Statistical and dynamical aspects of extreme heatwaves in the mid-latitudes

Robin Noyelle

université PARIS-SACLAY

Statistical and dynamical aspects of extreme heatwaves in the mid-latitudes

Aspects statistiques et dynamiques des vagues de chaleur extrêmes dans les moyennes latitudes

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°129, Sciences de l'Environnement d'Île de France (SEIF)
Spécialité de doctorat : Géosciences
Graduate School : Géosciences, Climat, Environnement, Planètes
Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **Laboratoire des sciences du climat et de l'environnement (Université Paris-Saclay, CNRS, CEA, UVSQ)**, sous la direction de **Pascal YIOU**, directeur de recherche CEA, et la co-direction de **Davide FARANDA**, directeur de recherche CNRS.

Thèse soutenue à Paris-Saclay, le 18 juin 2024, par

Robin NOYELLE

Composition du jury

Membres du jury avec voix délibérative

Pascale BRACONNOT Directrice de Recherche, CEA, Université Paris-Saclay	Présidente
Dorian S. ABBOTT Professeur associé, Université de Chicago	Rapporteur & Examinateur
Olivia ROMPPAINEN-MARTIUS Professeure, Université de Berne	Rapportrice & Examinatrice
Erika COPPOLA Directrice de recherche, International Centre for Theoretical Physics	Examinatrice
Erich FISCHER Professeur, ETH Zürich	Examinateur

THESE DE DOCTORAT

NNT : 2024UPASJ013

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences de l'Environnement d'Île-de-France (SEIF)

Titre: Aspects statistiques et dynamiques des vagues de chaleur extrêmes dans les moyennes latitudes

Mots clés: extrêmes climatiques, vagues de chaleur, algorithmes d'événements rares, théorie des valeurs extrêmes, dynamique atmosphérique

Résumé:

L'intensité et la fréquence des vagues de chaleur augmentent en réponse au réchauffement climatique d'origine anthropique. Cette thèse étudie les aspects statistiques et dynamiques des événements de chaleur extrêmes et très extrêmes dans les moyennes latitudes, en particulier en Europe. Elle aborde les questions des températures maximales de surface qui peuvent être atteintes pendant une vague de chaleur, la différence entre les mécanismes physiques conduisant à des vagues de chaleur extrêmes et des vagues de chaleur très extrêmes, la possibilité de simuler efficacement des vagues de chaleur très extrêmes dans un modèle climatique et l'évolution dynamique des vagues de chaleur extrêmes avec le réchauffement climatique.

La première partie de la thèse étudie les aspects statistiques des vagues de chaleur extrêmes. Elle aborde la question de la limite supérieure des températures de surface. L'approche est basée sur la théorie des valeurs extrêmes (TVE) et compare les résultats cette méthode aux processus physiques qui limitent fondamentalement l'augmentation des températures. Des lacunes de l'approche TVE traditionnelle sont démontrées et je propose une approche pour atténuer ces dernières en contraignant physiquement le fit des distributions de probabilité basées sur la TVE.

La deuxième partie de la thèse porte sur la question des mécanismes dynamiques à travers lesquels le système climatique s'organise pour produire des événements de chaleur intenses. Je montre tout d'abord dans un long run de contrôle d'un modèle climatique que les événements de chaleur extrêmes ont tendance à être typiques, c'est-à-dire à être plus semblables les uns aux autres que les événements de chaleur modérés. L'étude des extrêmes étant limitée par un fort problème de sous-échantillonnage, je détaille ensuite l'intérêt d'utiliser des algorithmes dits d'événements rares qui permettent d'échantillonner plus d'extrêmes que les simulations habituelles ne peuvent en fournir. J'applique un tel algorithme d'événements rares dans le modèle IPSL-CM6A-LR pour échantillonner des étés extrêmes et très extrêmes en Europe occidentale dans des conditions préindustrielles, actuelles et futures de forçages anthropiques. J'étudie en particulier les changements dans la dynamique conduisant à ces étés extrêmes au cours des différentes périodes. Je montre que, dans le modèle, le réchauffement climatique est associé à une diminution de la variabilité de la circulation atmosphérique mais à une augmentation de la variabilité thermodynamique.

Le travail présenté dans cette thèse démontre l'intérêt de faire le lien entre les approches physiques et statistiques pour l'étude des événements climatologiques extrêmes et très extrêmes. Je montre en particulier que l'utilisation de techniques comme les algorithmes d'événements rares permet de répondre à des questions physiques sur le système climatique qui sont hors de portée des méthodes classiques.

Title: Statistical and dynamical aspects of extreme heatwaves in the mid-latitudes

Keywords: climate extremes, heatwaves, rare events algorithms, extreme value theory, atmospheric dynamics

Abstract:

Heatwaves are increasing both in frequency and intensity as a result of anthropogenic global warming. This PhD studies statistical and dynamical aspects of extreme and very extreme heat events in the mid-latitudes with a particular focus on European heatwaves. It addresses the questions of the maximal near-surface air temperatures that can be reached during a heatwave event, the difference between the physical mechanisms leading to extreme vs very extreme heatwaves, the possibility to simulate efficiently very extreme heatwaves in a climate model and the dynamical evolution of extreme heatwaves with global warming.

The first part of the PhD investigates statistical aspects of extreme heatwaves. It addresses the question of the upper bound for near-surface air temperatures. The approach is based on Extreme Value Theory (EVT) and I compare the results of this method to the physical processes that fundamentally limit the increase of air surface temperatures. The shortcomings of the traditional EVT approach are demonstrated and I propose an approach to alleviate the latter by physically constraining the fit of the EVT-based probability distributions.

The second part of the PhD addresses the question of the dynamical mechanisms by which the climate system organizes to produce intense heat events. I first show in a long control run of a climate model that extreme heat events tend to be typical, i.e. to be more similar to each other than moderate heat events. Because the study of extremes is impaired by a strong under-sampling problem, I then detail the interest of using so-called rare events algorithms which allow to sample more extremes than regular simulations can provide. I apply such a rare events algorithm in the IPSL-CM6A-LR model to sample extreme and very extreme hot summers in Western Europe under pre-industrial, present and future conditions of anthropogenic forcings. In particular I investigate changes in the dynamics leading to these extreme summers in the different periods. I show that, in the model, global warming is associated to a decrease of the variability of the atmospheric circulation but to an increase of the thermodynamic variability.

The work presented in this thesis demonstrate the interest of bridging the gap between physical and statistical approaches for the study of extreme and very extreme climate events. I show in particular that using techniques like rare events algorithms allows to answer physical questions about the climate system that are out of reach for classical methods.

Résumé long

Il n'y a aujourd'hui aucun doute pour la communauté scientifique que la température globale de la Terre augmente en réponse aux émissions anthropogéniques de gaz à effet de serre et aux changements dans l'utilisation des sols. L'une des conséquences les plus marquantes de ce réchauffement est l'augmentation de la fréquence et de l'intensité des vagues de chaleur. Les surfaces continentales des latitudes moyennes boréales - qui englobent l'Europe, l'Amérique du Nord et la partie septentrionale du continent eurasien - se réchauffent en particulier plus rapidement que le reste du monde, et les extrêmes dans certaines de ces régions se réchauffent encore plus vite que la température moyenne. Au cours des 20 dernières années, des événements d'une intensité extrême et sans précédent ont été observés aux latitudes moyennes, notamment en 2003 et 2019 en Europe occidentale, en 2010 en Russie et en 2021 dans l'ouest du Canada. Ce dernier événement a surpris même la communauté scientifique travaillant sur les vagues de chaleur, avec des températures de surface maximales atteignant 49,6 °C à Lytton, en Colombie-Britannique (50°N), battant les précédents records de 5 °C.

Historiquement, les sciences du climat se sont surtout intéressées aux propriétés moyennes à long terme du système climatique. L'étude des moments d'ordre supérieur des distributions statistiques des variables atmosphériques et océaniques, et en particulier des extrêmes, est plus récente et a pris de l'importance en raison de la quantité croissante de données disponibles et des impacts observés des extrêmes sur les sociétés et les écosystèmes. Les extrêmes constituent un élément essentiel des propriétés statistiques du système climatique et, malgré leur rareté, peuvent avoir des conséquences considérables sur les multiples composantes du système climatique.

Dans cette thèse, je m'intéresse aux propriétés statistiques et dynamiques des événements de chaleur extrême et très extrême dans les moyennes latitudes. J'aborderai les questions suivantes :

1. **Quelles sont les températures maximales qui peuvent être atteintes pendant une vague de chaleur ?**

2. **Les mécanismes physiques qui conduisent aux événements de chaleur très extrême sont-ils similaires à ceux qui conduisent aux événements de chaleur extrême ?**

3. **Est-il possible de simuler efficacement des vagues de chaleur extrême à l'aide de modèles climatiques ?**

4. **Comment la dynamique des vagues de chaleur extrêmes évoluera-t-elle avec le changement climatique ?**

La thèse commence (chapitre 1) par une introduction générale sur la définition des événements extrêmes dans un contexte climatologique et sur le problème fondamental de sous-échantillonnage qui empêche l'étude de ces événements avec les méthodes habituelles employées en climatologie. Je détaille ensuite les mécanismes physiques connus qui conduisent à des événements de chaleur extrêmes dans le système climatique, en mettant particulièrement l'accent sur les vagues de chaleur des moyennes latitudes et de l'Europe, et je souligne les incertitudes qui subsistent à propos de ces événements. Enfin, je passe en revue l'évolution dynamique attendue des vagues de chaleur avec le réchauffement climatique. Le reste de la thèse est organisé en deux parties qui étudient respectivement les aspects statistiques et dynamiques des vagues de chaleur.

Dans la première partie (chapitres 2 et 3), j'étudie la question des températures maximales près de la surface qui peuvent être atteintes pendant une vague de chaleur. L'approche est principalement statistique, utilisant la théorie des valeurs extrêmes (EVT), et compare ses résultats aux processus physiques qui limitent fondamentalement l'augmentation de la température de surface de l'air (chapitre 2). En particulier, j'étudie les limites des estimations statistiques basées sur la théorie des valeurs extrêmes et je propose une approche pour atténuer leurs défauts en contraignant physiquement l'ajustement des distributions de probabilité (chapitre 3).

Dans la deuxième partie (chapitre 4, chapitre 5, chapitre 6 et chapitre 7), je m'intéresse aux mécanismes physiques dynamiques qui permettent au système climatique de produire des événements de chaleur extrêmes et très extrêmes. Comme l'étude des extrêmes est entravée par un fort problème de sous-échantillonnage, j'utilise des algorithmes dits d'événements rares qui permettent d'échantillonner plus d'extrêmes que les simulations régulières ne peuvent en fournir. Le chapitre 4 propose un examen de ces méthodes. Je détaille les principaux algorithmes utilisés dans des contextes géophysiques, j'étudie les principaux avantages et inconvénients de ces méthodes et je propose des solutions pour atténuer ces derniers. Le chapitre 5 étudie la dynamique typique conduisant à des extrêmes dans une simulation de contrôle du modèle climatique de l'IPSL (IPSL-CM6A-LR) dans une configuration préindustrielle – c'est-à-dire la dynamique qui est censée être échantillonnée par les algorithmes d'événements rares. Je propose l'idée que la dynamique typique est le concept pertinent pour étudier les mécanismes dynamiques conduisant aux extrêmes dans un contexte climatologique plutôt que météorologique. Dans les chapitres 6 et 7, j'applique un algorithme d'événements rares dans le même modèle pour échantillonner les étés extrêmement chauds en Europe occidentale dans le cadre de scénarios pré-industriels, actuels et futurs de forçages anthropiques. En particulier, j'étudie comment la dynamique menant à ces étés extrêmes change entre ces différentes périodes.

Cette thèse a été financée par le Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA).

“Beauty, I hear you ask, do not the Graces flee where integrals stretch forth their necks?”

– Ludwig Boltzmann

Remerciements

Même si je suis le seul auteur de cette thèse – et que j'en assume donc les erreurs et inexactitudes – le travail présenté ici n'aurait pas pu être réalisé sans l'aide et le soutien d'un grand nombre de personnes. Un travail scientifique est toujours le résultat d'un effort collectif et je souhaite ici remercier celles et ceux grâce à qui cette thèse a pu être menée à bien.

Je remercie d'abord les membres de mon jury, qui ont accepté de rapporter et d'examiner mon travail : Dorian Abbott et Olivia Martius-Romppainen en tant que rapporteurs, Erika Coppola et Erich Fischer en tant qu'examinateurs et Pascale Braconnot en tant qu'examinatrice et présidente du jury.

J'ai une pensée émue pour l'ensemble des membres du LSCE avec qui j'ai interagi pendant ces trois années et qui font de ce laboratoire un environnement de travail exceptionnel : les permanents de l'équipe ESTIMR et du thème 3 en général, ma marraine de thèse Isabelle Pison, Nikki Vercauteren et Eric Simonnet pour leur suivi, et Arnaud Caubel et Yann Meurdesoif pour leur soutien indispensable pour les simulations du modèle de l'IPSL. J'ai une pensée particulière pour les stagiaires, doctorants et post-doctorants avec qui j'ai passé ma thèse, qui m'ont soutenu et aidé et qui ont illuminé mes pauses déjeuner et certaines de mes soirées : Florentin, Bastien, Meriem, Stella, Nicolas, Julien, Lucas, Emma, Joséphine, Clara, Mireia, Laura, Germain, Clément, Flavio, Pradeebane, Camille, Phillipe, Simon, Ludyvine, Emeric, Yann, Xavier, Akim, Vivien, Valentin, Paula, Léna, Yann, Martin, Adrien, Burak, George, Lua, Nemo, Yoann, Grégoire, Laurie, Greta, Clément...

Je souhaite aussi remercier du fond du cœur Pascal Yiou et Davide Faranda, mes deux directeurs, pour avoir accepté d'encadrer ma thèse, m'avoir fait confiance au cours de ces trois années et avoir toujours mis en avant mon travail. Ce fut un privilège de travailler à leurs côtés.

J'ai interagi avec tellement de scientifiques exceptionnels qui ont nourri ma réflexion depuis le début de cette thèse qu'il m'est difficile de tous les nommer ici. J'ai ajouté des notes de bas de pages dans les chapitres de cette thèse pour remercier celles et ceux envers qui j'ai une dette particulière. Je souhaitais ici remercier en particulier mes co-auteurs : Yi Zhang, Philippe Naveau, Yoann Robin, Gianmarco Mengaldo, Chenyu Dong, Bénédicte Colnet.

A l'issue de ma soutenance j'ai exprimé publiquement ma profonde gratitude envers l'école de la République, qui m'a porté là où je suis aujourd'hui et grâce à qui j'ai pu faire des études de grande qualité. Je renouvelle ici cette gratitude.

Enfin, et peut-être surtout, je remercie mes proches pour leur soutien inébran-

lable. Mes amis et amies : Guillaume, Anne-Sophie, Bénédicte, Pierre, Félix, Corentin, Anatole, Paul, Lucas, Rayan, Elise, Samuel... Inès, pour tout. Ma famille, et en particulier mes parents et ma soeur qui, sans toujours comprendre mes choix, se sont faits un devoir de me soutenir en toutes circonstances.

Contents

Introduction

There is nowadays no doubt in the scientific community that the global temperature of the Earth is increasing as a result of anthropogenic emissions of greenhouse gases and land use changes. One of the most global and consequential result of this warming is the increasing frequency and intensity of heatwave events. Mid-latitudinal boreal land surfaces — encompassing Europe, North America and the northern part of the Eurasian continent — are in particular warming faster than the rest of the world and the extremes in some of these regions are warming even faster than the mean temperature. In the last 20 years, events of extreme and unprecedented intensities have been observed in the mid-latitudes, among them 2003 and 2019 in Western Europe, 2010 in Russia and 2021 in Western Canada. The latter has startled even the scientific community working on heatwaves, with maximum surface temperatures up to 49.6°C in Lytton, British Columbia (50°N), breaking previous records by up to 5°C.

Historically, climate science has been mostly interested in long term mean properties of the climate system. The study of higher order moments of the statistical distributions of atmospheric and oceanic variables, and in particular extremes, is more recent and took importance as a result of both an increasing amount of data and the large observed impacts of extremes on societies and ecosystems. Extremes are indeed a fundamental part of the statistical properties of the climate system and, despite their rarity, can have tremendous consequences on the multiple components of the climate system.

In this PhD, I am interested in the **statistical and dynamical properties of extreme and very extreme heat events in the mid-latitudes**. I will address the following questions:

1. **What is the maximal near-surface air temperature that can be reached by a heatwave event?**

2. **Are the physical mechanisms leading to *very* extreme heat events similar to those leading to extreme ones?**

3. **Is it possible to simulate efficiently extreme heatwaves with climate models?**

4. **How will the dynamics leading to extreme heatwaves change with climate change?**

The thesis begins (**chapter 1**) with a general introduction on the definition of extreme events in a climatological context and the fundamental under-sampling issue preventing the study of these events with usual methods employed in climatology. I then detail the known physical mechanisms leading to extreme heat events in the climate system with a particular focus on mid-latitude and European heatwaves and emphasize what is still uncertain about these events. Finally, I review the expected dynamical evolution of heatwaves with global warming. The rest of the thesis is organized in two parts investigating respectively the statistical and dynamical aspects of heatwaves.

In the first part (**chapter 2 and chapter 3**) I investigate the question of the maximal near-surface air temperature that can be reached during a heatwave event. The approach is mainly statistical, making use of Extreme Value Theory (EVT), and compare its results to the physical processes that fundamentally limit the increase of air surface temperature (**chapter 2**). In particular, I investigate the discrepancies of the statistical estimates based on EVT and propose an approach to alleviate their shortcomings by physically constraining the fit of the probability distributions (**chapter 3**).

In the second part (**chapter 4, chapter 5, chapter 6 and chapter 7**), I am more interested in the dynamical physical mechanisms which allow the climate system to produce extreme and *very* extreme heat events. Because the study of extremes is impaired by a strong under-sampling problem, I use so-called rare events algorithms which allow to sample more extremes than regular simulations can provide. **Chapter 4** proposes a review of these methods. I detail the main algorithms used in geophysical contexts, investigate the main advantages and drawbacks of these methods and propose some solutions to alleviate the latter. **Chapter 5** investigates the *typical* dynamics leading to extremes in a control run of the IPSL-CM6A-LR model in a pre-industrial configuration — which is the dynamics that is expected to be sampled by rare events algorithms. I propose the idea that the typical dynamics is the relevant concept for studying the dynamical mechanisms leading to extremes in a climatological rather than meteorological context. In **chapter 6 and chapter 7**, I apply a rare events algorithm in the IPSL-CM6A-LR model to sample extreme and very extreme hot summers in Western Europe under pre-industrial, present and future scenarios of anthropogenic forcings. In particular I investigate how the dynamics leading to these extreme summers changes between these different periods.

Chapter 1

Extreme heat events in the climate system

This introductory chapter is organized as follows. In section 1.1, I recall key defining elements on the climate system and its statistical properties. In section 1.2, I discuss the definition of extreme events in the climate system and the associated under-sampling issue hampering their study. Finally, in section 1.3, I review the literature on the dynamical mechanisms associated to heatwaves and how they may change in the future under global warming.

1.1 Definition elements

1.1.1 The climate system and its description

The last IPCC report glossary provides the following definition for the **climate system** (IPCC 2023): "The global system consisting of five major components: the atmosphere, the hydrosphere, the cryosphere, the lithosphere and the biosphere and the interactions between them.". In other words, the climate system consists of all components of the Earth and how they co-evolve at all time scales. As such, the climate system may be one of the most complex systems subject to scientific study, involving a wide range of disciplines such as physics, chemistry, meteorology, oceanography, biology etc.

The main physical properties of the climate system — especially with regards to the study of extreme events — can be summarized by a few ideas. Ghil and Lucarini (2020) begin their recent review of the physics of the climate system with a statement on its key properties: the climate system is "forced, dissipative, chaotic and out of equilibrium". The forced, dissipative and out of equilibrium properties are a result of the differences in the absorption of solar radiation both longitudinally, latitudinally and vertically. This results into compensating vertical and horizontal energy fluxes which give rise to the large scale circulation motions of the oceans and the atmosphere (Schneider 2006; Wunsch and Ferrari 2004), and more generally drives forced variability at long time scales. This has lead to a description of the climate system as a thermal engine transforming (radiative) heat from the Sun

into mechanical energy (Lorenz 1955; Lorenz 1967; Lucarini et al. 2014). Although average large scale circulations exist, due to internal instabilities and non-linearities, the climate system is also chaotic at a broad range of time scales resulting into internal (or natural) variability (Ghil 2001).

The evolution of the components of the climate system (e.g. atmosphere, oceans, land surfaces, ice masses) is mainly described using the continuum approximation to model these components as mathematical fields evolving in time according to partial differential equations (PDEs). The equations describing their evolution express in mathematical form the physical principles of the conservation of mass — including phase changes —, 3D momentum and energy, with a general state equation for fluid components (Peixoto and Oort 1992; Ghil and Lucarini 2020; Vallis 2017) [1]. Nonetheless, the source and sink terms in these equations — especially those who describe the interactions of two or more sub-systems, such as between the atmosphere and the oceans, or between liquid an vapor phases of water in clouds — are usually not known at time and spatial scales relevant for their parametrization in the evolution equations of the large scale circulation (a few dozens of kilometers and a few hours), which renders the mathematical description of the climate system fundamentally uncertain.

The use of a particular form of PDEs is a first approximation to describe the evolution of the climate system. Their analytic resolution is impossible in general and the scientific community nowadays mostly relies on the numerical integration of a spatially discretized form of these equations, which constitutes a second fundamental approximation. As a consequence, the temporal evolution of the climate system can be described very generally in the framework of dynamical systems:

$$\frac{dX}{dt} = b(X(t), t, \theta(t)) \tag{1.1}$$

where $X(t) \in \Omega \subset \mathbb{R}^d$ is the state vector of the system, encompassing all the physical elements describing the climate system at time t (especially the discretized version of the fields describing the atmosphere and the oceans), and b is the evolution equation which allows to compute the future evolution of the system, i.e. a climate model. It is convenient to note very generally $\theta(t)$ an ensemble of forcing parameters that may or may not depend on time. Whether these parameters are considered internal or external may be the subject of debate — for example the increase of greenhouse gases (GHG) due to human activities. The dimension d of the phase space $\Omega \subset \mathbb{R}^d$ of the full system is usually very high in modern climate models, probably of the order of 10^8 or higher even for so-called low-resolution model[2]. This spatially-discrete but time-continuous version of the evolution equations has also to be rendered time-discrete for practical numerical integration.

[1] For the atmosphere and the oceans, these equations are a special form of the Navier-Stokes equations. [2] The IPSL-CM6A-LR model (Boucher et al. 2020) has for example a spatial resolution of $142 \times 144 \times 79 \simeq 1.5 \times 10^6$ grid points for the atmosphere. This resolution has to be multiplied by the number of resolved 3D variables, and one has to add the number of surface 2D variables explicitly resolved by the model.

1.1.2 Mathematical definition of the climate

a) Ergodic theory and dynamical systems analysis

Although the definition of the climate system contains the word "climate", what comes first in the definition of this object of study is the "climate system" itself. The possibility to define precisely the notion of **climate** is a consequence of the bridge that was made during the 20th century between ergodic theory and dynamical systems analysis (Tantet et al. 2016).

The development of equilibrium statistical physics at the end of the 19th and the beginning of the 20th century (Maxwell 1860; Gibbs 1902; Boltzmann 1909) was based on the problem that there are too many particles to follow inside a gas — and therefore one cannot describe analytically their evolution. However, it is still possible to describe the macroscopic properties of the gas by assuming the existence of a probability distribution μ invariant under the dynamics of the system. This distribution is nowadays called the Boltzmann-Gibbs distribution and it depends on the energy and the temperature of the system (Landau and Lifshitz 1969). The relevance of this distribution to compute macroscopic quantities relies on an ergodic hypothesis that was later proved by Birkhoff (Birkhoff 1927; Birkhoff 1931). This hypothesis states that whatever the initial starting state X_0 of the system, the temporal average of a (macroscopic) quantity of interest $F : \mathbb{R}^d \to \mathbb{R}$ (called an **observable**) will be equal to its phase space average:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T F(X(t))dt = \int_\Omega F(X)\mu(dX) \tag{1.2}$$

where Ω is the phase space of the system and $X(t = 0) = X_0$. In other words, the Birkhoff theorem proves that the computation of temporal averages is equivalent to distribution averages of the system and therefore gives a practical way to compute the latter.

On the other hand, in dynamical system analysis, the pioneering work of Poincaré (Poincaré 1881) showed the existence of unstable (i.e. chaotic) solutions to differential equations. The study of chaos was subsequently mainly driven by mathematicians (Lyapunov 1892; Perron 1929; Kolmogorov 1954; Livi et al. 2003) until the work of Lorenz (1963) showed the relevance of chaos for forced-dissipative physical systems with the paradigmatic example of Rayleigh-Bénard convection[3] and thanks to the first developments of numerical integration. Lorenz importantly showed that such systems exhibit somewhat paradoxical properties: (i) the existence of aperiodic but bounded solutions with strong sensitivity to initial conditions and (ii) the convergence of these solutions towards a complicated topological object in the phase space (later called a **strange attractor**) which is neither a fixed point nor a periodic orbit. In other words, Lorenz both showed that there are inherent limitations to meteorological forecasts due to the chaotic nature of the atmosphere, but that at longer time scales the system still exhibits a form of order (cf. Fig. 1.1). The work of Anosov (1962) later showed more generally that if a system is hyperbolic

[3] Non-rotating plane setting in which the fluid is in contact with two reservoirs at different temperatures.

(i.e. sufficiently chaotic, Araújo and Viana (2009)), statistical properties are robust even though individual trajectories are sensitive to perturbations.

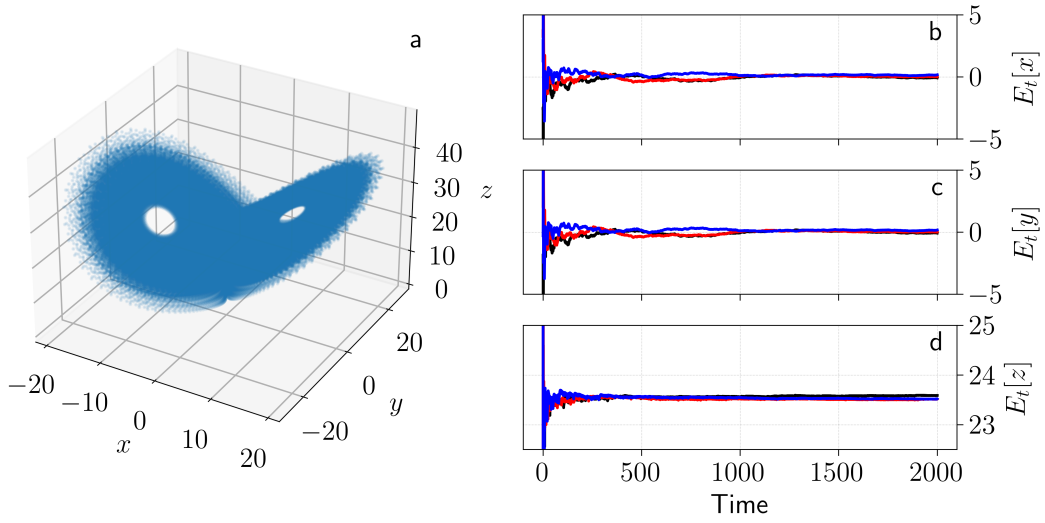


Figure 1.1: **Lorenz attractor and convergence towards the stationary probability distribution.** (a) Lorenz (1963) attractor, and empirical average for increasing time T in x-axis: (b) $E_T[x] = \frac{1}{T}\int_0^T x(t)dt$, (c) $E_T[y] = \frac{1}{T}\int_0^T y(t)dt$ and $E_T[z] = \frac{1}{T}\int_0^T z(t)dt$. The different colors show the convergence to a climatological average of trajectories with different starting conditions.

This paved the way to interpret dynamical systems in topological terms, i.e. with respect to the geometrical properties of their (strange) attractors and made a formal link with their statistical properties (Eckmann and Ruelle 1985). Ruelle and Takens (1971) demonstrated that strange attractors emerge in forced-dissipative systems far from equilibrium. The link between the dynamics of hyperbolic systems (Smale 1967) and their statistical properties was formalized by Sinai (1972) and Bowen and Ruelle (1975). Even though forced-dissipative systems are out-of-equilibrium — so that classical equilibirum statistical mechanics does not apply — a unique invariant probability distribution with physical relevance still exists: the so-called Sinai-Ruelle-Bowen (SRB) probability distribution (also called *measure*, Ruelle (2009)). This probability distribution corresponds to what is sometimes referred to as a Non-Equilibrium Steady State (NESS) (Gallavotti and Ruelle 1997). Importantly, the SRB probability distribution is invariant under the dynamics of the system: the probability of a set on the attractor is constant even if the set is deformed by the integration of the system.

The correspondence between time and phase-space averages — which is crucial to define meaningful macroscopic properties of a complex dynamical system — that was present in equilibrium statistical mechanics could now be expanded to out-of-equilibrium systems. The SRB probability distribution plays the role of the Gibbs probability distribution in equilibrium thermodynamics but has different mathematical properties. Because of the presence of forcing and dissipation, the attractor is

strange: its mathematical dimension is smaller than the dimension of the phase space and even non-integer. The attractor is not (even locally) a smooth manifold, so that one *cannot* write:

$$\mu(dx) = \mu(x)dx. \tag{1.3}$$

The attractor can be better conceptualized as the Cartesian product of a smooth manifold — corresponding to unstable (i.e. chaotic) directions of the flow — and a fractal set — corresponding to contracting directions of the flow arising as a result of dissipation (Lucarini et al. 2014).

The existence of SRB probability distributions for an important class of chaotic systems provides a rigorous framework for the study of non-equilibrium physical systems (Ruelle 1999; Gallavotti 2014). Although their properties are not intuitive, the existence of these probability distributions allows to compute phase-space averages: they give the weight(s) to be used to compute expectations. Again, thanks to the ergodicity hypothesis, one can compute the phase-space average of quantities of interest simply as the limit of time averages.

b) Definition of the climate in a stationary case

The last IPCC report glossary gives the following definition of the **climate** (IPCC 2023): "Climate in a narrow sense is usually defined as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period for averaging these variables is 30 years, as defined by the World Meteorological Organization (WMO). The relevant quantities are most often near-surface variables such as temperature, precipitation, and wind. Climate in a wider sense is the state, including a statistical description, of the climate system.". In this section, I give a precise mathematical sense to this classical definition of the climate that one can summarize as "the ensemble of statistical properties of the climate system".

The developments on ergodic theory and dynamical systems in the previous section allows to give a mathematically precise definition of the concept of climate as soon as one can justify that the mathematical framework presented can reasonably fit with the climate system. The relevance of the mathematical category of hyperbolic systems for the study of more general chaotic systems with a large number of degrees of freedom is based on an extension of the ergodic hypothesis: the chaotic hypothesis proposed by Gallavotti and Cohen (1995). This hypothesis states that, very generally, a system exhibiting chaotic motions should be considered as hyperbolic, and therefore one can assume the existence of an attractor and an SRB probability distribution for this system. Because of its chaoticity — especially for the atmosphere as experienced every day by meteorological operational forecasters (Bauer et al. 2015) — the climate system constitutes a good candidate to apply the chaotic hypothesis. As soon as one mentions climatological quantities, this is implicitly the hypothesis that is made and I will follow this practice here.

In a stationary context, **the climate is therefore the SRB probability distribution μ_θ associated with the climate system seen as a dynamical**

system. The stationary hypothesis implies in the notation used here that the external parameters θ do not depend on time and that the evolution function $b(X(t), t, \theta)$ does not depend explicitly on t (or its dependence is periodic). In other words, the climate is both the very complex attractor on which the trajectories of the climate system are converging (i.e. a topological object in the phase space) and the associated probability distribution that one can define on this object and which renders the computation of phase-space averages possible. The distribution of states inside the attractor is what is usually referred to as internal variability in climate science. The SRB probability distribution μ_θ therefore defines what states in the phase space can possibly be reached by the dynamics of the system.

Because of the high dimensionality of the climate system and therefore of its (strange) attractor, any attempt to find directly the probability distribution μ_θ is doomed to fail. One has to rely on the ergodic hypothesis to define climatological properties: phase-space averages using the probability distribution μ_θ are equivalent to long time averages whatever the starting point of the dynamics:

$$\mathbb{E}_{\mu_\theta}[F] = \lim_{T\to\infty} \frac{1}{T} \int_0^T F(X(t))dt = \int_\Omega F(X)\mu_\theta(dX). \tag{1.4}$$

This justifies the WMO definition of the climate, as soon as the averaging period is long enough with respect to the typical time scales of the quantities averaged. For components with a longer time scale than the atmosphere (for instance the oceans), the somewhat arbitrary 30 years averaging period can clearly not be a suitable approximation of the infinite time limit.

With this framework, the study of climate change is equivalent to finding how the climatological probability distribution μ_θ changes when the forcing parameters θ change. One would for example be interested in the change of climatological quantities:

$$\Delta F = \mathbb{E}_{\mu_{\theta+d\theta}}[F] - \mathbb{E}_{\mu_\theta}[F] \tag{1.5}$$

where both expectations on the right hand side are defined according to their respective probability distribution for the forcing parameters θ and $\theta + d\theta$. As a consequence climatology as an object of study is well defined and investigating how the climate will be in 100 years is meaningful — because it is a statistical and therefore boundary conditions problem — even though the state of the climate system cannot be predicted in the next 10 days — because it is an initial value problem in a chaotic system.

c) Definition of the climate in a non-stationary case

As explained above, the relevance of the definition of climate as the ensemble of statistical properties of the climate system crucially depends on the ergodic hypothesis to identify temporal and phase-space averages. Once this hypothesis is broken — and one could argue that it has always been the case throughout Earth's history — the use of temporal averages to define the climate bears little meaning (see Arguez and Vose (2011) for such limits of the 30 years average definition of the WMO).

In the context of climate change, there is anyway no doubt that the stationary approximation used to define the concept of climate is not valid anymore: increasing concentrations of GHGs in the atmosphere caused by human activities are equivalent to imposing a time-varying forcing to the climate system.

It is nonetheless still possible to give a sense to the notion of climate in a non-stationary case, i.e. for non-autonomous dynamical systems. Since the 1990s, mathematicians have clarified the understanding of non-autonomous dynamics and the concept of **snapshot attractors** has emerged (Crauel and Flandoli 1994; Arnold et al. 1995; Kloeden and Rasmussen 2011; Carvalho et al. 2012; Kuehn et al. 2015). In the field of climate science, Michael Ghil used the term **pullback attractors** to define a similar object (Ghil et al. 2008; Chekroun et al. 2011; Bódai et al. 2011; Bódai and Tél 2012; Drótos et al. 2015).

The general idea of snapshot or pullback attractors is to still be able to define a probability distribution with which computing phase-space averages is meaningful, but this probability distribution will evolve in time. The reader is invited to refer to the cited literature for the mathematical details of the definition. Here I will simply state that the way to describe a probability distribution μ_t evolving with time is to consider an (infinite) ensemble of trajectories evolving according to the dynamics of the system which are initialized infinitely far in the past. The snapshot/pullback attractor, and the associated probability distribution μ_t, of a forced dynamical system can therefore be understood as the ensemble of states in which the system can be at time t. The expectation of any observable is then defined simply as:

$$\mathbb{E}_{\mu_t}[F] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} F(X^i(t)) = \int_{\Omega_t} F(X)\mu_t(dX) \tag{1.6}$$

where the $X^i(t)$ are trajectories initialized infinitely far in the past and Ω_t is the snapshot/pullback attractor at time t. Note that as a consequence, the expectation of the observable F is evolving in time.

In the context of climate science, this is equivalent to defining the climate as an *ensemble average*. This has lead some authors to use the term parallel climate realizations (Tél et al. 2020) to describe the use of ensembles of a model. The practice of the ensemble average to define the climate is also possible in the stationary case and has been mentioned to define the climate (Leith 1975; Peixóto and Oort 1984; Peixoto and Oort 1992). It however becomes an absolute necessity in a non-stationary context where the ergodic hypothesis is not valid anymore. The ensemble of parallel members of a climate model is therefore the generalization of the Gibbs probability distribution from equilibrium statistical physics for a non-equilibirum system whose parameters are drifting in time.

A radical interpretation of these mathematical elements could lead someone to think that it is impossible to define the climate in a non-stationary model without running several members. It is actually more a question of time scales of the physical phenomena studied: if the forcing changes slowly so that the transients to the attractor are irrelevant for the dynamics then it is valid to compute 30 years averages to define climatological quantities (or at least it is a good approximation).

If on the other hand the time scales of the physical phenomena studied are of the same order of magnitude as the time scales of the forcing, then the only way to define a meaningful notion of climate is through an ensemble average. It is likely that given the typical time scales of the climate system (especially for the oceans) and the speed with which the anthropogenic forcings are imposed nowadays, the climate system is clearly in this realm and ensembles are absolutely needed to evaluate climate change. Climate scientists have indeed increasingly recognized the necessity to use large ensembles to define future evolutions of the climate (see for example the introduction of Maher et al. (2019) for a review of large ensembles of climate models).

1.2 What is an extreme event?

Although the notion of climate can be described as the ensemble of statistical properties of the climate system, historically, for feasibility reasons, its vast and rich properties have often been reduced to the study of the first order moments of the statistical distribution: average of temperatures, precipitations, winds, positions of the jets etc. The large scale circulations mentioned above is an example of such first order moments. It is only in the recent decades that the interest of the climate community has shifted towards higher order moments and quantiles of the climate variables distributions, and in particular towards extreme events. The increasing interest in extremes is partly due to an increasing amount of data and partly to the reconnaissance that extreme events can have disproportionately more effects on the system they affect than their frequency of occurrence would suggest (Ghil et al. 2011). The last IPCC report for example includes a whole chapter focusing on the projected changes of extreme events under climate change (Seneviratne et al. 2021). The very definition of what constitutes an extreme event in the climate system is nevertheless still not fixed. In this section I give a brief review of the issues associated to the definition and the study of extreme events.

1.2.1 Definition issues

The last IPCC report glossary (IPCC 2023) defines an **extreme weather event** as: "an event that is rare at a particular place and time of year. Definitions of 'rare' vary, but an extreme weather event would normally be as rare as or rarer than the 10th or 90th percentile of a probability density function estimated from observations.". On the other hand an extreme climate event occurs "when a pattern of extreme weather persists for some time". In the following, I will follow the practice of the IPCC report which generally calls extreme weather and extreme climate events as **extreme climate events**, in the sense of extremes happening in the climate system, whatever the time scales considered.

This first definition therefore relies on the rarity of the event occurring by fixing a somewhat arbitrary threshold at the 90th (or 10th) percentile of the empirical probability distribution of an observable. Alternatives exist, such as using a fixed

threshold associated to impacts (for example the 35°C threshold for wet bulb temperature, Lu and Romps (2023)) or defining as extreme any event which has important socio-economic consequences. In a cross-disciplinary review, McPhillips et al. (2018) showed that that there is a wide variety of definitions used in the literature on extremes in geosciences and a lack of coherence in what is or is not considered as an extreme event, especially with regards to whether the impacts should be included in the definition of the event. The WMO advises to define extreme events with respect to their magnitude (how much they depart from a baseline or threshold), duration and geographical extent (WMO 2023). However, apart from a general advice to base the thresholds on local climatological conditions, the organisation does not give precise guidance for defining when an event is considered extreme or not.

These definition issues are not minor because they can impact what kind of scientific conclusions one can draw. The use of a relative threshold with respect to climatology is probably the most employed metric in climate science, but the value of the quantiles considered can vary and this can have scientific consequences. For example when considering the 90th quantile as an extreme threshold for high temperature events, does one consider the quantile of the yearly distribution of temperature? This implies that on average 36-37 days per year would be considered extreme, very likely in summer due to the seasonal cycle, which means that around 1/3 of days in summer would be considered as extreme. But one may also consider the 90th quantile of the daily distribution of temperature, which implies that around 10 days per summer would be considered as extreme. On the other side of the extremeness spectrum, one may consider that an extreme event is an event occurring only every 10 years, which makes around 0.1 days per summer extreme. All these events are coherent with the threshold based definition but they will be associated with different degrees of challenges in their study and probably also to different physical conclusions. For example, it is mathematically possible to imagine that global warming would be associated to a change in the shape of the probability distribution of temperatures in summer so that the 90th percentile-like events of the past are *less* likely in the future, while making the 99th percentile-like events of the past *more* likely in the future. In such a case, should one conclude that extreme events have become more or less likely?

1.2.2 Extreme events in a climatological context

It is not possible to avoid the scientific issues associated to the very definition of extreme events: one will always be dependent on the arbitrariness of the threshold considered. To make this issue explicit and draw the link with what has been said above about the definition of climate, I use a general definition of an extreme event as a quantile of a certain order for a given observable:

Definition: a state X of the climate system is said to be **extreme at the α-th order with respect to the observable F** if

$$\mathbb{P}_\mu[F \geq F(X)] := \int_\Omega \mathbb{1}(F(Y) \geq F(X))\mu(dY) = 1 - \alpha \tag{1.7}$$

where the probability is taken with respect to the appropriate probability distribution μ defining the climate in the context considered. $\mathbb{1}$ is the indicator function: it is equal to 1 if $F(Y) \geq F(X)$ and 0 elsewhere. The value of $F(X)$ is then the **return level** for an extreme event of F with a **return period (or time)** of $1/(1-\alpha)$. In the following, I will distinguish between extreme and **very extreme events**. Although these denominations are somewhat arbitrary, I will consider that **very extreme events** are events of exceptional intensities, i.e. whose return time is equal or larger than 10 years.

This definition makes clear that being extreme is a property of a point (or a set of points) in the phase space of the climate system that depends on a given observable F and a relative threshold of the observable distribution. A state is therefore not extreme per se: it may be extreme with respect to one observable but not with respect to another. It also makes clear that being extreme is fundamentally a climatological statement: one needs to refer to the climatological probability distribution of the system considered.

Note that with this definition, the issue of defining a physically meaningful extreme stems from the definition of the observable F. For example the question of the duration of the event can be addressed by defining F with a temporal rolling mean:

$$F(X(t)) := \int_{u=-T}^{u=T} \tilde{F}(X(t+u))du \tag{1.8}$$

where $\tilde{F}$ is also an observable. For example, an extreme heat event can be defined as an event such that the 5-, 10- or 15-days rolling mean is above the 90th quantile of its climatological distribution.

Similarly, the properties of extremes of an observable $F_{\phi,\lambda}$ defined at the grid-point of latitude ϕ and longitude λ may defer from the properties of the extremes of the same observable at another grid point $F_{\phi',\lambda'}$. From a purely statistical point of view there is no reason to expect that the phenomenologies of the extremes are similar when one changes the location where the observable is computed: they are fundamentally different observables (e.g. Lucarini et al. (2023)).

Although coherent with the IPCC definition, this definition is not perfectly aligned with the practices of atmospheric scientists who would rather define extremes as a category of phenomenon with similar physical mechanisms. For example, heatwaves are considered as the same kind of events because they are associated to the same kind of atmospheric dynamics (see section 1.3) even though they correspond to extremes at different locations on Earth. The rationale to do so in atmospheric science is usually to increase the sample size of events and obtain statistically meaningful results, which clearly makes sense when few data is available. However, with the exponential increase in the quantity of data (especially from model outputs), it may be time to adopt a more restrictive definition of what is an extreme to make clear what is common and what is not among similar kind of events occurring at different places.

With this definition, for instance, extreme temperature events at different locations may not be considered as belonging to the same category because it would make more sense to define them with respect to a local temperature observable $T_{\phi,\lambda}$

(e.g. Stefanon et al. (2012) and Cattiaux and Ribes (2018)). This leads in fact to two scientific questions that are often mixed in the literature and that it may be worthwhile to separate: (i) what are the physical mechanisms to reach extremes of $T_{\phi,\lambda}$ and (ii) how those physical mechanisms change with ϕ and λ. In other words, it may be interesting to separate the threshold question (why is an extreme not like a regular event?) from the geographical question (is an extreme at two locations reached via the same physical mechanisms?).

1.2.3 The burden of rareness

Up to know I have considered statistical properties in a case where the system studied is sampled by an infinite number of trajectories. This is obviously never the case in practice, especially for a (very) high dimensional system such as the climate system for which we only have access to limited time series. I present here a discussion of the main challenge for the study of extreme (and rare) events: the lack of statistics.

Suppose one has an unfair coin and wishes to estimate the probability that the coin lands on heads. Suppose also that this coin is extremely unfair: its (unknown) probability of landing on heads is $p = 10^{-2}$. If one throws 100 times the coin, then the results may be 0, 1, 2, 3 heads, or, much less likely, more than 3 heads. The associated estimated probability may therefore be 0, 0.01, 0.02 and 0.03, which makes it erroneous by up to a factor 3. This intuitive example shows that one would need to throw a large number of times the coin to estimate precisely p: 100 times is far from enough and the estimation of the true probability is associated with a lot of uncertainty. This constitutes a general result for rare events: a brute force Monte Carlo simulation is inefficient and I now demonstrate that more rigorously.

Suppose one wants to estimate the probability p_A of a random process X_t reaching the set A using samples $(X_i)_{1\leq i\leq N}$ independent and identically distributed. These would typically correspond, for instance, to daily samplings from a climate model run. A represents an event of interest, for example extreme values of an observable F: $A = \{X \mid F(X) \geq a\}$ for a given threshold a. With a brute force Monte Carlo strategy, the naive estimator for the probability $p_A = \mathbb{P}[X_t \in A]$ is simply:

$$\hat{p}_A = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_A(X_i), \tag{1.9}$$

where $\mathbb{1}_A$ is the indicator function of the set A. In other words, one simply computes the empirical frequency with which the random process X_t reaches the set A. This estimator is unbiased ($\mathbb{E}[\hat{p}_A] = p_A$) and, thanks to the independence of each sample, its variance is:

$$\mathbb{V}[\hat{p}_A] = \frac{\mathbb{V}[\mathbb{1}_A]}{N} = \frac{p_A(1-p_A)}{N} \simeq \frac{p_A}{N} \tag{1.10}$$

using the fact that $p_A \ll 1$, i.e. assuming that the set A is rare. The relative error $\mathbb{RE}$ of this estimator is therefore:

$$\mathbb{RE} := \frac{\sqrt{\mathbb{V}[\hat{p}_A]}}{\mathbb{E}[\hat{p}_A]} = \frac{1}{\sqrt{p_A N}} \tag{1.11}$$

If one wants, for example, to estimate a return level for a return period of 1000 years, then $p_A \simeq 10^{-3}$ and to have a relative error of just 10% one needs $N \simeq 10^5$ which is a prohibitive amount of samples for any complex model, such as most climate models. When considering the expected value of any observable ψ in the rare set A estimated similarly by averaging $\psi(X)$ for events X that reach the set A:

$$\hat{\psi} = \frac{\sum_{i=1}^{N} \psi(X_i)\mathbb{1}(X_i \in A)}{\sum_{i=1}^{N} \mathbb{1}(X_i \in A)}, \tag{1.12}$$

the relative error of the estimation is also large for reasonable values of the standard deviation and the mean of ψ in set A (see section A.1.1 in annex for the derivation).

As a consequence, the study of extreme events is fundamentally limited by the lack of statistics and by the large variance associated to any estimator. This therefore renders the detailed study of their phenomenology difficult. Any procedure seeking to estimate quantities of interest for low probability events has to propose a way to reduce the variance of the naive estimator. The most obvious, but costly, approach is to **increase the sample size**: running a long simulation in a stationary context or a large ensemble in a non-stationary context (e.g. Suarez-Gutierrez et al. (2023), Bevacqua et al. (2023), and McHugh et al. (2023)) — see chapter 5. Another solution is to **make assumptions on the probability distribution of the extremes**, as for example done by Extreme Value Theory (EVT) (Coles 2001) — see chapter 2 and chapter 3. Similarly, but with different founding hypotheses, Large Deviation Theory (LDT) seeks to estimate the occurrence of large deviations of a sum of random variables from its mean via computation of the rate function from a limited sample (Gálfi et al. 2021).

However, neither EVT nor LDT allow one to simulate entire trajectories leading to extreme events. For the same reasons as presented above when estimating a very low probability, the system does not provide spontaneously numerous trajectories leading to rare events with a usual Monte Carlo simulation. The goal of a set of methods called **rare events algorithms** is to circumvent this issue by making the simulation of rare events of interest more likely by several orders of magnitude for a limited computational cost. Rare events algorithms may be seen as tools to explore the phase space of complex systems in regions of the attractor that are seldom visited, but still reachable. I present a review of the use of rare events algorithms in chapter 4 and I use a rare events algorithm to sample extreme summers in a climate model in chapters 6 and 7.

1.3 Extreme heat events in the climate system: definition, physical mechanisms and future evolution

In this section I present a review of the scientific literature on heatwaves with a special focus on mid-latitude and European heatwaves. A large part of the heatwave

literature has indeed focused on this region of the world (Perkins 2015), because of institutional, socio-economic but also scientific reasons. Europe has been a hot spot for extreme heat events in the last 30 years (Rousi et al. 2022; Faranda et al. 2023b): extremes have warmed faster than the mean (Huntingford et al. 2024) and faster than expected by most climate models (especially in Western Europe Patterson (2023) and Vautard et al. (2023)). Very impactful events have occurred in Europe in the last 20 years — like in 2003 in France (Stott et al. 2004; Black et al. 2004; García-Herrera et al. 2010) or 2010 in Russia (Dole et al. 2011; Otto et al. 2012; Trenberth and Fasullo 2012; Di Capua et al. 2021). Being situated at the exit of the North Atlantic storm track, Europe is a region where atmospheric wave breaking occurs and its meteorology is thus complex (Hoskins 1983). There is therefore a scientific challenge to better understand the dynamics of heatwaves in this region, explain why there are such discrepancies between models and observations and forecast the future evolution of heatwaves.

The interest of the meteorological and climatological communities in the dynamics of heatwaves is rather new. The lack of high-quality data and the limited size of data sets (cf. the discussion above on the burden of rareness) have been — and still are — major constraints for the scientific research on heatwaves. Most of the literature has been published after the record breaking 2003 summer in Western Europe, notably starting with the seminal paper of Stott et al. (2004) who proposed the first attribution statement of an extreme event. Numerous extreme and very extreme events have occurred since then, sparkling a strong interest in the statistics and dynamics of heatwaves. The current state of knowledge on heatwaves presented below is based on the reviews of Perkins (2015), Horton et al. (2016), and the more recent Domeisen et al. (2023) and Barriopedro et al. (2023).

1.3.1 What is a heatwave?

I begin the review with a definition of a heatwave event. In the following I use indifferently the denomination **heatwave** and **extreme heat event**. I will also use the denomination **extreme heatwaves** or **very extreme heat event** for heatwaves of exceptional intensity, i.e. with a return time equal or larger than 10 years. The last IPCC report glossary (IPCC 2023) defines a heatwave as "a period of abnormally hot weather, often defined with reference to a relative temperature threshold, lasting from two days to months". A heatwave is therefore a climatological extreme in the sense proposed above: it characterizes the state of the climate system and is defined with respect to the climatology of the near-surface temperature observable. A heatwave occurs when near-surface temperatures — typically 2-m air temperature — present large positive anomalies with respect to what should be expected at a certain location and period of the year.

Numerous metrics have been used in the literature to define and characterize heatwave events — not to say that each study redefines its own metrics —, with several authors advocating for a normalization of practices in order to obtain comparable results from one study to another (Russo et al. 2014; Perkins 2015; Perkins-Kirkpatrick and Lewis 2020; Barriopedro et al. 2023; Russo and Domeisen 2023).

Classically, the definition proposed is an event of at least n (with n typically between 2 to 6 days) consecutive days with temperatures exceeding a percentile α of the climatological distribution, typically the α = 90th or 95th percentile (Perkins and Alexander 2013) but with large data sets more stringent conditions like the 99th or 99.9th percentiles have been used. The choice of the threshold is usually also driven by the limited sampling and therefore the trade-off between having a large enough data set of extreme events and selecting events that are sufficiently extreme. However, the very estimation of these relative thresholds is not straightforward and may induce discontinuities in metrics at the beginning or end of the reference period (Diffenbaugh 2020; Brunner and Voigt 2024). This is especially the case in the presence of long-term trends such as global warming. Sippel et al. (2015) for example showed that normalizing temperature data relative to the local mean and variability of a reference period leads to the overestimation of probabilities of extremes. Whether authors consider daily maximum, mean or even minimum temperatures varies from one study to another. The daily maximum and mean temperatures are usually used for climate-oriented studies (Barriopedro et al. 2023) while minimum temperature is more relevant for impacts, especially on human health (Kovats and Hajat 2008). This classical definition encompasses an idea of intensity (above a high α percentile) and duration (n days) at the grid-cell level. Other definitions also take into account the spatial extent of the event (Perkins 2015).

The differences between studies usually do not arise in what constitutes a heatwave but in metrics to measure their characteristics. Metrics focusing on intensity, duration, frequency, timing and spatial extent have been used. Russo and Domeisen (2023) for example studied the differences in the observed temporal trends obtained for four intensity metrics and found vastly different results. Most of the discrepancies likely arise as a result of different needs by different scientific communities: the meteorological and physical climatology communities are interested in dynamical mechanisms leading to extremes, while the climatological impact community may be more interested in the impact-based metrics such as cumulative heat for example (Russo et al. 2015). Impact-based metrics notably include other quantities than temperature, such as humidity for heat stress and may use absolute rather than relative thresholds to define a heatwave — e.g. above 35°C rather than above the 90th percentile of the local temperature distribution (Barriopedro et al. 2023).

1.3.2 Dynamics of heatwave events

In this section I review the main physical mechanisms associated to the short term dynamics of heatwaves. I detail what are the known mechanisms by which the climate system organizes to produce extreme near-surface air temperatures at time scales of days to weeks. I therefore mainly detail the contributions of the atmosphere and land surfaces. Slower components of the climate system and external forcings (such as SSTs, aerosols, GHGs etc) are reviewed in the next sections. From now on, I will implicitly consider heatwaves in the mid-latitudes although most mechanisms also occur in other regions of the world, but with important regional variations. Recent studies on the polar and sub-polar regions have for example shown specific

dynamics — especially driven by advection through atmospheric rivers — that are seldom seen in the mid-latitudes (Hermann et al. 2020; Mattingly et al. 2023; Murto et al. 2023; Blanchard-Wrigglesworth et al. 2023; Wille et al. 2024a,b). I will also consider only summer heatwaves — therefore both relative and absolute temperature extremes — despite the fact that winter heatwaves are not as much studied but can also have strong impacts (Gloege et al. 2022).

It should be emphasized that a large part of what is known on the dynamics of heatwaves is based on the analysis of case studies of events that have occurred in the last 20 years in the mid-latitudes (e.g. Russo et al. (2015), Hauser et al. (2016), Hoy et al. (2017), Sánchez-Benítez et al. (2018), Wehrli et al. (2019, 2020), Sánchez-Benítez et al. (2022), Tuel et al. (2022), Serrano-Notivoli et al. (2023), Tripathy and Mishra (2023), and Rousi et al. (2023)), especially very intense events like 2003 in Western Europe or 2010 in Western Russia. Heatwaves rarely happen again at the exact same place and in the same configuration therefore it is usually assumed that knowledge about mechanisms occurring at one longitude/latitude can be transferred to another longitude/latitude in the mid-latitudes: the burden of rareness is overcome by extending the spatial range of study. In other words, there is a 'leap of faith' between numerous meteorological case studies and climatological properties of heatwaves in general. This may lead to conundrums about different dynamics leading to extreme temperatures that could actually be the result of either natural variability or regional variations in the dynamics of heatwaves.

a) Atmospheric dynamics

A heatwave is primarily an atmospheric-driven event (Horowitz et al. 2022): a specific atmospheric dynamics is a necessary condition for the appearance of a heatwave. The first studies investigating the atmospheric dynamics leading to extreme near-surface temperatures in the mid-latitudes early observed the co-occurrence of heatwaves with stationary anticyclonic structures disrupting the climatological westerlies. These anticyclones are usually embedded in a slow-moving or quasi-stationary larger scale wave structure, often called an atmospheric block (Xoplaki et al. 2003; Meehl and Tebaldi 2004; Stefanon et al. 2012; Pfahl and Wernli 2012; Schaller et al. 2018; Castañeda and Wang 2024), but it was later recognized that in more equatorward regions of the mid-latitudes — in the south of Europe in particular — heatwaves occur in conjunction with stationary subtropical ridges rather than traditional omega blocking patterns (Marshall et al. 2014; Sousa et al. 2018; Jiménez-Esteve et al. 2022). The main variable for this diagnosis is geopotential height, usually taken in mid-troposphere (500hPa) but the high pressure system extends over the full troposphere (Perkins 2015). Close to the surface on the other hand, a heat (or thermal) low can be present, creating local weak cyclonic circulation (Della-Marta et al. 2007; Fischer et al. 2007).

Lagrangian analyses. Heatwaves occur in the planetary boundary layer and in particular close to the ground where the dynamics is usually not geostrophic. Therefore, one cannot simply read the origin of air masses on a geopotential height map.

Lagrangian backward analyses (Santos et al. 2015) following particles of air inside the heatwave regions have proposed quantification of the different processes leading to large positive anomalies of near-surface air temperature. There are essentially three physical processes in the atmosphere which can lead to positive temperature anomalies:

- **Advection from hotter neighbouring regions** — usually from climatologically warmer regions,
- **Adiabatic warming by large-scale subsidence** at the core of the anticyclonic anomaly — also favoring clear skies and increased surface short wave insolation,
- **Diabatic warming** induced by various phenomenons such as turbulent sensible heat flux, latent heat release by condensation of water or short and long wave insolation.

Which of these mechanisms is the dominant driver of heatwaves varies greatly from one region to another, one heatwave to another and probably also with the intensity of the heatwave one is interested in (Lee and Grotjahn 2016; Quinting and Reeder 2017; Zschenderlein et al. 2019; Schielicke and Pfahl 2022). An important result from these studies is that in the mid-latitudes, hot events are usually not associated with strong heat advection from very remote regions ($>$ 2000 km) for particles close to the surface (Zschenderlein et al. 2018). Most of the air particles are already near to — especially to the east — or in the target area three days prior to the heatwaves, although transport above non-local but not far away dry soils may play an important role for advection of hot and dry air (see below). Remote advection is mainly present at mid- and upper-level, cf. Figure 1.2 for an example in Central Europe (Nakamura et al. 1997; Pfahl et al. 2015; Steinfeld and Pfahl 2019; Zschenderlein et al. 2020). Hotz et al. (2024) proposed a Lagrangian analysis of the vertical structure of recent extreme heatwaves revealing complex mechanisms that depend on the event study. In mid- and high-troposphere the anomalies are mostly advective, while they are adiabatic in low troposphere and diabatic in both low and high troposphere.

Large positive anomalies at the surface in the mid-latitudes are usually caused by strong adiabatic warming, and surface and radiative diabatic fluxes (Bieli et al. 2015; Zschenderlein et al. 2019). Much of the warming is therefore *in situ* — even though it is also clearly influenced by remote processes. Röthlisberger and Papritz (2023) recently proposed a quantification of the three physical processes presented above over the entire globe using ERA5 data in the 1979-2020 period. They showed strong geographical variations with a dominance of advection over mid-latitude oceans, adiabatic warming near mountain ranges and diabatic heating over tropical and subtropical land masses. Over Western Europe in particular, the anomalies with respect to the mean are around 8–10°C for maximum temperatures and decompose in around 2°C for the advective part, 4–6°C for the adiabatic part and 3–4°C for the diabatic part. They also show that most trajectories are local both in time and space.

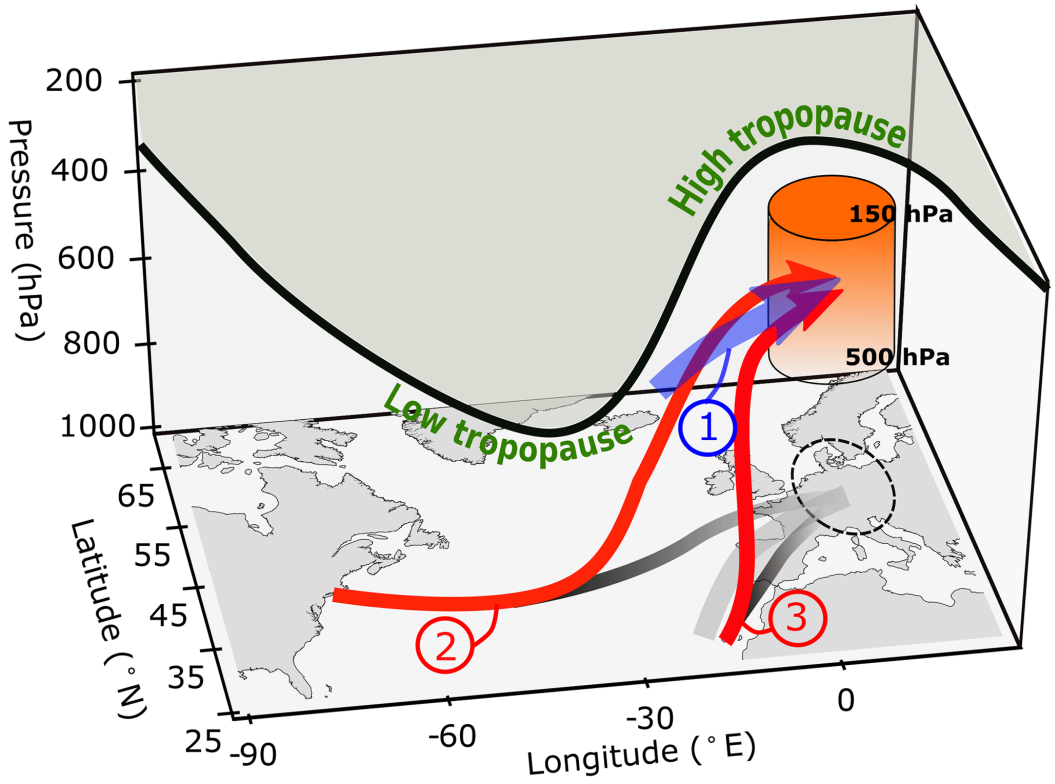


Figure 1.2: **Schematic representation of the origin of air parcels at mid and high-level in anticyclones during heatwave events in central Europe.** Source: Zschenderlein et al. (2020), Fig. 11.

Lagrangian studies use reanalyses or climate models outputs to compute backward trajectories. These simulations usually do not resolve convection explicitly and rely on parametrization schemes that are an imperfect representation of actual vertical movement of air in the atmosphere, especially during extreme events for which little is known. This may constitute an important source of uncertainties for the origin of warm anomalies at small scales. In general, processes occurring inside the planetary boundary layer (PBL) during extreme events are not fully understood and therefore well represented in models (Vautard et al. 2013). The dynamics of the PBL may play an important role in the intensification of extreme heatwaves by allowing a multi-day build-up of heat and entrainment of warm air over multiple days (Miralles et al. 2014). The use of convection permitting and/or high resolution models to better represent these mechanisms is a promising but stammering avenue of research (Sangelantoni et al. 2023; Williams et al. 2024).

Atmospheric blocking. The association between stationary anticyclonic structures, called atmospheric blocking, and mid-latitude heatwaves is well established (Kautz et al. 2022). What creates a heatwave is usually the persistence of this particular atmospheric structure: it is in general both a necessary and sufficient condition to reach extreme near-surface temperatures (Hoskins and Woollings 2015;

Pfleiderer and Coumou 2018; Röthlisberger and Martius 2019; Kautz et al. 2022). Blocking systems are long-lasting quasi-stationary and self-sustaining tropospheric flow patterns that are associated with a large meridional flow component and thus an interruption and/or deceleration of the zonal westerly flow (Fig. 1.3). Even though atmospheric blocking is usually understood as a so-called omega blocking pattern — with two low pressure systems on each flanks of the anticyclone —, dipole blocks and amplified ridges also fall into this category (Kautz et al. 2022). Various indices have been used to diagnose blocking situations (Woollings et al. 2018), but for most indices Europe is a dominant region of blocking due to the configuration of a strong, meridionally tilted storm track upstream of a large land mass.

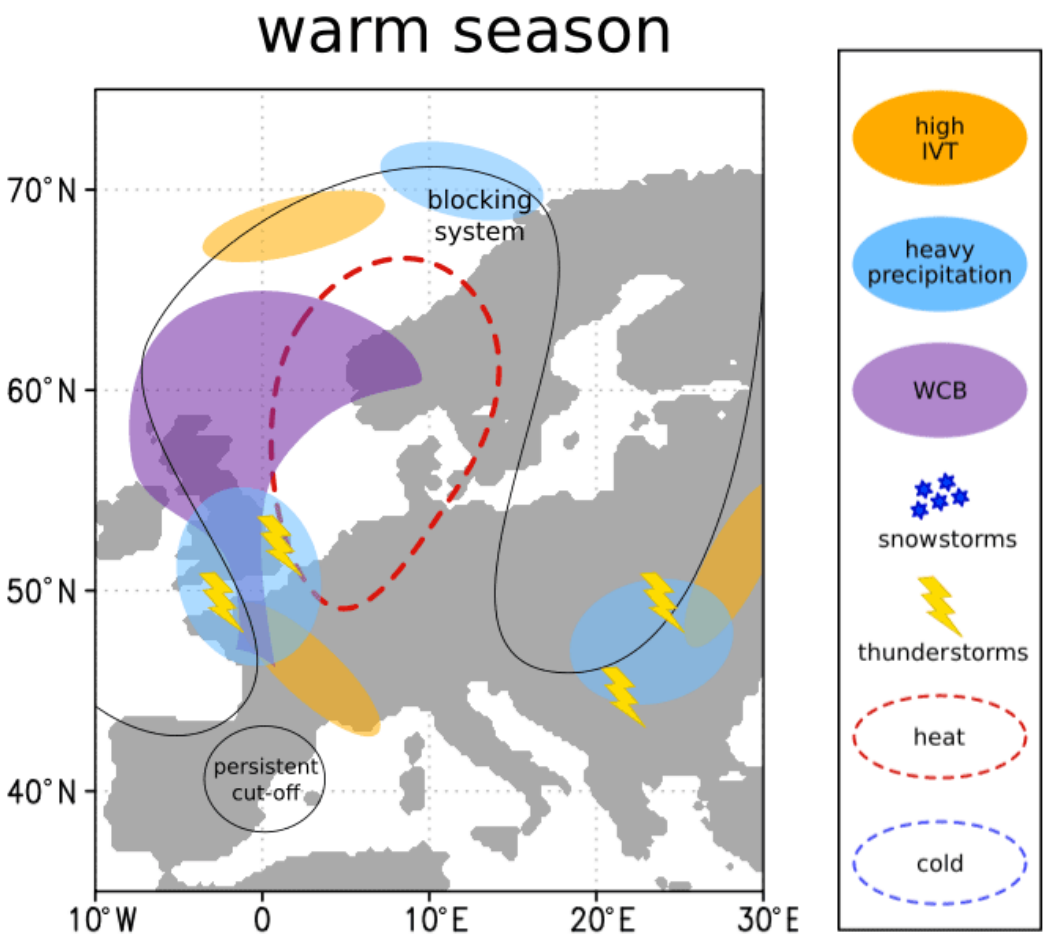


Figure 1.3: **Schematic representation of a typical blocking situation over Western Europe and associated meteorological impacts in summer.** Source: Kautz et al. (2022), adapted from Fig. 2.

The meteorological and climate literature on atmospheric blocking mechanisms is rich as these systems were studied as early as atmospheric data were available (Charney and DeVore 1979; Coughlan 1983). There is nevertheless limited understanding on the factors determining the onset and maintenance of these atmospheric

systems in general and in conjunction with large anomalies of near-surface temperature in particular (Barriopedro et al. 2023). These factors are numerous and may vary regionally, but they usually involve the interaction of atmospheric waves of different scales. Recent advances also underline the key role of latent heat release by condensation in moist ascending air streams in baroclinic systems upstream of the anticyclones for both formation and maintenance of blocking, especially for intense and large blocks (Pfahl et al. 2015; Steinfeld and Pfahl 2019) — see for example Schumacher et al. (2022a), Neal et al. (2022), and Baier et al. (2023) for the role of this mechanism to explain the intensity of the 2021 British Columbia extreme heatwave. Blocking situations may involve complex processes and origins of air parcels as shown by Zschenderlein et al. (2020): during European heatwaves, as much as half of air parcels inside the anticyclonic region have been diabatically heated by latent heat release. Two branches are identified in the North Atlantic: a remote branch — key for onset — above the western North Atlantic and related to the warm conveyor belt of upstream cyclones, and a nearby branch — key for maintenance — located in northwestern Africa which is diabatically heated by convection (cf. Fig. 1.2). Also, despite some recent progress, the frequency of blocking events remain underestimated by current climate models (Brunner et al. 2018; Kleiner et al. 2021).

Rossby waves. At synoptic and planetary scale, because of the presence of an atmospheric block, heatwaves are associated with an excursion of the extratropical jet from its climatological position, forming north-south meanders. These structures are called Rossby waves (Rossby 1939) and result from the conservation of so-called potential vorticity by atmospheric air parcels. Potential vorticity is a function of relative and planetary vorticity and of static stability. Because there is a north-south gradient of planetary vorticity on Earth, an air parcel that is displaced northward (southward) will begin to rotate anticyclonically (cyclonically), i.e. change its relative vorticity to conserve its potential vorticity, giving rise to Rossby waves. The large scale mid-latitude circulation can be decomposed in two groups of Rossby waves: (i) fast-moving synoptic-scale waves (sometimes called free waves) with high zonal wave numbers (≥ 6) propagating mainly in the longitudinal direction with a phase speed of 6-12 m/s, and (ii) slow-moving quasi-stationary planetary-scale Rossby waves with low zonal wave numbers (< 6) and close to zero phase speed (Schneider et al. 2015). The former can have various origins, including baroclinic development, diabatic heating in the tropics (Stan et al. 2017) or anomalous SSTs, while the latter are the climatological response to spatially inhomogeneous diabatic sources/sinks and orography and shape large scale seasonal mean circulation patterns (Hoskins and Karoly 1981; Held et al. 2002; White et al. 2021).

Rossby waves are ubiquitous in the atmosphere because they are its spontaneous response to perturbations. However, such potential vorticity anomalies in the atmosphere exist on a large spectrum from individual blocks and localized Rossby wave packets (RWP) to circumglobal Rossby waves (CGW). A theory of the linear behavior of Rossby waves propagation was developed in the second half of the 20th century (Hoskins and Karoly 1981; Branstator 1983; Held 1983; Hoskins and Ambrizzi 1993; Vallis 2017) and there has been a recent renewed interest for them because of their

links with extreme events (Schubert et al. 2011; Palmer 2013; Hoskins and Woollings 2015; Wirth et al. 2018). The latter can indeed arise from anomalously persistent and/or high amplitude Rossby waves.

The linear theory for the propagation of Rossby waves is based on a Reynolds decomposition of the atmospheric flow into the sum of a background flow and (small) perturbations. It introduces the concept of a refractive index for the atmosphere: Rossby waves are refracted towards regions of higher refractive index. Classical ray tracing allows to determine the expected propagation of Rossby waves for a given background flow. As the refractive index increases toward the equator, Rossby waves should be refracted towards the tropics where they would break and dissipate. But atmospheric jets can produce a local maximum of refractive index when jets are strong and narrow. There thus exists turning latitudes on both sides of the jet for waves of particular wavenumbers: between these two turning latitudes a waveguide exists which constrain the propagation of Rossby wave — and therefore synoptic-scale energy. As such, the atmospheric jets force Rossby waves to propagate zonally, enabling the propagation of energy over long distances and creating teleconnection patterns that synchronize the weather in remote regions of the world (Teng and Branstator 2019). This theory fundamentally depends on the so-called Wentzel-Kramers-Brillouin (WKB) approximation which states that the background flow should vary gradually relative to the scale of the waves, which is largely questionable in the atmosphere. An alternative perspective considers strong gradients of potential vorticity as waveguides: Rossby waves can propagate along a step function separating a region of high and a region of low PV (Platzman 1968; Swanson et al. 1997; Martius et al. 2010; Wirth 2020).

Whether the waveguide theories are correct for explaining the meanders of the jet is fundamentally a question of time scales. On climatological time scales, the causal relationship between waveguides and meridionally constrained Rossby waves has been well established (White et al. 2022). On subseasonal and synoptic time scales the correspondence is not so clear: the jet waveguides are impacted and shaped by the waves themselves. The waves are usually not small perturbations on the background flow and the separation between the two is difficult, particularly for high amplitude events that are associated to near-surface extremes (Wirth and Polster 2021). A full fundamental theory for the growth and propagation of Rossby waves on the spatial and temporal scales relevant to extreme events is lacking, particularly under strongly non-linear conditions. Although the role of amplified Rossby waves is well recognized for the emergence of persistent anticyclonic anomalies, two different mechanisms have been proposed for the formation of heatwaves blocking anticyclones (Horton et al. 2016): amplified quasi-stationary circumglobal Rossby waves (CGW) or high-amplitude transient non-circumglobal Rossby wave packets (RWP).

Circumglobal Rossby waves. The hypothesis of amplified CGW by waveguides for explaining heatwaves was first proposed by Petoukhov et al. (2013) and the associated physical mechanism is named quasi-resonant amplification (QRA). The authors suggested that a common mechanism for the generation of persistent, longitudinal, planetary-scale, high amplitude patterns of the atmospheric circulation

in the mid-latitudes with high zonal wave numbers (especially 5 and 7, cf. Fig. 1.4) results from trapping within mid-latitude waveguides of free synoptic waves. The quasi-stationary dynamical response with these high wavenumbers is usually weak but under the right conditions (especially a double jet), mid-latitude waveguides may favor a strong amplification of that response through quasi-resonance for preferred phases — and thus preferred location for extremes — of the wave patterns (so-called phase-locking behavior). This is similar to an oscillator forced near the natural frequency which enhances the amplitude of waves. This would create highly persistent configurations in summertime, reducing summer variability and favoring extremes at remote places at the same time — this mechanism being particularly relevant for concurrent heatwaves around the hemisphere. Several studies have investigated the relevance of this mechanism for recent heatwaves (Coumou et al. 2014; Petoukhov et al. 2016; Mann et al. 2017; Kornhuber et al. 2019; He et al. 2023; Screen and Simmonds 2014).

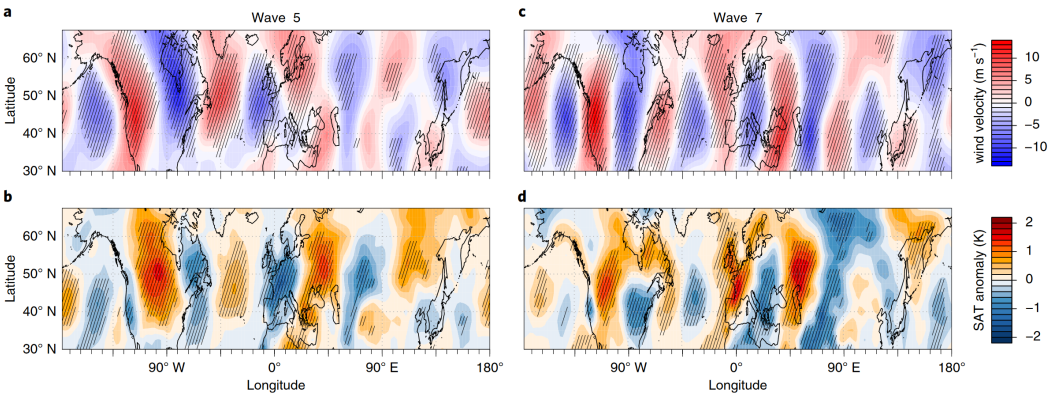


Figure 1.4: **Circumglobal Rossby waves and surface extremes.** (a,c) Composite plots of meridional wind velocity and (b,d) surface temperature anomalies during weeks of high-amplitude waves with wavenumbers 5 and 7. Statistically significant at 5% deviations from climatology are hatched. This analysis uses the NCEP reanalysis over the period 1979-2018. Source: Kornhuber et al. (2020), Fig. 2.

Although the link between preceding patterns of anomalous atmospheric planetary waves and surface heatwaves has been shown by several studies (Teng et al. 2013), the QRA mechanism has been criticized. In general, summer waveguides tend to be weaker than winter waveguides and not to be circumglobal (Teng and Branstator 2019). The contributions of CGW to temperature extremes is not higher than that of non-circumglobal patterns (Fragkoulidis et al. 2018; Röthlisberger et al. 2016). Although there is a statistically significant link between jet waviness on a hemispheric scale and monthly temperature anomalies (Screen and Simmonds 2014), and on synoptic time scales periods of frequently occurring temperature extremes have been linked to low storm track activity (Lehmann and Coumou 2015; Coumou et al. 2015; Pfleiderer et al. 2019), the link between jet waviness and the occurrence of weather extremes is stronger for regional scale than hemispheric jet

waviness (Röthlisberger et al. 2016). Jet waviness may be more linked to the occurrence of daily weather extremes via synoptic scale weather systems. Branstator and Teng (2017) for example analyzed correlation maps of the sub-seasonal variability of 200-hPa meridional wind to quantify the waveguidability and found that waviness in the upper troposphere is typically not stretched out circumglobally but organized in eastward propagating patches of limited spatial extent (Fragkoulidis et al. 2018). Also, the very definition of a waveguide may be problematic and some authors (Wirth and Polster 2021) have suggested that jets may arise as the consequence rather than the cause of waves with large amplitudes. Overall it is not clear that linear theory is still relevant for high amplitude events when the intensity of eddies is of the same order of magnitudes as the background flow.

Recurrent Rossby wave packets. Critics of the QRA hypothesis have pointed out that it is less likely than not that a persistent atmospheric blocking system arises as a result of a circumglobal phase-locked wave train. Another explanation for large and persistent anticyclonic anomalies is the recurrence of transient, fast-moving and amplified RWP that can arise from a variety of sources (Wirth et al. 2018). RWP are Rossby waves for which the amplitude has a local maximum and decays to smaller values at larger distances. RWP also propagate in the zonal direction, especially along waveguides, and they transfer energy from one individual trough or ridge to its downstream neighbor, a process called downstream development (cf. Fig 1.5 for an example).

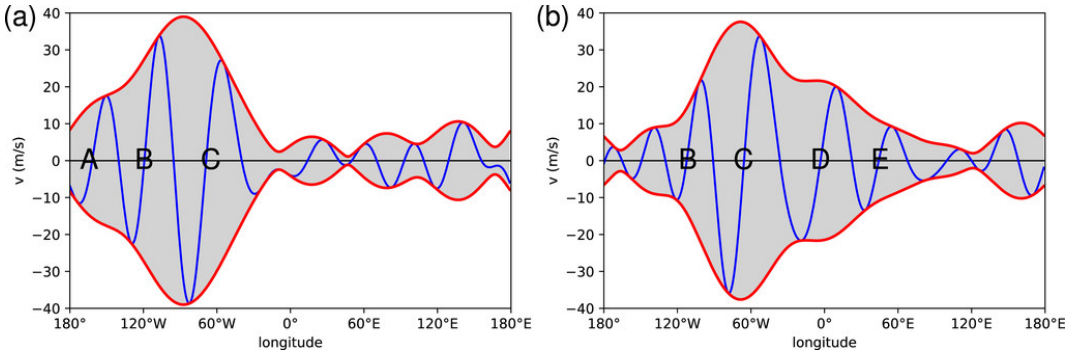


Figure 1.5: **Example of a Rossby Wave Packet.** (a) 0000 UTC 7 Aug and (b) 0000 UTC 9 Aug 2002 300hPa meridional wind restricted to zonal wave numbers 4-10 and averaged over 40°-60°N (blue) and envelope (red). Source: Wirth et al. (2018), Fig. 3.

Recurrent RWPs arising from multiple transient synoptic scale wave packets reamplifying in the same geographical region may result into a persistent anomalies (Tuel and Martius 2024). This can be seen on a Hovmöller diagram of the 250-hPa meridional wind but not on monthly composite maps (Röthlisberger et al. 2019). As such recurrent RWP can trigger persistent surface weather anomalies over multiple synoptic wavelengths while blocking is more local (one synoptic wavelength). Waves amplifying in the same phase upstream of the block could continuously reinforce the

block and lead to recurrent RWP conditions upstream. As explained above, diabatic processes and latent heat release associated with a series of transient synoptic cyclones upwind of the block can also contribute to the onset and maintenance of a stationary block and therefore a heatwave at the surface (Steinfeld and Pfahl 2019; Zschenderlein et al. 2020; Kautz et al. 2022).

There is indeed a statistical link between upper-tropospheric transient RWP and lower-tropospheric temperature extremes (Zschenderlein et al. 2018). The 2003 and 2010 heatwaves were not associated with CGW for example (Fragkoulidis et al. 2018). Recurrent RWP have also been linked to wet and dry spells (Ali et al. 2021). Jiménez-Esteve et al. (2022) showed on an idealized model simulation that mid-latitude heatwaves are associated with the eastward propagation and local amplification of Rossby wave packets. Pyrina et al. (2024) recently investigated the debate between CGW and non-circumglobal RWP for the occurrence of concurrent and non-concurrent heatwaves over the Northern and Southern hemispheres. In the clusters they identified, mid-latitude heatwaves typically occur at the leading edge of a RWP — where wave breaking takes place — to form or reinforce a block. These RWP were not associated to a specific zonal wave number. These results would tend to favor the recurrent RWP hypothesis. The authors nevertheless find that for high amplitudes CGW, concurrent heatwaves occur more often in the Northern Hemisphere when the dominant zonal wave number is 7.

b) Land surfaces

Heatwaves are extreme climate events in so far as they happen at the interface and through the interaction of several components of the climate system, the main ones being the atmosphere and the land surfaces. I now detail the crucial role of the latter in reaching extreme near-surface air temperatures. The atmosphere and the land surfaces are coupled through various exchanges of mass, energy and momentum happening in the PBL (Dickinson 1995). Here I mainly focus on the most important variable through which the atmosphere and land surfaces are coupled during heatwaves: soil moisture. It is only in the recent years that the importance of land surface evaporation for climate and climate extremes has been highlighted (Miralles et al. 2019). Soil moisture refers to the amount of water stored in the unsaturated soil zone, i.e. close to the surface. Soil moisture is the variable coupling land — and atmosphere — energy and water balances through the evapotranspiration (ET) term, i.e. the transfer of water from Earth's surface (open water and ice surfaces, bare soil and vegetation) to the atmosphere via evaporation and transpiration through the stomata of plant leaves. ET being highly sensitive to changes in radiation and temperature, it acts as a transmission belt of global and local changes throughout the entire water cycle, and therefore also the carbon cycles through its coupling to photosynthesis. Land surfaces represent a water and energy storage and therefore induce persistence and memory at longer timescales than the typical 5-10 days of the atmosphere (Lorenz et al. 2010).

The most direct effect of soil moisture for reaching high temperatures is through its role in the energy budget via the energy partitioning at the surface: high soil

moisture favors evaporation and surface cooling through latent heat flux. On the other hand, whenever soil moisture limits latent heat fluxes, more energy is available for sensible heating, inducing an increase of near-surface air temperatures (Seneviratne et al. 2010). Classically, a conceptual model where three soil moisture regimes are distinguished is used to present the interaction between soil moisture and the atmosphere:

- **dry regime**: soil moisture is lower than the wilting point, i.e. the amount of water under which plants wilt, and no or little ET happens,
- **wet regime**: soil moisture is higher than a critical value where the partitioning between latent and sensible heat fluxes is independent of the soil moisture content,
- **transitional regime**: soil moisture is on average between the wilting point and the critical value.

The wet regime is said to be energy-limited: soil moisture does not impact ET variability because incoming energy — mostly by short wave solar radiation — is the fundamental limiting factor. Dry and transitional regimes are both soil moisture-limited but only in the latter does soil moisture impact ET variability and therefore resulting energy fluxes to the atmosphere. In transitional regimes soil moisture - temperature interactions therefore increase summer temperature variability (Mueller and Seneviratne 2012).

In transitional regimes soil moisture deficit plays a key role for reaching extreme temperatures, under the appropriate anticyclonic atmospheric conditions (Alexander 2011; Perkins 2015; Gevaert et al. 2018). Crucially, there are large observational and modelling evidences for the asymmetric role of soil moisture on the temperature distribution: land - atmosphere coupling is stronger and more impactful during hot days than on average (Mueller and Seneviratne 2012; Seneviratne et al. 2013; He et al. 2022; Mukherjee et al. 2023). Models have for example shown that interactive or fixed soil moisture simulations have different behavior (Vogel et al. 2017): soil moisture - temperature feedbacks significantly contribute to the amplified warming of the hottest days compared to that of mean temperature. Most of Europe being in the transitional regime (Seneviratne et al. 2010), this mechanism is key for explaining the physical processes leading to increasing temperature extremes in this region. Similarly, anthropogenic interventions such as irrigation (Thiery et al. 2017), overexploitation of lands (Cowan et al. 2020) or heat release in cities (Chen et al. 2023) have been shown to have limited influence on annual mean temperature but to enhance extremes (up to 1°C). Strandberg and Kjellström (2019) for example showed in simulations that in regions where water is not a large constraint, afforestation leads to a stronger cooling in the maximum rather than mean local near-surface temperatures.

The interaction of favorable atmospheric conditions for heatwaves with low soil moisture essentially leads to the self-amplification of heatwave events. Pre-existing low soil moisture implies that a larger fraction of incoming radiation is employed

to warm up the environment, which leads to an accumulation of sensible heat in the atmosphere (Fischer et al. 2007). But more importantly, there is a feedback loop between high temperatures and low soil moisture. The persistent anticyclonic patterns lead to low relative humidity, high air temperature and reduced cloudiness, causing a high atmospheric demand for water that increases ET. As soil and vegetation dry out, latent heat fluxes will be reduced, implying the partitioning of already higher incoming solar energy towards more sensible heat fluxes. Air in the PBL therefore becomes drier and hotter, reinforcing the already high atmospheric evaporative demand (Teuling et al. 2013; Miralles et al. 2019). Cloud cover also declines as a result of drier air and limited latent heat flux, decreasing precipitations and increasing incoming shortwave radiation (Vogel et al. 2017; Vogel et al. 2018).

Importantly, studies have shown that the influence of soil moisture on heatwaves can be non-local through heat advections from dry neighboring regions (see also the discussion on the Lagrangian analyses above). Several works have shown the impact of remote dry spring conditions for favoring heat in the summer during particularly strong heatwave events, especially dry conditions in the south of Europe spreading in the north of Europe through atmospheric transport of anomalously warm and dry air (Vautard et al. 2007; Zampieri et al. 2009; Quesada et al. 2012). Physically, dry air forming over dry soils in the Mediterranean region induces less convection and diminished cloudiness. This air can then get transported northward in the PBL, increasing local temperature and vegetation evaporative demand, activating the feedback loop described above. Especially, summer heat is more sensitive to the occurrence of specific weather regimes in initially dry than wet cases, inducing asymmetry also in summer heat predictability. Several studies on so-called mega-heatwaves — i.e. heatwaves with an exceptional spatial and temporal extent — have shown such a mechanism of heat building by large-scale upwind horizontal advection (Miralles et al. 2014; Schumacher et al. 2019). Schumacher et al. (2019) for example showed using a Lagrangian tool tracking the origin of heat, the role of upwind land-atmosphere feedbacks in the 2003 and 2010 heatwaves: up to 30% of the advected sensible heat was caused by the drought upwind. Besides of the self-amplification mechanism, a self-propagation for droughts via an atmospheric bridge and through heat advection — and therefore potentially heatwaves — has been proposed (Schumacher et al. 2022b).

Several studies have suggested an influence of land surface anomalies not only on local and regional PBL properties (Ek and Holtslag 2004; Lemordant et al. 2016) — affecting its stability and growth and therefore entrainment of air from the top and avection from the surroundings — but also on the large scale tropospheric circulation which may act as an amplification mechanism (Douville 2002; Douville and Chauvin 2000; Koster et al. 2016; Wang et al. 2019; Teng et al. 2019; Domeisen et al. 2023). Local dry-hot surface conditions can influence the tropospheric circulation by producing an intense surface heat low and favoring anticyclonic circulation in the mid-troposphere, which could constitute a feedback between continental-scale circulation and extreme temperatures (Fischer et al. 2007; Haarsma et al. 2009; Zampieri et al. 2009). Mega-heatwaves in particular can create exceptionally hot and deep PBL (Miralles et al. 2014), which are likely to substantially modify the

global circulation as shown by model simulations (Merrifield et al. 2019), favoring non-local heatwave amplification in subsequent months. Inhomogeneous warming can also amplify quasi-stationary waves (Sato and Nakamura 2019), which may constitute a reinforcing mechanism for long lasting events. Koster et al. (2016) in particular showed in a simulation that an idealized diabatic heating anomaly creates a positive feedback loop enhancing further the overall dryness of the continental interior. Knowledge on the link between surface anomalies, mesoscale and synoptic circulations during heatwaves nevertheless remains low (Barriopedro et al. 2023).

Despite the now widely recognized role of soil moisture in shaping the right tail of the distribution of temperatures, the understanding of soil moisture - atmosphere feedbacks during heatwaves remains incomplete. Especially, it remains a large source of uncertainties in current and future climate model projections (Vogel et al. 2018; Luo et al. 2022; Al-Yaari et al. 2023). Al-Yaari et al. (2023) for example recently showed that in the CMIP6 ensemble over the historical period the mean intensity of heatwaves tends to be overestimated because of overly pronounced soil dessication during the events. Measurements and observations of surface hydrology are not as global as atmospheric observations and they more strongly depend on local contexts. The intricate relationships between different variables also limit the formal identification of causal links outside models, especially for non-local feedbacks (Miralles et al. 2019). There are therefore substantial discrepancies between models and observations (Hirschi et al. 2011; Dirmeyer et al. 2018). The crucial partitioning of surface energy into latent and sensible heat fluxes is in particular incorrectly represented which leads to errors expanding throughout the surface energy and water cycles.

Because soil moisture impact on ET is largely mediated by plants transpiration (Seneviratne et al. 2010), representing correctly vegetation characteristics are essential for these interactions (Kala et al. 2016). Modelling the response of large and diverse ecosystems to temperature and water stress remains a difficult problem due to the variety of plant physiological responses (De Kauwe et al. 2015). Because the coupling between the atmosphere and land surfaces is stronger during heatwave events, this can lead to substantial errors on the daily maximum temperatures. Stéfanon et al. (2012) for example showed with the example of the June and August 2003 heatwaves that accounting for interactive plant phenology would attenuate June heat wave but amplify the August heatwave by suppressing ET leading to enhanced daily maximum temperatures by up to 1.5°C.

The different mechanisms presented in the last two subsections are summarized in Figure 1.6.

1.3.3 Modulation at longer time scales

At longer time scales, heatwaves are modulated by various slow components of the climate system. However, much less is known about the exact importance of these mechanisms (Barriopedro et al. 2023). An important role is played by local and remote SSTs. In general, anomalous SSTs have been linked to changes in the general circulation of the atmosphere (Gastineau and Frankignoul 2015). Simulations and

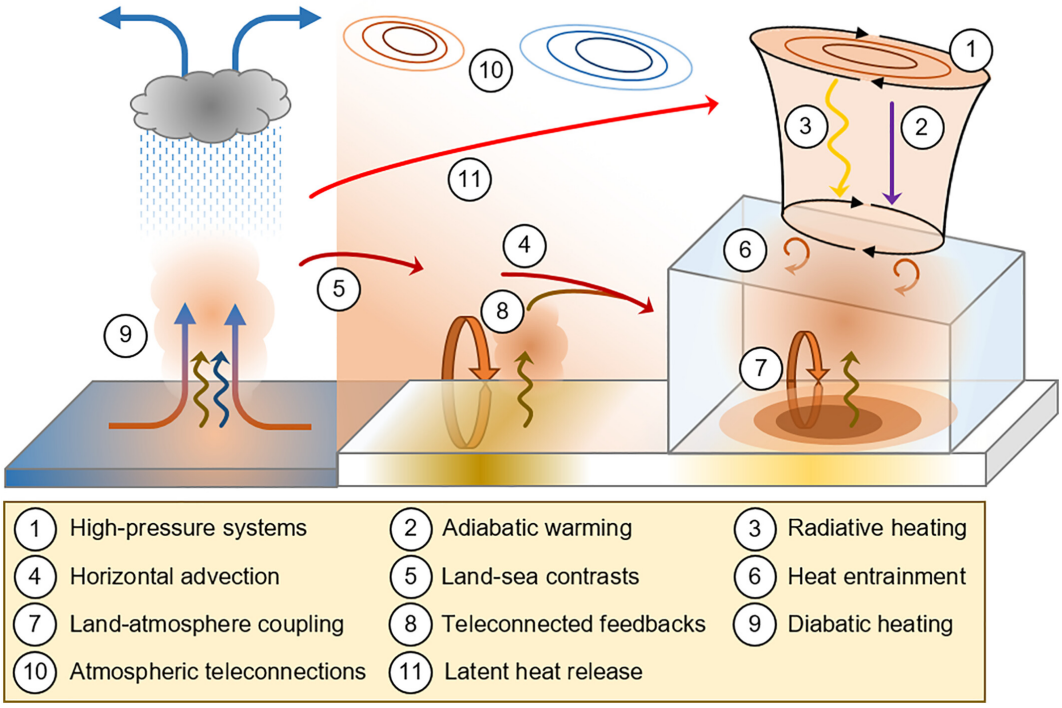


Figure 1.6: **Summary of the large-scale-to-regional drivers of heatwaves.** The figure excludes external forcings such as greenhouse gases, aerosols and land-use/land-cover changes affecting long-term trends. Source: Barriopedro et al. (2023), Fig. 8.

observations have suggested that anomalous patterns of SSTs could trigger anomalous Rossby wave trains, quasi-stationary hemispheric waves or shifts in mid-latitude jets position leading to heatwaves (Duchez et al. 2016; McKinnon et al. 2016b; Wolf et al. 2020; Luo et al. 2022; Deng et al. 2022; Luo et al. 2023; Wallberg et al. 2024). Tropical SST drivers for heatwaves are clearly established and involve interactions with large-scale modes of variability such as ENSO, IOD or MJO (Stan et al. 2017; Domeisen et al. 2023), although these drivers usually highly depend on the region of the world where the heatwave occur.

The role of natural and anthropogenic aerosols is also increasingly recognized as an important factor for explaining the intensity of recent heatwaves, and more generally regional patterns of warming (Schumacher et al. 2024). The role of aerosols during heatwaves involve complex atmospheric physics and chemistry, with direct effects on solar radiation and indirect effects from the interaction with clouds microphysics. Aerosols could either buffer (Dey et al. 2021) or intensify (Baró et al. 2017) heatwaves. Importantly, the role of the diminution of anthropogenic aerosol emissions — especially over the North Atlantic region — could have been in the past and be in the future stronger than the role of GHG increase (Dong et al. 2017; Samset et al. 2018; Zhao et al. 2019; Wang et al. 2023). Nevertheless, little is known on the mechanisms involving aerosols during heatwaves (Barriopedro et al. 2023).

1.3.4 Future evolution of heatwaves

a) Thermodynamical evolution

The question of the past and future evolution of heatwaves — especially under anthropogenic global warming — is crucial for societal adaptation. It is also a fundamental climatological question: how do extreme heat events in the climate system respond to a change in the external forcings? This question can be decomposed into two elements: the thermodynamical evolution and the dynamical evolution. The thermodynamical evolution refers to the global increase of near-surface temperatures. This evolution is well understood: global (GHG concentrations) antropogenic forcings have been undoubtedly recognized as the major drivers of the increase in frequency, duration and intensity of heatwaves worldwide. The continued warming of the Earth caused by enhanced trapping of the long-wave thermal radiation emitted by the planet results in local warming of the temperature distribution, which clearly increases the frequency of heat extremes *defined using fixed thresholds* (Seneviratne et al. 2021; Barriopedro et al. 2023; Domeisen et al. 2023).

Changes in temperature extremes have accelerated in the last decades at a faster rate than the mean temperatures (Perkins-Kirkpatrick and Lewis 2020; Barriopedro et al. 2023; Huntingford et al. 2024). The last IPCC report stated that it is virtually certain that there has been an increase in the number of warm days and nights since 1950 (Seneviratne et al. 2021). There has also been an increase in warm season heatwave characteristics such as magnitude, duration and frequency, even though confidence can vary locally. Importantly for societal adaptation, there has been a strong increase in record breaking events: for all continents the top three years for heatwave magnitude occurred during the 21st century (Barriopedro et al. 2023). The amount of record-breaking and record-shattering events is also expected to increase and depend on the rate of warming (Bador et al. 2016; Power and Delage 2019; Fischer et al. 2021).

Most of the changes in heatwaves characteristics are explained by the increase in the mean temperatures, the latter being the result of global warming with important regional variations (oceans vs land, equatorial vs polar regions) (Domeisen et al. 2023; Van Loon and Thompson 2023). A key question is whether the trends in extremes currently observed could also be driven by a change in variability and higher order moments (Simolo and Corti 2022). In general, a faster increase in the probability of extremes is not incompatible with a change of only the mean: for a fixed heatwave threshold and, for example, a Gaussian distribution for temperatures, the probability to be above this threshold will increase non-linearly with a linear increase of the mean. Several studies have emphasized the importance of increasing variability (Kodra and Ganguly 2014; Douville et al. 2016). But in many regions a shift in the mean can explain a large fraction of the observed changes (McKinnon et al. 2016a) and there is currently no clear evidence for the hypothesis of global increases in temperature variance. The change in mean temperature caused by increased radiative forcing dominates future changes of warm extremes. The contribution from increased variability is expected to be larger in transitional regions (see below) but projected changes in variability are small (Barriopedro et al. 2023).

Future heatwave characteristics may therefore have a similar relationship to the corresponding future climatology as today's characteristics do to today's climatology (Vogel et al. 2020).

Overall, there is strong evidence for the anthropogenic influence in the observed trends of temperature extremes at different temporal scales (Fischer et al. 2014; Christidis et al. 2015; Kim et al. 2016; Diffenbaugh et al. 2017; Dunn et al. 2020; Ossó et al. 2022; Skrzyńska and Twardosz 2023; Ruosteenoja and Jylhä 2023). Current levels of global warming may already be responsible for approximately 75% of hot days globally (Fischer and Knutti 2015). Quantifying the relative contribution of different anthropogenic forcings is nevertheless still problematic. Locally, land-use/land-cover practices may play an important role (Lejeune et al. 2018). Extensive practices such as irrigation or cropland intensification may have dampened the regional increases in hot extremes to the point of reversing the warming trend in some regions (Mueller et al. 2016; Thiery et al. 2020). Internal variability — especially multi-decadal modes of variability — may also have played and will play an important role and represent an important source of uncertainty for projections (O'Reilly et al. 2021; Barriopedro et al. 2023; Blanusa et al. 2023; Luo et al. 2023). At high latitudes, up to 25%-50% of the projected spread in warm season heatwave days may depend on natural variability (Perkins-Kirkpatrick and Gibson 2017). CMIP models do capture the observed changes in extreme indices for heatwaves, although with some tendency to overestimate the magnitude trends in hot days (Fan et al. 2020; Al-Yaari et al. 2023). There has nevertheless been only limited additional improvements in terms of mean biases or model spread between the difference versions of CMIP (Fan et al. 2020; Kim et al. 2020; Wehner et al. 2020).

b) Atmospheric circulation

Dynamical changes of the atmospheric circulation and/or of local feedback mechanisms with soil moisture could imply a faster (or slower) increase in extremes than the linear shift of the distribution of extreme temperatures would imply. This would translate into a change of the shape of the temperature distribution. The level of confidence on projected dynamical changes in general and for extremes in particular is however low (White et al. 2022). As stated by Shepherd (2014): "Nearly everything we have any confidence in when it comes to climate change is related to global patterns of near-surface temperature, which are primarily controlled by thermodynamics". The question of the evolution of atmospheric dynamics with climate change is hotely debated and no clear consensus has been reached for the moment (Coumou et al. 2015; Horton et al. 2015; Rogers et al. 2022; Rousi et al. 2022).

One large issue to reach consensus on this matter is that current climate models are biased in the way they represent the circulation, especially for extremes (e.g. Vautard et al. (2023) and Patterson (2023) for Western Europe). The limits of models include the limited horizontal resolution, the representation of air-sea interactions, the land-atmosphere coupling, the land-cover and land-use changes, the parametrization of boundary layer processes and surface fluxes, the hydrological cycles etc. Simulating correctly the mean state of the climate and at the same time

the multi-scale interactions leading to heatwaves in the present and in the future is therefore very challenging for current climate models (White et al. 2022). These limits explain why the dynamical evolution of the atmosphere under anthropogenic forcing is largely uncertain.

Even when considering only the average atmospheric circulation, there is no clear consensus at this point. One of the key question in the changes of the Northern Hemisphere circulation is the evolution of mid- and high-level jets (Barnes and Screen 2015). A lot of the debate on the changes in atmospheric circulation revolves around changes in mid-latitude jets positions, speed and waviness (White et al. 2022). There is also an ongoing debate about the specific mechanisms at play in connections to dynamical changes, for extreme events in particular (Barnes and Screen 2015; Francis 2017; Huguenin et al. 2020; Stendel et al. 2021). Even if some statistically significant trends in the atmospheric circulation have been observed in the recent decades (*detection* question), in general it is not clear whether they represent a forced response to anthropogenic forcings (*attribution* question). The detection of robust circulation trends is difficult and risky due to the limited period of observations and large internal variability. On the other hand, the attribution of dynamical changes is also problematic since they are often not well understood and/or robustly simulated by models (Shepherd 2014; Hoskins and Woollings 2015; Shaw 2019; Simmons 2022).

Several studies have shown that simulations of the CMIP ensemble of climate models do not project a dynamical change of weather patterns associated to heatwaves in the future (Cattiaux et al. 2012; Brunner et al. 2018; Ventura et al. 2023). In particular, the association with blocking events is still present as one could expect physically (Zhang et al. 2023). It is therefore reasonable to assume that, at least for limited global warming, the mechanisms by which high near-surface air temperatures are reached will be similar in the future — although their relative importance could change. The debate on the dynamical changes in the atmosphere is thus essentially focused on the question of whether the patterns associated with heatwaves will occur more often: will there be a change in the frequency of blocking events in particular and in the waviness of jets in general?

In the recent years there has been more circulation patterns favoring heatwaves over Western Europe, which partially explain the large increases in extremes in this region of the world (Horton et al. 2015; Jézéquel et al. 2018; Alvarez-Castro et al. 2018; Vautard et al. 2023; Singh et al. 2023; Patterson 2023; Faranda et al. 2023b; D'Andrea et al. 2024). Global change in weather patterns is however highly uncertain (Grotjahn et al. 2016) and there is for the moment no robust compelling evidence of increase in the amplitude or frequency of wavy patterns associated with heatwaves across the diversity of the metrics proposed (Simmons 2022; Barriopedro et al. 2023). Similarly, there is no obvious trend in blocking frequencies in summer. There may be an emerging consensus in models for decreasing blocking frequency and increasing blocking duration (Hoskins and Woollings 2015; Davini and d'Andrea 2020; Kautz et al. 2022), but with important regional variations. Moreover trends computed over the historical period are weak and contrast with observations. The observed evolution of blocking is not clear and may depend on the indices used for

diagnosis (Barnes et al. 2014; Woollings et al. 2018). Natural variability is high and may dominate the future evolution, at least in the next few decades.

The mechanism often invoked for potential changes in the dynamics of atmospheric jets is Arctic amplification (Previdi et al. 2021). Arctic amplification refers to the enhancement of near-surface air temperature change over the Arctic relative to lower latitudes and involve several physical mechanisms, mainly changes in the albedo due to sea-ice loss, but also energy transport from lower latitudes. As a consequence, the meridional near-surface temperature gradient between the Equator and the Pole is expected to decrease as a result of global warming. This would change the baroclinicity, reduce the speed of jets and more generally change the atmospheric circulation in boreal summer. On the other hand, due to enhanced deep convection and latent heating in the tropics, the meridional upper-level temperature gradient is expected to increase, which would strengthen the mid-latitude westerlies (Coumou et al. 2018). As such, the expected evolution of the mid-latitude jets is stuck in the so-called *tug-of-war* between these two opposite influences. For the moment, the atmospheric response to Arctic sea ice loss is not robust in observations and models (Barnes and Screen 2015; Screen et al. 2018) and observed changes may simply be the result of natural variability (Huntingford et al. 2019).

For heatwaves in particular, in addition to a climatological change in the mean position of the jets which would impact the mean summer temperature, two competing mechanisms for future dynamical changes have been proposed (Horton et al. 2016) (Fig. 1.7):

- Increasingly favorable conditions for the amplification of free waves leading to high amplitudes CGW by the QRA mechanism described above,
- Weakening of the summer storm track leading to more persistent weather patterns.

Both mechanisms would lead to an increase in the variability of the boreal summer and therefore would explain the recurrence of weather patterns leading to more heatwaves than a linear shift of the temperature distribution would suggest.

Several studies have claimed to have found recent statistically significant increase in the frequency of high amplitude events in July and August (Coumou et al. 2014; Rogers et al. 2022). This may arise as a result of Arctic amplification which would favor in particular the increase in frequency and persistence of double jet structures (Mann et al. 2017; Kornhuber et al. 2019; Rousi et al. 2022). However, model simulations and theoretical arguments suggest a reduced waveguidability for weakened jets under climate change, which is in contradiction to what one should expect if more CGW are to be seen in the future (Teng and Branstator 2019).

Another possibility is the weakened summer circulation in the North Atlantic caused by reduced low-level baroclinicity and therefore less or weaker synoptic-scale cyclogenesis in summer. This would lead to slower propagating synoptic weather systems, and which would favor more persistent weather patterns (Kornhuber and Tamarin-Brodsky 2021; Luo et al. 2024). Coumou et al. (2015) showed a significant weakening of the summer circulation in observations using a variety of metrics (zonal

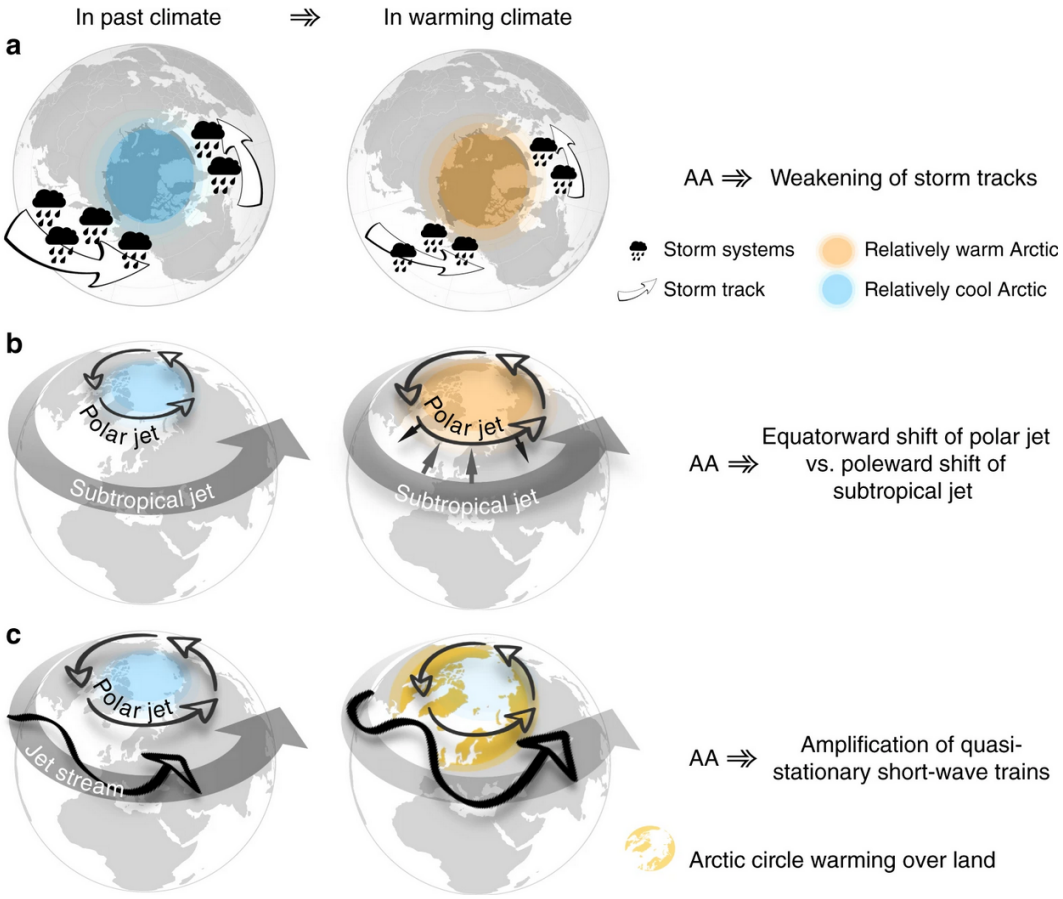


Figure 1.7: **Schematic representation of proposed mechanisms for the evolution of atmospheric circulation during boreal summer.** Source: Coumou et al. (2018), Fig. 3.

wind, eddy kinetic energy, amplitude of fast-moving Rossby wave) and the association between monthly heat extremes and weak circulation in the North Atlantic. This weakening is expected to continue with more global warming (Lehmann et al. 2014; Kornhuber and Tamarin-Brodsky 2021). These mechanisms are quite robust in so far as theoretical, observational and modeling evidence supports the hypothesis that summer storm tracks weaken with enhanced Arctic warming (Coumou et al. 2018).

Some studies have indeed detected enhanced predictability/persistence (Hoffmann 2018; Pfleiderer and Coumou 2018; Pfleiderer et al. 2019; Kornhuber and Tamarin-Brodsky 2021), especially for heatwaves, in the Euro-Atlantic region. These results have however been criticized (Huguenin et al. 2020), mainly because they are often metrics-dependent. Some studies (Dorrington et al. 2022) even found less persistent weather regimes in a warming worlds, with an increasing zonal flow.

Meridional temperature gradients are however not the only large scale changes to consider for the dynamics of mid-latitude jets. The tug-of-war also applies to land-sea temperature gradients which are expected to increase as the result of in-

creased radiative forcing and lower thermal capacity of the lands (Shaw and Voigt 2015). This change in thermal forcing may also be able to cause an amplification of quasi-stationary atmospheric waves leading to an intensification of summer trends in temperatures at certain locations (Sato and Nakamura 2019) via wavier jets (Moon et al. 2022).

c) Land surfaces

Besides atmospheric dynamics, an another important driver of the increased frequency and intensity of heatwaves in the future is the increasing occurrence of soil moisture limitations and droughts. Even though there are still large uncertainties on the response of soil moisture to rising GHG concentrations (Vogel et al. 2018), there is confidence in a northward shift of soil-moisture limited regimes in the mid-latitudes, creating a new transitional climate zone with strong land - atmosphere coupling in Central and Eastern Europe (Seneviratne et al. 2006). This implies that evapotranspiration variability would strongly increase under global warming, which translates into an increased variability of sensible heat fluxes and therefore near-surface air temperatures (Seneviratne et al. 2010; Perkins 2015; Gao et al. 2023). A preferential heating of the upper tail of the temperature distribution would thus largely be governed by increased coupling with (low) soil moisture (Diffenbaugh et al. 2007; Fischer et al. 2012; Dirmeyer et al. 2012; Al-Yaari et al. 2023; Huntingford et al. 2024). Soil moisture - temperature feedbacks may represent the main factor driving projected increases in hot extremes in mid-latitudes and explain why hot extremes are warming faster than the mean (Schwingshackl et al. 2018; Domeisen et al. 2023). Teng et al. (2016) for example showed in a model simulation that the amplification of future heat waves is largely driven by local energy fluxes and not induced by planetary wave events variability.

However, as explained above, soil moisture - temperature interactions are imperfectly represented in models which leads to large uncertainties in climate models projections (Vogel et al. 2018; Miralles et al. 2019). Typical regional errors including too persistent and/or intense heatwaves and overestimated variability in extreme temperatures have been linked to too strong land-atmosphere feedbacks (Mueller and Seneviratne 2014; Miralles et al. 2019; Van Loon and Thompson 2023). Vegetation effects — especially land-cover and land-use historical changes or parametrization of biogeophysical processes — are not well represented and can cause opposite responses in extremes depending on dominant mechanisms (Barriopedro et al. 2023). Climate models also often ignore future irrigation changes. Subtle effects like the impact of an increasing level of CO_2 in the atmosphere can also affect plant dynamics during extremes as shown by Lemordant et al. (2016): increased CO_2 levels change the seasonality of the water cycle through stomatal regulation and increased leaf area. The water saved during the growing season through higher water use efficiency could mitigate summer dryness and the impacts on heatwaves.

Summary

Context

In this introductory chapter I give definition elements for the climate system, for the concept of climate as the ensemble of statistical properties of the climate system and for extreme events in general. I discuss the under-sampling issue inherent to the study of extreme and rare events and present the three solutions that will be used in this thesis: Extreme Value Theory (EVT), long simulations and rare events algorithms. Finally, I review the physical mechanisms associated to heatwaves in the mid-latitudes and how they may change with anthropogenic global warming. I outline in particular the uncertainties related to the evolution of the atmospheric dynamics associated to heatwaves.

Goals

This thesis is about statistical and dynamical aspects of extreme heatwaves in the mid-latitudes. In the rest of the manuscript I will address the following questions:

1. **What is the maximal near-surface air temperature that can be reached by a heatwave event?** (chapter 2 and chapter 3)
2. **Are the physical mechanisms leading to *very* extreme heat events similar to those leading to extreme ones?** (chapter 5)
3. **Is it possible to simulate efficiently extreme heatwaves with climate models?** (chapter 4, chapter 6 and chapter 7)
4. **How will the dynamics leading to extreme heatwaves change with climate change?** (chapter 7)

Chapter 2

Maximal reachable temperatures for Western Europe in current climate

2.1 Introduction

In this chapter I present a work motivated by the recent paper of Zhang and Boos (2023) who proposed an estimate of a physically realistic upper bound for surface temperatures in the mid-latitudes. The question motivating this work is how more intense an observed extreme heat event could have been. I present a case study of the record breaking July 2019 heat event in Western Europe. The event was short but extremely intense in Western Europe, breaking previous records by a large margin (larger than 2°C for the Paris-Montsouris weather station, for example). I compare the physical estimate of the upper bound with a statistical estimate based on Extreme Value Theory (EVT). I show that the EVT-based upper bound likely underestimate the worst heatwave event physically possible.

The rationale for investigating the July 2019 event in Western Europe comes from the results of recent works (Patterson 2023; Vautard et al. 2023) which have shown that this region of the mid-latitudes has seen a more rapid trend in extreme temperatures than predicted by most climate models. Vautard et al. (2023) employed a flow analogues methodology and attributes a part of this trend to atmospheric patterns changes. Here I also condition on the atmospheric dynamics of the July 2019 event but rather investigate thermodynamical changes influencing the intensity of the event between two past periods (1940-1980 and 1981-2021).

2.2 Article published in *Environmental Research Letters*

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
21 June 2023

REVISED
4 August 2023

ACCEPTED FOR PUBLICATION
4 September 2023

PUBLISHED
18 September 2023

LETTER

Maximal reachable temperatures for Western Europe in current climate

Robin Noyelle[1,*], Yi Zhang[2,3], Pascal Yiou[1] and Davide Faranda[1,4,5]

[1] Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay & IPSL, Gif-sur-Yvette 91191, France
[2] Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, United States of America
[3] Miller Institute for Basic Research in Science, University of California, Berkeley, Berkeley, CA 94720, United States of America
[4] London Mathematical Laboratory, 8 Margravine Gardens London, W6 8RH London, United Kingdom
[5] Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, PSL Research University & IPSL, Paris, France
[*] Author to whom any correspondence should be addressed.

E-mail: robin.noyelle@lsce.ipsl.fr

Keywords: heatwaves, climate change, maximum temperatures, flow analogues

Abstract

Human bodies, ecosystems and infrastructures display a non-linear sensibility to extreme temperatures occurring during heatwave events. Preparing for such events entails to know how high surface air temperatures can go. Here we examine the maximal reachable temperatures in Western Europe. Taking the July 2019 record-breaking heatwave as a case study and employing a flow analogues methodology, we find that temperatures exceeding 50 °C cannot be ruled out in most urban areas, even under current climate conditions. We analyze changes in the upper bound of surface air temperatures between the past (1940–1980) and present (1981–2021) periods. Our results show that the significant increase in daily maximum temperatures in the present period is only partially explained by the increase of the upper bound. Our results suggest that most of the warming of daily maximum surface temperatures result from strengthened diabatic surface fluxes rather than free troposphere warming.

1. Introduction

Extreme weather events have strong impacts on societies and ecosystems [1]. Among them, heatwaves have been the focus of extensive attention due to their increasing frequency and intensity with global warming [2]. Most biological and technical systems have a limiting capacity to cope with extreme temperatures and can respond non linearly once certain thresholds are reached. These include human bodies [3–6], plants [7, 8], ecosystems [9] and infrastructures [10]. Urban areas are particularly at risk due to the so-called heat island effect [11] which enhances maximum temperatures during the day and minimum temperatures during the night. The capacity to predict and anticipate future maximum intensities of heatwaves is therefore of primary importance for adaptation to climate change.

The physical mechanisms leading to heatwaves in the midlatitudes combine specific atmospheric synoptic conditions and anomalously low soil moisture, which can interact to further increase the intensity of the event [11–14]. Heatwaves are associated with a slow moving, sometimes called quasi-stationary, high-amplitude Rossby wave [15] which is often embedded in a hemispheric pattern of wave number 5 or 7 [16–18]. Above the heatwave region, a blocking anticyclone builds up at mid- and upper-level troposphere in conjunction with a change in the jet stream's climatological path, towards a large poleward meridional meander. The anticyclone sustains the poleward advection of warm air along its western flank, adiabatic warming by subsidence and clear skies at its center. When linked to abnormally dry soils, these conditions promote warming through short-wave insolation, resulting in the allocation of incoming solar energy towards sensible rather than latent heat [19–26]. If the anticyclone stays stationary, extreme temperature can be reached. Although it is well established that global warming renders the temperature distribution hotter on average, how the coupled dynamics between the atmosphere, the oceans and the soils will evolve in response to this forcing during the hottest days is still debated [11, 14, 27–31].

One way to quantify the risk associated with heatwaves is to measure the tail probabilities of the temperature distribution, e.g. the risk of the maximum annual temperature reaching a certain level. A classical approach to deal with this assessment is extreme value theory (EVT). EVT has been used to determine statistical models for maxima (or minima) of climate variables of interest [32], such as temperatures. It is based on a convergence principle of the probability distribution of block maxima or peaks-over-threshold [33–35] towards a generalized extreme value (GEV) distribution. It allows to compute return values corresponding to very large return periods (i.e. longer than the period of observations), even in non stationary contexts [36]. This approach is for example extensively used by attribution methods to compare the probabilities of reaching extreme temperatures between a counterfactual world without climate change and a factual world with climate change [37–39]. This allows to estimate how climate change made a particular event more (or less) likely.

Fitting a GEV distribution to extreme temperatures usually gives estimates of the shape parameter that are robustly negative (e.g. [40, 41]). This implies that, in a stationary context, the distribution of annual maximum temperatures is bounded upwards. This upper bound is often the information asked to the scientific community by practitioners who want to prepare for the worst case scenario. However, recent intense heatwaves such as in 2010 in Russia [18], 2019 in Western Europe [42] and 2021 in the Pacific Northwest [43, 44] have challenged the reliability of this statistical upper bound by breaking it sometimes by a large margin. Another approach to estimate maximal reachable temperatures would be to run a long simulation with a climate model to sample more extreme events than those observed in the past. Apart from the inherent limitations of models to represent correctly the entire temperature distribution—especially at its tail [40, 45, 46] — this method is limited by the fact that estimating precisely small probabilities requires an extremely long simulation [47], which is too costly for most complex models. Various approaches have been suggested to address this problem, like rare events algorithms [48–50] or ensemble boosting [51].

Recently, [52] proposed a physics-based theory to address the question of the upper bound for extreme temperatures over midlatitude continental areas. They provide evidence for the hypothesis that moist convective instability sets a fundamental upper limit for surface temperatures. This allows for an estimate of maximal temperatures based on large scale circulation of the free-troposphere and surface specific humidity. In this paper, we make use of this theory to investigate the maximal reachable temperature in Western Europe in the current climate. We use the flow analogues attribution method [53–55]. We condition the estimate of maximal temperatures on the free-troposphere large scale circulation observed during the most intense event of the recent years in Western Europe, namely the July 2019 heatwave. 2019 saw two exceptional heatwaves in June and July [56] with maximum temperatures very unlikely in a world without anthropogenic climate change [45, 57].

We focus on the July 2019 heatwave—when most records of Western Europe cities were broken—and we estimate the maximal temperatures this event could have reached in the present climate. We also investigate how these upper bounds have changed with respect to the past and detail the physical mechanisms related to the changes in daily maximum of surface temperatures.

This paper is organized as follows: in section 2, we present the data used, the computation of the heatwave maximal temperature and the flow analogs methodology. Section 3 presents and discusses the results. Finally, the conclusions drawn from our analysis are presented in section 4.

2. Data and methods

We use the ERA5 reanalysis data set of the European Center for Medium Range Weather Forecasting (ECMWF) over the period 1940–2021 [58]. The data have a horizontal resolution of $0.25° \times 0.25°$. For illustration purposes, we additionally use daily temperature data from the Paris Montsouris weather station taken from the Infoclimat website (https://www.infoclimat.fr/climatologie/globale/paris-montsouris/07156.html).

To study how more intense an observed heatwave event could have been, we use the criteria developed by [52] who provided evidence for the hypothesis that convective instability fundamentally limits maximal surface air temperatures over midlatitude land. This hypothesis states that, during heatwaves, surface air temperature T reaches its maximal value when the air column is neutrally stratified. Accordingly, the heatwave stops when precipitation is triggered by the convective instability of the boundary layer. This happens when the surface air parcel moist static energy

$$\mathrm{MSE} = c_p T + L_v q + g z_s \tag{1}$$

is equal to the free-tropospheric saturation moist static energy MSE_a^*. Here c_p is the specific heat of air at constant pressure, L_v the latent heat of vaporization, q the specific humidity at the surface, g the gravitational acceleration and z_s the geopotential height of the surface. We choose to define the free-tropospheric situation as the state of the atmosphere at 500 hPa and MSE_a^* is therefore computed from equation (1) by

replacing T with the air temperature at 500 hPa T_{500}, q with the saturation specific humidity at T_{500}:

$$q_{\text{sat}}(T_{500}) \simeq \frac{\epsilon e_{\text{sat}}(T_{500})}{500\,\text{hPa}} \tag{2}$$

where ϵ is the molar ratio between water vapor and dry air and $e_{\text{sat}}(T_{500})$ is the saturation vapor pressure given by the Clausius–Clapeyron relation—and z_s with the geopotential height at 500 hPa z_{500}. The maximal surface temperature is therefore:

$$T_{s,\max} = T_{500} + \frac{L_v}{c_p}(q_{\text{sat}}(T_{500}) - q) + \frac{g}{c_p}(z_{500} - z_s). \tag{3}$$

This temperature is maximized if q is null. However, such a low specific humidity is unrealistic for midlatitude land because of the presence of surface waters and vegetation. In the following, except specified explicitly, we therefore use the hourly June–July–August–September (JJAS) minimum $q_{\min}$ ever observed over the period 1940–2021 at each grid point. Using q at 2 m to compute $T_{s,\max}$ is a conservative choice because when an air parcel rises, it entrains surrounding dryer air, which decreases its specific humidity. For T_{500} and z_{500} we use the daily mean values as the diurnal cycle does not influence strongly these quantities.

$T_{s,\max}$ defines a heatwave potential in the sense that it gives the maximal temperature potentially reachable given the synoptic situation. Whether this temperature will be reached is however not certain because, for example, the circulation pattern may change too fast for the temperature to have time to increase under the forcing of incoming energy fluxes. Letting TX1d be the observed daily temperature maximum at the surface, one can therefore define an index $\text{I}_{T\text{X1d}}$ for how much an observed heatwave is close to its maximal potential:

$$\text{I}_{T\text{X1d}} := \frac{T\text{X1d} - T_{\text{clim}}}{T_{s,\max} - T_{\text{clim}}}, \tag{4}$$

where T_{clim} is a climatological temperature and TX1d is the observed daily maximum temperature. Here we use the July averaged daily mean temperature as a climatological value. When this index is close to 1, the event is close to its maximal potential intensity: it could not have been more intense.

To determine how global warming may have impacted the physical mechanisms leading to the extremely intense event of July 2019 in Western Europe, we employ an attribution methodology. The most common attribution method of a specific event [37] consists in comparing the probabilities of an observable reaching a certain level in a counterfactual world and in a factual world. The difference between the two periods usually lies in their global mean temperature, i.e. worlds with or without anthropogenic global warming. The ratio of the probabilities gives how much more likely an observed event has been rendered because of global warming. This method however does not condition the result on the large scale circulation and therefore bears the risk of comparing atmospheric dynamics that are different even though they give rise to the same values for the observable of interest.

Here we make use of the flow analogues attribution methodology proposed by [55] and adapted from [53] (see also [59, 60]). The idea of the method is to compare the expectation of a variable of interest X in the counterfactual world ($F = 0$) and in the factual world ($F = 1$), conditional on the large scale circulation $\mathcal{C}(\zeta)$ of the observed event ζ:

$$\Delta_{\mathcal{C}(\zeta)}X = \mathbb{E}[X \mid F = 1, \mathcal{C}(\zeta)] - \mathbb{E}[X \mid F = 0, \mathcal{C}(\zeta)]. \tag{5}$$

In our analysis, ζ corresponds to the 25 July 2019 situation, the day where the heat peak was reached. The conditioning on the large scale circulation $\mathcal{C}(\zeta)$ should be understood as being in a reasonable vicinity of the circulation of the event rather than as a strict equality, which cannot be enforced. This framework allows to answer the question: how a similar large scale circulation pattern in the two worlds leads to different outcomes in an observable of interest? If the difference $\Delta_{\mathcal{C}}X$ is statistically significant, then one can say that in the factual world the event has been rendered more (or less) intense by $\Delta_{\mathcal{C}}X$. This method allows to disentangle the role of thermodynamics and dynamics (see [61] for the 2019 event). Whether the event has become more likely can in principle be obtained [53] if one can estimate the probabilities $\mathbb{P}[\mathcal{C}(\zeta) \mid F]$ and $\mathbb{P}[\mathcal{C}(\zeta) \mid X > x]$, where x is a given threshold, in the two worlds.

This method can be used with climate models output with or without anthropogenic forcing to define the factual and counterfactual worlds. However, models have known deficiencies, including biases [62] and incorrect dynamics of extremes under forcing over Western Europe [40, 45, 46]. As in [55] we therefore choose to rely on reanalysis outputs only, which, even though they also present limitations, are the closest we have to the actual past state of the atmosphere. Here, we define the counterfactual world as the past period (1940–1980) and the factual world as the present period (1981–2021).

To condition on the synoptic circulation, we select in both periods the best 40 analogs as the daily mean 500 hPa geopotential height (z_{500}) maps minimizing the pointwise Euclidean distance with the daily mean z_{500} map of the target day ζ. In the factual period we also impose that no analog can be found within a 20 calendar days window around the

25 July 2019 to avoid selecting the event itself. The reason for choosing to condition on z_{500} is that in the quasi-geostrophic approximation relevant for synoptic scales, z_{500} acts as an approximate streamfunction for the general circulation and has been shown to be a good conditioning field for temperatures [54]. Over the period 1940–2021, the troposphere has warmed and geopotential heights may therefore have inflated. To avoid this factor to bias the analogs selection, we remove a spatially uniform trend to the geopotential heights before computing the Euclidean distance. This ensures that the spatial gradients are preserved and therefore that analogs have a similar synoptic circulation. To compute the trend we average spatially the z_{500} maps and fit a second degree polynomial to this time series. The second order was chosen because of the so-called warming pause in the global temperature of the Earth in the period 1950–1970 [63]. The analogs are computed over the domain depicted with a dashed box in figure 1 panel (c). This domain is chosen because it encompasses the observed heatwave region at the surface and the relevant associated synoptic pattern, namely a large poleward meander of the jet stream associated with a strong anticyclonic anomaly over Western Europe. The analogs are chosen only in the months June, July, August and September to ensure closeness with the event. We verified that the results are not qualitatively sensitive to the choice of the number of analogs provided that we extract 20 or 30 analogs maps.

Forty-one years long factual and counterfactual periods ensure a large enough statistical sample of z_{500} maps to be able to select sufficiently representative analogs over such a limited geographical area. The counterfactual period can be assumed to satisfy the assumption of climate stationarity—for example with regards to the mean (not shown) or extreme (see e.g. figure 1 panel (a)) temperatures in summer—with respect to anthropogenic climate change. This period is also long enough for the interannual and interdecadal natural variability of the atmospheric circulation to be averaged over. The same is not necessarily true for the factual period for which there is a strong warming of Western Europe between the 1980 s and the 2010 s decades. A direct identification of the factual period and a stationary warmer world should therefore be made with care and it is more relevant to frame our results as showing a difference between a past world and a present world [64].

Once obtained the two sets of analogs for the factual and counterfactual periods, we average them out to search for significant differences in observables X of interest. To determine significant changes, we apply a two-tail Welch t-test [65], with different variances, at each grid point. We mark as significant only grid point changes for which the p-value of the test is below 0.05. In the following, as the reference event takes place in July we chose to display anomalies with respect to the July monthly mean and standard deviation over the period 1990–2021. Apart when specified explicitly, this convention concerns all anomalies.

3. Results and discussion

3.1. Description of the event and intensity potential

In the end of July 2019, Western Europe was hit by a heatwave which broke many records in France, Belgium, the Netherlands and Western Germany ([45], figure 1 panel (b)). The heatwave peaked on the 25 July and, for instance, the Paris Montsouris station recorded its largest temperature since 1900 with a new record of 42.6 °C, breaking the previous 1940 record by a margin of 2.2 °C (figure 1 panel (a)). The synoptic situation was characterized by a strongly meridional meander of the high level jet (figure 1 panel (c)) sustaining advection of warm air and anomalies of up to 7 K at 500 hPa above the North Sea (figure 1 panel (b)). Extreme surface temperatures were reached after an anomalously dry and hot month (30% of climatological precipitations and +2.4 °C temperature anomaly for the Paris region) following the previous June heatwave [56], which made the event develop under desiccated soils (figure 1 panel (d)). The event however only lasted a few days and daily maximum temperatures decreased by more than 10 K on the 26 July.

Figure 2 panel (a) shows the difference between the maximal surface air temperature $T_{s,\max}$ and the observed daily maximum temperature $TX1d$ for the 25 July 2019. To ensure realism in the value of $T_{s,\max}$, we use the hourly JJAS minimum specific humidity $q_{\min}$ ever observed over the period 1940–2021 at each grid point. Over the core heatwave region (Northern France and Benelux), the daily maximum temperature provided by the reanalysis data set is close to the maximal potential intensity by a margin of 4–6 K, and consequently the heatwave potential index is close to 90%.

Table 1 shows the estimation of $T_{s,\max}$ for several major urban areas in Western Europe which all set their 1940–2021 maximum temperature record during the July 2019 event. For all of them, we used a four grid point average encompassing most of the main city and its surroundings. We compare the estimated maximal surface temperature to the maximum value of a non-stationary GEV distribution fitted on the summer temperature maxima over the 1940–2021 period using the ERA5 reanalysis data set (see supplementary materials for more information). We emphasize that these values are conservative with respect to what can be actually measured by a weather station at the surface because they are averaged over a roughly 50 km by 50 km square. For example, the value for $TX1d$ for the Paris region with ERA5 gives 41.4 °C whereas the Paris Montsouris station measured 42.6 °C.

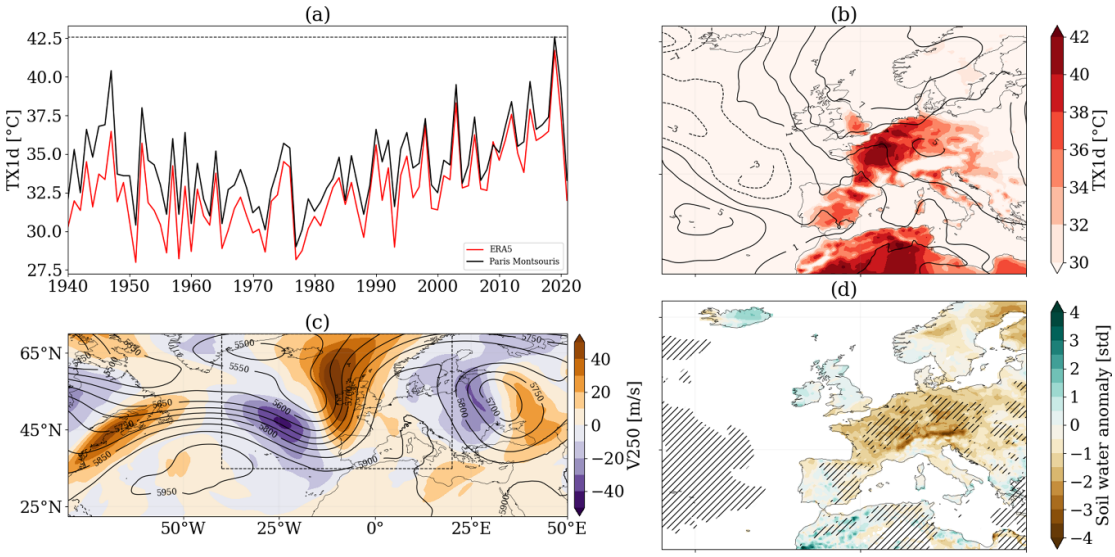


Figure 1. (a) Annual maximum temperature for the Paris Montsouris station (black) and the grid point where the station is located for ERA5 (red). The black dashed line corresponds to the temperature reached on the 25 July 2019 at the Paris Montsouris station. (b) Daily maximum 2 m air temperature (colors) and daily mean 500 hPa temperature anomalies (contours). (c) Daily mean meridional wind at 250 hPa (colors) and geopotential height at 500 hPa (contours). (d) Daily mean normalized soil moisture anomaly (colors) and hourly mean 2 m specific humidity at the time when maximum 2 m air temperature is reached (hashed). The hashed regions represent anomalies below −0.5 standard deviation. All maps are drawn for the 25 July 2019.

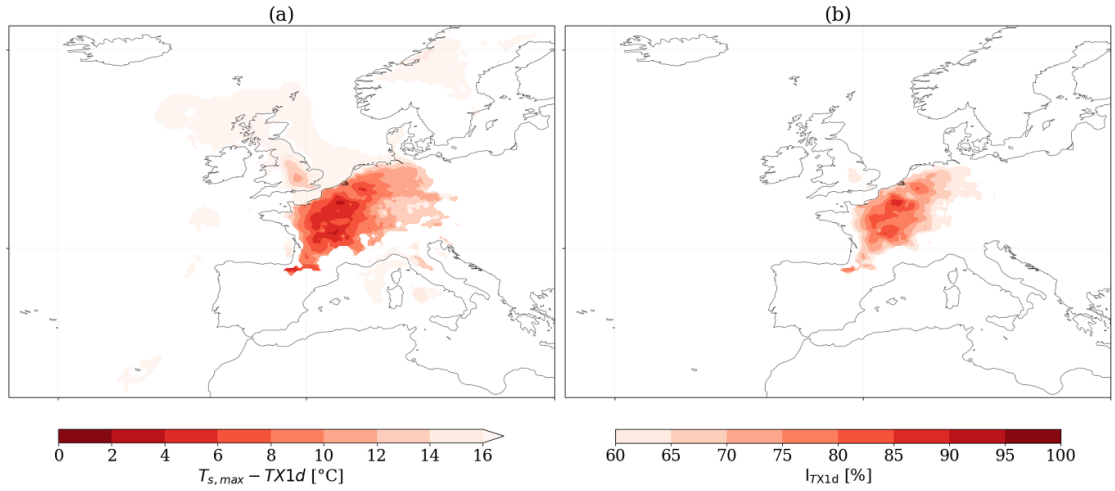


Figure 2. (a) Difference between the maximal air surface temperature $T_{s,\max}$ and the observed daily maximum temperature $TX1d$ for the 25 July 2019. (b) Heatwave potential index for the 25 July 2019. We show land grid points where the daily maximum temperature on the 25 July 2019 exceeded the 99th percentile of daily maximum temperature in July over the period 1990–2021. $T_{s,\max}$ is computed using the hourly summer minimum specific humidity $q_{\min}$ ever observed over the period 1941–2021 at each grid point.

Strikingly, even in the current climate, it is not possible to rule out the possibility of reaching 50 °C in all of these cities once we include the uncertainty on the minimum surface specific humidity. For example, for the Paris area the ERA5 dataset provides a daily maximum of 41.4 °C, while $T_{s,\max}$ is 46.6 °C [45.2, 53.8]. Reaching such temperatures would imply either a very intense drought, with very low values of surface specific humidity, and/or massive advection of heat from neighbouring regions. The non-stationary GEV estimate for the upper bound is significantly lower than $T_{s,\max}$ by 5 to 10 K. The median estimate is even lower than the 2019 value for Paris and Lille. For most cities, the estimated $T_{s,\max}$ is inside the uncertainty range of the GEV estimate, but much closer to the upper bound than to the lower one. This suggests that the statistical analysis of past time series of maximum temperatures underestimates the risk of reaching extreme temperatures.

3.2. Flow analogs analysis

Figure 3 panel (a) shows the difference between the factual and counterfactual periods of the analogs averaged daily maximum 2 m air temperature $TX1d$

Table 1. Maximal temperatures in °C for several urban areas in Western Europe. All of them reached their 1940–2021 maximum temperature record during the July 2019 event. See supplementary materials for the detail of the computation of the uncertainty range.

City	*TX*1d (ERA5)	$T_{s,\max}$ (realistic)	$T_{s,\max}$ (absolute)	GEV estimate
Paris	41.4	46.6 [45.2, 54.6]	54.2 [53.8, 54.6]	40.8 [40.0, 47.2]
Lille	39.7	47.1 [45.0, 55.0]	54.9 [54.6, 55.0]	39.2 [38.3, 48.5]
Bruxelles	38.9	47.1 [45.4, 54.9]	54.8 [54.5, 54.9]	42.8 [37.6, 49.7]
Amsterdam	35.6	47.8 [44.7, 56.2]	55.9 [55.7, 56.2]	37.3 [33.7, 41.5]
Köln	40.2	48.4 [45.9, 55.5]	55.3 [55.0, 55.5]	42.1 [39.6, 48.4]
Franckfurt	39.3	50.1 [46.4, 56.8]	56.6 [56.3, 56.8]	42.2 [39.3, 49.2]
London	35.6	48.6 [46.9, 56.0]	55.8 [55.6, 56.0]	42.9 [36.3, 51.0]
Luxembourg	37.8	45.4 [45.2, 54.1]	53.8 [53.5, 54.1]	41.1 [37.6, 47.5]

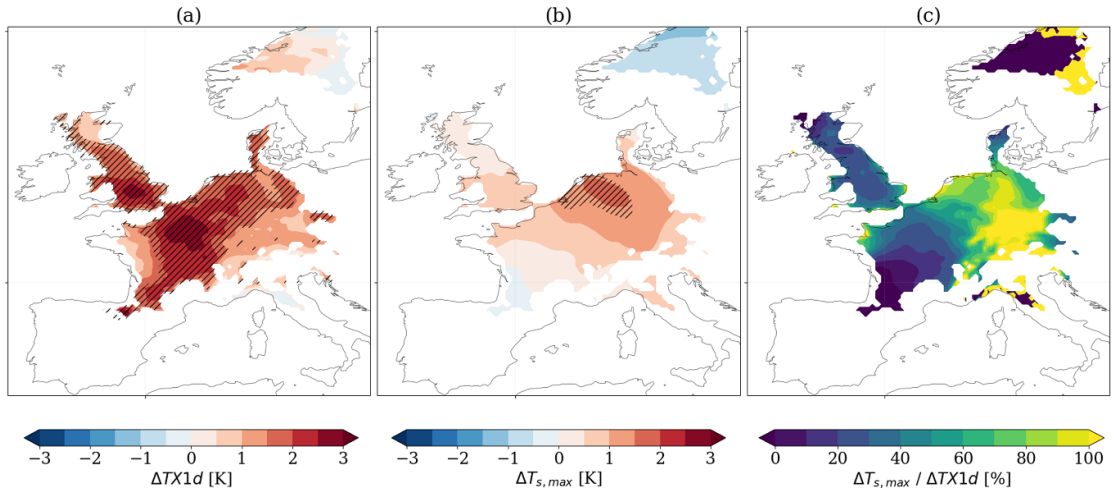


Figure 3. Difference between the factual (1981–2021) and counterfactual (1940–1980) analogs averaged (a) daily maximum 2 m air temperature *TX*1d and (b) absolute maximal surface temperature $T_{s,\max}$. Hashed grid points correspond to a significant difference at the 5% level. (c) Percentage of change in *TX*1d explained by the change in $T_{s,\max}$ between the two periods. For all panels we show grid points where the maximum temperature on the 25 July 2019 exceeded the 99th percentile of daily maximum temperature in July over the period 1990–2021 (heatwave event region).

over Western Europe. In the heatwave region (France, England, Benelux and Western Germany) *TX*1d is between 2 and 3 K more intense in the recent period. It should also be noted that this analogs difference is stronger by approximately 1 K than the difference in the daily mean temperature (see figure C1 in supplementary materials).

Figure 3 panel (b) displays the difference of the absolute maximal surface temperature $T_{s,\max}$. This quantity has not changed significantly over most of Western Europe—except a small region above the Netherlands which coincides with an increase of T_{500} (not shown). Figure 3 panel (c) shows the percentage of change in *TX*1d explained by the increase in $T_{s,\max}$. The latter corresponds to no more than 30% of the actual change between the two period over France (Paris region) but as much as 70% over Benelux and Western Germany.

The change in T_{500} is the dominant factor for the stronger increase in absolute $T_{s,\max}$ over the Netherlands, with an increase of 0.7 K, whereas only 0.4 K over Paris (see vertical profiles in figure C2). There is also a more significant drying of surface air over the Netherlands (figure C2 panels (b) and (d)). These two elements lead to a much stronger increase in the analogs averaged lifting condensation level (LCL) between the two regions: 21 hPa over the Netherlands vs 14 hPa over Paris, although the difference between the two periods is non significant at the 5% level. Dryer air parcels in a hotter troposphere follow the dry adiabatic temperature profile longer, which leads to increased surface temperature as can be seen when comparing white and magenta lines in figure C2 panels (a) and (c). These elements suggest that under the synoptic atmospheric conditions associated with the 25 July 2019 event, most of the observed change in 2 m air temperature maximum in France between the two periods is caused by surface processes—either local diabatic heating or advection from neighboring regions.

Figure 4 shows the analogs analysis for *TX*1d and z_{500} (first row), surface specific humidity q and T_{500} (second row). In both periods the averaged z_{500} pattern matches the event z_{500} pattern (figure 1 panel (c)) and the difference between the periods is minimal as can be seen by the fact that there are almost no significant differences in the z_{500} field (figure 4 panel (c)). We are therefore confident on the quality of the analogs and the relevance of the comparison

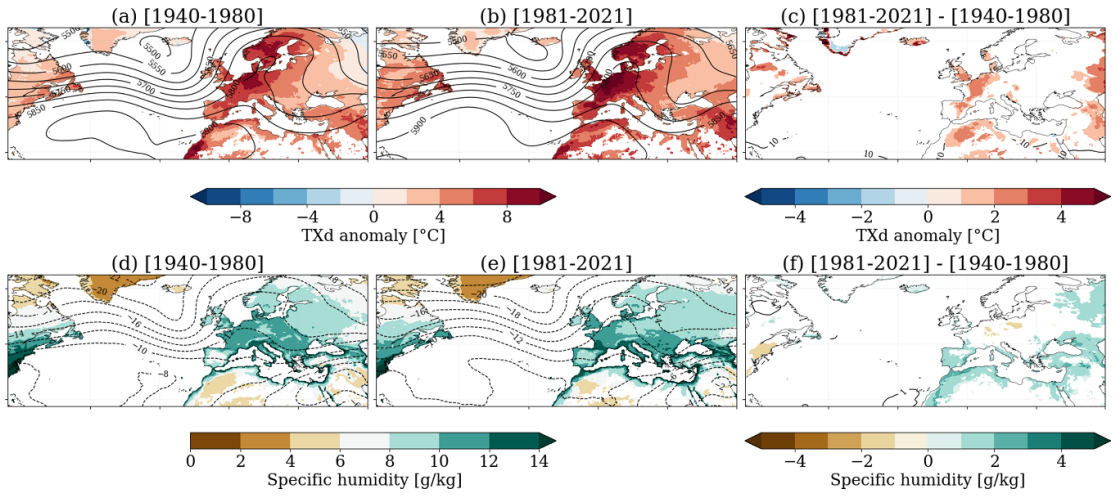


Figure 4. First row: analogs averaged *TX*1d anomalies (colors) and daily mean z_{500} (contours) for (a) the counterfactual and (b) the factual periods. (c) Difference between panels (a) and (b). For *TX*1d the anomalies are computed with respect to the 1990–2021 July monthly mean 2 m air temperature climatology. Second row: analogs averaged hourly mean 2 m specific humidity at the time when *TX*1d is reached (colors) and daily mean T_{500} (contours) for (d) the counterfactual and (e) the factual periods. (f) Difference between panels (d) and (e). In the last column, for all fields only the grid point with a significant difference at the 5% level are displayed.

between mean observables conditional on the synoptic circulation.

Figure 4 panel (f) shows no significant difference in the T_{500} fields between the two periods. This should however be taken with care because in equation (3) this variable is used to compute $e_{\text{sat}}(T_{500})$ which evolves strongly non-linearly with T_{500}. A small difference in T_{500}, even non significant, can therefore lead to significant $T_{s,\max}$ differences at the surface as shown in figure 3 panel (b). Finally, over the heatwave region the difference in specific humidity is not significant (figure 4 panel (f)), except over a small area in Western Germany.

In figure C3 in supplementary materials we present the same plot for upper layer soil moisture, which is a variable closely related to 2 m specific humidity (low soil moisture reduces the evapotranspiration potential). There is a general drying pattern over most of the heatwave region in the factual period, but it is significant only over the same region as specific humidity in figure 4 panel (f).

The flow analogs method for these two variables is however less robust because soil moisture may show a much slower dynamics than z_{500}. The distribution of soil moisture (and therefore specific humidity) is likely to depend on the full trajectory followed by the atmospheric dynamics rather than its last state. The association between the synoptic circulation $\mathcal{C}(\zeta)$ of the 25 July 2019 and dry soils/low surface specific humidity may therefore be a coincidence, especially for such a short extreme. As a consequence, it is possible that the framework presented in equation (5) may not be relevant for these variables.

Figure C4 in supplementary materials indeed shows that the JJAS daily minimum of hourly mean specific humidity at 2 m displays a significant increase between the two periods for Western Europe (figure C4 panel (c)), probably as a result of moisture advection from nearby oceans [66]. This is apparently in contradiction with the analogs analysis presented here only if one dismisses the conditioning on the z_{500} pattern. Nonetheless, whether the soil desiccation pattern observed between the two periods is a relevant feature of the change in the physical mechanisms leading to extreme temperatures in the two periods can be questioned.

Previous literature showed that there is a decrease of surface specific humidity during the hottest days [56, 67–69], but this may arise as both a cause and a consequence of intensifying surface temperature extremes [70]. We note also that although soil moisture is well represented in ERA5, the temporal drying trend may be underestimated [71].

Figure 5 shows the $T_{s,\max}$ indicator computed in both periods using either the actual specific humidity at the surface q or the 1940–2021 minimum specific humidity $q_{\min}$. As already shown above, figure 5 panel (f) shows a significant change in absolute $T_{s,\max}$ only over a limited region, which cannot explain the intensity change in *TX*1d between the two periods. Moreover, if one takes into account the moistening of minimum surface specific humidity over Western Europe in the factual period, this small difference may disappear completely as it would decrease $T_{s,\max}$ in the factual period. When one uses the actual q of the analogs, the difference is much stronger between the two periods: there is an increase of more than 3 K of $T_{s,\max}$ over a large region encompassing most of Germany. This region however does not coincide perfectly with the region with a significant change in *TX*1d in figure 4 panel (c).

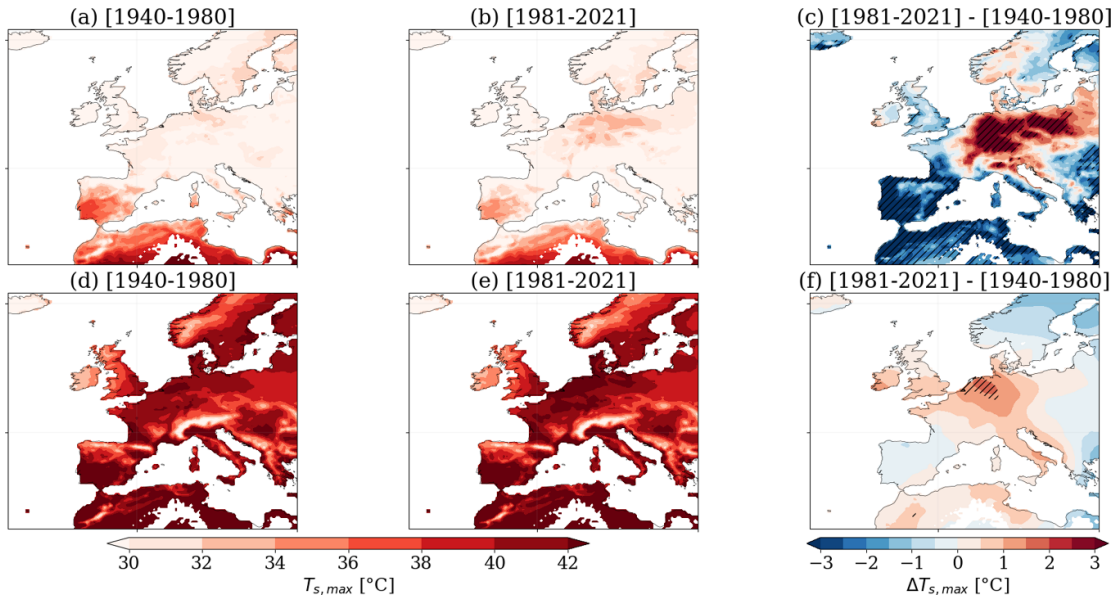


Figure 5. First row: analogs averaged $T_{s,\max}$ computed using the specific humidity of the analogs when $TX1d$ is reached for (a) the counterfactual and (b) the factual periods. (c) Difference between panels (a) and (b). Second row: analogs averaged $T_{s,\max}$ computed using the minimum specific humidity over the 1940–2021 period for (d) the counterfactual and (e) the factual periods. (f) Difference between panels (d) and (e). In the last column, the hashed grid points correspond to a significant difference at the 5% level.

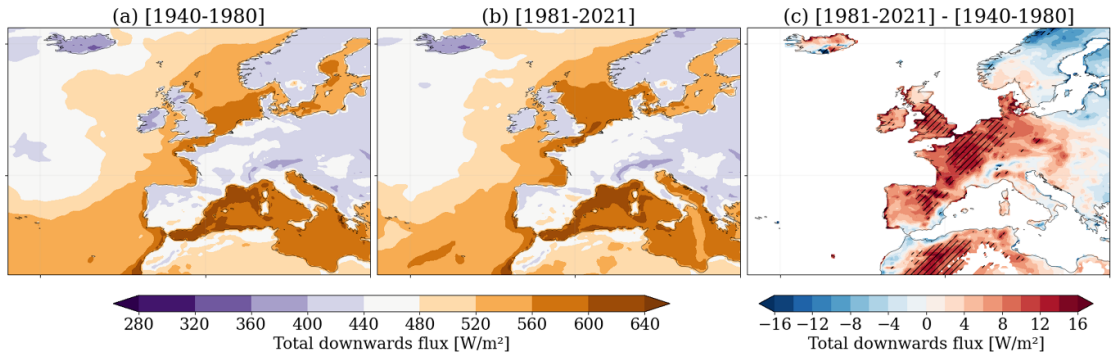


Figure 6. Analogs averaged of the sum of downwards energy fluxes at the surface for (a) the counterfactual and (b) the factual periods. (c) Difference between the two periods. In panel (c), the hashed grid points correspond to a significant difference at the 5% level.

To investigate this question, we analyze incoming energy fluxes. Even though the heatwave potential intensity has not changed much between the two periods, the actual capacity of a synoptic circulation to reach this maximal intensity may be greater in the factual period. Figure 6 displays the analogues average of the sum of daily mean incoming energy fluxes at the surface (downward positive). It includes latent and sensible heat fluxes, in addition to net shortwave radiation and downward longwave radiation. Figure 6 panel (c) shows that there is a significant and strong difference in incoming energy fluxes between the two periods which coincide with the heatwave region where the difference in $TX1d$ is significant. These fluxes are strongly coupled to the boundary layer and soils states—especially with regards to the partitioning between sensible and latent heat. Therefore one cannot claim that energy fluxes are a fully external cause of the increase of $TX1d$ between the periods. However, they are an indicator of the strong increase in local diabatic heating.

The observed increase in $TX1d$ could also be the result of increased sensible heat advection from neighboring regions with increasingly desiccated land surface [24, 72]. It is possible that this phenomenon of non-local self-propagation of dry and hot conditions played an important role for the 25 July 2019 event as shown by [56], especially with regards to the aftermath of the June heatwave which desiccated soils in the south of France and the Iberian peninsula. However, it is unlikely to be the case for all of the 40 analogues in both periods: as shown by [73], in Europe, hot events are generally associated with weaker horizontal transport, but strong adiabatic

warming and local temperature increase caused by enhanced surface fluxes.

We emphasize that the results presented here are very sensitive to changes in surface specific humidity. This variable has significantly increased in the factual period, which according to equation (3), should compensate the increase in T_{500} due to mean tropospheric warming with climate change. If we take the example of the Paris grid point, the change between the two periods in $T_{s,\max}$ due to the change in analogs averaged T_{500} is equal to +0.58 K. On the contrary, if we assume that minimum specific humidity is expected to shift by the same amount as the mean difference between the distribution of daily minimum of hourly mean specific humidity ($+0.52\,\mathrm{g\,kg^{-1}}$), then one could expect a change of $T_{s,\max}$ by −1.29 K, which would largely compensate the tropospheric warming. This simple reasoning however assumes that the minimum of specific humidity during the hottest days would increase as the rest of the distribution. This is in opposition to the analogs analysis presented here which shows in contrast drying during the days associated with a synoptic circulation similar to the 2019 event. This analysis therefore supports the idea that the increase in the variability in temperatures, especially at the tail, is the result of increased variability in soil moisture and surface specific humidity. As a consequence, under the appropriate atmospheric conditions—namely a large anticyclonic anomaly—the response in daily maxima of surface temperature is stronger in the recent period.

4. Conclusion

We investigated the maximal reachable temperatures in Western Europe in the current climate. We took the July 2019 record-shattering heatwave event as a paradigmatic example and investigated how the maximum temperatures reached during this event could have been higher. We found that even in the current climate, we cannot rule out the possibility that 50 °C cannot be reached in most urban areas of Western Europe. With a similar atmospheric pattern as the 2019 event, this would need a massive and large scale drought, with levels of surface specific humidity lower than the lowest observed in the last 80 years.

We then employed a flow analogues method to investigate how the upper bound of surface temperatures has changed between the past (1940–1980) and the present (1981–2021). We demonstrated that in the present period, the conditional expectation of daily maximum 2 m air temperature has increased significantly. This increase is stronger by around 1 K than the increase in the daily mean temperature. The change in the expected heatwave intensity potential $T_{s,\max}$ explains part of this increase over the Netherlands and Western Germany where a small increase in T_{500} combined with dry soils to increase $T_{s,\max}$ by more than 3 K in the recent period. The same does not apply over France and Belgium, where the change in $T_{s,\max}$ is small and insignificant with limited and non significant surface drying and no T_{500} difference. However, over these regions, incoming energy fluxes at the surface have increased strongly in the present period which may explain why higher $TX1d$ are reached compared to the past period even though the intensity potential has not changed. Our results suggest that most of the increase in observed daily maximum surface temperature may result from strengthened surface diabatic warming processes rather than free troposphere warming.

What would be the impact of a similar weather pattern as the July 2019 event in the future has been investigated by [74] under various warming scenarios and using a storyline approach. They found that an analogous event would entail peak temperatures around 50 °C in Central Europe under a high emission scenario. Combined with our results, these elements show that the past extreme temperatures are not a good guide of the worst case scenario under global warming. The fast increase of maximum surface temperatures already observed demands a massive update of adaptation strategies.

Data availability statement

The data that support the findings of this study are available upon request from the authors. ERA5 reanalysis data are available on the Copernicus website (https://cds.climate.copernicus.eu/cdsapp#!/home).

Acknowledgments

The authors would like to thank Robert Vautard and Fabio D'Andrea for fruitful discussions, and the two anonymous reviewers for their valuable comments.

Funding

The authors acknowledge the support of the Grant ANR-20-CE01-0008-01 (SAMPRACE: PY). This work has received support from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 101003469 (XAIDA) and by the Marie Sklodowska-Curie Grant Agreement No. 956396 (EDIPI). R N was supported by a doctoral Grant from CEA.

Conflict of interest

The authors declare no conflict of interest.

Code availability

The main results of this work were obtained using Python. The scripts are available upon request.

Authors' contributions

R N did the data collection and analysis. All authors contributed to the design of the analysis and the writing of the paper.

Appendix A. Non-stationary GEV estimation

If we consider the maximum $M_n = \max\{X_1, \ldots, X_n\}$ of n i.i.d. random variables $X_1, \ldots, X_n$, the extremal theorems [33, 34] state that if there exist sequences $a_n > 0$ and b_n so that the distribution $\mathbb{P}(\frac{M_n - b_n}{a_n} \leqslant z)$ converges towards $G(z)$ when n goes to infinity, then $G(z)$ is given by:

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\} & \text{if } \xi = 0 \end{cases} \tag{A.1}$$

where μ is called the location parameter, σ the scale parameter and ξ the shape parameter. This distribution is usually called the GEV distribution. Importantly, it is bounded upwards for $\xi < 0$ and the upper bound is given by:

$$z_{\max} = \mu - \frac{\sigma}{\xi}. \tag{A.2}$$

Here we estimate a non-stationary GEV by letting the parameter μ depend on another covariate T encompassing the warming of the Earth: $\mu = \mu_1 + \mu_2 T$. We chose T to be the annual average of 2 m air temperature of North-Hemisphere over land as described in [75]. The GEV distribution is then estimated using the maximum likelihood method and we compute the upper bound using the formula (A.2).

We then report the 'upper' upper bound for T in 2021:

$$z_{\max}(2021) = \mu_0 + \mu_1 T_{2021} - \frac{\sigma}{\xi}. \tag{A.3}$$

We compute the uncertainty range over this upper bound using $n = 1000$ bootstrap samples of the initial series. For each bootstrap series, we fit a non-stationary GEV law (with T as a covariate on μ) and compute the upper bound. We take the median values of these bootstrapped estimates as the best estimate and we define the lower (resp. higher) bound of the estimate as the 97.5% and 2.5% quantiles of the bootstraped $z_{\max}$.

Appendix B. Uncertainty in $T_{s,\max}$ estimation

For computing the uncertainty range of the absolute $T_{s,\max}$ for the urban areas, we use for T_{500} and z_{500} the spread of the 10 ensemble members of the ERA5 reanalysis simulation. The lowest (resp. highest) bound of the range uses the lowest (resp. highest) T_{500} and z_{500} over the members. The ensemble spread is small therefore the uncertainty range for the absolute $T_{s,\max}$ is around 1 K.

For the computation of the more realistic $T_{s,\max}$, we also need to take into account the uncertainty in the $q_{\min}$ value. To do so, we fit a stationary GEV law over the JJAS annual minimum of hourly 2 m air specific humidity over the 82 years between 1940–2021. This allows to give an approximate value to $P(q) := \mathbb{P}(q_{\min} \geqslant q)$ where q is a fixed threshold. We compute the uncertainty range over this estimated probability using $n = 1000$ bootstrap samples of the initial series. For each bootstrap series, we fit a GEV law and compute the return level q_{82} associated with a probability $P(q_{82}) = \frac{1}{82}$, similar to the probability of the minimum observed in the ERA5 dataset. We then take the 97.5% and 2.5% quantiles of the bootstraped q_{82} as the uncertainty range for $q_{\min}$. Due to the resampling procedure, the 2.5% quantile is negative for all cities. As this is not physical, we take this lower bound to be null. We then combine the lowest (resp. highest) T_{500} and z_{500} over the members to the highest (resp. lowest) $q_{\min}$ to get the uncertainty range. Due to the large uncertainty in $q_{\min}$, the uncertainty range for the realistic $T_{s,max}$ is also large.

Appendix C. Additional figures

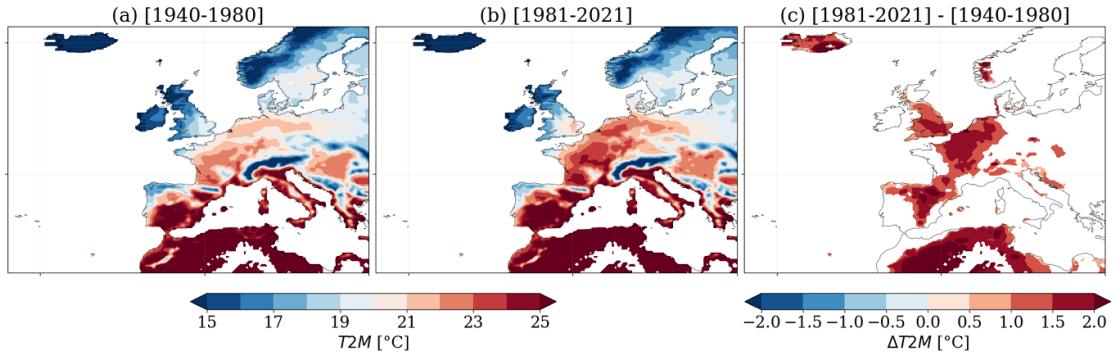


Figure C1. Analogs averaged daily mean 2 m air temperature T2M for (a) the counterfactual and (b) the factual periods. (c) Difference between the two periods of the averaged field. In panel (c), only the grid points with a significant difference at the 5% level are displayed.

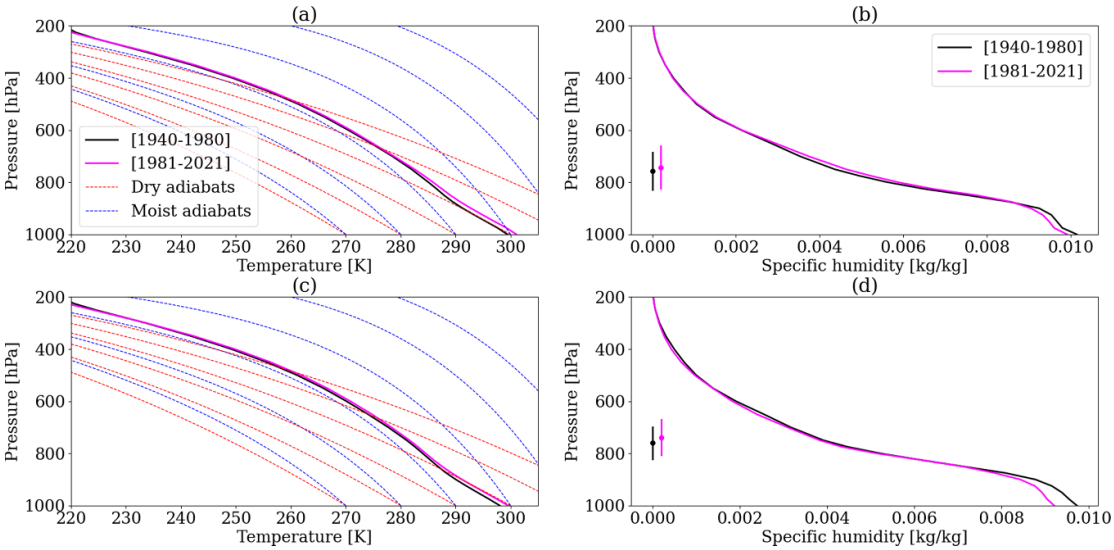


Figure C2. Vertical profiles of analogs averaged air temperature for (a) the Paris region (lat = [47.5°,49.5°], lon = [1°,5°]) and (c) the Netherlands region (lat = [51°,53.25°], lon = [5°,9°]). Vertical profiles of analogs averaged air specific humidity for (b) the Paris region and (d) the Netherlands region. The white and magenta dots show the mean lifting condensation level in the two periods. The associated vertical bars display a two standard deviation spread over the analogs.

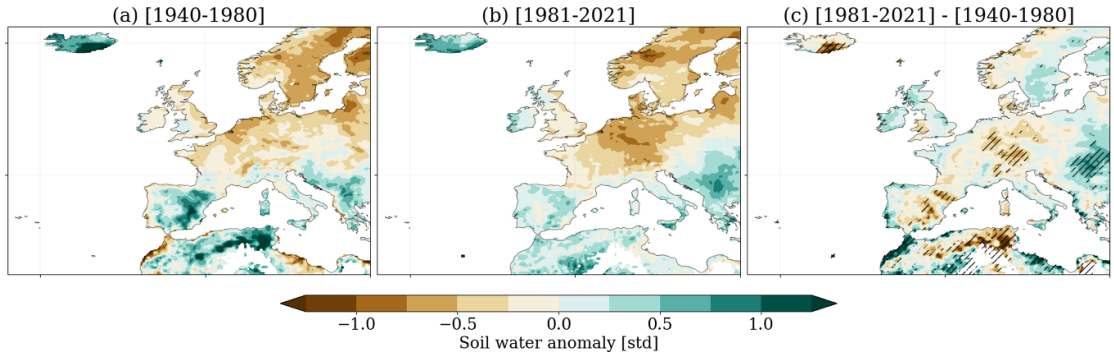


Figure C3. Analogs averaged daily mean upper layer soil moisture (10 cm) for (a) the counterfactual and (b) the factual periods. (c) Difference between the two periods of the averaged field. In panel (c), the hashed grid points correspond to a significant difference at the 5% level.

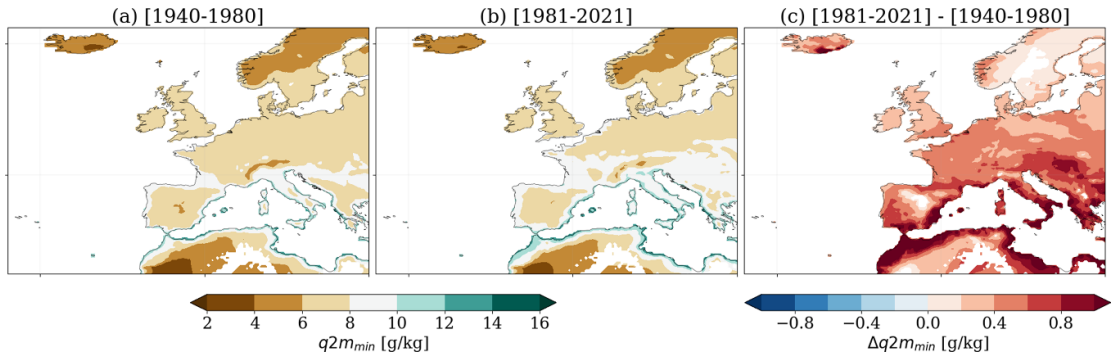


Figure C4. June-July-August-September average of daily minimum of hourly mean specific humidity at 2 m for (a) the counterfactual period and (b) the factual period. (c) Difference between the two periods of the averaged field. In panel (c), the colored grid points represent a significant difference at the 5% level.

References

[1] Pörtner H O *et al* 2022 IPCC sixth assessment report
[2] Seneviratne S I, Zhang X, Adnan M, Badi W, Dereczynski C, Di Luca A, Vicente-Serrano S M, Wehner M and Zhou B 2021 Weather and climate extreme events in a changing climate *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press)
[3] Campbell S, Remenyi T A, White C J and Johnston F H 2018 *Health Place* **53** 210–8
[4] Breshears D D, Fontaine J B, Ruthrof K X, Field J P, Feng X, Burger J R, Law D J, Kala J and Hardy G E S J 2021 *New Phytol.* **231** 32–39
[5] Huang W T K, Braithwaite I, Charlton-Perez A, Sarran C and Sun T 2022 *Environ. Res. Lett.* **17** 034017
[6] Masselot P *et al* 2023 *Lancet Planet. Health* **7** e271–81
[7] Hatfield J L and Prueger J H 2015 *Weather Clim. Extremes* **10** 4–10
[8] Brás T A, Seixas J, Carvalhais N and Jägermeyr J 2021 *Environ. Res. Lett.* **16** 065012
[9] Bastos A *et al* 2021 *Earth Syst. Dyn.* **12** 1015–35
[10] Zuo J, Pullen S, Palmer J, Bennetts H, Chileshe N and Ma T 2015 *J. Cleaner Prod.* **92** 1–12
[11] Barriopedro D, García-Herrera R, Ordóñez C, Miralles D and Salcedo-Sanz S 2023 *Rev. Geophys.* **61** e2022RG000780
[12] Perkins S E 2015 *Atmos. Res.* **164** 242–67
[13] Horton R M, Mankin J S, Lesk C, Coffel E and Raymond C 2016 *Curr. Clim. Change Rep.* **2** 242–59
[14] Domeisen D I, Eltahir E A, Fischer E M, Knutti R, Perkins-Kirkpatrick S E, Schär C, Seneviratne S I, Weisheimer A and Wernli H 2022 *Nat. Rev. Earth Environ.* **4** 1–15
[15] Petoukhov V, Rahmstorf S, Petri S and Schellnhuber H J 2013 *Proc. Natl Acad. Sci.* **110** 5336–41
[16] Coumou D, Petoukhov V, Rahmstorf S, Petri S and Schellnhuber H J 2014 *Proc. Natl Acad. Sci.* **111** 12331–6
[17] Kornhuber K, Coumou D, Vogel E, Lesk C, Donges J F, Lehmann J and Horton R M 2020 *Nat. Clim. Change* **10** 48–53
[18] Di Capua G, Sparrow S, Kornhuber K, Rousi E, Osprey S, Wallom D, van den Hurk B and Coumou D 2021 *npj Clim. Atmos. Sci.* **4** 1–14
[19] Seneviratne S I, Lüthi D, Litschi M and Schär C 2006 *Nature* **443** 205–9
[20] Fischer E M, Seneviratne S I, Lüthi D and Schär C 2007 *Geophys. Res. Lett.* **34** L05707
[21] Seneviratne S I, Corti T, Davin E L, Hirschi M, Jaeger E B, Lehner I, Orlowsky B and Teuling A J 2010 *Earth-Sci. Rev.* **99** 125–61
[22] Hirschi M, Seneviratne S I, Alexandrov V, Boberg F, Boroneant C, Christensen O B, Formayer H, Orlowsky B and Stepanek P 2011 *Nat. Geosci.* **4** 17–21
[23] Miralles D G, Van Den Berg M, Teuling A and De Jeu R 2012 *Geophys. Res. Lett.* **39** L21707
[24] Miralles D G, Teuling A J, Van Heerwaarden C C and Vilà-Guerau de Arellano J 2014 *Nat. Geosci.* **7** 345–9
[25] Rasmijn L, Van der Schrier G, Bintanja R, Barkmeijer J, Sterl A and Hazeleger W 2018 *Nat. Clim. Change* **8** 381–5
[26] Dirmeyer P A, Balsamo G, Blyth E M, Morrison R and Cooper H M 2021 *AGU Adv.* **2** e2020AV000283
[27] Petoukhov V, Petri S, Rahmstorf S, Coumou D, Kornhuber K and Schellnhuber H J 2016 *Proc. Natl Acad. Sci.* **113** 6862–7
[28] Kornhuber K, Petoukhov V, Karoly D, Petri S, Rahmstorf S and Coumou D 2017 *J. Clim.* **30** 6133–50
[29] Mann M E, Rahmstorf S, Kornhuber K, Steinman B A, Miller S K and Coumou D 2017 *Sci. Rep.* **7** 1–12
[30] Mann M E, Rahmstorf S, Kornhuber K, Steinman B A, Miller S K, Petri S and Coumou D 2018 *Sci. Adv.* **4** eaat3272
[31] Kornhuber K and Tamarin-Brodsky T 2021 *Geophys. Res. Lett.* **48** e2020GL091603
[32] Ghil M *et al* 2011 *Nonlinear Process. Geophys.* **18** 295–350
[33] Fisher R A and Tippett L H C 1928 Limiting forms of the frequency distribution of the largest or smallest member of a sample *Mathematical Proceedings of the Cambridge Philosophical Society* vol 24 (Cambridge University Press) pp 180–90
[34] Gnedenko B 1943 *Ann. Math.* **44** 423–53
[35] Coles S, Bawa J, Trenner L and Dorazio P 2001 *An Introduction to Statistical Modeling of Extreme Values* vol 208 (Springer)
[36] Cheng L, AghaKouchak A, Gilleland E and Katz R W 2014 *Clim. Change* **127** 353–69
[37] National Academies of Sciences Engineering and Medicine 2016 *Attribution of Extreme Weather Events in the Context of Climate Change* (The National Academies Press)
[38] Philip S *et al* 2020 A protocol for probabilistic extreme event attribution analyses *Adv. Stat. Climatol. Meteorol. Oceanogr.* **6** 177–203
[39] Van Oldenborgh G J *et al* 2021 *Clim. Change* **166** 13
[40] Van Oldenborgh G J, Wehner M F, Vautard R, Otto F E, Seneviratne S I, Stott P A, Hegerl G C, Philip S Y and Kew S F 2022 *Earth's Future* **10** e2021EF002271
[41] Auld G, Hegerl G and Papastathopoulos I 2021 Adv. Stat. Clim. Meteorol. Oceanogr. **9** 45–66
[42] Mitchell D, Kornhuber K, Huntingford C and Uhe P 2019 *Lancet Planet. Health* **3** e290–2
[43] Philip S Y *et al* 2022 *Earth Syst. Dyn.* **13** 1689–713
[44] Thompson V, Kennedy-Asser A T, Vosper E, Lo Y E, Huntingford C, Andrews O, Collins M, Hegerl G C and Mitchell D 2022 *Sci. Adv.* **8** eabm6860
[45] Vautard R *et al* 2020 *Environ. Res. Lett.* **15** 094077

[46] Patterson M 2023 *Geophys. Res. Lett.* **50** e2023GL102757
[47] Wouters J and Bouchet F 2016 *J. Phys. A: Math. Theor.* **49** 374002
[48] Ragone F, Wouters J and Bouchet F 2018 *Proc. Natl Acad. Sci.* **115** 24–29
[49] Yiou P and Jézéquel A 2020 *Geosci. Model Dev.* **13** 763–81
[50] Ragone F and Bouchet F 2021 *Geophys. Res. Lett.* **48** e2020GL091197
[51] Gessner C, Fischer E M, Beyerle U and Knutti R 2021 *J. Clim.* **34** 6619–34
[52] Zhang Y and Boos W R 2023 *Proc. Natl Acad. Sci.* **120** e2215278120
[53] Yiou P, Jézéquel A, Naveau P, Otto F E, Vautard R and Vrac M 2017 *Adv. Stat. Climatol. Meteorol. Oceanogr.* **3** 17–31
[54] Jézéquel A, Yiou P and Radanovics S 2018 *Clim. Dyn.* **50** 1145–59
[55] Faranda D, Bourdin S, Ginesta M, Krouma M, Messori G, Noyelle R, Pons F and Yiou P 2022 *Weather Clim. Dyn.* **3** 1311–40
[56] Sousa P M, Barriopedro D, García-Herrera R, Ordóñez C, Soares P M and Trigo R M 2020 *Commun. Earth Environ.* **1** 48
[57] Ma F, Yuan X, Jiao Y and Ji P 2020 *Geophys. Res. Lett.* **47** e2020GL087809
[58] Hersbach H *et al* 2020 *Q. J. R. Meteorol. Soc.* **146** 1999–2049
[59] Cadiou C, Noyelle R, Malhomme N and Faranda D 2023 *Asia Pac. J. Atmos. Sci.* **59** 83–94
[60] Faranda D, Pascale S and Bulut B 2023 *Environ. Res. Lett.* **18** 034030
[61] Xu P, Wang L, Huang P and Chen W 2021 *Atmos. Res.* **252** 105446
[62] Maraun D 2016 *Curr. Clim. Change Rep.* **2** 211–20
[63] Zhang L 2016 *Clim. Dyn.* **47** 3157–69
[64] van Haren R, van Oldenborgh G J, Lenderink G and Hazeleger W 2013 *Environ. Res. Lett.* **8** 014053
[65] Welch B L 1947 *Biometrika* **34** 28–35
[66] Chadwick R, Good P and Willett K 2016 *J. Clim.* **29** 7613–32
[67] Coffel E D, Horton R M, Winter J M and Mankin J S 2019 *Environ. Res. Lett.* **14** 084003
[68] Padrón R S, Gudmundsson L, Decharme B, Ducharne A, Lawrence D M, Mao J, Peano D, Krinner G, Kim H and Seneviratne S I 2020 *Nat. Geosci.* **13** 477–81
[69] McKinnon K A, Poppick A and Simpson I R 2021 *Nat. Clim. Change* **11** 598–604
[70] Merrifield A L, Simpson I R, McKinnon K A, Sippel S, Xie S P and Deser C 2019 *Geophys. Res. Lett.* **46** 14082–92
[71] Li M, Wu P and Ma Z 2020 *Int. J. Climatol.* **40** 5744–66
[72] Schumacher D L, Keune J, Van Heerwaarden C C, Vilà-Guerau de Arellano J, Teuling A J and Miralles D G 2019 *Nat. Geosci.* **12** 712–7
[73] Bieli M, Pfahl S and Wernli H 2015 *Q. J. R. Meteorol. Soc.* **141** 98–108
[74] Sánchez-Benítez A, Goessling H, Pithan F, Semmler T and Jung T 2022 *J. Clim.* **35** 2373–90
[75] Osborn T J, Jones P D, Lister D H, Morice C P, Simpson I R, Winn J, Hogan E and Harris I C 2021 *J. Geophys. Res.: Atmos.* **126** e2019JD032352

2.3 Perspectives

The first contribution of this work is to propose a comparison between a physical and a statistical estimate for the upper bound of surface temperatures in the mid-latitudes, taking the July 2019 event in Western Europe as a case study. I show the large difference between the two estimations, with physical upper bounds being 5–10°C above the statistical upper bounds. The fact that the median estimate of the latter is below the actual value observed in Paris in 2019 is worrying for the capacity of the GEV estimate to be able to provide a plausible worst case scenario. This casts some doubt on statements of the impossibility of an event occurring without global warming, even when the event is very extreme (Philip et al. 2021). Even though the physical upper bounds may appear unrealistic, their intensity seems plausible when compared to the 2019 event, and at least more plausible than the GEV estimates which are overpassed by actual events. This suggests that observed past events — and the statistics fitted on them — largely underestimate the worst possible events. I explore this discrepancy and propose a way to alleviate the problem in the next chapter.

The second contribution of this work is to explore the thermodynamical mechanisms leading to heat events under atmospheric conditions similar to the July 2019 event. The analogues analysis has shown that it can satisfactorily reproduce the pattern observed using analogs in both past periods (1940-1980 and 1981-2021) which allows to isolate the sole role of thermodynamics. I show that even when conditioning on similar atmospheric patterns as the one of the July 2019 event, the intensity of daily maximum temperatures has largely increased between these two periods. This increase is much larger than what would suggest the change in the thermodynamic limit provided by the upper bound of Zhang and Boos (2023), i.e. mainly by mid-tropospheric warming. I have therefore attributed this increase to strengthened surface heat fluxes between the two periods. As a consequence, it is likely that the rapid increase in the intensity of extremes in Western Europe is caused by a combination of both more favorable weather patterns (Vautard et al. 2023) and enhanced thermodynamic surface processes. The latter could be related to a northward shift of the transitional zone for soil moisture (Seneviratne et al. 2010). Whether these changes are forced by the increase of GHGs global levels — and the decrease of aerosols emission (Schumacher et al. 2024) — or whether they arise as a result of natural variability (Singh et al. 2023) is an important question for future research — especially with regards to adaptation to climate change in Europe.

Summary

Context and goals

Extreme Value Theory (EVT) is one of the most employed statistical method to estimate probabilities and return times for extreme events. For temperatures, it also gives an estimate of the upper bound for maximal values reachable. The reliability of this upper bound has been questioned by recent very intense events. The goal of this chapter is to investigate this discrepancy and give an estimate for how more intense an observed extreme heat event could have been.

Methods

I Investigate the question of the upper bound for surface temperatures in Western Europe with a statistical approach based on EVT and a physical approach based on the stability of the air column (Zhang and Boos 2023) using the July 2019 heatwave event as a case study. I also employ a flow analogues methodology to study the evolution of thermodynamical mechanisms influencing the intensity of the event for atmospheric patterns similar to those of the July 2019 event between two past periods (1940-1980 and 1981-2021).

Results

I show that the statistical upper bound provided by EVT is likely largely underestimated. The latter is 5–10°C lower than the upper bound based on the stability of the air column. For the Paris region grid points, the median estimate of the EVT-based upper bound is below the actual value observed in Paris in 2019.
With the flow analogues methodology, I also find that even when conditioning on similar atmospheric patterns as the one of the July 2019 event, the intensity of daily maximum temperatures has largely increased between the 1940-1980 period and the 1981-2021 period. This increase is much larger than what would suggest the mid-tropospheric warming observed between the two periods. I attribute this increase to strengthened surface heat fluxes.

Chapter 3

Integration of physical bound constraints to alleviate shortcomings of statistical models for extreme temperatures

3.1 Introduction

In this chapter I present a follow-up of the work presented in the previous chapter. I show in chapter 2 on a case study in Western Europe that the statistical and the physical estimates of the upper bound for surface temperatures vastly differ. The example of the Paris grid point showed that the statistical estimate probably largely underestimates the true upper bound. A recent work by Zeder et al. (2023) showed more generally using a large ensemble of a climate model and synthetic data that the GEV fits on a limited-size sample underestimate the probability of rare events — at least when using the most common fitting packages.

Here I propose to bridge the gap between the physical and statistical estimates of the upper bounds for surface temperatures. I take for granted that the GEV fits with a limited amount of data underestimates low probability events, i.e. the most extreme and impactful events. I seek to avoid this effect — i.e. also to avoid 'climatic surprises': events that are said to be impossible but that still happen (e.g. Philip et al. (2021)). To do so I propose to physically constrain the GEV fits by imposing that the upper bound of the GEV — which is computed as a function of the parameters of the distribution — is equal to an external physical estimate for this upper bound. The question was then which estimate of the upper bounds should be chosen.

The first idea was to impose a very large upper bound ($\sim 70°C$), way above the maximum air surface temperature ever observed on Earth, to ensure that climatic surprises are not possible. However I showed on synthetic GEV data that, on the

contrary, this tends to bias upwards the estimation of low probability events (see below). I therefore implement an adapted version of the upper bound proposed by Zhang and Boos (2023). Although the value of this upper bound can be discussed — in particular other mechanisms could limit the maximal surface temperatures before the convective limit is reached[1] — it gives a good first estimate of the (unknown) value one could impose to a GEV fit.

3.2 Article submitted to the *Journal of Climate*

I show here the manuscript submitted on the 29th of February 2024 to the *Journal of Climate*.

[1] I thank Jacopo Riboldi, Matthias Röthlisberger and Erich Fischer for the discussions we had on this topic.

Integration of physical bound constraints to alleviate shortcomings of statistical models for extreme temperatures

Robin Noyelle,[a] Yoann Robin,[a] Philippe Naveau,[a] Pascal Yiou,[a] and Davide Faranda[a,b,c]

[a] *Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL & Université Paris Saclay, Gif-sur-Yvette, 91191,France*

[b] *London Mathematical Laboratory, 8 Margravine Gardens London, W6 8RH, London, United Kingdom*

[c] *Laboratoire de Météorologie Dynamique/IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, 75005, Paris, France*

Corresponding author: Robin Noyelle, robin.noyelle@lsce.ipsl.fr

ABSTRACT: Heatwaves have devastating impacts on societies and ecosystems. Their frequencies and intensities are increasing globally with anthropogenic climate change. Statistical models using Extreme Value Theory (EVT) have been used for quantifying risks of extreme temperatures but recent very intense events have cast doubt on their ability to represent the tail probabilities of temperatures. Using outputs from a large ensemble of a climate model, we show that physics-based estimates of the upper-bound of temperatures in the mid-latitudes are 3–8°C higher than suggested by EVT-based models. We propose a new method to bridge the gap between the physical and statistical estimates by forcing the EVT-based models to have an upper bound coherent with the bound provided by the instability of the air column. We show that our method reduces the underestimation of tail risks while not deteriorating the performance of the statistical models on the core of the distribution of extreme temperatures.

SIGNIFICANCE STATEMENT: The usual application of extreme value theory to determine the probabilities of extreme temperatures tends to underestimate the risks. We propose a method to reduce this underestimation by imposing a physically relevant upper bound estimate based on the stability of the air column. The coupling between extreme value theory and physics knowledge alleviates many of the shortcomings of usual statistical approaches used in heatwave attribution.

1. Introduction

Heatwaves and their impacts have been the focus of extensive attention in the context of global climate change (National Academies of Sciences Engineering and Medicine 2016; Pörtner et al. 2022). There is indeed a clear signal pointing towards increasing frequency and intensity for these events worldwide (Seneviratne et al. 2021). Human bodies (Campbell et al. 2018; Breshears et al. 2021; Huang et al. 2022; Masselot et al. 2023), plants (Hatfield and Prueger 2015; Brás et al. 2021), ecosystems (Bastos et al. 2021) and infrastructures (Zuo et al. 2015) have a limited capacity to cope with extreme temperatures and can suffer large damages once certain thresholds are reached. The capacity to predict and anticipate future maximum intensities of heatwaves is therefore of primary importance for adaptation to climate change.

A simple way to quantify the risks associated with extreme temperatures is to run a long simulation with a climate model to sample more extreme events than those observed in the past. Apart from the inherent limitations of models to represent correctly the entire temperature distribution — especially at its tail (Naveau et al. 2018; Vautard et al. 2020; Van Oldenborgh et al. 2022; Patterson 2023; Vautard et al. 2023) — this method is limited by the fact that estimating precisely small probabilities requires extremely long simulations (Wouters and Bouchet 2016), which is too costly for most complex — and therefore realistic — models. Various approaches have been suggested to address this problem, like rare events algorithms (Ragone et al. 2018; Yiou and Jézéquel 2020; Ragone and Bouchet 2021) or ensemble boosting (Gessner et al. 2021; Fischer et al. 2023), but, although less costly, they still require extensive climate simulations.

An other classical approach is to measure the upper tail probabilities of temperatures distribution — typically the risk of the yearly maximum of daily maximum 2-m air temperature reaching a certain level — using results from Extreme Value Theory (EVT). EVT is a mathematical theory based on a convergence principle of the probability distribution of block maxima (Coles et al. 2001;

Beirlant et al. 2005) towards a Generalized Extreme Value (GEV) distribution. It has been used to devise statistical models for maxima (or minima) of climate variables of interest (Ghil et al. 2011; Katz et al. 2002), such as temperatures. It is extensively used by attribution methods to compare the probabilities of reaching extreme levels between a counterfactual world without climate change and a factual world with climate change (Hannart and Naveau 2018; Philip et al. 2020; Naveau et al. 2020; Kiriliouk and Naveau 2020; Van Oldenborgh et al. 2021; Worms and Naveau 2022).

A GEV distribution has three parameters: the location μ, scale σ and shape ξ parameters. Fitting such a distribution to extreme temperatures usually gives estimates of the shape parameter that are negative (Van Oldenborgh et al. 2022; Auld et al. 2021). This implies that the distribution of annual maximum temperatures is bounded upwards by a value:

$$B(\mu, \sigma, \xi) = \mu - \frac{\sigma}{\xi} \tag{1}$$

which depends on the parameters of the distribution. Observing a record temperature above this upper bound should then have a null probability, i.e. it should be impossible. The upper bound is an important quantity insofar as it provides an ultimate worst case scenario for societal adaptation (Palmer 2002; Fischbacher-Smith 2010; Scher et al. 2021). An underestimation of this quantity may therefore have daunting consequences in impact studies.

In the context of non-stationary data — as is most of climate data under climate change — it is customary to let at least one GEV parameter depend on an external covariate representing the effect of climate change, such as Global Mean Surface Temperature (GMST). Making the location parameter evolve linearly with GMST is a natural choice to take into account the shifting towards more intense values of temperatures distribution with anthropogenic global warming. One can also let the scale parameter evolve with the covariate. We emphasize that if some of the parameters of the distribution depend on GMST, then the statistical bound B will also depend on this covariate.

Recent intense heatwaves such as in 2010 in Russia (Di Capua et al. 2021), 2019 in Western Europe (Mitchell et al. 2019) and 2021 in the Pacific Northwest (Philip et al. 2021; Thompson et al. 2022) have nonetheless challenged the reliability of the estimation of the statistical upper bound by breaking it sometimes by a large margin (Fischer et al. 2021). Several issues indeed arise when applying EVT to empirical data: the pre-asymptotic nature of these data (Gomes 1984; Coles et al. 2001), the parametric hypotheses made for taking into account the non-stationarity

induced by global warming and the limited number of points to fit the distribution. For extreme temperatures, the latter is probably the most problematic insofar as one usually has no more than 70-80 historical years to fit a GEV distribution. The mis-specification of the statistical GEV model, upper bound included, may therefore lead to an underestimation of the tail risks of very intense events (Diffenbaugh 2020; Zeder et al. 2023).

The physical mechanisms leading to heatwaves in the mid-latitudes occur at various spatial and temporal scales (Perkins 2015; Horton et al. 2016; Domeisen et al. 2022; Barriopedro et al. 2023). In the mid- and upper-level troposphere, these events are associated to a quasi-stationary high-amplitude Rossby wave (Petoukhov et al. 2013) which can be embedded in a hemispheric pattern (Coumou et al. 2014; Kornhuber et al. 2020; Di Capua et al. 2021). A blocking anticyclone is situated above the heatwave region sustaining the poleward advection of warm air along its western flank and causing adiabatic warming by subsidence and clear skies at its center. The latter causes warming through short-wave insolation of the lower layers of the atmosphere, which can be amplified if soils are dessicated and water is limited for plants evapotranspiration. This results in the allocation of incoming solar energy towards sensible rather than latent heat, increasing further surface air temperature (Seneviratne et al. 2006; Fischer et al. 2007; Seneviratne et al. 2010; Hirschi et al. 2011; Miralles et al. 2012, 2014; Rasmijn et al. 2018; Dirmeyer et al. 2021).

If the anticyclone stays stationary, extreme temperatures can be reached. High temperature increases are fundamentally limited by the moist convective instability of the air column (Zhang and Boos 2023), which defines a relevant upper bound for surface temperatures in the mid-latitudes. During heatwaves, surface air temperatures reach their maximal values when the air column is neutrally stratified. Accordingly, temperatures drop when precipitations are triggered by the convective instability of the boundary layer.

As a consequence, one can derive an estimate of the maximum reachable temperature at the surface from characteristics of the free-troposphere (temperature and geopotential at 500hPa) and surface specific humidity. This allows to propose an estimate of the physical upper bound B_ϕ of temperature for mid-latitudes regions, for which Noyelle et al. (2023) showed that it is 5–10°C higher than statistical estimates of the upper bound in a case study for Western Europe. Additionally, as shown by Zhang and Boos (2023) this upper bound increases with global warming.

In this paper, we address the issue of the underestimation of tail risks for extreme temperatures by bridging the gap between the physical and statistical theories of extreme temperatures. We propose to estimate statistical GEV models on time series of yearly maxima where the upper bound of the distribution is imposed by the physical knowledge of the system:

$$B(\mu, \sigma, \xi) = B_\phi. \tag{2}$$

This reduces the space parameter dimension by one. We show that this method estimates correctly the GEV distribution parameters and reduces the underestimation bias of traditional fitting methods.

This paper is organized as follows. Section 2 presents the climate data and the statistical models used. In this section we detail how we compute an estimate of the physical upper bound for surface temperatures. The results of the analysis are presented in section 3. We first demonstrate the inadequacy of the estimation of the upper bound with traditional methods. Then we show the relevance of our approach on synthetic GEV data and on a large ensemble of climate model outputs. Finally, we discuss the results in section 4.

2. Data and methods

a. Climate model data

Here we use the large ensemble of the state-of-the-art IPSL-CM6A-LR model (Boucher et al. 2020) with the CMIP6 configuration under an extended historical simulations spanning historical forcing over the period 1850-2014 and RCP2-4.5 forcing over the period 2014-2059. The ensemble encompasses 31 independent members. For each member and each year, the GMST is computed as the area weighted global mean 2-m air temperature. For each member and each grid point over the Northern Hemisphere (30°N-80°N) we extract the yearly maximum of daily maximum 2-m air temperature. To compute the physical upper bound of surface temperature B_ϕ we additionally consider the yearly maximum over the months June, July, August and September of the daily mean air temperature at 500hPa (T_{500}), daily mean geopotential height at 500hPa (Z_{500}) and yearly minimum of daily mean 2-m air specific humidity. For the period 1850-1949, only 10 members

are available for the variables T_{500} and Z_{500}. At each grid point, we therefore resample randomly from these 10 members to create data series of yearly maxima of T_{500} and Z_{500}.

The use of climate data that we make here is not intended to give precise and relevant values for adaptation purposes but merely as an illustration of our method on physically relevant data points.

b. Statistical models

Classically, the distribution of yearly maxima of daily maximum 2-m air temperature (TXx) is modelled using a member of the so-called Generalized Extreme Value (GEV) family of distributions. The probability of a yearly maxima Z to be below a threshold z is expressed as $P(Z \leq z) = G(z)$ where:

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi(\frac{z-\mu}{\sigma})\right]^{-1/\xi}\right\}, & \text{if } \xi \neq 0, \\ \exp\left\{-\exp\left[-(\frac{z-\mu}{\sigma})\right]\right\}, & \text{if } \xi = 0, \end{cases} \tag{3}$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. We denote this model $\mathcal{M}_0$. The use of this family of distributions is motivated by the extremal theorems which state that at a suitable limit of an infinite number of samples, the distribution of maxima converges towards a member of the GEV family (Coles 2001). It is therefore implicitly assumed that the pre-asymptotic distribution of yearly maxima can be correctly represented by this distribution.

To account for non-stationarity in the temperatures distribution, one can first let the location parameter depend linearly on GMST:

$$\mu = \mu_0 + \mu_1 \text{GMST}. \tag{4}$$

and the scale parameter σ is constant. We denote this model $\mathcal{M}_1$. We also consider a statistical model $\mathcal{M}_2$ where the scale parameter is also dependent on GMST:

$$\begin{cases} \mu = \mu_0 + \mu_1 \text{GMST} \\ \sigma = \log\left(1 + \exp(\sigma_0 + \sigma_1 \text{GMST})\right). \end{cases} \tag{5}$$

The functional form for the scale parameter assures that it stays positive and — although it is non-linear — for typical values that we consider here the scale parameter depends almost

linearly on GMST. Importantly, for all these models if the shape parameter ξ is negative, the distribution is bounded upwards and the bound B depends on the three parameters (cf. Eq.1). We use the maximum likelihood method to provide estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ (and correspondingly for non-stationary ones) of the parameters of the GEV distribution. The estimator for the bound is then:

$$\hat{B} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}. \tag{6}$$

Note that in the non-stationary cases, the estimate $\hat{B}$ depends on the covariate GMST and therefore on time.

In the following, we mimic the practice of attribution studies (Philip et al. 2020) by fitting the statistical models with one ($\mathcal{M}_1$) and two non-stationary parameters ($\mathcal{M}_2$) on 70 TXx randomly resampled from the pool of the 31 climate model members over the period 1945-2014 using the ensemble mean GMST as a covariate (see Fig. B1a for its time evolution). The choice to use only 70 points was made in reference to the average number of data points available for most weather stations (which were installed after 1945). To obtain distributions of the quantities of interest displayed below, we repeat the fitting procedure 1000 times.

c. Computation of the physical upper bound

As in Zhang and Boos (2023) and Noyelle et al. (2023), the maximum reachable surface temperature is computed as the surface temperature for which the moist static energy of the surface air parcel is equal to the saturated moist static energy of the free-troposphere (considered to be at 500hPa):

$$T_{s,max} = T_{500} + \frac{L_v}{c_p}(Q_{sat}(T_{500}) - Q) + \frac{g}{c_p}(Z_{500} - Z_s). \tag{7}$$

where T_{500} is the air temperature at 500hPa, Q the surface specific humidity of the air parcel, Z_{500} the geopotential height at 500hPa and Z_s the elevation of the surface. $T_{s,max}$ is computed from these four variables using the latent heat of vaporization L_v, the specific heat of air at constant pressure c_p, the gravitational constant g and the equation $Q_{sat}(T_{500}) \simeq \frac{\epsilon e_{sat}(T_{500})}{500\text{hPa}}$ where $e_{sat}(T_{500})$ is the saturation vapor pressure given by the Clausius-Clapeyron relation.

The relation 7 is maximal when T_{500} and Z_{500} are maximized and Q is minimized. However, by plugging the yearly maximum value of the two first quantities and the yearly minimum of the last one, it is possible that we either overestimate the physical upper bound B_ϕ — because there could not be a meteorological situation that combines the extremized values of these three quantities at the same time — or that we underestimate it — because the yearly natural variability may not provide sufficiently extreme values of these three quantities.

To propose a reasonable value of the physical upper bound B_ϕ for each resampling of n values of TXx that may be computable in practice, we also resample the corresponding n yearly maxima of T_{500} ($TX_{500,i}$) and Z_{500} ($ZX_{500,i}$), and n yearly minima of Q (QN_i). For each year $1 \leq i \leq n$ resampled we compute the physical upper bound $B_{\phi,i}$ using equation 7 with the corresponding yearly maxima of T_{500} and Z_{500}. However, we treat differently the air specific humidity Q variable which, as shown by Noyelle et al. (2023), is critical to have a relevant estimation of the upper bound. For this variable we use the same value for each of the n resampled years. This value is defined as the covariate-trend adjusted minima $\widetilde{QN}_{min}$ of the $(QN_i)_{1\leq i\leq n}$ over the n resampled years. If the trend on QN is non significant, we use simply the minima of Q over the n resampled years. This procedure gives us n values of the physical upper bound $B_{\phi,i}$:

$$B_{\phi,i} = T_{s,max}(TX_{500,i}, ZX_{500,i}, \widetilde{QN}_{min}). \tag{8}$$

We then regress the $B_{\phi,i}$ on the covariate GMST (or RMST) which gives us a linear estimate for the physical upper bound

$$B_\phi(\text{GMST}) = A + B \times \text{GMST}. \tag{9}$$

This procedure is straightforward to apply and gives a reasonable estimate for the physical upper bound. It should nonetheless be stated that this value constitutes only a rough estimate of the true upper bound and is subject to uncertainties. We note that our procedure can be more easily conceptualized in a bayesian context for which one could define a prior distribution on $B_\phi(\text{GMST})$ (Robin and Ribes 2020).

During the fit of the statistical models presented in the preceding subsection, we then impose that the statistical upper bound be equal to the physical upper bound:

$$\hat{B} = \mu - \frac{\sigma}{\xi} = B_\phi. \tag{10}$$

Doing so is equivalent to imposing a constraint on the GEV parameters and therefore reduces the dimension of the parameters space by one. The statistical modesl for which the bound is imposed are denoted as $\mathcal{M}_0^B$, $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$.

3. Results

a. Shortcomings of the traditional fitting approach

We first show the inadequacy of the traditional fitting approach for estimating tail probabilities of the yearly maximum of daily maximum temperature (TXx) with a GEV distribution (see also Fischer et al. (2023); Zeder et al. (2023)). To do so, we use the outputs of the large ensemble (31 members). We fit the statistical models with one ($\mathcal{M}_1$) and two non-stationary parameters ($\mathcal{M}_2$) on 70 TXx randomly resampled from the pool of the 31 members over the period 1945-2014 using the ensemble mean GMST as a covariate. We then investigate whether the upper bounds B_1 and B_2 estimated with the two models are "true" upper bounds, i.e. whether they are exceeded on the full dataset over the period 1850-2059 for the 31 members. By doing so 1000 times, we have an estimate of the probability that the estimated upper bounds of the two models will be exceeded at least once (see appendix A for the detail of the computation).

Figure 1ab shows the result of this computation for every land grid points in the Northern Hemisphere mid-latitudes (30N-80N). For most of the grid points, the probability to exceed the upper bound estimated on the period 1945-2014 is above 40%. Remarkably, there is not a single grid point over the region studied for which the probability to exceed the upper bound is equal to 0 after 1000 resamplings. Model $\mathcal{M}_1$ (first column) performs better for this test than model $\mathcal{M}_2$ (second column) but both statistical models largely underestimate the value of the upper bound. This shows that, at least on climate model outputs, the estimation of the upper bound of the TXx distribution is not reliable. Figure 1cd gives an estimation of the empirical return time that corresponds to the values of the bounds B_1 and B_2. The return time — which should be infinite

— is in practice closer to 2000 years for model $\mathcal{M}_1$ (Fig. 1c) and to 200 years for model $\mathcal{M}_2$ (Fig. 1d). We also note that, for model $\mathcal{M}_1$, large regions in the eastern side of continental masses show a large return time of the upper bound (above 6000 years, colored in light blue), which suggests that for these regions the upper bound is only slightly underestimated. Finally, the ratio between the GEV and the empirical probabilities of a centennial event is depicted in Figure 1ef. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate (see 4 for the detail of the computation). For both models, the intensity of events which are predicted to be centennial according to the GEV estimate actually happen every 25 to 50 years, which make them 2 to 4 times more likely than predicted. Apart for a few grid points, this underestimation occurs almost everywhere.

Increasing the size of the sample for fitting the models only slightly improves the situation (see Fig. B2 for 100 years and Fig. B3 for 150 years) and for most grid points the probability to exceed the upper bound estimated is above 20% even with 150 years of data. Similarly, changing the fitting period towards a period with more variation in the covariate does not change the results significantly (Fig. B4 with 70 points resampled over the period 1956-2025). Finally, using the Regional Mean Surface Temperature (RMST) rather than GMST as a covariate only marginally changes the results (Fig. B1bcd for the regional covariates, and Fig. B5 for the results of the fits). Figure 2 shows that the statistical estimation of the upper bound is lower than the physical estimation of the upper bound by a margin of 3 to 8°C. Except for a small region in Western Russia, the statistical upper bound is systematically lower than the physical upper bound. We additionally note that the estimation of the physical upper bound is much more precise than the statistical upper bound, cf. Figure B6 in appendix showing the standard deviation of the estimation computed on the 1000 resamplings.

b. Imposing the upper bound on synthetic GEV data

We first demonstrate the interest of the method of imposing the upper bound in the GEV fit using synthetic data with parameters typical of the parameters found when fitting a GEV on TXx data (see appendix A for the detail). We simulate 1000 series of 50, 100 and 200 points distributed according to the model $\mathcal{M}_1$, i.e. with a non-stationarity on the location parameter only. We then fit the full model $\mathcal{M}_1$ and the model $\mathcal{M}_1^B$ where the (correct) upper bound is imposed during the

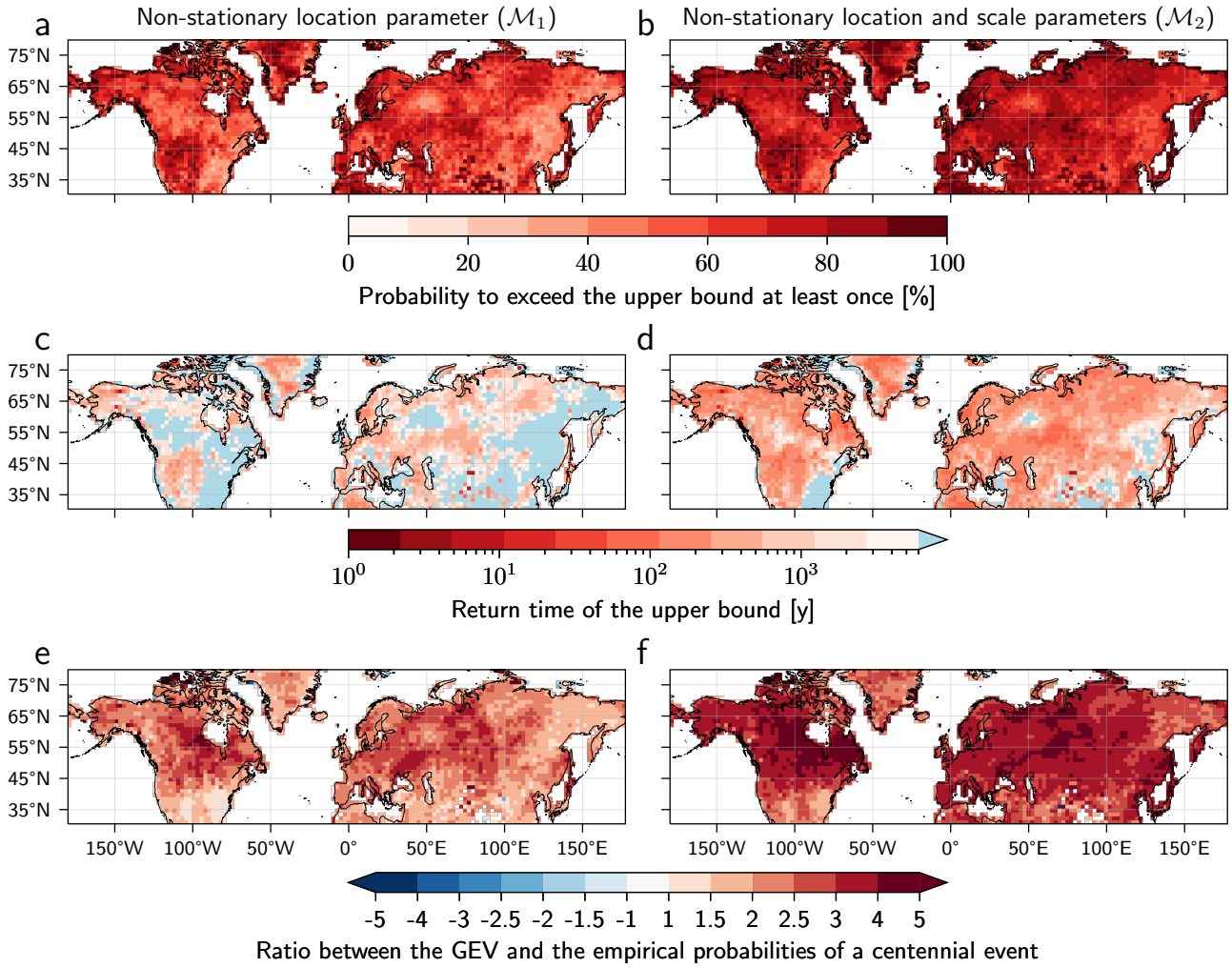


FIG. 1. Bias of tail risks of extreme temperatures with the traditional GEV fitting approach. First row (a-b): empirical probability that the statistical upper bound estimated using 70 points resampled on the period 1945-2014 over all 31 members is exceeded at least once on the full dataset (1850-2059). Second row (c-d): estimated return time of this upper bound on the full dataset. Grid points where only one TXx exceeds the upper bound (return time above 6000 years) are colored in light blue. Third row (e-f): ratio between the GEV and the empirical probabilities of a centennial event. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate. First column: statistical model with a linear dependence on GMST of the location parameter only ($\mathcal{M}_1$). Second column: statistical model with a linear dependence of both the location and scale parameters ($\mathcal{M}_2$).

fit. For comparison purposes, we additionally fit three other models derived from $\mathcal{M}_1$ where either the (correct) location ($\mathcal{M}_1^{\mu}$), scale ($\mathcal{M}_1^{\sigma}$) or shape parameter ($\mathcal{M}_1^{\xi}$) is imposed to the model.

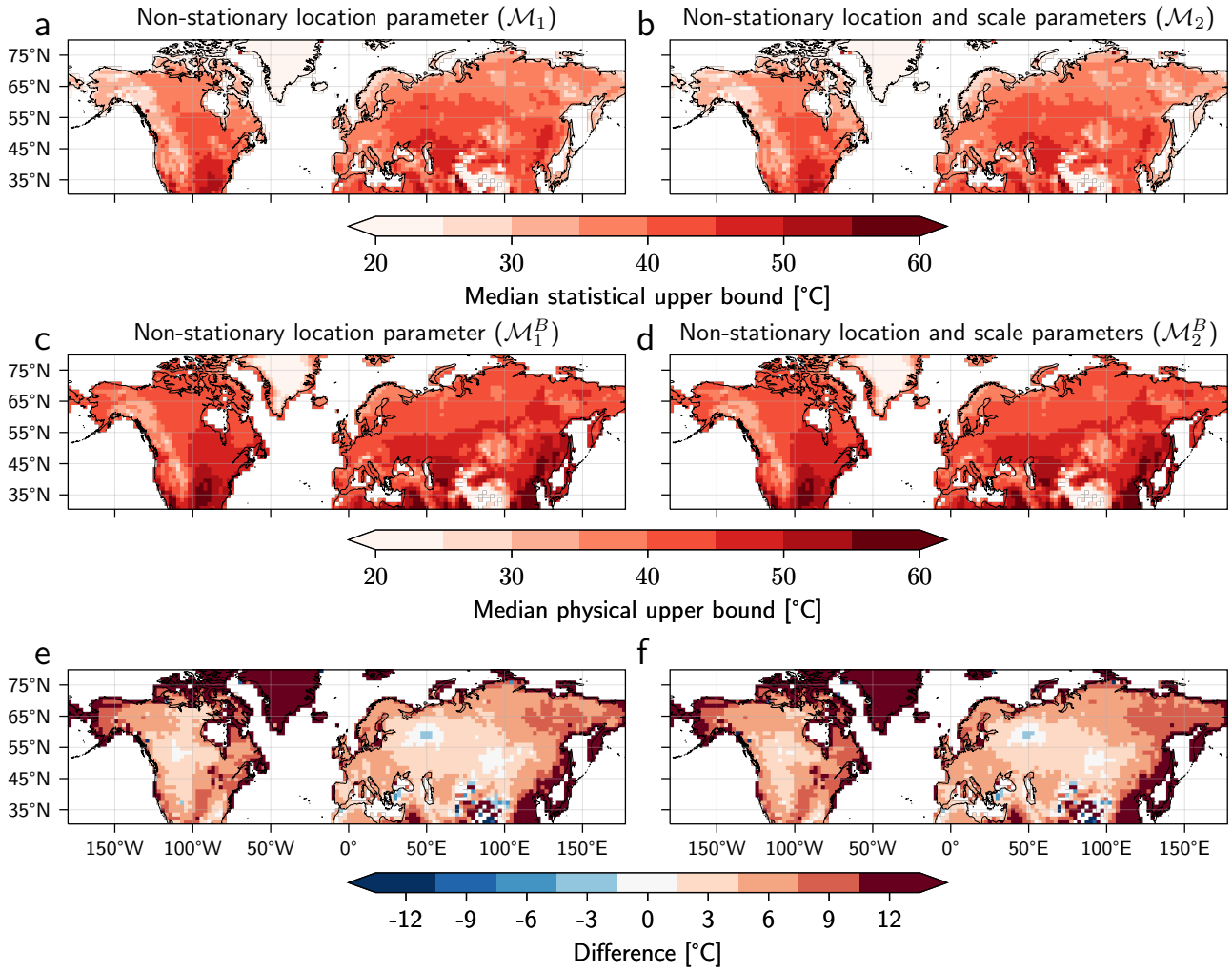


FIG. 2. Absolute values of the upper bounds. Panels a and b show the median statistical upper bound estimated with models $\mathcal{M}_1$ and $\mathcal{M}_2$ obtained after 1000 fits with data resampled over the period 1945-2014. Panels c and d show the corresponding median physical upper bound. Panels e and f show the difference between the two. For all the plots we chose the covariate to be +0.5K above the 1951-1980 average (corresponding roughly to the year 2000).

Figure 3 shows the results for the estimation of the different parameters (Fig. 3abcd) and several metrics (Fig. 3efghi). All models estimate correctly the location parameters μ_0 and μ_1 — with a slight overestimation of the first one and underestimation of the second one — but they almost all underestimate the scale parameter (Fig. 3c). Models $\mathcal{M}_1$, $\mathcal{M}_1^\mu$ and $\mathcal{M}_1^\sigma$ underestimate the shape parameter and there is a large spread around the true value (Fig. 3d). On the contrary, when we impose the upper bound the estimation of the shape parameter is remarkably precise.

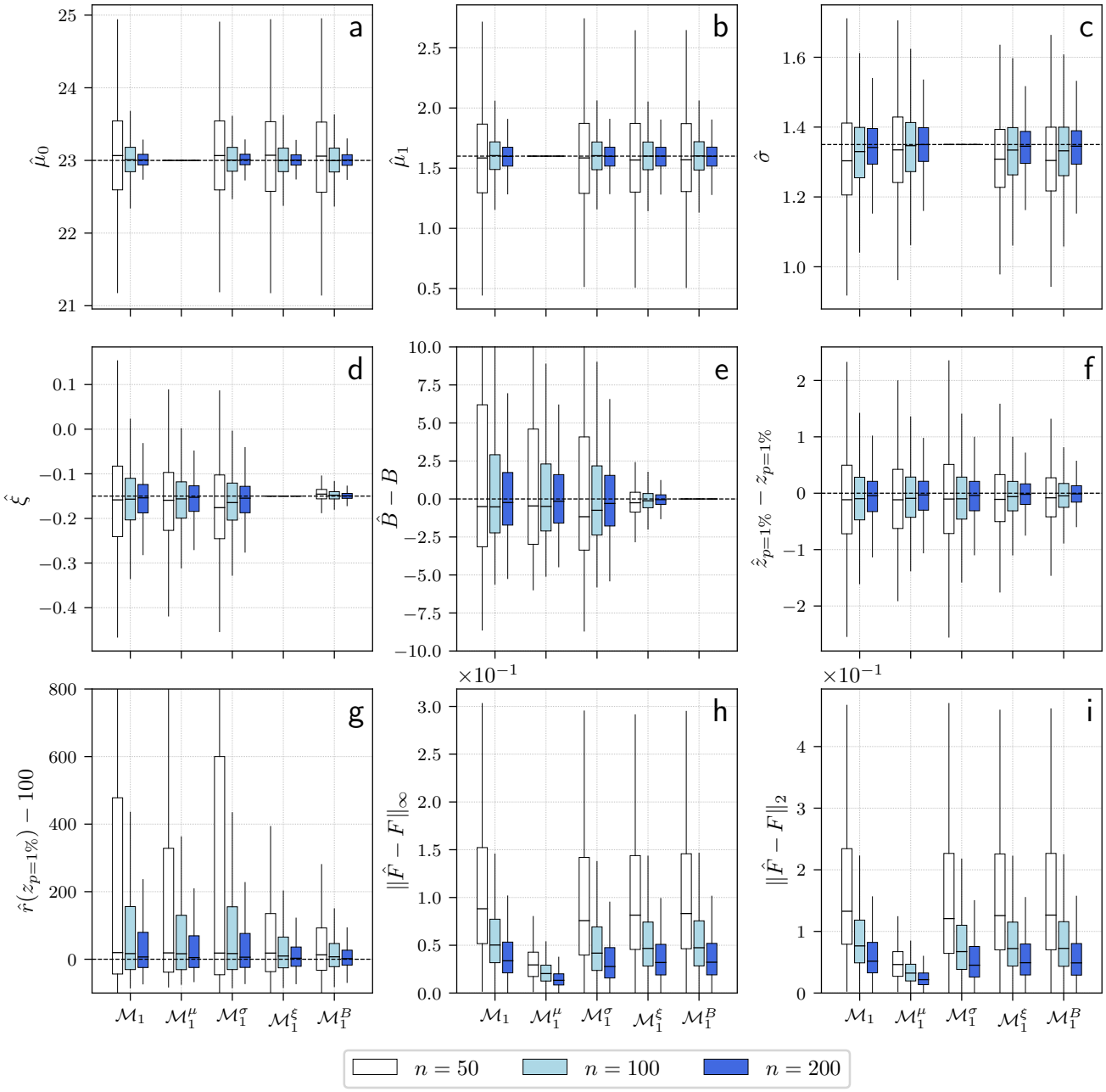


FIG. 3. Fit of the statistical models derived from $\mathcal{M}_1$ on synthetic GEV data. Estimation of the parameters of the model: (a) and (b) location parameters, (c) scale parameter and (d) shape parameter. Metrics to measure the quality of the fit: (e) bias in the upper bound, (f) bias in the return level at 1%, (g) bias in the return time for the true return level at 1%, norm (h) L_∞ and (i) norm L_2 between the estimated and the true CDFs. Each boxplot is derived from 1000 fits of randomly sampled GEV data from a $\mathcal{M}_1$ model with $n = 50, 100$ and 200 samples.

As a consequence, the upper bound is systematically underestimated for models $\mathcal{M}_1$, $\mathcal{M}_1^\mu$ and $\mathcal{M}_1^\sigma$. It is also the case for the model $\mathcal{M}_1^\xi$ but the estimation is much more precise. We note here

the paramount importance of correctly estimating the shape parameter for correctly estimating the upper bound and vice versa.

The return level at 1% (Fig. 3f) is underestimated and the corresponding return time (Fig. 3g) is overestimated for all models but the model with the upper bound imposed perform best both in bias and in spread. When comparing models in terms of the norms L_∞ and L_2 with respect to the true Cumulative Distribution Function (CDF), the model $\mathcal{M}_\mu$ performs best but the other models have similar performance. Increasing the sample size improves the estimation for all models, but for the estimation of tail probabilities the model $\mathcal{M}_1^B$ remains the best (e.g. Fig. 3g).

Appendix B provides similar estimations for the models $\mathcal{M}_0$ (Fig. B7) and $\mathcal{M}_2$ (Fig. B8) for which we have similar results as those presented here. Our results demonstrate that once one knows the upper bound of a bounded GEV distribution, imposing the value of the bound during the fit gives the best fitting results on the metrics displayed here. Nonetheless, when it comes to non-synthetic data, it is likely that we will make an error on the imposed value of the upper bound — which is unknown. To investigate this issue, we additionally fitted the model $\mathcal{M}_1^B$ with an error B_{err} on the bound B. Results are presented in d for an error of $B_{err} = +2.5$ (Fig. B9) and $B_{err} = +5$ (Fig. B10). In these cases, the scale parameter (Fig. B9c) tends to be underestimated and the shape parameter (Fig. B9d and B10d) to be overestimated. Accordingly, the return level at 1% (Fig. B9fc and B10f) is overestimated and the return time for the true return level at 1% is underestimated (Fig. B9g and B10g). We nevertheless note that the error made is small when compared to the spread around the estimation for the other models and that for $B_{err} = 2.5$ the overestimation of the return time is of the same order of magnitude as the underestimation of the full model $\mathcal{M}_1$. On the rest of the distribution, the performance of the biased model $\mathcal{M}_1^{B+B_{err}}$ is comparable to the other models (Fig. B9hi and B10hi). For the quality of the fit it is therefore preferable to impose an upper bound as close as possible to the correct upper bound but errors up to +5°C can be tolerated even though they tend to overestimate events at the upper tail of the distribution. We come back to this question in the discussion section.

c. Imposing the upper bound on a climate model outputs

We apply our approach to temperatures simulated by a climate model and determine the distribution of yearly maxima of daily maximum temperature at one grid point in Western Europe

(lon=0°, lat=49.44°N) for the 31 members of the IPSL model over the period 1850-2059. Figure 4a shows the evolution of these maxima with time. We note in particular that there is an extreme outlier on the year 2019 which reaches the value of 38.6°C (shown by a black triangle), shattering the previous record of all members by a margin of 1.9°C. We fit the models $\mathcal{M}_1$ and $\mathcal{M}_2$ with the GMST as covariate by resampling 1000 times over all members during the period 1945–2014 (70 years), as previously. Figure 4a shows the median, 5th and 95th percentiles of the estimated upper bounds B_1 and B_2 for both statistical models along with the estimated physical upper bound B_ϕ. The median estimates of upper bounds B_1 and B_2 are *below* 10 data points of the temperature time series. On the contrary, the physical upper bound is never reached and is above the 95-th percentile of the distribution of the upper bounds B_1 and B_2.

Figure 4b illustrates the issue of the underestimation of the upper bound with the case of the 2019 extreme event. In black we show the empirical distribution of maxima for years 2016-2021, for which the value of the covariate GMST is similar to the one of 2019 (14.32°C vs 14.24 to 14.38°C). As an illustration we show the PDF fitted on member 14 during the period 1945-2014 for models without the upper bound imposed ($\mathcal{M}_1$ and $\mathcal{M}_2$) and with the upper bound imposed ($\mathcal{M}_1^B$ and $\mathcal{M}_2^B$). The PDF of the former are very close and are not distinguishable on the graph. Remarkably, they both estimate an upper bound which is around 1°C *smaller* than the 2019 extreme outlier. To show that we did not choose member 14 for its underestimation of the bound, the two blue boxplots shows the distribution of 1000 estimations of the 2019 upper bound done by resampling as previously. More than 75% of these estimations are below the maximum value of 2019.

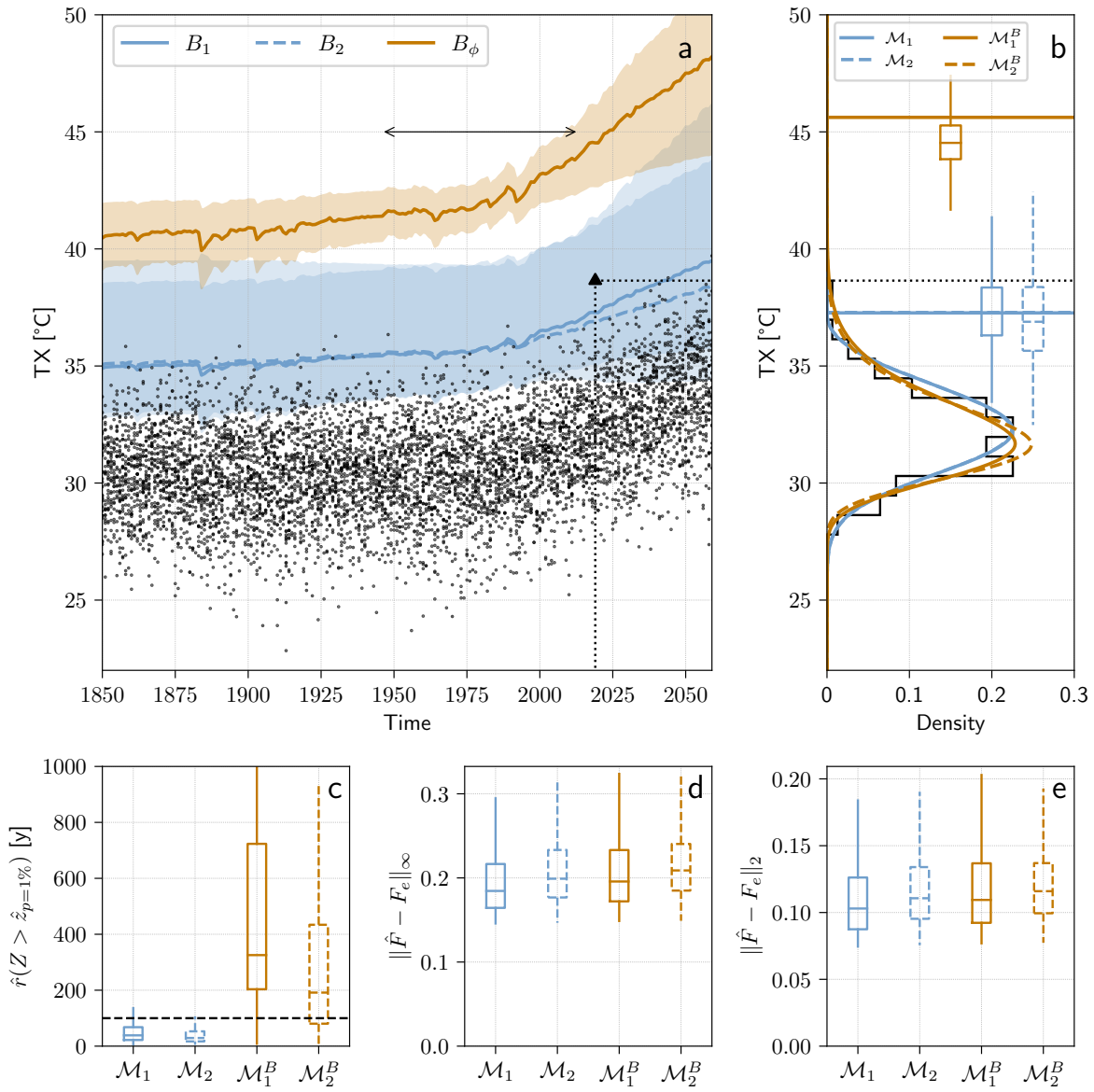


FIG. 4. Comparison between the statistical models with and without imposing the upper bound. (a) Yearly maxima of daily maximum temperature (TXx) at a grid point in Western Europe (lon=0°, lat=49.44°N) for the 31 members of the IPSL model (black), median, 5 and 95-th percentiles of the distribution of upper bounds B_1 and B_2 (blue plain and dashed) for statistical models $\mathcal{M}_1$ and $\mathcal{M}_2$ for 1000 fits with resampled data points on the period 1945-2014 (black arrow), and median, 5 and 95-th percentiles of the distribution of physical upper bounds B_ϕ (orange). The black triangle shows the value of the 2019 extreme outlier (38.6°C). (b) Histogram of TXx over the period 2016-2021 (black) and PDFs (colors) using the fit on member 14 for statistical models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$. The black dotted line shows the extreme 2019 event and the colored lines shows the physical upper bound and the estimated statistical upper bounds for the fit on member 14 (the blue plain and dashed lines are almost confounded). The boxplots shows the distribution of upper bounds for the different models for 1000 fits on resampled data. (c) Distribution of return time for the return level at 1% for all statistical models. Average norms (d) L_∞ and (e) L_2 between the estimated and the empirical CDFs.

Fig. 4c shows a comparison between the estimated return time in the full data set (1850-2059) of the 1% return level using the fits of the different models. Models $\mathcal{M}_1$ and $\mathcal{M}_2$ vastly overestimate this return time: the median return time empirical estimate for the return level at 1% according to their distribution is around 40 years. On the contrary, models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$ underestimate this return time, by around 250 years for the latter and 100 years for the former. This translates into an overestimation of tail risks of extreme temperatures for the two models for which the physical upper bound is imposed. Figure 4de show the performance of the different models on the rest of the distribution using average norms L_∞ and L_2 between the fitted CDF and the empirical CDF (see SI for the detail of the computation). Models have similar performance for both metrics, with a slightly better one for models $\mathcal{M}_1$ and $\mathcal{M}_1^B$. Models where the upper bound is prescribed perform slightly worse compared to traditional models. We provide in d the same figure using the RMST as covariate (Fig. B11), which gives similar results.

Finally, Figure 5 shows the same results as Fig. 1 for models where the upper bound is prescribed ($\mathcal{M}_1^B$ and $\mathcal{M}_2^B$). Remarkably, Fig. 5ab show that a for a majority of grid points the probability to exceed the upper bound is null after 1000 resamplings (grid points colored in blue). For grid points where the probability is not null, it is close to zero (e.g. in the Iberian peninsula or in North West America). There is however two regions over Western Russia and Central Eurasia where the probability to exceed the upper bound at least once is still above 60 %. Nevertheless, Fig. 5cd show that for the vast majority of these grid points, the return time of the estimated upper bound is very high: above 6000 years which implies that only a single TXx on the dataset is above this value. In Fig. B12 and B13 we provide the fits for example grid points in these two regions. They show that only one TXx in the 19th century is actually above the median estimate of the physical upper bound (respectively in 1850 and 1860). We also note that for those grid points the statistical and physical estimates of the upper bounds are much closer than for the grid point in Western Europe — although the physical upper bound is more precisely estimated.

Additionally, Fig. 5ef show that imposing the upper bound partially alleviates the issue of underestimation of centennial events. Contrary to Fig. 1ef, there is no uniform underestimation of the return level of the centennial events. In particular, coastal regions even tend to overestimate the risk of these events. This effect is probably due to the choice we made to take the lowest value of the surface specific humidity in the resampled data to compute an estimate of the physical

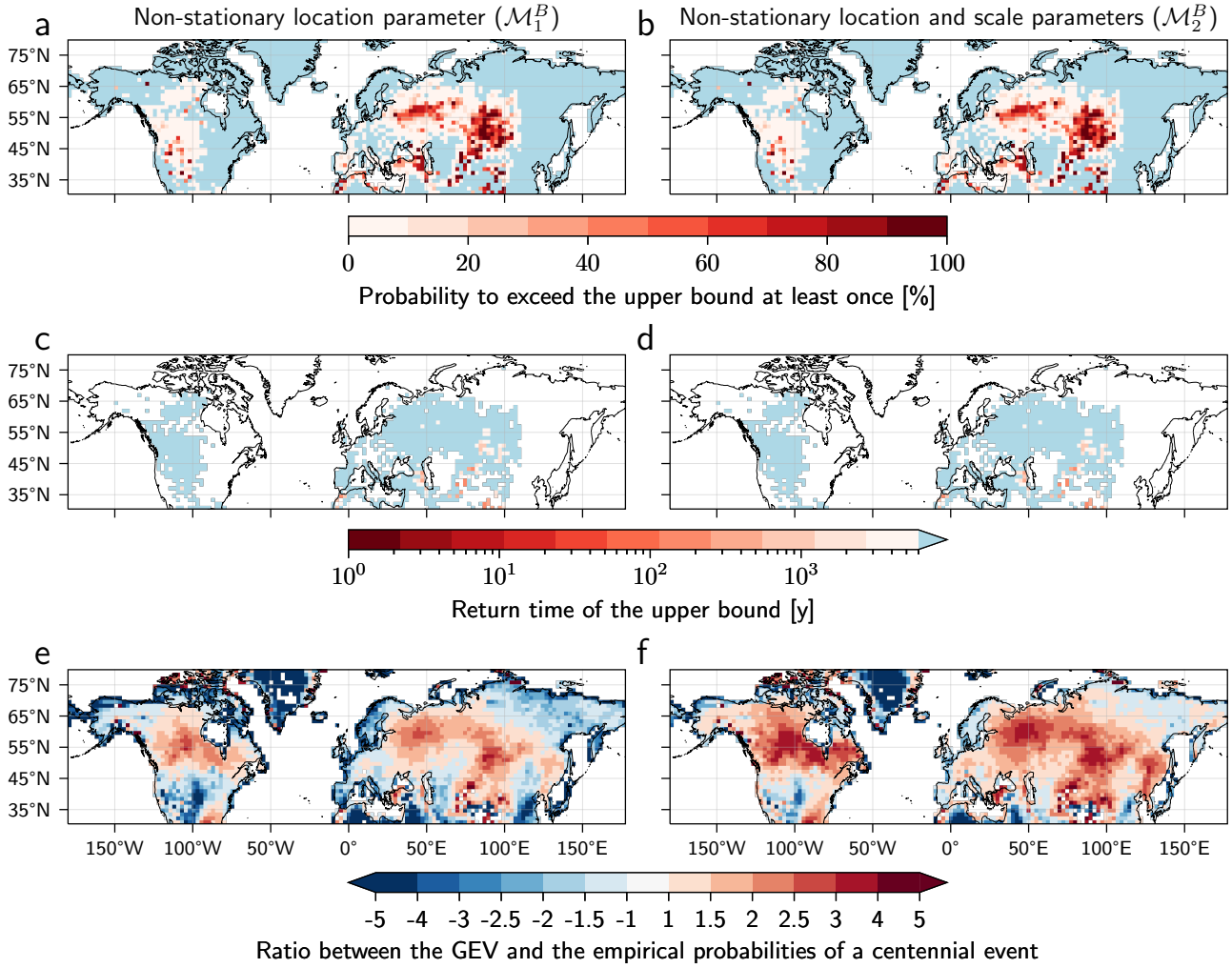


FIG. 5. Same as Figure 1 with models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$. First row (a-b): empirical probability that the statistical upper bound estimated using 70 points resampled on the period 1945-2014 over all 31 members is exceeded at least once on the full dataset (1850-2059). The grid points where the probability is null are colored in light blue. Second row (c-d): estimated return time of this upper bound on the full dataset. Grid points where only one TXx exceeds the upper bound (return time above 6000 years) are colored in light blue. Grid points where no TXx exceeds the upper bound are colored in white. Third row (e-f): ratio between the GEV and the empirical probabilities of a centennial event. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate. First column: statistical model with a linear dependence on GMST of the location parameter only ($\mathcal{M}_1^B$). Second column: statistical model with a linear dependence of both the location and scale parameters ($\mathcal{M}_2^B$).

upper bound. On the other hand, the overestimation bias over Europe for example is largely reduced and some regions even slightly overestimate the risk. One should note that there are

still large regions where our procedure underestimate the risk of centennial events, for example in Canada, Western Russia and Central Eurasia. This effect may be due to historical aerosols forcing which are used by the model and the choice we made to use data between 1945 and 2014 for the fit. This period indeed corresponds to higher anthropogenic aerosols levels in these regions, which are increasingly recognized to strongly influence the magnitude of extreme temperatures (Sillmann et al. 2013; Dong et al. 2017; Westervelt et al. 2020; Luo et al. 2020; Seong et al. 2021). Nevertheless, the overestimation is reduced to around 2.5 more events in the dataset than according to the GEV estimate (Fig. 5e). Finally, as previously, model $\mathcal{M}_2^B$ performs worse according to this metric than model $\mathcal{M}_1^B$.

4. Discussion and conclusions

We presented a new method to bridge the gap between physical knowledge and statistical estimates of tail probabilities of extreme mid-latitude surface temperatures. We proposed to incorporate a physical estimate of the upper bound during the fit of GEV distributions for yearly maximum of daily maximum 2-m air surface temperature. We showed on synthetic GEV data and on climate model outputs that doing so leads to more precisely fitted parameters — especially the crucial shape parameter, reduces the underestimation of tail probabilities and does not deteriorate the performance of the fit on the rest of the distribution. The underestimation of the upper bound is largely avoided and we additionally showed that our method improves the estimation of probabilities of centennial events, although this was not intended by the method. Our method also has the advantage of rendering the statistical fit more stable with respect to its tail properties. Figure B6 indeed shows that the standard deviation on the estimated upper bounds is much smaller for models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$ than for classical models $\mathcal{M}_1$ and $\mathcal{M}_2$. It especially ensures that the upper bound always exists — which even though is clear physically, is not often found in practice. We additionally showed that overestimating the upper bound of the distribution would tend to also overestimate tail probabilities and therefore risks of very high temperatures. This overestimation is however of the same order of magnitude as the underestimation associated with classical fits of GEV distribution.

Other metrics could have been chosen to quantify the quality of the fits with respect to data. We nevertheless emphasize that all classical metrics based on the log-likelihood, such as the Akaike information criterion (Akaike 1998), the Bayesian information criterion (Schwarz 1978) or the

Likelihood ratio test (Coles et al. 2001), are not adapted here because they give an infinitely bad score to classical models. The latter indeed underestimate the upper bound: a model which gives a zero probability to an event which still occurs has an infinite log-likelihood. To the best of our knowledge there is also no consensus on the metrics to evaluate the quality of the fit of a model in a non-stationary context.

Choosing the value of the upper bound to be imposed during the fit is the crucial element for the success of our procedure. The estimation procedure of the physical upper bound B_ϕ that we proposed here likely tends to overestimate this quantity because it combines extremized value of temperature and geopotential at 500hPa and surface specific humidity which are unlikely to happen at the same time in practice. As shown on synthetic GEV data and on climate model outputs, it is therefore possible that we overestimate tail risks at some places. However, given the fact that GEV estimates of tail probabilities may be used by practitioners for adaptation purposes, it is not only a scientific but also an ethical question to choose to over- or underestimate risks of very high temperature events. Because of the large consequences that an underestimation of tail probabilities could have on societies, we here advocate for a scientific choice which avoids this underestimation even though it comes at the risk of overestimation of tail probabilities.

Acknowledgments. The authors would like to thank Yi Zhang for discussions which led to the work carried out in this paper. This work has received support from the European Union's Horizon 2020 research and innovation programme under Grant agreement no. 101003469 (XAIDA), by the Integrated Training Network (ITN) under the Marie Sklodowska-Curie Grant agreement no. 956396 (EDIPI), and the French ANR grant ANR-20-CE01-0008-01 (SAMPRACE).

Data availability statement. Pre-processed climate model outputs and GEV data used in this paper are available at `https://zenodo.org/doi/10.5281/zenodo.10715007`. The analysis presented here is done using Python. The Python code is available at `https://doi.org/10.5281/zenodo.10679570`. It makes use of the SDFC package developed by Yoann Robin and available at `https://zenodo.org/doi/10.5281/zenodo.4263885`.

APPENDIX A

Computation details

a. Computation details for Figure 1

1) Probability of exceeding the upper bound

We here detail the procedure to compute the probability of exceeding the upper bound that is presented in Figure 1. This procedure is applied for each grid point over the Northern Hemisphere mid-latitudes (30N-80N). We first draw randomly 70 TXx from the pooled data of the 31 members between the period 1945-2014 (2170 data points) without replacement. We then fit the statistical model into consideration. If the shape parameter of the fit is negative (i.e. if there is indeed an upper bound), we consider the estimated statistical upper bound $\hat{B}_i$ and we count the number n_i of TXx in the full dataset (1850-2059, 6510 data points) that are above $\hat{B}_i$. Note that for models $\mathcal{M}_1$ and $\mathcal{M}_2$ the estimated upper bound depends on the covariate GMST (or RMST, see below) and therefore on time. We do this procedure $N = 1000$ times and the probability of exceeding the upper bound at least once is then computed as:

$$\hat{P}(Z > \hat{B}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(n_i > 0) \tag{A1}$$

where $\mathbf{1}(n_i > 0)$ is equal to 1 if $n_i > 0$ and equal to 0 otherwise.

2) Median return time of the upper bound

The median return time of the estimated upper bound is then computed as:

$$\hat{r}(Z > \hat{B}) = \left(\text{Med}\left(\frac{n_i}{N_d}\right)\right)^{-1} \tag{A2}$$

where $N_d = 6510$ is the size of the full dataset and Med is the median operation. $\frac{n_i}{N_d}$ is an estimation of the probability that the yearly maxima is above the bound $\hat{B}_i$ for the fit i. We then compute the median of this probability over the $N = 1000$ resampling and inverse it to give an estimate of the return time of the estimated upper bound.

3) Ratio between the GEV and the empirical probabilities of a centennial event

We proceed similarly to estimate the return time of a centennial event. For each fit i, we first find the return level $\hat{z}_{p=1\%,i}$ corresponding to the 99th quantile of the fitted GEV PDF. Note that for models $\mathcal{M}_1$ and $\mathcal{M}_2$ the estimated return level depends on the covariate GMST (or RMST, see below) and therefore on time. We then count the number $\tilde{n}_i$ of TXx in the full dataset (1850-2059, 6510 data points) that are above $\hat{z}_{p=1\%,i}$. The median empirical return time of the centennial return level is thus:

$$\hat{r}(Z > \hat{z}_{p=1\%}) = \left(\text{Med}\left(\frac{\tilde{n}_i}{N_d}\right)\right)^{-1} \tag{A3}$$

where as previously $N_d = 6510$ is the size of the full dataset and Med is the median operation. $\frac{\tilde{n}_i}{N_d}$ is an estimation of the probability that the yearly maxima is above the GEV centennial return level $\hat{z}_{p=1\%,i}$ for the fit i. We then compute the median of this probability over the $N = 1000$ resampling and inverse it to give an estimate of the empirical return time of the estimated centennial return level. If the fit corresponds to the true underlying distribution, we should find $\hat{r}(Z > \hat{z}_{p=1\%}) = 100$ by definition of the centennial return level. However this is not the case in practice and we quantify this difference by dividing by the true return time. The ratio presented in Figure 1 is then computed as:

$$\text{Ratio} = \begin{cases} -\hat{r}(Z > \hat{z}_{p=1\%})/100, & \text{if } \hat{r}(Z > \hat{z}_{p=1\%}) > 100, \\ 100/\hat{r}(Z > \hat{z}_{p=1\%}), & \text{if } \hat{r}(Z > \hat{z}_{p=1\%}) \leq 100. \end{cases} \tag{A4}$$

If for example $\hat{r}(Z > \hat{z}_{p=1\%}) = 50$, the events that are said to be centennial by the statistical model actually occur every 50 years in the dataset. Therefore the ratio is equal to 2: these events are twice *more* likely in the dataset than according to the GEV estimate. On the other hand, if $\hat{r}(Z > \hat{z}_{p=1\%}) = 200$, the events that are said to be centennial by the statistical model actually occur every 200 years in the dataset. Therefore the ratio is equal to -2: these events are twice *less* likely in the dataset that according to the GEV estimate.

b. Computation of RMST

To define the RMST covariates, we split the Northern Hemisphere mid-latitude regions (30N-80N) into three parts: North America (180W-30W), Europe (30W-50E) and Asia (50E-180E). For each grid points in these three regions, we first compute the area weighted 2-m air temperature average. The RMST is then the ensemble mean of these quantities. The results are presented in Fig. B1bcd.

c. Computation details for synthetic GEV data

1) Simulations

For the simulation of synthetic GEV data we use the following parameters:

$$\mathcal{M}_0 \begin{cases} \mu = 23 \\ \sigma = 1.35 \\ \xi = -0.15 \end{cases} \tag{A5}$$

$$\mathcal{M}_1 \begin{cases} \mu_0 = 23 \\ \mu_1 = 1.6 \\ \sigma = 1.35 \\ \xi = -0.15 \end{cases} \tag{A6}$$

$$\mathcal{M}_2 \begin{cases} \mu_0 = 23 \\ \mu_1 = 1.6 \\ \sigma_0 = 1.35 \\ \sigma_1 = 0.1 \\ \xi = -0.15. \end{cases} \tag{A7}$$

493 2) Norm metrics

494 The norm metrics L_∞ and L_2 are computed between the CDFs of the fitted distribution $\hat{F}$ and
495 the true distribution F as such:

$$\|\hat{F} - F\|_\infty = \sup_{t \in \mathbb{R}} |\hat{F}(t) - F(t)| \tag{A8}$$

496 and:

$$\|\hat{F} - F\|_2 = \sqrt{\int_{-\infty}^{+\infty} (\hat{F}(t) - F(t))^2 dt}. \tag{A9}$$

497 Numerically these two quantities are computed using a sampling at $dt = 10^{-2}$ for t between -200
498 and 200.

499 *d. Norm metrics for climate model outputs*

500 The distribution of return times presented in Figure 4 panel c and Figure B11 panel c are
501 computed the same way as for the return time presented above for Figure 1. The norm metrics L_∞
502 and L_2 are computed between the empirical CDF F_e of data and the fitted CDF $\hat{F}$ for each year
503 (using the 31 members) and then averaged over the full period (1850-2059).

APPENDIX B

Additional figures

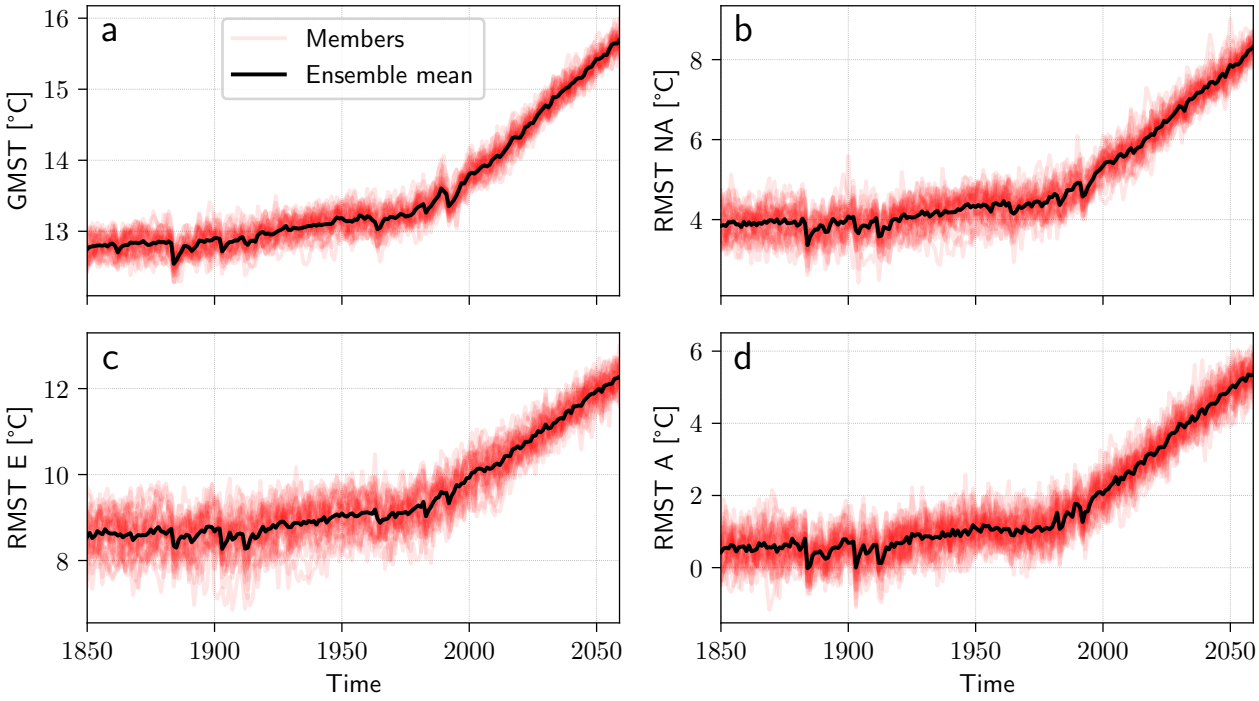


FIG. B1. Evolution of covariate for the GEV fit. (a) Ensemble mean (black) and members (red) GMST. (b) Same for RMST over North America (180W-30W, 30N-80N). (c) Same for RMST over Europe (30W-50E, 30N-80N). (d) Same for RMST over Asia (50E-180E, 30N-80N).

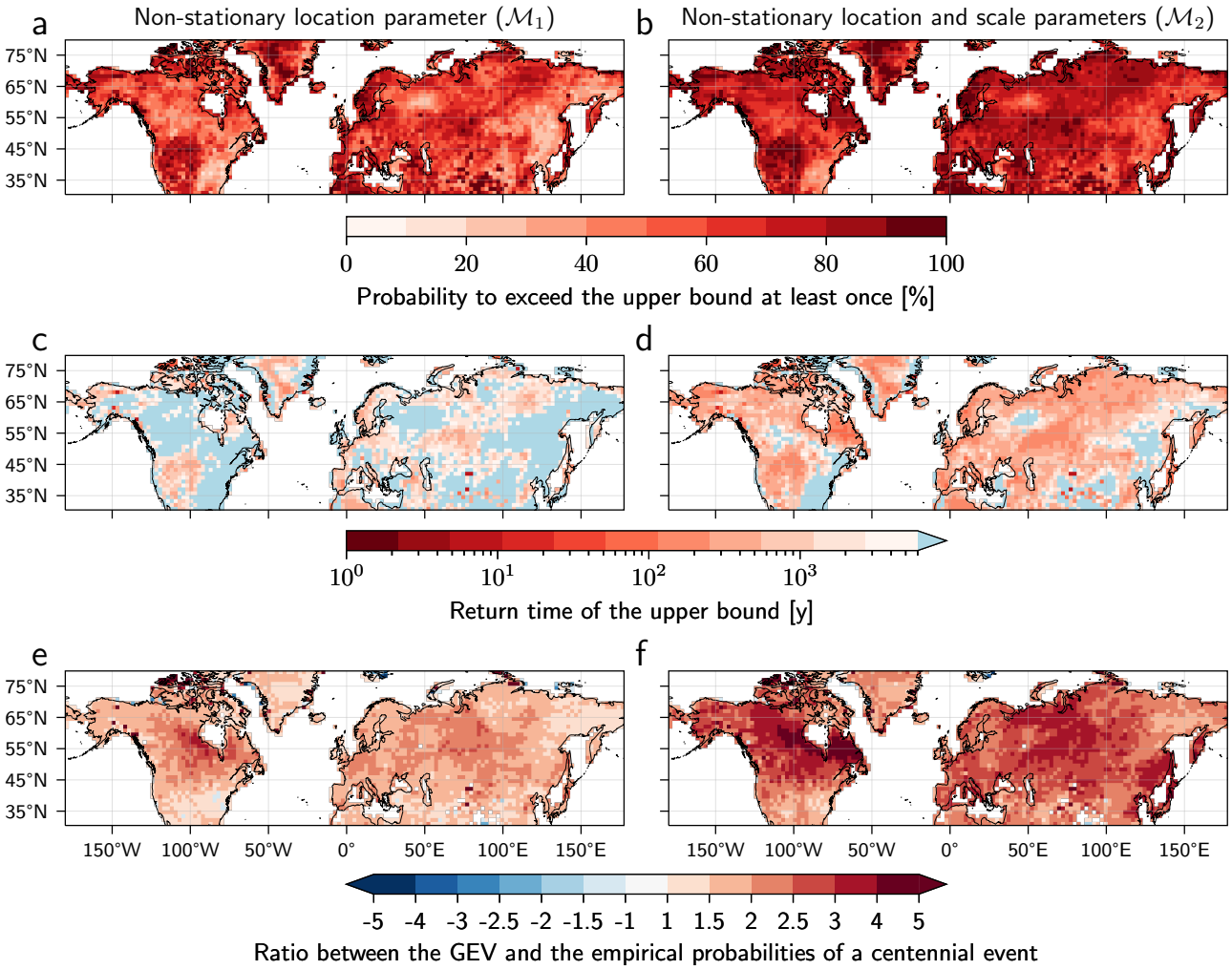


FIG. B2. Same as Fig. 1. with 100 points resampled over the period 1915-2014.

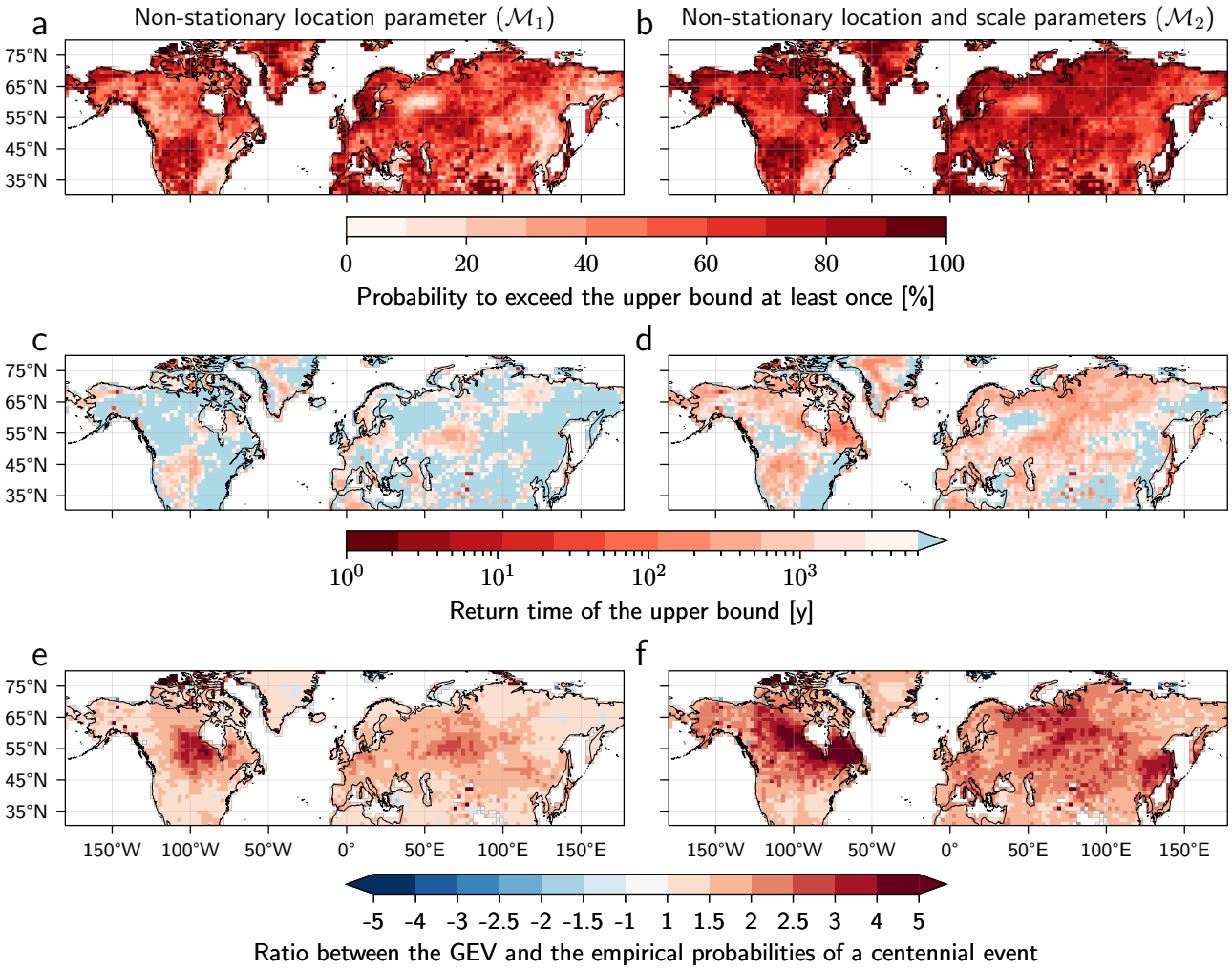


FIG. B3. Same as Fig. 1. with 150 points resampled over the period 1865-2014.

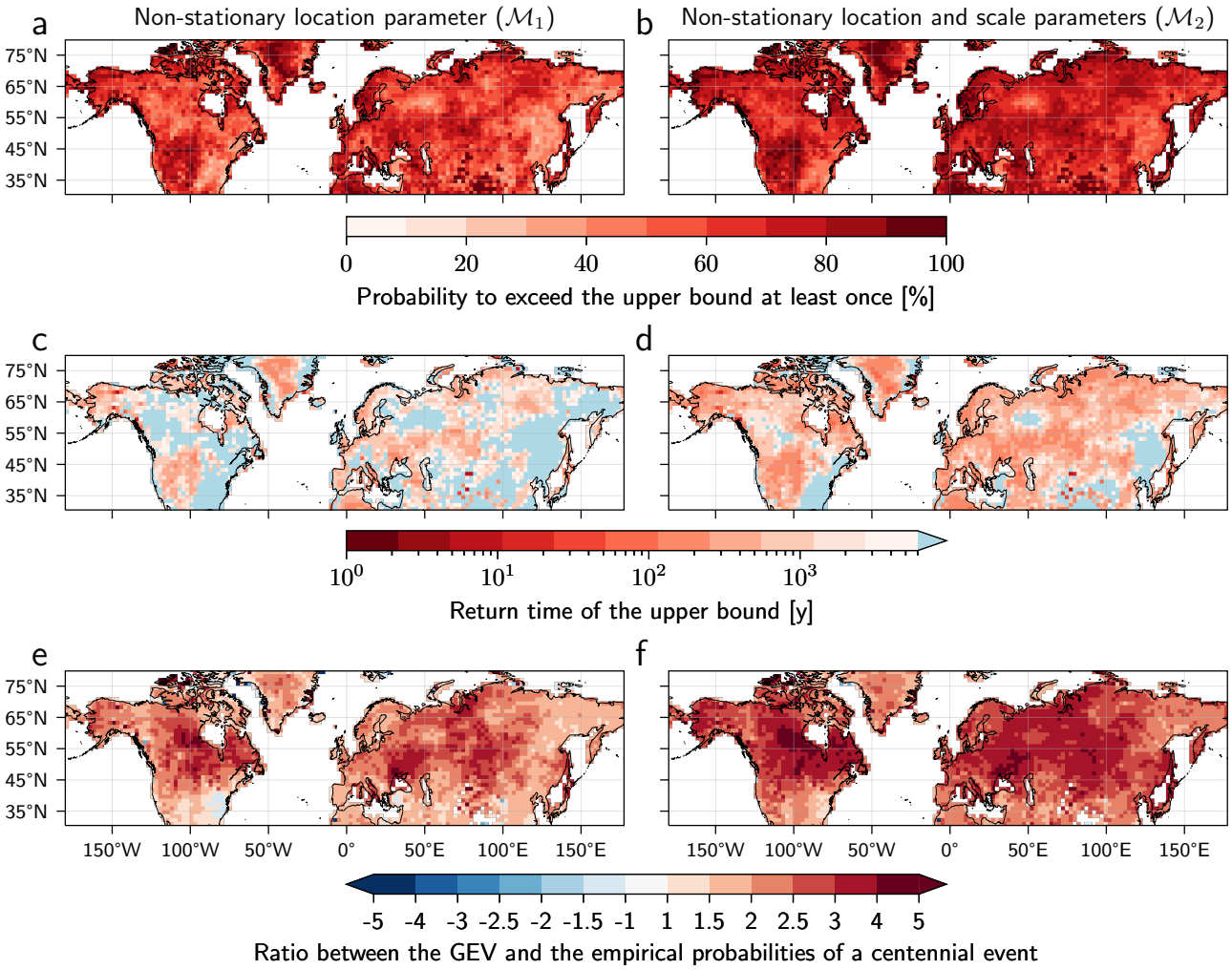


FIG. B4. Same as Fig. 1. with 70 points resampled over the period 1956-2025.

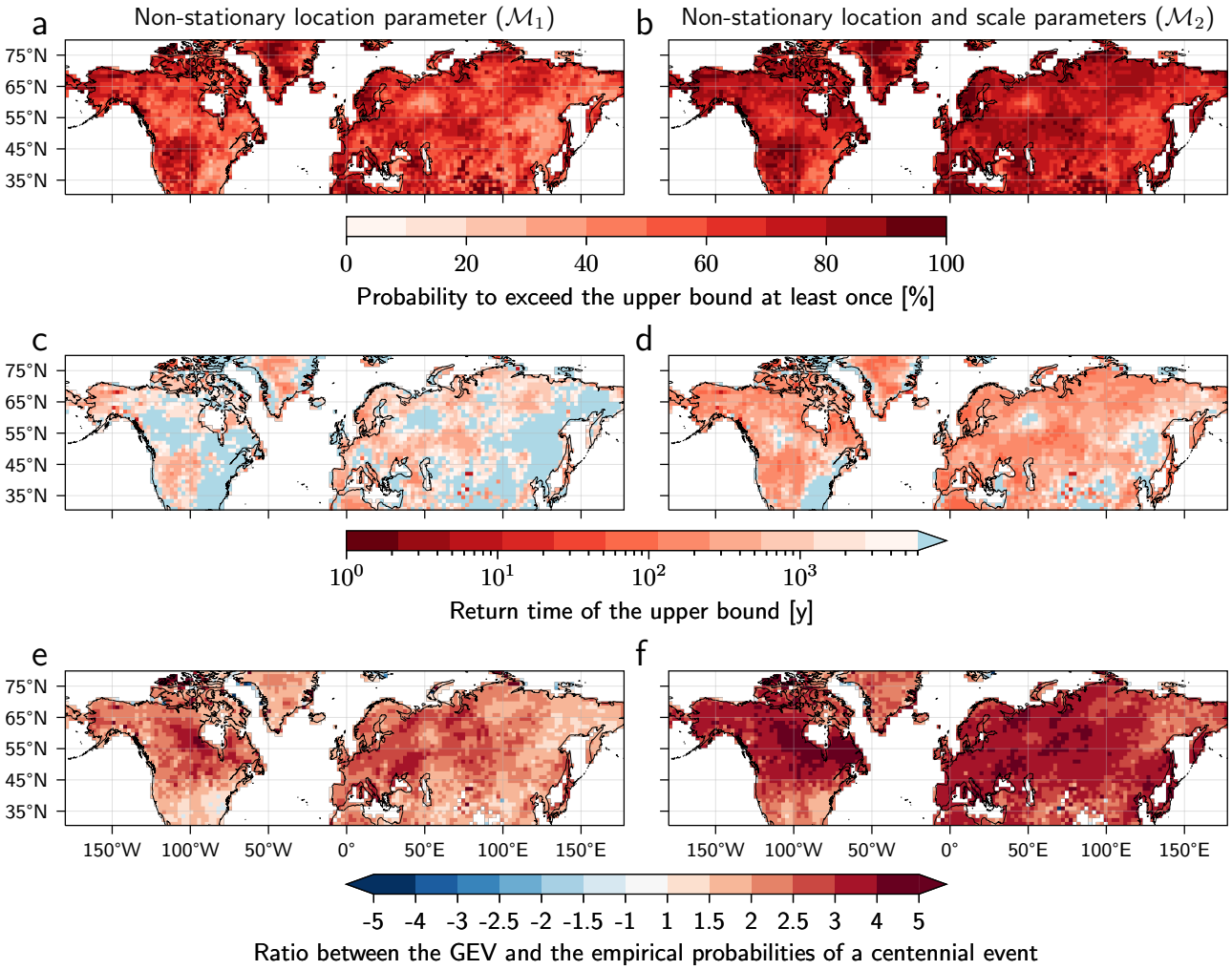


FIG. B5. Same as Fig. 1. with 70 points resampled over the period 1945-2014 with a regional covariate.

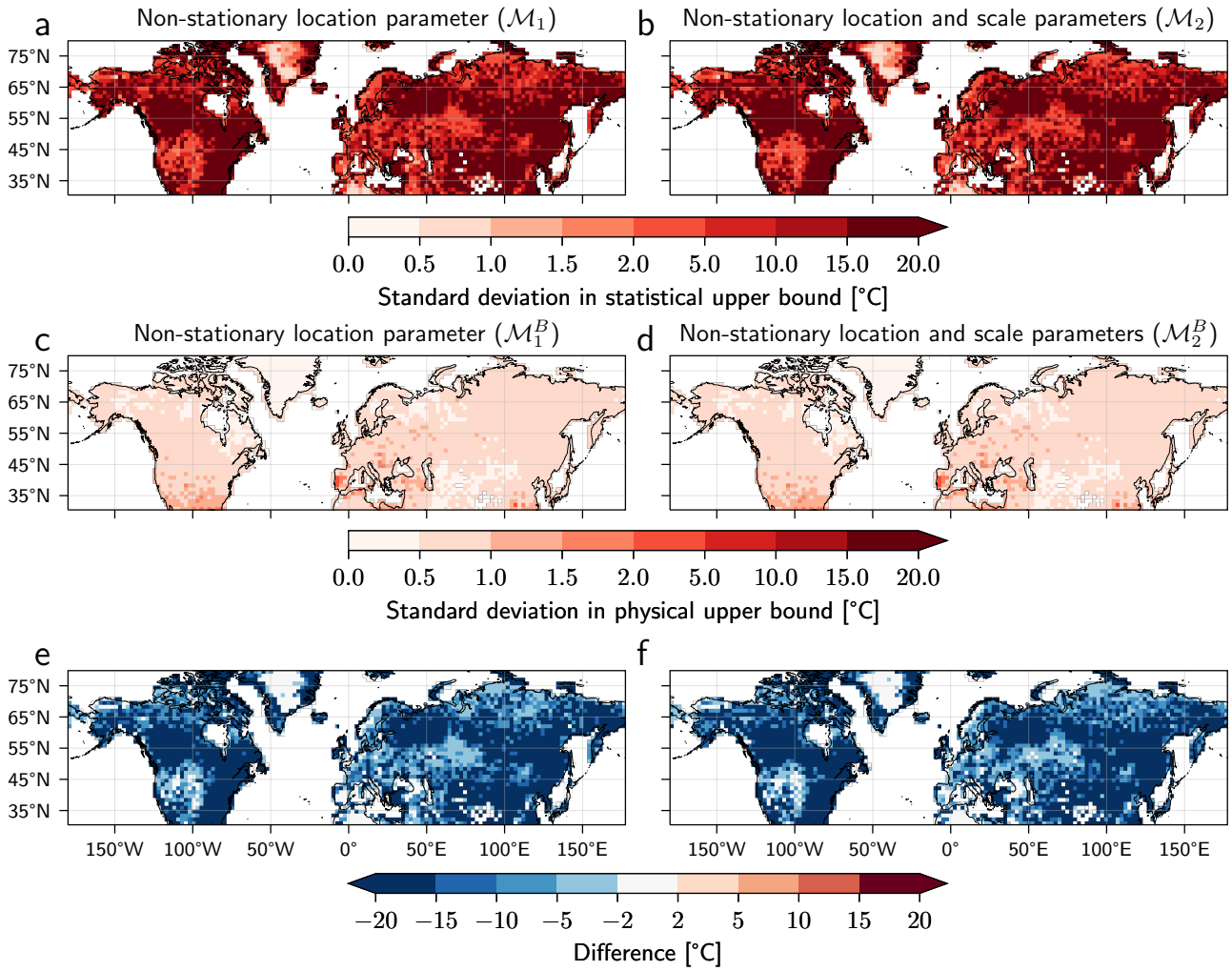


FIG. B6. Precision of the estimation of the upper bounds. Panels a and b show the standard deviation in the estimated statistical upper bounds found for models $\mathcal{M}_1$ and $\mathcal{M}_2$ obtained after 1000 fits with data resampled over the period 1945-2014. Panels c and d show the corresponding standard deviation in the estimated physical upper bounds. Panels e and f show the difference between the two. For all the plots we chose the covariate to be +0.5K above the 1951-1980 average (corresponding roughly to the year 2000).

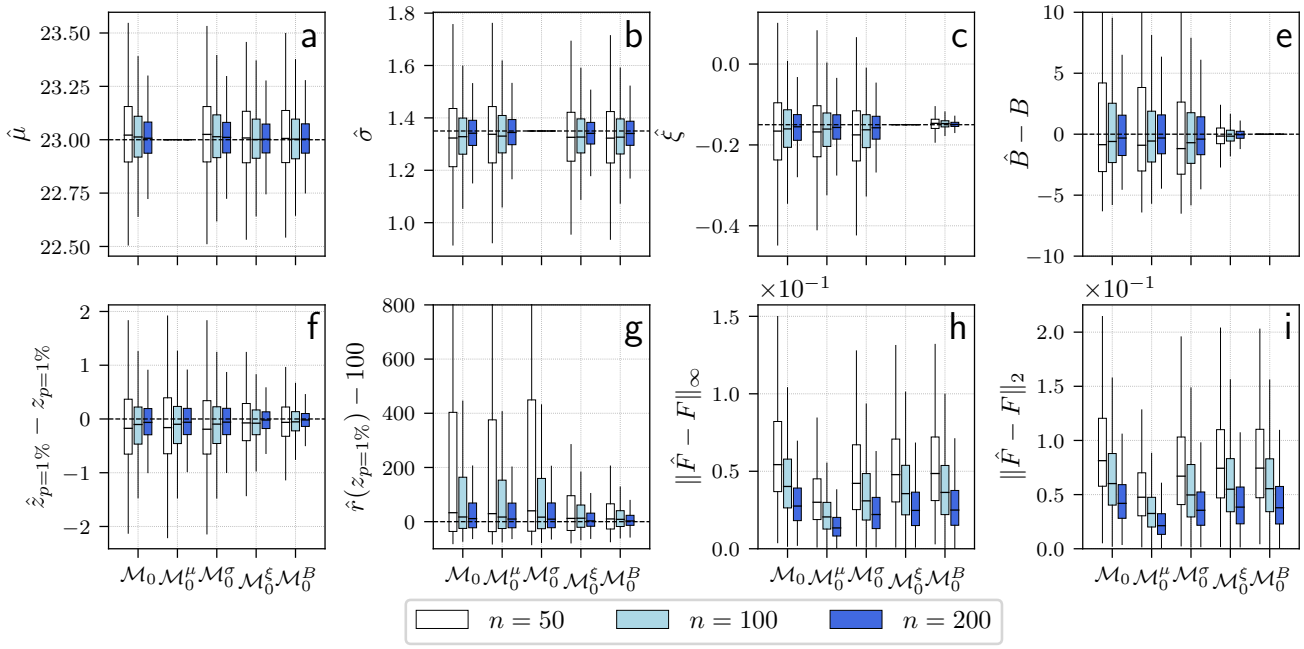


FIG. B7. Same as Figure 3 with synthetic data simulated according to model $\mathcal{M}_0$.

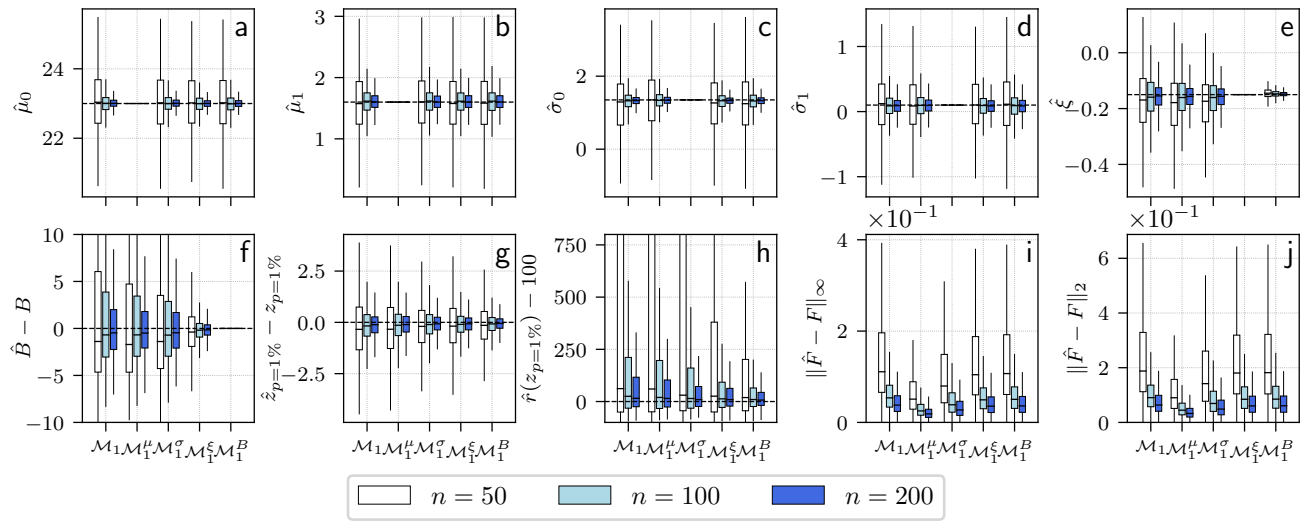


FIG. B8. Same as Figure 3 with synthetic data simulated according to model $\mathcal{M}_2$.

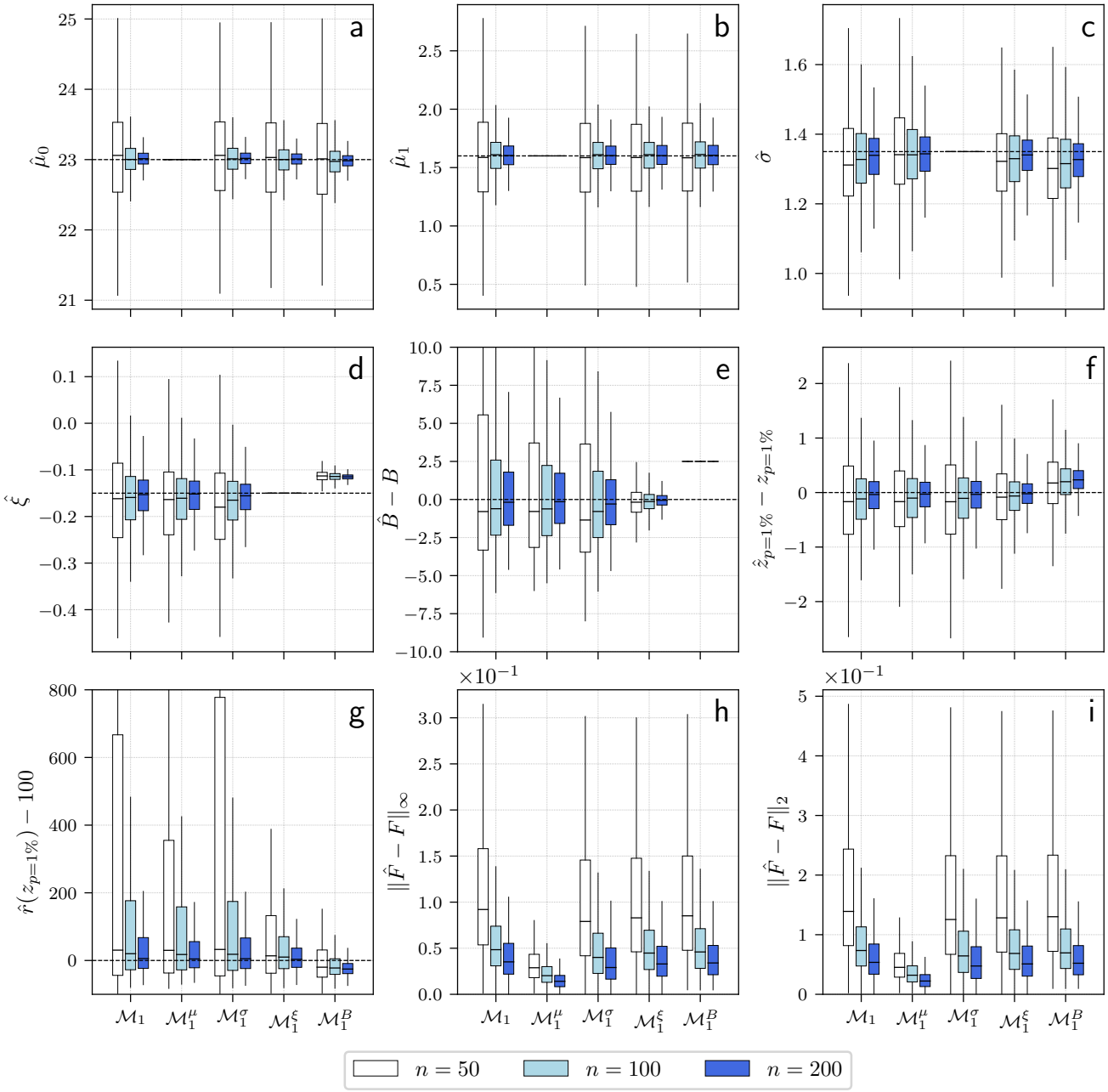


FIG. B9. Same as Figure 3 with an error $B_{err} = 2.5$ imposed for model $\mathcal{M}_1^B$.

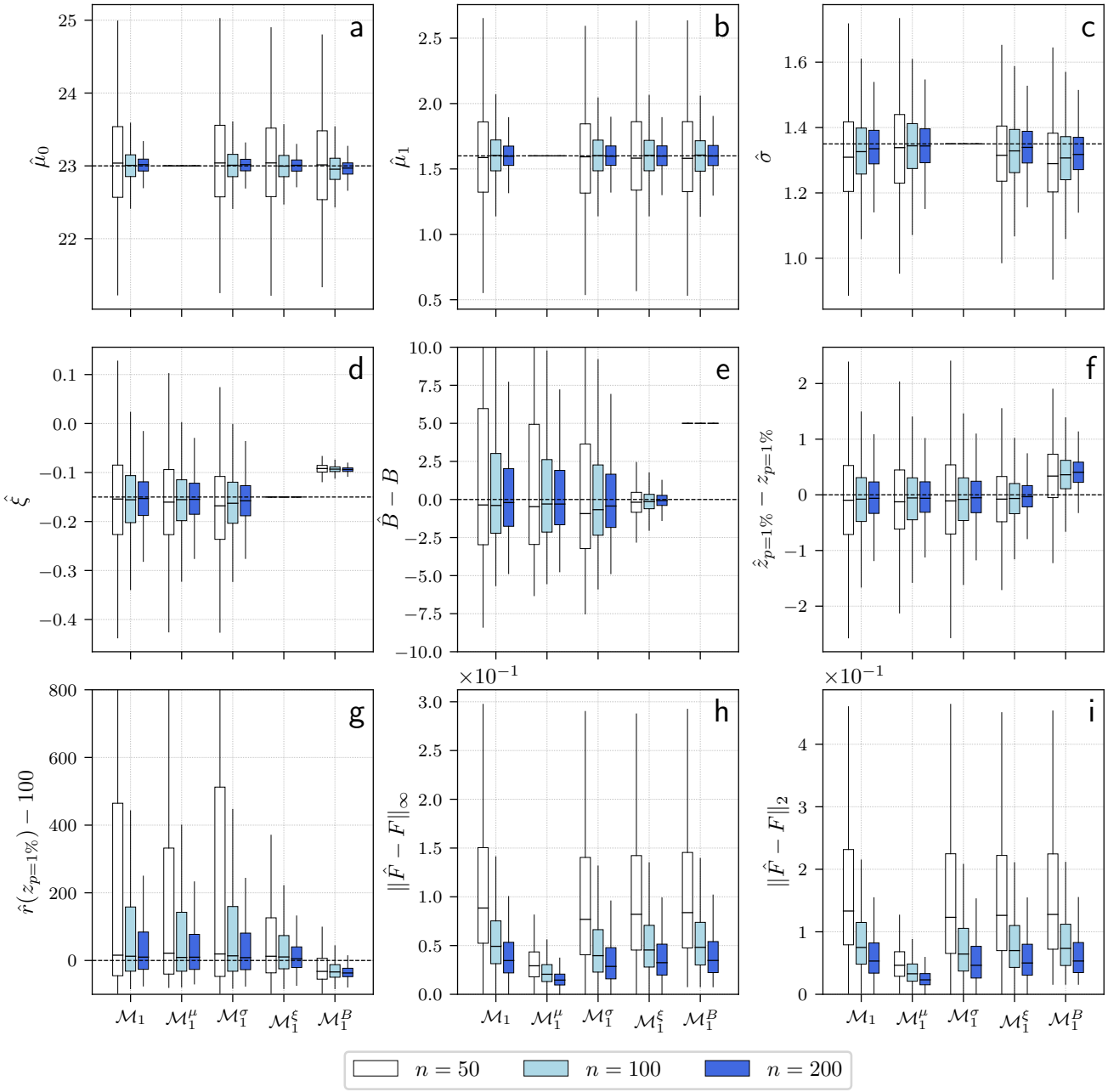


FIG. B10. Same as Figure 3 with an error $B_{err} = 5$ imposed for model $\mathcal{M}_1^B$.

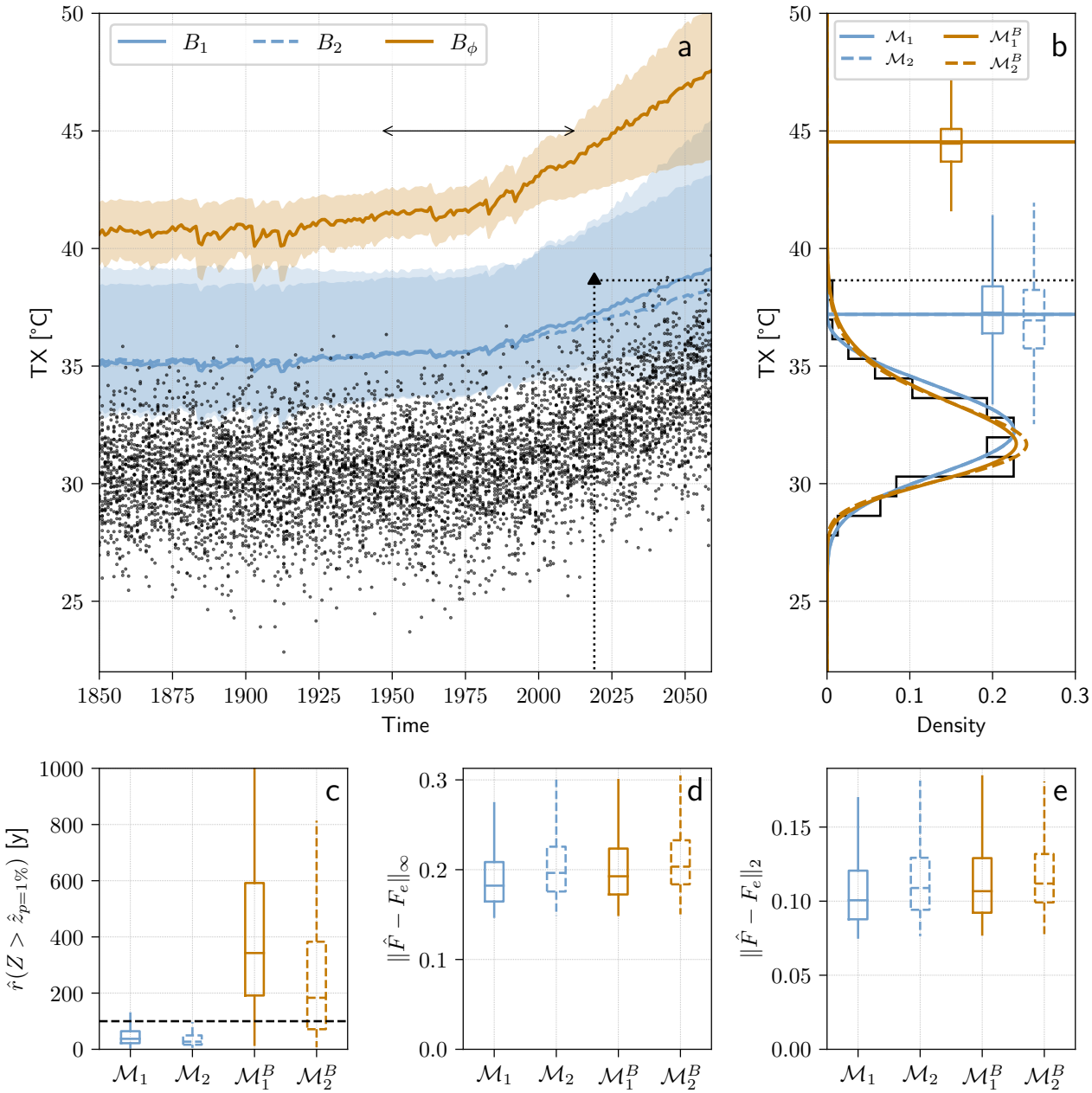


FIG. B11. Same as Fig. 4 with a regional covariate.

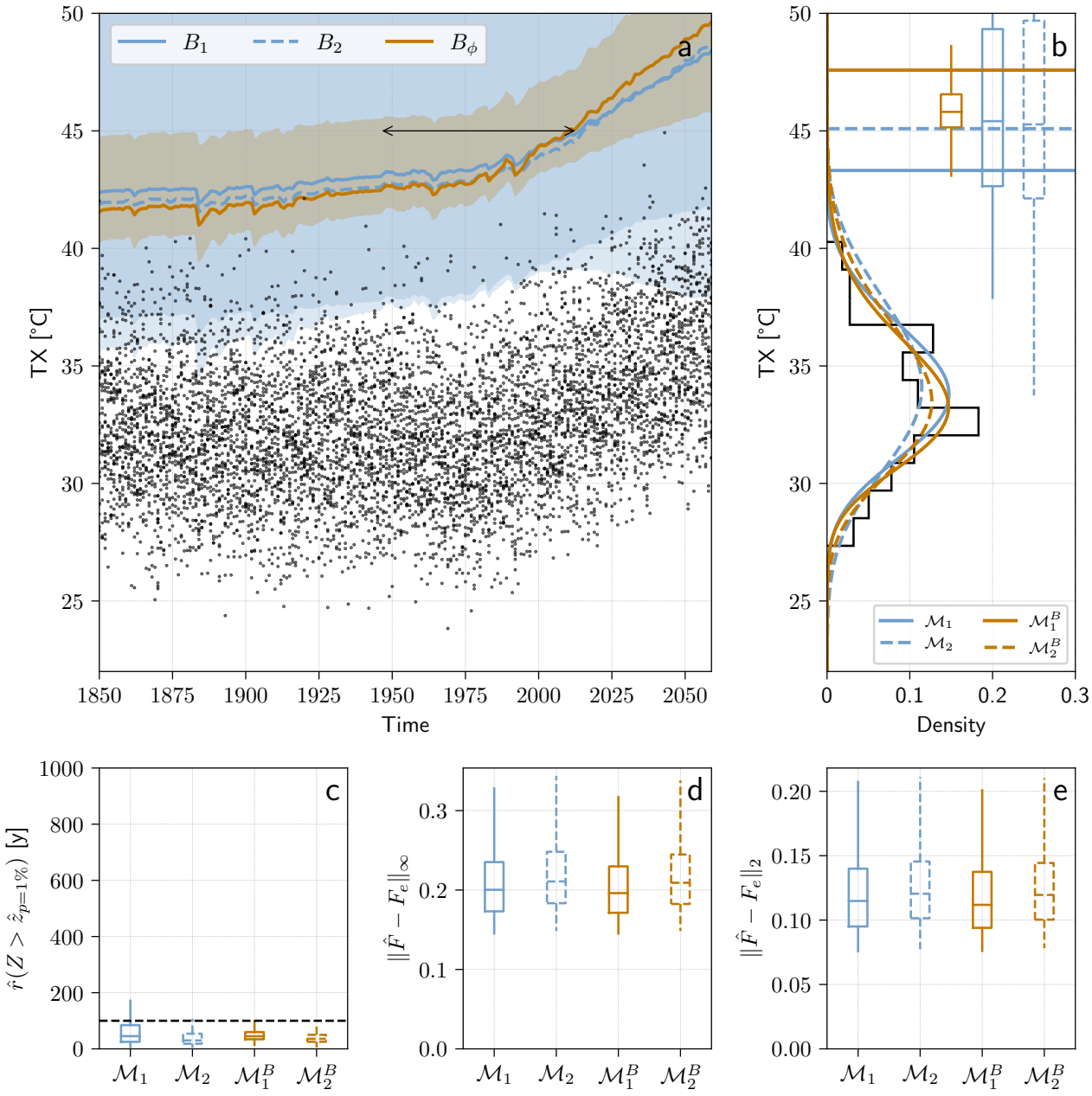


FIG. B12. Same as Fig. 4 with a grid point in Western Russia (lon=47.5°E, lat=57.04°N).

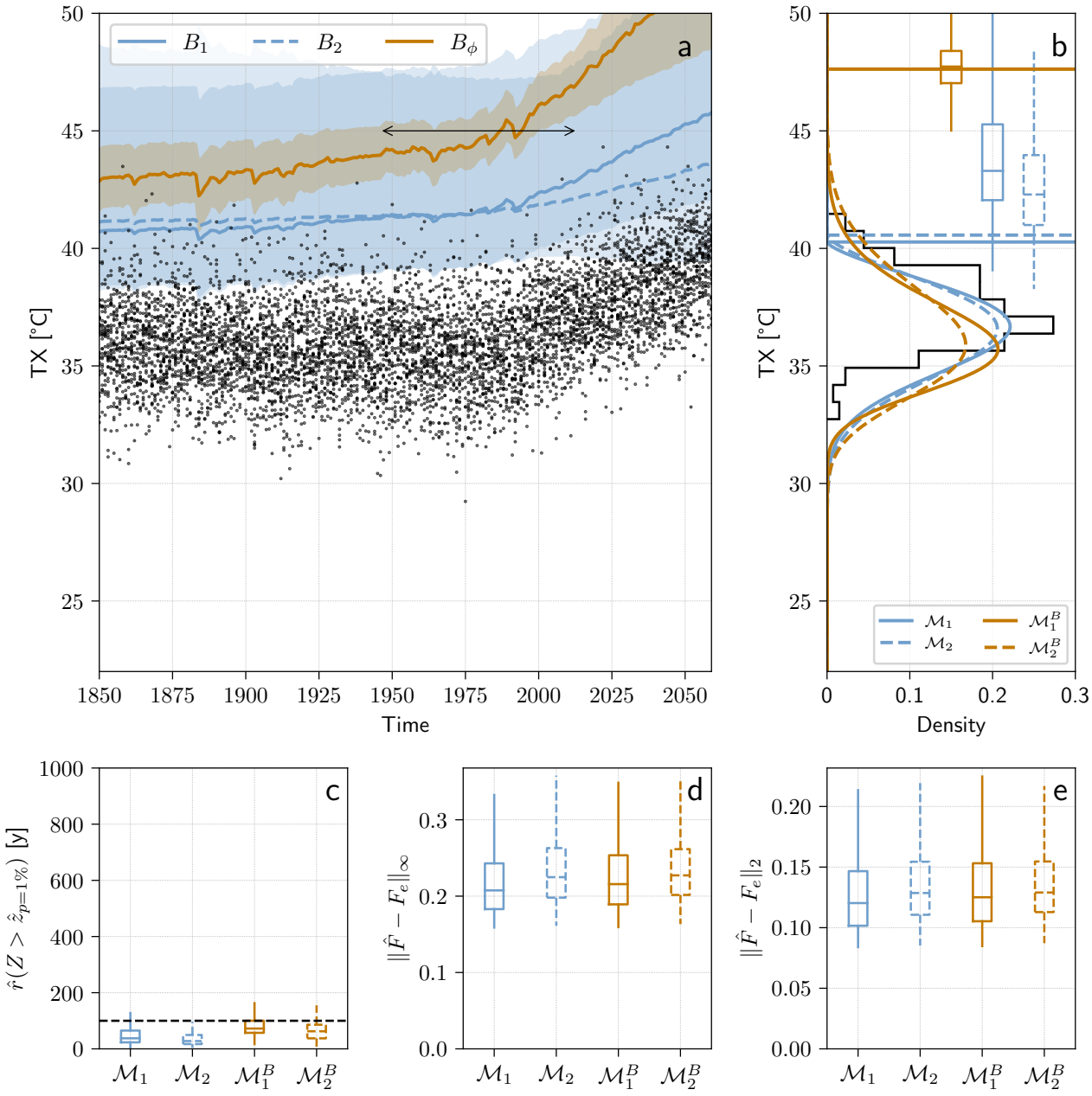


FIG. B13. Same as Fig. 4 with a grid point in Eastern Kazakhstan (lon=80°E, lat=49.44°N).

References

Akaike, H., 1998: Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike*, Springer, 199–213.

Auld, G., G. Hegerl, and I. Papastathopoulos, 2021: Changes in the distribution of observed annual maximum temperatures in europe. *arXiv preprint arXiv:2112.15117*.

Barriopedro, D., R. García-Herrera, C. Ordóñez, D. Miralles, and S. Salcedo-Sanz, 2023: Heat waves: Physical understanding and scientific challenges. *Reviews of Geophysics*, e2022RG000780.

Bastos, A., and Coauthors, 2021: Vulnerability of european ecosystems to two compound dry and hot summers in 2018 and 2019. *Earth system dynamics*, **12 (4)**, 1015–1035.

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers, 2005: *Statistics of Extremes: Theory and Applications*. https://doi.org/10.1002/0470012382.

Boucher, O., and Coauthors, 2020: Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, **12 (7)**, e2019MS002 010.

Brás, T. A., J. Seixas, N. Carvalhais, and J. Jägermeyr, 2021: Severity of drought and heatwave crop losses tripled over the last five decades in europe. *Environmental Research Letters*, **16 (6)**, 065 012.

Breshears, D. D., and Coauthors, 2021: Underappreciated plant vulnerabilities to heat waves. *New Phytologist*, **231 (1)**, 32–39.

Campbell, S., T. A. Remenyi, C. J. White, and F. H. Johnston, 2018: Heatwave and health impact research: A global review. *Health & place*, **53**, 210–218.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer, London, https://doi.org/10.1007/978-1-4471-3675-0, URL http://link.springer.com/10.1007/978-1-4471-3675-0.

Coles, S., J. Bawa, L. Trenner, and P. Dorazio, 2001: *An introduction to statistical modeling of extreme values*, Vol. 208. Springer.

Coumou, D., V. Petoukhov, S. Rahmstorf, S. Petri, and H. J. Schellnhuber, 2014: Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proceedings of the National Academy of Sciences*, **111 (34)**, 12 331–12 336.

Di Capua, G., S. Sparrow, K. Kornhuber, E. Rousi, S. Osprey, D. Wallom, B. van den Hurk, and D. Coumou, 2021: Drivers behind the summer 2010 wave train leading to russian heatwave and pakistan flooding. *npj Climate and Atmospheric Science*, **4 (1)**, 1–14.

Diffenbaugh, N. S., 2020: Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events. *Science Advances*, **6 (12)**, eaay2368.

Dirmeyer, P. A., G. Balsamo, E. M. Blyth, R. Morrison, and H. M. Cooper, 2021: Land-atmosphere interactions exacerbated the drought and heatwave over northern europe during summer 2018. *AGU Advances*, **2 (2)**, e2020AV000 283.

Domeisen, D. I., and Coauthors, 2022: Prediction and projection of heatwaves. *Nature Reviews Earth & Environment*, 1–15.

Dong, B., R. T. Sutton, and L. Shaffrey, 2017: Understanding the rapid summer warming and changes in temperature extremes since the mid-1990s over western europe. *Climate Dynamics*, **48**, 1537–1554.

Fischbacher-Smith, D., 2010: Beyond the worst case scenario:'managing'the risks of extreme events. *Risk Management*, **12**, 1–8.

Fischer, E., S. Sippel, and R. Knutti, 2021: Increasing probability of record-shattering climate extremes. *Nature Climate Change*, **11 (8)**, 689–695.

Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär, 2007: Contribution of land-atmosphere coupling to recent european summer heat waves. *Geophysical Research Letters*, **34 (6)**.

Fischer, E. M., and Coauthors, 2023: Storylines for unprecedented heatwaves based on ensemble boosting. *Nature Communications*, **14 (1)**, 4643.

Gessner, C., E. M. Fischer, U. Beyerle, and R. Knutti, 2021: Very rare heat extremes: quantifying and understanding using ensemble reinitialization. *Journal of Climate*, **34 (16)**, 6619–6634.

Ghil, M., and Coauthors, 2011: Extreme events: dynamics, statistics and prediction. *Nonlinear Processes in Geophysics*, **18 (3)**, 295–350.

Gomes, M. I., 1984: Penultimate limiting forms in extreme value theory. *Annals of the Institute of Statistical Mathematics*, **36 (1)**, 71–85.

Hannart, A., and P. Naveau, 2018: Probabilities of causation of climate changes. *Journal of Climate*, **31**, 5507–5524.

Hatfield, J. L., and J. H. Prueger, 2015: Temperature extremes: Effect on plant growth and development. *Weather and climate extremes*, **10**, 4–10.

Hirschi, M., and Coauthors, 2011: Observational evidence for soil-moisture impact on hot extremes in southeastern europe. *Nature Geoscience*, **4 (1)**, 17–21.

Horton, R. M., J. S. Mankin, C. Lesk, E. Coffel, and C. Raymond, 2016: A review of recent advances in research on extreme heat events. *Current Climate Change Reports*, **2 (4)**, 242–259.

Huang, W. T. K., I. Braithwaite, A. Charlton-Perez, C. Sarran, and T. Sun, 2022: Non-linear response of temperature-related mortality risk to global warming in england and wales. *Environmental Research Letters*, **17 (3)**, 034 017.

Katz, R. W., M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Advances in Water Resources*, **25 (8-12)**, 1287.

Kiriliouk, A., and P. Naveau, 2020: Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Annals of Applied Statistics*, **14 (3)**, 1342–1358.

Kornhuber, K., D. Coumou, E. Vogel, C. Lesk, J. F. Donges, J. Lehmann, and R. M. Horton, 2020: Amplified rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nature Climate Change*, **10 (1)**, 48–53.

Luo, F., L. Wilcox, B. Dong, Q. Su, W. Chen, N. Dunstone, S. Li, and Y. Gao, 2020: Projected near-term changes of temperature extremes in europe and china under different aerosol emissions. *Environmental Research Letters*, **15 (3)**, 034 013.

Masselot, P., and Coauthors, 2023: Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in europe. *The Lancet Planetary Health*, **7 (4)**, e271–e281.

Miralles, D. G., A. J. Teuling, C. C. Van Heerwaarden, and J. Vilà-Guerau de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature geoscience*, **7 (5)**, 345–349.

Miralles, D. G., M. Van Den Berg, A. Teuling, and R. De Jeu, 2012: Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters*, **39 (21)**.

Mitchell, D., K. Kornhuber, C. Huntingford, and P. Uhe, 2019: The day the 2003 european heatwave record was broken. *The Lancet Planetary Health*, **3 (7)**, e290–e292.

National Academies of Sciences Engineering and Medicine, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, Washington, DC, https://doi.org/10.17226/21852.

Naveau, P., A. Hannart, and A. Ribes, 2020: Statistical Methods for Extreme Event Attribution in Climate Science. *Annual Reviews of Statistics and its Application*, https://doi.org/10.1146/annurev-statistics-031219-.

Naveau, P., A. Ribes, F. Zwiers, A. Hannart, A. Tuel, and P. Yiou, 2018: Revising Return Periods for Record Events in a Climate Event Attribution Context. *Journal of Climate*, **31 (9)**, 3411–3422, https://doi.org/10.1175/JCLI-D-16-0752.1.

Noyelle, R., Y. Zhang, P. Yiou, and D. Faranda, 2023: Maximal reachable temperatures for western europe in current climate. *Environmental Research Letters*, **18 (9)**, 094 061.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **128 (581)**, 747–774.

Patterson, M., 2023: North-west europe hottest days are warming twice as fast as mean summer days. *Geophysical Research Letters*, **50 (10)**, e2023GL102 757.

Perkins, S. E., 2015: A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, **164**, 242–267.

Petoukhov, V., S. Rahmstorf, S. Petri, and H. J. Schellnhuber, 2013: Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. *Proceedings of the National Academy of Sciences*, **110 (14)**, 5336–5341.

Philip, S., and Coauthors, 2020: A protocol for probabilistic extreme event attribution analyses. adv stat climatol meteorol oceanogr 6: 177–203.

Philip, S. Y., and Coauthors, 2021: Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the us and canada june 2021. *Earth System Dynamics Discussions*, 1–34.

Pörtner, H.-O., and Coauthors, 2022: Climate change 2022: Impacts, adaptation and vulnerability. *IPCC Sixth Assessment Report*.

Ragone, F., and F. Bouchet, 2021: Rare event algorithm study of extreme warm summers and heatwaves over europe. *Geophysical Research Letters*, **48 (12)**, e2020GL091 197.

Ragone, F., J. Wouters, and F. Bouchet, 2018: Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, **115 (1)**, 24–29.

Rasmijn, L., G. Van der Schrier, R. Bintanja, J. Barkmeijer, A. Sterl, and W. Hazeleger, 2018: Future equivalent of 2010 russian heatwave intensified by weakening soil moisture constraints. *Nature Climate Change*, **8 (5)**, 381–385.

Robin, Y., and A. Ribes, 2020: Nonstationary extreme value analysis for event attribution combining climate models and observations. *Advances in Statistical Climatology, Meteorology and Oceanography*, **6 (2)**, 205–221.

Scher, S., S. Jewson, and G. Messori, 2021: Robust worst-case scenarios from ensemble forecasts. *Weather and forecasting*, **36 (4)**, 1357–1373.

Schwarz, G., 1978: Estimating the dimension of a model. *The annals of statistics*, 461–464.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, **99 (3-4)**, 125–161.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land–atmosphere coupling and climate change in europe. *Nature*, **443 (7108)**, 205–209.

Seneviratne, S. I., and Coauthors, 2021: *11 Chapter 11: Weather and climate extreme events in a changing climate*.

Seong, M.-G., S.-K. Min, Y.-H. Kim, X. Zhang, and Y. Sun, 2021: Anthropogenic greenhouse gas and aerosol contributions to extreme temperature changes during 1951–2015. *Journal of Climate*, **34 (3)**, 857–870.

Sillmann, J., L. Pozzoli, E. Vignati, S. Kloster, and J. Feichter, 2013: Aerosol effect on climate extremes in europe under different future scenarios. *Geophysical Research Letters*, **40 (10)**, 2290–2295.

Thompson, V., and Coauthors, 2022: The 2021 western north america heat wave among the most extreme events ever recorded globally. *Science advances*, **8 (18)**, eabm6860.

Van Oldenborgh, G. J., and Coauthors, 2021: Pathways and pitfalls in extreme event attribution. *Climatic Change*, **166 (1-2)**, 13.

Van Oldenborgh, G. J., and Coauthors, 2022: Attributing and projecting heatwaves is hard: We can do better. *Earth's Future*, **10 (6)**, e2021EF002 271.

Vautard, R., and Coauthors, 2020: Human contribution to the record-breaking june and july 2019 heatwaves in western europe. *Environmental Research Letters*, **15 (9)**, 094 077.

Vautard, R., and Coauthors, 2023: Heat extremes in western europe are increasing faster than simulated due to missed atmospheric circulation trends.

Westervelt, D. M., and Coauthors, 2020: Local and remote mean and extreme temperature response to regional aerosol emissions reductions. *Atmospheric Chemistry and Physics*, **20 (5)**, 3009–3027.

Worms, J., and P. Naveau, 2022: Record events attribution in climate studies. *Environmetrics*, https://doi.org/10.1002/env.2777, https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.2777.

Wouters, J., and F. Bouchet, 2016: Rare event computation in deterministic chaotic systems using genealogical particle analysis. *Journal of Physics A: Mathematical and Theoretical*, **49 (37)**, 374 002.

Yiou, P., and A. Jézéquel, 2020: Simulation of extreme heat waves with empirical¡? xmltex\break?¿ importance sampling. *Geoscientific Model Development*, **13 (2)**, 763–781.

Zeder, J., S. Sippel, O. C. Pasche, S. Engelke, and E. M. Fischer, 2023: The effect of a short observational record on the statistics of temperature extremes. *Geophysical Research Letters*, **50 (16)**, e2023GL104 090.

Zhang, Y., and W. R. Boos, 2023: An upper bound for extreme temperatures over midlatitude land. *Proceedings of the National Academy of Sciences*, **120 (12)**, e2215278 120.

Zuo, J., S. Pullen, J. Palmer, H. Bennetts, N. Chileshe, and T. Ma, 2015: Impacts of heat waves and corresponding measures: a review. *Journal of Cleaner Production*, **92**, 1–12.

3.3 Perspectives

In this work I make a proposition to avoid the shortcomings of GEV fits for estimating low probability events as outlined by Zeder et al. (2023). Extreme value theory (EVT) is a powerful and versatile theory which gives general results for any underlying distributions. Nonetheless, it is clear at this point that the convergence to a correct estimate of an EVT-based distribution for low probability events with the typical sample size of climate data (around 70-80 or less yearly maxima for most weather stations) is not ensured. I propose a way to alleviate this issue and avoid the catastrophic error of giving a null probability to events that can actually happen, i.e. to avoid 'climatic surprises'.

Although the proposition made here has interesting properties when applied to synthetic GEV data, when applied to a large ensemble of a climate model it also has several drawbacks. In particular, although it avoids the systematic underestimation of the probability of centennial events, it now tends to overestimate the probability of some events. The value of the upper bound imposed on the GEV fits is a key element of the precision of the method as shown in synthetic GEV data. It is therefore crucial to estimate the upper bound as precisely as possible. The main contribution of this work is to propose the idea of imposing the upper bound, not the method by which I estimate this upper bound physically using a modified version of the formula proposed by Zhang and Boos (2023). As stated in the manuscript, the method proposed would be more suited in a Bayesian context: the upper bound is itself unknown and therefore constitutes a random variable. One should therefore have a prior distribution estimate of this value. Whether this estimation is based on the method of computation proposed here or on another — especially outside the mid-latitudes — is a secondary, yet important, point.

Finally, I want to emphasize that, as detailed in the conclusion of the article, under- or overestimating the probability of very extreme events is not a symmetric error. Infrastructures are often built with respect to a very low probability threshold and/or a worst case scenario. Therefore, underestimating the probability of a very extreme event is much worse (especially if the event is said to be impossible) than overestimating it (Taleb 2010). The questions of very extreme events and the estimation of worst case scenarios are therefore not only technical but also have strong societal implications (Sutton 2019).

Summary

Context and goals

The EVT-based fits on limited-size samples underestimate the probability of very intense events. In particular, some of these events are said to be impossible according to the statistical distribution but still occur in models and in the real world.

Methods

I propose the idea of physically constraining the fit of EVT-based statistical distributions for extreme temperatures. The upper bound of the distribution is imposed by a physically realistic value derived from properties of the mid-atmosphere and surface humidity. I use outputs of a large ensemble of a state-of-the-art climate model over historical and near future periods to check the accuracy of this new fit.

Results

Imposing the upper bound partially alleviates the issue of the underestimation of very extreme events — typically centennial events — and, more importantly, avoids the appearance of climatic surprises, i.e. events that are said to be impossible but that still occur. The method proposed tends to overestimate very extreme events if the upper bound imposed is too high. However, under- or over-estimating very extreme events is not a symmetric error. While the former leads to linearly increasing costs of preparation, the latter can lead to catastrophic damages on societies that would face events they thought to be impossible.

Chapter 4

Rare events algorithms and their application to geophysical systems

4.1 Introduction

This chapter aims to present a review of both the theoretical paradigms currently available to understand the trajectories leading to rare events in dynamical systems in general and geophysical fluid systems in particular, and the techniques that have been employed in geophysical fluid models to sample these trajectories. Due to the practical orientation of this PhD thesis, I refer only briefly to the mathematical literature or to works with applications outside climate or meteorological sciences.

It is only recently that theoretical advances and algorithmic techniques have been developed to understand better not only the return statistics of extreme events, but also the dynamical path leading to extremes in physical systems (e.g. Vanden-Eijnden et al. (2010), Grafke and Vanden-Eijnden (2019), and Bouchet et al. (2019)). I call generally the techniques aiming to simulate rare events as **rare events algorithms**, even though this term is usually restricted to importance sampling or splitting methods (Rubino, Tuffin, et al. 2009). Rare events algorithms have been applied in many different systems: percolation (Adams et al. 2008), molecular dynamics and computational chemistry (E et al. 2002; Noé et al. 2009; Vanden-Eijnden et al. 2010), turbulence (Grafke et al. 2013), planetary dynamics (Abbot et al. 2021) etc. With the exception of turbulence problems, most of them have not been applied to systems which bear similarities with geophysical systems, namely: time- and space-continuity, high dimensionality and no spectral gap.

This chapter is organized as follows. Section 4.2 presents a review of the application of rare events algorithms to geophysical systems. Section 4.3 details mathematical concepts that I keep referring to when it comes to understanding the path leading to extreme events in dynamical systems. Section 4.4 gives a description of the algorithms employed in the literature, with a focus on so-called splitting methods. As splitting methods are by far the most employed, section 4.5 presents

the current challenges associated with their application and some propositions to overcome them.

The reader not interested in the details of the implementation of rare events algorithms and their technical shortcomings may skip this chapter. The algorithm employed in chapters 6 and 7 is presented in section 4.4.3c, but its main characteristics are recalled shortly in chapter 6.

4.2 Review of the use of rare events algorithms in geophysical systems

In the context of geophysical systems, rare events algorithms have been used for two main purposes. The first one concerns the study of transitions from a fixed point to another fixed point, or from a basin of an attractor to the basin of another attractor. In other words, this first application deals with the question of **multistability** of geophysical systems. The second purpose seeks to investigate the occurrences of **extreme values of an observable** of interest, without changing the attractor itself. Its purpose is to sample regions of the system's attractor which are seldom visited but that do not lead a priori to a bifurcation in the overall properties of the system. The review is summarized in table 4.1.

4.2.1 Sampling rare transitions between metastable states

Complex systems can have multiple states of equilibrium for the same values of the forcing parameters. Similarly, non-equilibrium systems can have several attractors corresponding to disjoint basins of attraction. The question that naturally arises is therefore how and how often does the system go from one of these states to another. Indeed, if the parameters of the system do not change, then each state is stable and it is not possible to go spontaneously from one attractor to another. If now a more realistic model with — for example — a stochastic noise which may represent disturbance from inside or outside of the system is considered, then one recognizes intuitively that a sufficiently strong noise could force a transition into another basin of attraction. It is therefore natural to understand this crossing of the barrier between basins of attraction as a rare event and to use rare events algorithms to sample these transitions.

Weare (2009) was one of the first to develop such an approach, using path sampling methods. He studied the bimodality of the Kuroshio current using a parallel marginalization step in conjunction with a hybrid Monte Carlo scheme to improve samples generated by standard particle filters. Laurie and Bouchet (2015) studied rare transitions in a barotropic quasi-geostrophic model with a stochastic forcing. They used the instanton formalism and apply a numerical optimization algorithm to compute the most probable rare transition between two states: vortex dipoles and zonal jets. Bouchet et al. (2019) studied multistability of jet dynamics in a barotropic

β-plane model of atmospheric dynamics. They showed the presence of very rare transitions between an equilibrium with 2 and an equilibrium with 3 jets and the typical phenomenology associated. In Simonnet et al. (2021) the same authors make a full statistical mechanics study of the same barotropic β-plane quasi-geostrophic model.

Reviewing numerical computation of rare events via large deviation theory, with a particular focus on instantons simulation, Grafke and Vanden-Eijnden (2019) show the effectiveness of the instanton method on a stochastic version of the Charney and DeVore (1979) model, which exhibits two fixed points corresponding to zonal and blocked flows of the mid-latitude atmospheric circulation. They show that they are also able to study the relative stability of the two configurations: the zonal state being exponentially preferred in the low-noise limit. Finkel et al. (2020) used a framework inspired by Transition Path Theory (TPT) to study a low-order model of Sudden Stratospheric Warming (SSW) composed of two stable states. In Finkel et al. (2021a) the same authors studied a more complex model of SSW with 75 degrees of freedom, and applied a procedure they call Dynamic Galerkin Approximation (DGA) (Finkel et al. 2021b) with only short simulations to compute the quantities of interest from TPT. Baars et al. (2021) used a modified version of the Adaptive Multilevel Splitting (AMS) algorithm (see below), what they call Trajectory-Adaptive Multilevel Sampling (TAMS) algorithm, to apply it to a high-dimensional system. They apply their method on a 2D model of the Atlantic Meridional Overturning Circulation (AMOC) to compute transition paths and probabilities of collapse. Cini et al. (2024) also studied the noise-induced collapse of the AMOC driven by internal climate variability and used the Giardina-Kurchan-Tailleur-Lecomte (GKTL) rare events algorithm (see below) in an intermediate complexity model.

4.2.2 Sampling extreme values of an observable

The second purpose for which authors used rare events algorithms in the geophysical context is to simulate trajectories which lead to extreme values of an observable of interest. This purpose is different from investigating multistability for at least three reasons. The first reason is that 'extreme states' are usually 'unstable': contrary to navigating from one basin of attraction to another, one has to deal with the fact that the system tends to be repelled by these states (else they would not be rare). The second reason is that when one investigates multistability, especially in simple models, the two fixed points or two basins of attractions that constitute the starting and the ending points of the rare trajectories are generally known. In the contrary for extreme values of an observable, one does not always have strong clues on the phenomenology of the extreme states, i.e. also on their position in the phase space. Finally, several separated regions of the phase space may a priori correspond to extreme values of the same observable.

Wouters and Bouchet (2016) used a genealogical particle analysis algorithm to study the Lorenz (1996) toy model of the atmosphere and showed several implementation issues on a Orstein-Uhlenbeck process for which exact probabilities can be calculated. They used an analog of energy and computed the probability that it reaches extreme values. Ragone et al. (2018) used the Giardina-Kurchan-Tailleur-

Lecomte (GKTL) algorithm to study extremes of summer temperatures with the intermediate complexity climate model PlaSim. They show that the algorithm is able to increase massively the amount of available statistics of warm extremes in a perpetual summer setting of the model. In Ragone and Bouchet (2021) the same authors used this methodology with — for the first time — a state-of-the-art climate model (CESM1.2.2) and are also able to produce hundreds of times more extremes than a control run with the same computational cost. Moreover, the abundance of statistics allow them to identify precisely teleconnection patterns associated with warm summers in Europe.

Working on previous ideas from Hoffman et al. (2006a) and Hoffman et al. (2006b), Plotkin et al. (2019) investigated the process of rapid intensification of two tropical cyclones by applying an action minimization algorithm to nudge an atmospheric forecasting model into forming more intense tropical cyclones. The algorithm is able to determine the optimal (small) perturbations to make the tropical cyclones more intense. Remarkably, the optimal perturbations obtained are far from trivial and give interesting insights in the phenomenology of tropical cyclones rapid intensification. Webber et al. (2019) also investigated the use of a rare events algorithm to study tropical cyclones. More generally, they emphasized the need of adaptation of rare events algorithms to mesoscale weather and contrary to Plotkin et al. (2019) investigate not only the likeliest path but statistics of extreme mesoscale weather. They used what they call the Quantile Diffusion Monte Carlo (QDMC) algorithm to sample extreme tail behavior of two historical tropical cyclones.

Yiou and Jézéquel (2020) used an importance sampling algorithm associated with a Stochastic Weather Generator (SWG) based on circulation analogues to simulate an ensemble of heat waves with a large return period. Using this Markov chain model and biasing the probability matrix in favours of hot days, they are able to simulate unseen but possible trajectories of the atmospheric circulation leading to heatwaves in a computationally efficient way. Yiou et al. (2023) used this method to sample extreme 15-days heatwaves in a SWG for a period of time corresponding to the 2024 Paris Olympics. Carney et al. (2020) compared two rare events algorithms with EVT and brute force Monte Carlo simulation estimations of small probabilities. Using three dynamical models (the Ornstein-Uhlenbeck process, the Lorenz (1996) system and a simplified GCM (Plasim)) as test beds, they study how well the rare event probability estimation of each technique compares to a gold standard afforded by a very long run control. They find that classical extreme value theory methods outperform the other methods for estimating small probabilities. Wouters et al. (2023) tested the application of a rare events algorithm in an intermediate complexity model to sample seasonal precipitation extremes. Sauer et al. (2024) used the GKTL rare events algorithm in an intermediate complexity model (PlaSim-LSG) to sample extreme negative summer pan-Arctic sear ice area anomalies.

Gessner et al. (2021) developed a storyline approach based on the resampling of a 5000-yr pre-industrial climate model simulation (CESM1) called ensemble boosting generating large samples of reinitialized extreme heatwaves, but without providing the associated probabilities. In Gessner et al. (2022), the same authors extended the method to generate coherent climate model-based drought storylines of different

intensities and for different locations. Fischer et al. (2023) provided a review of the method with examples for very intense heatwave events.

Paper	Phenomena	Purpose	Method	Algorithm
Sauer et al. (2024)	Sea ice	Extremes	Splitting	GKTL
Cini et al. (2024)	AMOC	Multistability	Splitting	GKTL
Wouters et al. (2023)	Precipitations	Extremes	Splitting	GKTL
Fischer et al. (2023)	Heatwaves	Extremes	Splitting	Ensemble boosting
Yiou et al. (2023)	Heatwaves	Extremes	Biased dynamic	Biasing
Gessner et al. (2022)	Droughts	Extremes	Splitting	Ensemble boosting
Simonnet et al. (2021)	Jets dynamics	Multistability	Splitting	AMS
Gessner et al. (2021)	Heatwaves	Extremes	Splitting	Ensemble boosting
Ragone and Bouchet (2021)	Heatwaves	Extremes	Splitting	GKTL
Baars et al. (2021)	AMOC	Multistability	Splitting	AMS
Finkel et al. (2021a)	SSW	Multistability	Direct computation	DGA
Carney et al. (2020)	Heatwaves	Extremes	Splitting	GKTL, DMC
Yiou and Jézéquel (2020)	Heatwaves	Extremes	Biased dynamic	Biasing
Finkel et al. (2020)	SSW	Multistability	Direct computation	Discret. of PDEs
Grafke and Vanden-Eijnden (2019)	Synoptic circulation	Multistability	Direct computation	gMAM
Bouchet et al. (2019)	Jets dynamics	Multistability	Splitting	AMS
Webber et al. (2019)	Tropical cyclones	Extremes	Splitting	QDMC
Plotkin et al. (2019)	Tropical cyclones	Extremes	Biased dynamic	Action minimization
Ragone et al. (2018)	Heat waves	Extremes	Splitting	GKTL
Wouters and Bouchet (2016)	Lorenz96	Extremes	Splitting	DMC
Laurie and Bouchet (2015)	Synoptic circulation	Multistability	Direct computation	MAM
Weare (2009)	Kuroshio	Multistability	Biased dynamic	Particle filters

Table 4.1: Summary of the applications of rare events algorithms in geophysical systems.

4.3 Mathematical elements

This section introduces theoretical tools for understanding the simulation of rare events in geophysical systems. I first describe the formalism of Stochastic Differential Equations (SDEs) and discuss why this formalism is suited to study rare events. Then I present two theories aiming to understand the dynamical path leading to rare events: Transition Path theory (TPT) and instanton theory. These two theories are not mutually exclusive and may be seen as two ways of describing the same object. Finally, I give an example of the key quantities of interest of these two theories with a 1D Brownian motion, also known as an Orstein-Uhlenbeck process.

4.3.1 Stochastic differential equations formalism

Classicaly, the dynamics of the atmosphere and the oceans is described using Partial Differential Equations (PDEs). Models are therefore usually fully deterministic, even though there is a growing interest in representing sub-grid processes using stochastic terms (Palmer 2019). The results that I present in the following are mathematically

valid for systems with an explicit stochastic term in the limit of small noise. It is therefore reasonable to consider them in the context of meteorological, oceanic and climate models for which it is natural to assume that simulation results should be robust to the addition of such a small noise.

Using a suitable basis of functions and truncating at a certain threshold, PDEs can be transformed into Ordinary Differential Equations (ODEs) for which there is no explicit spatial continuity. When simulating rare events numerically, this difference does not exist anymore as any PDEs has to be truncated to be simulated on a computer. By adding a stochastic term to these ODEs, one obtains Stochastic Differential Equations (SDEs). I choose to use this formalism because it has a broad generality and it is one of the most employed in the literature. The results are also stronger on the mathematical side in this case (Grafke and Vanden-Eijnden 2019).

Let $X_t \in \mathbb{R}^d$ represents the state of the system in the phase space at time t, with d the number of dimensions needed to describe the system. Consider a SDE in X_t composed of a deterministic part $b : X_t \mapsto b(X_t) \in \mathbb{R}^d$ and an additive stochastic term proportional to $\sigma : X_t \mapsto \sigma(X_t) \in \mathbb{R}^{d\times d}$:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t. \tag{4.1}$$

In the language of SDEs, b is called the drift and σ the diffusion matrix. $W_t \in \mathbb{R}^d$ is a Wiener process, also called a Brownian motion, so that dW_t is what is commonly referred to as a white noise in physics. If the diffusion matrix $\sigma(X_t)$ is not diagonal, the added noise may be the same for several variables of the system. Entire lines of the diffusion matrix may be null to represent the effect of stochasticity on only some components of the system.

This equation is supposed to be able to represent the evolution of the state of the system. The drift term $b(X_t)$ represents the deterministic part of the dynamics, typically the Navier-Stokes equations in a GCM. The diffusion term $\sigma(X_t)dW_t$ represents the effect of unresolved processes on the dynamics of the system — for example sub-grid processes in the context of geophysical flows simulation. The form 4.1 assumes that the stochasticity is additive with respect to the deterministic part.

I am interested in statistics of observables computed on this system. Observables are smooth functions of phase-space variables $f : \mathbb{R}^d \to \mathbb{R}$ which are of physical interest. Mean temperature over a given area may be seen, for example, as an observable of a GCM output. Associated with equation 4.1 is an operator $\mathcal{L}$ which describes the evolution of observable functions forward in time following a trajectory. By definition:

$$\mathcal{L}f(x) = \frac{d}{dt}\mathbb{E}[f(X_t)|X_0 = x]|_{t=0}. \tag{4.2}$$

Ito's lemma — which is the equivalent of the chain rule for SDEs — allows to represent $\mathcal{L}$ as a partial differential operator:

$$\mathcal{L}f(x) = \sum_i b_i(x)\frac{\partial f(x)}{\partial x_i} + \sum_{i,j} D_{ij}\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \tag{4.3}$$

where $D_{ij} = \frac{1}{2}(\sigma\sigma^T)_{ij}$. The generator $\mathcal{L}$ has the remarkable property of providing statistics along simulated paths as the solutions to PDEs, i.e. without stochastic terms.

The stochastic process described by equation 4.1 admits a time-dependent probability density $\rho(x,t)$ which can be derived from the generator. The density spreads out from its initial point over time according to the Fokker-Planck equation, which can be written in terms of the adjoint $\mathcal{L}^*$ of the generator:

$$\frac{\partial \rho(x,t)}{\partial t} = \mathcal{L}^* \rho(x,t) \tag{4.4}$$

where $\mathcal{L}^*$ is defined by: $< \mathcal{L}f, g >=< f, \mathcal{L}^* g >$ for two observables f and g and $< \cdot, \cdot >$ is the euclidean inner product in $L^2(\mathbb{R}^d)$. $\mathcal{L}^*$ can then be shown to be equal to:

$$\begin{aligned} \mathcal{L}^* f(x) &= \sum_i \frac{\partial}{\partial x_i} \left[-b_i(x) f(x,t) + \sum_j \frac{\partial}{\partial x_j} (f(x,t) D_{ij}(x)) \right] \\ &= \nabla \cdot [-b(x) f(x,t) + \nabla \cdot (f(x,t) D(x))] \end{aligned} \tag{4.5}$$

Assuming that the process is ergodic (cf. the discussion in chapter 1 section a)), the density forgets the initial condition and stabilizes into a long-term stationary probability density $\rho_S(x)$ solution of $\mathcal{L}^* \rho_S = 0$. This probability can be approximated by either simulating the SDE for a long time and making statistics, or directly solving the stationary PDE $\mathcal{L}^* \rho_S = 0$ with the normalization: $\int \rho_S(x) dx = 1$.

4.3.2 Transition path theory

Initially developed for molecular simulation (Vanden-Eijnden et al. 2010), Transition Path Theory (TPT) has since been used in various domains to understand the statistics of trajectories in the phase space leading to rare events of interest. I present the key concepts of the theory as in Vanden-Eijnden (2006). In this PhD I do not use techniques from TPT, but as it is one of the few available theories to describe mathematically the dynamical path leading to extreme values of an observable in a dynamical system, I think it is worth detailing here.

The stationary density ρ_S defined above is an equilibrium quantity characterizing the long-term occupation statistics in the phase space. But it is not enough to describe some events of interest which are characterized by a dynamics. These events are called **transition paths**: they are trajectories beginning inside a set A and ending in another set B of the phase space. Typically, transition paths would correspond to the trajectories leading to extremes of an observable of interest. Two quantities describe those trajectories: forward and backward committor functions.

The forward committor q^+ describes the progress of a stochastic trajectory traveling from set A to set B:

$$q^+(x) = \mathbb{P}[X_t \text{ reaches set } B \text{ before set } A | X_0 = x] \tag{4.6}$$

so that $q^+(x \in A) = 0$ and $q^+(x \in B) = 1$. The forward committor can be shown to obey the following boundary value problem:

$$\begin{cases} \mathcal{L}q^+(x) = 0 \text{ if } x \in (A \cup B)^c, \\ q^+(x) = 0 \text{ if } x \in A, \\ q^+(x) = 1 \text{ if } x \in B. \end{cases} \tag{4.7}$$

While q^+ describes the future of a transition, the backward committor q^- describes its past. It is defined as:

$$q^-(x) = \mathbb{P}[X_t \text{ comes from set } A \text{ rather than set } B | X_0 = x] \tag{4.8}$$

so that $q^-(x \in A) = 1$ and $q^-(x \in B) = 0$. Then q^- solves the following boundary value problem:

$$\begin{cases} \tilde{\mathcal{L}}q^-(x) = 0 \text{ if } x \in (A \cup B)^c, \\ q^-(x) = 1 \text{ if } x \in A, \\ q^-(x) = 0 \text{ if } x \in B \end{cases} \tag{4.9}$$

where $\tilde{\mathcal{L}}$ is the time-reversed generator, which evolves observables backward in time. It can be shown to be such that for all smooth-function f, $\tilde{\mathcal{L}}f = \frac{1}{\rho_S}\mathcal{L}^*(\rho_S f)$ with ρ_S the equilibrium density of X_t.

A trajectory X_t is said to undergo a transition at time t if it is on the way from set A to set B. The trajectory is then said to be a reactive trajectory[1]. The probability density of a reactive trajectory $\rho_R(x)$, that is the probability of observing the system X_t at the location x during a transition, is proportional (up to a normalization constant) to the product $\rho_S(x)q^-(x)q^+(x)$. This density is large in regions of phase space that are highly sampled by reactive trajectories. This is how TPT gives information about precursors: indicating regions of the phase space that are usually visited by the system over the course of a transition path. The strength of TPT is to transform a differential equation with a stochastic term in two more classical boundary value problems for deterministic PDEs that can in principle be solved by discretization. Once the boundary values problems for the forward and backward committor functions are solved, one has enough information to characterize transitions from set A to set B.

The direction and intensity of this transition is specified by the reactive current. The probability current $\boldsymbol{J}$ is a vector field that satisfies a continuity equation with the time-dependent density ρ:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho = -\nabla \cdot \boldsymbol{J}. \tag{4.10}$$

At equilibrium, i.e. when $\rho = \rho_S$, $\boldsymbol{J}$ is therefore non-divergent: $\nabla \cdot \boldsymbol{J} = 0$. The reactive current on the other hand $\boldsymbol{J}_{AB}$ between set A and set B is also defined as

[1] This terminology comes from computational chemistry.

a flux of probability, but restricted to reactive paths. It therefore has a divergence: A is a source and B a sink. $\boldsymbol{J}_{AB}$ is defined implicitly by saying that given any surface $\mathcal{S}$, the surface integral of $\boldsymbol{J}_{AB}$ over $\mathcal{S}$ gives the probability flux of reactive trajectories across $\mathcal{S}$. More precisely if $\mathcal{S}$ is any surface enclosing set A but not set B, with outward normal vector $\boldsymbol{n}$, then the flux $\oint_{\mathcal{S}} \boldsymbol{J}_{AB} \cdot \boldsymbol{n} dS = R_{AB}$ is the number of forward transitions per unit time, called the transition rate between A and B.

A streamline of $\boldsymbol{J}_{AB}$ is therefore similar to a typical — or most probable — transition path (see next section). It will almost surely not be realised by any real transition path, but it gives the common geometric features in phase space of all transition paths. It can be shown (Metzner et al. 2006) that $\boldsymbol{J}$ and $\boldsymbol{J}_{AB}$ can be expressed in terms of the more fundamental quantities defined above as:

$$\begin{cases} \boldsymbol{J} = \rho_S b - \nabla \cdot (\rho_S D), \\ \boldsymbol{J}_{AB} = q^+ q^- \boldsymbol{J} + \rho_S D (q^- \nabla q^+ - q^+ \nabla q^-). \end{cases} \tag{4.11}$$

The forward transition rate, captures the frequency of transitions between A and B rather than the overall time spent in each. Since a $B \to A$ transition must occur between every two $A \to B$ transitions, $R_{AB} = R_{BA} = R$, called more generally the transition rate. In the climate community working on extremes, the quantity of interest is often the return time associated with a particular event. Here, the return time is expressed as the inverse of the transition rate.

In general in non-linear systems, the forward and backward transitions to and from an extreme events are not the reverse of one another. One question is for example the speed of the transition. The equilibrium statistic $\mathbb{E}[q^+(X_t)] = \int q^+(x)\rho_S(x)dx$ partially answers this as when it is inferior (superior) to 1/2 it indicates more time spent on the way to A (B).

Another typical quantity of interest, coming from chemistry (Vanden-Eijnden 2014), is the rate constant k_{AB}, which is larger if $A \to B$ transitions happen faster than $B \to A$ transitions. One therefore normalizes the transition rate by the overall time spent having come from A, which is $\mathbb{E}[q^-(X_t)]$:

$$k_{AB} = \frac{R}{\mathbb{E}[q^-(X_t)]}. \tag{4.12}$$

$1/k_{AB}$ then estimates the total transition time between entering A (having last visited B) and next reentering B.

Although initially designed for studying the statistical behavior of transitions in the ergodic limit of a stationary system, TPT was recently extended to periodically forced dynamics and time-dependent finite-time systems by Helfmann et al. (2020).

4.3.3 Instanton theory

TPT sees the path leading to extreme events in a probabilistic sense: it gives information on the probability of the system being in certain regions of the phase space when transitioning from set A to set B. In high-dimensional systems, such as the climate system, it is probably hopeless to be able to compute such probability

distributions on the full phase space. Instanton theory takes a more deterministic point of view and derives explicit equations for the most likely path leading to rare regions of the phase space.

Here I present the instantons as seen from the Freidlin-Wentzell theory point of view (Freidlin and Wentzell 1984), i.e. using a large deviations theory perspective. Their properties and equations can also be derived using the Martin-Siggia-Rose-Janssen-de Dominicis (MSRJD) formalism, developed in the early 1970s to calculate statistical properties of classical systems using a path integral formulation (Grafke et al. 2015).

In the SDE 4.1 presented above, one can assume that the noise is small and rewrite the diffusion matrix as:

$$\sigma(X_t) = \sqrt{\epsilon}\tilde{\sigma}(X_t) \tag{4.13}$$

as in the limit $\epsilon \to 0$ is considered. One can also assume that the noise covariance $\tilde{D} = \tilde{\sigma}\tilde{\sigma}^T = 2D$ is invertible.

The Freidlin-Wentzell theory has its roots in the application of Large Deviation Theory (LDT) to dynamical systems under random perturbations. In general, a family of random processes X_t^ϵ defined on $t \in [0, T]$ is said to fulfill a large deviation principle (LDP), if:

$$\mathbb{P}[X_t^\epsilon \in \Omega] \asymp \exp\left(-\frac{1}{\epsilon} I_T(X_t^\epsilon)\right) \tag{4.14}$$

where $\asymp$ is to be understood as the ratio of the logarithms of both sides tending to unity as $\epsilon \to 0$. The left-hand side of this equation describes the probability of ending up in a given set of interest Ω which would typically correspond to extreme values of an observable F. The right-hand side displays the functional I_T which is called the rate function. In the following, the ϵ is dropped for simplicity but one should keep in mind that the results presented are valid in the limit $\epsilon \to 0$.

The Freidlin-Wentzell theory gives the expression of this rate function:

$$I_T(X) = \begin{cases} \int_0^T L(X, \dot{X}) dt & \text{if the integral exists,} \\ +\infty & \text{otherwise,} \end{cases} \tag{4.15}$$

where

$$L(X, \dot{X}) = \frac{1}{2} ||\dot{X} - b(X)||^2_{\tilde{D}} \tag{4.16}$$

is the equivalent of a Lagrangian for the system. The norm $||f||^2_\chi =< f, \tilde{D}^{-1} f >$ is induced by the noise covariance $\tilde{D}$, with $< \cdot, \cdot >$ the inner product in the space considered.

In Eq. (4.14), when $\epsilon \to 0$, the probability is dominated by the process $\hat{X}$ which minimizes the rate function. The minimizer $\hat{X}$ represents the trajectory with the maximum likelihood of realization with the given boundary conditions. Under the stated conditions, the probability distribution accumulates near $\hat{X}$: this most likely of the least likely trajectories is called the *instanton*.

It is possible to associate a Hamiltonian $H(X, P)$ to the Lagrangian 4.16 via a Legendre transform:

$$H(X, P) = \sup_{\dot{X}} \left[< P, \dot{X} > -L(X, \dot{X}) \right] \tag{4.17}$$

which here gives:

$$H(X, P) =< b(X), P > +\frac{1}{2} < P, \tilde{D}(X)P > \tag{4.18}$$

where $P = \frac{\partial L}{\partial \dot{X}}$ is the conjugate momentum to X. Now, the minimizer $\hat{X}$ can also be expressed as the solution of Hamilton's equations:

$$\begin{cases} \dot{X} = \partial_X H = b(X) + \tilde{D}P, \\ \dot{P} = -\partial_X H = -(\nabla_X b(X))^* P + < P, \nabla_X \tilde{D}(X)P > \end{cases} \tag{4.19}$$

where $(\nabla_X b(X))^*$ is the adjoint of $\nabla_X b(X)$. This formalism is still valid when the system considered is modeled with PDEs and in this case, $\tilde{D}P$ denotes the convolution product between $\tilde{D}$ and P in the phase space. These equations are called instanton equations.

The first equation is simply the original SDE where the stochastic term has been replaced by the deterministic term $\tilde{D}P$ which represents the 'optimal' forcing of the system leading to the extreme event of interest. It can indeed be shown (Grafke and Vanden-Eijnden 2019) that one also obtains the instanton equations when trying to reach the set Ω while minimizing the integral of the square added noise. In other words, the instanton equations can also be obtained as the result of an optimal control problem. The second equation is the equation for the forcing term P and depends mostly on the adjoint $(\nabla_X b(X))^*$ of the model. The very existence of those two Hamiltonian equations is an important result in so far as it is valid for any SDE of the form 4.1, even if the deterministic part does not derive from a Hamiltonian system as in most discretized form of geophysical equations.

The minimizer $\hat{X}$ also needs to fulfill given boundary conditions at $t = 0$ and $t = T$ on X or on a functional F. Solving numerically this boundary value problem is especially difficult as there is no initial condition for P. One would make repeated guess for the initial value of P, hoping that the forward integration in time would give the good values for the boundary conditions on X at $t = T$. To make this problem easier, one can implement the constraint directly in the optimization problem using a Lagrange multiplier $\lambda \in \mathbb{R}$, making the problem unconstrained:

$$\hat{X} = \arg\min_X \left\{ I_T(X) + \lambda(F[X(t = T)] - a) \right\}, \tag{4.20}$$

where $F[X(t = T)] = a$ is the final constraint on the observable F. This results in the same instanton equations but with different boundary conditions: $X(0) = X_0$ and $P(T) = \lambda \nabla_X F[X(t = T)]$. For the instanton leading to extremes of an observable of the form $F[X] = \int_0^T f(X)dt$, then the equation for P becomes

$$\dot{P} = -(\nabla_X b(X))^* P + < P, \nabla_X \tilde{D}(X)P > +\lambda \nabla_X f(X) \tag{4.21}$$

and the boundary conditions are $X(0) = X_0$ and $P(T) = 0$. The correspondence between λ and a is recovered by the condition $X(t = T) = a$ or $\int_0^T f(X)dt = a$. $P(a) = \mathbb{P}[F[X(t = T)] = a]$ or $\mathbb{P}[\int_0^T f(X)dt) = a]$ is estimated using equation 4.14 with the minimizer $\hat{X}$ to compute the value of the rate function.

This framework holds rigorously in the limit $\epsilon \to 0$, but it is possible to loosen conditions on the stochastic process and consider the case ϵ fixed and $\lambda \to \infty$ (Grafke and Vanden-Eijnden 2019). The reason is that for large λ only extreme events of the process are considered and a large deviation principle might hold for the observable even for finite noise.

The instanton formalism gives powerful results about the path in the phase space leading to extreme events in dynamical systems. If the hypotheses of the Freidlin-Wentzell theory apply, then the path leading to any extreme is unique, or more precisely dominated exponentially in probability by one unique path called the instanton[2]. However, one may note that this path depends on two parameters: the departure point X_0 and the time of integration T. The trajectory usually called the instanton is often not dependent on time T as one wishes to take the path minimizing the action for any T, including T being infinite. Although the instanton is the unique dominating path leading to an extreme event, it depends on its starting point so that the different paths may change for different starting points sampling the underlying attractor of the dynamical system. In an attractor with a strong mixing property, it may nevertheless be possible to loosen this condition.

4.3.4 Instantons in deterministic systems

Most climate models are not explicitly stochastic and one may wonder how the previous framework may apply in this case. In this section I propose an adaptation to the fully deterministic case based on the derivation by Lucarini et al. (2023)[3]. The idea here is that the explicit stochasticity in equation 4.1 is replaced by an 'implicit' stochasticity arising from the position of the starting point on the attractor of the system (on which an SRB probability distribution can be assumed to exist, cf. chapter 1 section a)).

Let us consider a chaotic dynamical system evolving continuously in time. Let $X_t \in \mathbb{R}^d$ be the state vector at time t. It is assumed that X evolves according to the following ordinary differential equation:

$$\frac{dX}{dt} = b(X) \tag{4.22}$$

where, as previously, $b : X \mapsto b(X) \in \mathbb{R}^d$ defines the dynamics of the system. It is assumed that the transients have died out and all trajectories considered belong to the attractor of the system and that there is a unique climatological probability distribution μ on this attractor. I am again interested in the statistics of observables

[2] Note that the uniqueness is guaranteed by the uniqueness of the minimizer of the rate function. If there are several minimizers, then several paths are possible. It may for example be the case if there are symmetries in the system and/or the observable extremized. [3] I thank Valerio Lucarini for our discussions on this subject.

f computed on the attractor. For $q \in \mathbb{R}$, one can define the set Ω_q as the set of points X on the attractor such that $f(X) \in [q, q + dq]$. It is explicitly assumed that f admits a large deviations principle, i.e. the set Ω_q of interest is exponentially rare with respect to the probability distribution μ with increasing q. Then, one can write:

$$\mathbb{P}[f(X_t) = q] = \mu(\Omega_q) \asymp \exp\left(-\min_{X \in \Omega_q} I(X))\right) \tag{4.23}$$

where $\asymp$ means that the ratio of the logarithms of both sides tends to unity when $q \to +\infty$. Here q plays the role of the large deviations parameter $1/\epsilon$ of the previous section, even though it does not appear explicitly as such in Eq. (4.23). The functional I is the rate function and is given by the Legendre transform (Dematteis et al. 2019):

$$I(X) = \max_P(\langle P, X\rangle - S(P)) \tag{4.24}$$

of $S(P) = \log \mathbb{E}_\mu[e^{\langle P,X\rangle}]$, the cumulant generating function of X under the probability distribution μ with $\langle \cdot\, , \cdot \rangle$ the inner product in $\mathbb{R}^d$. When $q \to +\infty$ in Eq. (4.23), the probability is exponentially dominated by the point $\widehat{X}$ which minimizes the rate function: $\widehat{X} = \underset{X \in \Omega_q}{\operatorname{argmin}}\ I(X)$.

Whether this mathematical framework fully applies to climate models can naturally be questioned. Nonetheless, it provides good reasons to expect that, when investigating more and more intense extremes, the dynamical path leading to these extremes tend to be unique. I explore this idea with a long control run of a climate model in chapter 5.

4.3.5 Example with the Orstein-Uhlenbeck process

In this section I illustrate the concepts defined above with a simple model. I consider the following one dimensional SDE:

$$dX_t = -\alpha X_t dt + \sigma dW_t, \tag{4.25}$$

with α and σ two positive real numbers. This SDE is usually called an Orstein-Uhlenbeck (OU) process, a Brownian motion or even a red noise process. It was for example used to model the variability of the ocean in the seminal paper of Hasselmann (1976). The deterministic drift $b(X) = -\alpha X$ models an exponential decay towards 0 while the diffusion term σdW_t provides random variability and drives the system out of equilibrium.

For the OU process (4.25), the Fokker-Planck equation is given by:

$$\frac{\partial \rho}{\partial t} = \alpha \frac{\partial (x\rho)}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 \rho}{\partial x^2} \tag{4.26}$$

which can be solved explicitly in general. With an initial condition constituted for example by a Dirac distribution centered at x_0, the solution is given by:

$$\rho(x,t) = \frac{1}{\sqrt{\pi \frac{\sigma^2}{\alpha} e^{-2\alpha t}(e^{2\alpha t}-1)}} \exp\left(-\frac{(x - x_0 e^{-\alpha t})^2}{\frac{\sigma^2}{\alpha} e^{-2\alpha t}(e^{2\alpha t}-1)}\right). \tag{4.27}$$

This PDF converges exponentially fast towards the stationary PDF given by:

$$\rho_S(x) = \sqrt{\frac{\alpha}{\sigma^2 \pi}} e^{-\frac{\alpha}{\sigma^2}x^2}. \tag{4.28}$$

Figure 4.1a displays the evolution of one example trajectory of the OU process with $\alpha = 1$ and $\sigma = 1$. The integration of the SDE Eq. (4.25) was made using the order (3.0,1.5) strong stochastic Runge-Kutta methods for SDEs proposed by Rößler (2010) with a time step $dt = 0.001$ and a total integration time $T = 1$. Figure 4.1b shows the corresponding stationary PDF.

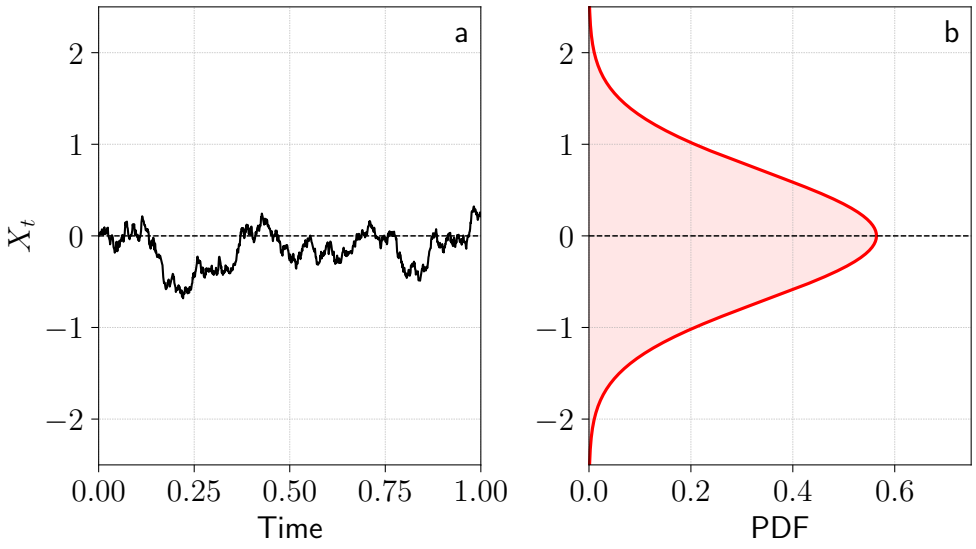


Figure 4.1: **One example trajectory and PDF of the Orstein-Uhlenbeck process.** (a) An example trajectory of the Orstein-Uhlenbeck process ($\alpha = 1$ and $\sigma = 1$) and (b) associated stationary PDF.

Transition Path Theory. One can define the set A as $A =]-\infty, 0]$ and the set B as $B = [a, +\infty[$ for $a > 0$. Schematically, the process starts at 0 and I am looking for the typical transition path for reaching a certain threshold a. Then using the definition above, the boundary value problem for the forward committor function q^+ gives:

$$\begin{cases} \mathcal{L}q^+(x) = -\alpha x \frac{\partial q^+(x)}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 q^+}{\partial x^2} = 0 \text{ if } 0 < x < a, \\ q^+(x) = 0 \text{ if } x \leq 0, \\ q^+(x) = 1 \text{ if } x \geq a. \end{cases} \tag{4.29}$$

Similarly, the boundary value problem for the backward committor function q^- is:

$$\begin{cases} \mathcal{L}q^-(x) = \frac{1}{\rho_S(x)}\left[\alpha\frac{\partial}{\partial x}(x\rho_S(x)q^-(x)) + \frac{1}{2}\sigma^2\frac{\partial}{\partial x^2}(\rho_S(x)q^-(x))\right] = 0 \text{ if } 0 < x < a, \\ q^-(x) = 1 \text{ if } x \leq 0, \\ q^-(x) = 0 \text{ if } x \geq a. \end{cases} \tag{4.30}$$

Solving those two problems one finds $q^+(x)$ and $q^-(x)$ for $x \in [0, 1]$:

$$\begin{cases} q^+(x) = \frac{\int_0^x e^{\frac{\alpha}{\sigma^2}y^2}dy}{\int_0^a e^{\frac{\alpha}{\sigma^2}y^2}dy}, \\ q^-(x) = 1 - \frac{\int_0^x e^{\frac{\alpha}{\sigma^2}y^2}dy}{\int_0^a e^{\frac{\alpha}{\sigma^2}y^2}dy} = 1 - q^+(x). \end{cases} \tag{4.31}$$

The quantities of interest of TPT for $x \in [0, 1]$ are:

$$\begin{cases} \rho_R(x) = \frac{q^-(x)q^+(x)\rho_S(x)}{\int q^-(y)q^+(y)\rho_S(y)dy} \\ J(x) = 0 \\ J_{AB}(x) = \frac{1}{2}\sqrt{\frac{\alpha\sigma^2}{\pi}}\frac{1}{\int_0^a e^{\frac{\alpha}{\sigma^2}y^2}dy} \end{cases} \tag{4.32}$$

In the particular case of the OU process studied here, I have shown that $q^+ = 1 - q^-$. This is a direct consequence of the fact that it is a gradient system: $-\alpha X_t = -\nabla V(X_t)$ with $V(X_t) = \alpha X_t^2$. For gradient systems, the path leading from set A to set B is the minimum energy path and is the same as the path leading from B to A, therefore $q^+ = 1 - q^-$. This is however not true in general and especially for geophysical dynamical systems.

The fact that $J(x) = 0$ is natural when considering a 1D process in so far as a non-divergent J in 1D implies J being constant which leads to $J = 0$ for a flux with no source and no sink. One can also remark that J_{AB} does not depend on x — and therefore also has no divergence in 1D — but is not null as it has a source (set A) and a sink (set B).

Figure 4.2 displays the forward q^+ and backward q^- committor functions, and the PDF of a reactive trajectory ρ_R for the OU process with $\alpha = 1$ and $\sigma = 1$ and for the set A and B defined above with $a = 2.5$. One can remark on the figures that the threshold $q^+(x) = 0.5$ is not reached until $x \sim 2.4$, illustrating the difficulty for the system to reach the 2.5 threshold as shown by the low probability associated with this value in Figure 4.1b.

With these numerical values, one finds $J_{AB} = R_{AB} = R = 2.44 \times 10^{-3}$ and $k_{AB} = \frac{R}{\mathbb{E}[q^-(X_t)]} = 2.45 \times 10^{-3}$. The return time of the event is then $\frac{1}{R} = 409$ and the total transition time between entering A and reentering B is $\frac{1}{k_{AB}} = 408$.

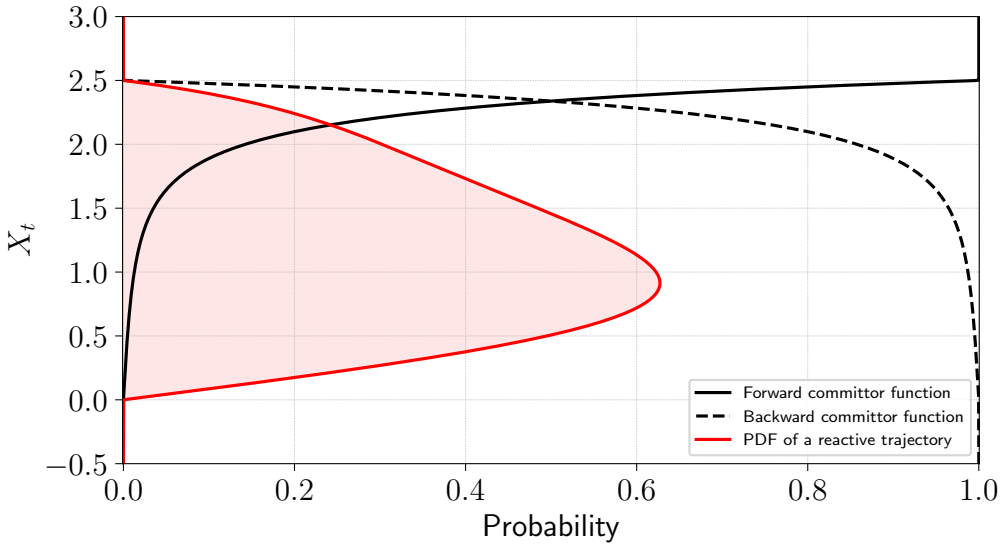


Figure 4.2: **Key functions of interest of Transition Path Theory for the Orstein-Uhlenbeck process.** Forward (plain black) and backward (dashed black) committor functions and PDF of a reactive trajectory for the Orstein-Uhlenbeck process ($\alpha = 1$ and $\sigma = 1$). The sets A and B for computing these functions are defined as $A =]-\infty, 0]$ and $B = [2.5, +\infty[$.

Instanton analysis. For the OU process, the instanton equations are:

$$\begin{cases} \dot{X} = -\alpha X + \sigma^2 P, \\ \dot{P} = \alpha P. \end{cases} \tag{4.33}$$

I consider the instanton solution for the boundary conditions $X(0) = 0$ and $X(T) = a$ for a given a. Then, the equation for P gives: $P(t) = P(0)e^{\alpha t}$, which in the equation for X with the two boundary conditions gives:

$$X(t) = a\frac{\sinh(\alpha t)}{\sinh(\alpha T)}. \tag{4.34}$$

This formula gives the explicit expression of the most likely trajectory leading to $X(t = T) = a$ when starting from $X(0) = 0$. One way of checking whether it is correct is the so-called instanton filtering procedure (Grafke et al. 2015). In a long simulation of a SDE, it consists in averaging trajectories starting at $X(0) = 0$ for which $X(t) \sim a$. Figure 4.3a presents such a filtering procedure for the OU process with $\alpha = 1$ and $\sigma = 1$. I simulated 1000 trajectories which have $X(T) \in [a - 0.1, a + 0.1]$ with $a = 2.5$. The black curves is the mean of these trajectories and the red curve is the theoretical result for the instanton. The two curves overlap almost perfectly.

One should note that the instanton trajectory depends critically on the observable chosen. If for example one considers the observable $F[X_t] = \int_0^T X dt$ then the

instanton equations are:

$$\begin{cases} \dot{X} = -\alpha X + \sigma^2 P, \\ \dot{P} = \alpha P + \lambda. \end{cases} \tag{4.35}$$

In this case, the solution of the two equations with the condition $\int_0^T X dt = a$ is

$$X(t) = \frac{\alpha a}{\frac{3}{2} + \frac{1}{2}e^{-2\alpha T} - 2e^{-\alpha T} - \alpha T} \left[\sinh(\alpha t)e^{-\alpha T} + e^{-\alpha t} - 1\right]. \tag{4.36}$$

Figure 4.3b presents the same filtering procedure as above but this time choosing 1000 trajectories for which $\int_0^T X(T) \in [a - 0.05, a + 0.05]$ with $a = 1.5$. Again, the theoretical prediction is almost identical to the empirical mean found with the filtering procedure.

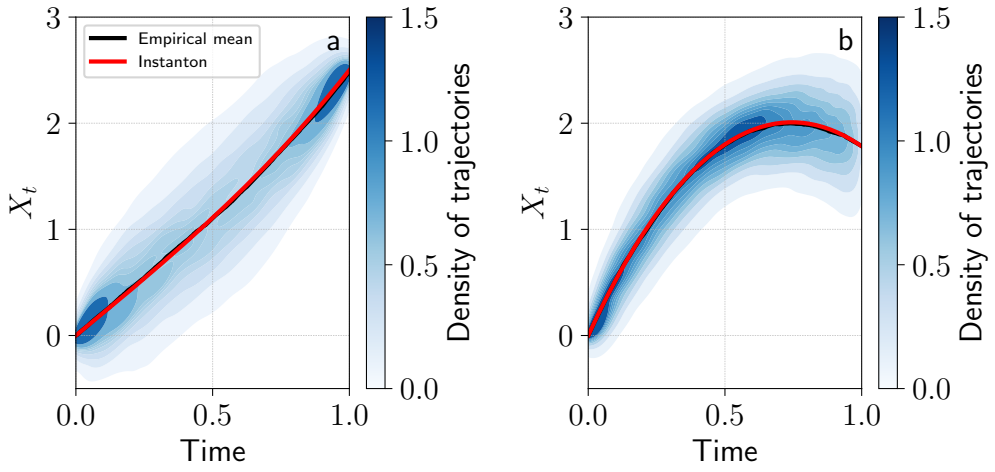


Figure 4.3: **Empirical and theoretical trajectories leading to extremes for the Orstein-Uhlenbeck process.** (a) Empirical PDF (colours) of 1000 trajectories of the Orstein-Uhlenbeck process ($\alpha = 1$ and $\sigma = 1$) for which $X(t = T) \in [a - 0.1, a + 0.1]$ with $a = 2.5$. The empirical mean of these trajectories (black) and the associated instanton for the observable $F[X_t] = X(t = T) = a$ (red) are plotted for comparison. (b) Same but for trajectories with $\int_0^T X(t)dt \in [a - 0.05, a + 0.05]$ with $a = 1.5$ (colours), their empirical mean (black) and the instanton for the observable $F[X_t] = \int_0^T X(t)dt = a$.

4.4 Rare events algorithms

I propose here to classify rare events algorithms into three main categories: **direct computation methods, biased dynamic methods and splitting methods**. Those are gross classifications as there is no official denomination and in practice

some algorithms overlap two categories. Direct computation methods aim at solving directly equations associated to key quantities of interest for sampling the dynamics leading to extremes, either deterministically (instantons) or probabilistically (TPT). Biased dynamic methods refer to methods that directly modify the dynamics of the system to make a rare event more likely to appear. Splitting methods run several members of a model in parallel and periodically kill members that move away from the target region of the phase space associated with the rare event studied, and clone members that come closer to this region.

Here I present the main algorithms used in the geophysical dynamical systems literature (cf. table 4.1). A large number of rare events algorithms exist in the literature (e.g. string method and nudged elastic band method for transition path sampling in molecular systems, Vanden-Eijnden et al. (2010)) but they are often designed for systems with a low number of degrees of freedom, discrete in time and/or deriving from a gradient, which is generally not the case for geophysical systems and therefore of lesser interest to this thesis.

4.4.1 Direct computation methods

a) TPT-based methods

As shown above, in TPT the quantities of interest (ρ_S, q^+, q^- and $\boldsymbol{J}_{AB}$) are solutions to PDEs involving the generator $\mathcal{L}$. Those equations can be solved explicitly by spatial discretization. This is the approach taken by Finkel et al. (2020) using a finite volume scheme to directly discretize the adjoint $\mathcal{L}^*$ as a matrix on a regular grid with a model with 3 dimensions.

However when the phase space has more dimensions this procedure leads to difficulties related to the burden of dimensionality: the volume expands quickly with the dimension and the discretization in the phase space is not feasible anymore. In Finkel et al. (2021a), the authors expand TPT and generalize committors functions to forecast not only where the trajectory ends up, but what happens along the way using what they call forecast/aftcast functions. These functions also obey PDEs which are still local in state space. Based on previous works especially in molecular dynamics (Thiede et al. 2019; Strahan et al. 2021; Finkel et al. 2021b), the authors note that those PDEs can still not be solved easily by discretization, but that the essential property of spatial locality allows for data-driven approximation with a large ensemble of short trajectories. They use a Galerkin approximation with a suitably defined set of basis functions and therefore reduce the dimension of the problem. The PDEs to be solved then becomes a system of linear equations that can be solved more easily.

As noted by Finkel et al. (2021a), the main important limitation of the method is the data generation step. One indeed needs to use a long stationary trajectory to first sample the attractor. This is intended both to seed initial data points for short trajectories and provide a ground truth for validating the accuracy of the method. This long data set may be difficult to obtain from a real model. The authors note that using splitting methods to sample more efficiently the attractor could help to

solve the first point. Even though this method gives precise and direct estimates of the quantities of interest, which allow for a thorough investigation of the properties of reactive trajectories, there are limitations associated with the sampling of the phase space and the Galerkin approximation to make it feasible for a high-dimensional system.

b) Instanton methods

I follow the presentation of Grafke and Vanden-Eijnden (2019) for the two main methods used to find instantons: numerical action minimization and direct integration of instanton equations.

The main algorithm used in fluid mechanics to find numerically instantons is the geometric Minimum Action Method (gMAM) (Heymann and Vanden-Eijnden 2008). For an arbitrary action functional, general methods of minimization have already been proposed in the past in the context of gradient systems, such as in chemistry with the so-called string method (E et al. 2004). The gMAM method can be viewed as a generalization of the string method for non-gradient fields. Its starting point is a modified action functional similar to the Freidlin-Wentzell action functional 4.15 for which the minimizers are identical to the infinite time minimizers of the original action. An important class of problems that can be solved efficiently by the gMAM are noise-driven transitions between stable fixed points in the context of metastability, as there is perfect knowledge of starting and ending points. This method can however not be applied for extremes if one has no idea of their position in the phase space.

The alternative to minimizing the action functional is to solve directly the corresponding instanton equations 4.19. This is especially relevant when looking for instantons leading to extremes of an observable. Problems of this type are indeed difficult to frame in the context of minimization of the action functional, as the final condition is not fixed but only subject to a constraint. One could rely on shooting methods (Keller 2018) but it is usually hopeless in high-dimensional systems. One can rather consider the dual problem. The only input being a restriction on a single degree of freedom of the final condition, the instanton formalism provides the most probable final state which fulfills the constrain, the most probable evolution in time from a given initial condition into this state, and the force that was necessary to achieve this evolution (via the value of the action).

The seminal paper of Chernykh and Stepanov (2001) proposed an algorithm to solve iteratively the instanton equations with a final constraint. The implementation proceeds as follows. One has to first begin with a guess either for $P(x,t)$ or for $X(x,t)$. Suppose an initial guess is made for $X(x,t) = X^0(x,t)$ (for example $X^0(x,t) = 0$). One then chooses a value for λ and iterates over k until $X^k(x,t)$ has converged to a fixed value:

1. Solve equation for P backward in time using X^k as input and $P(T) = \lambda \nabla_X F[X^k(t = T)]$ as the initial condition. The solution of this equation is P^k.

2. Solve equation for X forward in time using P^k as input and $X(0) = X_0$ as the initial condition. The solution of this equation is X^{k+1}.

Having found the fixed point of the iterative algorithm, one can recover the optimal forcing by evaluating $\tilde{D}P$. Using this algorithm avoids having to rely on shooting methods but has a drawback: instead of choosing directly the value a of the observable, its dual λ is prescribed and the corresponding value of a is obtained a posteriori. In practice, this is usually not a problem, typically one is interested in the complete distribution $P(a)$ and therefore producing instantons for a whole range of λ similarly covers a whole range of a. This procedure is quite straightforward to apply but there are two main computational challenges: (i) numerical instability and (ii) convergence issues related to the fact that one needs sometimes to take $T \to \infty$ in order to compute the minimizer (Grafke et al. 2014).

Finally one should note that they are links between the instanton methods and other rare events algorithms once the instanton is known (which is however not the case for most complex systems). When looking at biased dynamic methods, the instanton can be considered as the optimal tilt in the small-noise limit and used as such to 'push' the system in the direction of interest. For splitting methods, the instanton can be used in the reweighting process (Ebener et al. 2019) to orient the parallel trajectories into following its path in the phase space.

With the exception of Laurie and Bouchet (2015) and Grafke and Vanden-Eijnden (2019) who computed instanton for low-dimensional geophysical dynamical systems, to the best of my knowledge no studies tried to compute explicitly the instanton path in a geophysical system, not to mention a state-of-the art climate or meteorological model. The main issue is of course the need of having an ajoint model to compute the 'optimal' noise, i.e. integrate the equation for the conjugate momentum in the instanton equations. Even if one has access to such an adjoint, the Chernykh-Stepanov algorithm may not converge easily on a high-dimensional, chaotic system. Nonetheless, there are interesting works which applied such a computation to high-dimensional systems (Grafke et al. 2015; Grafke and Vanden-Eijnden 2019; Schorlepp et al. 2022).

The explicit computation of instantons is a strong numerical challenge. Nevertheless, the paradigm set by instanton theory for explaining the uniqueness of the route to extreme events in stochastic dynamical systems can definitely not be ignored. The practice of making composite maps of similar events in climate and meteorological science finds a strong justification in this paradigm and corresponds to the instanton filtering procedure detailed above. It is therefore natural to demand that rare events algorithms in general sample the instanton leading to extreme events, i.e. the most probable path (cf. chapter 5).

4.4.2 Biased dynamic methods

Biased dynamic methods aim at modifying directly the dynamics of the system in a controlled way in order to make rare events happen more often. Those methods have been extensively studied and employed on discrete stochastic systems (Bucklew

2004). They are more difficult to employ when it comes to time-continuous systems for which the entire phase space is not known — even after a long control run — and therefore not mapped with a transition probability matrix.

These methods are close to the idea of control methods and some authors (Hartmann et al. 2019) proposed an adaptive importance sampling method where the stochastic differential equation is modified to include a control drift term. An optimal drift term can then be found by solving a stochastic control problem. It is however still doubtful that this procedure is feasible for high-dimensional systems.

a) Finite states phase space

In the context of a discrete stochastic model such as a Stochastic Weather Generator (SWG), Yiou and Jézéquel (2020) proposed an importance sampling procedure to alter the probability matrix of the SWG in the direction of warmer temperatures. To do so, one considers the K best analogues of the atmospheric circulation X_t at time t. The values of the observable of interest F — temperature in Yiou and Jézéquel (2020) — of the K analogues are sorted in decreasing order so that the ranks are written R_k. The weights to modify the probability matrix are then chosen as:

$$w_t^k = Ae^{-\alpha R_k} \tag{4.37}$$

and A is a normalization constant so that $\sum_k w_t^k = 1$. In this context, the weight values do not depend on time and not on the unit of the variable either, so that this procedure does not need major adaptation to simulate any observable.

The next state of the trajectory, that is one of the K best analogues, is chosen at random by sampling the dates of the analogues with the weights w_t^k so that if $\tilde{F}_t$ is the simulated observable among the analogue observables F_t^k then $\mathbb{P}[\tilde{F}_t = F_t^k] = w_t^k$. The expected value of $\tilde{F}$ is then:

$$\mathbb{E}[\tilde{F}_t] = A\sum_{k=0}^{K} e^{-\alpha R_k} F_t^k = A\sum_{k=0}^{K} e^{-\alpha R_k} \text{sort}(F_t)^k \tag{4.38}$$

where $\text{sort}(F_t)$ are the sorted values of F_t in descending order. This allows to select the circulation analogues that favor the highest value of observable F_t. The α parameter controls the strength of the selection procedure. The analogue importance sampling tilts the trajectories towards the largest values of F_t.

The probability distribution of the simulations is controlled by the value of α. According to Yiou and Jézéquel (2020) a formulation of the expected probability associated with each simulation can be obtained heuristically. If one note Q the smallest number so that:

$$A\sum_{k=1}^{Q} e^{-\alpha k} > 1 - \epsilon \tag{4.39}$$

where $\epsilon > 0$ is a small number, then the probability of dynamic trajectories with parameter value α is close to $(Q/K)^M$, where M is the average number of independent days during the simulated season and K is the number of analogues.

This formulation may suffer from numerical problems for large values of α and small values of K. An alternate empirical estimation of the probability distribution of the average of the trajectory is to consider that the seasonal average of the observable closely follows a Gaussian distribution. Then an empirical estimate of the probability is obtained by comparing the quantiles of the observed distribution of the average of the observable and the mean of simulated averages. With a chosen value of $K = 20$, Yiou and Jézéquel (2020) show that the heuristic and empirical approximations give similar estimates of the probability or return periods for values of $\alpha \leq 0.5$.

Once one has access to a discrete stochastic model of the system of interest — such as a SWG for the atmosphere — this method is efficient to simulate a large number of plausible physical trajectories. When biasing in favour of high values of an observable of interest F, this method samples unseen but possible trajectories with extreme values of the time-integrated observable. Even though the number of possible trajectories is high with such a SWG, when strongly biasing in favor of an observable, the number of trajectories leading to extreme values may become small as the algorithm will loop over the small number of states which have the highest values of the observable. In other words, this method is limited to states which have already been seen and cannot generate other states with intense values of the observable.

b) Infinite states phase space

Plotkin et al. (2019) proposed a method to maximize the intensity of two simulated cyclones by pushing the system into unseen states. This method is based on minimizing an action and is close to the one proposed in E et al. (2004). As Plotkin et al. (2019) explicitly modify the dynamics of their model, this method fits better with the biased dynamic category than with the direct computation methods category.

The main idea is to add small perturbations to true model trajectories in order to achieve a transition into a rare state while minimizing the magnitude of the perturbations added to keep trajectories physically relevant. Plotkin et al. (2019) note that this problem is close to the one of data assimilation, with a different score function. This framing of the problem is also close to the instanton formalism, which explicitly looks for the 'optimal' noise to push the system in the direction of interest.

The objective is to minimize the perturbations added to a true model trajectory in order to obtain a trajectory that results in a lower value of an observable F at the final state X_N with N the number of optimization time steps. Then the method seeks to minimize the following action functional:

$$I = \sum_{i=0}^{N-1} \frac{1}{2R_i^2} \tilde{\eta}_i^T D^{-1} \tilde{\eta}_i + \frac{1}{R_f^2} F(X_N) \tag{4.40}$$

with $X_i = \tilde{b}(X_{i-1}) + \tilde{\eta}_i$, $\tilde{b} = bdt$ being the recurrence function and $\tilde{\eta}_i = dW_i$ the perturbation added at time i. The initial state is restricted to stay close to state $\bar{X}_0$: $X_0 = \bar{X}_0 + \tilde{\eta}_0$. D is the background covariance matrix, i.e. the time-averaged outer product of the differences between 12- and 24-hr forecasts with the model over the course of several months in their case. R_i and R_f are constants that balance running

cost against final cost: they determine the relative importance of the running cost at each time step and of the final cost.

This method needs a model with adjoint capabilities as shown by the presence of the derivative of the recurrence function when maximizing I. In Plotkin et al. (2019) they adapt the WRF Variational Data Assimilation model, initially used for assimilation purposes. They optimize over perturbations to zonal and meridional wind, surface pressure, temperature, water vapor mixing ratio and geopotential. Optimizing the perturbations over a limited set of variables could be a way ahead for finding more easily instanton-like trajectories in climate and meteorological models. The method proposed by Plotkin et al. (2019) is indeed natural in this context.

This scheme allows the authors to study the phenomenology of the optimal perturbations, in particular the one that lead to rapid intensification of tropical cyclones. Remarkably, they show that the optimal perturbations need not be symmetric contrary to what has been thought before. However, this scheme does not allow to access the probabilities of the new intensification levels attained thanks to the optimal perturbations.

4.4.3 Splitting methods

As shown in table 4.1, splitting methods are by far the most employed methods for simulating rare events in geophysical dynamical systems. In the following I detail the different algorithms employed in the literature. As these methods are both promising and challenging, I shall present the main issues associated with their use in the next section, illustrating them with simple models.

The general idea of splitting methods is to start N parallel members of a model with different initial conditions. The members evolve according to the model and at each resampling time some of them are killed and some of them are cloned, depending on weights computed according to their past evolution. The weights are to be chosen carefully so that high values of the weights are equivalent to the trajectories moving close to the rare event of interest. This general method is also known in the literature as genealogical particle analysis, genetic algorithm or "go with the winners" algorithm (Grassberger 2002; Bouchet et al. 2019).

The algorithms are generally guided by a one-dimensional coordinate $\theta : \mathbb{R}^d \mapsto \mathbb{R}$ from the phase space to $\mathbb{R}$ that is high in some regions of the phase space and low in other regions. Where θ is high, the algorithms should exhibit a greater propensity toward splitting. Where θ is low, the algorithms should exhibit a greater propensity toward killing. θ is therefore reminiscent of the probability of a reactive trajectory ρ_R in the framework of TPT. This PDF is however unknown in most cases and is usually determined heuristically — for example by choosing the very observable one seeks to extremize. θ is known as a reaction coordinate, an order parameter or a score function. I call it here a **score function**. The particular choice of a score function is crucial for the efficiency of the algorithm for computing rare events statistics.

In the following, I use the standard vocabulary of climate science. A **simulation** corresponds to the integration of a climate model over a given amount of time. An

ensemble simulation is a simulation run with several **members** in parallel, i.e. several trajectories in the phase space that may or may not start from the same initial conditions. Note that in the rare events literature, what is called 'members' in the climate literature is usually called 'trajectories' or 'particles'. I use either members or trajectories in the following, but it should be emphasized that the word 'trajectories' refers to phase space trajectories $t \mapsto X_t$ and not to Lagrangian trajectories of tracers or particles of air for example.

a) Ensemble boosting

Ensemble boosting is a method proposed in climate sciences by Gessner et al. (2021). The idea of the method is close to the other rare events algorithm presented below, although it may not always be used iteratively in practice. The ensemble boosting method seeks to answer the question: how much an event that has occurred could have been more intense? It is therefore a storyline approach and — at least for now — does not allow to recover associated probabilities of extremes (see below for a discussion on this point).

The algorithm is depicted schematically in Figure 4.4. One starts from a long and/or multi-member simulation of a climate model. Then one selects N extreme events that have happened during the long simulation. The idea of the method is to resample these extremes by starting a multi-member simulation some days/weeks before the moment where the extremes have occurred and add a small perturbation to the original simulation. Because of the chaoticity of the system, the different members eventually separates and by choosing a proper lead time before the event, one can sample events that are more intense than the original event (Fischer et al. 2023). This allows to have access to physically consistent storylines for very intense events.

As it has been used up to now, the algorithm does not allow to recover the climatological expectations and therefore the probabilities of the extreme events sampled — hence the use of the term 'storyline' to describe the trajectories sampled by the algorithm. It should nevertheless be noted that under mild assumptions, it may be possible to recover the initial probabilities (see A.2.1 in annex for a proposition and also the recent work of Finkel and O'Gorman (2024)).

b) Diffusion Monte Carlo

I present a version of the Diffusion Monte Carlo (DMC) algorithm close to the one formalized in Webber et al. (2019). DMC has been the subject of mathematical analyses (Del Moral 2004), which describe the convergence and asymptotic behavior of the algorithm as the ensemble of size N approaches infinity. DMC estimates are unbiased and converge as the number of members N tends to infinity under mild integrability conditions. These theoretical results are general, holding true for systems with arbitrarily high dimension d. Any quantity that can be estimated by direct sampling can also be estimated by DMC. Estimates can include functions that depend on the entire path from time 0 until a later time t_i.

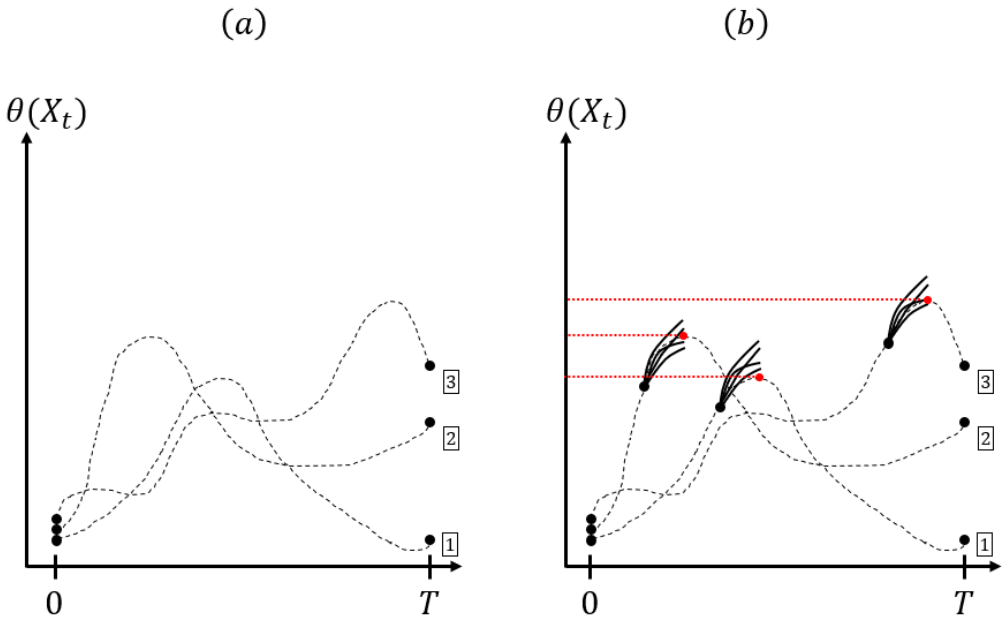


Figure 4.4: **Schematic representation of the ensemble boosting rare events algorithm.** (a) Initial unperturbed integration of a 3-members ensemble simulation. (b) The three extremes of the θ metric are resampled by starting new ensembles of the climate models a few days/weeks before the occurrence of the extreme after adding a small perturbation to the original simulation.

Let's note $(X_t^n)_{1\leq n\leq N}$ the N members of the simulation. I consider a finite series of resampling times $0 = t_0 < t_1 < \cdots$ that may or may not be separated homogeneously (see below for a discussion). In the following I note $X_i^n := X_{t_i}^n$. I consider a family $(V_i)_i$ of splitting functions which are functions of the score function θ.

The algorithm is depicted schematically in Figure 4.5. The implementation of the algorithm proceeds as follows. First, initialize N independent members $(X_0^n)_{1\leq n\leq N}$. Then iterate the following procedure over i:

1. **Reweighting**: for each member n

   - if $i = 0$, define initial weights as:

   $$w_0^n = e^{V_0(X_0^n)}, \tag{4.41}$$

   - if $i > 0$, define weights as:

   $$w_i^n = \bar{w}_{i-1} e^{V_i(X_i^n) - V_{i-1}(\hat{X}_{i-1}^n)}. \tag{4.42}$$

   Compute the average weight: $\bar{w}_i = \frac{1}{N}\sum_{n=1}^N w_i^n$.

2. **Resampling**: create an ensemble of updated members $(\hat{X}_i^n)_{1\leq n\leq N}$ consisting of N_i^n copies of each member X_i^n. The numbers N_i^n must be randomly chosen to ensure: $\sum_n N_i^n = N$ and $\mathbb{E}[N_i^n] = \frac{w_i^n}{\bar{w}_i}$.

3. **Simulation**: run the model from time t_i to time t_{i+1}.

If the splitting functions $(V_i)_i$ are chosen correctly, the simulation with the DMC algorithm oversamples regions where the score function θ is large and undersamples regions where θ is small. The distribution of members converges to the distribution X_{t_i} weighted by a likelihood ratio $L_i(X) = \frac{e^{V_i(X)}}{\mathbb{E}[e^{V_i(X_{t_i})}]}$ which favors high values of θ. For any observable ψ, to estimate its climatological expectation $\mathbb{E}_\mu[\psi(X_{t_i})]$, DMC uses the following estimator:

$$\mathbb{E}_\mu[\psi(X_{t_i})] \simeq \frac{\bar{w}_i}{N} \sum_{n=1}^{N} \psi(\hat{X}_i^n) e^{-V_i(\hat{X}_i^n)}. \tag{4.43}$$

Using this formula with the observable $\psi = \mathbb{1}(f(X_t) \geq a)$ for $a \in \mathbb{R}$ and an observable f, one then has access to the probability of f taking extreme values ($\mathbb{E}_\mu[\mathbb{1}(f(X_t) \geq a)] = \mathbb{P}_\mu[f(X_t) \geq a]$). One should note that the DMC algorithm applies a selection procedure before the first simulation step. In practice if one does not want to apply this first selection step, it can be avoided by choosing $V_0 = 0$.

There are several practical ways of performing the resampling step. I present two of them used in the literature:

- For the first one, proposed by Ragone et al. (2018), the number of copies produced by each member is:

$$c_i^n = \lfloor \tilde{w}_i^n + u_i^n \rfloor \tag{4.44}$$

  where $\tilde{w}_i^n = \frac{w_i^n}{\bar{w}_i}$ to ensure that $\sum_n \tilde{w}_i^n = N$, $\lfloor . \rfloor$ is the integer part and the u_i^n are N independent random numbers sampled from a uniform distribution between 0 and 1. When $c_i^n = 0$ the member is killed. After the killing process the total number of members is $N_i = \sum_{n=1}^N c_i^n$. Then one computes the difference: $\Delta N_i = N_i - N$:

  - if $\Delta N_i > 0$, then ΔN_i members are randomly selected (without repetition) among the remaining members and killed,
  - if $\Delta N_i < 0$, then ΔN_i members are randomly selected (with repetition) among the remaining members and cloned.

- For the second one, proposed by Webber (2019) and named sorted stratified resampling scheme, one proceeds as follows:

  1. Reindex the weights and the members $(w_i^n, X_i^n)_{1\leq n\leq N}$ so that

$$\theta(X_i^1) \leq \cdots \leq \theta(X_i^N) \tag{4.45}$$

2. Construct the empirical quantile function q_i for the ensemble $(w_i^n, X_i^n)_{1\leq n\leq N}$ as: $q_i(x) = \theta(X_i^n)$ for

$$\frac{\sum_{j=1}^{n-1} w_i^j}{\sum_{j=1}^{N} w_i^j} \leq x < \frac{\sum_{j=1}^{n} w_i^j}{\sum_{j=1}^{N} w_i^j}. \tag{4.46}$$

3. Select updated members $\hat{X}_i^n = q_i(\frac{n-1+u_i^n}{N})$ where u_i^n are N independent random variables following a uniform law over $[0, 1]$.

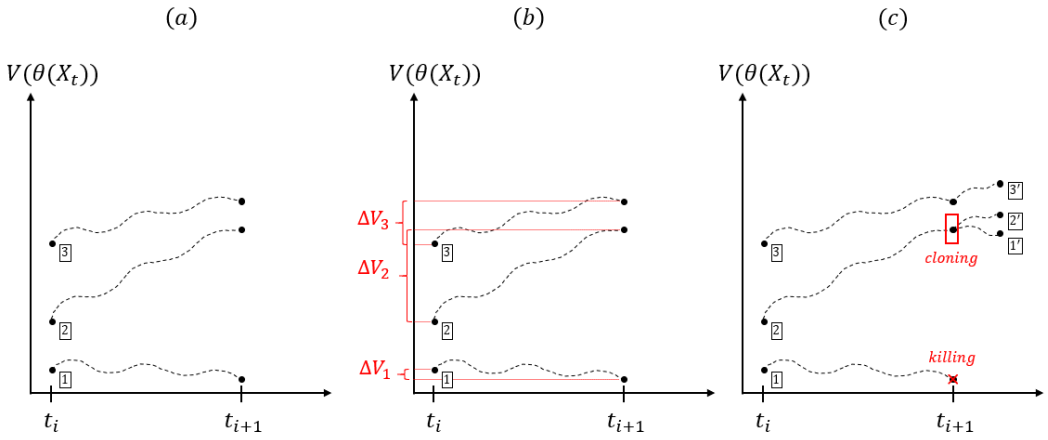


Figure 4.5: **Schematic representation of the Diffusion Monte Carlo rare events algorithm.** (a) Simulation step where the different members evolve from time t_i to time t_{i+1}.(b) Reweighting step. The ΔV_n represent the arguments of the exponential weights updated at this step. (c) Killing and cloning step when the members with the lowest weights are killed (member 1 here) and members with the highest weights are cloned (member 2 here).

The splitting functions $(V_i)_i$ are the most important parameters for determining the dynamics of DMC. It is therefore also the key weakness of the algorithm as it is quite sensitive to the particular splitting functions that are used. The most straightforward example of a splitting function is $V_i(X) = C\theta(X)$ where $C > 0$ is a splitting parameter chosen by the user that controls how many times a single particle can be split to create new copies. The main issue with this scheme is that the number of copies per member may become too much biased in favor of high values of θ for certain non-linear processes. In this case, one faces the problem of **extinction**, that is the fact that the final members selected by the algorithm are too similar to one another. While the algorithm remains unbiased also in this case, the estimator 4.43 may have a high variance, which renders the algorithm useless. As a consequence, the practical application of DMC requires tuning to avoid both extinction and low selection.

To counteract these issues, Webber et al. (2019) proposed to adapt the DMC algorithm to make it more robust, in the context of mesoscale extreme weather

events (tropical cyclones in their case). Their key idea is to adaptively rescales at each resampling time the score function θ to match a target distribution ν_i, which would typically be the normal distribution. It is the rescaled score function θ'_i that is used for splitting and killing of simulations. They call their algorithm Quantile Diffusion Monte Carlo (QDMC).

QDMC is distinguished from standard DMC by an adaptation step. After estimating the distribution of $\theta(X_{t_i})$, QDMC builds a transformation $\theta'_i = \gamma_i(\theta)$ so that the distribution of $\theta'_i(X_{t_i})$ approximates a target distribution ν_i. In particular, QDMC builds a transport function from distribution θ_i to the target distribution ν_i of the form

$$\gamma_i(y) = F_{\nu_i}^{-1}(F_{\theta_i}(y)). \tag{4.47}$$

Here, F_{θ_i} is the cumulative distribution function for the distribution θ_i and $F_{\nu_i}^{-1}$ is the quantile function for the distribution ν_i, i.e. the inverse of the cumulative distribution function F_{ν_i}. The typical target distribution is $\nu_i = \mathcal{N}(0, 1)$.

Intuitively, it is natural to ask for the rare events algorithms to be insensitive to a monotonic bijective change of the score function θ, which is not the case for the DMC algorithm but is for QDMC. QDMC definitely makes a step in the direction of such a more flexible rare events algorithm, even though the choice of some parameters remain arbitrary (see the discussion below).

c) Giardina-Kurchan-Tailleur-Lecomte algorithm

As first formulated by Giardina' et al. (2006), the Giardina-Kurchan-Tailleur-Lecomte (GKTL) algorithm aims to compute large deviation functions. It was for example used by Tailleur and Kurchan (2007) and Laffargue and Tailleur (2014) to compute rare trajectories using finite-time Lyapunov exponents, leading to what the authors call Lyapunov Weighted Dynamics (LWD). Giardina et al. (2011) reviewed the algorithm and its applications for various dynamical systems. The name 'GKTL' was employed by Ragone et al. (2018) and I follow this denomination.

In the GKTL algorithm the weights are chosen to favour members which exhibit large values of a time-averaged observable F, rather than values at resampling times as in DMC. The algorithm is nonetheless almost identical to the DMC algorithm if one chooses a primitive $\theta = \int F$ of the observable F as the score function and constant splitting functions $V_i(\theta) = k\theta$. I describe the algorithm as presented in Ragone et al. (2018).

The algorithm is depicted schematically in Figure 4.6. The implementation of the algorithm proceeds as follows. First, one initializes N independent members $(X_0^n)_{1\leq n\leq N}$. Then one iterates the following procedure over the resampling times $t_i = i\tau$ with τ the resampling duration, from $t_1 = \tau$ until $t_i = T/\tau$ (with T an integer multiple of τ):

1. **Simulation**: run each member from time $t_{i-1} = (i-1)\tau$ to time $t_i = i\tau$.

2. **Reweighting**: Assign to each member $X_{t_i}^n$ the weight:

$$w_i^n = \frac{e^{k\int_{t_{i-1}}^{t_i} F(X_t^n)dt}}{R_i} \tag{4.48}$$

with R_i the normalization factor so that: $\sum_n w_i^n = N$: $R_i = \frac{1}{N}\sum_{n=1}^{N} \frac{e^{k\int_{t_{i-1}}^{t_i} F(X_t^n)dt}}{R_i}$.

3. **Resampling**: create an ensemble of updated members $(\hat{X}_i^n)_{1\leq n\leq N}$ consisting of N_i^n copies of each member X_i^n. The numbers N_i^n must be randomly chosen to ensure: $\sum_n N_i^n = N$ and $\mathbb{E}[N_i^n] = w_i^n$ (see above for the resampling schemes).

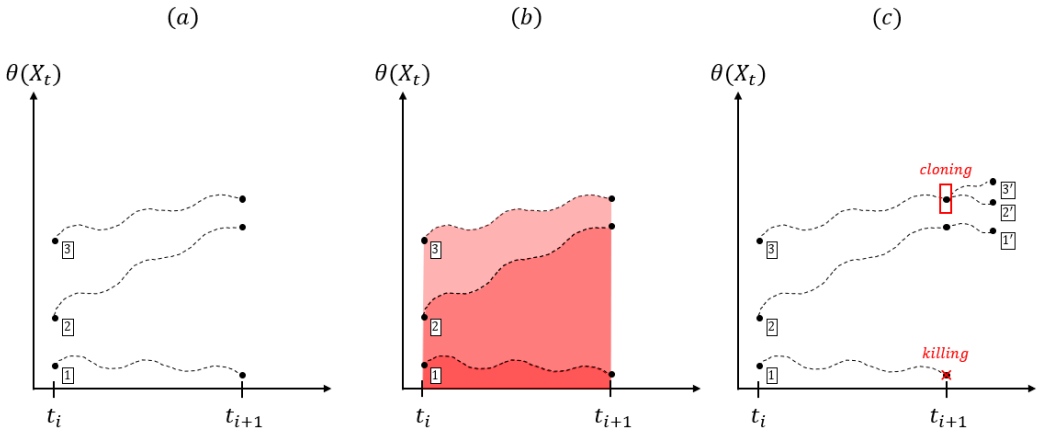


Figure 4.6: **Schematic representation of the Giardina-Kurchan-Tailleur-Lecomte rare events algorithm.** (a) Simulation step where the different members evolve from time t_i to time t_{i+1}. (b) Reweighting step. The red areas under the different curves represent the arguments of the exponential weights updated at this step. (c) Killing and cloning step when the members with the lowest weights are killed (member 1 here) and members with the highest weights are cloned (member 3 here).

The parameter k is a control parameter chosen by the user: positive (negative) values of k favor members with high (low) values of the time-averaged F observable. The highest the absolute value of k, the more selective the algorithm will be — with the risk of selecting only one or a few members at each resampling time. Ragone et al. (2018) emphasize that the N initial conditions should be independent and provide a reasonable sampling of the attractor of the system, even though this may be difficult to enforce in a high-dimensional system with N low. The version presented here assumes that resampling times are distributed homogeneously, more precisely that they are integer multiple of the resampling time τ. In principle this strategy could also be implemented with heterogeneous resampling times.

As in DMC, the distribution of members converges to the climatological distribution weighted by a likelihood ratio $L(X) = \frac{e^{k\int_0^T F(X_t)dt}}{\mathbb{E}_\mu[e^{k\int_0^T F(X_t)dt}]}$ which favors extreme values of the time-averaged observable F. For any observable ψ an estimator of $\mathbb{E}_\mu[\psi(\{X_t\}_{0\leq t\leq T})]$ is then:

$$\mathbb{E}_\mu[\psi(\{X_t\}_{0\leq t\leq T})] \simeq \frac{1}{N}\sum_{n=1}^{N}\psi(\{X_t\}_{0\leq t\leq T})w_n^{-1} \tag{4.49}$$

with

$$w_n = \Pi_{i=1}^{T/\tau} w_i^n. \tag{4.50}$$

d) Adaptive Multilevel Splitting

Multilevel splitting was the first idea developed to sample rare events and dates back to the work of Kahn and Harris (1951). Here I present a modern version of the algorithm, namely the Adaptive Multilevel Splitting (AMS) proposed by Cérou and Guyader (2007). Cérou et al. (2019) reviewed the algorithm, including recent theoretical advances such as unbiasedness and central limit theorems for its variance.

The general idea is to decompose a (small) probability p of transitioning from a set A to a set B into a chain of products of (larger) conditional probabilities. As in TPT, a reactive trajectory is a particular realisation of this transition. I present a version of the algorithm slightly modified from the original algorithm of Cérou and Guyader (2007). In this version, trajectories have a fixed duration T (Cérou et al. 2019; Bouchet et al. 2019; Rolland 2021). This description is closed to the Trajectory-Adaptive Multilevel Sampling (TAMS) algorithm described by Baars et al. (2021). For simplicity, I present a version of the algorithm where only one trajectory/member is resampled at each step, but note that it can be adapted straightforwardly for multiple resampling at each step.

The algorithm is depicted schematically on Figure 4.7. The implementation of the algorithm proceeds as follows. First, one initializes N independent members $(X_0^n)_{1\leq n\leq N}$ all starting in the set A. Contrary to DMC and GKTL, one then runs all members until $t = T$ is reached, or the member ends up in a target set B. Each member X_t^n has a certain maximum value of the score function θ in $[0, T]$ which can be noted Q^n. Set $k = 1$ and $w_0 = 1$. One then repeats the following steps until $\min_n Q^n \geq z_{max}$ with a user-chosen value of z_{max} or the number of iterations k is greater than a user-chosen value K:

1. **Identify the member with minimal score function**: consider l so that $Q^l = \min_n Q^n$, which is the index of the member for which the maximum value of the score function over $[0, T]$ is minimal. This member is stored and assigned the weight w_{k-1}. Since the member X_t^l has the smallest value of Q^n, all other members n have some time t_j for which $\theta(X_{t_j}^n) \geq Q^l$.

2. **Update the weight**:
$$w_k = \left(1 - \frac{1}{N}\right) w_{k-1} \tag{4.51}$$

3. **Resample the member with minimal score function**:

   - Select randomly a member X_t^r with $r \neq l$ among the remaining members and set $(\tilde{X}_t^l)_{0 \leq t \leq t_{min}} = (X_0^r, ..., X_{t_{min}}^r)$ where t_{min} is the smallest time for which $\theta(X_{t_{min}}^r) \geq Q^l$.
   - Simulate the rest of the trajectory starting from $X_{t_{min}}^r$ until again the member reaches either $t = T$ or the set B. This member has a new maximum value of the score function $\tilde{Q}^l$ which is always greater than or equal to Q^l.
   - Set $X^l = \tilde{X}^l$ and $Q^l = \tilde{Q}^l$.

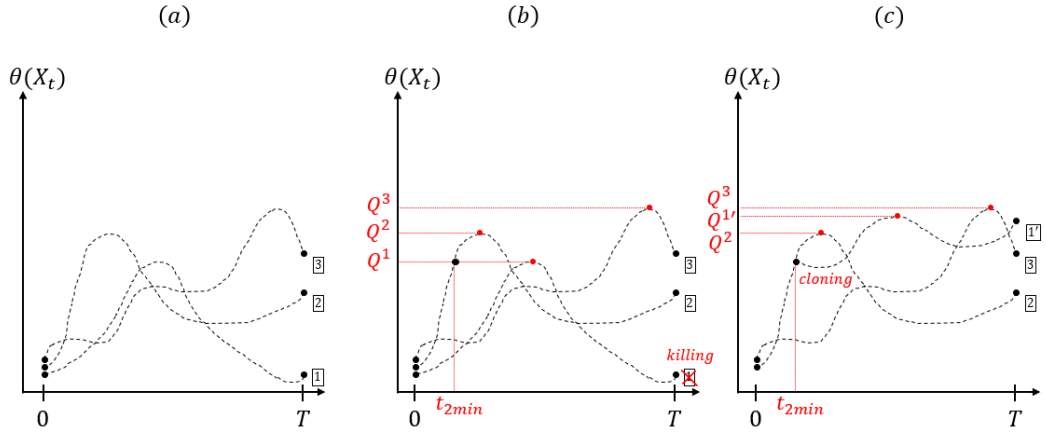


Figure 4.7: **Schematic representation of the Adaptive Multilevel Splitting rare events algorithm.** (a) Initial run of all members over the time period from 0 to T. (b) Identification of the member with the minimal value of the maximal score function value (member 1 here). This member is then killed. (c) Resampling of the killed member starting from the time t_{2min} corresponding to the minimal time for which the member from which it is resampled (member 2 here) reached the value of the maximum score function of the killed member. The member from which the new member is resampled is chosen randomly among the remaining ones.

The weights computed at each step represent the probability of a member reaching iteration $k+1$. At the end of the iteration loop, the remaining members are assigned the weight w_K. This gives an unbiased estimator of the transition probability:

$$p_{AB} \simeq \frac{N_B w_K}{N} = \frac{N_B}{N} \prod_{k=1}^{K} (1 - \frac{1}{N}) \tag{4.52}$$

where K is the final number of iterations and N_B is the number of members that reached B. The climatological expectation of any observable ψ can be computed using the following formula:

$$\mathbb{E}_\mu[\psi(X_t)] \simeq \frac{1}{N} \sum_{k=0}^{N+K-1} \psi(X_t^k) \frac{w_k}{\sum_{k=0}^{N+K-1} w_k} \tag{4.53}$$

where for $k \geq K$, $w_k = w_K$. The summation is computed over all the discarded ($0 \leq k < K$) and the remaining members ($K \leq k < K + N$).

As it is made clear by the separation between the set A and the set B, the AMS algorithm is especially suited to estimate transition probabilities between two sets, especially between two fixed points or two basins of attraction. It has to be adapted for studying extreme values of an observable ψ. In this latter case, one uses ψ itself as the score function. It is not mandatory to explicitly prescribe the set B and therefore each member is run until time T, whatever the values reached by ψ. The probability of ψ reaching values above a threshold $a \in \mathbb{R}$ is then computed using the formula 4.53 with the observable $\mathbb{1}(\psi(X_t) \geq a)$.

Contrary to the DMC and GKTL algorithms, the AMS algorithm does not run several members in parallel and selects at regular time step. The AMS algorithm first run all members for a fixed time and then re-run some parts of the members that have not reached high enough values of the score function. Therefore, the computation and space requirements are different from DMC and GKTL (especially in the number of initial members) and the total size is only partially controlled by the K parameter.

The AMS algorithm may also be subject to the extinction problem if one tries to reach high values of the observable ψ. Indeed, up to a point, because members with minimal score function are resampled from other members, the algorithm has a tendency to replace all members by a single one reaching the highest values of the score function and selecting time of resampling only at the very end of the simulation. Therefore the total number of iterations K should remain reasonable. A good approximation for the number of iterations (Lestang et al. 2018) is that to reach events with a probability $p \simeq 10^{-\beta}$ one should set approximately $K = N\beta$ iterations.

4.4.4 Example with the Orstein-Uhlenbeck process

To illustrate the DMC, GKTL and AMS algorithms, I show their application on the Orstein-Uhlenbeck (OU) process presented above with $\alpha = 1$ and $\sigma = 1$. I seek to estimate paths leading to extreme (positive) values of the observable $F : X_t \to X_t$ over a finite simulation time $T = 1$ and their corresponding probabilities. The three algorithms are therefore run using the simple score function $\theta = F : X_t \mapsto X_t$.

The integration is made using the order (3.0,1.5) strong stochastic Runge-Kutta methods for SDEs proposed by Rößler (2010) with a step $dt = 0.001$ and a total integration time $T = 1$. For both DMC and GKTL algorithms, I choose to implement 10 homogeneous resampling times between 0 and T: $t_i = i \times \frac{T}{10}$. DMC and GKTL algorithms are run with $N = 1000$ parallel members. I choose for the splitting parameters $C = 3.5$ for DMC and $k = 10$ for GKTL. For the AMS algorithm, the number of initial members is $N_{ini} = 200$ and I choose $K = 1500$. The values of

the parameters were chosen by trial and error to obtain similar extreme values with the three algorithms (see below).

Figure 4.8 presents the results of the application of these algorithms. The first (resp. second and third) column displays the application of the DMC (resp. GKTL and AMS) algorithm. The first row shows the evolution of all members between each resampling times for DMC and GKTL — i.e. before the resampling step and therefore including members which are subsequently killed — and all discarded members for AMS. The second row displays the final members selected by the algorithms, i.e. the members that have survived the selection process. The third row compares the probabilities of the OU process reaching a threshold a for $a \in [0, 3.5]$ at $t = T$ starting from $X(t = 0) = 0$. The exact probability is computed using formula 4.27 and is depicted in red. The black dots correspond to the probability estimated by the corresponding algorithm.

Figure 4.8hij demonstrate the efficiency of the rare events algorithms: they are able to sample paths leading to extreme events with a probability as low as 10^{-6} while running only $N = 1000$ members. The estimation of the corresponding probabilities is accurate and begins to deviate from the true probability only for high threshold ($a > 3$) for which the variance of the algorithms increases again. Running the same algorithms several times would allow to partially control this issue.

One may note that the final trajectories selected are not similar for the three algorithms. For DMC, the trajectories are overall pretty close to the instanton presented in Figure 4.3a. If I had chosen a stronger C parameter, the algorithm would have reached lower probabilities but the selection would also have been stronger and therefore the extinction problem may have been more important, leading to an increase in the variance of the estimated probabilities. For the GKTL algorithm, because the weights are computed using the full integral of $\theta(X_t) = X_t$, it is not anormal that the algorithm selects concave trajectories because the instanton is different than for DMC as presented in the difference between Fig.4.3a and Fig.4.3b. I used the GKTL algorithm to estimate maximum values of X_t and not integrals of an observable as the algorithm is intended to be used, which my explain the somewhat less precise estimation of low probabilities of X_T.

Lastly, it should be noted that the AMS algorithm produces more trajectories (2200 vs 1000) than the DMC and GKTL algorithms because of its different resampling strategy. This is a little more space consuming but also allows precise estimates of the corresponding probabilities. The final trajectories selected (Fig. 4.8f) seem to be in between the two different instantons in Fig. 4.3. Overall each algorithm selects the instanton corresponding to different observables: maximizing the value at the end of the trajectory for DMC and maximizing the integral over the trajectory for GKTL. For the AMS algorithm, if one was to increase the total number of trajectories simulated K one would sample trajectories that look more like those selected by DMC (not shown). Even though each algorithm provides unbiased estimates of the mean of any observable along the selected path, it is clear that the most likely trajectory for reaching extremes of an observable or for moving from one fixed point to another may not be the one sampled by the chosen algorithm. This point is crucial when interpreting the physics of the trajectories sampled by the rare

events algorithm.

The example of the application of splitting methods on the OU process shows the strength of the rare events algorithms but also substantial challenges for choosing the parameters of the algorithms. In the next section, I shall discuss more this issue and propose some guidelines and solutions when it is already possible.

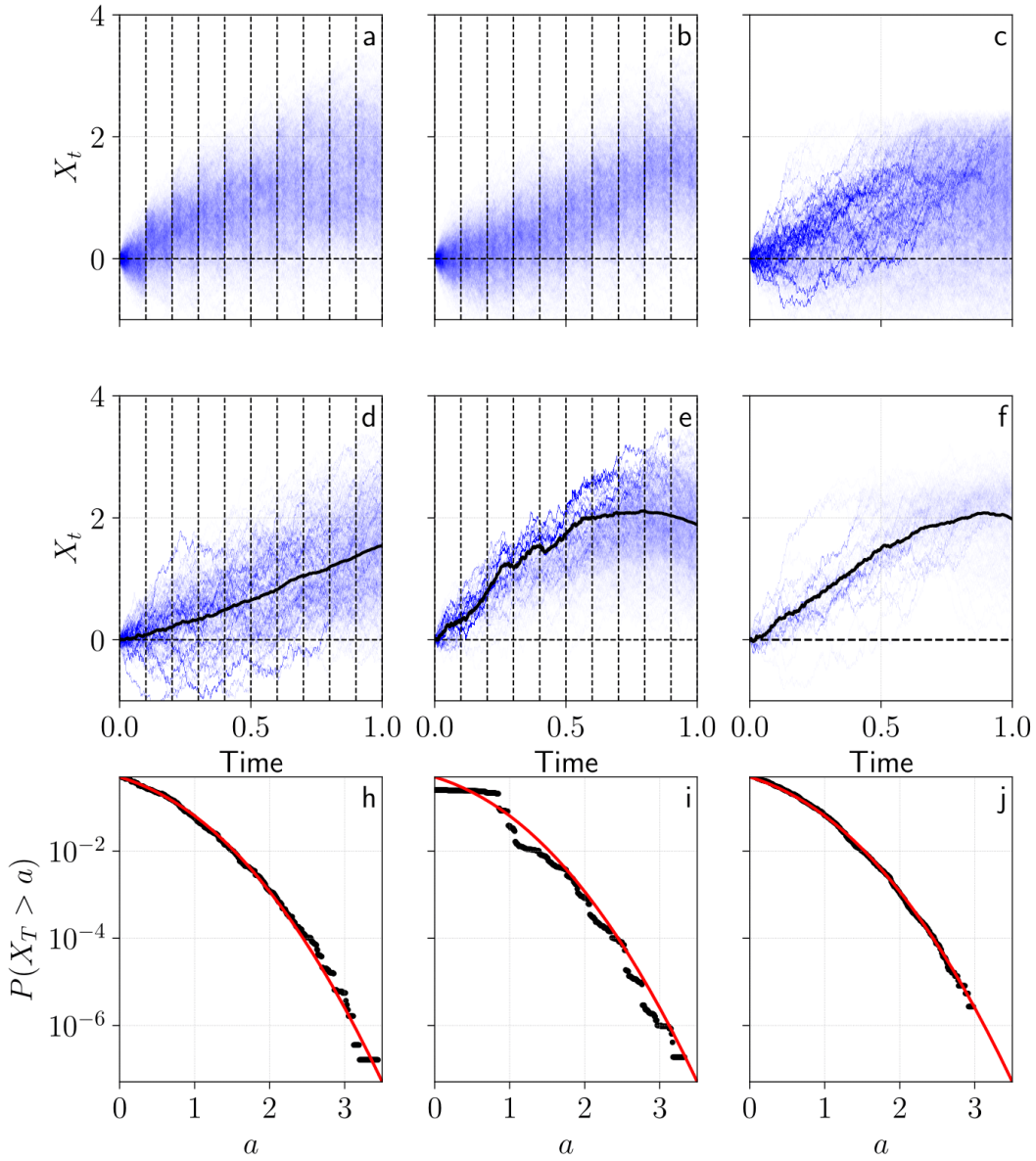


Figure 4.8: **Application of the DMC, GKTL and AMS rare events algorithms to the Orstein-Uhlenbeck process.** First (resp. second and third) column displays the application of the DMC (resp. GKTL and AMS) algorithm to the OU process ($\alpha = 1$ and $\sigma = 1$). (a, b and c) Evolution of all members between each resampling times for DMC and GKTL and all discarded trajectories for AMS. Vertical dashed lines represents the resampling times for DMC and GKTL. (d, e and f) Evolution of the final members selected by the algorithms. The black line is the average of final members. (h, i and j) Probabilities of the OU process reaching a threshold a for $a \in [0, 3.5]$ at $t = T$ starting from $X(t = 0) = 0$. The exact probability is depicted in red and the black dots correspond to the probability estimated by the corresponding algorithm.

4.5 Challenges for the application of splitting methods

Splitting methods have found promising applications for sampling rare events in geophysical fluid systems. The algorithms presented above greatly increase the available statistics for rare events of interest while also providing unbiased probability estimates. However, there are also several drawbacks and opened questions concerning their implementation that prevent a more widespread use of these methods. In this section, I shall try to point out some of them and propose some solutions.

4.5.1 Score function

The efficiency of splitting methods for sampling extreme events are based on the existence of a score function θ which maps the phase space and gives a 'distance' from the class of rare events one wishes to reach. Even though these algorithms are unbiased in the large ensemble limit, their performance depends dramatically on the quality of the score function and/or splitting functions used. On the theoretical side, the conditions that have to apply on the score function mainly consist in having a non-zero gradient everywhere on the phase space, always pointing in the direction associated with the rare event of interest (Cérou et al. 2011). These conditions are difficult to ensure in practice and authors have mainly chosen score functions heuristically using their physical knowledge of the key processes of the system studied.

For the DMC and GKTL algorithms, when looking to sample extreme values of an observable, the natural score function would be the observable itself — as used directly in the GKTL algorithm. For instance, if one seeks to sample extreme heatwaves, it is natural to choose 2-m air temperature as the score function (cf. chapters 6 and 7). However this choice does not ensure the performance of the algorithm to be good in every possible system. One could imagine that high values of this observable are not reached gradually and therefore the algorithms may not be able to select correctly the relevant trajectories. I give here an example of this behavior. Let us define the following stochastic systems of equations:

$$\begin{cases} dX_t = -\alpha X_t dt + \sigma dW_t \\ U_t = X_t \cos(\omega t). \end{cases} \tag{4.54}$$

Suppose one has access only to U_t and wants to sample its extreme values with a rare events algorithm. There are two time scales involved in the system: $\tau_X = 1/\alpha$ the time scale of X_t and $\tau_{\cos} = 2\pi/\omega$ the time scale of the cosinus oscillation. Depending on the ratio between these two time scales, the variable U_t itself may not be used as a good score function to sample its own extremes:

- if $\tau_X \ll \tau_{\cos}$ then the cosinus is a slow variable that is almost not changing on the time scales of X and therefore using U directly as a score function can sample extremes (even though these extremes may not be maximally extreme if for example $\cos(\omega t)$ is close to 0),

- if $\tau_X \gg \tau_{\cos}$ then the cosinus is a fast variable on the time scales of X and U cannot be used directly as a score function: the fast oscillation of the cosinus imposes a time scale to U and do not let enough time to members selected by the rare events algorithm to separate so that the selection procedure leads the system towards extreme values of U.

This issue is illustrated in Figure 4.9. Fig. 4.9a shows a case where $\tau_X \gg \tau_{\cos}$. I first run an initial trajectory and then 100 trajectories are initiated at $t = 0.25$. Because the time scale of the OU process in X is slower than the oscillation of the cosinus, all trajectories follow the oscillation of the cosinus and as a consequence, if one were to use U as a score function to sample extremes, it would not be possible to sample extremes: the trajectories resampled have no time to separate and reach even greater value of the observable. Depending on the phase of the cosinus oscillation, high values of U are not a good score function to sample even more extreme values of U in the future. On the other hand, Fig. 4.9b shows a case where $\tau_X \ll \tau_{\cos}$. In this case, the slow oscillation of the cosinus is almost irrelevant for reaching high values of U: only the dynamics in X is important and this dynamics can be sampled by a rare events algorithms. Compared to the first case, in the second case trajectories resampled have enough time to separate and reach higher values of the score function without being forcibly embedded in the dynamics of the cosinus oscillation.

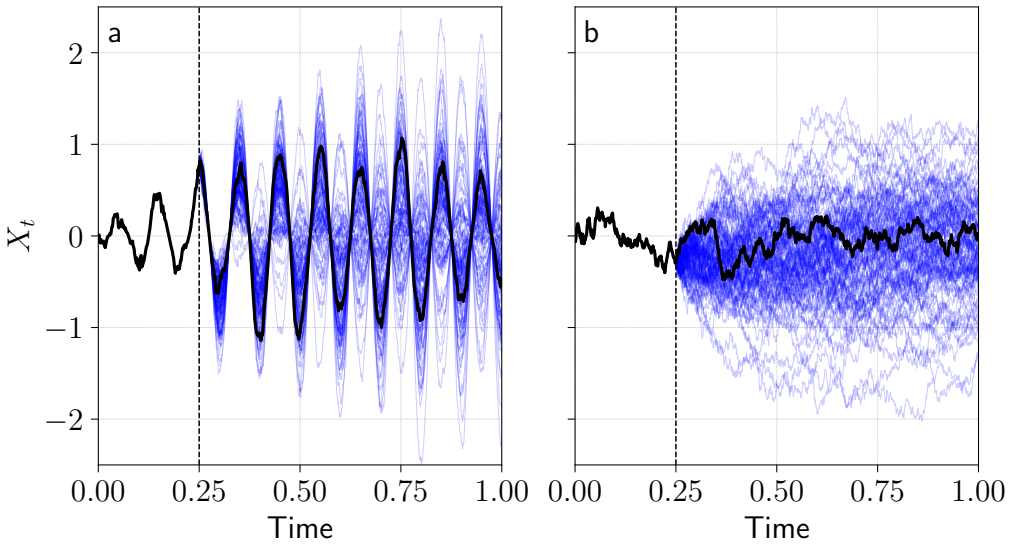


Figure 4.9: **Illustration of the time scales issue when choosing the score function.** (a) Evolution of one trajectory with $\tau_X = 1$ and $\tau_{cos} = 0.1$ (black) and 100 similar trajectories starting at $t = 0.25$ from the initial trajectory (blue). (b) Evolution of one trajectory with $\tau_X = 1$ and $\tau_{cos} = 10$ (black) and 100 similar trajectories starting at $t = 0.25$ from the initial trajectory (blue).

This simple system is an example of a more general problem. If the typical time scale $\tau_{ext}(\psi)$ associated with an extreme of the observable ψ is smaller than $\tau_{sep}(\psi)$ the typical time scale of separation of the observable ψ for two trajectories initialized closely (i.e. close to the Lyapunov time scale of the system), then one cannot simulate the extreme events of ψ using $\theta = \psi$ as a score function. One needs to find another score function $\tilde{\theta}$ for which $\tau_{ext}(\psi) > \tau_{sep}(\tilde{\theta})$ and ideally $\tau_{ext}(\psi) \gg \tau_{sep}(\tilde{\theta})$. Whether this score function always exists and is easy to find may be a difficult question to answer. As a consequence, the DMC and GKTL algorithms are not suited to sample extremes which have a typical time scale inferior to the Lyapunov time scale of the system, which is around 5–10 days for the atmosphere. To give an example, using cumulative precipitations as a score function, the GKTL algorithm is suited to study monthly or longer anomalies but not daily extremes — unless one is capable of providing a score function which can 'predict' the occurrence of short extremes weeks in advance. For short extremes, the ensemble boosting approach is more relevant as one already knows that an extreme has occurred (see also Finkel and O'Gorman (2024)). See also Lestang et al. (2018) who found a similar issue with the AMS algorithm when sampling extreme pressure fluctuations in a turbulent channel flow.

For the AMS algorithm, the optimal score function is the committor function (Giardina et al. 2011). However this function is unknown as it is precisely the function one wishes to estimate with the algorithm. A good score function should therefore display the same isosurfaces as the one of the committor functions but using a non-optimal score function leads to reactive trajectories that do not correspond to the most probable one (Rolland and Simonnet 2015). A recent proposition made by Lucente et al. (2021) is to use machine learning methods to learn an approximate committor function that would then be used to guide the algorithm, provide better statistics of rare events and therefore learn a more accurate committor function (Miloshevich et al. 2023). This reinforcing feedback loop is appealing on the paper but its applicability to high-dimensional geophysical systems has still to be demonstrated.

More generally, if one already has clues on the phenomenology of the rare events under study (e.g. low soil moisture and persistent anticyclonic conditions leading to heatwaves in Europe, cf chapter 1 section 1.3), it may be possible to build an approximate score function with a low-dimensional projection of the system phase space. One could for example use Fourier modes, wavelets basis or Empirical Orthogonal Functions (EOFs) and couple them with time-dependent splitting functions such as in Webber et al. (2019) or Bréhier and Lelièvre (2019).

Another possibility illustrated by Tailleur and Kurchan (2007) and Laffargue and Tailleur (2014) is to use a 'ubiquitous' measure of rarity in the phase space, grounded in dynamical systems theory, such as finite-time Lyapunov exponents in their case. Using this kind of measures does not target a specific rare event but rather uses properties of the phase space to sample ubiquitously 'rare events', i.e. regions of the phase space with low values of the climatological probability distribution. Other examples of such interesting indices in the phase space include local dimension and/or analogs quality (Faranda et al. 2017a). To the best of my knowledge, this has never been done for complex high-dimensional systems. This proposition may be

interesting to explore the variability of the climate attractor but it does not ensure that rare values of the observable of interest will be reached.

Overall, the procedure to find relevant score functions for targeting rare events of interest in geophysical dynamical systems is still empirical. One may notice that this issue is fundamentally hard or even impossible in so far as it consists in predicting the direction that the model will follow before doing the actual calculation. This definitely has strong theoretical limitations especially because of mixing properties of chaotic systems such as most geophysical fluid ones (Katok and Hasselblatt 1997). Nevertheless, the iterative process consisting in simulating some rare events and using the knowledge obtained on the phenomenology leading to these extremes to sample more efficiently other rare events is promising, especially in combination with so-called reinforcement learning algorithms able to infer patterns in large dataset (Lucente et al. 2021).

4.5.2 Resampling times and total integration time

Resampling times. To the best of my knowledge, there is no theoretical principles to guide the choice of resampling times for the DMC and GKTL algorithms. Intuitively, there is however a clear trade-off between too short and too long resampling times.

Selections should indeed not be performed too frequently, as cloning increases correlations between the members and therefore reduces the effective number of independent members, increasing the estimator variance: this is the already-mentioned extinction problem. For deterministic systems and stochastic systems with a weak noise, cloned members may also not have the time to separate one from another, making the selection step irrelevant.

On the contrary, if not performing selections frequently enough, the members distribution relaxes to the unbiased climatological probability distribution, leading to the poor brute force Monte-Carlo variance. Those two requirements usually translate into resampling intervals close to the Lyapunov time and/or mean autocorrelation time of the system. However, the system studied may behave differently depending on the region of the phase space where it is evolving (the finite-time Lyapunov time may be different from the mean Lyapunov time). For example, with the Orstein-Uhlenbeck starting from any position $X_0 = x$, the mean converges to 0 as $\mathbb{E}[X_t] = xe^{-\alpha t}$ and the variance converges to 1 as $\mathbb{V}[X_t] = 1 - e^{-2\alpha t}$.

Therefore, even though the easiest way to define the resampling times is to use a homogeneous scheme — given a resampling time τ and a total integration time T that is an integer mutliple of τ, one then defines the resampling times as $t_i = i\tau$ — it may not be the most appropriate choice, especially to sample extreme values of an observable which are situated in phase space regions intrinsically repellent for the system. Webber et al. (2019) proposed an alternative time-heterogenous resampling strategy: they define resampling times $(t_i)_{1\leq i\leq K}$ so that

$$\int_0^{t_1} e^{2\alpha t} dt = \int_{t_1}^{t_2} e^{2\alpha t} dt = \cdots = \int_{t_{K-1}}^{t_K} e^{2\alpha t} dt \tag{4.55}$$

with α the decaying parameter in the Orstein-Uhlenbeck process and more generally an indicator of the rate of decrease of the variance of the observable of interest.

With the time-homogeneous strategy the natural choice of splitting functions for DMC is $V_i(x) = C\theta(x)$ whereas for the time-heterogeneous strategy, Webber et al. (2019) propose to define splitting functions as $V_i(x) = Ce^{\alpha(t_i-1)}\theta(x)$. Therefore, with the time-heterogeneous strategy, strength and frequency of splitting increase exponentially with time. This translates into more clones for members with high values of the score function θ as the last resampling time is approached (see also (Wouters and Bouchet 2016)).

Total integration time. The total integration time of splitting methods is not often mentioned as a relevant parameter to investigate. There are indeed few theoretical requirements associated with its value. Intuitively, it is clear that if the total integration time is too short, the noise to be added to reach an extreme value of an observable or another basin of attraction is unrealistically large. However, if the total integration time is too long, then the trajectory may not be representative of the actual path followed by the system to reach a rare region. The instanton is indeed the Freidlin-Wentzell action minimizer over all integration times but it is not possible in practice to let $T \to \infty$. Authors often use heuristics to choose this total integration time T, the main one being that T should be of the order of 5 to 10 times the auto-correlation time associated with the observable of interest when sampling its extreme values.

There is actually a trade-off between the number of resampling times K and the total integration time T. To reach a given extreme value of an observable on a dynamical system, the best strategy would be to select frequently (K large) and let the system evolve for a long time (T large). However, such a strategy encounters rapidly the extinction problem: because at each resampling time one overwrites the killed members with the cloned members, the number of effectively different members decreases rapidly and the variance of the estimator becomes large.

Selecting frequently with a small total integration time (K large, T small) would reduce the extinction problem in so far as having smaller resampling times means less spread in the weights, weaker selection and therefore less extinction. However, a small T gives the system less time to reach a high value of an extreme. On the contrary, selecting less frequently but over a long period of time (K small, T large) would let enough time to the system to reach high values of the observable of interest and avoid the extinction problem but not selecting strongly enough let the system comes back to its climatological probability distribution which do not display extreme values.

To illustrate this trade-off, I run the DMC algorithm on the OU process for pairs of the K and T parameters: $K = 2, 4, \ldots, 20$ and $T = 0.25, 0.50, \ldots, 2.50$. For each pair ($K$,$T$), the DMC algorithm is run 50 times with homogeneous resampling times, $N = 500$ trajectories and $C = 3.5$. Figure 4.10 displayed the results for metrics of interest. Fig. 4.10a shows the mean value reached by the trajectories at $t = T$. As expected, the highest values are reached for K and T as large as possible. The associated probabilities computed by the algorithm is shown in Fig. 4.10b. For each

pair (K, T) the average $\hat{p}$ of probabilities estimated by the 50 runs of the algorithm is shown. Fig. 4.10c illustrates the extinction problem. It shows the mean extinction rate for the 50 runs, defined as the proportion of initial trajectories that have been killed along the selection process. This extinction rate increases with K and T. Finally, Fig. 4.10d shows the relative error $\sqrt{\mathbb{V}(\hat{p})}/\hat{p}$ for the estimated probability $\hat{p}$, where the variance is computed over the 50 runs. The relative error is minimum for K small and T large.

From Figure 4.10, it seems that the best trade-off for a splitting parameter C for the DMC algorithm is given by an integration time as long as possible — although this may be cost-limited in practice for complex geophysical models — and an intermediate number of resampling times, around 10 in this case. Choosing such values allows an equilibrium between reaching extreme values of an observable while at the same time keeping the estimation of the corresponding probabilities precise. These observations point towards a classical bias-variance dilemma for choosing the implementation parameters of the rare events algorithm: selecting more often decrease the bias (with respect to an extreme levels one wishes to attain) but increases the variance of the estimated probability.

4.5.3 Number of parallel members

Theoretical analyses (Del Moral 2004) ensure that the algorithms presented above converge to the true probability of the rare events when the number of members run in parallel N goes to infinity. However, to the best of my knowledge there is no indication for the actual number of parallel members N to run for practical applications, especially in high-dimensional systems. Authors often use around 100 members but the reasons are mostly because of computational cost issues rather than a thorough theoretical analysis. In practice, the larger N the better: a too small value of N not only gives results with large variance but also triggers extinction by loss of diversity (all members collapse to a single one). An alternative is also to run the algorithm several times with similar initial conditions, which weakens the extinction risk and also allows to estimate a variance of the estimator associated with each algorithm. For a total fixed computational cost, there is therefore a trade-off to find between the number of times the algorithm is run and the number of members by run.

4.5.4 Sensitivity to parameters

Every algorithm presented above has user-tuned parameters (see table 4.2), especially when it comes to the intensity of the selection (C parameter in splitting functions for DMC and k parameter for GKTL). These parameters can require delicate tuning as the algorithms can become unbalanced. Members with the highest positions could be split into dozens or hundred of clones thus members become highly correlated leading to volatile and error-prone estimates for rare event probabilities. QDMC is a possible response to this over-selecting problem, transforming every distribution of the score function at resampling times to a normal distribution and

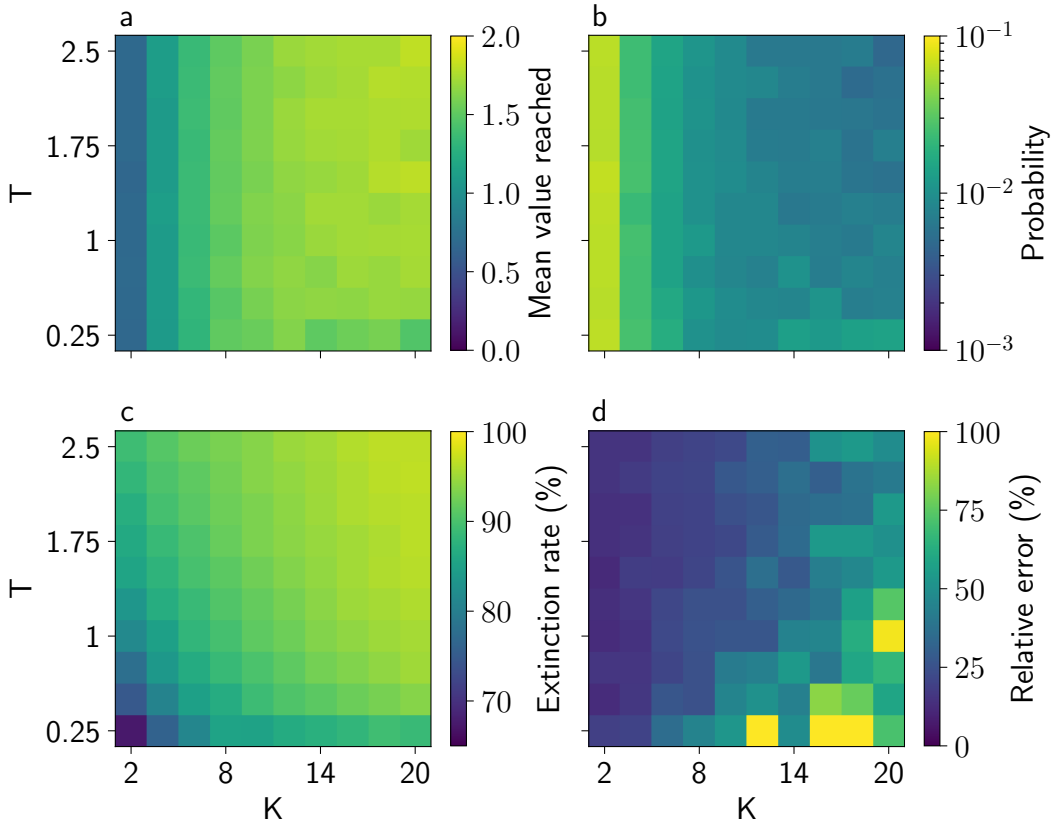


Figure 4.10: **Illustration of the trade-off between the number of resampling times K and the total integration time T with the DMC algorithm and the Orstein-Uhlenbeck process.** For each pair (K,T), the DMC algorithm is run 50 times with homogeneous resampling times, $N = 500$ trajectories and $C = 3.5$. (a) Mean value $a_{(K,T)}$ reached by the members at $t = T$. (b) Mean estimated probability $\hat{p}$ given by the algorithm for the mean value $a_{(K,T)}$: $\hat{p} \simeq \mathbb{P}[X_T > a_{(K,T)}]$ (c) Proportion of initial trajectories discarded by the algorithm. (d) Relative error $\sqrt{\mathbb{V}(\hat{p})}/\hat{p}$ for the estimated probability $\hat{p}$.

therefore rendering every selection step equivalent. However such a normalisation procedure does not exist for the GKTL algorithm. In general, tuning may become much more difficult when sampling a process with poorly known dynamics.

4.5.5 Sensitivity to initial conditions

The question of initial conditions for the members of the splitting algorithms is seldom treated in the literature. When one studies multistability, it is rather understandable as one wishes to go from one fixed point to another and therefore taking initial conditions close to the starting fixed point is natural. It is also the case when studying transition between two basins of attraction, even though it can be trickier in this case — if the initial conditions on one attractor are close to the separation

Algorithm	Parameters
DMC	- total integration time T - number of members N - resampling times $(t_i)_i$ - splitting functions $(V_i)_i$ - score function θ - initial conditions - splitting parameter C
GKTL	- total integration time T - number of members N - resampling times $(t_i)_i$ - score function θ - initial conditions - splitting parameter k
AMS	- total integration time T - number of members N - score function θ - initial conditions - number of iterations K

Table 4.2: User-chosen parameters for the three splitting algorithms presented

between the two basins of attraction it may be easier to cross the frontier.

When studying extreme values of an observable, the question seems however more relevant as the attractor may be extended in a high-dimensional space and the instanton leading to rare regions may be different starting from one point or another of the attractor. On paper, initial conditions should sample uniformly the attractor. However, it is not clear how to ensure such a condition on a high dimensional attractor. In particular, certain region of the phase space may be more important to sample than others, for example sea surface temperature of the North Atlantic for temperature extremes in Western Europe.

One solution could be to condition the results of the algorithm on a region of the phase space $\mathcal{X}_0$ where the initial conditions are drawn. This is equivalent to estimating conditional means $\mathbb{E}[F(X_t)|\mathcal{X}_0]$ and making the choice of the initial conditions set $\mathcal{X}_0$ clear. This is particularly relevant when there are variables in the system that have typical time scales much slower than the time scale of the extremes sampled by the rare events algorithms.

4.5.6 Multiple time scales

The path in the phase space leading to rare events may involve several time scales that are not taken into account by the algorithms. For example, in the climate system, to reach the most intense heatwaves in British Columbia, one may need first an intense ENSO event (time scale of several years), low soil moisture conditions due to a Spring drought (time scale of several months) and a persistent anticyclone (time scale of several weeks). If one uses a rare events algorithm with a total integration time of one or two months, it would not be able to grasp the full potential of the rare event.

The importance of taking into account multiple time scales phenomena to simulate rare events is a profound question about the very nature of extreme events in geophysical dynamical systems. Sapsis (2021) proposed to differentiate between extreme events due to (i) weak energy transfers between wavenumbers without the occurrence of positive Lyapunov exponents and (ii) violent energy transfers between degrees of freedom associated with intermittently positive Lyapunov exponents in the underlying dynamics. I could rephrase this distinction by saying that events of type (i) result from co-occurrence between phenomena with different time scale while events of type (ii) result from an intrinsic instability of the system. This distinction has also consequences when using splitting methods to sample extreme events.

In the second case, the question of multiple time scales phenomena is not a strong issue as what is at stake is to find the right way to guide the system towards the instability region to sample extreme events (however this may be difficult in practice). In the first case however, the algorithm has to sample phenomena occurring at different time scale and therefore may have to select trajectories based on a complex and maybe time-dependent score function.

Webber et al. (2019) mentioned that applying rare event sampling in a dynamical downscaling context should be a multitiered process: a GCM can be run either directly or with a splitting method. Then starting from the initial conditions selected from GCM output, a regional weather model can be run either directly or with a splitting method. To the best of my knowledge there is no practical application which used such a framework and there is also no theoretical results taking into account those multiple time scales with many phenomena interacting, but this idea seems of definite interest for efficiently sampling extreme events in geophysical dynamical systems.

To illustrate this issue, I propose to apply the DMC rare events algorithm to the following system of two OU processes with different time scales:

$$\begin{cases} dX_t^1 = -\alpha X_t^1 dt + \sigma dW_t^1, \\ dX_t^2 = -\alpha\beta X_t^2 dt + \sqrt{\beta}\sigma dW_t^2. \end{cases} \tag{4.56}$$

These two processes are independent one from another and have the following relationship between their PDF : $\rho^1(x,t) = \rho^2(x, t/\beta)$. In particular, they have the same stationary PDF. The processes X_t^1 and X_t^2 are in fact the same OU process with different time scales: if for example one chooses $\beta = 0.01$ then X_t^2 is 100 times 'slower' than X_t^1. X_t^2 is intended to model a slow variable in the system (for example the oceans) while X_t^1 represents the fast variable of the system (for example the atmosphere).

I am looking for extremes of an observable which would depend on both of these variables: I choose the observable $F : (X_t^1, X_t^2) \mapsto X_t^1 + X_t^2$. Because X_t^1 and X_t^2 are independent, their joint distribution is $\rho_{(X^1,X^2)} = \rho^1 \times \rho^2$. The probability of F taking values above a threshold a over a time T can therefore be computed as the convolution product of ρ^1 and ρ^2:

$$\mathbb{P}[F \geq a] = \int_{-\infty}^{+\infty} \rho^1(x^1, T) \left(\int_{a-x^1}^{+\infty} \rho^2(x^2, T) dx^2 \right) dx^1 \tag{4.57}$$

where ρ^1 and ρ^2 are given by formula (4.27).

I first model what would happen if one applies a rare events algorithm for maximizing this observable while only taking into account the typical time scale of the X_t^1 process. To do so I run the DMC algorithm with 10 homogeneous resampling times over a total integration time $T = 1$, with a splitting strength $C = 3.5$ and $N = 1000$ members. The score function is F itself. For initial conditions, I take $X_0^1 = 0$ and X_0^2 distributed according to its stationary PDF. Results of this first strategy are depicted in Figure 4.11.

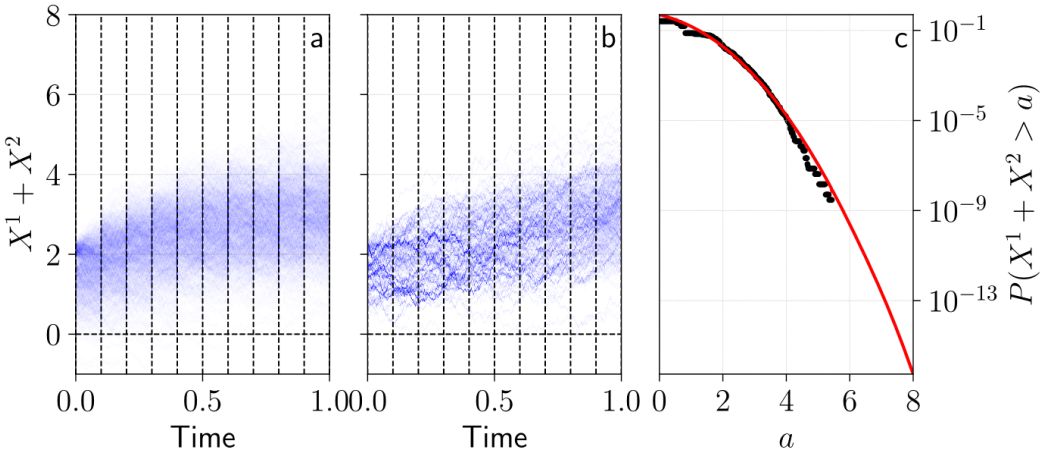


Figure 4.11: **Reaching extreme values of an observable with two Orstein-uhlenbeck processes without preliminary selection.** Application of the DMC algorithm to the system 4.56 with 10 homogeneous resampling times over a total integration time $T = 1$, with a splitting strength $C = 3.5$ and $N = 1000$ trajectories. The score function is F itself. For initial conditions, I take $X_0^1 = 0$ and X_0^2 distributed according to its stationary PDF. (a) Evolution of the observable $F(X_t^1, X_t^2) = X_t^1 + X_t^2$ for all members between each resampling times. Vertical dashed lines represents the resampling times. (b) Evolution of the final members selected by the algorithm. (c)Probabilities $\mathbb{P}[F \geq a]$ for $a \in [0, 8]$. The exact probabilities are depicted in red and the black dots correspond to the probabilities estimated by the algorithm.

Secondly, I propose an algorithm in two steps:

1. The first step consists in finding high values of the variable X_t^2 by running the DMC algorithm with 10 homogeneous resampling times over $T' = 100$, a splitting strength $C = 3.5$ and $N = 1000$ trajectories. The score function is $F' : (X_t^1, X_t^2) \mapsto X_t^2$. For initial conditions, I take X_0^1 and X_0^2 distributed according to their stationary PDF.

2. The second step consists in finding high values of the variable X_t^1 by running the DMC algorithm with 10 homogeneous resampling times over $T = 1$, a splitting strength $C = 3.5$ and $N = 1000$ trajectories. The score function is

F. For initial conditions, I take $X_0^1 = 0$, but I take the $X_{T'}^2$ found by the first stage.

To compute the expectation of any observable G, formula (4.43) has to be modified to take into account the biased initial conditions selected by the first stage:

$$\mathbb{E}[G(X_{t_i})] \simeq \frac{\bar{w}_i}{N} \sum_{n=1}^{N} \frac{G(\hat{X}_i^n)}{e^{V_i(\hat{X}_i^n)}} \times \frac{\bar{w}_{T'}}{e^{V_{T'}(\hat{Y}_{T'}^n)}} \tag{4.58}$$

where $\bar{w}_{T'}$ is the mean weight computed by the first stage at its last resampling time and $e^{V_{T'}(\hat{Y}_{T'}^n)}$ is the exponential of the splitting function computed by the first stage at its last resampling time for member n. This is equivalent to giving an initial weight to each member in the short simulation.

Results of this second strategy are depicted in Figure 4.12. Comparing with Figure 4.11, the two-stages algorithm is able to estimate precisely much lower probabilities: there is a gain of almost 4 orders of magnitude. One has to note that the two-stage algorithm is not simply two DMC algorithms with the same score function $\theta = F$ applied one after the other, which would not allow to give such low probability (not shown). By targeting both X^1 and X^2, the score function F would indeed not be specific enough. Coupled with resampling times at the first stage too large for the process X^1, the PDF $\rho_{(X^1,X^2)}$ would come back to its unbiased stationary PDF between each resampling times. In the contrary, the two-stage algorithm targets first the slow variable X^2 with the adapted score function $\theta' = F' : (X_t^1, X_t^2) \mapsto X_t^2$, forgetting the fast variable X_t^1 and selecting 'optimal' initial conditions for the second stage of the algorithm targeting specifically the X_t^1 variable. This strategy could be applied broadly in complex geophysical systems with phenomena of different time scales resulting of extremes of an observable of interest.

4.5.7 Deterministic systems

In the case of stochastic systems, splitting methods functions naturally because once a member is cloned, the resulting members naturally diverge one from another because of the different realization of the noise applied on the system. This is however not the case for deterministic systems for which a perturbation has to be added after the cloning step. The issue when doing so is that there is also a trade-off on the amplitude of the noise to be added: if the noise is too small then the members will not rapidly differentiate one from another, while if it is too high it will push the trajectories out of the attractor and the full trajectory will not be physical anymore.

For high-dimensional real systems such as climate models, there is usually no difference as uncertainties arising from sub-grid processes are high enough so that adding a random noise to separate members should keep each member physically plausible. This may however not always be the case for simpler systems. Wouters and Bouchet (2016) proposed to apply the noise a few time steps before the resampling times so that when the rebranching is done at the resampling times the trajectory has converged to the attractor. They also mention that this does not affect the results when compared with a long brute force Monte Carlo simulation.

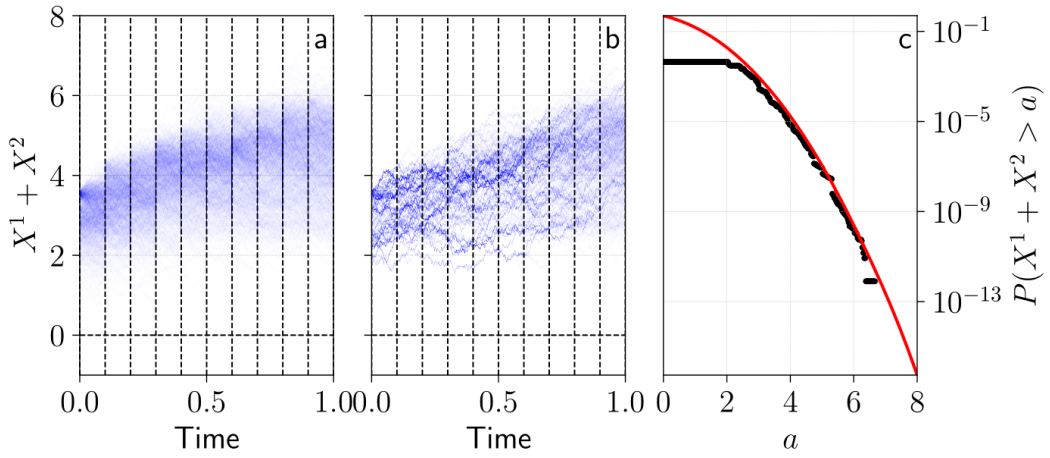


Figure 4.12: **Reaching extreme values of an observable with two Orstein-uhlenbeck processes with preliminary selection.** Application of the DMC algorithm to the system 4.56 with 10 homogeneous resampling times over a total integration time $T = 1$, with a splitting strength $C = 3.5$ and $N = 1000$ trajectories. The score function is F itself. For initial conditions, I take $X_0^1 = 0$ and X_0^2 selected by the first stage. (a) Evolution of the observable $F(X_t^1, X_t^2) = X_t^1 + X_t^2$ for all members between each resampling times. Vertical dashed lines represents the resampling times. (b) Evolution of the final members selected by the algorithm. (c) Probabilities $\mathbb{P}[F \geq a]$ for $a \in [0, 8]$. The exact probabilities are depicted in red and the black dots correspond to the probabilities estimated by the algorithm.

Summary

Context

The study of rare and extreme events is hindered by their under-sampling in real world data and in model simulations. One of the way to overcome this issue is to use so-called rare events algorithms, which are general methods aiming to sample extremes efficiently in dynamical systems.

Goals

In this chapter I present a review of these methods, with a particular focus on methods that have and can be applied in geophysical systems. I first present a review of the application of rare events algorithms for geophysical systems. This review then presents key mathematical concepts to frame the simulation of the paths leading to extreme events in a more general, physical and mathematical context. I explain in particular the concept of instanton, or most probable path leading to an extreme event. I then present the algorithms employed to study geophysical systems with a particular emphasis on so-called splitting methods for which I give the details of the algorithms employed. I finally present the challenges associated to their application in complex climate models and propose some solutions to alleviate these challenges.

Chapter 5

Investigating the typicality of the dynamics leading to extreme temperatures in the IPSL-CM6A-LR model

5.1 Introduction

In this chapter[1] I investigate the idea of the typical dynamics — or instanton — leading to extreme events introduced in chapter 4 section 4.3.3: the dynamics of extremes tend to concentrate around a most probable path. The rationale for this investigation is twofold. First, there is a practical aspect: as shown in chapter 4 section 4.4.4 on the simple Orstein-Uhlenbeck process, it is natural to impose that the dynamics sampled by rare events algorithms is the most probable dynamics associated with the extreme event one seeks to sample. Investigating this typical dynamics leading to extreme events is therefore a prerequisite to the application of rare events algorithms.

Second, I want to propose here the idea that the concept of typical dynamics is the fundamental concept to understand the dynamics leading to extreme events in a *climate* context. A lot of extreme events have been investigated in the *meteorological* context, cf. section 1.3 for examples. However, how the dynamics associated with an extreme event should be conceptualized in a climate context? Up to now, extremes have been mainly treated statistically in a climate context (e.g. chapters 2 and 3). One may be interested in the probability that an observable F — near-surface temperature for example — be above a certain level a:

$$\mathbb{P}_\mu[F \geq a] := \int_\Omega \mathbb{1}(F(X) \geq a)\mu(dX), \tag{5.1}$$

where X is the state vector of the system and μ the appropriate probability

[1] I am deeply indebted to Valerio Lucarini and Vera Melinda Galfi for the ideas developed in this chapter. I warmly thank them for our discussions on this subject. I also thank Tobias Grafke for discussing these results with me.

distribution. I propose that the dynamical equivalent of equation 5.1 is the typical dynamics associated with events reaching level a for observable F:

$$t \mapsto \mathbb{E}_\mu[X_t \mid F(X_0) \geq a]. \tag{5.2}$$

In this definition, the typical dynamics is a function that depends on a centered time so that the extreme event happens at $t = 0$. The output of this function is *a trajectory in the full phase space of the system* $t \mapsto X_t$, which corresponds to the most probable path followed by events that have reached level a. In other words, the typical dynamics is the climatological dynamical concept associated to extreme events (which is itself a static property of points in the phase space as defined in chapter 1 section 1.2.2).

The definition proposed in equation (5.2) is actually straightforward for a climate scientist: it is simply the definition of lagged composite maps for different events that have reached the level a, and therefore corresponds to a common practice in climate science. The added value of this definition is to formalize the statistical object that is estimated through composite maps: the most probable path leading to an extreme. It should be emphasized that this definition allows to conceptualize the effect of, for example, climate change on the dynamics leading to extreme events: one would typically be interested in how the typical dynamics changes under the influence of an external forcing.

5.2 Article published in *Climate Dynamics*

In this chapter I investigate the concept of typical dynamics in a 2000-y control run of the IPSL climate model under pre-industrial stationary forcings. This work was very much inspired by the recent paper of Lucarini et al. (2023) who looked at the typicality of the dynamics leading to extreme events similar to the 2021 heatwave event in Western Canada (Philip et al. 2021). If the typical dynamics concept is relevant, one should expect that when the threshold for defining extremes increases, the typicality also increases: extreme events tend to look more and more alike. This the idea behind the so-called *perfect storm*: all extreme event ingredients align perfectly to form a *very* extreme event. As a consequence, all events that have reached very extreme levels of an observable have had these same ingredients and are therefore very close. Here I investigate this prediction by looking at the dynamics of independent events that reach extreme temperatures at different locations in Europe in the 2000-y pre-industrial control run of the IPSL-CM6A-LR climate model.

Climate Dynamics
https://doi.org/10.1007/s00382-023-06967-5

ORIGINAL ARTICLE

Investigating the typicality of the dynamics leading to extreme temperatures in the IPSL-CM6A-LR model

Robin Noyelle[1] · Pascal Yiou[1] · Davide Faranda[1,2,3]

Received: 23 March 2023 / Accepted: 17 September 2023

Abstract
Determining the underlying mechanisms leading to extreme events in dynamical systems is a challenging task. Under mild hypotheses, large deviations theory predicts that, as one increases the threshold defining an extreme, dynamical trajectories which reach the extreme look more and more like one another: they converge towards a typical, i.e. most probable, trajectory called the instanton. In this paper, we use a 2000-year simulation of the IPSL-CM6A-LR model under a stationary pre-industrial climate to test this prediction on the case of hot extremes. We investigate whether the physical mechanisms leading to extreme temperatures at four locations in Europe are more similar with increasing extreme temperatures. Our results show that most physical variables exhibit the expected convergence towards a most probable trajectory, with some geographical and temporal variations. In particular, we observe the presence of a cut-off low in some trajectories, which suggests the existence of multiple pathways leading to extreme temperatures. These findings confirm the relevance of instanton dynamics to understand the physical mechanisms driving extreme events in climate models.

Keywords Extreme events · Large deviations theory · Instanton · Typicality · Atmospheric dynamics

1 Introduction

Extreme weather events can have tremendous impacts on societies and ecosystems (Pörtner et al. 2022). Among them, heatwaves have been the focus of extensive attention due to their increasing frequency with global warming (Seneviratne et al. 2021). Their impacts include adverse health effects, increased power consumption, infrastructure damages, forest fires, droughts and crop failures (Koppe et al. 2004; Zuo et al. 2015; Yaghmaei 2020). In the last decade, extensive research has been conducted to better understand the dynamics leading to heatwaves (Perkins 2015; Horton et al. 2016; Domeisen et al. 2022a). The general scenario combines specific atmospheric synoptic conditions and anomalously low soil moisture, which can interact to further increase the intensity of the event.

Heatwaves in the midlatitudes are usually associated with a slow moving, sometimes called quasi-stationary, high-amplitude Rossby wave (Petoukhov et al. 2013). This structure is often embedded in a hemispheric pattern of wave patterns 5–7, which can trigger extreme heat and rainfalls simultaneously at different places (Coumou et al. 2014; Kornhuber et al. 2020; Di Capua et al. 2021). The mechanisms and causes of the amplification of such wave patterns are still discussed, especially their dependence on climate change (Screen and Simmonds 2014; Petoukhov et al. 2016; Kornhuber et al. 2017; Mann et al. 2017, 2018; Kornhuber and Tamarin-Brodsky 2021). Over the heatwave region, an anticyclone builds up—a situation called 'blocked'—at mid- and upper-level troposphere in conjunction with a change in the jet stream from its climatological path towards a large poleward meander. The anticyclone sustains the poleward advection of warm air along its western flank, adiabatic warming by subsidence and clear skies at its center. Those

✉ Robin Noyelle
robin.noyelle@lsce.ipsl.fr

Pascal Yiou
pascal.yiou@lsce.ipsl.fr

Davide Faranda
davide.faranda@lsce.ipsl.fr

[1] Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay and IPSL, 91191 Gif-sur-Yvette, France

[2] London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK

[3] Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, PSL Research University and IPSL, Paris, France

Published online: 09 October 2023

conditions favor warming by short-wave insolation, especially at the peak of the seasonal cycle. Close to the ground, a positive feedback loop is initiated as anomalously dry soils favor the partition of incoming solar energy into sensible rather than latent heat, which enhances surface evaporation and may reinforce the anticyclonic structure (Hirschi et al. 2011; Miralles et al. 2012, 2014; Rasmijn et al. 2018; Dirmeyer et al. 2021).

These mechanisms all played a role in the record-breaking heatwaves of 2003 in western Europe (García-Herrera et al. 2010) and 2010 in Russia (Dole et al. 2011; Otto et al. 2012; Trenberth and Fasullo 2012; Di Capua et al. 2021). The same mechanisms are also present in the exceptionally intense heatwave that occurred in Western North America in June 2021. Previous records of temperature were broken by up to 5 °C (Philip et al. 2021) making it one of the most intense heatwaves ever recorded (Thompson et al. 2022). Due to its exceptional nature, this event triggered extensive research. Dry soils (Zhang et al. 2021, 2022) likely combined with an intense omega blocking anticyclone resulting from a wave breaking event, associated with southern excursion of the polar vortex (Overland 2021; Neal et al. 2022), and interacting with an atmospheric wave emanating from the tropical Pacific (Bartusek et al. 2022). Some authors also suggested the role of latent heat release through moisture advection by an unusual atmospheric river through the North Pacific (Qian et al. 2022; Lin et al. 2022; Mo et al. 2022). Lucarini et al. (2023) however showed using a long simulation in a pre-industrial climate that this event was *typical* with respect to other intense events simulated by the model at this location.

Extreme Value Theory (EVT) has been used to determine statistical models for maxima (or minima) of climate variables of interest (typically temperature or precipitation) (Ghil et al. 2011). EVT is based on a convergence principle of the probability distribution of maxima or peaks-over-threshold (Coles et al. 2001). It allows to compute return values corresponding to very large return periods (i.e. longer than the period of observations), even in non stationary contexts (Cheng et al. 2014). In practice, EVT has been an efficient framework to estimate statistical variations on short lived extremes (e.g. the highest daily temperature). In principle multi-variate EVT provides a framework to investigate events that combine several variables (e.g. temperature, precipitation and the atmospheric circulation) (Tawn 1990), but such analyses are not designed to investigate long lasting events, for which the time persistence is a key factor in the extreme (although using the so-called extremal index partially alleviates this issue, e.g. Moloney et al. 2019).

Large Deviations Theory (LDT) is however the key statistical framework employed in statistical physics (Touchette 2009) and has begun to gain momentum in climate sciences (Ragone et al. 2018; Galfi et al. 2021). In contrast to EVT, it provides asymptotic laws at the leading exponential order for extremes of sums of random variables. One of its applications is the path integral formalism in stochastic dynamical system (Martin et al. 1973; Janssen 1976; Dominicis 1976; Freidlin and Wentzell 1987) which uses the theory of large deviations to describe the dynamics leading to extremes as an optimal realisation of the noise. It predicts that, under mild hypotheses, the dynamics leading to extremes of any observable concentrates around a single most probable trajectory, usually called the *instanton*, as the noise amplitude decreases (Chetrite and Touchette 2015; Dematteis et al. 2019a; Grafke and Vanden-Eijnden 2019). Recent applications of instanton theory in complex systems have proved to give valuable insights in the typical dynamics leading to extremes (Bouchet et al. 2019; Finkel et al. 2021; Dubrulle et al. 2022; Fuchs et al. 2022). Dematteis et al. (2019b) presented an extension of this framework for extremes in systems with a general stochastic component. Lucarini et al. (2023) showed that this leads to a concentration of trajectories leading to extremes around a typical one for non-stochastic dynamical system with a physical invariant measure.

In this paper, we investigate the typicality of the dynamics leading to extremes in a long simulation of a climate model. In particular, we address the question whether there is a concentration of trajectories leading to extreme 2 m air temperature events.

As a case study, we take the 2000 years pre-industrial control run of the IPSL-CM6A-LR Earth System Model (ESM) (Boucher et al. 2020). We examine the characteristics of trajectories leading to heatwaves as their duration and intensity vary. We evaluate how typical the heatwaves are by examining the key dynamic factors that contribute to their formation. Lastly, we study how the dynamics change when the location at which the heatwaves are observed is altered.

This paper is organized as follows: in Sect. 2, we introduce the methodology employed to isolate the dynamics leading to extremes and the normalized variance metrics used to measure convergence. Section 3 presents the results of the analysis, focusing first on one observable and then on three observables at different locations. In Sect. 4 we discuss the LDT framework in a non-stochastic dynamical system. Finally, the conclusions drawn from our analysis are presented in Sect. 5.

2 Data and methods

2.1 Climate model data

We use the output of the pre-industrial control run of the IPSL-CM6A-LR model (Boucher et al. 2020) for the Coupled Model Intercomparison Project phase 6 (CMIP6) (Eyring et al. 2016). This simulation is 2000 years long and

represents a stationary climate with a CO_2 concentration corresponding to pre-industrial levels. The model has a horizontal atmospheric resolution of 2.5° in longitude and 1.3° in latitude. During the 2000 years, the global mean 2-meter air temperature of the Earth slightly drifts by 0.25K. We assume that this drift can be neglected when it comes to studying extremes in the midlatitudes, because we are interested in deviations that can exceed several K. We use the sea level pressure (SLP), upper level soil moisture (SM), air temperature at 2-m (T2M), air temperature at 850 hPa (T850), geopotential height at 500 hPa (Z500) and the meridional component of the wind at 250 hPa (V250) fields. These variables are sampled with a daily frequency. Due to a technical issue, approximately 1/4 of the data for the Z500 variable are missing. In the following, we therefore present results for Z500 only for the dates available in our data set. We have checked that when restricting to the period when we have the Z500 variable available for the other variables do not affect the results. We therefore present the results for those other variables over the whole 2000 years.

2.2 Extracting the typical dynamics

We define some mathematical notions first. Let us consider a chaotic dynamical system evolving continuously in time with $x(t) \in \mathbb{R}^d$ its state vector at time t, where d is the number of dimensions needed to describe the system. We assume that x evolves according to the following ordinary differential equation

$$\frac{dx}{dt} = b(x) \tag{1}$$

where $b : x \mapsto b(x) \in \mathbb{R}^d$ defines the dynamics of the system. In the following we always assume that the transients have died out and all trajectories considered are supposed to belong to the attractor of the system. We assume that there is a unique physical invariant measure μ on this attractor. We are interested in the statistics of observables computed on the attractor. Observables are smooth functions of phase-space variables $F : \mathbb{R}^d \to \mathbb{R}$ which have a physical interest.

Our observables will be of the following type:

$$F_r(x_t) = \frac{1}{r}\int_{-r/2}^{r/2} f(x_{t+t'})dt' \tag{2}$$

where f is also an observable and $r \in \mathbb{R}$ is the size of the rolling mean window. In the following, f is the daily mean temperature over a specific area. For $q \in \mathbb{R}$, we define the set Ω_q as the set of points x on the attractor such that $F_r(x) \in [q, q + dq]$. In other words, we are interested in the states of the system for which the value of the observable F_r is close to q, where q will correspond to extreme quantiles of the distribution of F_r. Here we investigate the typical state—also called the instanton—$\hat{x}_q := \mathbb{E}[x \mid x \in \Omega_q]$ conditional on $F_r(x) = q$. To do so, we employ an instanton filtering procedure adapted from Grafke et al. (2013). It consists of averaging independent events x taken from a long simulation of the dynamical system for which the value of the observable $F_r(x)$ is close to an extreme level. To assess the convergence towards the typical state, we compare the conditional variance $\mathbb{V}[x \mid x \in \Omega_q]$ to the unconditional variance $\mathbb{V}[x]$.

Contrary to Lucarini et al. (2023), we do not consider events above a certain level $\mathbb{E}[x \mid F_r(x) \geq q]$. The rationale of this choice is twofold. First we want to investigate the convergence properties for different values of q for a fixed number of independent events in each case. Second, if a large deviations principle applies for sufficiently extreme values of q, then the events for which $F_r(x) \sim q$ are exponentially favored compared to the ones with $F_r(x) > q$, therefore we should not expect a large difference between $\mathbb{E}[x \mid F_r(x) \simeq q]$ and $\mathbb{E}[x \mid F_r(x) \geq q]$. This is however a limit result and we may see some differences with a finite data set. Morevoer, we cannot a priori rule out the possibility of a non linear dependance of the dynamics $\mathbb{E}[x \mid F_r(x) = q]$ with respect to q when reaching extremes. Therefore, the dynamical mechanisms needed to reach high temperatures may be different than the ones to reach *very* high temperatures.

2.3 Experimental set-up

In the following, we consider four observables derived from T2M: T2M at three grid points situated in Southern (38–39°N, 5.25–3.75°W), Western (49–50°N, 1.25–3.75°E) and Northern Europe (59–60°N, 13.75–16.25°E) and T2M averaged over a region extending in Western and Central Europe (46–53.5°N, 0–25°E). These observables are named respectively S, W, N and WCE. We want to investigate the highest values reached by these observables, therefore we restrict the analysis to the three months of the meteorological summer: June, July and August (JJA). Except specified explicitly, we do not consider detrended or deseasonalized variables. To investigate how the dynamics can change for longer events, we consider the extremes of the observable after applying a rolling mean window of $r = 1$, 5 and 15 days. The size of these windows were chosen to investigate both short and long lasting heatwaves events.

We consider the time series of one given observable F_r among S, W, N and WCE regions during the summer months for a given rolling mean window r. For a quantile q_α of a given order α of the empirical distribution of F_r, we select the pool of the $n = 50$ independent events x_i for which the values of their observable $F_r(x_i)$ are the closest to the value of the quantile q_α. In other words, we find the dates of the n-nearest neighbors of the quantile q_α of the observable. These events are searched over any of the days in the JJA

months. To ensure that these events are independent one from another, we impose that for a nearest neighbor to be chosen, it must be separated by more than 15 days from any nearest neighbor already present in the pool. The choice of this timescale was made with regards to the typical chaotic timescale of the atmosphere (around 10 days). Our procedure is equivalent to defining the set Ω_q as:

$$\Omega_{q_\alpha} := \Omega_\alpha = \{x \mid F_r(x) \in [q_\alpha - \eta, q_\alpha + \eta]\} \tag{3}$$

for η as small as possible to ensure that the number of elements in Ω_α is $\#\Omega_\alpha = n$. We consider the quantiles of order $\alpha = 0.75$, 0.95, 0.99 and 0.999 of the empirical distributions of observables S, W, N and WCE. These quantiles are arbitrary and are chosen to ensure homogeneity between the different observables for which the extreme values may be very different.

Figure 1 presents the histograms of the empirical distributions of the four observables (rows) for the different rolling mean windows (columns). The vertical lines in color present the quantiles for the different orders considered. The horizontal lines of the same colors present the spread η of the observables for the points found with our procedure. For the smallest orders (0.75, 0.95 and 0.99), the spread is barely visible, which means that the values of the observable for points in Ω_α are very close to the value of the quantile. For the order 0.999, the spread is bigger but still small compared to the standard deviation of the full empirical distribution.

For each value of the α-th order quantile, a given observable F_r and a rolling mean window r, we therefore consider fields (SLP, SM, T2M, T850, Z500 and V250) $\psi_{\alpha,r,F_r}(\phi,\theta,t,m)$ that have four arguments: the longitude ϕ and latitude θ, the time t—where the time is expressed relative to the day t when the observable is such that $F_r(x_t) \simeq q_\alpha$—, and the number m among the points in Ω_α ($m = 1, 2, \ldots, n$). In the following we skip the α, r, F_r indices for simplicity, but note that the quantities introduced are always relative to a given value of this triplet. For any field ψ, we use the term composite to denote the average $\widehat{A}^r_\psi$ of ψ over the points in Ω_α and the rolling mean window r:

$$\widehat{A}^r_\psi(\phi,\theta,t) = \frac{1}{n}\sum_{m=1}^{n}\frac{1}{r}\sum_{t=-r/2}^{r/2}\psi(\phi,\theta,t,m). \tag{4}$$

2.4 Reduction of variance for extremes

To measure the clustering of points in Ω_α, we compare the conditional variance $\mathbb{V}[x \mid x \in \Omega_q]$ to the unconditional variance $\mathbb{V}[x]$ over the summer:

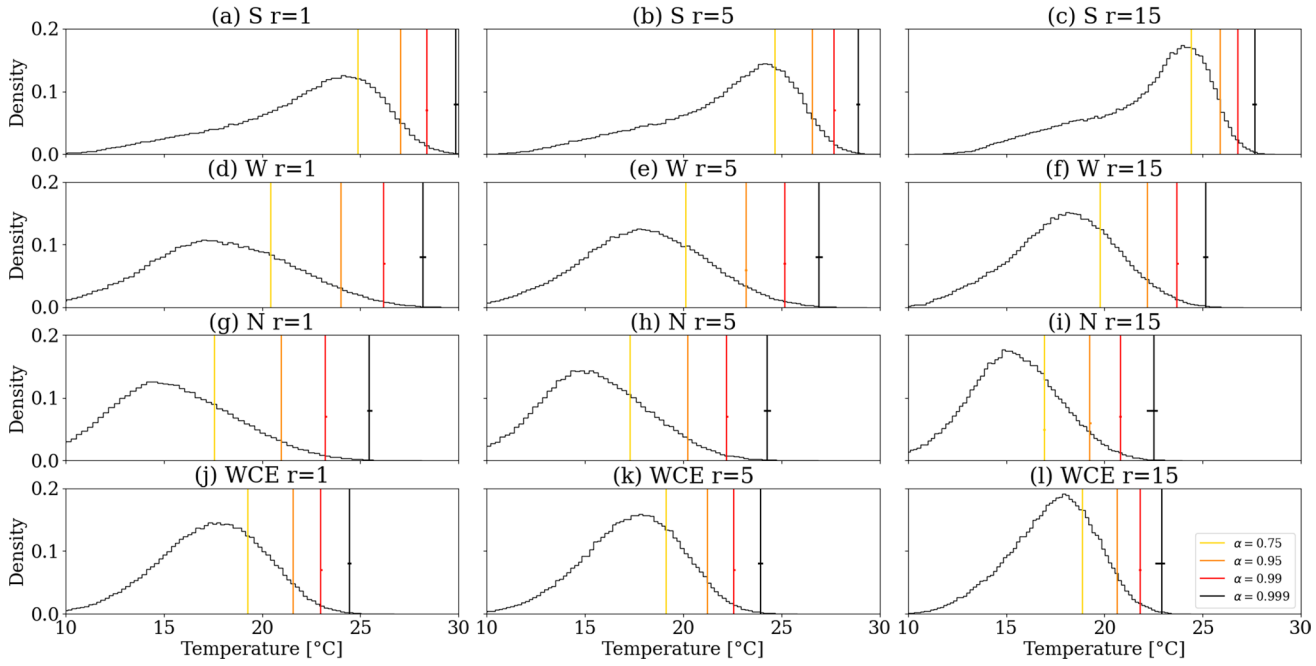


Fig. 1 Histograms of the observables and associated quantiles. Rows: observables derived from the T2M variable at three grid points situated in Southern Europe (S: 38–39°N, 5.25–3.75°W), Western Europe (W: 49–50°N, 1.25–3.75°E) and Northern Europe (N: 59–60°N, 13.75–16.25°E) and over one grid box extending in the Western and Central Europe (WCE, 46–53.5°N, 0–25°E). Columns: rolling mean windows for the computation of the observable of $r = 1$, 5 and 15 days. The vertical bars show the value of the quantiles of order α of the time series for $\alpha = 0.75$, 0.95, 0.99 and 0.999. The horizontal lines show the spread of the observables for the points in Ω_α

$$\widehat{V}^r_\psi(\phi,\theta,t) = \mathbb{V}_m\left[\frac{1}{r}\sum_{t=-r/2}^{r/2}\psi(\phi,\theta,t,m)\right] / \mathbb{V}_{\psi^r,clim}(\phi,\theta) \quad (5)$$

where $\mathbb{V}_m$ is the variance over the points in Ω_α and $\mathbb{V}_{\psi^r,clim}(\phi,\theta)$ is the climatological variance over the whole summer (JJA) of the field ψ after applying a rolling average with window r. The variance in Eq. 5 is normalized in the sense that it is compared to the variance observed over the whole summer. In the following, it is thus expressed in percentage. The lower the value of this percentage, the smaller the variance between the points in Ω_α. If we were to select n days randomly over the full data set, we could expect to find a variance close to $\mathbb{V}_{\psi^r,clim}$ (after the rolling average). Therefore, the closer $\widehat{V}^r_\psi$ is to one (or 100%), the less specific is the dynamics of the averaged field $\widehat{A}^r_\psi$ at (ϕ,θ,t). The closer it is to zero, the more concentrated are the n points in Ω_α. To test whether $\widehat{V}^r_\psi$ is significantly different than 1, we assume that the climatological variance is well estimated and we use a χ^2 test at the 5% level with the null hypothesis being that the unconditional variance is not significantly higher than the conditional variance. With $n = 50$, this condition is equivalent to $\widehat{V}^r_\psi < 0.7$.

We will note the normalized variance $\widehat{V}^{r=1}_\psi(\phi,\theta,t)$ as $\widetilde{V}_\psi(\phi,\theta,t)$. We emphasize that in general $\widehat{V}^r_\psi(\phi,\theta,t)$ is not the rolling mean of $\widetilde{V}_\psi(\phi,\theta,t)$ over a window r. In order to have simple diagnosing metrics, we also average the $\widetilde{V}_\psi$ variance spatially:

$$\langle\widetilde{V}_\psi(t)\rangle = \sum_{\phi,\theta}\widetilde{V}_\psi(\phi,\theta,t)\cos(\theta) \,/\, \sum_{\phi,\theta}\cos(\theta) \quad (6)$$

over either the Northern Hemisphere ($\theta \in [22.5°\text{N}, 90°\text{N}]$ and $\phi \in [0°\text{E}, 360°\text{E}]$) or the Euro-Atlantic sector ($\theta \in [22.5°\text{N}, 70°\text{N}]$ and $\phi \in [80°\text{W}, 50°\text{E}]$). For simplicity, we drop the index ψ in the following where it is not ambiguous.

3 Results

3.1 Concentration of trajectories leading to extreme temperatures in Western Europe

In this section and the next one we present the results for the extremes of T2M at the grid point in Western Europe (observable W). Figure 2 shows the composite maps of anomalies of T2M and Z500 over the Euro-Atlantic sector for the $n = 50$ points for which the values of their observable is the closest to the quantile and over the rolling window $r = 5$ days for the four quantiles considered. The anomalous T2M values and their spatial extension increase as the α -th order quantile increases. As α increases, the deviation in Z500 also increases, reaching a maximum of 160 m at the center of the high-pressure system located above North-Western Europe. There is also a warm anomaly in Eastern North-America and a cold anomaly in Western Russia. Figure 11 in annex shows the same results over the whole Northern Hemisphere. Anticyclonic anomalies are present all over the Hemisphere, with a distinct wave number 4 hemispheric pattern at mid-troposphere for the highest quantiles. These structures coincide with warm anomalies at the ground.

Figure 3 shows in contours the composite T2M field and in colors the normalized variance $\widehat{V}^r$. For all values of α,

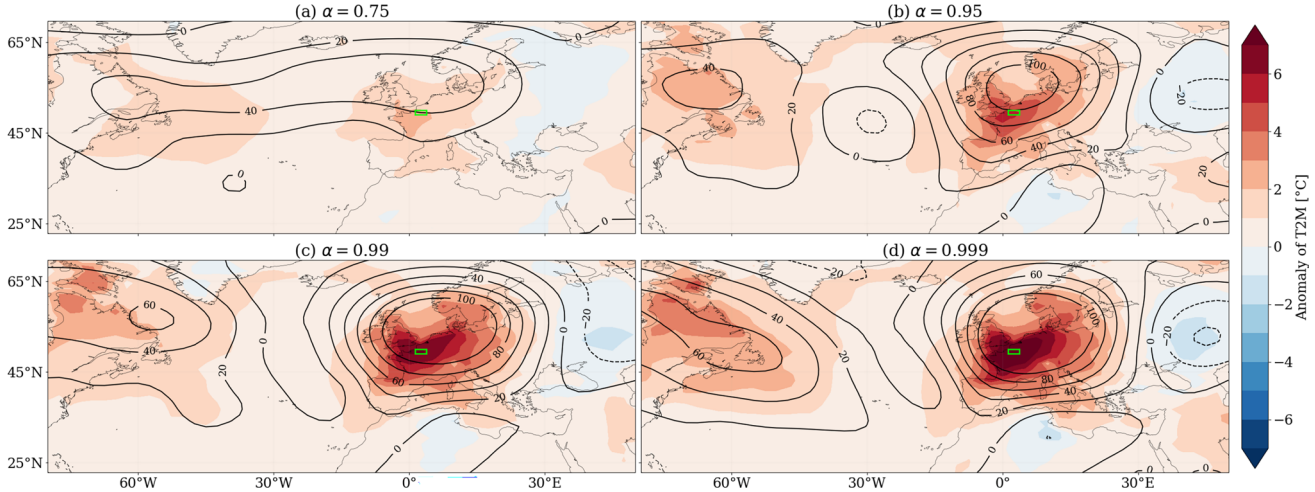


Fig. 2 Composite maps $\widehat{A}^r$ of anomalies of T2M and Z500 for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. Colors: anomaly of air temperature at 2 m (T2M (K)). Contours: anomaly of geopotential height at 500 hPa (Z500 (m)). Anomalies are computed with respect to the JJA average. The green box displays the location where the observable is computed

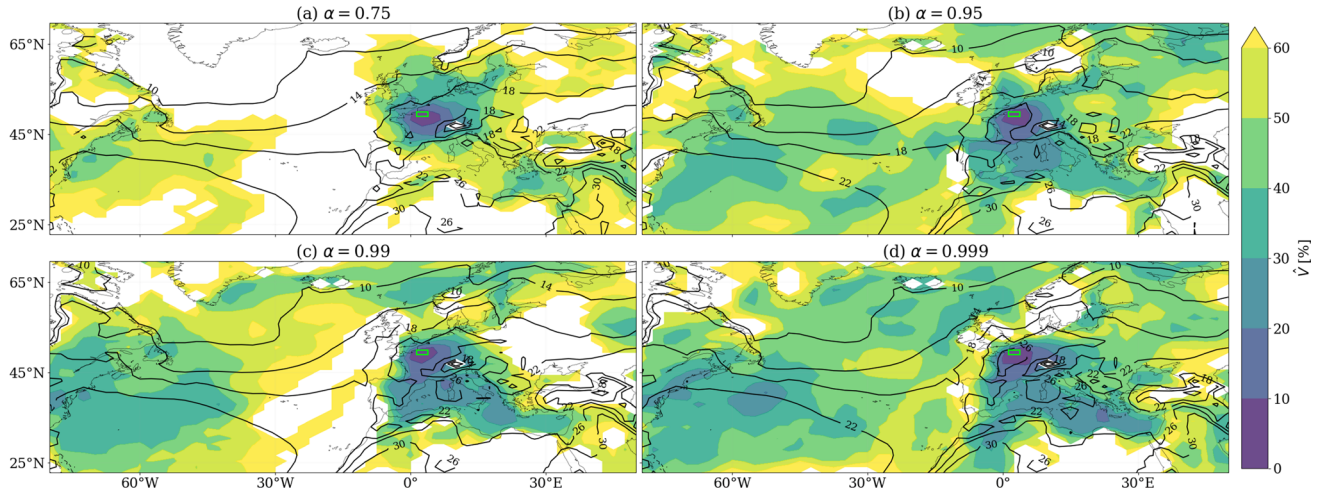


Fig. 3 Composite $\widehat{A}^r$ (contours, °C) and normalized variance $\widehat{V}^r$ (colors) of the T2M field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

the lowest values of the normalized variance are located in Western Europe, i.e. around the location where the observable is computed. When α increases, the normalized variance decreases: for example, the normalized variance exceeds 60% in most of the North Atlantic region for $\alpha = 0.75$, and it decreases to less than 40% for $\alpha = 0.999$. Figure 4 presents the same analysis for Z500. The extension of regions of highest variance also decreases when increasing the order α. The regions with the lowest variance is again centered at the location where the observable is computed. For $\alpha = 0.999$ for example, a large region of very low normalized variance ($\widehat{V}^r < 20\%$) embraces most of Western Europe. The decrease of variance is not uniform, with high variance remaining downstream of the anticyclonic region and a smaller localized region west of the Iberian peninsula. We come back to this latter structure below. We finally remark that on Fig. 4, for the highest quantiles the composite Z500 field shows that the large anticyclone over Western Europe is not centered just above the location of the observable, but is rather situated to its south-east.

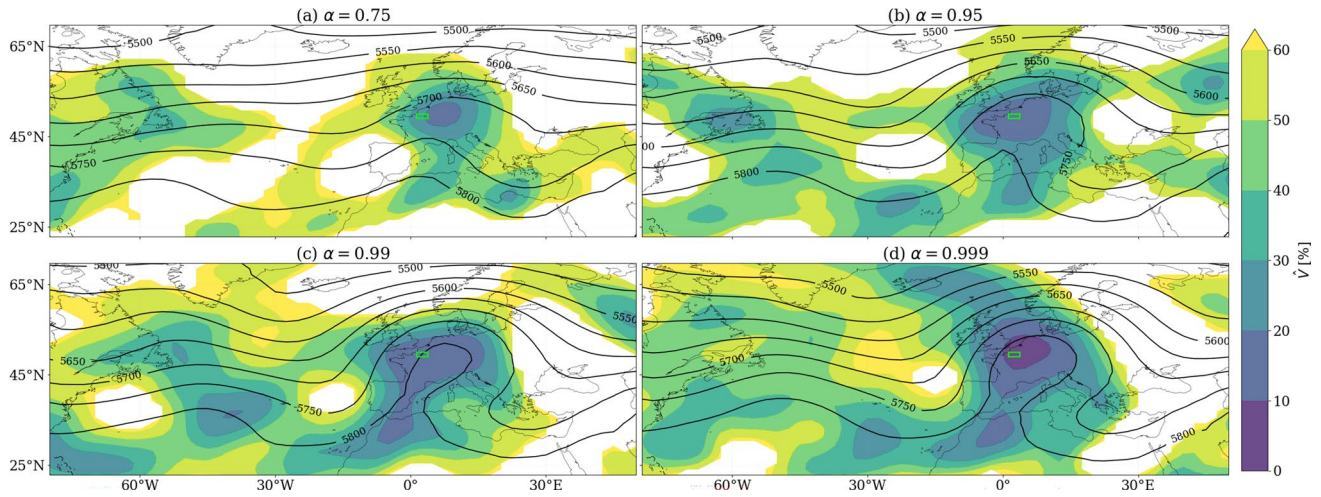


Fig. 4 Composite $\widehat{A}^r$ (contours, m) and normalized variance $\widehat{V}^r$ (colors) of the Z500 field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

Figures 12 and 13 in annex display the same analysis over the whole Northern Hemisphere. The reduction of variance is seen over remote regions of the atmosphere, for example with a region of low variance ($\widehat{V}^r < 20\%$) in the North-Eastern Pacific for $\alpha = 0.999$ for Z500 (Fig. 13 panel d). We emphasize that for all those maps, the number of points in Ω_α is always the same ($n = 50$ for T2M, $n \sim 30$ for Z500 due to the missing data). The outcome shown implies that the dynamics leading to extreme temperatures concentrates within a large geographical region.

The results for $r = 1$ day and $r = 15$ days for the composite anomaly maps of T2M and Z500 are similar to Fig. 2 and are thus not shown. As can be seen in Figs. 14 and 15 in annex for $r = 1$ day and in Figs. 16 and 17 in annex for $r = 15$ days, there is also a reduction of variance when α increases but the results depend on r. The decrease of variance is stronger for higher values of r. This suggests that longer extremes of temperatures tend to be more typical than shorter ones (Galfi and Lucarini 2021). We also note that the region of high variance situated on the South-West of the anticyclone is not present for $r = 15$ (Fig. 17). It is not clear however whether this is the result of a different dynamics for longer extremes and we investigate this point further below.

The analyses presented so far support the typicality interpretation of extreme events: the higher the value of the quantile q_α, the stronger the concentration of trajectories reaching this quantile around a single trajectory. To check whether our visual inspection is correct, we present in Fig. 5 the temporal evolution of the normalized variance $\langle \widetilde{V}(t) \rangle$ averaged over both the Euro-Atlantic sector (plain lines) and the entire Northern Hemisphere (dashed lines) from $t = -15$ to $t = +15$ days with respect to the event for all the variables and all the rolling windows r. The general picture drawn above is validated by the results presented in this figure: overall, the higher the value of the order α for the quantiles of the observable empirical distribution, the smaller the value of the normalized variance $\langle \widetilde{V}(t) \rangle$. This result holds when averaging over either only the Euro-Atlantic sector—where the observable is computed and the synoptic situation is the most relevant for the event—or the entire Northern Hemisphere. However, when we focus on the details of the different panels, we see that this behavior is less clear for some variables. The convergence is clear for T2M (panels a–c) and T850 (panels (m–o)) for which the normalized variance $\langle \widetilde{V}(t) \rangle$ is a decreasing function of α for almost all values of t. For dynamical variables such as Z500, we see that the convergence is stronger for higher values of the rolling window (compare for example panel f and panel d). This could be explained by the fact that a long extreme needs a persistent anticylonic circulation, which is more likely to be typical than for a short extreme. For the SM and SLP variables the situation is the most blurred. Higher values of α broadly correspond to smaller values of the normalized variance $\langle \widetilde{V}(t) \rangle$, but the ordering of the quantiles changes with both r and t.

We also note that the dates at which these extremes are reached are less dispersed around the peak of the T2M seasonality when α increases: the standard deviation in the calendar days of the points in Ω_α goes from 23.1 days for $\alpha = 0.75$ to 15.5 days for $\alpha = 0.999$.

In Fig. 4 there is a small isolated region of high variance situated west of the Iberian peninsula for all quantiles. We investigated this discrepancy by looking at individual events. It turns out that the synoptic dynamics at the mid-troposphere associated with some of the high temperature events are characterized by the presence of a cut-off low around this location. Table 1 in Appendix presents the percentage of such events. The presence of a cut-off low was investigated in a semi-objective way by looking at the existence of an isolated minimum of the stream function at 500 hPa located within [−30°S, +5°N] and [−40°W, +0°E] from the location where the observable W is computed and not embedded in the upper-level jet (250 hPa) (Muñoz et al. 2020). We however note that it is sometimes difficult to distinguish between a true cut-off low and a large meander of the jet caused by a deep, almost isolated, low pressure system over the Atlantic. The percentage of cut-off lows vary between 15 to 30%, with an average around 20%, depending on the quantiles and the size of the rolling mean window considered but without any clear trend (the differences may be due to the limited sample size). These cut-offs are not visible on the averaged maps (e.g. Fig. 4) not only because they represent only 20% of the events but also because their characteristic size is of the same order as the variance in the location of their center. Therefore even when considering only events with a cut-off low, they tend to be averaged out. The fact that even for very high quantiles there is a substantial amount of events with such a cut-off (around 20%) seems in contradiction with the unique path hypothesis. It therefore suggests that there may be at least a bi-modality in the typical dynamical paths to reach extremes for the W observable. We however note that if there are indeed two such paths, the convergence of trajectories that we showed above using all trajectories suggests that they are close to one another in the phase space.

3.2 Typical dynamics of extreme temperatures in Western Europe

In this section, we briefly present the key dynamical features associated to the instanton leading to extremes of observable W in the IPSL-CM6A-LR model. The concept of instanton is dynamic and it would be more rigorous to speak of the instanton as a function $t \mapsto \mathbb{E}[x_t \mid F(x_0) = q]$. For practical reasons, we present here only the composites $\widehat{A}^r$ and we refer

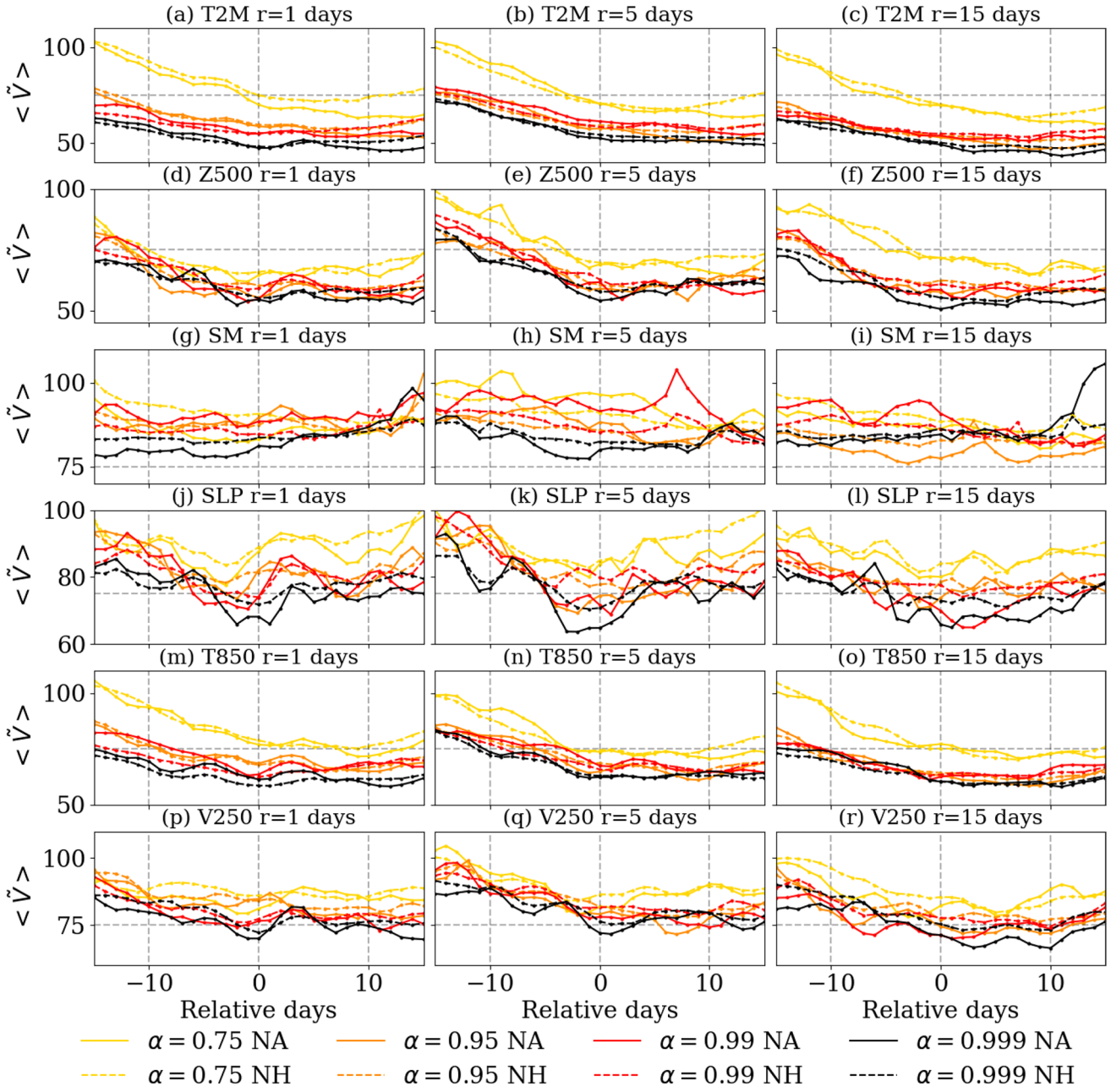


Fig. 5 Evolution of the normalized variance $\langle \widetilde{V}(t) \rangle$ averaged over the Euro-Atlantic sector (plain lines) and the Northern Hemisphere (dashed lines) for the different variables (W observable). The normalized variance is expressed in %. The colors show the α-th order quantile of the observable empirical distribution. The time is expressed relative to the day when the observable is such that $F_r(x_t) \simeq q_\alpha$. The gray horizontal dashed line shows the 75% level

the reader to supplementary material for the videos of the composites $t \mapsto \widehat{A}^{r=1}(t)$.

Figure 6 shows composites (colors) and zones of high variance ($\widehat{V}^r > 70\%$, hatches) for the other variables studied: anomaly of soil moisture (SM, panel a), anomaly of sea-level pressure (SLP, panel b), air temperature at 850 hPa (T850, panel c) and meridional wind speed at 250 hPa (V250, panel d). This figure is drawn for $r = 5$ days and $\alpha = 0.999$.

Panel a shows the large area of anomalously dry soils over most of Western and Central Europe. The small variance over this region demonstrate the key importance of local dry soils for reaching very high temperatures. The convergence is stronger to the east of the location of the observable, which agrees with the results of Zschenderlein et al. (2019). This pattern is similar to the one found by Faranda et al. (2022) using the SPEI index (Beguería et al. 2014). We also

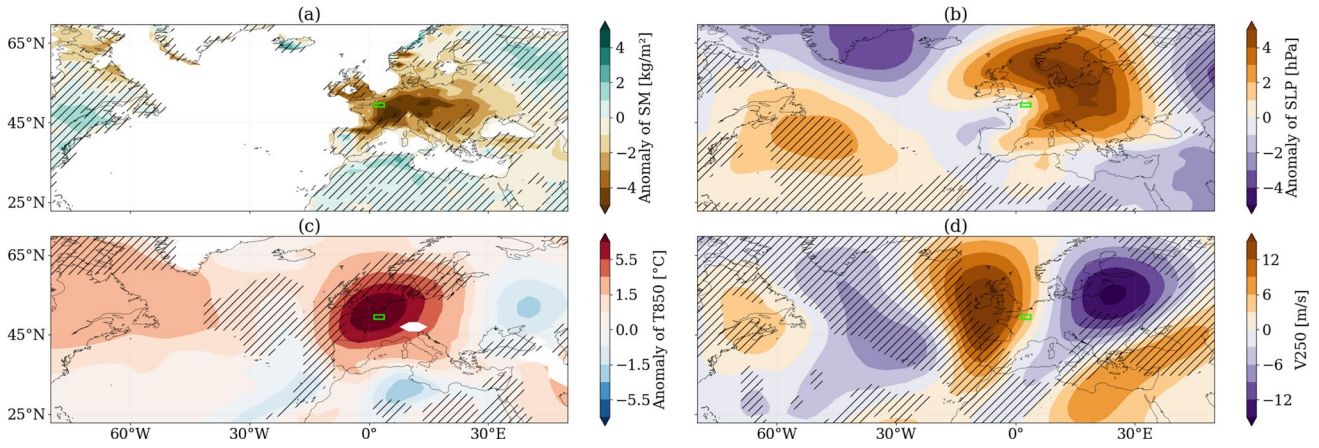


Fig. 6 Composite maps $\widehat{A}^r$ for SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the W observable's empirical distribution. The hatched areas correspond to $\widehat{V}^r > 70\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** soil moisture (SM), **b** sea-level pressure (SLP) and **c** temperature at 850 hPa (T850) with respect to their average over the summer (JJA), and **d** meridional wind speed at 250 hPa (V250). The green box displays the location where the observable is computed

note that when checking at $t = -15$d (i.e. 15 days before the event, not shown), the anomaly is still strong ($\simeq$4 kg/m^2) and the variance still small ($\widetilde{V} < 30\%$), which supports our preceding statement for the role of dry soils. Contrary to other fields (Fig. 6 panels b–d), the region of low variance is concentrated around the observable and do not extend over large regions.

Panel b shows a large positive anomaly of SLP northeast of the observable location and the low-level thermal low associated with the heatwave in Western Europe. The normalized variance is low over these regions. The positive anomaly situated above the Western Atlantic ocean may not be a relevant feature of the typical dynamics because the variance is high at this place, contrary to the anomalous low pressure over Greenland. Panel c shows the composite map of T850. The large intrusion of warm air from the south is a key feature of the dynamics. Even though the variance is low over mainland Western Europe, we detect a region of high variance extending from the west of the Iberian peninsula to Ireland and southern Norway and Sweden, i.e. to the western flank of the anticylonic structure (Fig. 4 panel d). This tongue of high variance is located at the highest gradient of T850. It is therefore difficult to decide between the two following explanations for this feature: either different dynamical mechanisms between the points in Ω_α (e.g. advection of warm air from the tropics) or a slightly displaced anticyclone which, combined with the strong gradients, would display such a strong variance tongue.

Finally, panel d shows the dynamics in the upper troposphere with the meridional wind speed at 250 hPa. The situation is characterized by a strongly meridional circulation west of the observable, which is consistent with the anticyclonic situation presented in Fig. 2. The regions of high variance are more extended than for the other variables. Some regions of high variance also coincide with regions with the highest gradients of V250, e.g. west of the positive anomaly centered on Ireland and above the Balkans and the Black Sea. The latter suggests the presence of an arm of the jet stream oriented to the north. We checked this explanation using the zonal wind speed at 250 hPa (not shown) and found that the synoptic situation over the event corresponds indeed to a splitting of the jet caused by the large blocked anticyclonic situation over Western Europe.

Figure 7 presents the same analysis over the entire Northern Hemisphere. For soil moisture (panel a), the only region outside the Euro-Atlantic sector where there are important anomalies and low variance is the South-Western coast of the USA. This region is also characterized by positive T2M temperature anomalies as can be seen in Fig. 11. Panel b shows a good agreement among points in Ω_α for the low-level low over Greenland and the Ural. We also notice a positive SLP anomaly over the North-Pacific, but this region is associated with high variance. Apart from the high variance tongue in Western Europe, and high variance around the Arctic region, panel c shows a good agreement between points in Ω_α for the T850 field, including over most of the Atlantic and Pacific oceans. Finally, panel d shows a wave pattern 6–7 extending over the entire Northern Hemisphere upper-troposphere. It should be noted, however, that the anomaly is stronger and the variance is smaller in the upstream area compared to the downstream area of the observable. Specifically, there is little structure visible above the Western Pacific region. This situation may be the result of either an hemispheric quasi-stationary pattern (Coumou et al. 2014; Kornhuber et al. 2020) or a transient Rossby wave packet (Fragkoulidis et al.

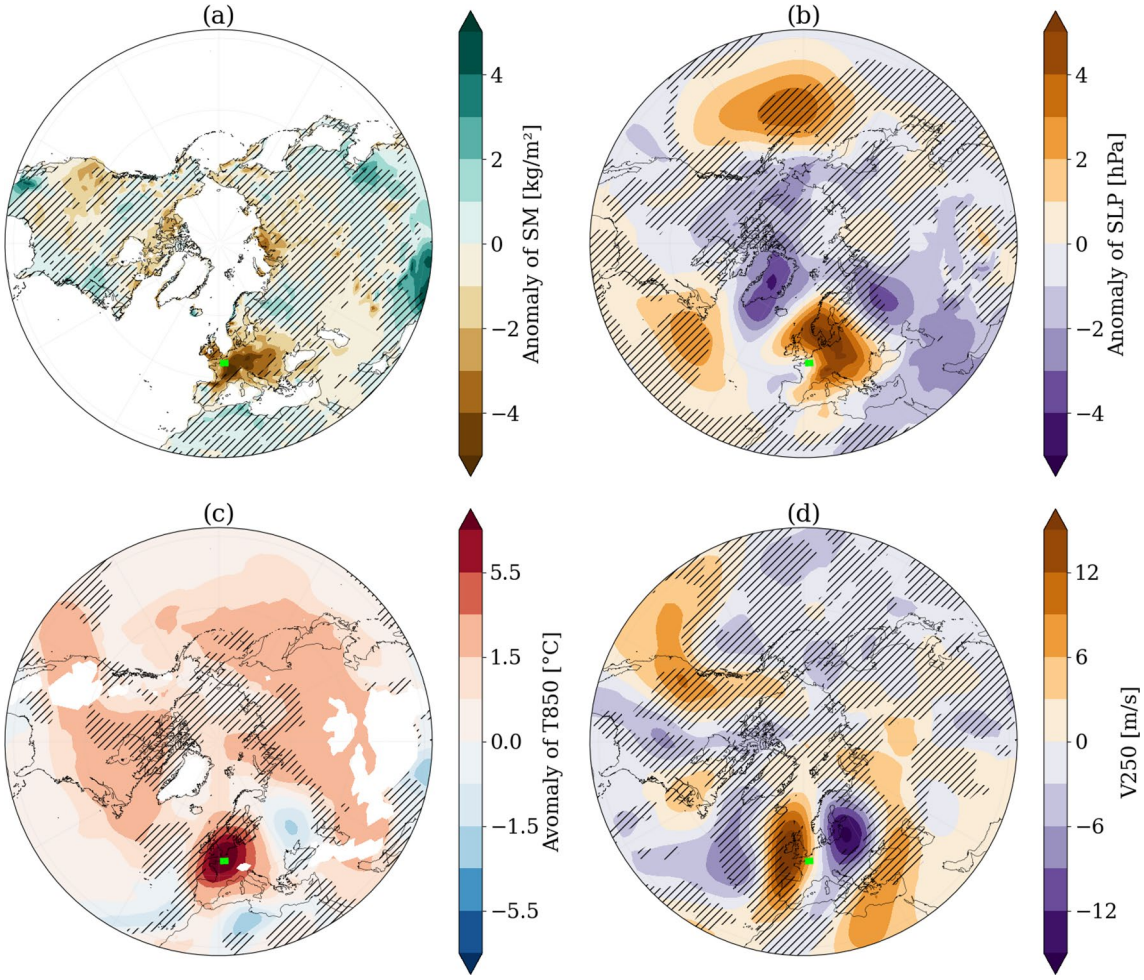


Fig. 7 Composite maps $\widehat{A}^r$ for SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the W observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 70\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** soil moisture (SM), **b** sea-level pressure (SLP) and **c** temperature at 850 hPa (T850) with respect to their average over the summer (JJA), and **d** meridional wind speed at 250 hPa (V250). The green box displays the location where the observable is computed

2018) depending on the points in Ω_α, which may explain the relatively higher variance associated with this field.

Up to this point we analyzed the typical dynamics $\mathbb{E}[x \mid F_r(x) = q]$ conditional on the value reached by an observable F_r. We demonstrated the concentration of the trajectories as the value q of the observable reaches extremes. We can now sketch the mechanisms associated with extreme temperatures at the grid point considered in Western Europe (observable W). The mechanisms summarized in the reviews of Perkins (2015), Horton et al. (2016) and Domeisen et al. (2022a) are present. To ensure very high temperatures, one needs dry soils, a large mid-troposphere anticyclone centered slightly to the south-east of the location of the extreme and an upper level Rossby wave train of 6–7 wave number. This situation ensures both the advection of warm air from the south at the west flank of the anticyclone, subsidence and associated adiabatic heating at the center of the anticyclone, and clear skies caused by the high-pressure system that allows for more radiative heating of the lowest layer of the atmosphere in conjunction with reduced water evaporation.

3.3 Results for the other locations

In this section we present the same analysis applied to the three other observables: T2M at two grid-points situated in the south (observable S) and north of Europe (observable N), and T2M averaged over a large area in Western and Central Europe (WCE observable). The results for the evolution of their normalized variance $\langle \widetilde{V}(t) \rangle$ is presented in Figs. 18, 19 and 20 respectively in annex. As above, the general picture of decreasing variance with increasing α is still valid but there are substantial inter-location variations. We note for example that the dynamical signal represented

by the variables Z500, SLP and V250 is much clearer for locations farther to the North (Fig. 18 vs Fig. 19 panels d–f for example). On the contrary, soil moisture plays a bigger role for the locations situated farther to the south (Fig. 18 vs Fig. 19 panels g–i). This suggests that extremes of temperature at places situated to the south are more "local" in the sense that they require less large scale organized circulation to be reached, which is consistent with the results of Sousa et al. (2018) who showed that a ridge situation better described the occurrence of heatwaves in southern Europe than the blocked situation as in northern Europe. This interpretation however needs to be validated at other locations and in particular may arise simply as a result of the fact that the Iberian peninsula is surrounded by water masses. We also note that these dynamical differences between lower and higher latitudes may be related to the skewness differences in the distribution of the summer temperature between observables (Fig. 1).

The observable over a large area (observable WCE) has a similar behavior to the W and N observables, with a stronger concentration of trajectories for dynamical variables than for soil moisture. We also note that the absolute values of the normalized variances for this observable are smaller than the one for the observable W which is situated around the same latitudes. It suggests that the typicality of the dynamics leading to anomalies of T2M over a large geographical area is stronger than for a localized observable. This seems reasonable in so far as it is less likely to have anomalies over a larger than a smaller area situated inside the larger one, hence the smaller number of synoptic conditions which can lead to an extreme for an extended observable.

Figure 8 presents the composite maps and the normalized variance maps for the S observable for the order $\alpha = 0.999$ and a rolling mean window of $r = 5$ days. The situation over the North-Hemisphere is presented in annex in Fig. 21. The synoptic situation is characterized by a large anticyclone centered south-east of the observable (panel b) and a positive SLP anomaly extending from Algeria to Northern France (panel d). The upper-level circulation displays a short wavelength Rossby wave (panel f, wave number 6–7). We however note that its amplitude is smaller than for the W observable. As previously, this dynamics leads to high T2M and T850 values and it has been anticipated by dry soils over most of Southern Europe and Northern Africa. The variance is lowest close to the location where the observable is computed. We note a large region of high variance downstream of the observable for V250 (panel f), but upstream for SLP and Z500 (panels b and d). As for the W observable, we see a high variance region at the west of the maximum gradients of T850 (panel e). The percentage of cut-offs associated with

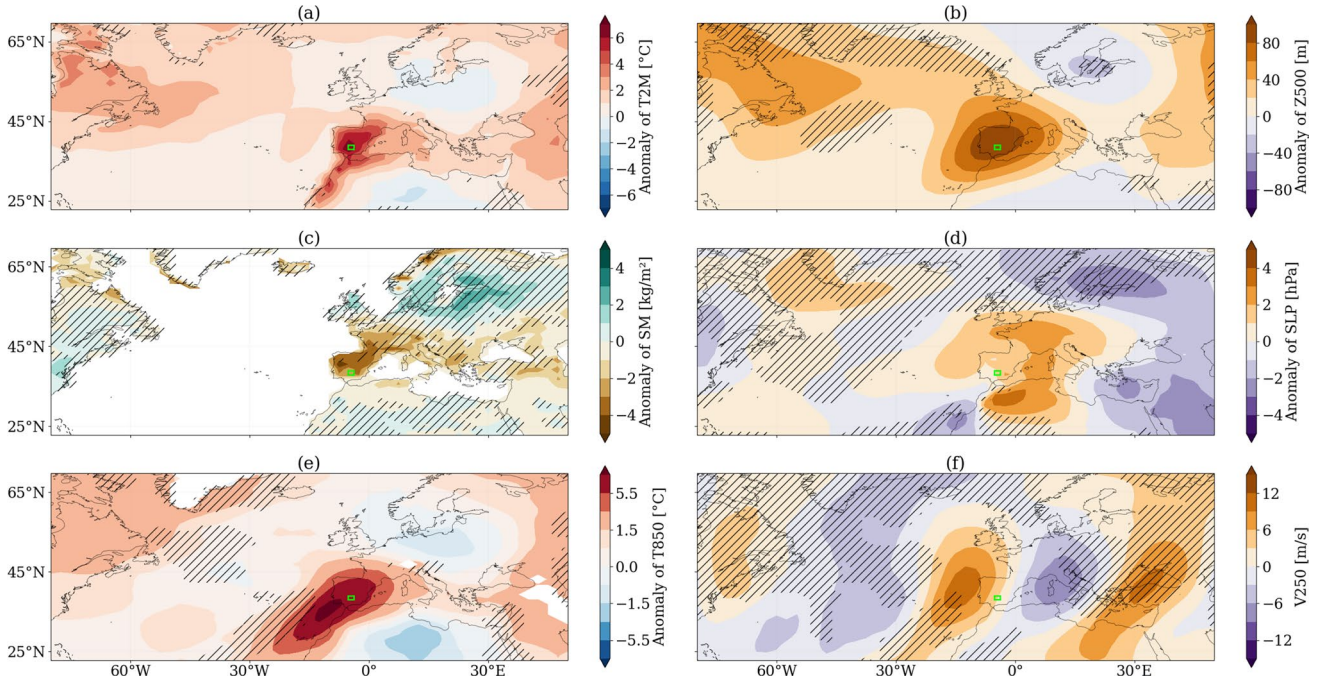


Fig. 8 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the S observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 70\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

these events (Table 1) is between 15 and 30%, similar to the W observable.

Figure 9 presents the same analysis for the N observable. The situation over the North-Hemisphere is presented in annex in Fig. 22. The synoptic situation is characterized by an anticyclone centered south of the observable (panel b), positive anomalies of SLP north of the observable (panel d) and a large amplitude Rossby wave at 250 hPa (panel f). The soil moisture situation (panel c) is almost the opposite of the one of the observable S, with a moist Southern Europe and a dry Northern Europe. The dry soils over Northern Europe, as for the region of anomalous T2M (panel a), extends over a large region encompassing Northern Europe and the Scandinavian and Baltic areas. This is in opposition to the extension of the anomaly in Fig. 8 panel a, which is confined to the Iberian peninsula. We also note the region of low variance associated with strong negative anomaly of soil moisture east of the observable (panel c) (Zschenderlein et al. 2019). The percentage of cut-offs is similar to the W and S observables but when looking at individual events we remark that a majority (> 50%) of these cut-offs are embedded in a so-called modon structure (Butchart et al. 1989) with a blocking high above Scandinavia and a symmetric low above the eastern Mediterranean, splitting the jet into two branches. Again, for the reasons explained above, these structures are averaged out on composite maps.

Figure 10 presents the results for the WCE observable. The situation over the Northern Hemisphere is presented in annex in Fig. 23. Contrary to the observables S, W and N, WCE has a large spatial extension as it encompasses most of West and Central Europe. In this case the synoptic situation is characterized by an anticyclone centered just above the observable (panel b), contrary to the precedent ones. The upper-level circulation looks very similar to the previous ones, with a short wavelength Rossby wave (panel f, wave number 6–7). The soil moisture anomalies (panel c) extend over a large region of Northern, Central and Eastern Europe and this feature is consistent across points in Ω_α. We also note a region of high variance above Southern Sweden for T850 (panel e), which is a feature similar to what was found for observable W (cf. Fig. 6 panel c). This feature could reflect different advecting dynamics of warm air above the boundary layer. The percentage of cut-offs is much higher than for the other observables (around 50%). As for the N observable most of these cut-offs are embedded in a modon structure.

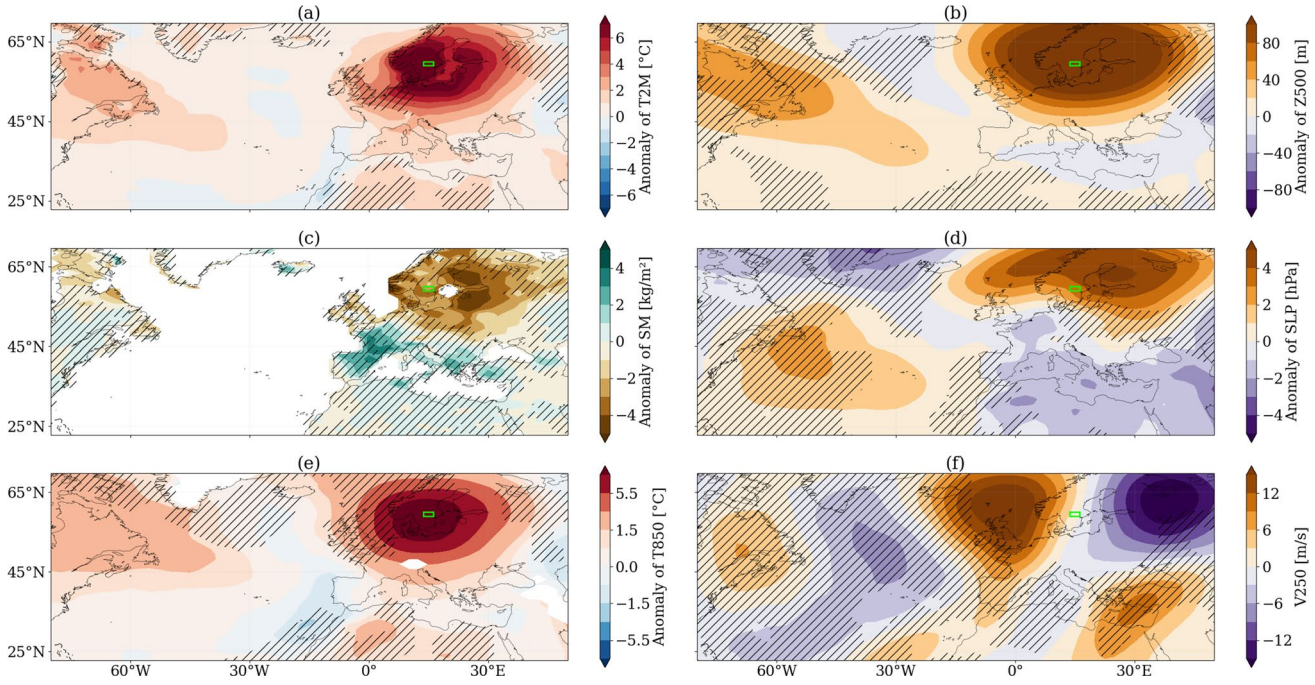


Fig. 9 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the N observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 70\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

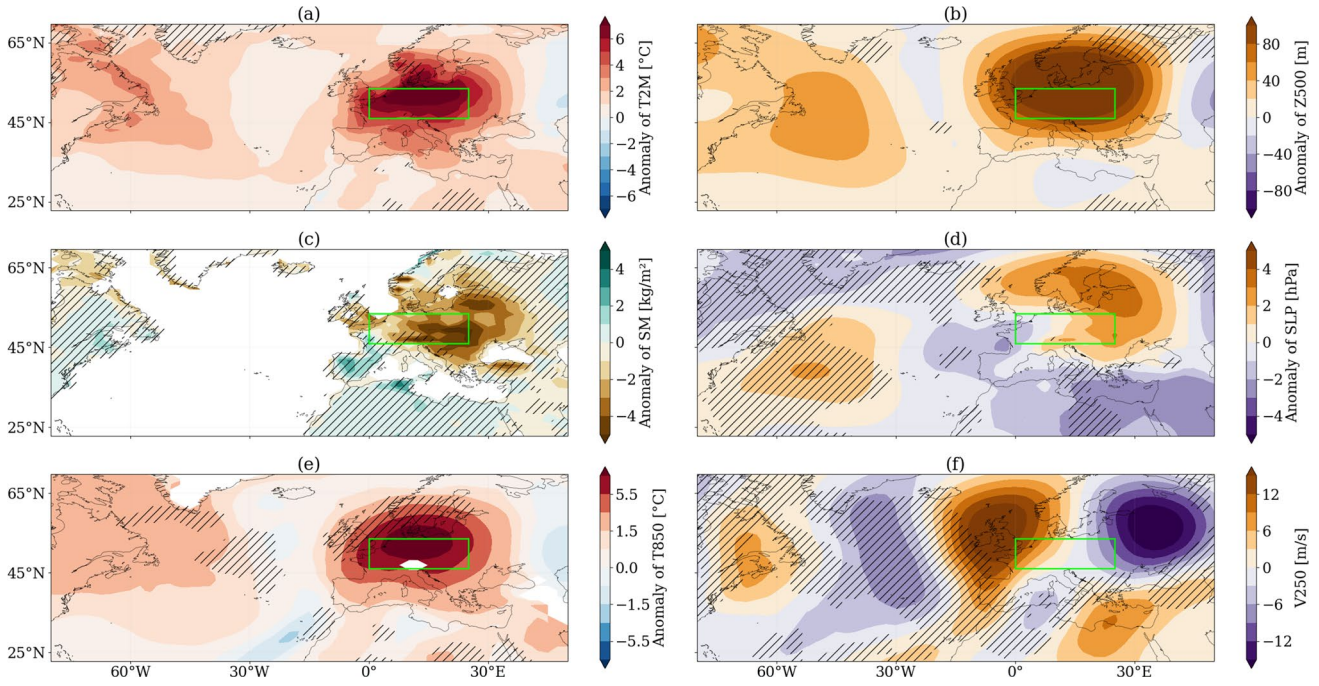


Fig. 10 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the WCE observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 70\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

4 Discussion

Our results show that for all the observables studied here the variance between trajectories reaching the extreme decreases as the level of extremeness of the observable increases. This indicates that there is a concentration of trajectories around a most probable one for extremes of 2-m air temperature in the IPSL-CM6A-LR model. One may wonder why it should be the case. In classical presentations of instanton dynamics (e.g. Grafke and Vanden-Eijnden 2019), the system is supposed to be explicitly stochastic and the probability of an observable can be expressed as a path integral. The path integral then becomes dominated by the contribution of minimum action trajectories in the limit of vanishing noise. In this framework, the variance around the most probable path reaching an extreme level is of the magnitude of the variance of the noise. There is therefore no reason to expect such a variance decrease. In deterministic chaotic systems—such as climate models—with a physical measure μ, Dematteis et al. (2019b) and Lucarini et al. (2023) showed that one can also find a large deviation principle where the role of the small parameter is played by the increasing thresholds reached by an observable. We follow here the presentation of Lucarini et al. (2023).

We consider the chaotic dynamical system defined in Sect. 2, with a unique physical invariant measure μ and an attracting set Ω. We are interested on extremes of observables $F : \Omega \to \mathbb{R}$. For $q \in \mathbb{R}$, we consider the set $\tilde{\Omega}_q$ as the set of points x on the attractor Ω such that $F(x) \geq q$. Then, under minimal assumptions (Dematteis et al. 2019b), one can write:

$$\mathbb{P}(F(x) \geq q) = \mu(\tilde{\Omega}_q) \asymp \exp\left(-\min_{x \in \tilde{\Omega}_q} I(x)\right) \tag{7}$$

where $\asymp$ means that the ratio of the logarithms of both sides tends to unity when $q \to +\infty$. Here q plays the role of the large deviations parameter even though it does not appear explicitly as such in Eq. (7). The functional I is called the rate function and is given by the Legendre transform:

$$I(x) = \max_{p} \left(\langle p, x \rangle - S(p)\right) \tag{8}$$

of

$$S(p) = \log \int_{\Omega} e^{\langle p,x \rangle} \mu(dx) \tag{9}$$

the cumulant generating function of x under the measure μ with $\langle \cdot\,, \cdot \rangle$ the inner product in $\mathbb{R}^d$. The large deviation principle is thus based on the uncertainty of the points in Ω which are distributed according to the measure μ.

When $q \to +\infty$ in Eq. (7), the probability is exponentially dominated by the point $\hat{x}$ which minimizes the rate function: $\hat{x} = \text{argmin}_{x \in \tilde{\Omega}_q} I(x)$. The minimizer $\hat{x}$ represents physically the state with the maximum likelihood of realization. Under the stated conditions, the probability measure accumulates near $\hat{x}$: this *most likely* of the *least likely* states is usually called the *instanton*. The physical intuition behind this result is the following: because extremes of an observable are rare (i.e. have low probability), the system has a set of paths with small probability measure in the phase space to reach those extremes. If the system had many paths to reach the extremes, then these would not be rare. This idea of uniqueness of limit behaviors in a dynamical system is reminiscent of the uniqueness of the limit distribution in EVT (Coles et al. 2001).

In our context, the long simulation of the model gives an empirical probability distribution for the states of the system in the phase space and the instanton can therefore be considered as the optimal initial conditions defining a trajectory which leads to the intense heatwave. Although we have no means to demonstrate that the average trajectories shown in this work correspond to the true instantons, the conditional dynamics sampled with the method presented here may be considered as a reasonable approximation. The long simulation of the model acts as a random sampling of states distributed according to the measure μ of the system. The fact that the probability measure concentrates around a most probable state for high thresholds q implies that states reaching the threshold will be on average closer to the most probable state when the threshold increases. Hence, the mean distance from this state—measured by our normalized variance—will decrease with increasing threshold.

5 Conclusions

In this paper we have investigated the typicality of the dynamics leading to extreme events using air temperature at 2-m (T2M) at four locations in Europe in the IPSL-CM6A-LR model (Boucher et al. 2020) pre-industrial control run as our observables of interest. Using the 2000-year simulation, we employed an instanton filtering procedure (Grafke et al. 2013) consisting in averaging trajectories which reach a similar extreme value of these observables.

We have shown that the variance between trajectories reaching the extreme decreases as the level of extremeness of the observable increases. In other words, the more intense the extreme of T2M, the more likely that the trajectories reaching this extreme all look the same. We demonstrated this convergence on all variables considered: air temperature at 2-m (T2M), geopotential height at 500 hPa (Z500), upper level soil moisture (SM), sea-level pressure (SLP), temperature at 850 hPa (T850) and meridional wind speed at 250 hPa (V250) for a grid point observable in Western Europe (observable W). Remarkably, the variance decreases even at places far from the location of the observable, which suggests a hemispheric dynamics leading to very intense heatwaves. We also showed a stronger decrease of variance for higher values of the rolling mean window r, indicating that the typicality is more easily reached when looking at longer time averages as suggested by Lucarini et al. (2023).

The instanton dynamics found with our analysis is consistent with the mechanisms identified by previous literature for heatwaves dynamics in mid-latitudes (Perkins 2015; Horton et al. 2016; Domeisen et al. 2022a). In the IPSL-CM6A-LR model with a pre-industrial CO_2 level, very high air temperature at 2-m are reached by a combination of dry soils, a large mid-troposphere anticyclone and an upper level Rossby wave of 6–7 wave number. This situation ensures both the advection of warm air from the south at the west flank of the anticyclone over the whole troposphere, subsidence, associated adiabatic heating, and clear skies which favors the radiative heating of the lowest layers of the atmosphere in combination with reduced water evaporation.

We investigated the instanton hypothesis for three other observables: T2M at two grid points situated in the south (observable S) and north (observable N) of Europe and T2M averaged over an area covering most of Western and Central Europe (observable WCE). We found a similar convergence mechanism of different trajectories, but with some discrepancies. The convergence of variance is much stronger for observable N than observable S for dynamical variables (SLP, Z500 and V250), and also stronger for observable WCE than observable W for most variables. We showed that the global dynamics described above is similar for reaching extremes at these locations. Overall we found that the instanton hypothesis is consistent with our results, but the convergence is stronger for observable farther to the north, observables computed on extended spatial locations and for longer extremes (higher values of the rolling mean window r). We also investigated the local discrepancies on the

convergence of variance and found that a substantial amount of dynamical paths (around 20% for the grid point observables) are associated to the presence of cut-off lows. This suggests that there may be a multi-modality in the dynamical paths reaching very intense hot events. This is however not clear what are the precise effects—either dynamical or thermodynamical—of the presence of these cut-offs on the intensity of the observed heatwaves.

The observed discrepancies in the decrease of the variance for some variables at certain location however do not allow to disprove the instanton hypothesis. Indeed, with an order $\alpha = 0.999$ for the quantile of the observable's empirical distribution over the JJA months and 2000 summers, it corresponds to roughly choosing $n = 50$ heatwaves with a return time around 40 years. This amount of data is greater than what has been recorded through observations since the start of the satellite era. Nonetheless, this may not be sufficient to ensure full convergence and this could explain why the variance remain high at certain locations. A much longer data set would be needed to investigate the dynamics leading to more extreme temperatures, especially daily records. For such very intense events, the bi-modality suggested here may disappear. Even if our analysis suggests that very high extremes are reached by a strengthening of the mechanisms leading to high extremes, it cannot be completely ruled out that the mechanisms can change if one wants to reach even higher values. One possible explanation could be the release of latent heat from tropical air in conjunction with an atmospheric river over the North Atlantic, which seems to be one of the reinforcing structures of the 2021 North-Western America heatwave (Qian et al. 2022; Lin et al. 2022; Mo et al. 2022). To the best of our knowledge, such a dynamics has never been observed for heatwaves in Western Europe.

In this paper we did not study long term potential precursors of heatwaves such as anomalous sea surface temperature (SST) patterns or large-scale modes of climate variability (Domeisen et al. 2022a). Anomalous SST patterns are known to be present in observed heatwaves (Black and Sutton 2007; Duchez et al. 2016; McKinnon et al. 2016), and a convergence of the surface oceanic dynamics to reach extreme land temperature is therefore likely in climate models. It may also be the case for large-scale modes of climate variability such as the El Nino-Southern Oscillation (ENSO) (Martija-Díez et al. 2021), the Atlantic Multidecadal Oscillation (AMO) or the Pacific Decadal Oscillation (PDO). These processes are nonetheless suggested by the large regions of low variance for T2M above the Atlantic and the Pacific oceans found with our analysis.

One may wonder how the paradigm of the typical dynamics leading to heatwaves is consistent with the presence of different clusters of heatwaves demonstrated by several studies on empirical data (e.g. Stefanon et al. 2012; Gibson et al. 2017; Wang et al. 2018; Keellings and Moradkhani 2020; Mondal and Mishra 2021). As we have shown here, the typical dynamics highly depends on the observable considered. It may therefore be possible that the typical dynamics leading to extremes of neighboring grid points is very similar and changes dramatically when crossing relevant physical barriers, such as mountains (Lucarini et al. 2023). Moreover, if one considers less extreme values of the distribution of grid point observables, it is possible that the typicality is not reached for this observable but it may be reached for another observable encompassing a broader region. One could therefore compare the event observed to the typical events for the observable which maximizes its spatiotemporal rarity as proposed by Cattiaux and Ribes (2018) in the context of attribution of extreme events.

Despite the lack of data for extremely rare events, the typical dynamics hypothesis offers a significant simplification by predicting that studying such events is equivalent to studying a single trajectory (assuming that their is no multi-modality). Therefore, this framework is encouraging for gaining predictive power on the dynamics leading to extreme events and it may explain why summer heatwaves are among the most predictable meteorological extremes on subseasonal timescales (Vitart and Robertson 2018; Vitart et al. 2019; Domeisen et al. 2022b). Here we studied the extremes of air temperature close to the surface. For other variables of interest our preceding statement of the validity of the instanton hypothesis should also be tested. More generally, this paper documents a method to study the dynamics leading to extreme events in non-equilibrium physical systems. Our framework suggests a connection between the statistical method of studying extreme events and the in-depth examination of specific events through case studies.

Appendix A: Additional figures

See Figs. 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23 and Table 1.

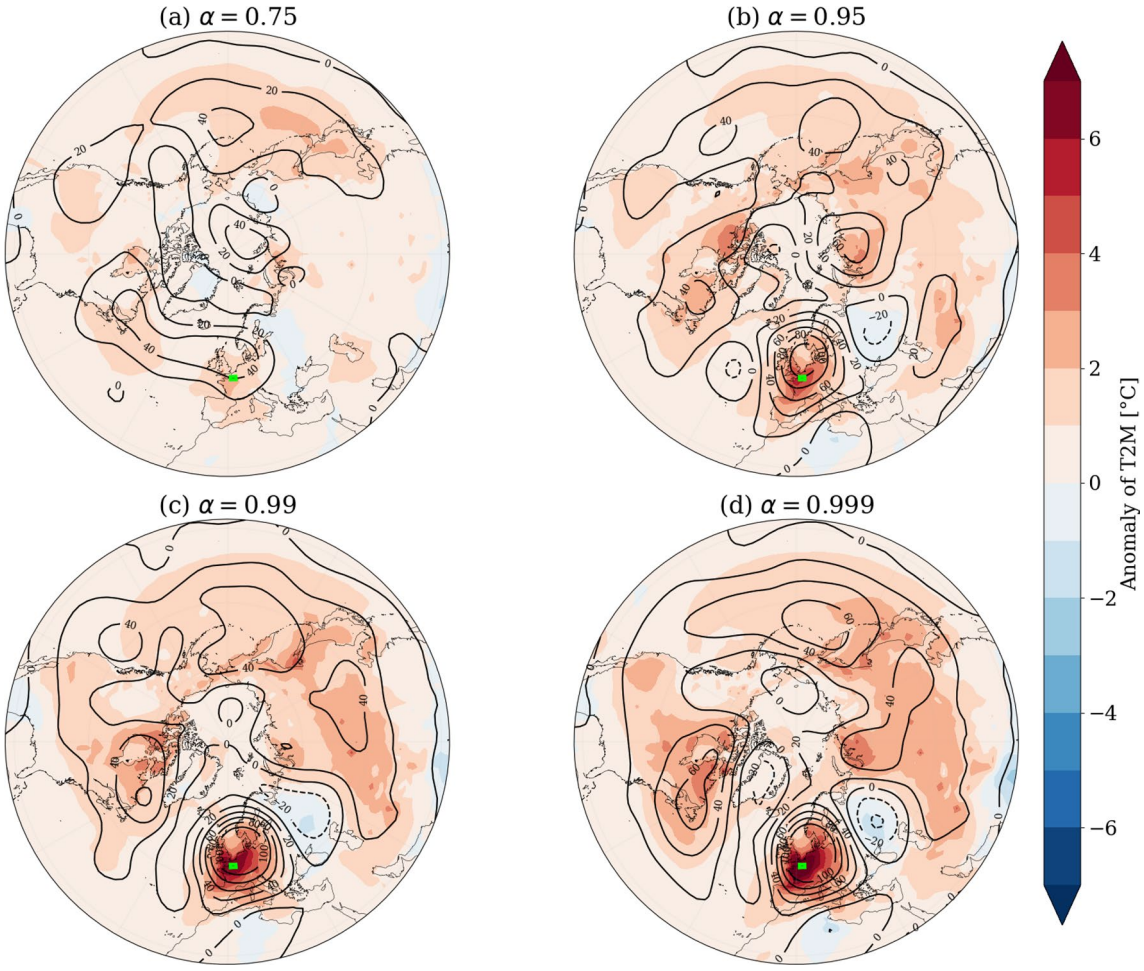


Fig. 11 Composite maps $\widehat{A}^r$ of anomalies of T2M and Z500 for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. Colors: anomaly of T2M (K). Contours: anomaly of Z500 (m). Anomalies are computed with respect to the JJA average. The green box displays the location where the observable is computed

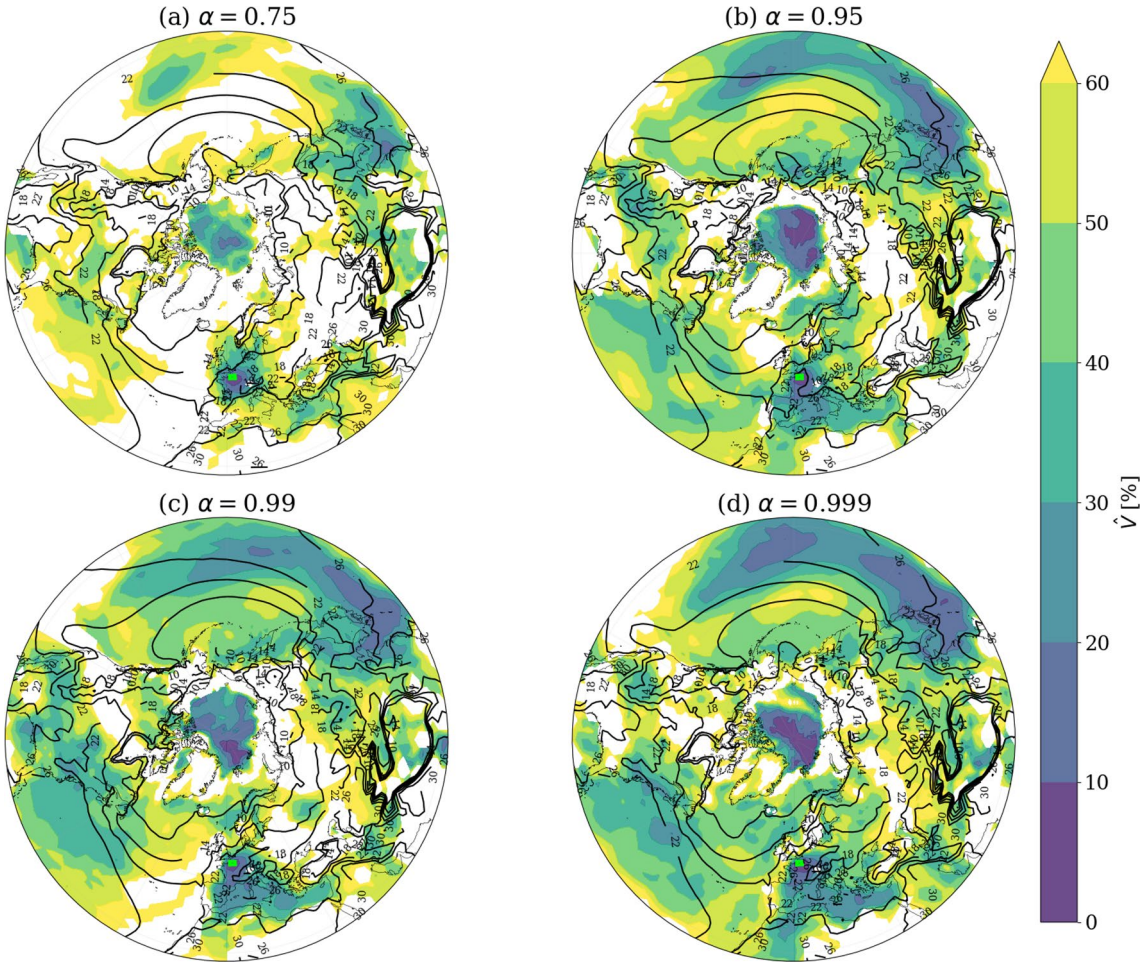


Fig. 12 Composite $\widehat{A}^r$ (contours, °C) and normalized variance $\widehat{V}^r$ (colors) of the T2M field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

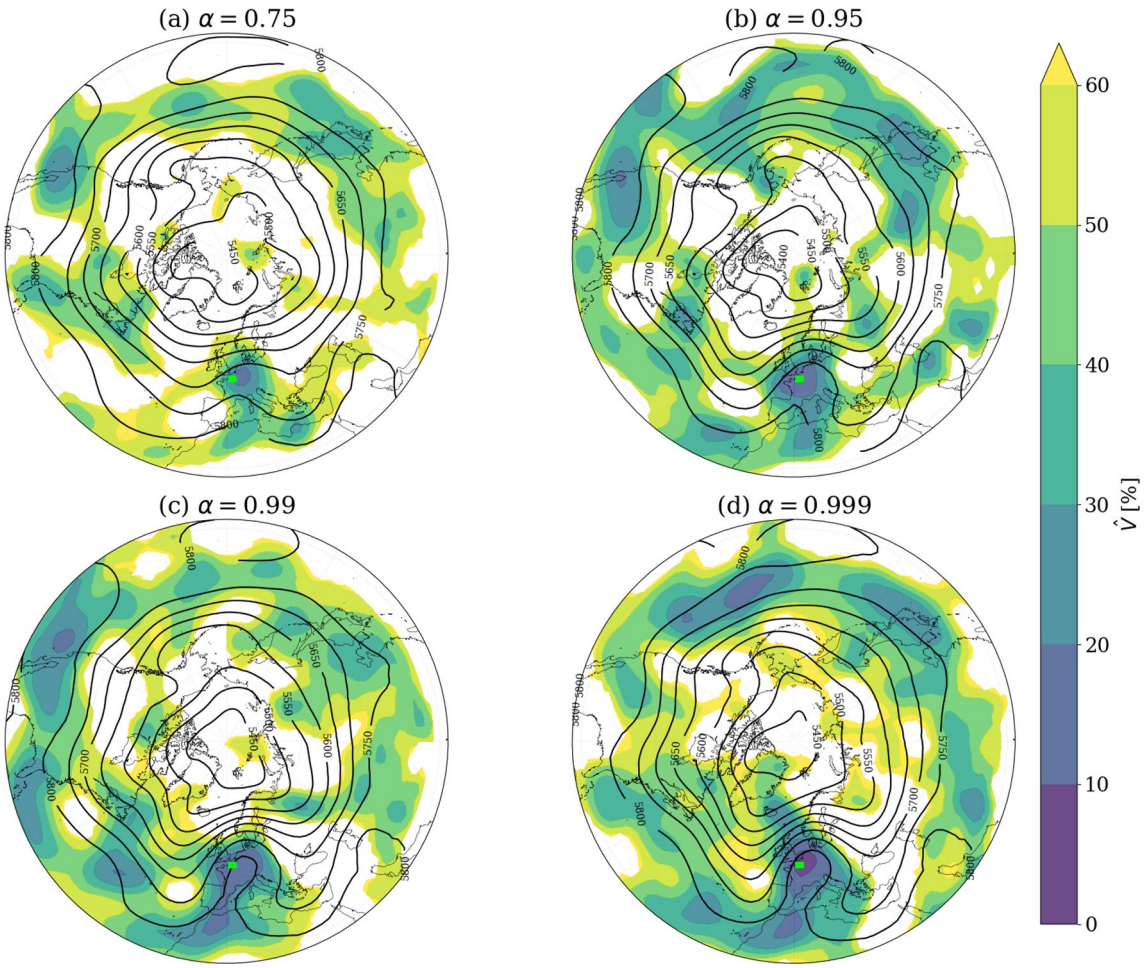


Fig. 13 Composite $\widehat{A}^r$ (contours, m) and normalized variance $\widehat{V}^r$ (colors) of the Z500 field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 5$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

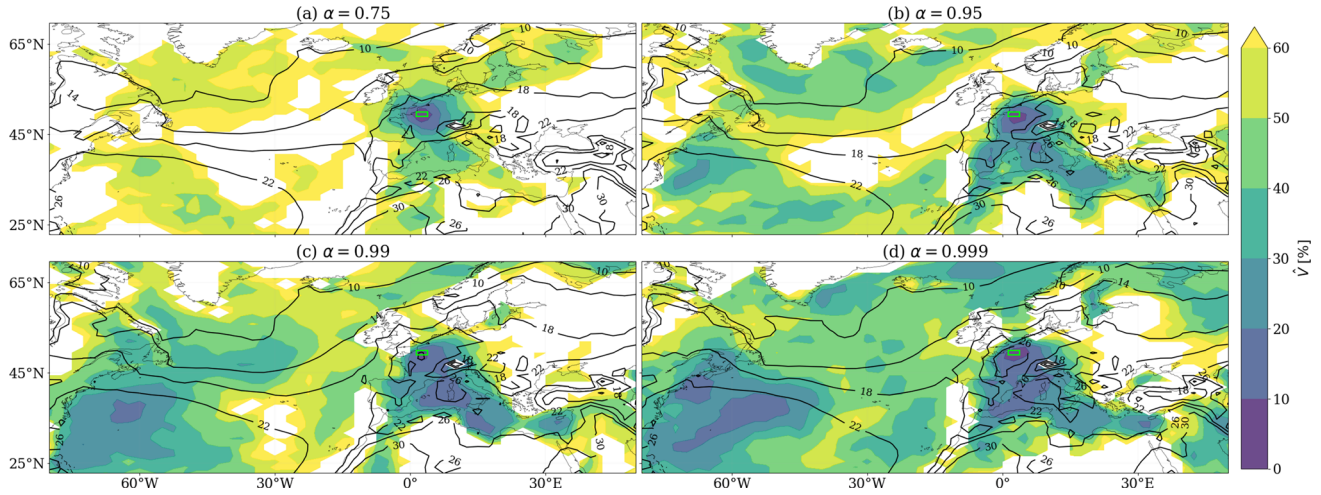


Fig. 14 Composite $\widehat{A}^r$ (contours, °C) and normalized variance $\widehat{V}^r$ (colors) of the T2M field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 1$ day. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

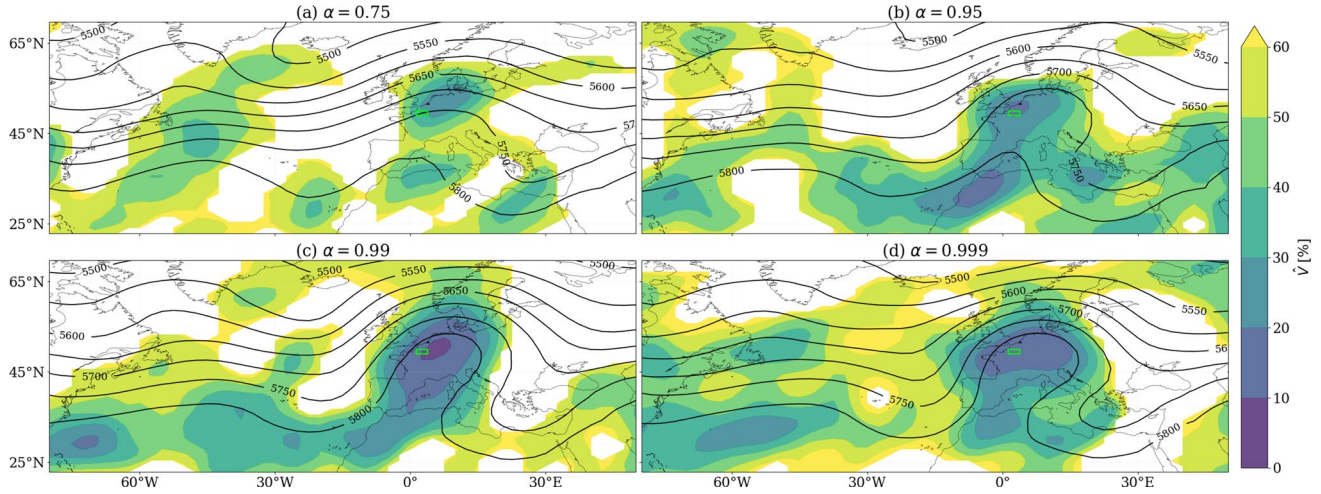


Fig. 15 Composite $\widehat{A}^r$ (contours, m) and normalized variance $\widehat{V}^r$ (colors) of the Z500 field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 1$ day. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

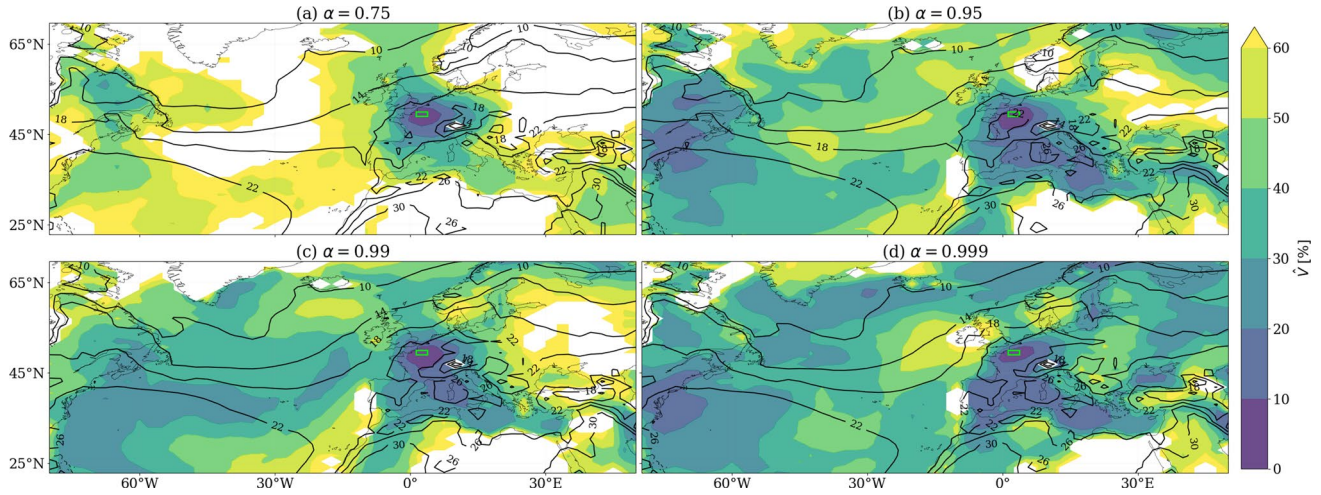


Fig. 16 Composite $\hat{A}^r$ (contours, °C) and normalized variance $\hat{V}^r$ (colors) of the T2M field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 15$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

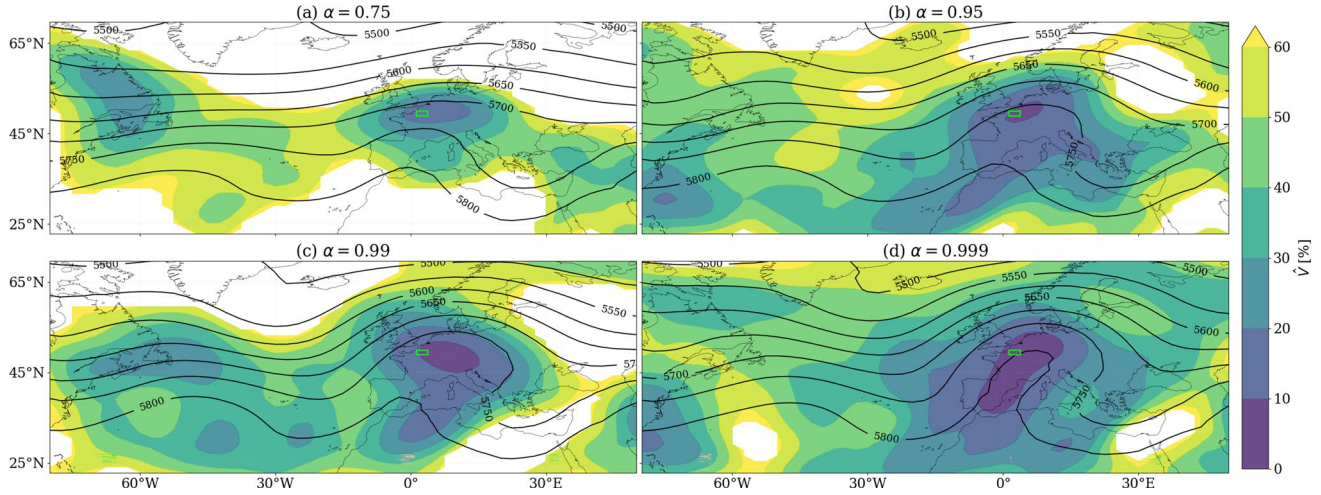


Fig. 17 Composite $\hat{A}^r$ (contours, m) and normalized variance $\hat{V}^r$ (colors) of the Z500 field for increasing α-th order of the quantile of the W observable's empirical distribution. The figure is computed for a rolling mean window of $r = 15$ days. The normalized variance is computed after averaging over the rolling window and is expressed in %. The blank regions correspond to a normalized variance not significantly different than 100% at the 5% level. The green box displays the location where the observable is computed

Table 1 Percentage of cut-off lows in the dynamics leading to extremes

Observable	$r = 1$			$r = 5$			$r = 15$		
	0.95 (%)	0.99 (%)	0.999 (%)	0.95 (%)	0.99 (%)	0.999 (%)	0.95 (%)	0.99 (%)	0.999 (%)
S	22	30	22	22	16	24	24	22	16
W	22	16	20	14	30	28	16	22	24
N	28	26	14	20	20	20	16	32	20
WCE	56	44	50	38	60	48	42	32	58

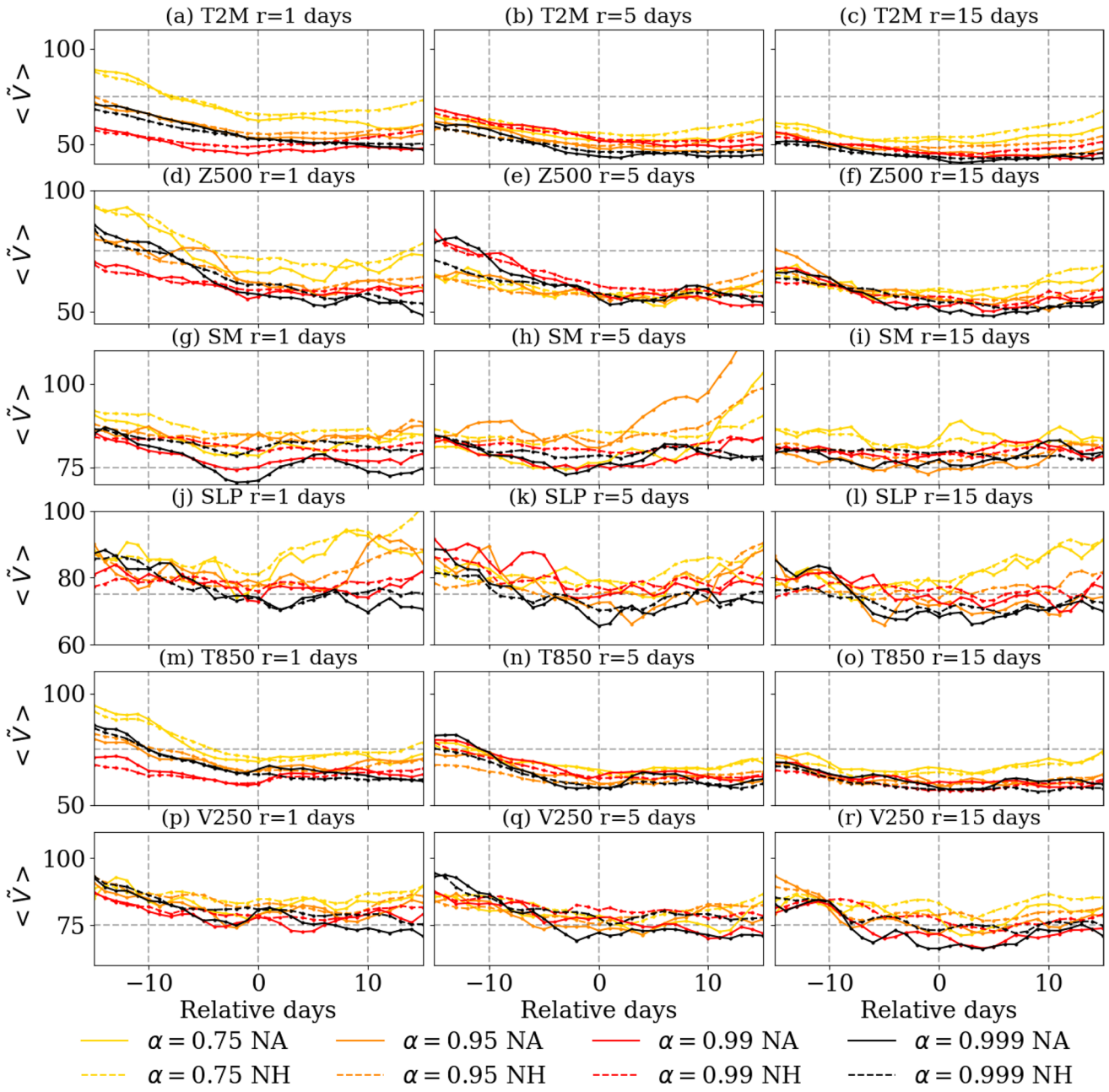


Fig. 18 Evolution of the normalized variance $< \widetilde{V}(t) >$ averaged over the Euro-Atlantic sector (plain lines) and the Northern Hemisphere (dashed lines) for the different variables (S observable). The normalized variance is expressed in %. The colors show the α-th order quantile of the observable's empirical distribution. The time is expressed relative to the day when the observable is such that $F_r(x_t) \simeq q_\alpha$. The gray horizontal dashed line shows the 75% level

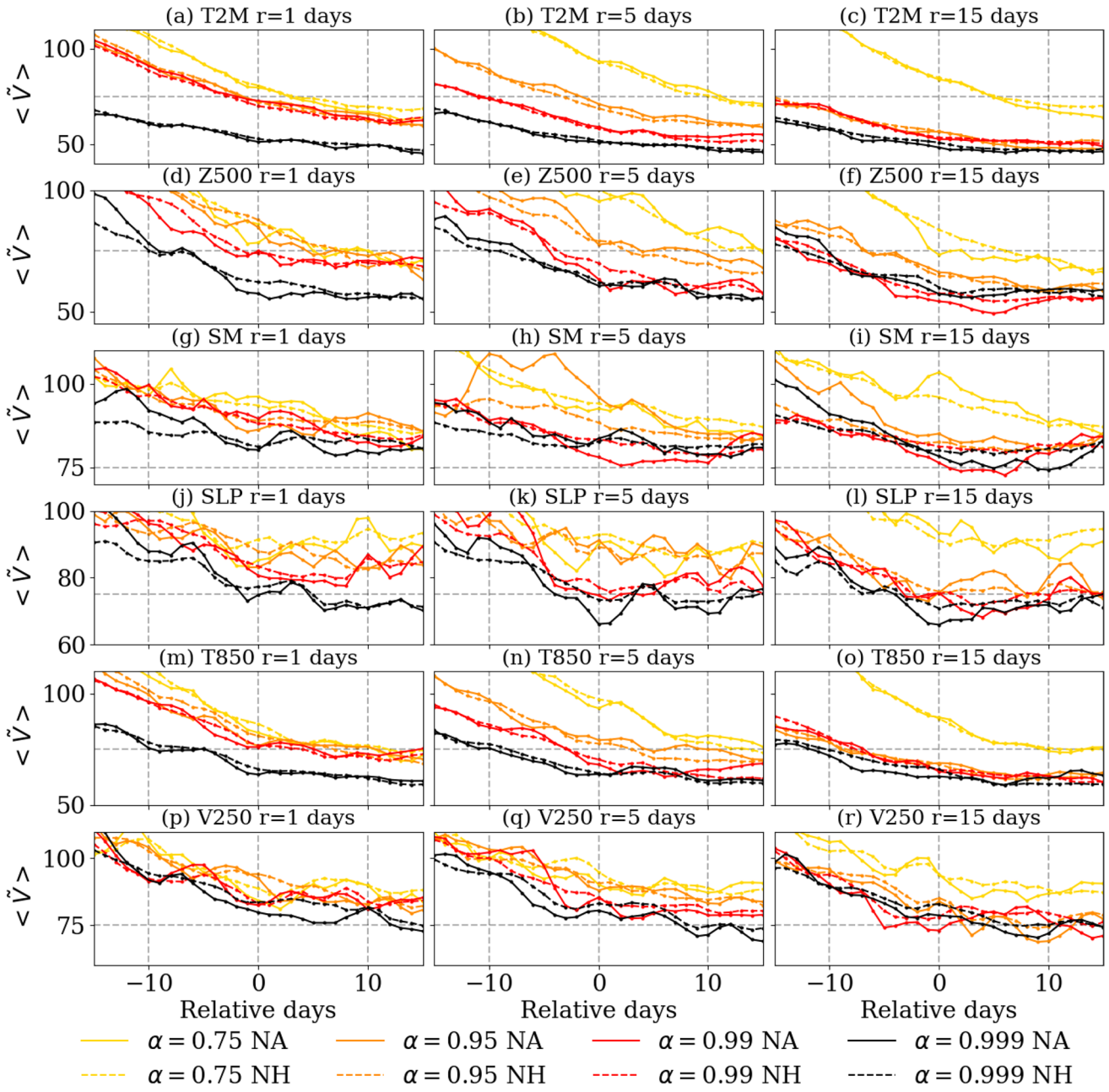


Fig. 19 Evolution of the normalized variance $< \widetilde{V}(t) >$ averaged over the Euro-Atlantic sector (plain lines) and the Northern Hemisphere (dashed lines) for the different variables (N observable). The normalized variance is expressed in %. The colors show the α-th order quantile of the observable's empirical distribution. The time is expressed relative to the day when the observable is such that $F_r(x_t) \simeq q_\alpha$. The gray horizontal dashed line shows the 75% level

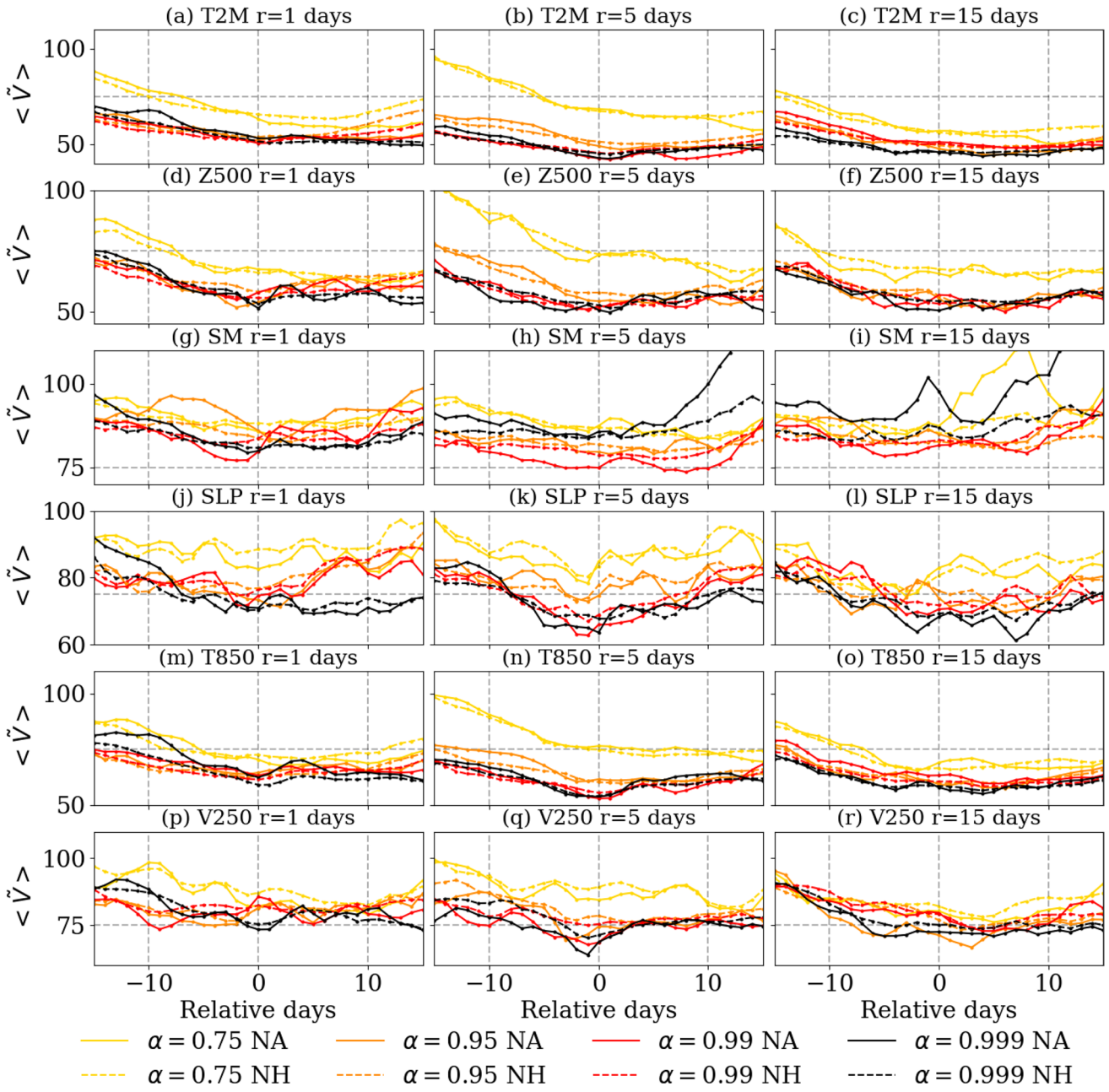


Fig. 20 Evolution of the normalized variance $< \widetilde{V}(t) >$ averaged over the Euro-Atlantic sector (plain lines) and the Northern Hemisphere (dashed lines) for the different variables (WCE observable). The normalized variance is expressed in %. The colors show the α -th order quantile of the observable's empirical distribution. The time is expressed relative to the day when the observable is such that $F_r(x_t) \simeq q_\alpha$. The gray horizontal dashed line shows the 75% level

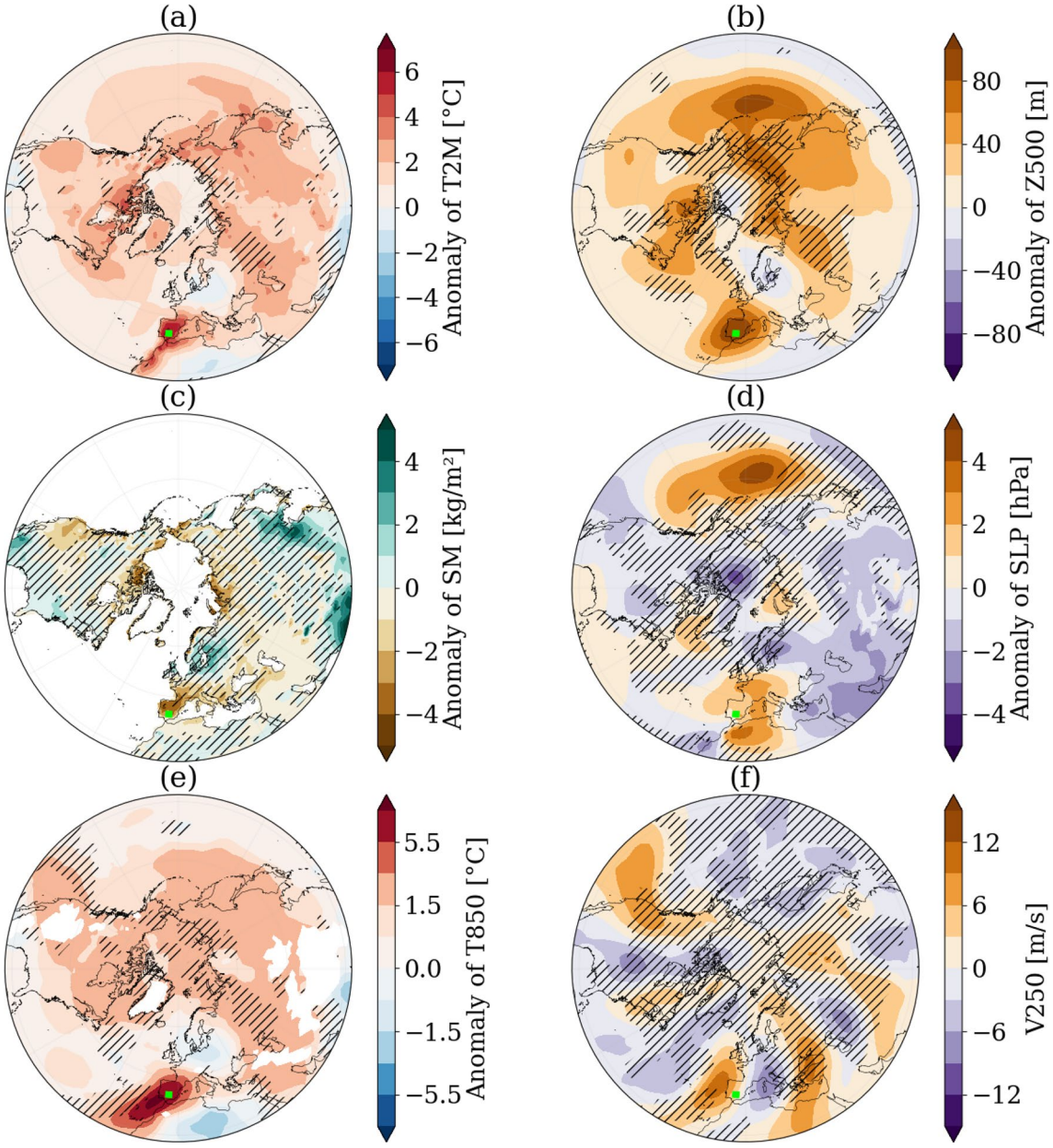


Fig. 21 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the S observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 50\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

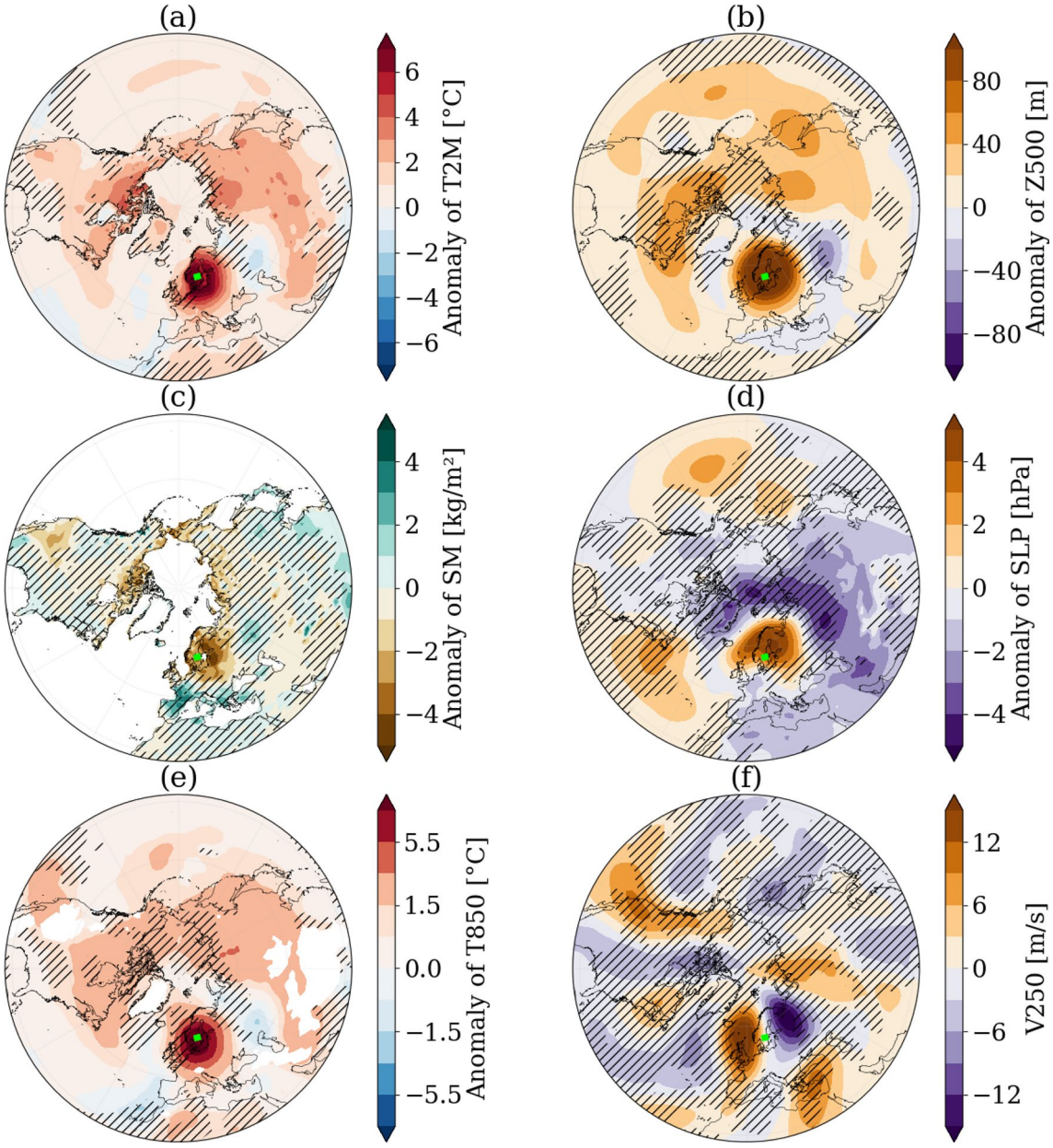


Fig. 22 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the N observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 50\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

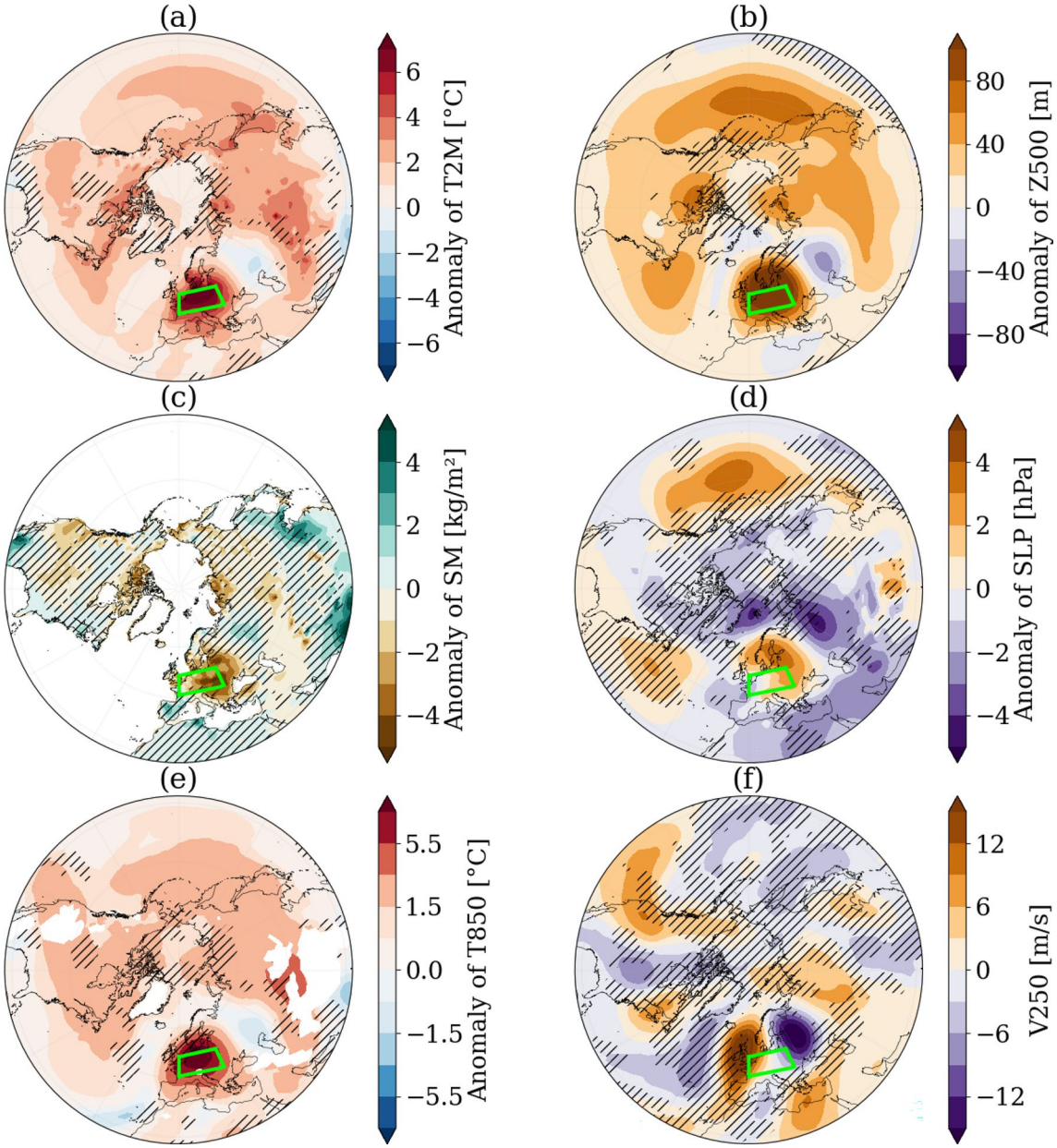


Fig. 23 Composite maps $\widehat{A}^r$ for T2M, Z500, SM, SLP, T850 and V250 for the quantile of order $\alpha = 0.999$ of the WCE observable's empirical distribution. The hatch areas correspond to $\widehat{V}^r > 50\%$. The figure is computed for a rolling mean window of $r = 5$ days. Anomaly of **a** air temperature at 2-m (T2M), **b** geopotential height at 500 hPa (Z500), **c** soil moisture (SM), **d** sea-level pressure (SLP) and **e** temperature at 850 hPa (T850), and **f** meridional wind speed at 250 hPa (V250). The anomalies are computed with respect to their average over the summer (JJA). The green box displays the location where the observable is computed

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00382-023-06967-5.

Acknowledgements The authors would like to thank V. Lucarini, M. Galfi, G. Messori, A. Caubel and T. Grafke for fruitful discussions, and M.A. Foujols for running the IPSL-CM6A-LR pre-industrial control run simulation. The authors also thank the two anonymous reviewers for their constructive comments.

Author contributions RN did the data analysis. All the authors contributed to writing and reviewing the article.

Funding This work has received support from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101003469 (XAIDA), the Grant ANR-20-CE01-0008-01 (SAMPRACE) and the Marie Sklodowska-Curie Grant Agreement No. 956396 (EDIPI). RN was supported by a doctoral Grant from CEA.

Data availability The pre-industrial control run of the IPSL model is available upon request.

Code availability The main results of this work were obtained using Python. The scripts are available upon request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical standard Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Bartusek S, Kornhuber K, Ting M (2022) 2021 north American heatwave amplified by climate change-driven nonlinear interactions. Nat Clim Change 12:1143–1150

Beguería S, Vicente-Serrano SM, Reig F et al (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int J Climatol 34(10):3001–3023

Black E, Sutton R (2007) The influence of oceanic conditions on the hot European summer of 2003. Clim Dyn 28(1):53–66

Boucher O, Servonnat J, Albright AL et al (2020) Presentation and evaluation of the IPSL-CM6A-LR climate model. J Adv Model Earth Syst 12(7):e2019MS002,010

Bouchet F, Rolland J, Simonnet E (2019) Rare event algorithm links transitions in turbulent flows with activated nucleations. Phys Rev Lett 122(7):074,502

Butchart N, Haines K, Marshall J (1989) A theoretical and diagnostic study of solitary waves and atmospheric blocking. J Atmos Sci 46(13):2063–2078

Cattiaux J, Ribes A (2018) Defining single extreme weather events in a climate perspective. Bull Am Meteor Soc 99(8):1557–1568

Cheng L, AghaKouchak A, Gilleland E et al (2014) Non-stationary extreme value analysis in a changing climate. Clim Change 127:353–369

Chetrite R, Touchette H (2015) Nonequilibrium Markov processes conditioned on large deviations. Ann Henri Poincaré 16:2005–2057

Coles S, Bawa J, Trenner L et al (2001) An introduction to statistical modeling of extreme values, vol 208. Springer, Berlin

Coumou D, Petoukhov V, Rahmstorf S et al (2014) Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. Proc Natl Acad Sci 111(34):12,331-12,336

Dematteis G, Grafke T, Onorato M et al (2019a) Experimental evidence of hydrodynamic instantons: the universal route to rogue waves. Phys Rev X 9(4):041,057

Dematteis G, Grafke T, Vanden-Eijnden E (2019b) Extreme event quantification in dynamical systems with random components. SIAM/ASA J Uncertain Quantif 7(3):1029–1059

Di Capua G, Sparrow S, Kornhuber K et al (2021) Drivers behind the summer 2010 wave train leading to Russian heatwave and Pakistan flooding. npj Clim Atmos Sci 4(1):1–14

Dirmeyer PA, Balsamo G, Blyth EM et al (2021) Land-atmosphere interactions exacerbated the drought and heatwave over northern Europe during summer 2018. AGU Adv 2(2):e2020AV000,283

Dole R, Hoerling M, Perlwitz J et al (2011) Was there a basis for anticipating the 2010 Russian heat wave? Geophys Res Lett 38(6):L06702

Domeisen DI, Eltahir EA, Fischer EM et al (2022a) Prediction and projection of heatwaves. Nat Rev Earth Environ 4:36–50

Domeisen DI, White CJ, Afargan-Gerstman H et al (2022b) Advances in the subseasonal prediction of extreme events: relevant case studies across the globe. Bull Am Meteorol Soc 103:E1473–E1501

Dominicis CD (1976) Techniques de renormalisation de la théorie des champs et dynamique des phénomenes critiques. J Phys Colloq 37:247

Dubrulle B, Daviaud F, Faranda D et al (2022) How many modes are needed to predict climate bifurcations? lessons from an experiment. Nonlinear Process Geophys 29(1):17–35

Duchez A, Frajka-Williams E, Josey SA et al (2016) Drivers of exceptionally cold North Atlantic ocean temperatures and their link to the 2015 European heat wave. Environ Res Lett 11(7):074,004

Eyring V, Bony S, Meehl GA et al (2016) Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. Geosci Model Dev 9(5):1937–1958

Faranda D, Pascale S, Bulut B (2022) Persistent anticyclonic conditions and climate change exacerbated the exceptional 2022 European-mediterranean drought. Environ Res Lett 18:034030

Finkel J, Webber RJ, Gerber EP et al (2021) Learning forecasts of rare stratospheric transitions from short simulations. Mon Weather Rev 149(11):3647–3669

Fragkoulidis G, Wirth V, Bossmann P et al (2018) Linking northern hemisphere temperature extremes to Rossby wave packets. Q J R Meteorol Soc 144(711):553–566

Freidlin MI, Wentzell AD (1998) Random perturbations of dynamical systems. Springer New York

Fuchs A, Herbert C, Rolland J et al (2022) Instantons and the path to intermittency in turbulent flows. Phys Rev Lett 129(3):034,502

Galfi VM, Lucarini V (2021) Fingerprinting heatwaves and cold spells and assessing their response to climate change using large deviation theory. Phys Rev Lett 127(5):058,701

Galfi VM, Lucarini V, Ragone F et al (2021) Applications of large deviation theory in geophysical fluid dynamics and climate science. La Rivista del Nuovo Cimento 44(6):291–363

García-Herrera R, Díaz J, Trigo RM et al (2010) A review of the European summer heat wave of 2003. Crit Rev Environ Sci Technol 40(4):267–306

Ghil M, Yiou P, Hallegatte S et al (2011) Extreme events: dynamics, statistics and prediction. Nonlinear Process Geophys 18(3):295–350

Gibson PB, Perkins-Kirkpatrick SE, Alexander LV et al (2017) Comparing Australian heat waves in the cmip5 models through cluster analysis. J Geophys Res Atmos 122(6):3266–3281
Grafke T, Vanden-Eijnden E (2019) Numerical computation of rare events via large deviation theory. Chaos Interdiscip J Nonlinear Sci 29(6):063,118. https://doi.org/10.1063/1.5084025
Grafke T, Grauer R, Schäfer T (2013) Instanton filtering for the stochastic burgers equation. J Phys A Math Theor 46(6):062,002
Hirschi M, Seneviratne SI, Alexandrov V et al (2011) Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nat Geosci 4(1):17–21
Horton RM, Mankin JS, Lesk C et al (2016) A review of recent advances in research on extreme heat events. Curr Clim Change Rep 2(4):242–259
Janssen HK (1976) On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. Z Phys B Condens Matter 23(4):377–380
Keellings D, Moradkhani H (2020) Spatiotemporal evolution of heat wave severity and coverage across the united states. Geophys Res Lett 47(9):e2020GL087,097
Koppe C, Kovats S, Jendritzky G et al (2004) Heat-waves: risks and responses. EUR/03/5036810, World Health Organization. Regional Office for Europe
Kornhuber K, Tamarin-Brodsky T (2021) Future changes in northern hemisphere summer weather persistence linked to projected arctic warming. Geophys Res Lett 48(4):e2020GL091,603
Kornhuber K, Petoukhov V, Karoly D et al (2017) Summertime planetary wave resonance in the northern and southern hemispheres. J Clim 30(16):6133–6150
Kornhuber K, Coumou D, Vogel E et al (2020) Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. Nat Clim Change 10(1):48–53
Lin H, Mo R, Vitart F (2022) The 2021 western north American heatwave and its subseasonal predictions. Geophys Res Lett 49(6):e2021GL097,036
Lucarini V, Melinda Galfi V, Riboldi J et al (2023) Typicality of the 2021 western north America summer heatwave. Environ Res Lett 18(1):015,004
Mann ME, Rahmstorf S, Kornhuber K et al (2017) Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. Sci Rep 7(1):1–12
Mann ME, Rahmstorf S, Kornhuber K et al (2018) Projected changes in persistent extreme summer weather events: the role of quasi-resonant amplification. Sci Adv 4(10):eaat3272
Martija-Díez M, Rodríguez-Fonseca B, López-Parages J (2021) Enso influence on Western European summer and fall temperatures. J Clim 34(19):8013–8031
Martin PC, Siggia E, Rose H (1973) Statistical dynamics of classical systems. Phys Rev A 8(1):423
McKinnon KA, Rhines A, Tingley M et al (2016) Long-lead predictions of eastern united states hot days from pacific sea surface temperatures. Nat Geosci 9(5):389–394
Miralles DG, Van Den Berg M, Teuling A et al (2012) Soil moisture-temperature coupling: a multiscale observational analysis. Geophys Res Lett 39(21)
Miralles DG, Teuling AJ, Van Heerwaarden CC et al (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat Geosci 7(5):345–349
Mo R, Lin H, Vitart F (2022) An anomalous warm-season trans-pacific atmospheric river linked to the 2021 Western North America heatwave. Commun Earth Environ 3(1):1–12
Moloney NR, Faranda D, Sato Y (2019) An overview of the extremal index. Chaos Interdiscip J Nonlinear Sci 29(2):022,101
Mondal S, Mishra AK (2021) Complex networks reveal heatwave patterns and propagations over the USA. Geophys Res Lett 48(2):e2020GL090,411
Muñoz C, Schultz D, Vaughan G (2020) A midlatitude climatology and interannual variability of 200-and 500-hPa cut-off lows. J Clim 33(6):2201–2222
Neal E, Huang CS, Nakamura N (2022) The 2021 pacific northwest heat wave and associated blocking: meteorology and the role of an upstream cyclone as a diabatic source of wave activity. Geophys Res Lett 49(8):e2021GL097,699
Otto FE, Massey N, van Oldenborgh GJ et al (2012) Reconciling two approaches to attribution of the 2010 Russian heat wave. Geophys Res Lett 39(4)
Overland JE (2021) Causes of the record-breaking pacific northwest heatwave, late June 2021. Atmosphere 12(11):1434
Perkins SE (2015) A review on the scientific understanding of heatwaves-their measurement, driving mechanisms, and changes at the global scale. Atmos Res 164:242–267
Petoukhov V, Rahmstorf S, Petri S et al (2013) Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. Proc Natl Acad Sci 110(14):5336–5341
Petoukhov V, Petri S, Rahmstorf S et al (2016) Role of quasiresonant planetary wave dynamics in recent boreal spring-to-autumn extreme events. Proc Natl Acad Sci 113(25):6862–6867
Philip SY, Kew SF, van Oldenborgh GJ et al (2021) Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the US and Canada June 2021. Earth Syst Dyn Discuss 13:1689-1713
Pörtner HO, Roberts DC, Adams H et al (2022) Climate change 2022: impacts, adaptation and vulnerability. IPCC sixth assessment report
Qian Y, Hsu PC, Yuan J et al (2022) Effects of subseasonal variation in the east Asian monsoon system on the summertime heat wave in Western North America in 2021. Geophys Res Lett 49(8):e2021GL097,659
Ragone F, Wouters J, Bouchet F (2018) Computation of extreme heat waves in climate models using a large deviation algorithm. Proc Natl Acad Sci 115(1):24–29
Rasmijn L, Van der Schrier G, Bintanja R et al (2018) Future equivalent of 2010 Russian heatwave intensified by weakening soil moisture constraints. Nat Clim Change 8(5):381–385
Screen JA, Simmonds I (2014) Amplified mid-latitude planetary waves favour particular regional weather extremes. Nat Clim Change 4(8):704–709
Seneviratne SI, Zhang X, Adnan M et al. (2021) Weather and climate extreme events in a changing climate. In: Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, USA, pp 1513–1766
Sousa PM, Trigo RM, Barriopedro D et al (2018) European temperature responses to blocking and ridge regional patterns. Clim Dyn 50(1):457–477
Stefanon M, D'Andrea F, Drobinski P (2012) Heatwave classification over Europe and the mediterranean region. Environ Res Lett 7(1):014,023
Tawn JA (1990) Modelling multivariate extreme value distributions. Biometrika 77(2):245–253
Thompson V, Kennedy-Asser AT, Vosper E et al (2022) The 2021 Western North America heat wave among the most extreme events ever recorded globally. Sci Adv 8(18):eabm6860
Touchette H (2009) The large deviation approach to statistical mechanics. Phys Rep 478(1–3):1–69
Trenberth KE, Fasullo JT (2012) Climate extremes and climate change: the Russian heat wave and other climate extremes of 2010. J Geophys Res Atmos 117(D17)
Vitart F, Robertson AW (2018) The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. npj Clim Atmos Sci 1(1):1–7

Vitart F, Cunningham C, DeFlorio M et al (2019) Sub-seasonal to seasonal prediction of weather extremes. In: Sub-seasonal to seasonal prediction. Elsevier, Amsterdam, pp 365–386

Wang P, Tang J, Wang S et al (2018) Regional heatwaves in China: a cluster analysis. Clim Dyn 50(5):1901–1917

Yaghmaei N (2020) Human cost of disasters: an overview of the last 20 years, 2000–2019. UN Office for Disaster Risk Reduction, Geneva

Zhang F, Biederman JA, Dannenberg MP et al (2021) Five decades of observed daily precipitation reveal longer and more variable drought events across much of the Western United States. Geophys Res Lett 48(7):e2020GL092,293

Zhang W, Hari V, Wang SS-Y et al (2022) Fewer troughs, not more ridges, have led to a drying trend in the Western United States. Geophys Res Lett 49(1):e2021GL097,089

Zschenderlein P, Fink AH, Pfahl S et al (2019) Processes determining heat waves across different European climates. Q J R Meteorol Soc 145(724):2973–2989

Zuo J, Pullen S, Palmer J et al (2015) Impacts of heat waves and corresponding measures: a review. J Clean Prod 92:1–12

5.3 Perspectives

The analyses carried out in this chapter show that the typical dynamics concept, i.e. the idea that *very* extreme events tend to look alike dynamically, is a relevant property of extremes, at least for surface temperatures in the long control run of the IPSL climate model. As a consequence — and somewhat paradoxically — even if one samples a few very extreme events, it may still be possible to have a correct first approximation of the atmospheric dynamics associated in so far as it is more likely than not that these events followed the typical dynamics. In other words, the burden of rareness is partially alleviated by this concentration property in the tail of the distribution.

As explained in the paper, there are nonetheless several discrepancies associated with the typical dynamics concept when one looks at different fields. It would therefore be important to validate this analysis on (i) other models and (ii) other observables. For example, even if the concentration of trajectories around a most probable one happens for all observables, the intensity of the extreme at which this happens may be so high that it lacks all practical relevance. This may for instance be the case for extreme precipitations, which can be associated to various atmospheric dynamics. Comparing how the typical dynamics — when it is relevant — changes from one model to another is another interesting avenue of research, especially in the context of intercomparison projects such as CMIP6. As explained in the introduction of this chapter, the concept of typical dynamics also paves the way for investigating rigorously how the dynamics leading to extreme events may change with climate change. It nevertheless should be mentioned that even though the typical, i.e. most probable, dynamics is identified with the average of paths in the phase space of trajectories reaching an extreme, there is no guarantee that this path is physical, i.e. that it can be reproduced by the model starting from optimal conditions. I think it is reasonable to assume that it is the case, at least up to small perturbations of the model. This issue is common and applies more generally to each analysis based on composites.

Finally, one of the initial motivation for this work was to investigate whether the dynamics leading to *very* extreme events was similar to the dynamics leading to 'regular' extremes. One could indeed see a shift to a different most probable path when imposing that events should be above a more extreme quantile of the observable studied. In other words, other physical mechanisms may have to be put into action to reach very extreme rather than extreme events: this is the idea behind so-called **dragon-kings** (Sornette and Ouillon 2012). Such a phenomenon has been observed in more simple systems and is usually called dynamical phase transition (Nyawo and Touchette 2017). My results on surface temperature in Europe in a pre-industrial configuration do not suggest such a phase transition: the very extreme events studied are similar to the extreme ones with more intense anomalies. This was also the conclusion reached by Gessner et al. (2021) using the ensemble boosting method. It may however be the case for other observables and/or other locations. Precipitations in the tropics are a good candidate: moderate extremes may be more likely reached by local thunderstorms but more intense extremes may need a dynamical phase

transition, such as a tropical cyclone for example (Peters et al. 2012).

Summary

Context and goals

How the dynamics leading to extremes should be understood in a general climatological context is not often discussed. Here I study the idea of a typical, most probable dynamics leading to extreme and very extreme events in the climate system that was proposed by previous works.

Methods

I use the 2000-y pre-industrial control run of a climate model and investigate how the dynamics of the model leading to extreme heat events at several locations in Western Europe change when one looks at more and more intense events. I propose a metric to quantify the typicality of these events.

Results

I show that, for most of these events, the variance of the dynamics leading to extremes tend to decrease as the intensity of the events considered increases. In other words, very extreme events tend to have a more similar dynamics than extreme ones. This idea supports the existence of a most probable path leading to extremes — the so-called instanton. I propose that the concept of the instanton should be considered as the climatological object defining the dynamics leading to extremes. For extreme heat events in this model, in Western Europe and under this configuration, my results also suggest that the dynamics leading to very extreme events corresponds to an amplification of the mechanisms leading to extreme events rather than the appearance of different mechanisms.

Chapter 6

Statistical and dynamical aspects of very extreme summers sampled in the IPSL-CM6A-LR climate model with a rare events algorithm

6.1 Introduction

In this chapter I implement the GKTL rare events algorithm — presented in chapter 4 section 4.4.3c and employed for the first time in climate science by Ragone et al. (2018)[1] — on the state-of-the-art IPSL-CM6A-LR climate model in an atmosphere-land configuration to sample extreme and very extreme summers. I address the question of the statistical and atmospheric dynamics characteristics of these summers. In particular, I detail what are the physical mechanisms which render the summers sampled extreme.

This chapter is organized as follows. In section 6.2 I first present the simulations made with the rare events algorithm and the tools employed to diagnose the atmospheric dynamics characteristics of the rare summers sampled. Section 6.3 shows the results obtained. The section begins by a statistical description of the extreme summers, then investigate the mean spatial structures of the centennial events sampled by the algorithm and finally details their atmospheric dynamics. In particular, I stress the dynamical role of cut-off lows to reach very extreme temperatures. Finally, section 6.4 summarizes the results obtained and discusses the advantages and drawbacks with regards to using the GKTL rare events algorithm to sample very intense extremes.

[1] I thank Francesco Ragone and Freddy Bouchet for our discussions on the use of rare events algorithms.

6.2 Methods

6.2.1 Rare events algorithm

I use the GKTL algorithm presented in chapter 4 section 4.4.3c. I briefly recall how the algorithm works and what kind of extreme events one may expect it would sample. The idea of the algorithm is to run an ensemble simulation of a climate model with N members. A member corresponds to one integration of the model, i.e. to one trajectory $t \mapsto X(t)$ in the phase space. Contrary to a regular ensemble, the rare events algorithm ensemble makes members interact during the course of the simulation. Between two resampling times t_i and t_{i+1}, I run the N ensemble members $(X_n)_{1 \leq n \leq N}$ in parallel. The integration is stopped at t_{i+1} and I compute a score function $F(X_n(t)) \in \mathbb{R}$, that allows to assign a weight $w_n(i)$ to each member n:

$$w_n(i) = \frac{e^{k \int_{t_i}^{t_{i+1}} F(X_n(t))dt}}{\frac{1}{N} \sum_{i=1}^{N} e^{k \int_{t_i}^{t_{i+1}} F(X_n(t))dt}} \tag{6.1}$$

where k is a control parameter which determines the strength of the selection. If k is positive (negative), the weight is high for high (low) values of the time-averaged score function F. The higher the absolute value of k, the more members with an extreme value of the time-averaged score function will be favored. Then each member is either killed or cloned depending on the value of its weight: the strongest the weight — and therefore the integral of the score function F — the highest the number of descendants of each member (see below for the discussion of the number of descendants per member). If the weight of one member is too small, i.e. if it performs poorly according to the score function F, there is a chance that the member is killed and therefore that it disappears from the ensemble. The number of descendants for each member is stochastic, but on average it is proportional to its weight. After several resampling times, the distribution of $\int F(X(t))dt$ in the ensemble is biased towards the right ($k > 0$) or left ($k < 0$) tail of the climatological distribution.

Here I employ the same resampling procedure as in Ragone et al. (2018), see chapter 4 section 4.4.3c for the detail. Note that the total number of members N stays constant during the full process, therefore at each resampling time the number of killed members is equal to the number of cloned members. In the simulations presented here the resampling is done every 5 days and a random perturbation on the potential temperature field is added to let the cloned members diverge from their parent as the model is fully deterministic. The random perturbation is computed as follows: $\eta r_{i,j,k} \theta_{i,j,k}$ on every value $\theta_{i,j,k}$ of the potential temperature at each grid point and levels indexed by i, j, k, where $r_{i,j,k}$ is a random number uniformly distributed in $[-1, 1]$ and $\eta = 10^{-4}$. This implies that the perturbation is at maximum of the order of 0.01% of the initial value. The typical precision obtained with climate models are much lower than this value, therefore in the following I consider that all simulations obtained are physical, i.e. that they could have been reached by the

model and are not an artifact of the random perturbations added[2].

The choice of the score function is crucial because it determines what kind of extremes are selected. By construction *the GKTL algorithm does not favor members that will experience an extreme event, but select retrospectively members which have experienced an extreme event* in the last resampling period. Nonetheless, by the choice of the score function and the physics of the system, favoring members which have experienced an extreme may actually favor the appearance of other similar extremes in the future. This is especially the case for heatwaves and the associated feedback loop with soil moisture: heatwaves favor soil desiccation which may favor later more intense heatwaves under the right atmospheric conditions (cf. chapter 1 section 1.3). Here I choose four different score functions:

- 2-m air temperature in °C (T2M) at grid point 1 (49.5°N, 2.5°E) with control parameter $k = 0.04/°C/day$,
- geopotential height at 500hPa (Z500) at grid point 1 with control parameter $k = 0.0015/m/day$,
- upper-level soil moisture at grid point 1 (SM1) with control parameter $k = -0.05/kg.m^2/day$,
- upper-level soil moisture at grid point 2 (SM2) (49.5°N, 7.5°E) with control parameter $k = -0.05/kg.m^2/day$.

In the following, I mention these simulations as the 'biased simulations' — biased in the sense that they favor a kind of extremes. I call them according to the variable they extremize, respectively biased T2M, Z500, SM1 and SM2. For each score function, I run 9 independent simulations using $N = 100$ members each. The simulations are independent in the sense that there is no inter-simulation interaction between the members. Because at each resampling time the number of descendants of each member is stochastic and the added perturbations are also stochastic, each simulation is different. There is a trade-off between the number of members per simulation and the number of simulations one can run (for a fixed computational cost). The number of members cannot be too low else the algorithm would be biased or have a high variance, but if it is too high then one needs to reduce the number of simulations made, which may also increase the variance. I choose the 9 simulations $\times$ 100 members per simulation trade-off by trial and error. I additionally run a control ensemble of 900 independent members with the same starting conditions and a random perturbation for differentiation.

The rationale for using the first score function is straightforward: 2-m air temperature is the variable classically used to define heatwaves. I therefore aim to select longer and more intense heat events with the algorithm. The rationale for using the second and third score functions is to sample respectively atmospheric and surface conditions which are typically associated with intense heat events: by maximizing

[2] Note that I chose the η value as large as possible for trajectories to separate rapidly, but at the same time not too large to let each simulation still be physical.

the geopotential height at 500hPa I try to sample more anticyclonic structure and by minimizing[3] upper-level soil moisture I pre-condition dry soils which would favor more intense sensible heat fluxes and therefore heatwaves. The last score function used is not, contrary to the others, computed at the same grid point but at a slightly longitudinally shifted ones which corresponds to the grid point with the strongest correlation with hot summers at the (49.5°N, 2.5°E) grid point in the work presented in chapter 5.

Contrary to previous applications (Ragone et al. 2018; Ragone and Bouchet 2021) I choose grid point score functions rather than extended spatial averages. By doing so I seek to impose a minimum amount of constraint on the full system and recover the spontaneous temporal and spatial scales of heat events in the grid point studied. When using extended spatial averages, one bears the risk of mixing up the dynamics of heat events occurring on one side or on another side of the spatial domain. As a result, composite maps could combine several dynamics that do not occur at the same time for individual events.

If it is assumed that the distribution of the score functions are Gaussian, then one can estimate the value of the summer averaged anomaly a that are sampled by the algorithm (Ragone and Bouchet 2020):

$$a \simeq 2\tau_c \sigma^2 k \tag{6.2}$$

where τ_c is the autocorrelation time of the score function and σ is its (daily) standard deviation. With the values of k chosen here, for summer anomalies one has: $a_{T2M} = 2.7°\text{C} = 2.6\text{std}$, $a_{Z500} = 54\text{m} = 2.6\text{std}$, $a_{SM1} = -13\text{kg/m}^2 = -5.8\text{std}$ and $a_{SM2} = -11\text{kg/m}^2 = -4.7\text{std}$. The distributions of the score functions are actually not Gaussian (especially for soil moisture), but this approximation gives an order of magnitude of the results that can be expected.

Here however I did not choose the values of k using this approximation. The rule of thumb I employed was to take the first five days of the control simulation — i.e. the first resampling period — and find the value k for each score function so that the average maximum number of descendants over all members was around 5. The rationale for this rule is to ensure that even if one member has 5 descendants and each of its descendants also has 5 descendants, in a period of 10 days — which is around the Lyapunov time scale of the atmosphere and therefore the time taken by members to drift apart after adding the random perturbations —, one member cannot, on average, populate more than 25% of the full ensemble. This is to avoid that a member which would be very extreme at the beginning but then falls back very quickly to the climatology would reduce too much the diversity inside the ensemble. I therefore chose values of k which are as strong as possible while avoiding the over-selection problem which would lead to too much extinction.

As explained in chapter 4 section 4.4.3, rare events algorithms do not only produce physically coherent very intense events, they also allow to compute the unbiased — i.e. climatological — expectations for these events. For any observable ψ (i.e. a smooth enough function), the climatological expectation of ψ using the rare events

[3] $k < 0$ for this score function.

algorithm ensemble is:

$$\mathbb{E}_\mu[\psi] = \frac{1}{N}\sum_{i=1}^{N} w_n^{-1}\psi(X_n(t)) \tag{6.3}$$

where $\mathbb{E}_\mu$ is the expectation with respect to the climatological probability distribution μ, $w_n = \Pi_i w_n(i)$ and $X_n(t)$ is the state vector of member n at time t. It is important to note that even if the weights w_n are computed according to the score function F, the formula is valid for the expectation of *any* other observable ψ.

In particular, if one wants to know the climatological properties of events for which the score function is above a certain level a, one computes:

$$\mathbb{E}_\mu[\psi \mid F(X(t)) \geq a] = \frac{\mathbb{E}_\mu[\psi \times \mathbb{1}(F(X(t)) \geq a)]}{\mathbb{E}_\mu[\mathbb{1}(F(X(t)) \geq a)]}, \tag{6.4}$$

or in plain words, one computes the observable only for members that have reached the level a and divide by the probability to reach the level a ($\mathbb{E}_\mu[\mathbb{1}(F(X(t)) \geq a)] = \mathbb{P}_\mu[F(X(t)) \geq a]$). For the reasons explained above (cf. chapter 1 section 1.2.3), these expectations are unprecisely estimated as long as a corresponds to a high quantile of the distribution of F. With the rare events algorithm, there are more members that reach high quantiles, but they come with a certain weight. Therefore the expectations in this equation are computed using formula (6.3) and with the weights w_n. Note that when N_s rare events algorithm simulations are run, each give an estimation for $\mathbb{E}_\mu[\psi \mid F(X(t)) \geq a]$ using equation (6.3). They are then simply averaged to give a final estimation of $\mathbb{E}_\mu[\psi \mid F(X(t)) \geq a]$.

In the following, two kind of averages are shown:

1. rare events algorithm ensemble averages, i.e. regular ensemble averages with equal weights for every ensemble member even if the simulation has been biased to produce extremes. Such a statistic is 'technical' in the sense that it displays what kind of trajectories are sampled by the algorithm and should be taken carefully when trying to infer climatological results,

2. climatological averages corresponding to centennial events, i.e. choosing a so that $\mathbb{P}_\mu[F(X(t)) \geq a] \simeq 10^{-2}$ and applying formula (6.4).

In the second case, it is important to note that some members sampled by the rare events algorithm are therefore discarded (because $\mathbb{1}(F(X_n(t)) \geq a) = 0$). The results obtained using centennial events only or with a regular average on the biased simulations are very similar (not shown here).

How to obtain statistical significance statements with the outputs of the rare events algorithm simulations — especially how to combine the results from the different simulations and take into account the fact that some members are very correlated — is not clear currently in the literature. In the following I therefore mainly show results normalized by removing the average and dividing by the standard deviation obtained on the control simulation, and I take a high anomaly threshold to interpret the results (> 0.5 std in absolute value). For comparison, if one considers

summer averages, with 900 members the significance at 5% with a Student t-test is obtained as long as the normalized anomaly is greater in absolute value than 0.07 std.

6.2.2 Configuration of the model

I use the IPSL-CM6A-LR model (Boucher et al. 2020) in its atmosphere - land surfaces configuration (LMDZOR) under pre-industrial forcing conditions. The model has a horizontal resolution of 2.5° in longitude and 1.27° in latitude and 79 vertical layers. I start by selecting a year in a 2000-y control run of the fully coupled model. This year is selected randomly and does not present any particular feature. A 4-months spin-up simulation (January to April) of the uncoupled model with oceanic and ice conditions of the initial coupled simulation is then run to let the atmosphere adapt to the uncoupling with the oceans and to provide initial conditions for the rare events simulation in summer. Random perturbations are added at the beginning of May to create a 100-member starting ensemble. The perturbations are added in May to avoid having to simulate the 100 members from January to April while still letting enough time for the different members to separate until the beginning of the rare events simulation in June. All the simulations begin on the 1st of June from those same 100 members and end on the 28th of August (90 days). To create the ensemble, a similar perturbation is added on all potential temperature levels as explained previously. Except specified explicitly, the variables used in the following are daily means. The ensemble simulations are launched through a tool developed by Arnaud Caubel and Yann Meurdesoif (`https://zenodo.org/doi/10.5281/zenodo.7848738`)[4].

I have made the explicit choice of having similar oceanic conditions for all simulations. As a result, the expectations obtained in the following (either with or without the rare events algorithm) should be more rigorously defined as conditional expectations with respect to the oceanic conditions used. In principle I could have started from a larger set of initial conditions encompassing different oceanic conditions and claim to have reached 'true' climatological expectations. In practice, first it is much more computationally expensive to do so. Second, because there is the extinction issue with the rare events algorithm (cf. chapter 4 section 4.4.3), in the end of the simulation there are only a few different remaining starting members in the ensemble (see below). This is almost equivalent to a start from fixed oceanic conditions and I therefore prefer to make this choice explicit. In a second step, one could also apply the same procedure starting from different initial oceanic conditions to evaluate the oceanic impact on the extreme summers sampled.

6.2.3 Atmospheric dynamics diagnostics

In this chapter I seek to understand the atmospheric dynamics mechanisms leading to centennial-like extreme summers. To do so I employ metrics used in the meteorological literature to diagnose the transfer of energy from eddies to the mean flow (**E**-vector), Rossby waves dynamics and cut-off low presence. To diagnose wave-like

[4] I warmly thank them for their help in launching simulations with the IPSL model.

behaviors I follow the common practice in the literature (e.g. Jiménez-Esteve et al. (2022)) which considers high-level tropospheric winds between 300 and 200hPa. Here I employ the 200hPa level as it is a standard output of the model. This altitude is nonetheless a little bit too high because in the mid-latitudes it is almost in the stratosphere. I therefore also present the results using the wind at 500hPa in supplementary materials for comparison. For the simulations analyzed in the next chapter I do have access to vertical distributions of the wind and cross-sections show that the 200hPa wind field behaves similarly as the 300hPa wind field (cf. figure 7.16). In this section I detail how the diagnostic metrics are computed.

E-vector. The propagation and tilt of transient eddies can be diagnosed using the so-called **E**-vector (Hoskins et al. 1983; Trenberth 1986; Schemm et al. 2018):

$$\mathbf{E} = (E_x, E_y) = \left(\frac{1}{2}\overline{v^{*2} - u^{*2}}, -\overline{u^* v^*} \right) \tag{6.5}$$

where u^* and v^* are transient wind components and the overbar denotes a temporal average. The horizontal component E_x is proportional to the group speed of transient eddies and therefore diagnose the direction of propagation of wave energy. The meridional component E_y shows the tilt of transient eddies: equatorward (resp. poleward)-pointing **E** indicate anticyclonically (resp. cyclonically) tilted eddies, i.e. also anticyclonic (resp. cyclonic) wave breaking. Up to a good approximation, the divergence of **E** corresponds to the transfer of zonal momentum from the eddies towards the mean flow, and the reverse for convergence. Here the transients u^* and v^* are defined as anomalies with respect to a 7-day rolling average of the wind fields at 200hPa. It is common to use 6-hourly data and consider the daily average to compute the overbar (Schemm et al. 2018; Riboldi et al. 2022). Here I only have access to daily data therefore the overbar is approximated by computing a 3-day rolling average of $v^{*2} - u^{*2}$ and $u^* v^*$.

Space-time spectral analysis. The characteristics of Rossby waves are diagnosed by performing a longitude-time Fourier decomposition of the meridional wind anomaly field at 200hPa (Randel and Held 1991; Riboldi et al. 2022). The anomalies are computed with respect to the daily ensemble average of the control simulation. With the temporal and spatial resolution of the model, it is possible to resolve harmonics of minimal zonal wavelength of 5° and minimal frequencies of 2 days. This somewhat limits the precision of the analysis proposed here but is still sufficient to resolve the largest Rossby waves of synoptic scales (> 1000km), which are the most relevant for the dynamics of heatwaves. For the sake of representation, in the following the results are interpolated below these two limits.

At each latitude ϕ, the meridional wind anomaly field is decomposed as a linear superposition of monochromatic zonally propagating waves. The Fourier coefficients $\hat{V}(k_\phi, \omega)$ with zonal wavenumber $k_\phi \geq 0$ and angular frequency ω are given by:

$$\hat{V}(k_\phi, \omega) = \frac{\sqrt{2\Delta t}}{N_\lambda \sqrt{N_t}} \sum_{n_t=1}^{N_t} \sum_{n_\lambda=1}^{N_\lambda} V(n_t, n_\lambda) e^{-i(\omega n_t \Delta t + 2\pi k_\phi n_\lambda / N_\lambda)} \tag{6.6}$$

where $V(n_t, n_\lambda)$ is the meridional wind anomaly at longitude n_λ and time n_t, $N_\lambda = 144$ is the number of longitudes, $N_t = 90$ is the number of days in the summer and $\Delta t = 1d$ is the temporal resolution of the simulation. The periodogram $P(k_\phi, \omega)$ is obtained by computing the square of the modulus of the Fourier coefficients $\hat{V}(k_\phi, \omega)$ and applying a smoothing Gaussian kernel in the ω dimension. Finally, the power spectral density $\rho(k_\phi, c_p)$ in the wavenumber k_ϕ-phase speed c_p space is recovered as:

$$\rho(k_\phi, c_p) = P(k_\phi, \omega) \frac{k_\phi}{2\pi R_E \cos\phi} \tag{6.7}$$

where R_E is the radius of the Earth and the results are interpolated from $c_p = -30m/s$ to $c_p = 30m/s$ in steps of 1 m/s.

This procedure is applied for all latitudes between 35° and 65°N and average the results to obtain a space-time spectrum for the summer. The ensemble average is then computed simply as the average over ensemble members of this procedure (except for the biased simulations for which to each member is applied the weight corresponding to centennial events as explained above).

Additionally, at all latitudes the enveloppe of Rossby wave packets is calculated as the modulus of the complex-valued Hilbert transform of meridional wind anomalies at 200hPa (Zimin et al. 2003).

Amplitude-phase decomposition. To identify the phase-shift of quasi-stationary Rossby waves, a similar procedure as Jiménez-Esteve et al. (2022) is employed on the meridional wind at 200hPa. The area-weighted latitudinal mean between 35° and 65°N is first computed and a 7-day running mean is applied to filter out transient eddies. Contrary to Jiménez-Esteve et al. (2022), I present results with and without removing the climatological mean of the area-weighted latitudinal mean, computed here using the control simulation ensemble average. Even though anomalies have no preferred phase in the control simulation (by definition of anomalies), this may not be the case in the biased simulations. I intend to use this analysis to diagnose the shift in the amplitude and/or phase of quasi-stationary Rossby waves between the control and biased simulations.

The temporally and spatially averaged field $\tilde{V}_{200}$ obtained is then decomposed into its Fourier components:

$$\tilde{V}_{200}(\lambda, t) = \sum_{k_\phi=0}^{+\infty} A(t) \cos(k_\phi \lambda + \Phi(t)) \tag{6.8}$$

with $A(t)$ the amplitude and $\Phi(t)$ the phase. In the following the 2D amplitude A – phase Φ histograms is computed to diagnose the shift occurring for the extreme summers sampled by the rare events algorithm.

Cut-off low frequency. Several algorithms exist to detect the presence of so-called cut-off lows — i.e. isolated minima of potential vorticity (PV) or geopotential height in the mid- to high-level troposphere — especially based on PV anomalies

(Wernli and Sprenger 2007; Favre et al. 2012; Pinheiro et al. 2017). Here the presence of 500hPa cut-off lows is diagnosed using an adaptation of the detection algorithm proposed by Muñoz et al. (2020). The algorithm does not rely on PV, and is better adapted to mid-level systems which are physically more relevant in the context of surface heatwaves. A cut-off low at 500hPa is considered to be present at a particular grid point if all of the following criteria are fulfilled:

1. Local geopotential height minimum: the 500hPa geopotential height of the grid point is at least 10m lower than the geopotential height in at least six of the eight surrounding grid points,

2. Isolation from the main westerly wind: there is an easterly flow in at least one of the four grid points located poleward of the grid point,

3. Cold core and thickness ridge eastward of the low: the difference between the geopotential height thickness between 500 and 850hPa at the grid point and at its immediate eastward neighbor is positive,

4. Frontal zone on the eastern flanck of the low: the thermal front parameter (TFP) on the immediate eastward neighbor of the grid point is negative. The TFP is computed as the change of the temperature gradient in the direction of the temperature gradient (at 500hPa):

$$\mathrm{TFP} = -\nabla|\nabla T| \cdot \left(\frac{\nabla T}{|\nabla T|}\right). \tag{6.9}$$

With the outputs of the model, the algorithm tends to detect the presence of cut-off lows at high latitudes, in particular northward of the eddy-driven jet. Such systems may indeed fill all the preceding criteria but they do not really qualify as an isolated minimum of geopotential height in the middle-high troposphere. Another criteria could have been added to impose that the grid points of cut-offs are below the jet position to correct for this behavior. However, as I am more interested in the dynamics of cut-off lows situated well below the jet, this error does not impact strongly the analysis.

6.3 Results

6.3.1 Grid point statistics

This section presents statistics for the grid points where the score functions are extremized. Figure 6.1 shows the empirical probability density functions (PDFs) of summer averaged 2-m air temperature (Fig. 6.1a), geopotential height at 500hPa (Fig. 6.1b), upper-level soil moisture (Fig. 6.1c) at grid point 1 and upper-level soil moisture at grid point 2 (Fig. 6.1d). When compared to the control PDFs, the PDFs of the score functions that are extremized by the rare events algorithm — i.e. for example T2M for the biased T2M simulation — demonstrate the efficient sampling of the tail of the summer averages distribution. For all score functions the rare events

algorithm sample extreme summers that are unprecedented in the 900-members control simulation. The choice of score functions closely related to the dynamics of heatwaves results in the sampling of extreme summers also for observables for which the algorithm does not extremize — i.e. for example Z500 for the biased T2M simulation. Figure 6.1b shows that the far tail of the extreme Z500 distribution is better sampled when maximizing for temperature rather than directly Z500, at least in grid point 1. This result is unexpected in so far as the Gaussian approximation presented above suggested an expected anomaly of the same order of magnitude for both simulations ($a_{T2M} = 2.7$std vs $a_{Z500} = 2.6$std). For the interpretation of Figure 6.1c and d, one should note that there is a hard coded limit of 8.14 kg/m^2 in the model for the upper-level soil moisture for the grid points considered. Therefore, summer averages close to this value correspond to extremely dry summers, almost the dryest that are possible in the model[5]. Finally, Figure 6.1a and b show that minimizing soil moisture at grid point 2, i.e. remotely with respect to where the T2M and Z500 variables are maximized, is more efficient for sampling more intense summer averaged T2M and Z500 than minimizing soil moisture at grid point 1. I come back below to this shift between soil moisture and temperature anomalies for extreme summers.

The time series of daily distributions presented in Figure 6.2 show how extreme are the summers sampled. Even the ensemble mean of the biased simulations is extreme compared to the control simulation: Figure 6.2a, c and d show that the ensemble mean in the middle of the summer is close to the 95% quantile of the control simulation and therefore the 95% quantile (or 5% quantile for soil moisture) of the biased simulations samples events that are seldom sampled by the control simulation. When maximizing for Z500 (Fig. 6.2b), the ensemble mean is higher but not extremely higher than the control ensemble mean. Figure 6.2a could give the false impression that the extremely hot summers sampled by the algorithm are constituted by members which are individually always moderately above the control mean and that extreme summers are reached this way. This is however not the case, and on individual time series (see Figure A.2a in supplementary materials for an example), there is a large variability in the temperature time series. The typical pattern is more of an alternation between intense heatwave events and returns to the climatological mean.

Figure A.2 also shows the extinction problem that has been mentioned before: once the simulation has run and the members are reconstituted from the 1st of June, only a few initial members have survived (4.4 on average). The evolution of the number of effectively different members after the simulation has run is shown in Figure 6.3. It illustrates the extinction issue: up to the middle of the simulation (mid-summer), there are only around 10 effectively different members. This shows why one needs to run several different 100-member simulations with the same parameters of the algorithm. One should also not that a selection is done at the last step of the algorithm which explains why there are around 65 and not 100 effectively

[5] This hard coded limit is justified physically by the fact that not all water contained in the soils can be used by plants for evapotranspiration, and a limited, but positive, quantity of water is always present.

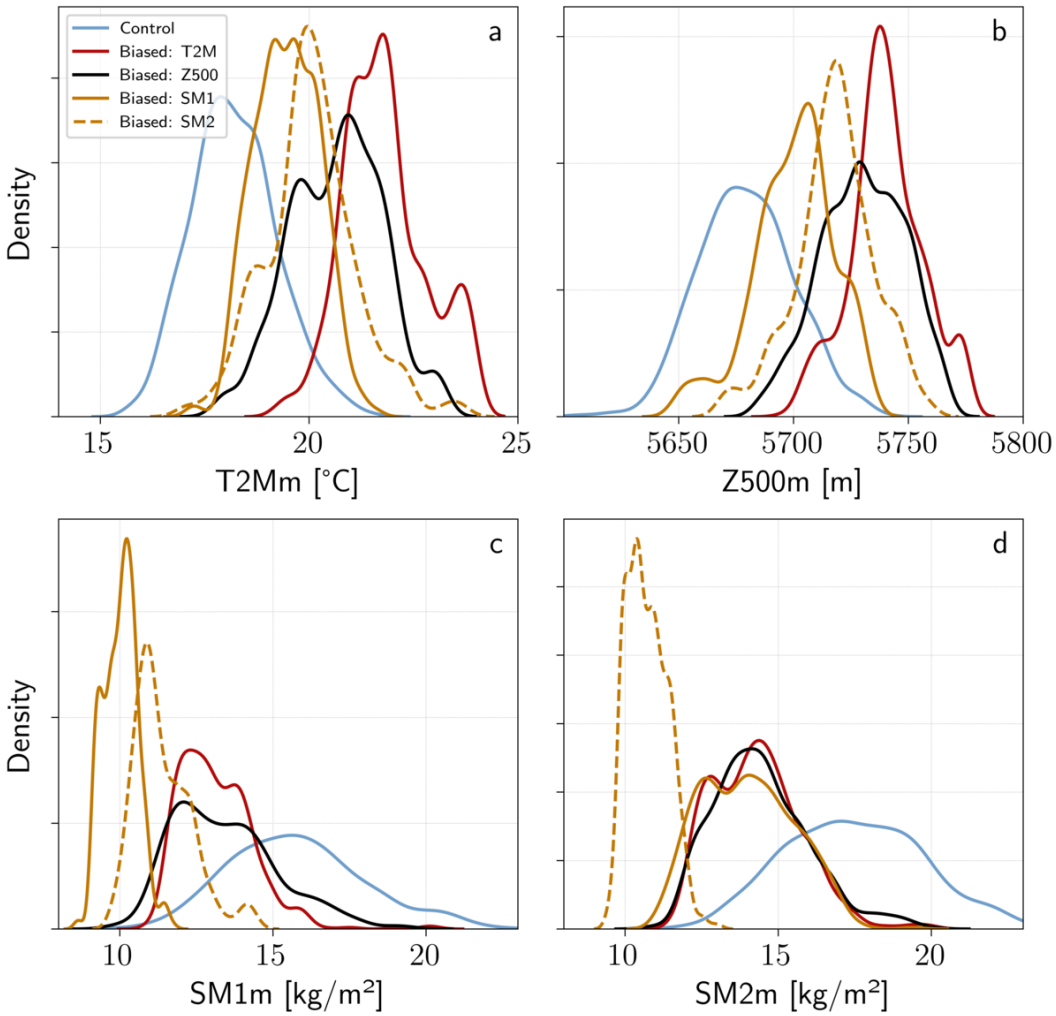


Figure 6.1: **Empirical probability density functions (PDFs) of summer averaged grid point observables.** PDFs of the summer averaged (a) 2-m air temperature, (b) geopotential height at 500hPa, (c) upper-level soil moisture at grid point 1 and (d) upper-level soil moisture at grid point 2. The PDFs are obtained by a kernel smoothing of the empirical histograms.

different members at the end of the simulation. This ratio of 65 surviving members vs 35 killed members is the approximate ratio at each resampling time during the simulation.

This issue of the substructure (Röthlisberger et al. 2020) of extremely hot summers is explored further in my simulations by counting the number and mean length of heatwaves during the summer. A heatwave is defined classically as a continuous period of time of at least 3 days for which the daily mean temperature is above the 90% daily climatological quantile of the temperature distribution. The latter is computed using the control simulation. All such events are computed at grid point 1. Figure 6.4 depicts the result of this computation. Figure 6.4a shows that for

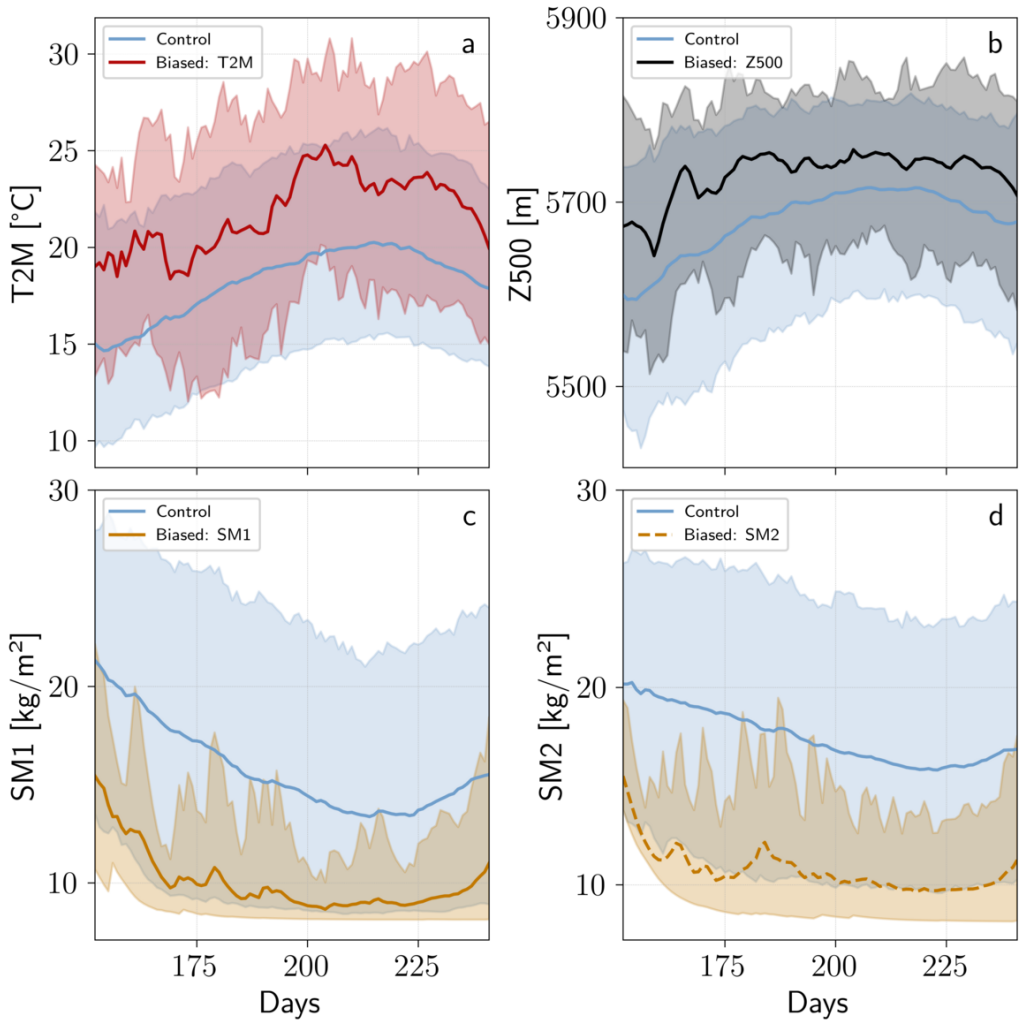


Figure 6.2: **Time series of the ensemble mean and the 5 and 95% quantiles of grid point observables.** (a) 2-m air temperature, (b) geopotential height at 500hPa and (c) upper-level soil moisture at grid point 1 and (d) upper-level soil moisture at grid point 2. For all the plots the plain/dashed line shows the ensemble mean and the shading the 5 and 95% quantiles of the variable displayed. For the biased simulations, all members of all simulations are pooled together to compute the 5 and 95% quantiles.

biased T2M simulations the mean number of heatwaves in the summer goes from around 1.3 in the control to around 4.8 in the biased simulation, while their mean duration goes from 4.6 to 7.3 days. The algorithm therefore selects both more heatwave events and longer heatwave events. When maximizing the geopotential height (Figure 6.4b) the mean duration increases to 5.7 days while the number of events increases to less than 3.7, illustrating the key role of high geopotential anomalies to reach high temperatures. When minimizing the local soil moisture (Figure 6.4c) on the other hand, the mean duration of heatwave events does not change much with

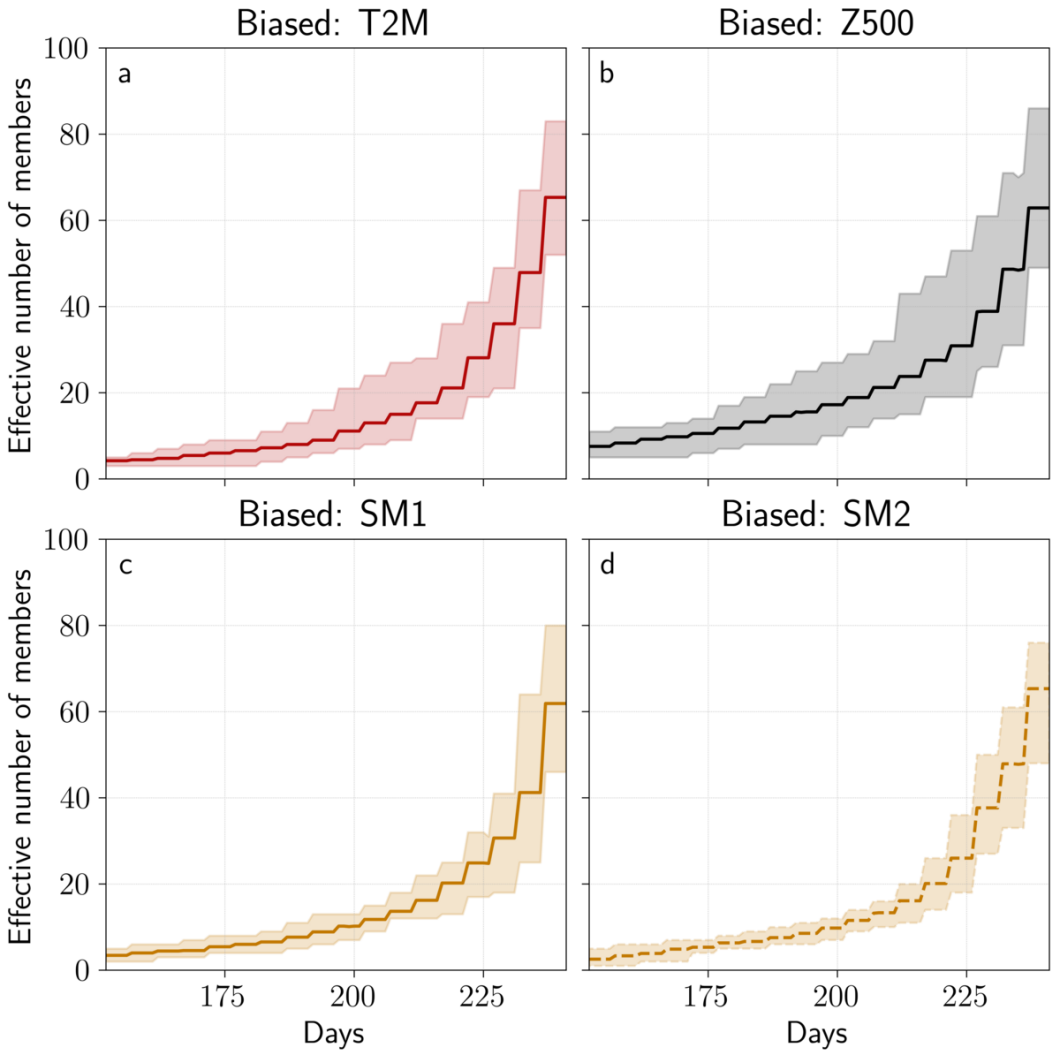


Figure 6.3: **Evolution of the number of effectively different members during the simulation.** The effective number of members represents the number of members in each 100-member simulation which are truly unique at time t. The computation is done after the simulation is run and the members are reconstructed from the 1st of June, hence the increasing function. The jumps show the moments when the resamplings occur. The shadings show the minimum and maximum of the effective number of members over the $N_s = 9$ simulations.

respect to the control (4.8 vs 4.6), but the number of events increases to around 2.3 per summer. When minimizing soil moisture at grid point 2 however, there is an increase in both the duration (6.2 days) and the number of heatwaves (3.0 per summer). Neither the biased Z500 nor the SM1 and SM2 biased simulations are enough to reach metrics as high as the ones of the biased T2M simulation. This results is strongly in favor of the need for combined non-local dry soils and high local geopotential height to obtain long and intense heatwaves.

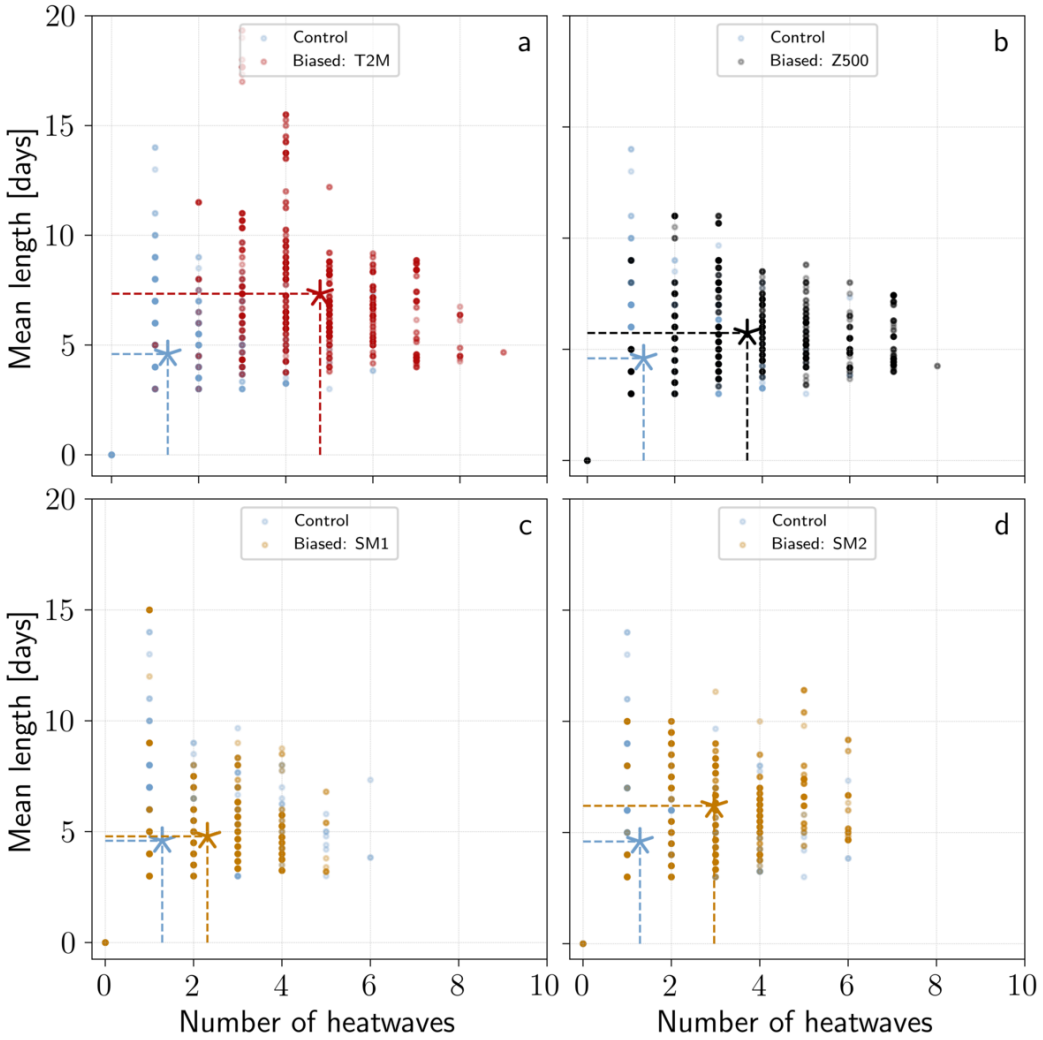


Figure 6.4: **Number of heatwaves per summer and mean length.** All heatwaves are computed at grid point 1. For all plots a heatwave is a continuous period of time of at least 3 days when daily mean temperature is above the 90% climatological quantile. The stars show the ensemble mean.

As explained above, the algorithm provides weights to compute climatological averages of observables of interest. This allows in particular to recover the probabilities of extreme summers. Figure 6.5 shows the probabilities for the simulations run here, each using their own score function as the observable of interest (i.e. using the observable on which the rare events algorithm is expected to be the most efficient). For the control simulation, the shadings represent the uncertainty on the estimation of the probability. The uncertainties are computed using bootstrap on the original 900 members and the 5–95% quantiles obtained are shown. For the biased simulations, the shadings show the minimum and maximum estimation of the probability over the $N_s = 9$ simulations. All simulations sample very extreme summers, with probabilities ranging from 10^{-2} to 10^{-6}, i.e. return times between one hundred and

one million years. This demonstrates a large efficiency gain compared to the control simulation, which, with 900 members, can only sample precisely summers with a 100 years return time. The variations from one simulation to another of the estimated probability are nonetheless important, one order of magnitude on average. Although the algorithm is asymptotically unbiased, one could wonder whether such low probabilities sampled here are really correct. The only way to prove this is the case would be to run a much larger number of control members, until even very low probability events happen — but this entails a corresponding high computational cost. Although this is not a formal proof, one can see in Figure 6.5 that the probabilities given by the rare events algorithm are in the continuity of the control simulation probabilities and therefore the two results seem at least coherent on the region where they overlap.

Insofar as summer averages (i.e. 90 days averages) are considered, by the central limit theorem, one may expect that the associated distribution is Gaussian. However, the central limit theorem is valid only in the vicinity of the average (Touchette 2009) and the tails of the distribution — i.e. the regions sampled here — must be described by large deviations theory (Gálfi et al. 2021). To illustrate the deviations from Gaussianity, the mean and standard deviation of a Gaussian distribution are fitted on the control run members and it is shown in dashed line in Figure 6.5 the associated extrapolation of the probability. Although this extrapolation is correct up to probabilities as low as 10^{-1}, it diverges for lower probabilities. Figure 6.5a for example shows that an extreme summer with average temperature around 23°C are 10 times more likely than predicted by the Gaussian approximation. The Gaussian approximation is nevertheless on the range of uncertainty of the algorithm when running $N_s = 9$ simulations. For the other simulations (Fig. 6.5b, c and d), this is the contrary: extreme summers are less likely than predicted by the Gaussian. This is especially clear for soil moisture, which is, contrary to the Gaussian distribution, bounded downwards in the model as explained above. The uncertainty ranges for these simulations also make clear that the tail of the summer average quantities are not Gaussian.

It should be noted however that the performance of the algorithm rapidly deteriorates when one tries to estimate the probabilities of observables that are not closely related to the summer averaged score function. Figure A.3 in supplementary materials shows a similar computation of probabilities as in Figure 6.5 but for low soil moisture at grid point 1 for biased T2M and Z500 simulations (Fig. A.3ab) and for high 2-m air temperature at grid point 1 for biased SM1 and SM2 simulations (Fig. A.3cd). The estimations are incompatible with the control estimations by one to two order of magnitudes. Although the GKTL algorithm is unbiased in the limit of an infinite number of members simulated, with $N = 100$ members its performance is better than the control simulation only when one focuses on the observables closely related to the score function. I come back to this issue in the discussion section.

To explore the link between the summer averaged quantities in the different simulations, a scatter plot of their cross distributions is shown in Figure 6.6. Figure 6.6b and d show the strong link between high summer-averaged geopotential heights and both high surface temperature and low soil moisture (although the correlation

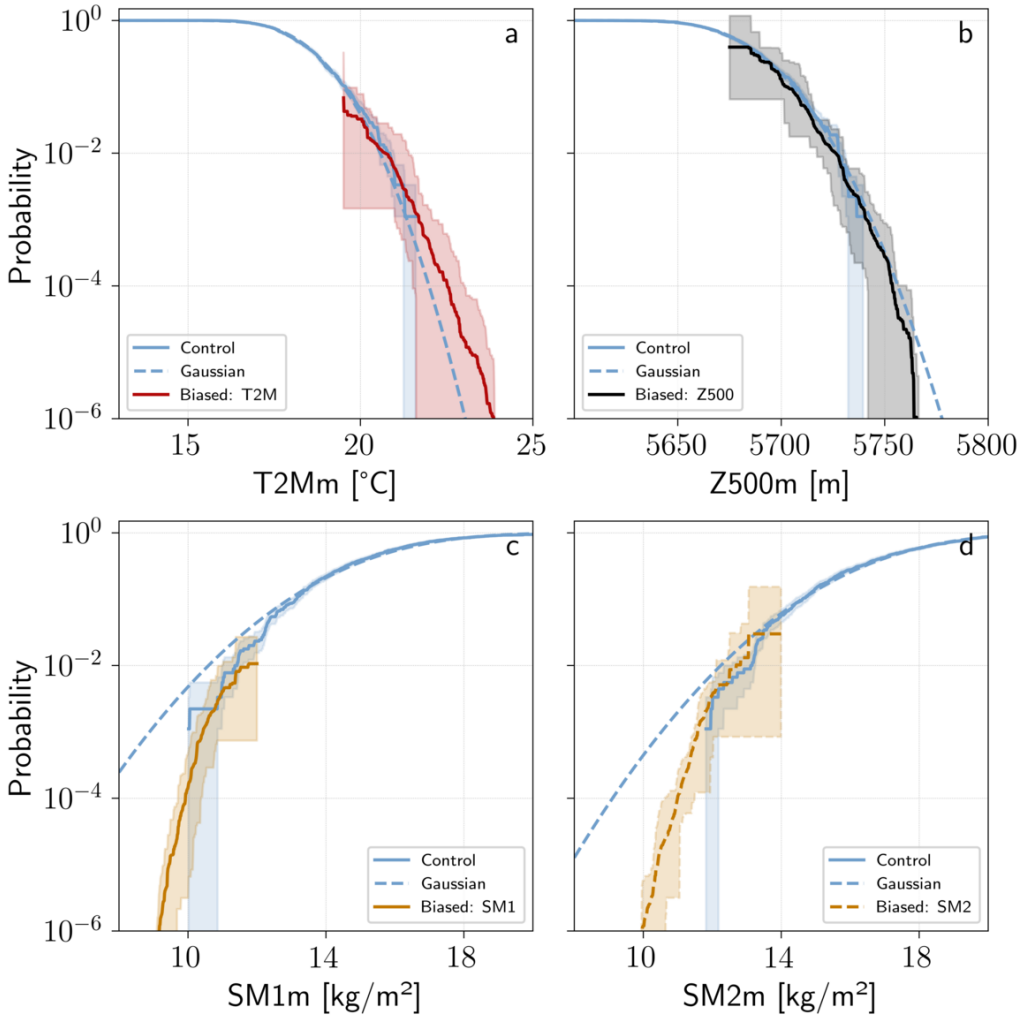


Figure 6.5: **Probabilities of an extreme summer according to the naive estimator (control simulation) and the rare event estimator.** (a) Probability of the summer averaged 2-m surface temperature and (b) probability of the summer averaged geopotential height at 500hPa to be above a certain threshold at grid point 1. (c) Probability of the summer averaged upper-level soil moisture at grid point 1 and (d) probability of the summer averaged upper-level soil moisture at grid point 2 to be below a certain threshold. The blue dashed lines represent the corresponding probabilities for a Gaussian distribution fitted on the control simulation by the method of moments. The shadings for the control simulation show the 5–95% quantiles of the estimated probability obtained using bootstrap on the 900 members. For the biased simulations the shadings show the minimum and maximum of the estimated probability over the $N_s = 9$ simulations.

is not as strong in the second case). This link is conserved in the biased simulations, which, when maximizing for either temperature or geopotential height, sample the

tail of the joint PDFs of these observables. As illustrated by Figure 6.6a and c, the link between soil moisture and temperature is however not as obvious. Although in the control simulation there is a small correlation between those two quantities, the correlation is lost in the biased simulations. Very hot summers sampled by the rare events algorithm are dryer than the average (see also Figure 6.1c and d), but there is a wide range of different possible temperatures for the same level of dryness. A similar phenomenon is observed for very dry summers. Interestingly, contrary to geopotential height and surface temperature in Figure 6.6b, the algorithm does not sample the combined tail of the PDFs (except for a few members in the biased SM2 simulation): there are no both very hot and very dry summers. This suggests that the well-known correlation between these two variables may not be present in the far tail of their joint distribution. One hypothesis to explain such a surprising result could be that when one member reaches very high temperatures at grid point 1, this tends to trigger convection and therefore rain, which is not favored when one wants to minimize soil moisture. Thus, dry summers may have to be hot but not too hot.

6.3.2 Summer averaged spatial structures

In this section I investigate the summer averaged spatial structures associated to the extreme summers sampled by the algorithm. As explained in section 6.2, I consider here centennial-like events, i.e. extreme summers that have a probability below 10^{-2} to occur. Contrary to the preceding section, in this section I therefore show climatologically relevant averages, i.e. conditional on reaching a threshold corresponding to centennial events.

The normalized anomalies of upper-level meridional wind (200hPa) and 2-m air temperature are shown in Figure 6.7. The normalization is computed by removing the climatological mean and dividing by the climatological standard deviation estimated on the control simulation at each grid point. Figure 6.7a shows the results for the biased T2M simulation. In an approximately 1000km-diameter circle centered around grid point 1, the summer averaged temperature is 3 standard deviations above the climatology. However, this strong anomaly is concentrated close to grid point 1: although there are corresponding cold and warm anomalies downstream, those are much less intense than in Western Europe. The upper-level wind pattern is also concentrated close to Western Europe, with a large and very anomalous meander around grid point 1. The meander is tilted eastward over Eastern Europe, suggesting a recurrence of wave-breaking phenomena. Similarly, the positive anomaly above the East Mediterranean region is the result of a northward shift of the subtropical jet (not shown). Interestingly, the anomalies upstream are small and there are no hemispheric pattern contrary to what has been suggested by some authors (see chapter 1 section a)). The size of most anomalies is of 1 to 2 synoptic wavelengths. These observations extend to the other simulations, with the exception that the anomalies of both V200 and T2M are smaller when minimizing soil moisture either at grid point 1 or 2. Finally, there is a longitudinal shift of anomalies between the four simulations. For example, contrary to Figure 6.7a, in Figure 6.7c only the western facade of France and the Iberian peninsula present positive

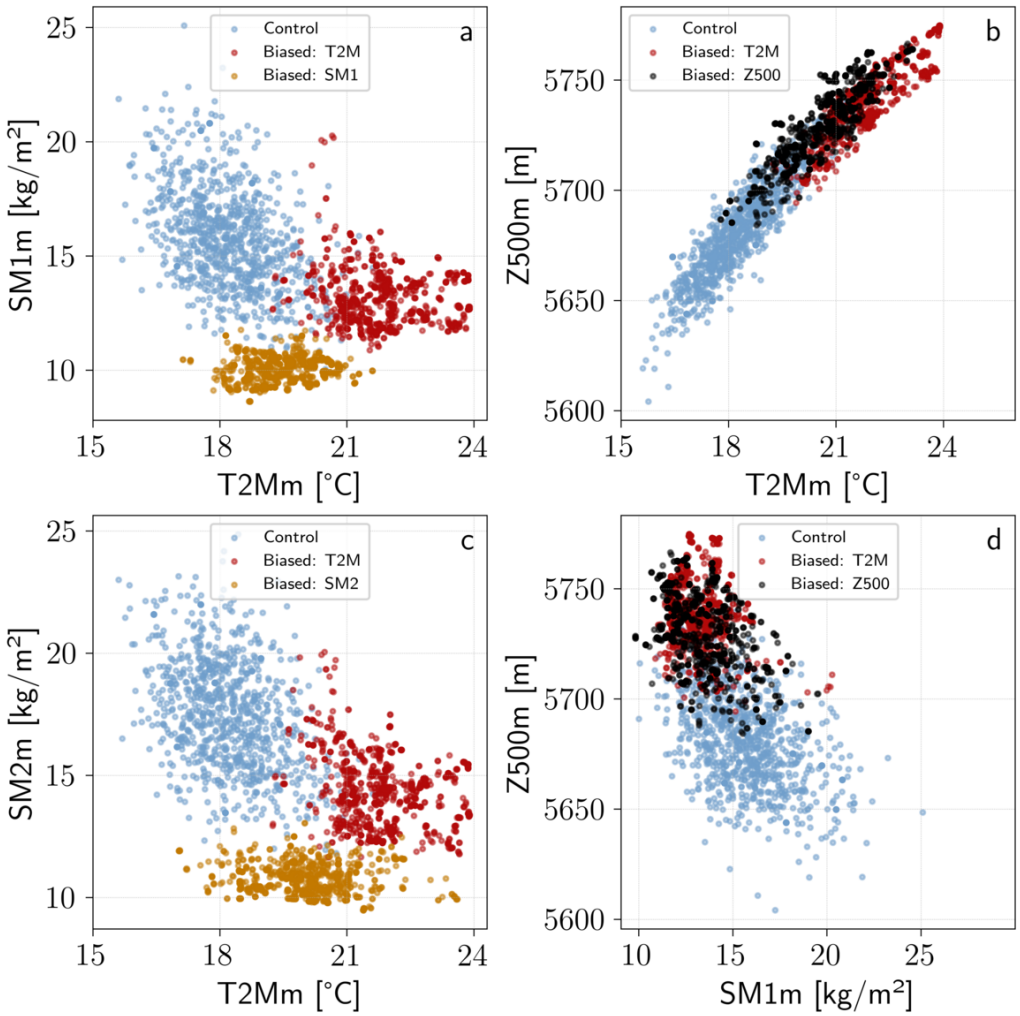


Figure 6.6: **Scatter plot of summer averaged grid point observables.** (a) Summer averaged 2-m air temperature vs. upper-level soil moisture at grid point 1. (b) Summer averaged 2-m air temperature vs. geopotential height at 500hPa at grid point 1. (c) Summer averaged 2-m air temperature vs. upper-level soil moisture at grid point 2. (d) Summer averaged upper-level soil moisture vs. geopotential height at 500hPa at grid point 1.

anomalies of T2M. There is a corresponding shift in anomalies of V200. It should be also mentioned that there are almost no significant anomalies in the tropics and not at all in the southern hemisphere for the fields displayed (not shown), which suggests a climatological atmospheric dynamics leading to extreme summers mostly confined to the mid-latitudes.

Figure 6.8 shows the average of the **E**-vector at 200hPa for the summer (see fig. A.4 for the **E**-vector at 500hPa). The anomalous southward pointing vector over Eastern Europe in Figure 6.8a shows the anticyclonic wave-breaking already mentioned. Interestingly, there is also a region of anticyclonic wave breaking west

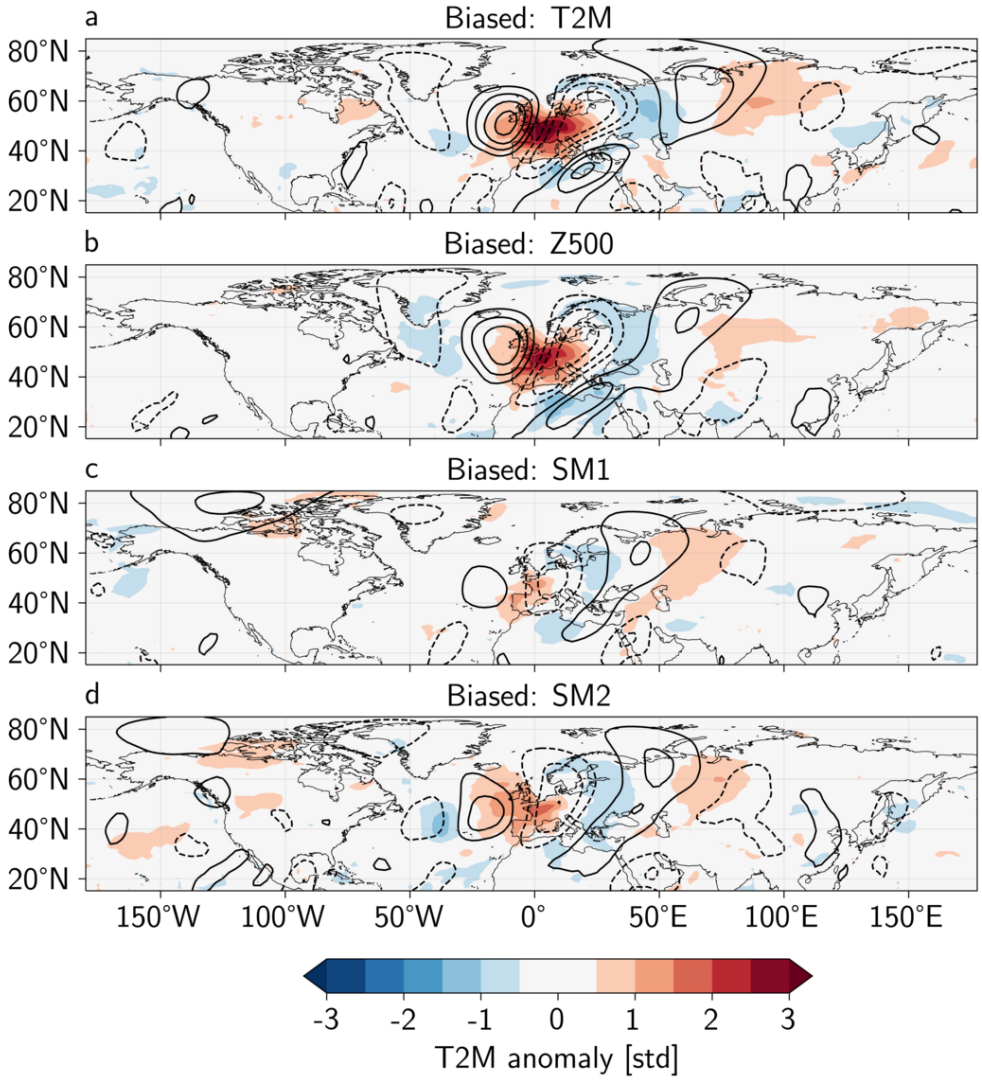


Figure 6.7: **Summer averaged normalized anomalies of 2-m air temperature (colors) and meridional wind at 200hPa (contours) for centennial events.** For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

of the Iberian peninsula. Especially, Figure 6.8b, c and d show slightly shifted anticyclonic wave breaking regions compared to the control simulation. I come back to this observation below. Finally, all simulations show less wave-breaking in the entrance of the storm track region, associated to a more zonal wave guide. For these extreme summers, waves tend therefore to break at preferred regions of the Euro-Atlantic sector, leading to the anomalies presented in Figure 6.7.

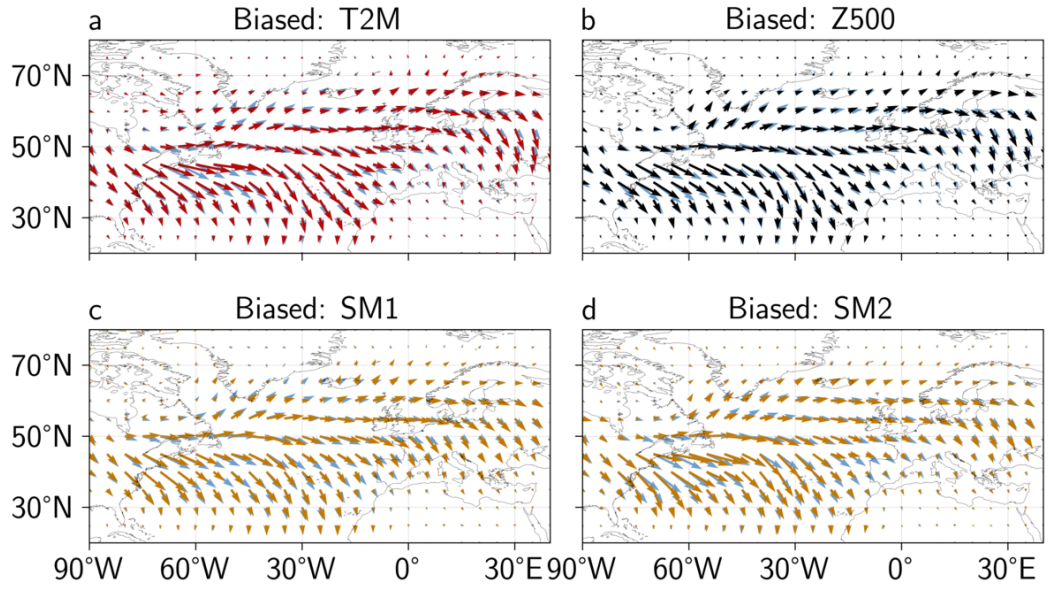


Figure 6.8: **Summer averaged E-vector at 200hPa for centennial events.** In each panel the blue arrows show the climatological **E**-vector computed on the control simulation.

Figure 6.9 shows the normalized anomalies of geopotential height at 500hPa and upper-level soil moisture. All simulations have a large anticyclonic anomalies above Western Europe, but the center of the anomalies is slightly shifted from one simulation to another: for Figure 6.9b, corresponding to the simulation maximizing Z500, as expected the anomaly is maximum above grid point 1, whereas it is shifted to the east for Figure 6.9a, to the west for Figure 6.9d and to the south-west for Figure 6.9c. Again, the Z500 anomalies are maximum above Western Europe and do not extend around the entire Northern Hemisphere. It should be also noted for all regions, the small negative Z500 anomaly south-east of the anticyclonic region, which corresponds to the isolation of low pressure systems by anticyclonic wave breaking above Eastern Europe. The strongest soil moisture anomalies in Figure 6.9 are also restricted to the European region and present an eastward shift with respect to the region of maximum anticyclonic anomaly. For Figure 6.9a in particular, the anomaly of soil moisture at grid point 1 — i.e. where the temperature is maximized — is not as strong as the soil moisture anomalies encompassing Germany and Poland regions.

One hypothesis to explain this shifted pattern is related to the advection of dry and hot air at the core of the heatwave regions. Previous studies (cf. chapter 1 section a)) computing Lagrangian backward trajectories in reanalysis and climate model outputs have shown that most of the air parcels during heatwave events in Western Europe come from either the heatwave region itself or its immediate east. This would plead in favor of the shifted low soil moisture pattern to be a precursor of intense heat events at grid point 1 in so far as low soil moisture promote high sensible heat fluxes and diabatic warming. It is however difficult to know whether

this pattern is a cause or a result of heat events.

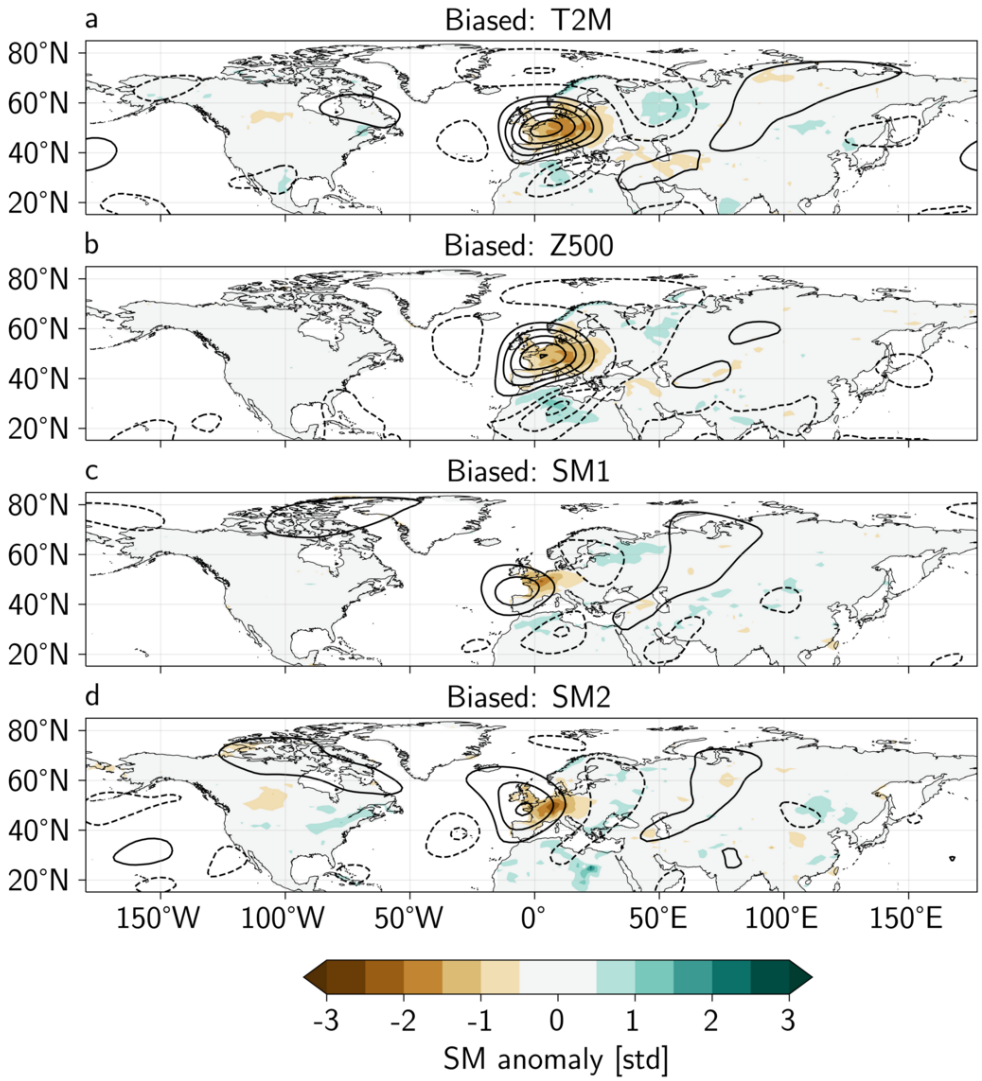


Figure 6.9: **Summer averaged normalized anomalies of upper-level soil moisture (colors) and geopotential height at 500hPa (contours) for centennial events.** For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

An idea of the origin of air parcel in a Eulerian framework can be given by analyzing the vertical structure of the anomalies of geopotential height and sea-level pressure, even though I did not run a Lagrangian particle tracing analysis which would be necessary to validate this result. Figure 6.10 shows the anomalies of sea-level pressure (SLP), geopotential height at 850hPa (Z850) and geopotential

height at 200hPa (Z200). All panels illustrate the westward tilted vertical structure of the summer anticyclone, with for example for Figure 6.10a the maxima of SLP anomalies situated above Poland while the maxima of Z200 situated above Southern UK. This vertical structure suggests, at grid point 1, advection from the east on the lower layers of the atmosphere, while from the south-west on the upper-levels. This reveals a non-barotropic vertical structure of the anticyclone leading to extreme temperature events, contrary to what is sometimes suggested in the literature (cf. chapter 1 section a)).

6.3.3 Atmospheric dynamics

In this section I investigate the atmospheric dynamics associated with extreme summers. Figure 6.11 shows the wavenumber-phase speed spectra at 200hPa obtained for centennial events and the difference with the climatological spectrum (see Fig. A.5 for the spectra at 500hPa). The spectra are different from one biased simulation to another. The biased T2M and Z500 simulations are similar in their increase in stationary and/or westward moving waves with low zonal wavenumbers (4 to 6), highlighting the persistence of anticyclonic blocking patterns. On the other hand, the biased Z500 simulation is associated with a decrease of fast-moving high zonal wavenumbers waves which is not the case for the biased T2M simulation. The results for soil moisture minimizing simulations is more surprising: for both of them there is an increase in stationary/westward-moving waves with high zonal wavenumbers (8 for both, 6 for SM1 and 5 for SM2) and a decrease of eastward moving waves with similar zonal wave numbers. The differences with the climatological spectrum are nevertheless small (around 10%) and the raw spectra are similar from one simulation to another.

I investigate the properties of slow-moving quasi-stationary waves through the amplitude-phase histogram detailed in section 6.2. The results are presented in Figure 6.12 for meridional wind speed anomalies at 200hPa for centennial events (see Fig. A.7 for meridional wind speed anomalies at 500hPa). Figure A.8 in supplementary material shows its aggregated version along the amplitude and phase axis (Fig. A.9 for the 500hPa level). Simulations with an extremization of the score function do not differ strongly in the quantity of energy distributed in the different zonal wave-numbers: they have approximately the same aggregated spectrum for each of them. However, they differ strongly in the phase of the different wave numbers. Figure 6.12b2 for example shows a 2-3 times increase in the frequency of waves with phase around $\pi/2$ compared to the control for wave number 4. There is a similar behavior for the other simulations but the preferred phase changes from a zonal wave-number to another. For example, for $k_\phi = 3 - 5$ biased simulations have a different preferred phase (which is coherent with the longitudinal shift that was described in Figure 6.7) but for $k = 6$ they all have the same preferred range of phases. The same decomposition using the raw meridional wind field at 200hPa is presented in Figure A.6 (cf. Fig. A.10 for meridional wind at 500hPa and Fig. A.11 and A.12 for the aggregated spectra). There is no clear shift in neither the amplitude nor the phase of the stationary waves in the biased simulations, except

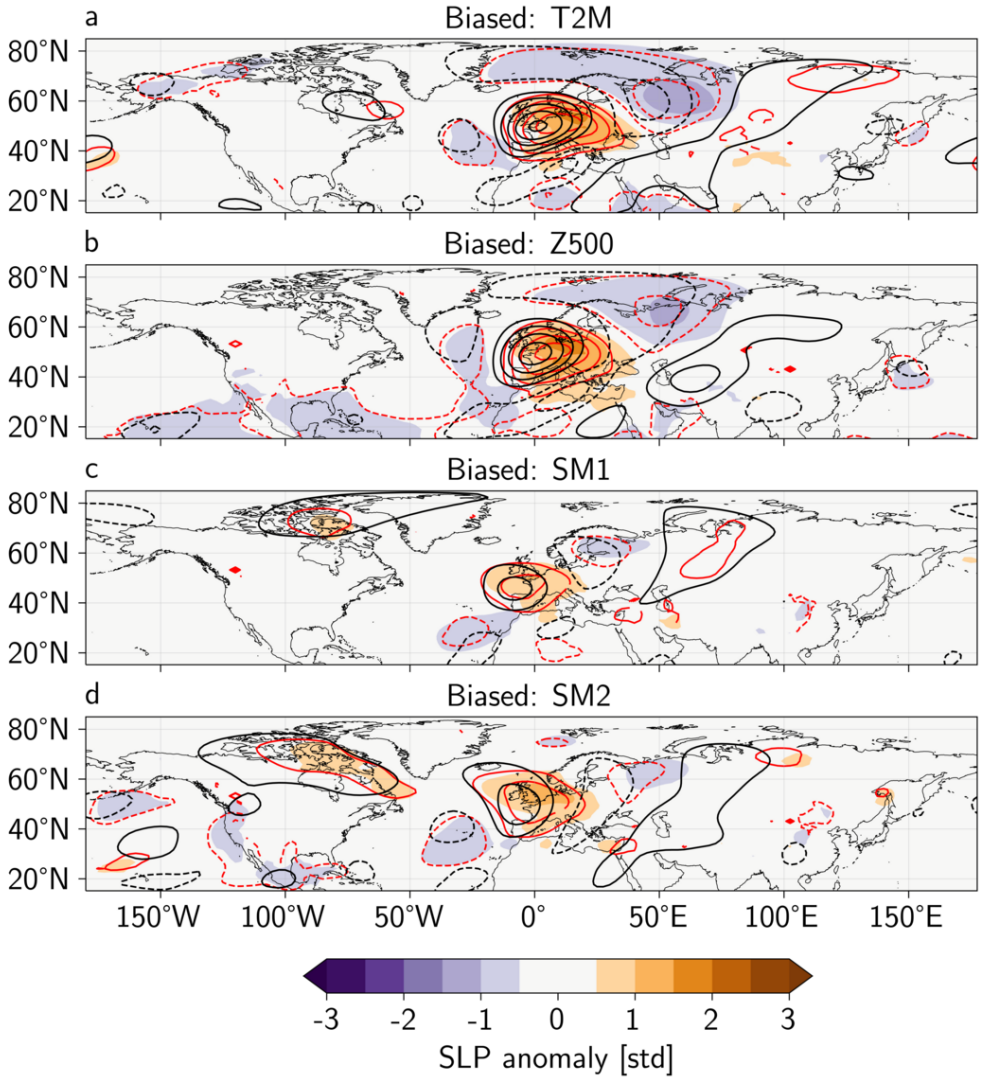


Figure 6.10: **Summer averaged normalized anomalies of sea-level pressure (colors), geopotential height at 850hPa (red contours) and at 200hPa (black contours) for centennial events.** For all fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

maybe for wavenumbers $k_\phi = 4$ and $k_\phi = 5$ for the biased T2M simulation (Fig. A.6b2 and b3). In the aggregated spectrum (Fig. A.11b–c2) a small shift in the phase for all simulations is seen for wavenumber $k_\phi = 6$.

Whether the excitation of certain modes and their phase-locking behavior is a result of an underlying physical amplification mechanism of the system or the simple consequence of the Fourier decomposition of anomalies leading to the patterns of

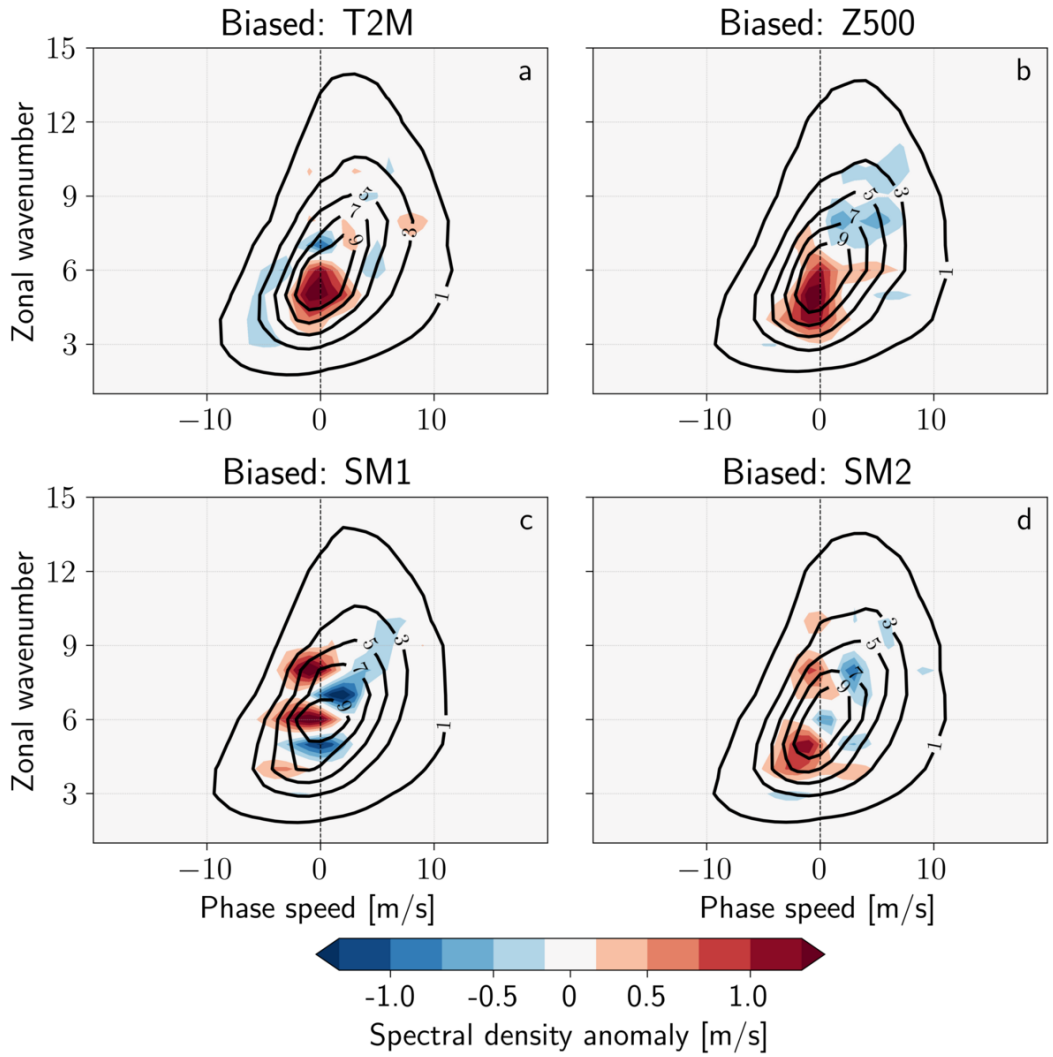


Figure 6.11: **Wavenumber-phase speed spectra of meridional wind anomalies at 200hPa for centennial events.** Raw spectrum in m/s (contours) and difference with the climatological spectrum (colors).

Figure 6.7 is not clear though. In particular, with this procedure it is assumed that the stationary hemispheric Fourier modes exist and are physically relevant. However, when the original field is reconstructed with for example modes $k_\phi = 3$ to $k_\phi = 10$ (not shown), it is found that they — as expected — correspond to a localized anomaly above Western Europe (i.e. the pattern of Figure 6.7). In other words, even though it is mathematically correct to decompose the field on hemispheric scale patterns, most of the dynamics is local and confined to the North Atlantic in practice. It is therefore difficult to interpret physically hemispheric modes (not to mention the fact that these modes only have a zonal and no meridional component).

On the other hand, Figure 6.13 shows the composite Hovmöller plot of anomalies of meridional wind speed at 200hPa averaged between 35°N and 65°N and normalized anomalies of 2-m air temperature averaged between 45°N and 55°N for the heatwave

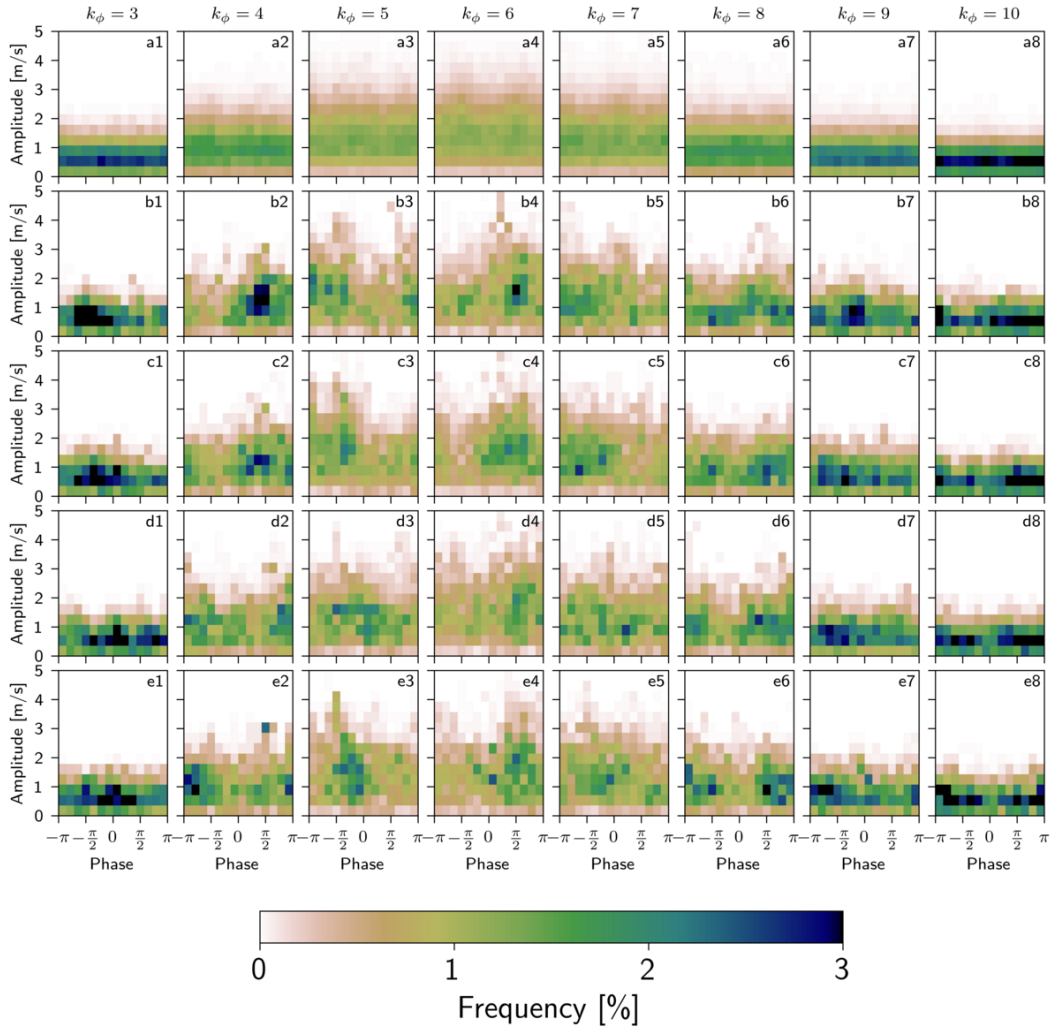


Figure 6.12: **Amplitude-phase histograms of meridional wind speed anomalies at 200hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) control simulation, (b) biased T2M simulation, (c) biased Z500 simulation, (d) biased SM1 simulation and (e) biased biased SM2 simulation.

events that occur in each simulations. As previously, a heatwave event is defined as a continuous period of time of at least 3 days for which the temperature is above the 90% climatological quantile. In all simulations the atmospheric dynamics is similar: there is a Rossby wave packet of 2-synoptic wavelength propagating to the east and amplifying locally in Western Europe. Because there are a lot of heatwave events in biased simulations, it is even possible to see the recurrence of new Rossby wave packets at 6-8 days after the heatwave (Fig. 6.13bce1). The temperature anomalies on the other hand are restricted to the region of amplification of the Rossby wave packet, i.e. where the score functions of the rare events algorithm are computed. For the biased T2M simulation (Fig. 6.13b2), the temperature anomalies are larger

during the heatwave and extend both before and after, which may be linked to a near zero phase speed of the Rossby wave packet. A similar mechanism, although less intense, may be happening for both the biased Z500 and SM2 simulations (Fig. 6.13ce2).

Indeed, when one looks at daily maps (see Figure 6.14 for two example hot days), the explanation in terms of hemispheric scale patterns is rather doubtful. Figure 6.14a1 for example shows the omega blocking-type structure above Western Europe that is amplified only locally, as shown by the envelope on Figure 6.14b1. On the other hand, one also observes not only wave-like but also vortex-like structures as in Figure 6.14a2 with an isolated minimum of Z500 — i.e. a cut-off low — west of the Iberian peninsula which is not embedded in a hemispheric Rossby wave nor a local Rossby wave packet pattern (Fig. 6.14b2).

I investigate further the role of these structures (as already mentioned in chapter 5). The frequency and frequency anomaly of cut-offs low over the Euro-Atlantic sector for the biased simulations is presented in Figure 6.15. All biased simulations have anomalously high frequency of cut-off lows over the Atlantic. Biasing for 2-m air temperature at grid point 1 lead for example to a 2-3 times increase in the frequency of cut-off lows west of the Iberian peninsula. For Z500 and SM1 biased simulations on the other hand, the frequency anomaly is not as high and is concentrated on the center of the North Atlantic. The biased SM2 simulation has the strongest anomaly, with a 4-5 times increase of cut-off lows with respect to the climatology around the 30°W-40°N grid point.

The composite atmospheric situation associated to the presence of a cut-off low west of the Iberian peninsula is shown in Figure 6.16 for the biased T2M simulation only. Figure 6.16a shows the well isolated minimum of Z500, which is not embedded in the jet, the latter being actually mostly zonal and situated much more to the north (50 to 60°N). The meridional wind speed on Figure 6.16b makes clear that the cut-off low is not embedded in a Rossby wave train (at least at 200hPa) and therefore justifies the use of the term 'vortex' to designate this structure. Although the **E**-vector composite suggests an anticyclonic wave breaking origin for the cut-off low, it is actually very difficult to validate this hypothesis in so far as, once isolated, cut-off lows behave very erratically and because their size is of the same order of magnitude as the variance in their position, they tend to vanish on composite maps. The dynamical consequence of the presence of such a cut-off low is likely to be the advection of negative PV anomalies and hot air to its north-east flank (i.e. above Western Europe) at mid-troposphere favoring the appearance and maintenance of a blocked anticyclone which can subsequently break above Eastern Europe (as suggested also by the southward pointing **E**-vector in the Baltic region).

Above Western Europe, a 1.5 standard deviation localized maxima of air temperature at 850hPa is present, associated with a similar anomaly of T2M. There is a coincidence between the region where the T850 anomaly associated to cut-off lows is maximum and the region where anomalous trends are found in Europe is striking (Patterson 2023; Vautard et al. 2023). The anomalous temperatures above the boundary layer are the result of south-west advection by the cut-off low, but whether the anomaly is mainly due to advective, adiabatic or diabatic (especially

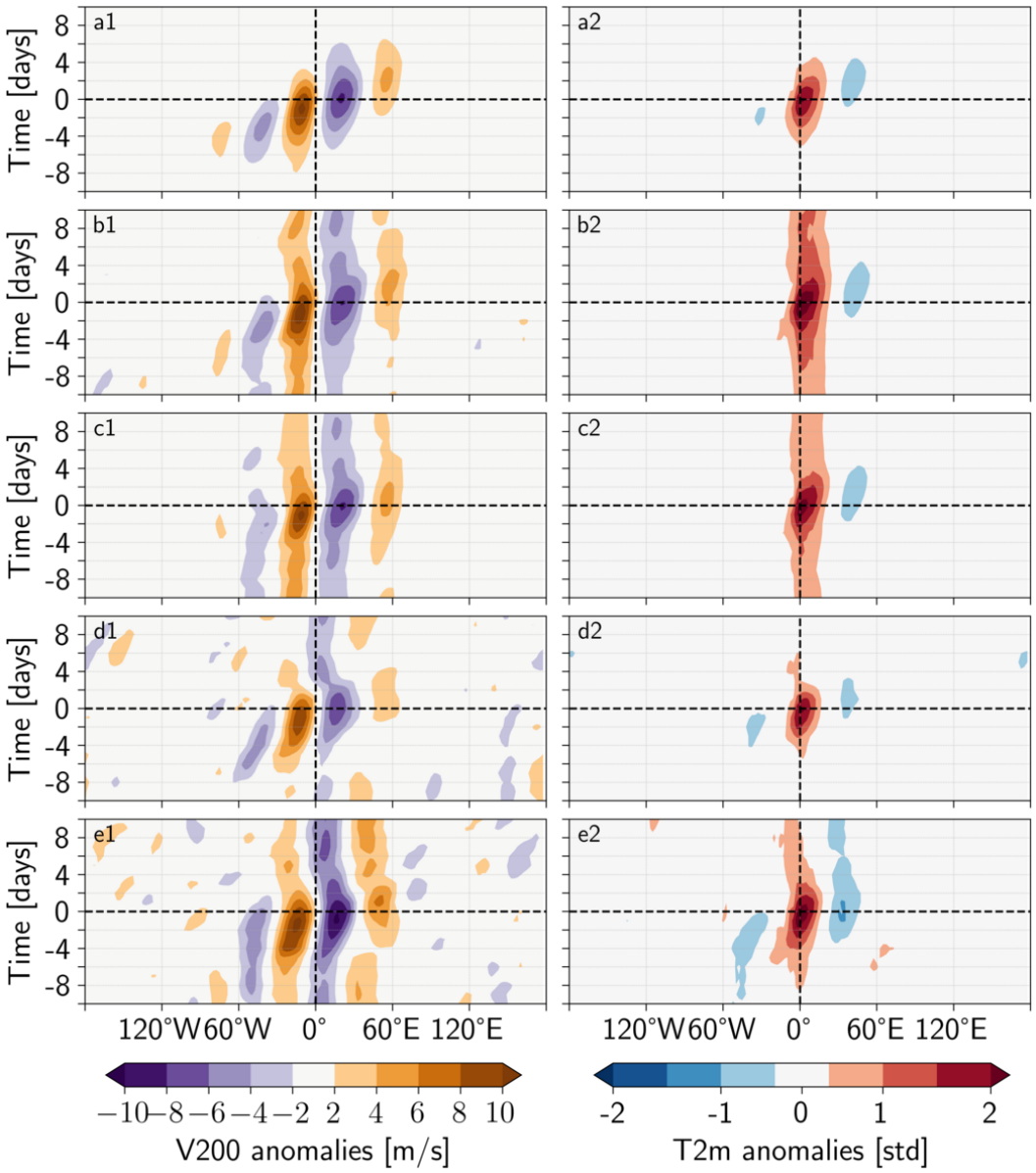


Figure 6.13: **Composite Hovmöller plot of the atmospheric and surface dynamics during heatwave events.** First column: meridional wind speed at 200hPa anomalies averaged between 35°N and 65°N. Second column: 2-m air surface temperature normalized anomalies averaged between 45°N and 55°N. The time is expressed relative to the hottest day of each heatwave event. Rows: (a) control simulation, (b) biased T2M simulation, (c) biased Z500 simulation, (d) biased SM1 simulation and (e) biased biased SM2 simulation. Number of events: (a) $n = 1141$, (b) $n = 4307$, (c) $n = 3290$, (d) $n = 2080$ and (e) $n = 2670$.

lattent heat release by precipitation) mechanisms is not clear. The three mechanisms probably play a role, but one would need a Lagrangian analysis to quantify

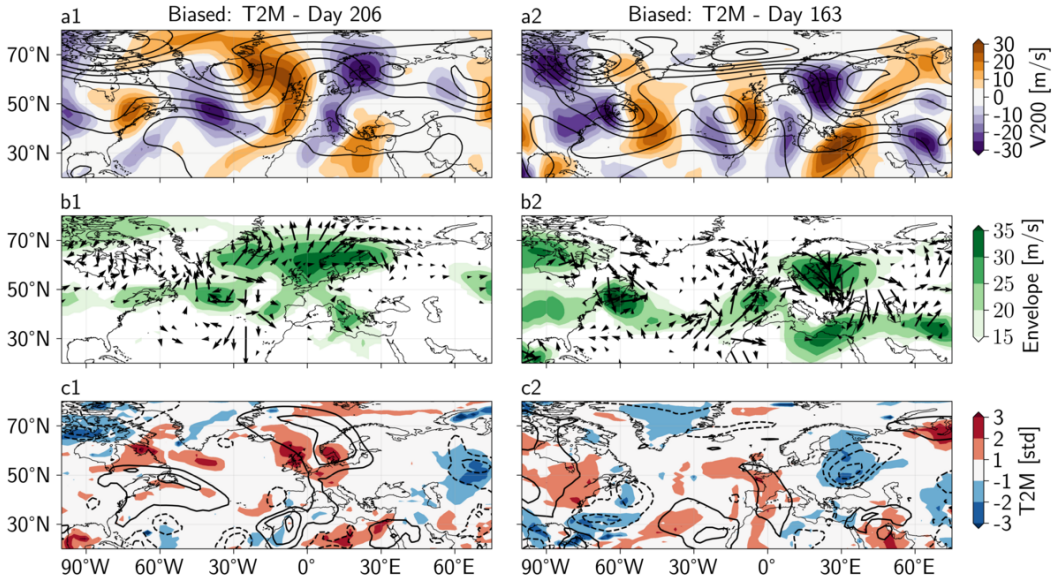


Figure 6.14: **Snapshot of the atmospheric dynamics for two heatwave days in Western Europe.** The two example days are sampled from the same member of the same biased T2M simulation. The first (resp. second) column shows the snapshot dynamics for day 206 (resp. 163). (a1-a2) 200hPa meridional wind (colors) and geopotential height at 500hPa (contours). The geopotential height contours are drawn every 75m starting at 5200m. (b1-b2) Rossby wave packet envelope (colors) and **E**-vector (arrows). The envelope is drawn where it exceeds 15 m/s and the arrows where the norm of the **E**-vector exceeds 15 m^2/s^2. (c1-c2) 2-m air temperature normalized (colors) and 500hPa air temperature (contours) anomalies. The anomalies are computed with respect to the daily ensemble mean and standard deviation from the control simulation. The contours are drawn every 1 standard deviation, starting from +/- 1. Plain (dashed) lines represent positive (negative) values of the standard deviation.

their respective importance. Although there is a 4mm precipitations contour east of the cut-off low, this value is quite small and probably not enough to explain such large anomalies above Western Europe. An other explanation may be the lifting of the hot boundary layer created over Spain and Morocco above Western Europe [6]. If one plots the same analysis on the control simulation (Figure A.13 in supplementary materials), the temperature anomalies are lower, with no strong anomaly over Western Europe. The atmospheric dynamics in particular is quite different, with weaker meridional winds associated to the cut-off and no anticyclonic wave breaking above Eastern Europe. This suggests that the cut-off low-heatwave association is not so straightforward and may occur only when some other conditions — especially with regards to the synoptic dynamics or soil moisture — are present.

I have shown that the presence of an Iberian cut-off low can be linked dynamically

[6] I thank Jacopo Riboldi for our discussions on this hypothesis.

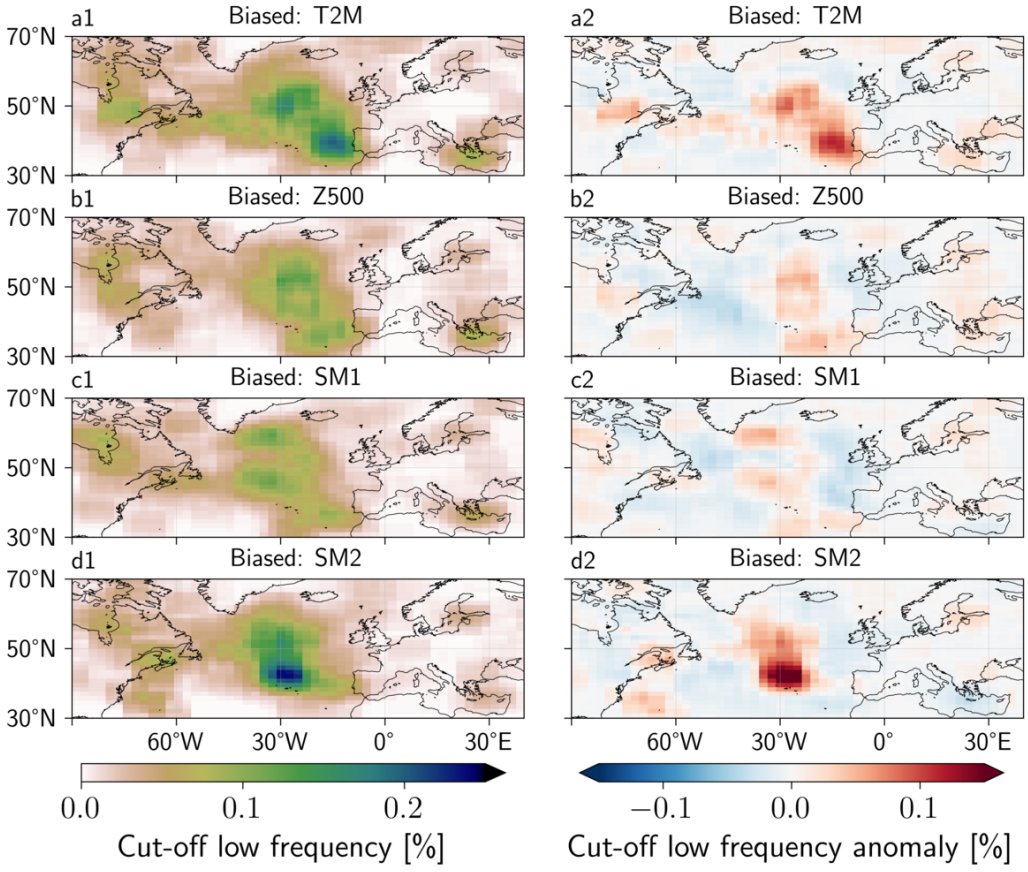


Figure 6.15: **500hPa cut-off lows frequency and frequency anomaly.** The anomaly is computed with respect to the control simulation.

to the occurrence of abnormally hot conditions in Western Europe, at least in the biased T2M simulation. On the other hand Figure 6.17 shows whether the presence of a heatwave in grid point 1 (defined as above) was preceded by the presence of an Iberian cut-off low. The occurrence of heatwaves are centered by considering the time when the maximum temperature during the heatwave is reached at $t = 0$. For each event and each time t, one counts whether there has been a cut-off west of the Iberian peninsula in the $t + 10$ days before (hence a strictly increasing cumulative frequency). Figure 6.17a shows that for the control simulation, around 20% of heatwaves have had a cut-off low west of the Iberian peninsula in the 10 days before — which is a proportion similar to the one found in chapter 5. For the biased Z500 and SM2 simulations the proportion is similar. However, it reaches 25% for biased T2M and goes as low as 13% for SM1. This suggests different dynamics for the heatwaves in those two cases, probably more driven by advective dynamics and adiabatic warming in the first case and local diabatic warming in the second case.

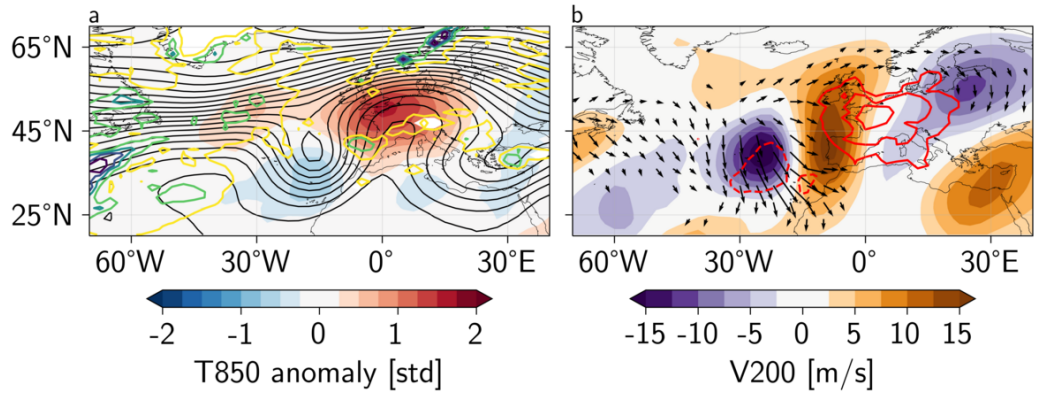


Figure 6.16: **Atmospheric dynamics associated to cut-off lows west of the Iberian peninsula.** The composites are made for all events (n = 1887) in the biased T2M simulation for which a cut-off low is detected in the 33.5°N-42.5°N - 22.5°W-10°W area. (a) Geopotential height at 500hPa (black contours), 3-day average precipitations (colored contours) and 3-day average air temperature at 850hPa anomaly (colors). The geopotential height contours are drawn every 20 meters starting at 5200m. The precipitations contours are drawn every 2mm starting at 2mm. (b) 3-day average 2-m air temperature anomaly (contours), **E**-vector (arrows) and meridional wind speed at 200hPa (colors). The **E**-vector is drawn for norms above 10 m²/s². The anomalies are computed with respect to the daily ensemble mean and standard deviation of the control simulation. For panel (b), the anomalies are drawn every 0.5 standard deviation, starting at +/- 0.5 std. Plain (dashed) lines represent positive (negative) values of the standard deviation.

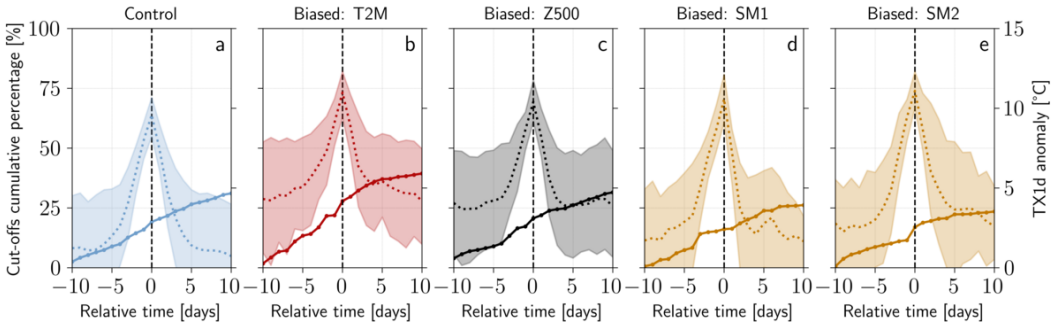


Figure 6.17: **Occurrence of an Iberian cut-off low when a heatwave is happening at grid point 1 at $t = 0$.** The cumulative percentage of cut-offs at time t gives the percentage of heatwave events which have had at least one Iberian cut-off in the last $t + 10$ days. The shadings for the anomaly of maximum daily temperature TX1d shows the 25 and 75% quantiles. Number of events: (a) n = 1141, (b) n = 4307, (c) n = 3290, (d) n = 2080 and (e) n = 2670.

6.4 Conclusions and discussion

I have shown that the use of a rare events algorithm biasing for different score functions allows to vastly increase the amount of extreme summers simulated using the state-of-the-art IPSL climate model. For a similar computational cost as a control simulation, it is possible to sample events that are 10^2-10^4 times less likely than the ones sampled in the original control simulation. This allows to reach precise climatological results on the dynamics leading to extreme summers. As such the rare events algorithm is clearly an improvement compared to the brute force sampling by a long simulation.

I have shown here that extremely hot summers in the IPSL climate model under pre-industrial forcings are the result of synoptic rather than planetary scale atmospheric anomalies. Centennial-like events correspond to a large positive temperature anomaly that is centered in Western Europe but that does not extend at the planetary scale. Vertically, the anticyclonic structure above the heatwave region is not barotropic: dry soils and high sea-level pressures are situated to the east of the maximum 2-m air anomaly while the 200hPa anticyclone is situated to its west. 20-25% of heatwave events are associated to an Iberian cut-off low, reinforcing the anticyclone in mid-troposphere. Statistically I have additionally shown that extremely hot summers are around 10 times more likely than predicted by a Gaussian approximation.

My results on this pre-industrial model simulation thus strongly favor the hypothesis of localized and recurrent dynamics via Rossby wave packets to explain the appearance of hot and very hot summers in Western Europe. It is nevertheless difficult to validate these results with observational data in so far as the undersampling issue for very rare and intense events also applies to them: there is not enough cases to compare to. My results may also be strongly model-dependent and it seems necessary to apply the same methodology at the same place for different models — which could also constitute a test bed for comparing models on their capacity to sample the physical mechanisms leading to very extreme events.

Nevertheless, the use of a rare events algorithm comes with several drawbacks that should be stated here. The main issue of the method is the extinction problem. Because some members are killed and others are cloned at each resampling times, when the simulations are reconstructed from the 1st of June, a lot of diversity in the ensemble is lost at the beginning of the simulation: around 95% of the initial members are completely discarded. This may make the results obtained with the algorithm depend strongly on a few initially selected members and therefore imposes that the algorithm is run several times. Here I have done so 9 times to balance the number of score functions I could explore and the precision of the results, and it is likely that one cannot go much below this number (Ragone and Bouchet (2021) chose to repeat the simulation 10 times). Another consequence of the extinction problem is that even if the model allows to obtain precise results on the tail of the distribution of the summer averaged score functions, the estimation of probabilities for observables that are not closely related to the score functions used are imprecise. In other words, the price of having a gain in the variance at the tail of a local score

functions is paid by increasing the variance almost everywhere else.

The other physical limit on the use of the GKTL algorithm concerns the kind of extremes that are sampled. As explained in section 6.2, the algorithm favors long lasting events which create large summer-averaged anomalies: the algorithm samples large deviations from the summer mean. As a consequence (see also chapter 4 section 4.4.4), the algorithm is not fitted to sample short, very intense events. At the beginning of this work, the rationale for sampling extremely low soil moisture was to favor the typical conditions which lead to very short and intense events. However, even though these kind of events exist in the biased simulations, their intensity is only slightly higher than the intensity of the strongest events in the control simulations (around 1°C, cf. Figure A.14 for the summer maximum of daily maximum temperature).

Another technical drawback is that the way the IPSL model — and probably most climate models — is currently implemented makes it unfitted to run a multitude of members over a limited time span[7]. The 5 days resampling period imposes that all members are run for 5 days, then stopped, resampled and rerun. On the other hand, control simulations are run by continuous periods of one month and do not require resampling. With the current implementation, the control simulation is run in 8 hours per month, for a total of 24 hours for $N = 100$ members for the summer. The biased simulations are run in 2 hours per 5 days period, for a total of 36 hours for $N = 100$ members for the summer. The rare events algorithm therefore makes the integration of the model around 50% longer. Most of the additional computational cost is not caused by the resampling procedure — which is almost immediate — but on loading the starting files of the model at the beginning of each 5 days simulation period. However, it should be noted that this time may be reduced by a change in the implementation of the model in the future. In particular, it may be possible to run the algorithm inside the simulation without having to stop the simulation at each resampling time, which would allow to avoid loading the starting files and therefore save time.

Thus, even though the computational gain is in principle of the order of 10^2-10^4 with respect to a control simulation, in practice with the rare events algorithm there is an additional computational cost increase of 50% and one needs to run around 10 times the algorithm with the same k parameter to avoid the extinction problem. As a consequence, the real computational gain is more around 6-600 — which is still a substantial gain.

Finally, it should be emphasized that even though there is a net computational gain compared to a control simulation, the gain is actually *local*. As I have shown there is no free lunch: the model samples very well the tail of a local score function but loses quality in its estimation almost everywhere else. In particular, one cannot really use a biased simulation to estimate climatological results far from the location where the score function is computed. As a consequence, if a 1000-y long control simulation can be used to obtain climatological results — including for extreme

[7] Note that I run the model on 712 CPUs thanks to an ensemble simulation tool which allows to launch simulations by pools of 10 ensemble members.

events — everywhere on Earth, one would have to run the rare events algorithm for each score function of interest (especially at various grid points) which may largely compensate the computational gain initially obtained by the rare events algorithm (not to mention the additional storage cost due to the multiple rare events simulations). As a consequence, it seems clear that rare events algorithms cannot replace long control run or large ensemble simulations. The latter are actually more computationally competitive when one considers their multiple potential applications. The typical use case of rare events algorithm is for case studies of extremes that are of particular interest for a physical problem (like an unexplained bias in the control simulation) or a practical application (like the sampling of very intense extremes for the development of storylines (Shepherd et al. 2018)).

Summary

Context and goals

Extreme events are under-sampled by brute force and/or ensemble simulations. The use of rare events algorithms allow to sample much more extremes than are provided in a regular simulation. These large samples of extremes can be used to provide physical insights in the dynamical mechanisms leading to extremes. In this chapter I study in particular the dynamical mechanisms leading to extreme hot summers in Western Europe.

Methods

I apply a rare events algorithm to a climate model to sample extreme and very extreme summers under pre-industrial forcings. I investigate the dynamics leading to extremely hot summers in Western Europe by biasing an ensemble simulation towards favoring summers with high near-surface temperatures, high geopotential height and low soil moisture.

Results

The rare events algorithm allows to recover climatological probabilities and I show that it samples events with probabilities of occurrence between 10^{-2} and 10^{-6}. I show that hot summers are characterized by both more and longer heatwaves. These summers arise as a combination of local non-barotropic high geopotential height and non-local soil moisture anomalies. The atmospheric dynamics of these summers is mainly driven by a recurrence of short Rossby wave packets rather than circumglobal wave patterns.

Chapter 7

Statistical and dynamical evolution between present and future very hot summers in the IPSL-CM6A-LR climate model

7.1 Introduction

In this chapter I expand the work presented in chapter 6 under present and future scenarios of external forcings. I seek to investigate statistical and dynamical changes in very hot summers between pre-industrial, present, SSP2-4.5 and SSP3-7.0 conditions. The target time period for the SSPs (O'Neill et al. 2014) is the last decade of the 21st century. The main forcing changes between these four configurations are the level of greenhouse gases, especially CO_2, aerosols concentration and land-use/land-cover changes which lead to a global warming of around +1.6°C in the present, +3.9°C in SSP2-4.5 and +5.2°C in SSP3-7.0 with respect to the pre-industrial simulation — for the IPSL model. I implement the same GKTL rare events algorithm as previously. In this chapter, only temperature at one grid point is used as a score function due to limited computational resources.

This chapter is organized as follows. In section 7.2 I present the changes made in the rare events algorithm and the configuration of the model with respect to the previous chapter. Section 7.3 displays the results obtained. The section begins by a statistical description of the extreme summers. I then display the changes in the summer climatologies of key thermodynamical and dynamical variables. In particular, I show an increase in the thermodynamic variability and a decrease in the dynamic variability between the present and the future scenarios. I then investigate the mean spatial structures of the centennial events sampled by the algorithm and I finally detail their atmospheric dynamics. Section 7.4 summarizes the results obtained and the conclusions reached.

7.2 Methods

In this chapter, the same rare events algorithm as in chapter 6 is employed. I therefore do not detail again how the algorithm works and the choices made for its implementation.The reader is invited to refer to section 6.2 of the previous chapter. Here I apply the GKTL algorithm with present and two future scenarios for external forcings on the climate system. The present starting conditions are extracted from a 32-member ensemble simulation of the fully coupled IPSL model (Boucher et al. 2020) under extended historical forcings between 2015 and 2025. For future scenarios, an intermediate warming scenario (SSP2-4.5, +3.9°C in the model) and a high warming scenario (SSP3-7.0, +5.2°C in the model) with respect to the pre-industrial simulation are chosen. For both warming scenarios, the starting conditions are chosen from a 11-member ensemble of the fully coupled model over the period 2090-2100.

To be comparable, simulations in the past, present and future should occur in similar phases of the slow components of the climate system. It is not possible to fully ensure that it is the case due to the limited sampling of the variability in the fully coupled simulations I have access to. Therefore, only two modes that are thought to be the most important for both the climate system as a whole and hot summers in Western Europe are chosen: the El-Nino Southern Oscillation (ENSO) and the Atlantic Meridional Oscillation (AMO), cf. chapter 1 section 1.3.3. The ENSO index in the model is computed as the 1-y rolling mean of the spatial average over the 190°E–240°E and 5°S–5°N area of the sea surface temperatures (SSTs). The AMO index in the model is computed as the 1-y rolling mean of the spatial average over the 75°W–7°W and 25°N–60°N area of the SSTs. Both indices are then normalized by removing the mean and dividing by the standard deviation. The cross-distribution of these indices for the pre-industrial, present and future simulations are presented in Figure 7.1. The black dots show the ENSO-AMO indices for the pre-industrial simulations of chapter 6. The AMO is in a slightly negative phase (AMO index around -1) while the ENSO is in a strong negative phase (ENSO index around -2). I find years as close as possible to this combination in the present and future scenarios as shown by the red dots in Fig. 7.1bcd. Nonetheless, one should note that in the 2000-y pre-industrial control run, the Pearson correlation between the ENSO index over the summer and the summer averaged temperature at grid point 1 is almost null ($\rho = 0.01$, $p > 0.05$) while the correlation with the AMO index is only moderate ($\rho = 0.24$, $p < 0.001$).

The score function used for each simulation is the same as previously: 2-m air temperature in °C (T2M) at grid point 1 (49.5°N, 2.5°E) with control parameter $k = 0.04$/°C/day. With the value of k chosen here, under a Gaussian approximation, the summer averaged anomalies expected to be sampled by the algorithm are $a_{\text{present}} = 2.3°\text{C} = 2.5\text{std}$, $a_{\text{SSP2}} = 2.5°\text{C} = 2.5\text{std}$ and $a_{\text{SSP3}} = 2.6°\text{C} = 2.6\text{std}$. Due to limited computational resources, here I only run T2M biased simulations. As shown in chapter 6, this score function is well suited to sample extremely hot summers. As previously, the model is run in its atmosphere - land surfaces configuration (LMDZOR). First, a 4-months spin-up simulation (January to April) of the uncou-

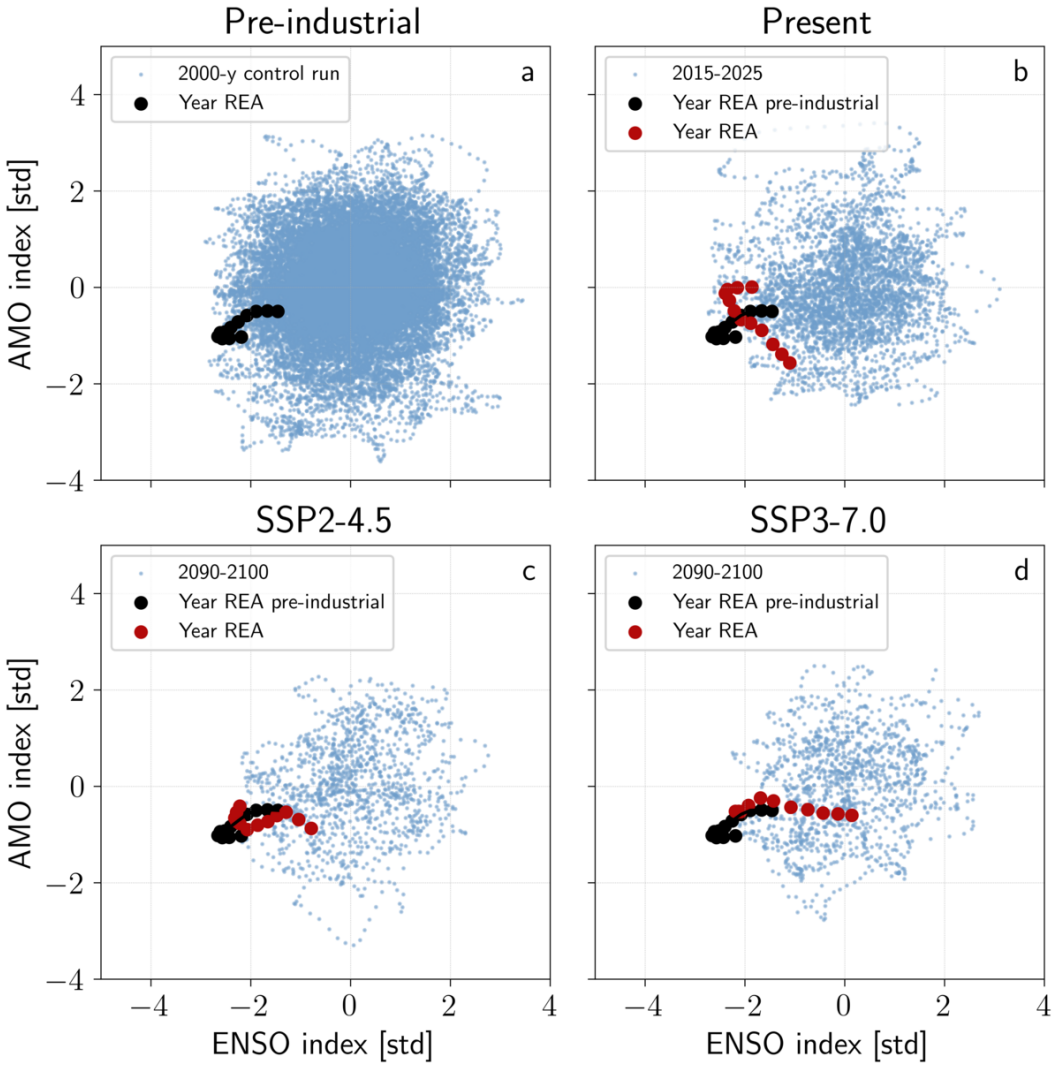


Figure 7.1: **ENSO-AMO indices distribution and values for the simulation years.** ENSO-AMO index distribution for (a) pre-industrial 2000-y control run, (b) 32 members between 2015-2025, (c) 11 members between 2090-2100 under SSP2-4.5 and (d) 11 members between 2090-2100 under SSP3-7.0. In all panels, the black dots show the values of the ENSO-AMO indices for the 12 months of the year chosen for initial conditions of the rare events algorithm (REA) in the pre-industrial simulation. For panels (b), (c) and (d) the red dots show the values of the ENSO-AMO indices for the 12 months of the year chosen for initial conditions of the rare events algorithm (REA) in the corresponding period.

pled model with oceanic and ice conditions of the initial coupled simulation is run to let the atmosphere adapt to the uncoupling with the oceans and to provide initial conditions for the rare events simulation in summer. Random perturbations on potential temperature at all levels and grid points are then added at the beginning of May to create a 100 members starting ensemble. The perturbations are added in

May to avoid having to simulate the 100 members from January to April while still letting enough time for the different members to diverge until the beginning of the rare events simulation in June. All the simulations begin on the 1st of June from those same 100 members and end on the 28th of August (90 days).

I run $N_s = 9$ rare events simulations with $N = 100$ members each for the present and the future scenarios. A 900-member control ensemble without the resampling procedure is additionally run for each period. In the following, I mainly consider centennial events, i.e. choosing a level a so that $\mathbb{P}_\mu[F(X(t)) \geq a] \simeq 10^{-2}$ where F is the summer averaged 2-m air temperature at grid point 1 and μ is the climatological probability distribution. These centennial events correspond to the following levels: $a_{\text{present}} = 22.2°\text{C}$, $a_{\text{SSP2}} = 25.5°\text{C}$ and $a_{\text{SSP3}} = 27.5°\text{C}$ (cf. Fig. 7.6).

Here I do not detail again how the atmospheric dynamics diagnostics are computed. The reader is invited to refer to section 6.2. For consistency with the analysis of chapter 6 the 200hPa level is mainly used to investigate the dynamics of Rossby wave and the corresponding figures at 500hPa are shown in supplementary materials. As previously, except specified explicitly, daily averaged outputs of the model are considered.

7.3 Results

7.3.1 Grid point statistics

This section presents statistics mainly for grid point 1, i.e. where the summer averaged 2-m air temperature is maximized in the biased simulations. Figure 7.2 shows the empirical histograms of the control and the biased simulations for summer averaged 2-m temperature (T2M, Fig. 7.2a), geopotential height at 500hPa (Z500, Fig. 7.2b), soil moisture at grid point 1 (SM1, Fig. 7.2c) and soil moisture at grid point 2 (SM2, Fig. 7.2d). As expected, summer averaged temperatures are stronger in the future: T2Mm = 19.8°C for the present, T2Mm = 23.1°C for SSP2-4.5 and T2Mm = 25.1°C for SSP3-7.0. This makes these summers respectively 1.6°C, 4.9°C and 6.9°C hotter than the control summers in the pre-industrial simulations of chapter 6, demonstrating an amplification with respect to global warming of respectively +0.0°C, +1.0°C and +1.7°C in the three periods. There is also a small increase in the standard deviations of the summer averaged temperature (0.93°C in the present, 0.99°C in SSP2-4.5 and 1.03°C in SSP3-7.0) but only the difference between present and SSP3-7.0 is significant at the 5% level with a Bartlett test (Bartlett 1937), even with 900 control members, indicating only a modest increase of the standard deviation of summer averaged temperatures at grid point 1. There is however a significant increase ($p < 0.001$) in the standard deviation of the summer distribution of daily temperatures (3.6°C in the present, 4.0°C in SSP2-4.5 and 3.8°C in SSP3-7.0), the SSP2-4.5 being the more variable. As the distribution of daily mean temperatures over the summer is not Gaussian, the Levene test (Levene 1960) is used for the significance of the change in the variance. The fact that the moderate warming scenario is the more variable may be a result of the transition zone

northward shift and the feedback with soil moisture that leads to stronger variability when soils are neither too dry nor too moist (Seneviratne et al. 2010), cf. next section for more elements on this point. Additionally, there is a constant amplitude of the anomaly corresponding to a centennial event for the summer averaged temperature at grid point 1: the anomaly is around +2.6°C in the pre-industrial simulation, and +2.4°C in the present, in SSP2-4.5 and in SSP3-7.0.

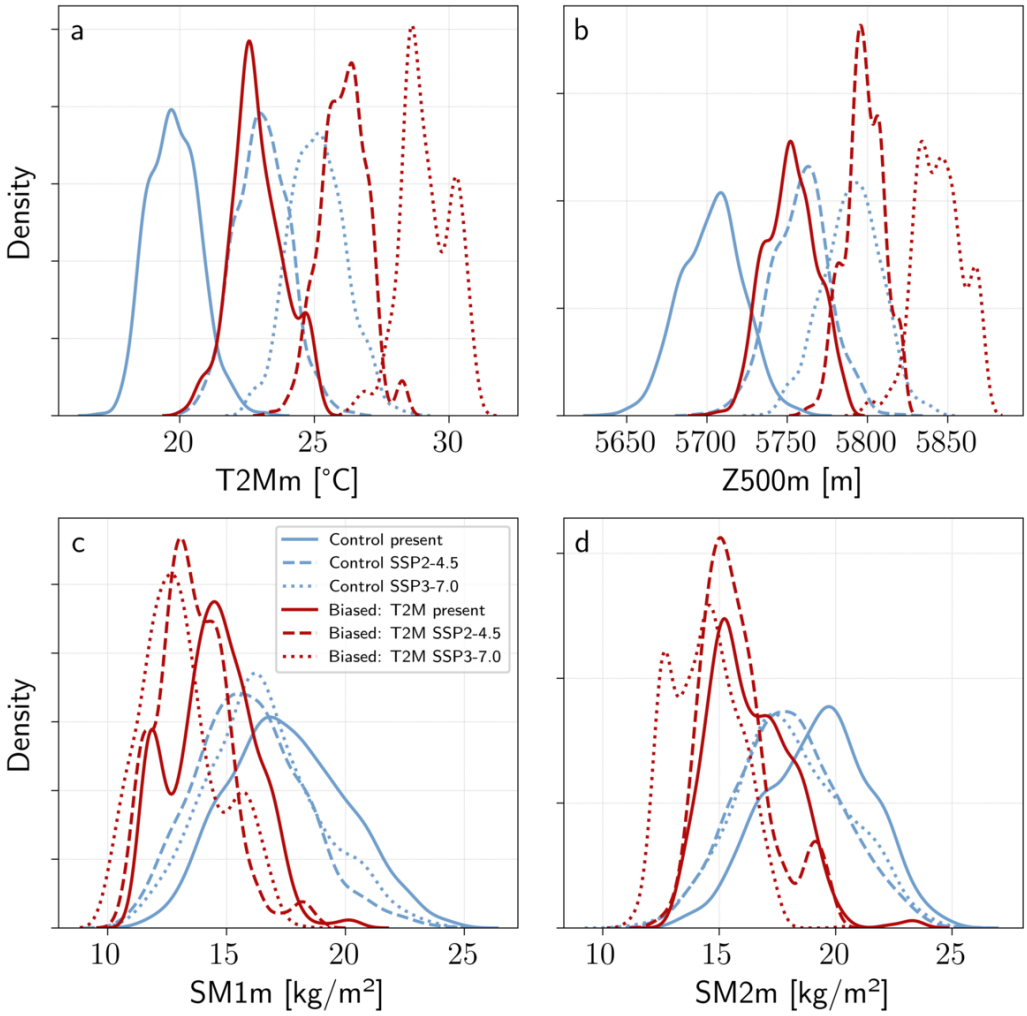


Figure 7.2: **Empirical probability density functions (PDFs) of summer averaged grid point observables.** PDFs of the summer averaged (a) 2-m air temperature, (b) geopotential height at 500hPa, (c) upper-level soil moisture at grid point 1 and (d) upper-level soil moisture at grid point 2. The PDFs are obtained by a kernel smoothing of the empirical histograms.

As expected, the biased T2M simulations sample much more extreme summer averaged temperatures and geopotential height, including extremes which are unprecedented in the control simulations. For example, the present biased T2M simulation displays summers which are as intense as the average ones expected with the

SSP2-4.5 scenario at the end of the 21st century. The summer averaged Z500 also increases in the two scenarios with respect to present conditions, while the summer averaged soil moisture decreases but more moderately, being in particular always bounded downwards by the 8.14 kg/m² limit (see chapter 6). As a consequence, contrary to other observables, the control distributions of the two future scenarios for soil moisture are very close. The difference between the biased simulations in the present and in the future are therefore smaller than the differences observed for temperatures. Biased T2M simulations are dryer than their climatology but only moderately dryer and similarly when they are compare between each other.

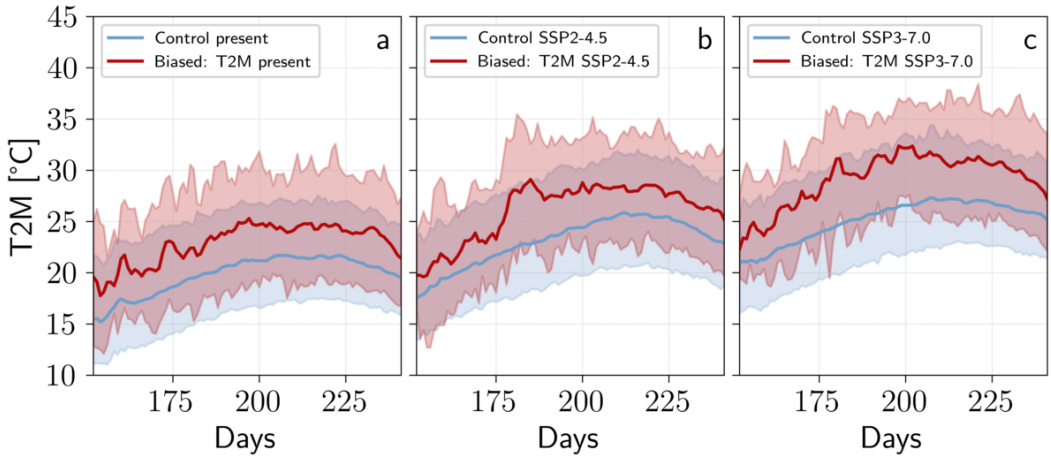


Figure 7.3: **Time series of the ensemble mean and the 5 and 95% quantiles of 2-m air temperature at grid point 1.** For all the plots the plain line shows the ensemble mean and the shading the 5 and 95% quantiles of the variable displayed. For the biased simulations, all members of all simulations are pooled together to compute the 5 and 95% quantiles.

The average time series of temperature at grid point 1 are shown in Figure 7.3. As previously, the average of the rare events algorithms simulations are several degrees above the climatological average obtained with the control simulations. The 95% quantile of the biased T2M simulations therefore sample very extreme daily temperatures. This is also shown by Figure 7.5 which displays the distribution of the summer maxima of maximum daily temperature (TXx) in the biased and control simulations. On average these maxima are around 3°C higher in the biased simulation, but the overall maxima of TXx in the biased simulation are not higher — even a little bit lower for the SSP2-4.5 and SSP3-7.0 — than the maximum over the 900 members in the control simulations. As explained previously, the GKTL algorithm is better suited to sample long persisting anomalies (i.e. large deviations) rather than short intense events.

Figure 7.3 could give the impression that members selected are continuously hotter than the control. This is however false in general as shown by Figure 7.4 displaying the number of heatwaves per summer and their average length. As previously, a heatwave is defined as a continuous period of time of at least 3 days length for which the daily mean temperatures are above the 90% climatological quantile

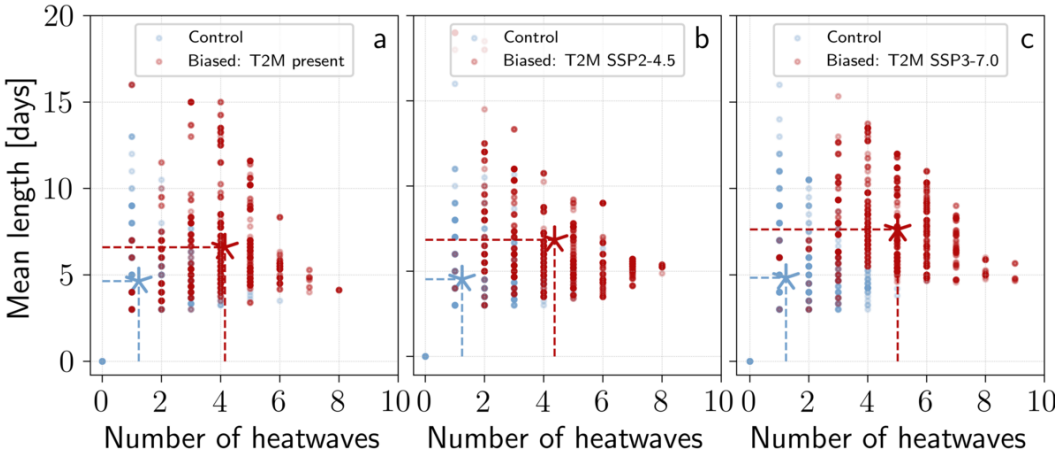


Figure 7.4: **Heatwaves number per summer and mean duration at grid point 1.** For all plots a heatwave is a continuous period of time where daily mean temperatures are above the 95% climatological quantile. The stars show the ensemble mean.

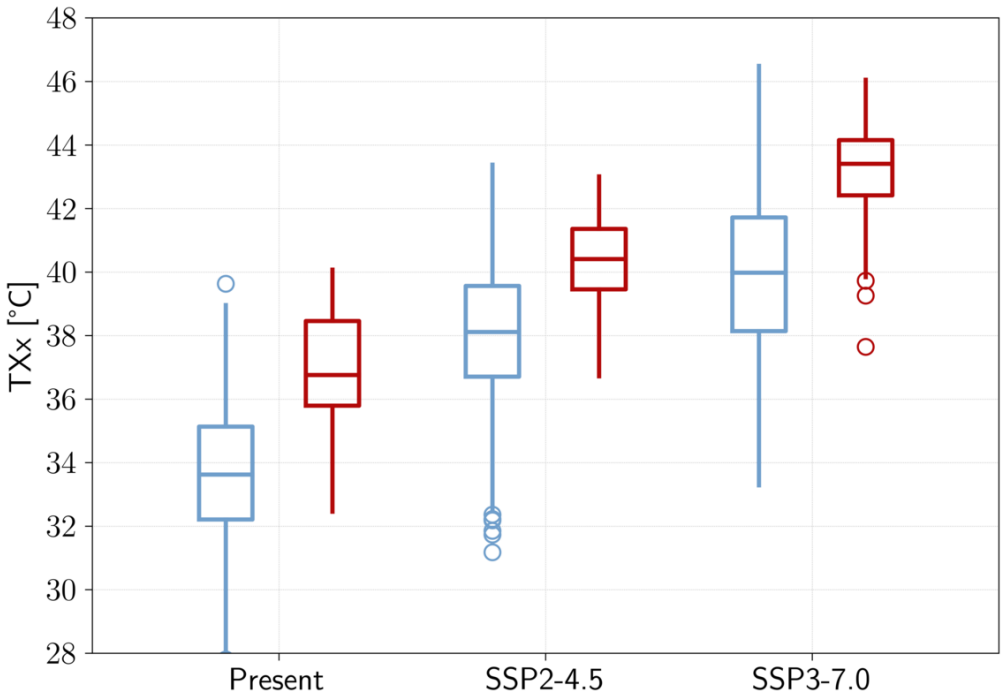


Figure 7.5: **Comparison of the distribution of summer maxima of maximum daily temperatures (TXx) at grid point 1.** The boxplots show the 25 and 75% quantiles, the median and the outliers.

computed on the control simulations. Biased T2M simulations have both more heatwaves and longer heatwaves. The number and mean length of heatwaves is constant in the control simulation for the present and the two future scenarios: around 1.2

heatwaves per summer with an average length of 4.6 days in the present, 4.5 days in SSP2-4.5 and 4.8 days in SSP3-7.0. There is however an increase for biased simulations in the number of heatwaves from present to future simulations (4.1 per summer for present, 4.4 for SSP2-4.5 and 5.0 for SSP3-7.0), and also an increase in mean duration (6.6 days for present, 6.8 for SSP2-4.5 and 7.6 for SSP3-7.0). It is however not clear whether these increases are due to a change in the properties of the system or to the variability of the sampling caused by the rare events algorithm between the three periods.

The intensity of the selection for the three periods is shown in Figure A.15. The final intensity of the extinction (around 65 effective members) is equivalent to the one in the pre-industrial simulations (cf. Fig. 6.3). This is expected in so far as the four simulations have a similar value of the parameter k in the GKTL algorithm that controls the strength of the selection — which, as shown in section 7.2, should lead to similar anomalies of the summer averaged temperature at grid point 1.

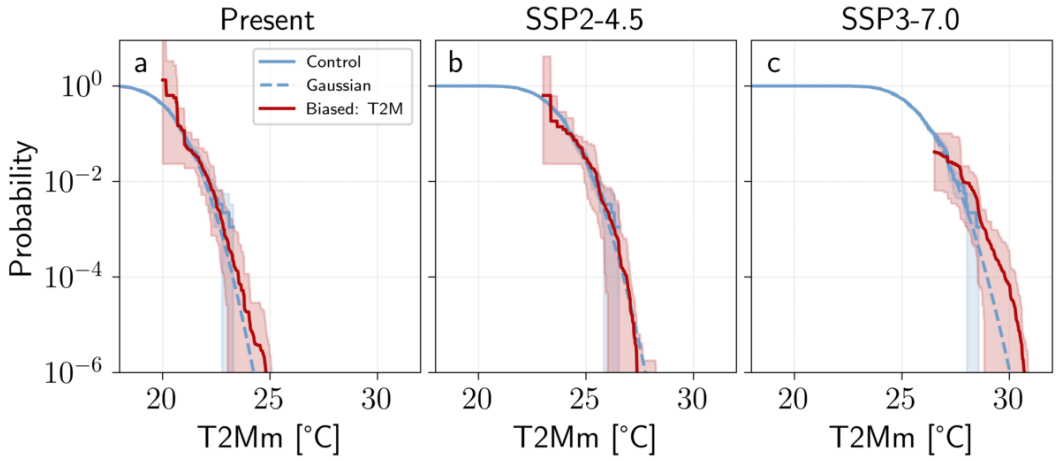


Figure 7.6: **Probabilities of an extreme summer according to the naive estimator (control simulation) and the rare event estimator.** Probability of the summer averaged 2-m surface temperature to be above a certain threshold at grid point 1. The blue dashed lines represent the corresponding probabilities for a Gaussian distribution fitted on the control simulation by the method of moments. The shadings for the control simulation show the 5–95% quantiles of the estimated probability obtained using bootstrap on the 900 members. For the biased simulations the shadings show the minimum and maximum of the estimated probability over the $N_s = 9$ simulations.

Figure 7.6 shows the mean probabilities of the summer averaged temperature to be above a given level at grid point 1 given by the algorithm for the three periods. As previously, the algorithm provides accurate probabilities as low as 10^{-6}, i.e. events with a return period of one million years. The uncertainty range from one simulation to another is small – around a few degrees. Contrary to Figure 6.5a, the present and SSP2-4.5 simulations follow closely the Gaussian extrapolation for low probabilities. This is however not the case for SSP3-7.0, which shows a similar departure from Gaussianity as in Figure 6.5a. This is likely related to the bimodality in the biased

temperature distribution in Figure 7.2b. It should nevertheless be noted that the Gaussian extrapolation is still in the uncertainty range. Whether this change is the result of a truly different mechanism is difficult to confirm with the simulations run here. Increasing the number N_s of simulations could lead to a clearer view on this topic. Physically, the pre-industrial and SSP3-7.0 simulations have in common a low level of anthropogenic atmospheric aerosols, which could explain a change in the tail behavior (cf. chapter 1 section 1.3.3).

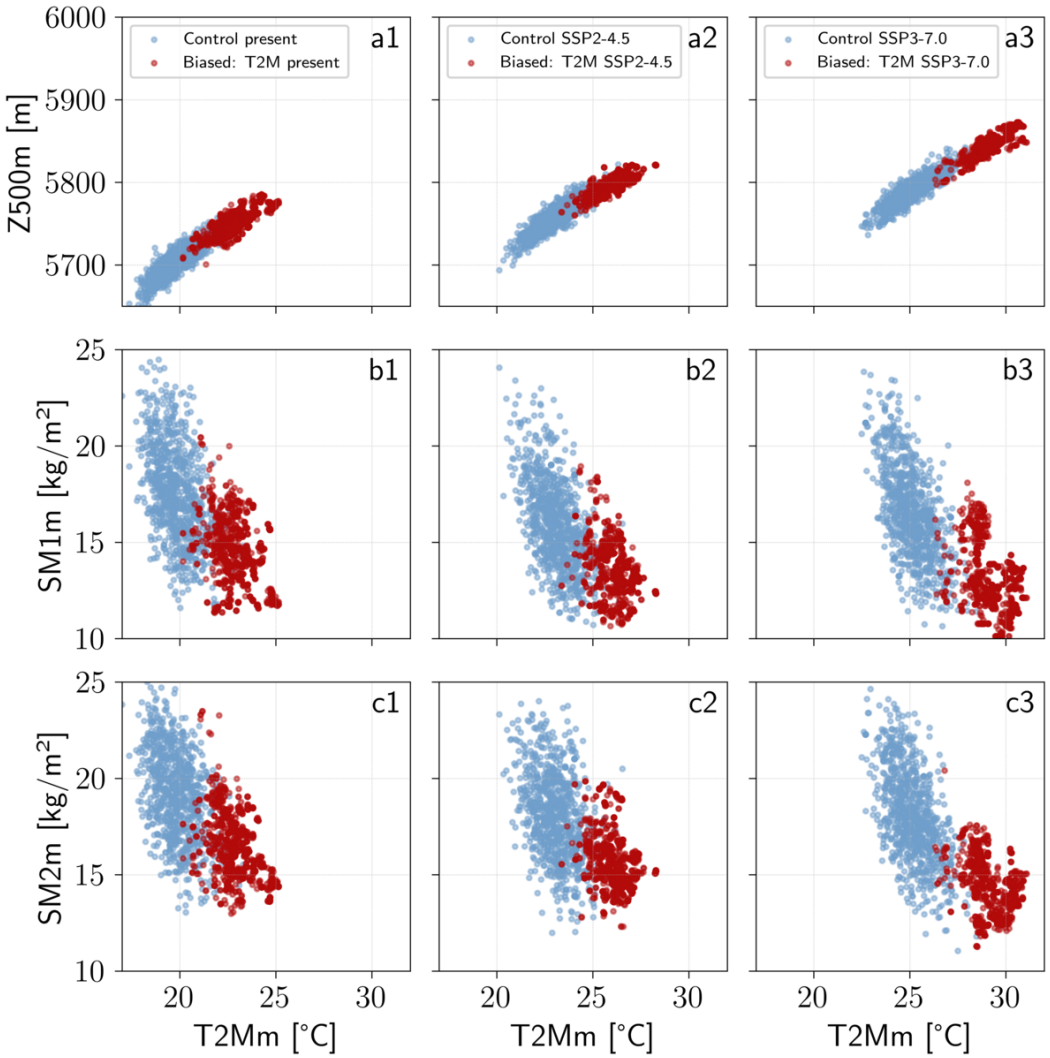


Figure 7.7: **Scatter plot of summer averaged grid point observables.** (a) Summer averaged geopotential height at 500hPa at grid point 1, (b) summer averaged upper-level soil moisture at grid point 1 and (c) upper-level soil moisture at grid point 2 vs summer averaged 2-m air temperature at grid point 1.

Finally, Figure 7.7 explores the cross-distribution between summer averaged 2-m air temperature at grid point 1 (T2M) and summer averaged (a) geopotential height at 500hPa (Z500), (b) upper-level soil moisture at grid point 1 (SM1) and (c) upper-

level soil moisture at grid point 2 (SM2). As previously, there is a strong link between T2Mm and Z500m in both the control and the biased simulations and this link is preserved in the future scenarios. For soil moisture, similar to the pre-industrial simulations, hot summers are dryer but not much dryer than the climatology and the algorithm does not explore the joint tail of the distribution. The future scenarios are dryer and hotter than the present, but at grid point 1 — as stated previously — the distributions of soil moisture in SSP2-4.5 and SSP3-7.0 are quite similar and the difference is mainly a shift in the temperature distribution (Fig. 7.7b2 and b3). At grid point 2 the situation is similar, except for SSP3-7.0 for which some members of the biased simulation sample both very hot and very dry summers.

7.3.2 Average thermodynamic and dynamic evolution between present and future scenarios

In the next sections I investigate the evolution of thermodynamical and dynamical mechanisms leading to centennial hot summers between the present and the future scenarios simulations. The fields shown are mainly normalized anomalies which are computed with respect to the climatology of each period. As a consequence, for the correct interpretation of the maps shown, in this section I first detail the summer averaged evolution of thermodynamical and dynamical fields between the control simulations of the three periods. The thermodynamical fields refer to variables which are related to diabatic energy fluxes (2-m air temperature, soil moisture, sensible and latent fluxes) while dynamical fields refer to variables related to the atmospheric dynamics (zonal and meridional winds). I show essential figures for the understanding of the results obtained for centennial events. The rest of the figures are shown in supplementary materials.

The differences in the summer averages, standard deviation in the summer averages and standard deviation in the full summer distribution are investigated. The standard deviation in the summer averages is computed between the $N = 900$ members of the control simulation. The standard deviation in the full summer distribution is computed between the pooled data of the $N = 900$ members and the 90 days of the control simulation. The statistical significance at 5% is evaluated with a Welch t-test (Welch 1947) for the summer averages, and an F-test for the standard deviations. By doing so it is assumed that the distribution of summer averages are close to a Gaussian, which is expected because of the central limit theorem. For the distribution of daily mean temperatures however the Gaussian hypothesis is clearly wrong and a Levene test (Levene 1960) is therefore used. For the summer averages, with 900 members the statistical significance is obtained for differences larger than 0.07 std. For the standard deviations, the changes in the percentage with respect to the present are shown: $\Delta\sigma/\sigma_p$ where $\Delta\sigma$ is the change in standard deviation between the two periods and σ_p is the standard deviation estimated on the present simulations. With 900 members in each simulation, with an F-test, the significance at 5% is obtained with a difference larger than 7% for the standard deviation in the summer averages and with a difference larger than 1% for the standard deviation in the full summer distribution. Climatologies shown are restricted to the Northern

Hemisphere mid-latitudes in so far as this is the region of interest for the extremes of near-surface air temperature as shown by chapter 6.

The differences in summer-averaged 2-m air temperature and upper-level soil moisture between the present and the future simulations is shown in Figures A.16 and A.17. As expected, future summers are much hotter in the future scenarios almost everywhere on the Northern Hemisphere, with stronger warming over the land and close to the pole rather than over the oceans and close to the Equator. Future summers are also significantly dryer in almost all of the European continent. Latent heat fluxes are moderately decreasing over Western Europe (Figure A.18), the strongest decrease being observed north of the Black Sea. This is likely the results of the competition between less available water but more potential for evapotranspiration due to the increasing mean temperatures caused by the radiative effect of GHGs. On the other hand, there is a consistent increase in sensible heat fluxes over the European continent (Figure A.19).

Figure 7.8 shows a significant increase of around 15-30% in the standard deviation of the summer averaged temperature over France and the Balkanic area in SSP3-7.0 with respect to the present. This is in contrast to the results of the preceding section on grid point 1 and the difference may arise as a result of the tempering influence of the nearby sea for this grid point. It is likely that there is also an increase for SSP2-4.5 but it is non-significant at the 5% level with 900 members. Figure A.22 in supplementary materials shows that this is the case in the summer distribution of daily 2-m air temperature. For soil moisture, Figure A.20 for the summer averages and Figure A.23 for the full distribution show a decrease of variability in the Mediterranean basin but a small increase in Western Europe (5-10%) that is non-significant in the summer averages. On the other hand, the variability in both latent heat (Fig. A.21, cf. Fig. A.24 for the full summer distribution) and sensible heat fluxes (Fig. 7.9, cf. Fig. A.25 for the full summer distribution) is strongly increasing, around 30-40% for SSP3-7.0 in Western Europe.

The evolution of the summer averaged 500hPa zonal and meridional winds is shown in Figures A.26 and A.27 (see Figures A.28 and A.29 for the winds at 200hPa). There is both a northward shift and a decrease in intensity of the jet over North-America that is extending in the western part of the Atlantic basin. Over Europe there are only minor changes in SSP2-4.5, and a moderate decrease (1-2 m/s) of zonal winds in SSP3-7.0. In the latter, there is also an increase in meridional winds over Morocco and Spain, which may favor warm advection at mid-troposphere in Western Europe.

The major change in the dynamics is however not on the mean circulation but on the variability. Figures 7.10 and 7.11 show that over the Euro-Atlantic region there is a clear tendency towards a decrease in the standard deviation of the summer averaged fields. Figure A.32 and Figure A.33 show that this is even clearer in the standard deviation of the summer distribution of daily fields. As a whole, there is a decrease of variability in the atmospheric dynamics between 10 to 20% in SSP3-7.0 with respect to the present simulation. Figures A.30, A.31, A.34 and A.35 show a similar — although more moderate — decrease of the variability in the Euro-Atlantic region of the winds at 200hPa.

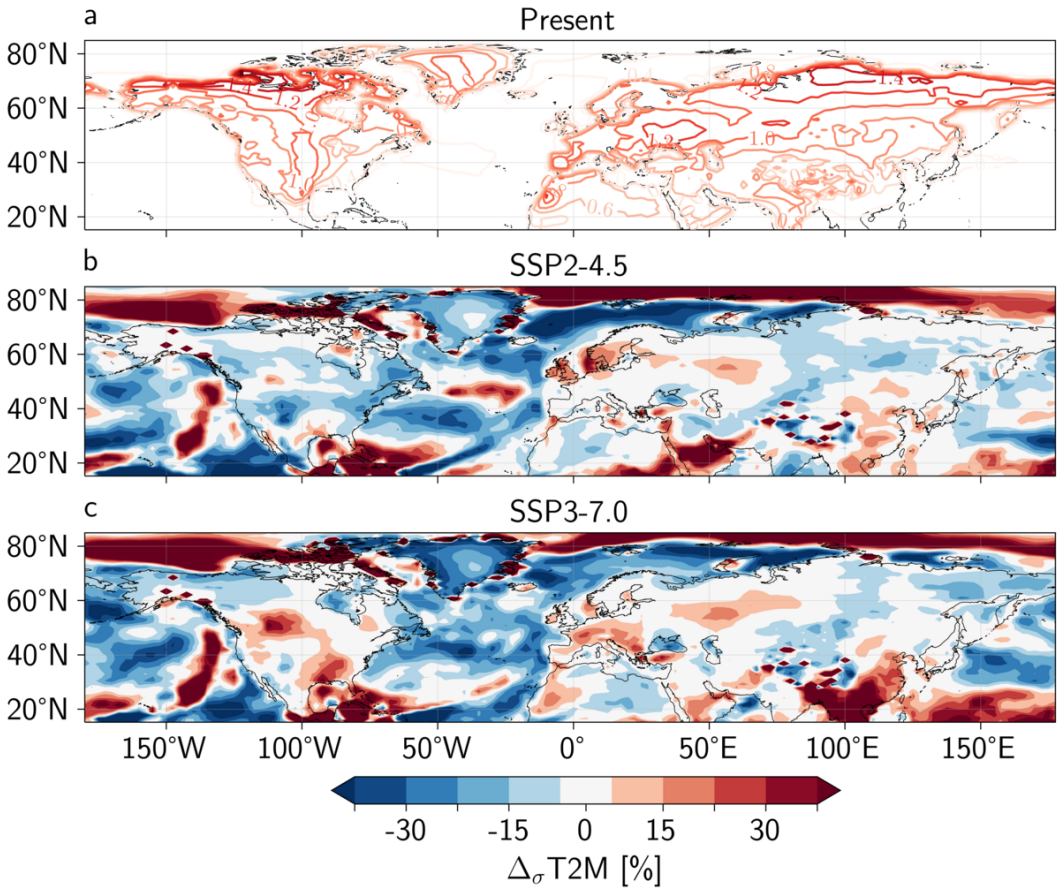


Figure 7.8: **Standard deviation of the summer averaged 2-m air temperature in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

To summarize this section, apart from the mean warming and drying between the present and future scenarios, the climatological changes over the Atlantic basin and Western Europe in the summers simulated can be described as two major evolutions: (i) an increase in the variability of the surface/thermodynamical variables (2-m air temperature, upper-level soil moisture, latent and sensible heat fluxes) and (ii) a decrease in the variability of the dynamical variables (500 and 200hPa zonal and meridional winds). I come back below to these changes to interpret the results obtained with the rare events algorithm.

7.3.3 Summer averaged spatial structure

In this section I investigate the summer averaged spatial structure of centennial-like events. Figure 7.12 shows in contour the normalized summer averaged meridional

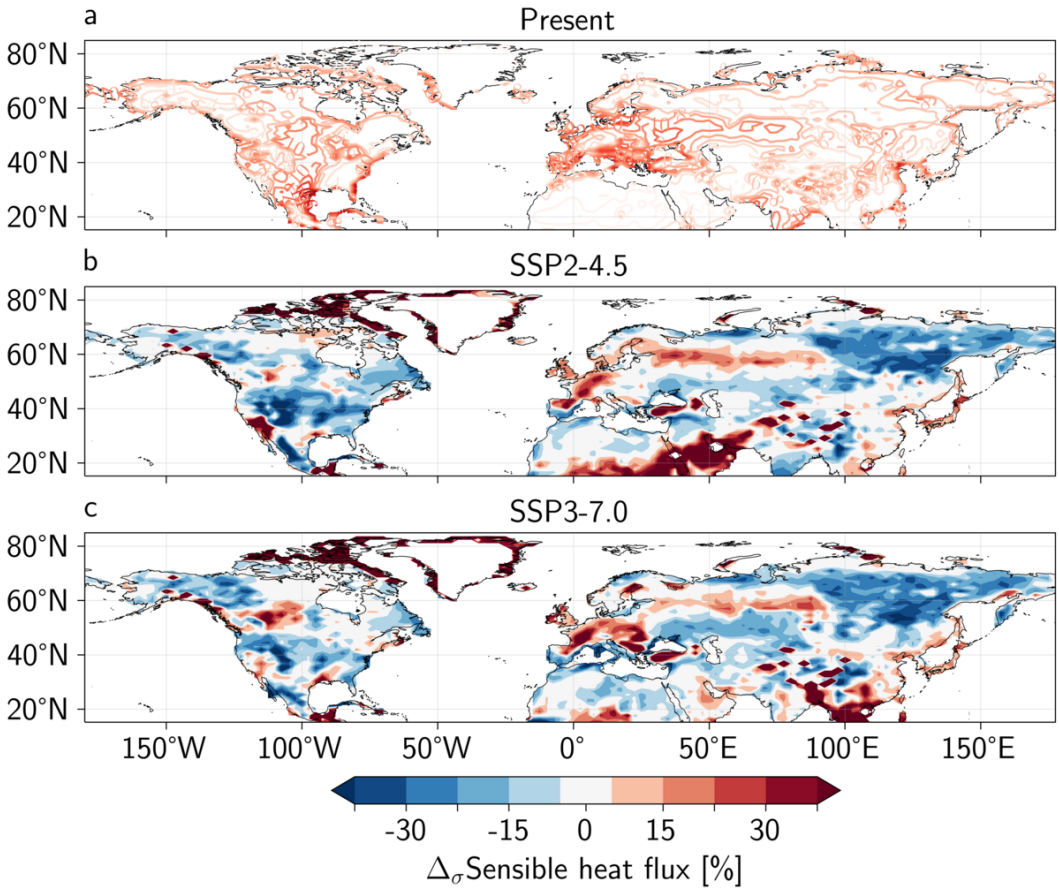


Figure 7.9: **Standard deviation of the summer averaged sensible heat flux in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

wind at 200hPa and in colors the normalized 2-m air temperature. The average patterns of temperature are very similar in the past (cf. Fig. 6.7a), present and future. In all periods, the strongest anomalies are concentrated in Western Europe and only moderately extend in the rest of the Eurasian continent. The atmospheric dynamics at 200hPa is also very similar, with the strongest meander concentrated around grid point 1 and not extending throughout the whole Northern hemisphere. There also seems to be a decrease of the spatial extension of the V200 and T2M anomalous patterns – compare for example Fig. 7.12a and Fig. 7.12b and c in Eastern Siberia. I come back to this observation below.

The summer averaged **E**-vector at 200hPa for centennial events is shown in Figure 7.13 (cf. Figure A.36 for the 500hPa **E**-vector). For all periods the difference with the control simulation is not as strong as in the pre-industrial case (cf. Figure 6.8). For the present simulation there is for example more anticyclonic wave breaking

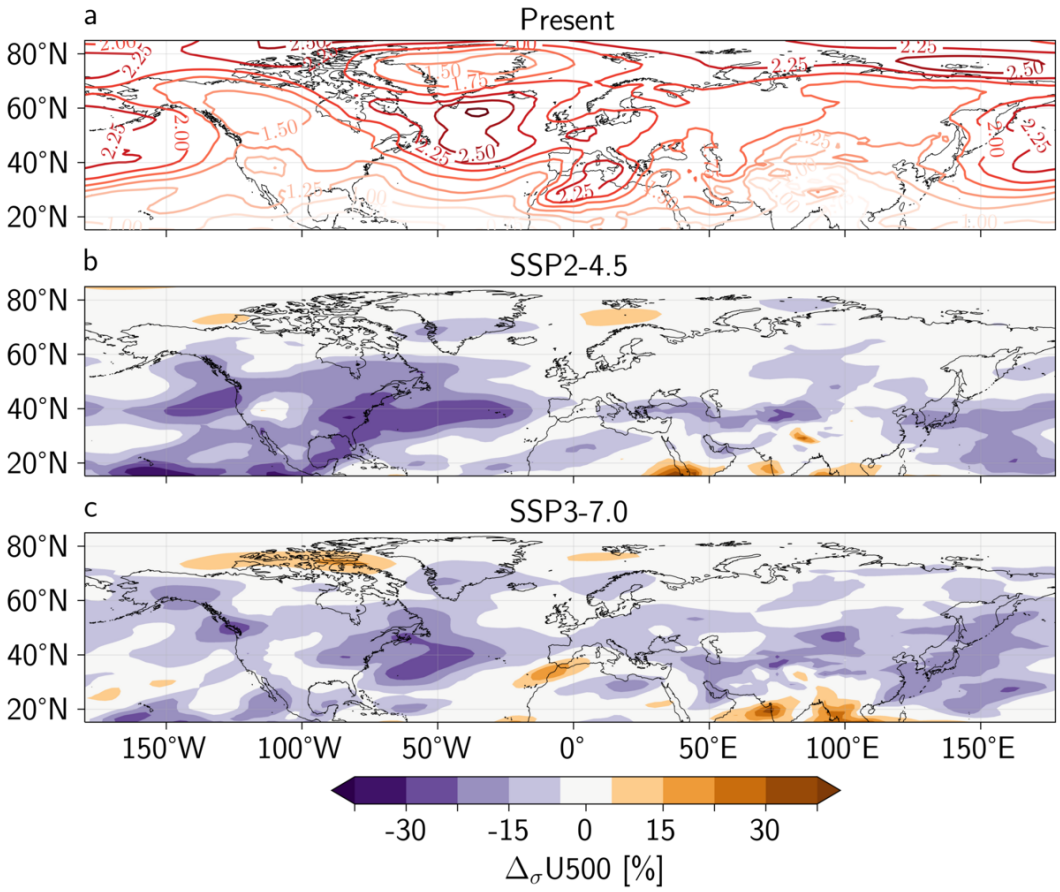


Figure 7.10: **Standard deviation of the summer averaged 500hPa zonal wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

in the Western part of the North Atlantic in the biased rather than the control simulation, which is the opposite of what was found for the pre-industrial case. Similarly, the strength of the anticyclonic wave breaking over Eastern Europe is smaller than in the pre-industrial simulation. Finally, there is a slight northward shift and decrease of the intensity of the eddies from the present to the future scenarios.

Figure 7.14 shows the summer averaged anomalies of Z500 (contours) and upper-level soil moisture (colors). As previously, the patterns are close to the ones obtained for the pre-industrial case (cf. Fig. 6.9) and there is minimal change from the present to the future scenarios. There is nevertheless a decrease of the spatial area with very low (< -1.5 std) soil moisture over Central Europe – see especially Fig. 7.14b for SSP2-4.5 — which is likely due to the model lower bounds for soil moisture. The strong anomalies of Z500 are concentrated in Western Europe for all periods

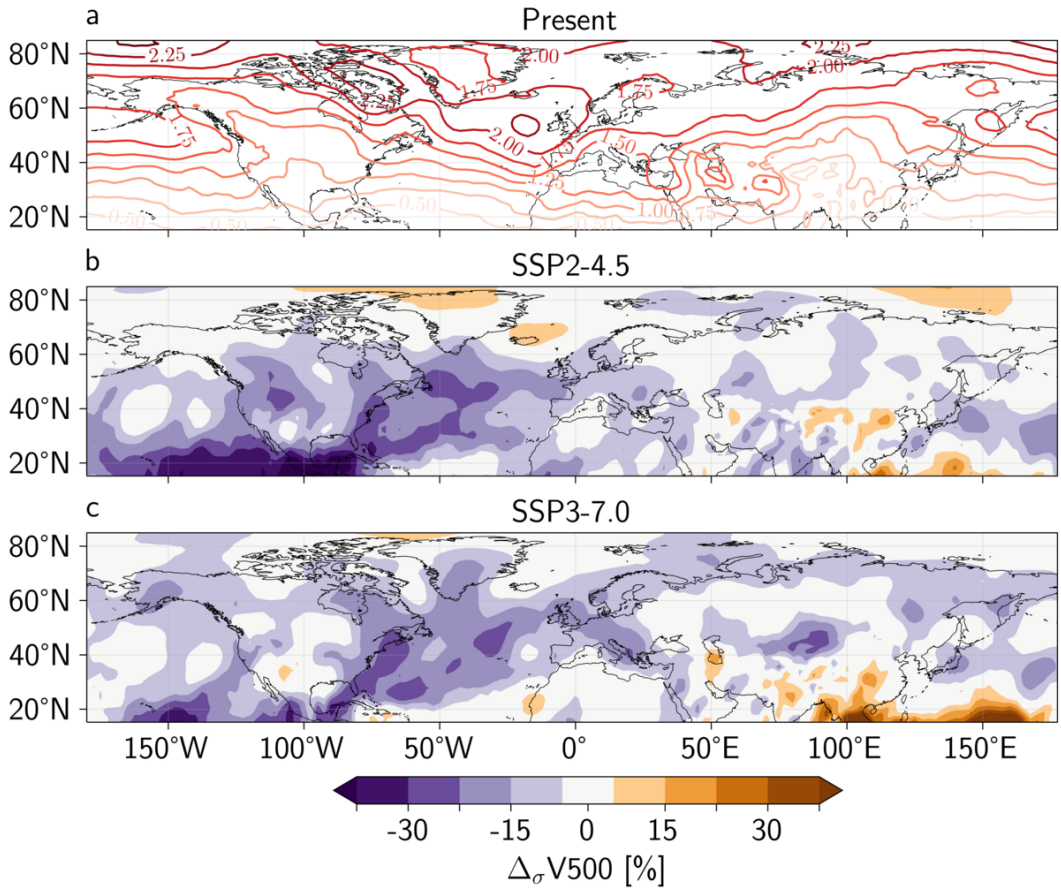


Figure 7.11: **Standard deviation of the summer averaged 500hPa meridional wind in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

and extend only moderately elsewhere. As noted previously, most of the strongest anomalies of soil moisture are found east of the grid point where the temperature is maximized. For SSP3-7.0, there is an extended but moderate negative anomaly of Z500 East of Greenland, and more generally in the polar and sub-polar regions, that is not present in the other periods.

Figure 7.15 shows a proxy of the vertical structure of the summer averaged atmosphere using SLP, Z850 and Z200 fields as in Figure 6.10. Similar to the pre-industrial simulation, the vertical structure is not barotropic but rather slightly tilted westward. In Europe the atmospheric patterns are remarkably similar between the past, present and future periods. The only major change is the low pressure system west of Greenland for SSP3-7.0 as noted previously.

For the simulations run in this chapter I have access to the outputs at various

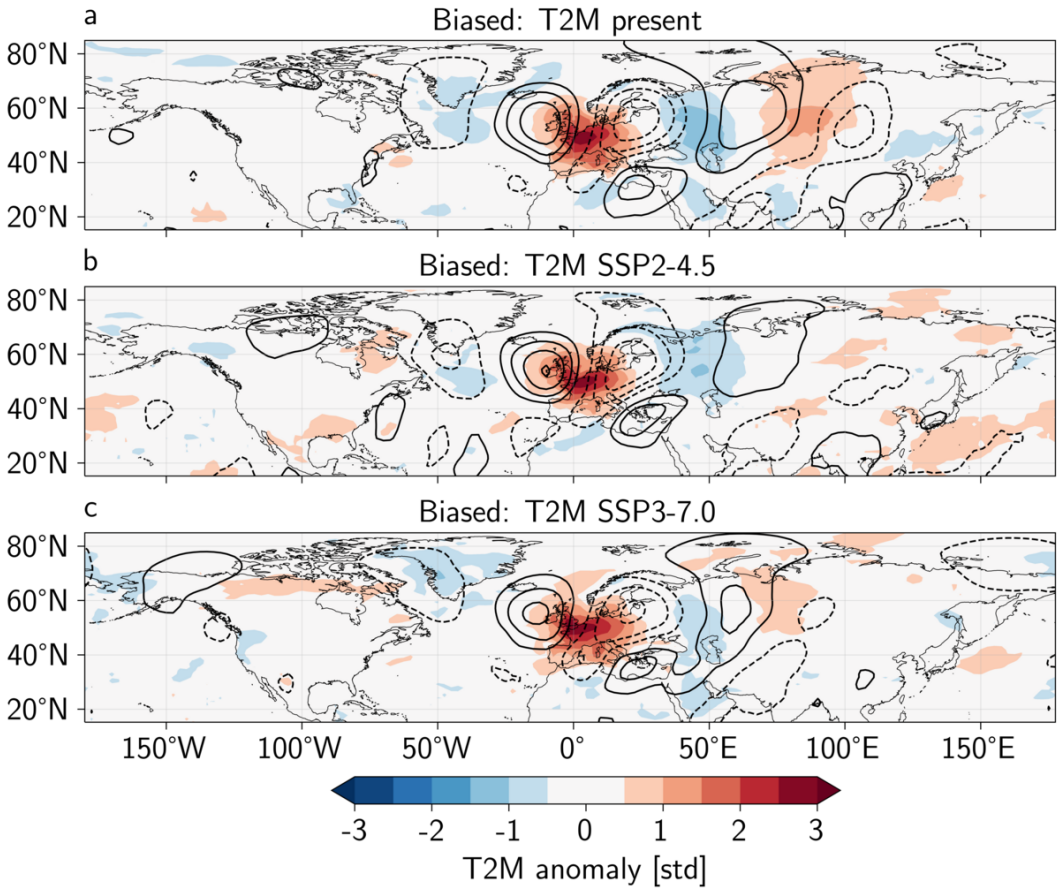


Figure 7.12: **Summer averaged normalized anomalies of 2-m air temperature (colors) and meridional wind at 200hPa (contours) for centennial events.** For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

levels in the atmosphere. Figure 7.16 therefore presents a longitudinal cross-section averaged between 40 and 60°N of the summer averaged normalized anomalies of air temperature (first row) and meridional wind (second row). The strong anomalies of air temperature extend over the troposphere up to the 300hPa level, with a tilted westward structure. Above the 200hPa level, a negative anomaly of temperature is present above the heatwave region in the stratosphere as a result of the anticyclonic anomaly below. The amplitude and spatial extension of the positive temperature anomaly close to grid point 1 are similar in the three periods but there is a major change in the longitudinal structure of anomalies. Whereas in the present there is an alternating pattern of 5 regions with significant air temperature anomalies, it is reduced to around 4 in SSP2-4.5 and to 2 in SSP3-7.0. This shows that extremely hot summers are more dynamically local in the future than in the present. This

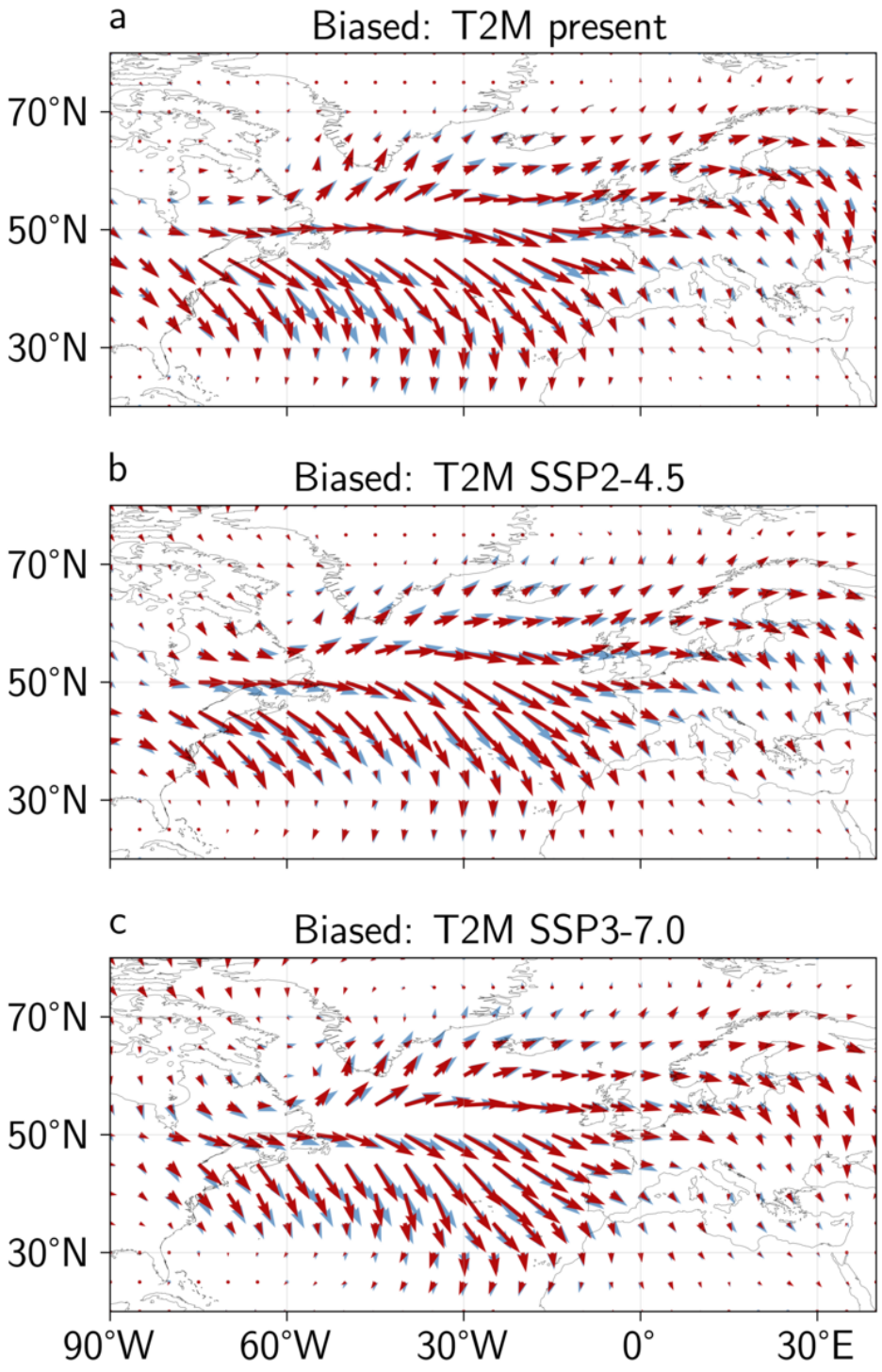


Figure 7.13: **Summer averaged E-vector at 200hPa for centennial events.** In each panel the blue arrows show the climatological **E**-vector computed on the control simulation.

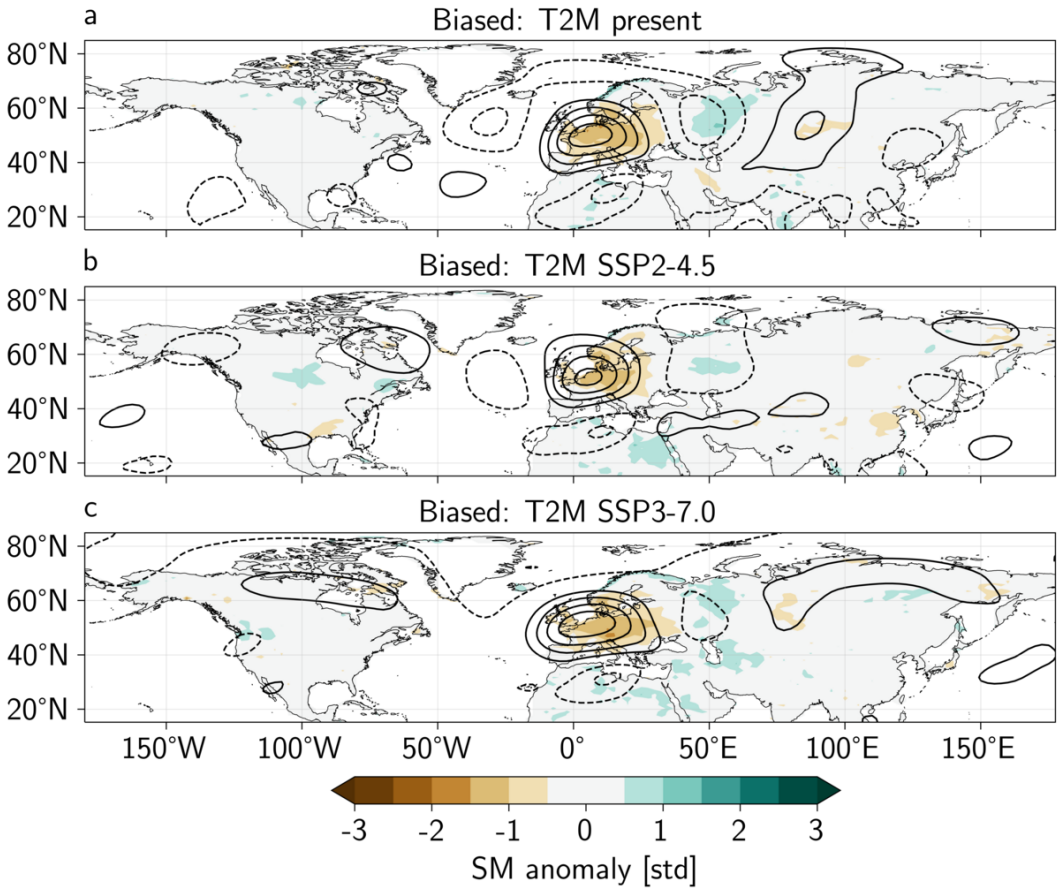


Figure 7.14: **Summer averaged normalized anomalies of upper-level soil moisture (colors) and geopotential height at 500hPa (contours) for centennial events.** For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

results is confirmed by the second row showing meridional wind speed anomalies. The anomalies extend barotropically over the entire troposphere and the beginning of the stratosphere in all periods. However, the anomalies are more local and do not extend as far longitudinally in the future scenarios compared to the present simulations.

Figure 7.17 shows the same vertical cross-section with specific humidity (first row) and geopotential height (second row). The increasing locality of the anomaly with future scenarios is striking for the geopotential height (Fig. 7.17a2 vs Fig. 7.17b2 and c2). The westward tilted structure of the anticyclone is also confirmed in this plot: anomalies of the geopotential height are on the east of grid point 1 close to the ground but above it or slightly to its west at 200hPa. The vertical pattern of anomalies of specific humidity corroborates the hypothesis stated in chapter 6: the

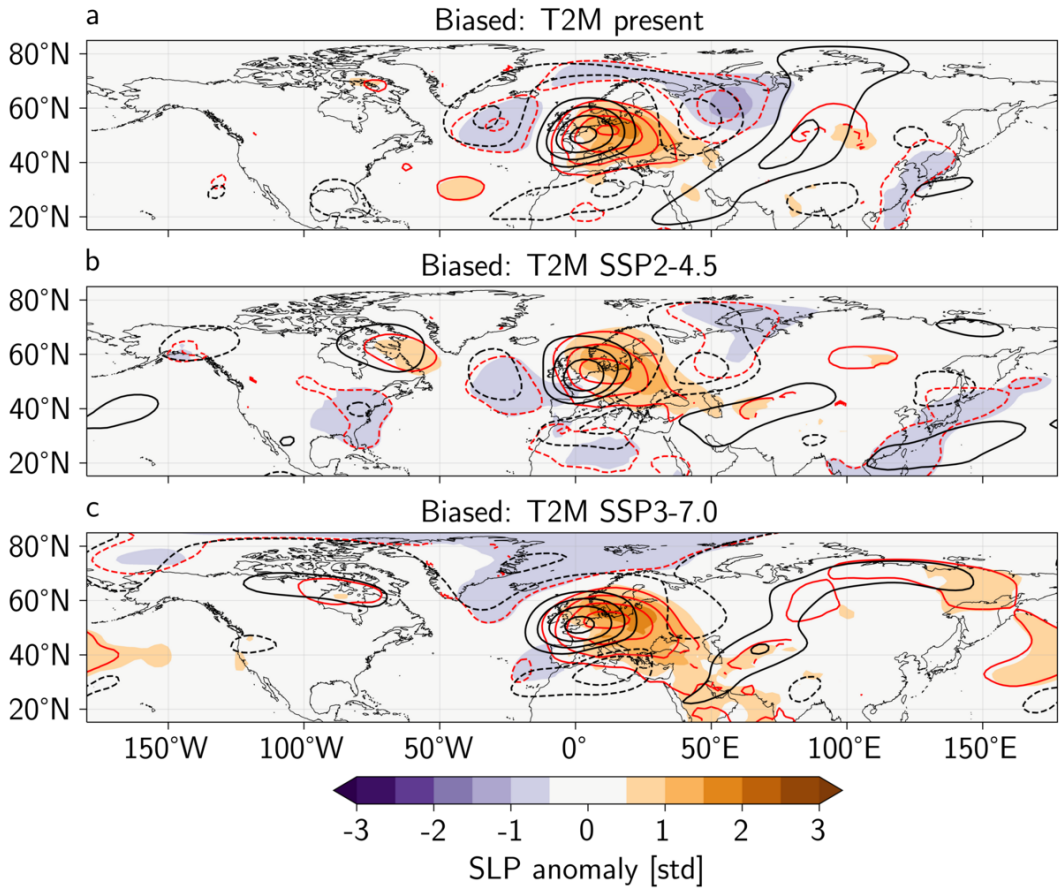


Figure 7.15: **Summer averaged normalized anomalies of sea-level pressure (colors), geopotential height at 850hPa (red contours) and at 200hPa (black contours) for centennial events.** For all fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The contours are drawn every 0.5 standard deviation starting at +/- 0.5. Plain (dashed) lines represent positive (negative) values of the standard deviation.

strongest positive anomalies are to the west of grid point 1 and above the boundary layer (above the 850hPa level), while the strongest negative anomalies are to the east of grid point 1 between 850 and 1000hPa. This strongly suggests advection from the south west above the boundary layer but from the east close to the surface.

The desiccation of soils during heatwaves results in the reallocation of incoming solar energy from latent towards sensible heat, enhancing diabatic heating of the lower layers of the atmosphere. Figures A.37 and A.38 show respectively the summer averaged normalized anomalies of latent and sensible heat fluxes. Consistent with the low soil moisture patterns of Figure 7.14, the latent heat fluxes anomalies are negative and the sensible heat fluxes are positive east of grid point 1 — but only moderately strong at grid point 1 where the temperature is yet maximized. Heat fluxes are especially strong for SSP3-7.0 for latent heat fluxes. This observation is

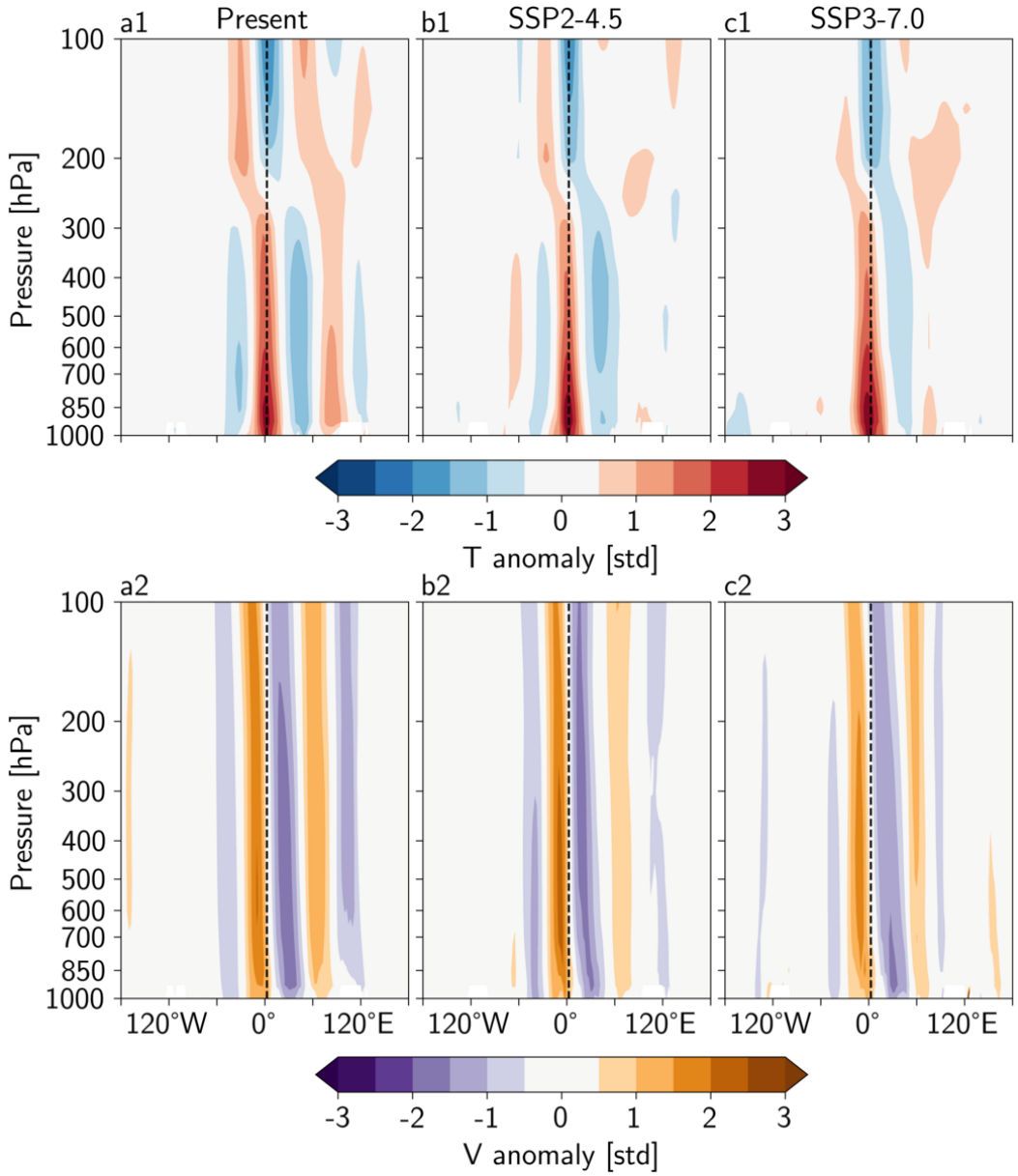


Figure 7.16: **Vertical cross-section of summer averaged normalized anomalies of air temperature and meridional wind for centennial events.** Columns corresponds to the biased T2M simulation in (a) present, (b) SSP2-4.5 and (c) SSP3-7.0. For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The cross-section is computed after a latitudinal average between 40 and 60°N. The black dashed line shows the longitude of grid point 1.

coherent with the increasing locality of the extreme temperature events in the future scenarios: it suggests that future extremes are more driven by local diabetic rather than adiabatic warming caused by specific atmospheric patterns. These observations

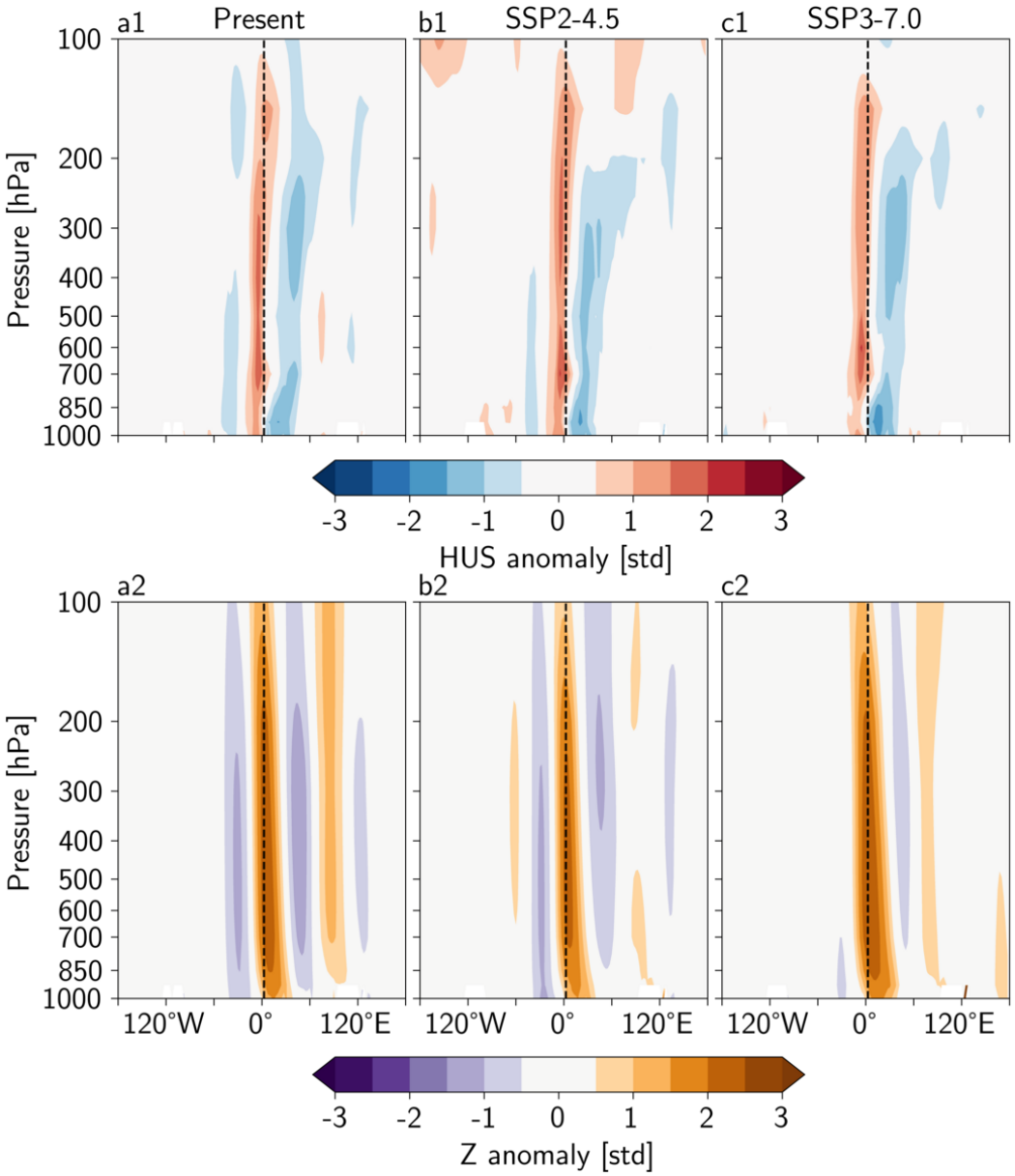


Figure 7.17: **Vertical cross-section of summer averaged normalized anomalies of air specific humidity and geopotential height for centennial events.** Columns corresponds to the biased T2M simulation in (a) present, (b) SSP2-4.5 and (c) SSP3-7.0. For both fields, normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation. The cross-section is computed after a latitudinal average between 40 and 60°N. The black dashed line shows the longitude of grid point 1.

with the biased simulations are coherent with the increasing variability in thermodynamic variables while increasing variability in the dynamic variables found in the previous section. As a result, the anomalies of the boundary layer height are stronger

in the future rather in the past (cf. Figure A.39 in supplementary materials).

7.3.4 Atmospheric dynamics

In this section I investigate the atmospheric dynamics leading to the anomalies and the average spatial patterns shown in the previous sections. Figure 7.18 shows the wave-number-phase speed spectra of meridional wind anomalies at 200hPa for centennial events (cf. Figure A.40 for meridional wind speed anomalies at 500hPa). The difference with the pre-industrial case (Fig. 6.11a) is quite striking: although the increase in westward moving/quasi-stationary waves with 4-5 wavenumbers is still seen for the present simulation it mostly disappears for SSP2-4.5 and SSP3-7.0 simulations. On the contrary, these simulations have weak anomalies (<10%) compared to their control simulations.

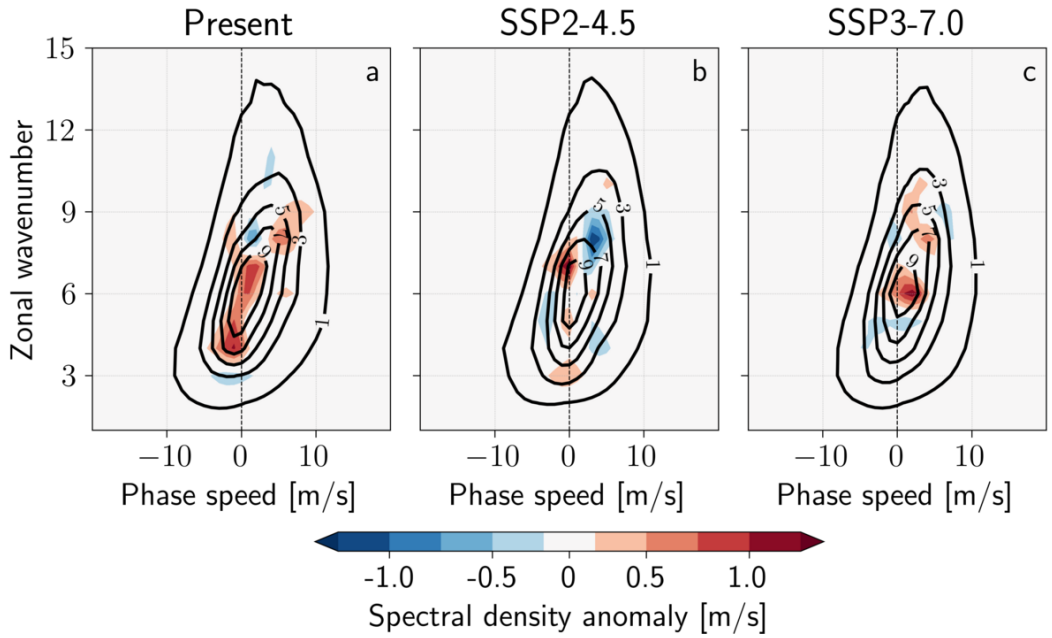


Figure 7.18: **Wavenumber-phase speed spectra of meridional wind anomalies at 200hPa for centennial events.** Raw spectrum (contours) and difference with the climatological spectrum (colors).

The distribution of quasi-stationary Rossby waves is investigated through an amplitude-phase diagram of meridional wind speed anomalies at 200hPa for wavenumbers $k_\phi = 3 - 10$ in Figure 7.19 (cf. Fig. A.41 for the 500hPa version). The 2D histograms are shown only for the biased simulations: by definition the anomalies for the control simulations do not have preferred phases. The aggregated version of these histograms along the amplitude and phase axis are presented in Figure A.43 (cf. Fig. A.44 for the 500hPa version). The distributions are similar for all periods and all wavenumbers, including when compared to the pre-industrial period (cf. Fig. 6.12). There is a moderate difference for wavenumber $k_\phi = 6$ for the present simulation, which has preferred phase around 0 rather than $\pi/2$ as in

the two other periods, but this may only be due to the variability of the trajectories sampled and not to a real physical difference. There is not a strong difference for this wavenumber in the aggregated spectrum (cf. Fig. A.43a2).

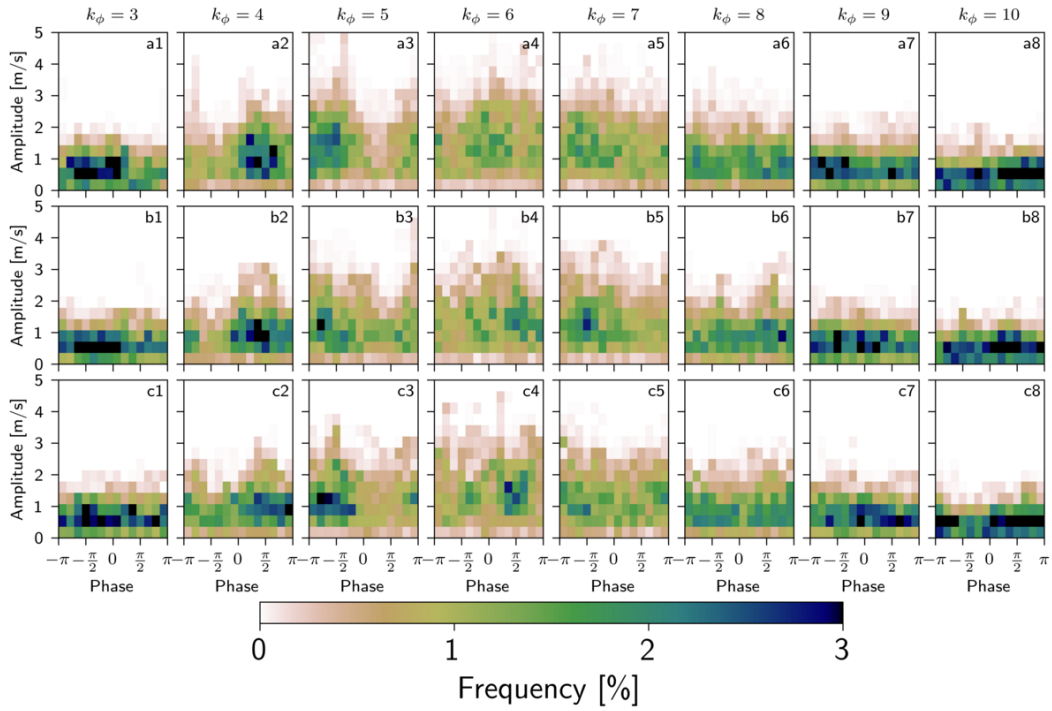


Figure 7.19: **Amplitude-phase histograms of meridional wind speed anomalies at 200hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

The histograms for the non-anomalous fields are presented in Figure A.42 (Fig. A.45 for the 500hPa level) and their aggregated versions are shown in Fig. A.46 for the 200hPa level and Fig. A.47 for the 500hPa level. The climatological histograms computed on the control simulations are provided in supplementary material (cf. Fig. A.48 for the 200hPa level and Fig. A.49 for the 500hPa level). The stationary waves are constant from the pre-industrial to present and future scenarios, for either the control or the biased simulations – the latter being closed to the climatology of the former. This is expected as they mainly arise as a result of the orography and land-oceans distribution of the Northern Hemisphere. The only difference is again a phase shift for wavenumber $k_\phi = 4$ for the present biased T2M simulation as already noted for the pre-industrial biased T2M simulation. This shift is clear in the aggregated form of the spectrum (cf. Fig. A.46a2). For the two future scenarios there is more spread in the phase for this wavenumber than in their respective control simulation, but there is not clearly a phase shift of the distribution.

On the contrary, the composite Hovmöller plots of both meridional wind speed anomalies at 200hPa and 2-m air temperature normalized anomalies for heatwave events in the biased simulations (Figure 7.20) show that heatwaves arise as a result

of the amplification of a short Rossby wave packet (RWP) in all periods. The RWP is also stronger and more extended in the present simulation compared to the future ones, which reinforces the interpretation in terms of locality of future temperature anomalies. This is however not seen on the control simulations (cf. Figure A.50), for which there are nonetheless around four times less heatwaves than in the biased simulations.

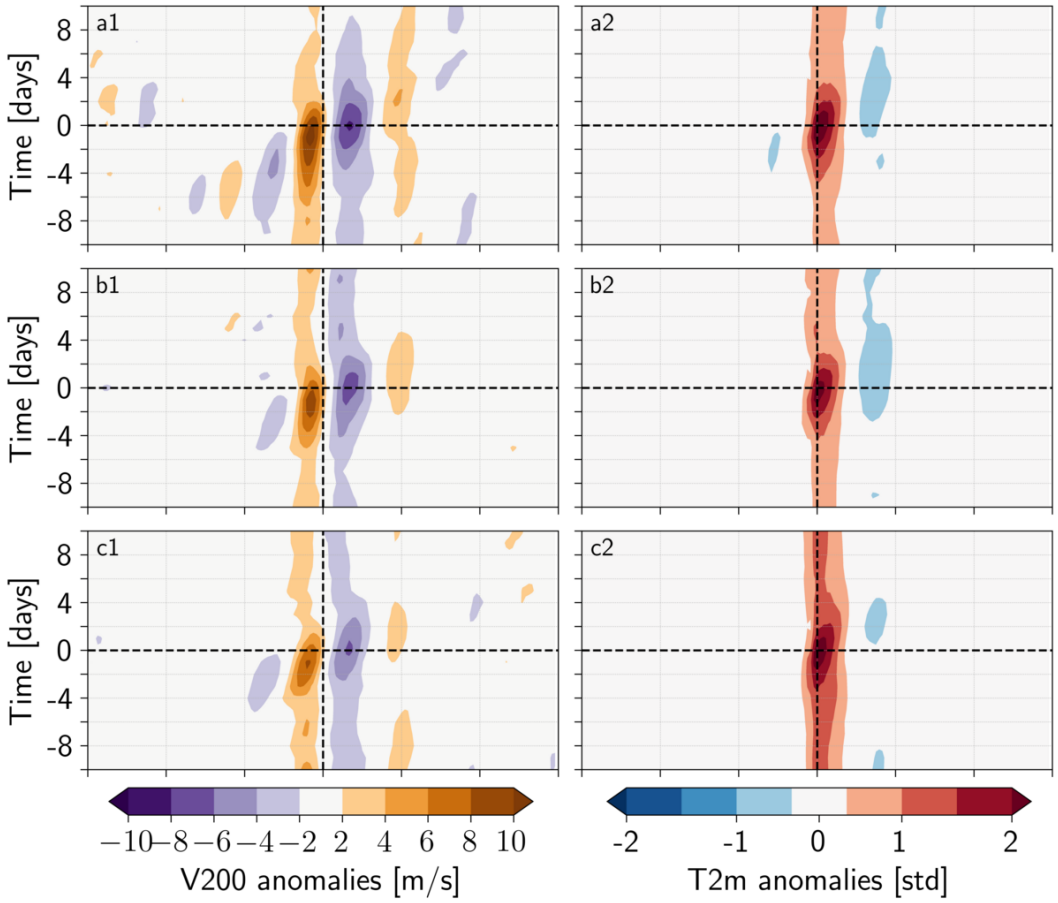


Figure 7.20: **Composite Hovmöller plot of the atmospheric and surface dynamics during heatwave events for the biased simulations.** First column: meridional wind speed at 200hPa anomalies averaged between 35°N and 65°N. Second column: 2-m air surface temperature normalized anomalies averaged between 45°N and 55°N. The time is expressed relative to the hottest day of each heatwave event. Rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0. Number of events: (a) $n = 3733$, (b) $n = 3929$ and (c) $n = 4513$.

As previously, I investigate the presence of cut-off lows in the biased T2M simulations. Figure 7.21 presents the frequency and frequency anomalies of cut-off lows for the biased simulations in the present and two future scenarios. For the present and SSP2-4.5 scenario there is still a positive anomaly of cut-off lows west of the Iberian peninsula, but it is much reduced compared to the pre-industrial case (cf. Fig. 6.15a1–2). Overall the total number of cut-off lows over the North-Atlantic

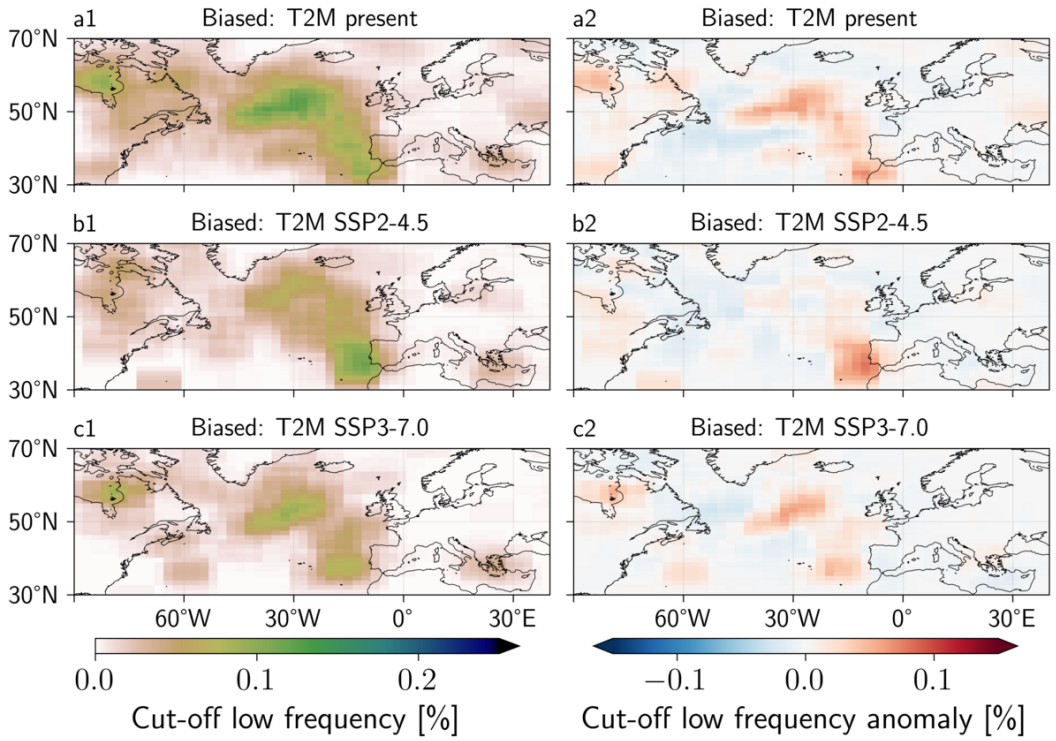


Figure 7.21: **500hPa cut-off lows frequency and frequency anomaly.** The anomaly is computed with respect to the control simulation.

basin is diminishing when evolving from pre-industrial to present and then future scenarios. For SSP3-7.0, the positive anomaly of cut-off lows frequency west of the Iberian peninsula is also smaller and slightly shifted westward.

This is confirmed by Figure 7.22 which shows the cumulative frequency of cut-off lows when a heatwave is occurring at grid point 1 in the control and biased T2M simulations. In the control simulations, there is a clear decrease of the cumulative percentage of the number of cut-off lows west of the Iberian peninsula from the present to the SSP2-4.5 and SSP3-7.0. For the present this number is already lower than the value in the pre-industrial period (cf. Fig. 6.17): around 13% vs around 20%. For the biased T2M simulation, the cumulative number of cut-offs when an heatwave is occurring at grid point 1 is similar but slightly higher (SSP2-4.5 and SSP3-7.0) to the control simulations. The role of cut-off lows for reaching high temperatures in Western Europe is therefore still present but much reduced in the future scenarios according to the simulations made here. This results is also coherent with the increased variability in thermodynamics and decreased variability in dynamics observed previously.

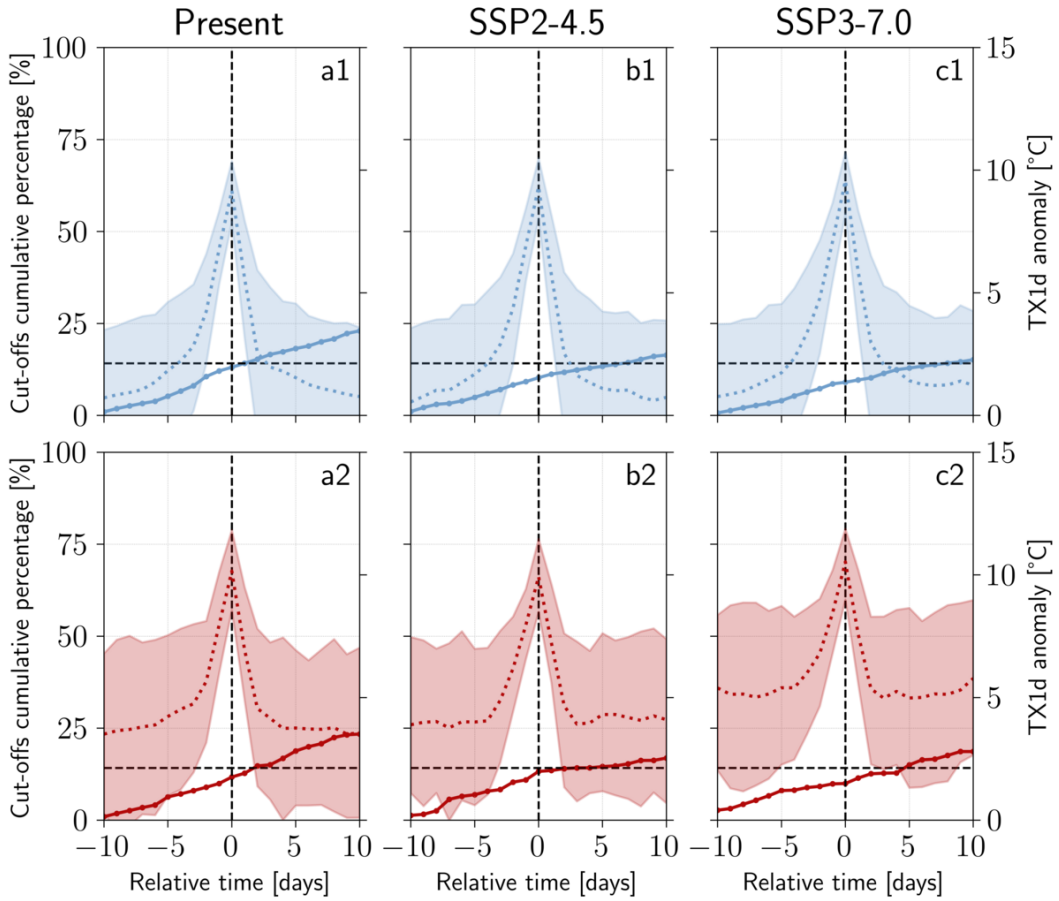


Figure 7.22: **Occurrence of an Iberian cut-off low when a heatwave is happening at grid point 1 at $t = 0$.** A heatwave is a continuous period of time when the daily mean temperature is above the 90% quantile of the daily distribution of T2M for at least three days. The cumulative percentage of cut-offs at time t gives the percentage of heatwave events which have had at least one Iberian cut-off in the last $t + 10$ days. The shadings for the anomaly of maximum daily temperature TX1d shows the 25 and 75% quantiles. The dashed horizontal line shows the cut-offs cumulative percentage at $t = 0$ for the present control simulations. Number of events: (a1) $n = 1089$, (b1) $n = 1091$, (c1) $n = 1085$, (a2) $n = 3733$, (b2) $n = 3929$ and (c2) $n = 4513$.

7.4 Conclusions and discussion

In this chapter I have expanded the application of the GKTL rare events algorithm presented in chapter 6 to present conditions and two future scenarios for the external forcings of the IPSL-CM6A-LR model in its atmosphere - land surfaces configuration. As previously, I have shown that the algorithm is able to sample efficiently very rare summers for a much reduced computational cost (see the 6.4 for the discussion on this point). This allows to draw physical conclusions on the thermodynamical and dynamical mechanisms leading to centennial-like hot summers in Western Europe

in the model and how they change with climate change.

Statistically, I found that the anomaly corresponding to a centennial hot summer is constant in the pre-industrial, present and two future scenarios: around +2.4°C with respect to the mean at grid point 1. I however obtained conflicting results on the shape of the right tail of the distribution of summer averaged temperature at grid point 1: there seems to be a departure from Gaussianity in the pre-industrial and SSP3-7.0 simulations but not in the present and SSP2-4.5 simulations which both closely follow the Gaussian approximation fitted on the control. Also the number and mean duration of heatwave events — i.e. events with at least three days above the 90% climatological quantile of daily mean near-surface temperatures — is similar in all periods. Consequently, the GKTL rare events algorithm sample similarly intense hot summers at grid point 1, compared to the control simulations, in the past, present and future configurations. There is nevertheless more and longer heatwaves for more intense global warming scenarios.

Dynamically, I found mostly similar results as the one obtained in the pre-industrial case: the vertical structure of the anticyclone above the hot grid point is not barotropic, the anomalous low soil moisture pattern associated with the hot summers is shifted to the east and the atmospheric pattern of winds associated with the hot summers is of synoptic rather than planetary scale and related to the local amplification of short Rossby wave packets. The dynamics sampled with the rare events algorithm is mostly similar to the typical dynamics found in chapter 5 on a 2000-y pre-industrial control run of the same model. This gives confidence in the relevance of the dynamics sampled by the algorithm in the pre-industrial period — and therefore also in other periods.

However, I also found important dynamical changes happening between the four periods. The main result obtained on the control simulations is a consistent increase in the summer variability of thermodynamical variables (2-m air temperature, upper-level soil moisture and as a consequence surface sensible and latent heat fluxes) and decrease in the summer variability of dynamical variables (winds at 200 and 500hPa). As a result, I showed that the typical longitudinal extension of the anomalous patterns of winds and geopotential height extending through the troposphere during hot summers is reduced in future warming scenarios. Also, the number of cut-off lows — especially Iberian cut-off lows — during heatwave events halved in future simulations compared to the pre-industrial case. These results point toward a more important role of local diabatic fluxes rather than the large scale organization of the atmosphere for creating hot summers in a warmer world (see also Teng et al. (2016)). This conclusion is likely to be extended to heatwaves in general in so far as hot summers appear primarily as a result of a succession of heatwaves during the summer.

It should be noted here that the decrease of variability in the dynamics alone should decrease the intensity of heatwaves (the temperature distribution would shrink), while it is the contrary for the increase of variability in thermodynamical variables (the temperature distribution would spread). The fact that the intensity of the anomaly of a centennial-like event is mostly constant through out the simulations may result as an equilibrium between these two competing mechanisms.

Validating this hypothesis is not straightforward, but one way of investigating this question could be to nudge wind patterns corresponding to centennial-like summers in the present with the thermodynamics of the future, and vice versa. How the anomaly corresponding to a centennial-like events change in the two cases could show whether the 'compensating hypothesis' is correct.

To conclude these two chapters on the statistical and dynamical aspects of very hot summers in Western Europe in the IPSL-CM6A-LR model, the natural question one may ask is how much these results are a faithful representation of the actual climate system. The IPSL-CM6A-LR is biased (Boucher et al. 2020) — although the biases are small in Europe — and more generally climate models have difficulties to represent faithfully how intense extreme events can get (cf. chapter 1 section 1.3). Moreover, here I do not simulate the oceans whereas one should expect a feedback of the oceanic circulation on the atmosphere especially considering the intensity of the anomalies simulated in the atmosphere. As explained previously, the main issue is however that in reanalysis data there are no, or very few, events which are as extreme as the ones sampled here. As a consequence, the burden of rareness also applies to real world data and there is simply no data to know whether the dynamics sampled by the model is correct or not. What makes the results obtained here not completely worthless is that I mainly sample a lot of moderately intense heatwaves and the dynamics sampled by the model is mostly coherent with the results found in the literature on heatwaves (cf. chapter 1 section a)). Consequently, it is probably reasonable to assume that the dynamics sampled here could happen in the climate system and that the simulations made here can be used for adaptation purposes.

Summary

Context and goals

How the atmospheric dynamics may change in the future is a climatological question whose answer can have tremendous impacts on societal adaptation to climate change. In this chapter I investigate the evolution of physical mechanisms leading to heatwaves with global warming.

Methods

I apply a rare events algorithm in a climate model to sample extreme hot summers under present and end-of-the-century conditions for two scenarios of anthropogenic forcings (SSP2-4.5 and SSP3-7.0).

Results

In the control simulations in the three periods, I show both a decrease in the variability of the atmospheric circulation over the North Atlantic in summer and an increase in the variability of thermodynamical surface quantities such as soil moisture, temperature and local energy fluxes. With the simulations maximizing near-surface air temperature in Western Europe, I demonstrate that the dynamics leading to heatwaves tend to be more local and less large scale-organized. In the future, high temperatures are still reached via a large anticyclone over Western Europe, but anomalies do not extend as far longitudinally as in the present and arise as a result of an increase in the variability of thermodynamical variables such as soil moisture and energy fluxes.

Chapter 8

Involvement in other scientific projects

8.1 Introduction

During my PhD I was involved in several collective scientific projects not directly related to the main topics of this thesis. In this chapter I give some details on these projects and on my contribution to each of them. For the projects which lead to the publication of an article in which I am first author the published version of the article can be found in annex to this thesis.

8.2 First author articles

8.2.1 Extreme rainfalls in Australia

This work was published in the *Asia-Pacific Journal of Atmospheric Sciences* and is available in annex B. In February and March 2022, the eastern coast of Australia recorded an unprecedented amount of precipitation with extended floods and damages to properties. We investigated the statistical and dynamical aspects of this event with the aim of providing an attribution statement. We employed both a classical statistical approach based on the difference in return times after the fit of a GEV in a past and present period and a more recent analog attribution methodology which conditions the attribution on the occurrence of similar atmospheric patterns (see also below). The main variable of interest is the 3-day averaged precipitations. By defining the 1948-1977 period as counterfactual and the 1990-2019 period as factual, we find that the dynamics of the event consisted in an unprecedented combination of several factors: a tropical atmospheric river, the presence of the Coral low pressure system and a blocking anticyclone offshore Eastern Australia.

An original contribution of the paper is to propose both a past vs present and an El Nino vs La Nina discrimination to define factual and counterfactual periods. Our main finding is however that no clear attribution statements can be made, both because of the unprecedented nature of this event, the lack of long high quality avail-

able data and the dependence of the results on the La Nina phase of the El Nino Southern Oscillation (ENSO). This study does not make use of climate models to provide a strong attribution statement. The biases in the representation of the current and future evolution of ENSO in climate models (e.g. Sobel et al. (2023)) indeed makes the reliance on models for attribution difficult in this region of the world. My contribution to this paper was to run the analogs attribution methodology, propose the El Nino vs La Nina discrimination for the attribution methodologies and write the paper.

8.2.2 Evolution of the jet stream variability

This work was published in *Climate Dynamics* and is available in annex C. During my research internship before the beginning of my PhD I worked on a simplified model of the jet stream (Faranda et al. 2019). I investigated the patterns of variability of a simple reduction of the low-level eddy-driven jet based on ERA5 reanalysis data and found intriguing results. We wanted to continue this work further on and I therefore supervised two interns — Vivien Guette and Akim Viennet — during the first year of my PhD. The idea of their work was to investigate with new metrics (Faranda et al. 2017b) from dynamical systems theory the evolution of the variability of the low-level eddy driven jet stream over the Euro-Atlantic basin. Our work led to the publication provided in annex.

Although the eddy-driven jet is mostly zonal in climatological mean, it actually meanders at meteorological time scales. Large meanders of the jet are often associated with extreme events in the mid-latitudes (cf. chapter 1 section a)). How the spatial and temporal variability of the jet will evolve in the future is therefore a crucial and hotly debated question. In this work, we studied the eddy-driven jet stream over the Euro-Atlantic sector and assess its dynamical properties in ERA5 and ERA20C reanalysis data set using indicators from dynamical system theory and more classical metrics of its mean characteristics. The ERA20C reanalysis data set spanning the 1900-2010 period is in particular used to evaluate the long term evolution of variability metrics. There are important inter-decadal fluctuations but apart from the mean speed of the jet — which is increasing almost linearly with time — it is difficult to see a large evolution over the 20th century despite a global warming of around 1°C. We therefore proposed a linear model to control for global modes of variability and aerosols emissions to disentangle the sole impact of global warming from the impact of natural variability of the climate system on the jet. Our main result is that there is a statistically significant signal in the direction of increasing speed and stability of the low-level jet. Our metrics show a consistent decrease in the variability of the jet, especially in summer. This suggests a zonalisation of the jet under global warming. Interestingly, our results are coherent with the results of the control simulations presented in chapter 7 which also showed a reduced variability of the atmospheric circulation in the future in summer. My contribution to this work was to design the study — especially the variability metrics and the linear model to control for natural variability —, supervise the interns who worked on this project and write the paper.

8.3 Co-author articles

8.3.1 Climate-change attribution retrospective using the analogs method

This work was published in *Weather and Climate Dynamics* (Faranda et al. 2022). We address the question of the role of climate change in key extremes that occurred in year 2021 over Europe and North America: the Winter Storm Filomena, the French spring cold spell, the Westphalia floods, the Mediterranean summer heat wave, Hurricane Ida, the Po Valley tornado outbreak, Medicane Apollo and the late-autumn Scandinavian cold spell. We used the ERA5 reanalysis dataset from 1950 to present and employed an analog methodology finding most similar sea-level pressure patterns to the extreme events of interest in the factual and counterfactual worlds. We then compute significant shifts in the spatial characteristics, persistence, predictability, seasonality and other characteristics of these analogues. We also diagnose whether in the present climate the analogues of the studied events lead to warmer/cooler or dryer/wetter conditions than in the past.

The main conclusion of the article is that most of the extreme events we investigate are significantly modified in the present climate with respect to the past, because of changes in the location, persistence and/or seasonality of cyclonic/anticyclonic patterns in the sea-level pressure analogues. My contribution in this paper was to run the analogs attribution methodology for the Westphalia floods case and write the corresponding part in the paper.

8.3.2 Ensembles of simulation of extreme heatwaves for Paris Olympics

This work was published in *npj Climate and Atmospheric Science* (Yiou et al. 2023). We employed the rare events algorithm based on analogs of the atmospheric circulation described in chapter 4 section a). We investigated the most extreme heatwaves in Ile-de-France that are physically plausible, under climate change scenarios, for the decades around 2024, finding analogues in the CMIP6 runs of different models.

The main conclusion of this article is that the 2003 record can be exceeded by more than 4°C in Ile-de-France before 2050. As explained above, the algorithm employed here is well suited when a large dataset such as CMIP6 is available and we want to investigate long-lasting events. My contribution to this paper was in the writing of the paper and the analysis of the outputs of the analogs rare events algorithms.

8.3.3 A statistical physics and dynamical systems perspective on geophysical extreme events

This work is currently under review at *Physical Review E* (Faranda et al. 2023a). This paper is a perspective follow-up of discussions held during the UNDERPIN

symposium in Erice, Italy in 2022 that I had the opportunity to attend. The symposium brought together climatologists, statistical physicists and mathematicians on the subject of extreme events. The contributions cover a broad spectrum from statistical physics and dynamical systems theory. The paper delineates knowledge gaps, presenting some related challenges and new formalisms which arise in the study of geophysical extreme events and may help better understand them. My contribution in this paper is a joint work with Caroline Wormell in which we proposed an estimator of the phase space gradient of a quantity called local dimension which measures the clustering of points in the phase space of a general dynamical system. We additionally propose an application to atmospheric circulation data with the NCEP reanalysis dataset. This work is still exploratory but could in principle be applied to explain what kind of structures in a large scale synoptic pattern make a pattern rare with respect to the climatology.

8.3.4 Dynamical changes in tropical Indo-Pacific

This work is currently under review at *Nature Geoscience*. This project lead by Chenyu Dong and Gianmarco Mengaldo investigates thermodynamical and dynamical changes in the tropical Indo-Pacific region in the ERA5 reanalysis dataset and relates this changes to the evolution of extreme events. The method is based on the detection of analogs of different fields (mainly total column water vapor and 850hPa streamfunction) and the investigation of their increasing or decreasing frequency. The method was proposed by Faranda et al. (2023b) and is here applied to another region of the world. We show that, over the past 80 years, the strengthening of the Pacific Walker circulation — consisting of rising air motion over Southeast Asia and descending air motion over the eastern Pacific — has led to a steady increase in weather patterns associated with extreme weather in the region, specifically heatwaves and extreme precipitation. This result is coherent with a more La Nina-like phase of the ENSO oscillation in the recent years. My contribution to this work was to make part of the code for the computation of the streamfunction and the detection of analogs increasing/decreasing in frequency. I was also largely involved in the writing of the paper and the discussion on the detection vs attribution results that we show.

Conclusion

Summary and conclusions reached

In this PhD I have been interested in **the statistics and dynamics of extreme and very extreme heat events in the mid-latitudes in general and in Europe** in particular. I have addressed the questions of **the maximal near-surface air temperature that can be reached during a heatwave event, the difference between the physical mechanisms leading to extreme vs very extreme heat events, the possibility to simulate efficiently extreme heatwaves in a climate model and the evolution of the climate system dynamics leading to extreme heatwaves with global warming.** To frame my work in a general physical and climatological context, I explored and proposed methods to explore the under-sampled regions of the climate attractor, i.e. provide climatological results for rare and extreme events.

In chapter 2, I investigated the question of the upper bound for surface temperatures in Western Europe with a statistical approach based on Extreme Value theory (EVT) and a physical approach based on the stability of the air column (Zhang and Boos 2023) using the July 2019 event as a case study. I showed that the statistical upper bound provided by EVT is likely largely underestimated. I also employed a flow analogues methodology to study the evolution of thermodynamical mechanisms influencing the intensity of the event and found that even when conditioning on similar atmospheric patterns as the one of the July 2019 event, daily maximum temperatures have largely increased between the past (1940-1980) and the present (1981-2021). This increase is much larger than what would suggest the mid-tropospheric warming observed between the two periods. I have attributed this increase to strengthened surface heat fluxes. The question nevertheless remains of the part of the observed change that is a result of a forced dynamics — i.e. by anthropogenic climate change — and the part that is the result of natural variability.

In chapter 3, I proposed the idea of physically constraining the fit of EVT-based statistical distributions for extreme temperatures. More specifically, the upper bound of the distribution is imposed by a physically realistic value derived from properties of the mid-atmosphere and surface humidity. I showed using outputs of a state-of-the-art climate model over the historical and near future periods that this partially alleviates the issue of the underestimation of very extreme events — typically centennial events — and, more importantly, avoids the appearance of climatic surprises, i.e. events that are said to be impossible but that still occur. I argued

that even though the method I proposed tends to overestimate very extreme events if the upper bound imposed is too high, under- or over-estimating very extreme events is not a symmetric error. While the former leads to linearly increasing costs of preparation, the latter can lead to catastrophic damages on societies that would face events they thought to be impossible.

In chapter 4, I proposed a review of so-called rare events algorithms, which are methods aiming to sample extremes in a dynamical system, and their applications in geophysical systems. I showed that there is a growing interest for these methods in the geophysical context thanks to their capacity to sample more efficiently rare and very rare events that are seldom observed even in long simulations. Nevertheless, there are currently many challenges to their application in complex climate models and I proposed some solutions to alleviates the latter.

In chapter 5, I studied the idea of a typical dynamics leading to extreme and very extreme events in the climate system. I used the 2000-y pre-industrial control run of a climate model and investigated how the climate system dynamics of the model leading to extreme heat events at several locations in Western Europe change when one looks at more and more intense events. I showed that, for most of these events, the variance of the dynamics leading to these extremes tend to decrease as the intensity of the events considered increases. In other words, very extreme events tend to have a more similar dynamics than extreme ones. This idea parallels the idea of the instanton, i.e. the existence of a most probable path leading to extremes. I proposed that the concept of the instanton should be considered as the climatological object defining the dynamics leading to extremes and could be used to investigate changes induced by external forcings such as anthropogenic climate change. For extreme heat events in this model, in Western Europe and under this configuration, my results also suggest that the dynamics leading to very extreme events corresponds to an amplification of the mechanisms leading to extreme events rather than the appearance of different mechanisms. This may however not be the case for other types of extremes.

In chapter 6, I applied a rare events algorithm to a climate model to sample extreme and very extreme summers under a pre-industrial configuration. I investigated the dynamics leading to extremely hot summers in Western Europe by biasing an ensemble simulation towards favoring summers with high near-surface temperatures, high geopotential height and low soil moisture. The rare events algorithm allows to recover climatological probabilities and I showed that I sampled events with probabilities of occurrence between 10^{-2} and 10^{-6}. I showed that hot summers are characterized by both more and longer heatwaves. These summers arise as a combination of local non-barotropic high geopotential height and non-local soil moisture anomalies. The atmospheric dynamics of these summers is mainly driven by a recurrence of short Rossby wave packets rather than circumglobal wave patterns.

In chapter 7, I applied the same rare events algorithm with present and two future scenarios of anthropogenic forcings, especially increased GHG concentrations. I sampled extremely hot summers in these three periods and provide an unbiased estimation of their (low) probabilities. In the control simulations under these three

configurations, I showed both a decrease in the variability of the atmospheric circulation over the North Atlantic in summer and an increase in the variability of thermodynamical surface quantities such as soil moisture, temperature and local energy fluxes. For the simulations maximizing near-surface air temperature in Western Europe, I demonstrated that the dynamics leading to heatwaves tend to be more local and less large scale-organized. In the future, high temperatures are still reached via a large anticyclone over Western Europe, but anomalies do not extend as far longitudinally as in the present and arise as a result of an increase in the variability of thermodynamical variables such as soil moisture and energy fluxes.

Perspectives

In the different chapters of this manuscript I have discussed perspectives of the work proposed. In this last section I summarize the main elements of these perspectives and propose more general ideas inspired by the analyses and methods I have used.

Extreme temperatures, EVT and risk assessment. The use of EVT in climate science is motivated both by its strong and universal mathematical foundations and by its capacity to give probabilities to unobserved events. It is therefore extensively used for, for instance, attribution studies and more generally risk assessment studies. My and others (Diffenbaugh 2020; Zeder et al. 2023) works have shown that with the typical size of climate relevant samples, the evaluation of EVT-based statistical distributions leads to a systematic underestimation of the probabilities and intensities of very extreme heat events. This strongly appeals for cautiousness in the use of these methods to provide present and future worst-case scenarios for adaptation. Physically constraining the statistical results of EVT seems extremely important, and justify the use of so-called storylines approaches (Shepherd et al. 2018), which, although they are usually unable to provide climatological probabilities, are able to exhibit actual physical trajectories leading to extreme and very extreme events. The physical constraint that I proposed in chapter 3 is only a first step in this direction and needs to be refined for temperature extremes and to be extended for other kind of extremes (precipitations, winds etc). In terms of epistemological position, it seems to me that one should be extremely cautious when stating that a particular return level for an extreme event is impossible: despite the modern extensive and real-time system of measurements of the Earth, the attractor of the climate system — and therefore the possibility of very intense events — is fundamentally under-sampled.

Typicality of extreme dynamics. The idea of a convergence of the climate system dynamics leading to very extreme events is mirroring the results of EVT which finds a universal convergence towards a similar law for maxima of random variables. It therefore suggests a paradoxical simplification of the dynamics of very extreme events and a partial overcoming of the burden of rareness. I provided some evidence for such a mechanism for extreme temperatures in the long control run of a

climate model, although I also showed that there are several practical discrepancies (presence of cut-off lows for some but not all extremes, no convergence of the variance for some fields etc). It would be interesting to investigate whether this idea of typical dynamics is also relevant for other types of extremes at other locations in the world, and especially how the typical dynamics vary spatially (Lucarini et al. 2023). This avenue of research is more theoretical but could also have practical applications. It could for example constitute a measure for comparing the dynamics sampled by different climate models. Additionally, it could provide the most probable dynamics for extremes of interest, similar to storylines, for adaptation purposes.

Use of rare events algorithms in a climate context. In the conclusion of chapter 6 I provided a critical review of the use of rare events algorithms based on my experience in applying a rare events algorithm with the IPSL model. I showed that the use of the GKTL algorithm to sample unprecedented, large, summer-long deviations of an observable is feasible and much more efficient than a regular brute force simulation. The algorithm works particularly well for score functions which have a 'red noise-like' behavior, i.e. for anomalies which results in the sum of small deviations over a long period of time. Near-surface air temperature and soil moisture fall into this category. The fact that I also succeeded in sampling large deviations of the geopotential height at 500hPa is nevertheless encouraging for applying the same method to other atmospheric dynamics-based score functions. On the other hand, the GKTL algorithm is not suited to sample short very intense events because the different members cannot be selected along the way and do not have enough time to diverge once a small perturbation has been added (cf. the discussion of this issue in chapter 4 section 4.4.3). For these applications, methods such as ensemble boosting are better suited (Fischer et al. 2023).

On the other hand, one cannot go over the fact that the use of a rare event algorithm for a particular score function of interest is still computationally expensive and even though it allows to reduce the variance in the tail of one observable, precision is lost almost everywhere else. In other words, rare events simulations are one-purpose simulations while a large ensemble and/or long simulation of a climate model can be used for multiple purposes. This needs to be taken into account when evaluating the opportunity to use a rare events algorithm. Moreover, one will always face the problem of evaluating whether the extreme summers sampled via the algorithm are physically realistic. The trajectories are physical in the sense that they could have been spontaneously simulated by the model, had I let it enough time to do so. However it is not clear whether the physics of the model realistically simulate what would happen in the climate system for such rare events. The fact that I did not simulate the oceans is a first limit insofar as with such strong anomalies of the atmospheric circulation the oceans will probably have responded differently than simulated in the control coupled simulation from which the boundary files are used. A second limit is the absence of dynamics in the vegetation: during extremely dry summers, it is likely that in the real climate system non-linear new phenomena would have appeared — forest fires for example. Finally, it is doubtful that the physics of the model — especially parametrization schemes — are still physically

relevant given the extremeness that I sampled. All these elements call again for cautiousness in the transposition of the trajectories sampled in the real world and my position is that rare events algorithms are first and foremost tools to explore limit properties *of climate models*.

In this direction, an interesting avenue of research for the application of rare events algorithms is the conditioning of extreme events to better understand the interaction of different components of the model. It is indeed common practice to do experiments in models such as changing one variable manually and measuring how the full system reacts. For example, one would impose low soil moisture (Fischer et al. 2007) and investigate the impacts on the heatwaves simulated by the model. However, doing so is artificial: there is no clear indication that it would have happened this way in the climate system. Another slow variable could cause both low soil moisture and high temperatures. With a rare events algorithm on the other hand, one can simulate physically coherent trajectories of low (or high) soil moisture as starting conditions and then run the model classically. The initial trajectories simulated are physically possible therefore if another variable is influencing both low soil moisture and high temperatures, it will be correctly sampled by the model. It would be interesting to compare systematically the results of this procedure with the procedure of imposing low soil moisture by hand. The rare events-based procedure could very much benefit the study of oceanic oscillations which arise as a result of interactions with the atmosphere: the oceans are often modelled as a red noise-like process driven by the white noise imposed by the atmosphere (Hasselmann 1976). Sampling extreme ENSO events with the GKTL algorithm is an interesting avenue of research for example.

Dynamics of heatwaves in the past, present and future. I showed that the use of rare events algorithms can be framed to answer physical questions about the dynamics of very intense events. My results obtained in the atmosphere-land surfaces configuration of the IPSL climate model clearly points in the direction of reduced atmospheric variability for the atmospheric circulation in the boreal summer but increase variability in surface energy fluxes. This paints a picture of more local, less large-scale organized heatwaves in the future compared to the present.

Nevertheless, from a physical point of view the result obtained in the present and the two SSPs used here are unsatisfactory. The main reason is that two many different forcings are applied at the same time on the system and therefore it is difficult to distinguish which forcings is changing which component of the system. Especially, it is not clear at this point whether the changes observed arise as a result of global warming of the Earth by the increase of GHG concentrations or reduced aerosol levels, globally and/or regionally. Both elements likely play a role but it is not possible to distinguish the two. It would be interesting — but probably also costly — to run the same simulations with GHG and aerosols forcings only to separate their respective dynamical role.

Similarly, as I explained in chapter 6, I ran simulations for a particular combination of the ENSO-AMO oceanic forcings. It would be valuable to know what is the impact of other phases of these oscillations on the dynamics of European heatwaves.

A large positive phase of the AMO oscillation could for example largely influence the results found here.

Bibliography

Abbot, D. S., R. J. Webber, S. Hadden, and J. Weare (2021). "Rare Event Sampling Improves Mercury Instability Statistics". In: *arXiv:2106.09091 [astro-ph]*. arXiv: 2106.09091 (→ page 101).

Adams, D. A., L. M. Sander, and R. M. Ziff (2008). "Harmonic measure for percolation and ising clusters including rare events". In: *Physical review letters* 101.14, p. 144102 (→ page 101).

Alexander, L. (2011). "Extreme heat rooted in dry soils". In: *Nature Geoscience* 4.1, pp. 12–13 (→ page 26).

Ali, S. M., O. Martius, and M. Röthlisberger (2021). "Recurrent Rossby wave packets modulate the persistence of dry and wet spells across the globe". In: *Geophysical Research Letters* 48.5, e2020GL091452 (→ page 25).

Alvarez-Castro, M. C., D. Faranda, and P. Yiou (2018). "Atmospheric dynamics leading to West European summer hot temperatures since 1851". In: *Complexity* 2018, pp. 1–10 (→ page 32).

Anosov, D. V. (1962). "Roughness of geodesic flows on compact Riemannian manifolds of negative curvature". In: *Soviet Mathematics-Doklady*, 3:1068–70 (→ page 5).

Araújo, V. and M. Viana (2009). "Hyperbolic Dynamical Systems". In: *Encyclopedia of Complexity and Systems Science*. Ed. by R. A. Meyers. New York, NY: Springer New York, pp. 4723–4737. DOI: 10.1007/978-0-387-30440-3_279 (→ page 6).

Arguez, A. and R. S. Vose (2011). "The definition of the standard WMO climate normal: The key to deriving alternative climate normals". In: *Bulletin of the American Meteorological Society* 92.6, pp. 699–704 (→ page 8).

Arnold, L., C. K. Jones, K. Mischaikow, G. Raugel, and L. Arnold (1995). *Random dynamical systems*. Springer (→ page 9).

Baars, S., D. Castellana, F. W. Wubs, and H. A. Dijkstra (2021). "Application of adaptive multilevel splitting to high-dimensional dynamical systems". en. In: *Journal of Computational Physics* 424, p. 109876. DOI: 10.1016/j.jcp.2020.109876 (→ pages 103, 105, 130).

Bador, M., L. Terray, and J. Boé (2016). "Emergence of human influence on summer record-breaking temperatures over Europe". In: *Geophysical Research Letters* 43.1, pp. 404–412 (→ page 30).

Baier, K., M. Rubel, and A. Stohl (2023). "The 3-Week-Long Transport History and Deep Tropical Origin of the 2021 Extreme Heat Wave in the Pacific Northwest". In: *Geophysical Research Letters* 50.24, e2023GL105865 (→ page 21).

Barnes, E. A., E. Dunn-Sigouin, G. Masato, and T. Woollings (2014). "Exploring recent trends in Northern Hemisphere blocking". In: *Geophysical Research Letters* 41.2, pp. 638–644 (→ page 33).

Barnes, E. A. and J. A. Screen (2015). "The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it?" In: *Wiley Interdisciplinary Reviews: Climate Change* 6.3, pp. 277–286 (→ pages 32, 33).

Baró, R., R. Lorente-Plazas, J. Montávez, and P. Jiménez-Guerrero (2017). "Biomass burning aerosol impact on surface winds during the 2010 Russian heat wave". In: *Geophysical Research Letters* 44.2, pp. 1088–1094 (→ page 29).

Barriopedro, D., R. García-Herrera, C. Ordóñez, D. Miralles, and S. Salcedo-Sanz (2023). "Heat waves: Physical understanding and scientific challenges". In: *Reviews of Geophysics* 61.2, e2022RG000780 (→ pages 15, 16, 21, 28–32, 35).

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests". In: *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* 160.901, pp. 268–282 (→ page 220).

Bauer, P., A. Thorpe, and G. Brunet (2015). "The quiet revolution of numerical weather prediction". en. In: *Nature* 525.7567. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7567 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Atmospheric dynamics;Climate sciences Subject_term_id: atmospheric-dynamics;climate-sciences, pp. 47–55. DOI: 10.1038/nature14956 (→ page 7).

Bevacqua, E., L. Suarez-Gutierrez, A. Jézéquel, F. Lehner, M. Vrac, et al. (2023). "Advancing research on compound weather and climate events via large ensemble model simulations". In: *Nature Communications* 14.1, p. 2145 (→ page 14).

Bieli, M., S. Pfahl, and H. Wernli (2015). "A Lagrangian investigation of hot and cold temperature extremes in Europe". In: *Quarterly Journal of the Royal Meteorological Society* 141.686, pp. 98–108 (→ page 18).

Birkhoff, G. D. (1931). "Proof of the ergodic theorem". In: *Proceedings of the National Academy of Sciences* 17.12, pp. 656–660 (→ page 5).

Birkhoff, G. D. (1927). *Dynamical systems*. Vol. 9. American Mathematical Soc. (→ page 5).

Black, E., M. Blackburn, G. Harrison, B. Hoskins, J. Methven, et al. (2004). "Factors contributing to the summer 2003 European heatwave". In: *Weather* 59.8, pp. 217–223 (→ page 15).

Blanchard-Wrigglesworth, E., T. Cox, Z. I. Espinosa, and A. Donohoe (2023). "The largest ever recorded heatwave—Characteristics and attribution of the Antarctic heatwave of March 2022". In: *Geophysical Research Letters* 50.17, e2023GL104910 (→ page 17).

Blanusa, M. L., C. J. López-Zurita, and S. Rasp (2023). "Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes". In: *Climate Dynamics* 61.3, pp. 1931–1945 (→ page 31).

Bódai, T., G. Károlyi, and T. Tél (2011). "A chaotically driven model climate: extreme events and snapshot attractors". English. In: *Nonlinear Processes in Geophysics* 18.5. Publisher: Copernicus GmbH, pp. 573–580. DOI: 10.5194/npg-18-573-2011 (→ page 9).

Bódai, T. and T. Tél (2012). "Annual variability in a conceptual climate model: Snapshot attractors, hysteresis in extreme events, and climate sensitivity". en. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22.2. Publisher: American Institute of PhysicsAIP, p. 023110. DOI: 10.1063/1.3697984 (→ page 9).

Boltzmann, L. (1909). *Wissenschaftliche Abhandlungen: Bd. 1865-1874*. Vol. 215. Chelsea Publishing Company, Incorporated (→ page 5).

Boucher, O., J. Servonnat, A. L. Albright, O. Aumont, Y. Balkanski, et al. (2020). "Presentation and evaluation of the IPSL-CM6A-LR climate model". In: *Journal of Advances in Modeling Earth Systems* 12.7, e2019MS002010 (→ pages 4, 188, 218, 244).

Bouchet, F., J. Rolland, and E. Simonnet (2019). "Rare Event Algorithm Links Transitions in Turbulent Flows with Activated Nucleations". In: *Physical Review Letters* 122.7. Publisher: American Physical Society, p. 074502. DOI: 10.1103/PhysRevLett.122.074502 (→ pages 101, 102, 105, 123, 130).

Bowen, R. and D. Ruelle (1975). "The Ergodic Theory of Axiom A Flows". In: *The Theory of Chaotic Attractors*, pp. 55–76 (→ page 6).

Branstator, G. (1983). "Horizontal energy propagation in a barotropic atmosphere with meridional and zonal structure". In: *Journal of Atmospheric Sciences* 40.7, pp. 1689–1708 (→ page 21).

Branstator, G. and H. Teng (2017). "Tropospheric waveguide teleconnections and their seasonality". In: *Journal of the Atmospheric Sciences* 74.5, pp. 1513–1532 (→ page 24).

Bréhier, C.-E. and T. Lelièvre (2019). "On a new class of score functions to estimate tail probabilities of some stochastic processes with adaptive multilevel splitting". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.3. Publisher: American Institute of Physics, p. 033126. DOI: 10.1063/1.5081440 (→ page 138).

Brunner, L., N. Schaller, J. Anstey, J. Sillmann, and A. K. Steiner (2018). "Dependence of present and future European temperature extremes on the location of atmospheric blocking". In: *Geophysical research letters* 45.12, pp. 6311–6320 (→ pages 21, 32).

Brunner, L. and A. Voigt (2024). "Pitfalls in diagnosing temperature extremes". In: *Nature Communications* 15.1, p. 2087 (→ page 16).

Bucklew, J. (2004). *Introduction to rare event simulation*. Springer Science & Business Media (→ page 120).

Carney, M., H. Kantz, and M. Nicol (2020). "Analysis and simulation of extremes and rare events in complex systems". In: *Advances in Dynamics, Optimization and Computation: A volume dedicated to Michael Dellnitz on the occasion of his 60th birthday*. Springer, pp. 151–182 (→ pages 104, 105).

Carvalho, A., J. A. Langa, and J. Robinson (2012). *Attractors for infinite-dimensional non-autonomous dynamical systems*. Vol. 182. Springer Science & Business Media (→ page 9).

Castañeda, V. and L. Wang (2024). "The role of climatological state in supporting US heat waves through Rossby waves packets". In: *Journal of Geophysical Research: Atmospheres* 129.4, e2023JD039212 (→ page 17).

Cattiaux, J. and A. Ribes (2018). "Defining single extreme weather events in a climate perspective". In: *Bulletin of the American Meteorological Society* 99.8, pp. 1557–1568 (→ page 13).

Cattiaux, J., P. Yiou, and R. Vautard (2012). "Dynamics of future seasonal temperature trends and extremes in Europe: a multi-model analysis from CMIP3". In: *Climate Dynamics* 38.9, pp. 1949–1964 (→ page 32).

Cérou, F. and A. Guyader (2007). "Adaptive Multilevel Splitting for Rare Event Analysis". In: *Stochastic Analysis and Applications* 25.2. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07362990601139628, pp. 417–443. DOI: 10.1080/07362990601139628 (→ page 130).

Cérou, F., A. Guyader, T. Lelievre, and D. Pommier (2011). "A multiple replica approach to simulate reactive trajectories". In: *The Journal of chemical physics* 134.5, p. 054108 (→ page 136).

Cérou, F., A. Guyader, and M. Rousset (2019). "Adaptive multilevel splitting: Historical perspective and recent results". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.4. Publisher: American Institute of Physics, p. 043108. DOI: 10.1063/1.5082247 (→ page 130).

Charney, J. G. and J. G. DeVore (1979). "Multiple Flow Equilibria in the Atmosphere and Blocking". EN. In: *Journal of the Atmospheric Sciences* 36.7. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1205–1216. DOI: 10.1175/1520-0469(1979)036<1205:MFEITA>2.0.CO;2 (→ pages 20, 103).

Chekroun, M. D., E. Simonnet, and M. Ghil (2011). "Stochastic climate dynamics: Random attractors and time-dependent invariant measures". en.

In: *Physica D: Nonlinear Phenomena* 240.21, pp. 1685–1700. DOI: 10.1016/j.physd.2011.06.005 (→ page 9).
Chen, B., C. Wu, X. Song, Y. Zheng, M. Lu, et al. (2023). "Anthropogenic heat release due to energy consumption exacerbates European summer extreme high temperature". In: *Climate Dynamics* 61.7, pp. 3831–3843 (→ page 26).
Chernykh, A. I. and M. G. Stepanov (2001). "Large negative velocity gradients in Burgers turbulence". In: *Physical Review E* 64.2. Publisher: American Physical Society, p. 026306. DOI: 10.1103/PhysRevE.64.026306 (→ page 119).
Christidis, N., G. S. Jones, and P. A. Stott (2015). "Dramatically increasing chance of extremely hot summers since the 2003 European heatwave". In: *Nature Climate Change* 5.1, pp. 46–50 (→ page 31).
Cini, M., G. Zappa, F. Ragone, and S. Corti (2024). "Simulating AMOC tipping driven by internal climate variability with a rare event algorithm". In: *npj Climate and Atmospheric Science* 7.1, p. 31 (→ pages 103, 105).
Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Vol. 208. Springer (→ page 14).
Coughlan, M. (1983). "A comparative climatology of blocking action in the two hemispheres". In: *Aust.-Meteor. Mag.* 31, pp. 3–13 (→ page 20).
Coumou, D., G. Di Capua, S. Vavrus, L. Wang, and S. Wang (2018). "The influence of Arctic amplification on mid-latitude summer circulation". In: *Nature Communications* 9.1, p. 2959 (→ pages 33, 34).
Coumou, D., J. Lehmann, and J. Beckmann (2015). "The weakening summer circulation in the Northern Hemisphere mid-latitudes". In: *Science* 348.6232, pp. 324–327 (→ pages 23, 31, 33).
Coumou, D., V. Petoukhov, S. Rahmstorf, S. Petri, and H. J. Schellnhuber (2014). "Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer". In: *Proceedings of the National Academy of Sciences* 111.34, pp. 12331–12336 (→ pages 23, 33).
Cowan, T., G. C. Hegerl, A. Schurer, S. F. Tett, R. Vautard, et al. (2020). "Ocean and land forcing of the record-breaking Dust Bowl heatwaves across central United States". In: *Nature communications* 11.1, p. 2870 (→ page 26).
Crauel, H. and F. Flandoli (1994). "Attractors for random dynamical systems". In: *Probability Theory and Related Fields* 100, pp. 365–393 (→ page 9).
D'Andrea, F., J.-P. Duvel, G. Rivière, R. Vautard, C. Cassou, et al. (2024). "Summer Deep Depressions Increase Over the Eastern North Atlantic". In: *Geophysical Research Letters* 51.5, e2023GL104435 (→ page 32).
Davini, P. and F. d'Andrea (2020). "From CMIP3 to CMIP6: Northern Hemisphere atmospheric blocking simulation in present and future climate". In: *Journal of Climate* 33.23, pp. 10021–10038 (→ page 32).
De Kauwe, M. G., S.-X. Zhou, B. E. Medlyn, A. J. Pitman, Y.-P. Wang, et al. (2015). "Do land surface models need to include differential plant species responses to drought? Examining model predictions across a mesic-xeric gradient in Europe". In: *Biogeosciences* 12.24, pp. 7503–7518 (→ page 28).
Del Moral, P. (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Vol. 88. Springer (→ pages 124, 141).
Della-Marta, P. M., J. Luterbacher, H. von Weissenfluh, E. Xoplaki, M. Brunet, et al. (2007). "Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability". In: *Climate Dynamics* 29, pp. 251–275 (→ page 17).
Dematteis, G., T. Grafke, and E. Vanden-Eijnden (2019). "Extreme event quantification in dynamical systems with random components". In: *SIAM/ASA Journal on Uncertainty Quantification* 7.3, pp. 1029–1059 (→ page 113).
Deng, K.-Q., C. Azorin-Molina, S. Yang, C.-D. Hu, G.-F. Zhang, et al. (2022). "Shifting of summertime weather extremes in Western Europe during 2012–2020". In: *Advances in Climate Change Research* 13.2, pp. 218–227 (→ page 29).
Dey, S., R. K. Choudhary, A. Upadhyay, and S. Dash (2021). "Aerosol-modulated heat stress in the present and future climate of India". In: *Environmental Research Letters* 16.12, p. 124022 (→ page 29).
Di Capua, G., S. Sparrow, K. Kornhuber, E. Rousi, S. Osprey, et al. (2021). "Drivers behind the summer 2010 wave train leading to Russian heatwave and Pakistan flooding". In: *npj Climate and Atmospheric Science* 4.1, p. 55 (→ page 15).
Dickinson, R. E. (1995). "Land-atmosphere interaction". In: *Reviews of geophysics* 33.S2, pp. 917–922 (→ page 25).
Diffenbaugh, N. S. (2020). "Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events". In: *Science Advances* 6.12, eaay2368 (→ pages 16, 253).
Diffenbaugh, N. S., J. S. Pal, F. Giorgi, and X. Gao (2007). "Heat stress intensification in the Mediterranean climate change hotspot". In: *Geophysical Research Letters* 34.11 (→ page 35).
Diffenbaugh, N. S., D. Singh, J. S. Mankin, D. E. Horton, D. L. Swain, et al. (2017). "Quantifying the influence of global warming on unprecedented extreme climate events". In: *Proceedings of the National Academy of Sciences* 114.19, pp. 4881–4886 (→ page 31).
Dirmeyer, P. A., B. A. Cash, J. L. Kinter III, C. Stan, T. Jung, et al. (2012). "Evidence for enhanced land–atmosphere feedback in a warming climate". In: *Journal of Hydrometeorology* 13.3, pp. 981–995 (→ page 35).
Dirmeyer, P. A., L. Chen, J. Wu, C.-S. Shin, B. Huang, et al. (2018). "Verification of land–atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations". In: *Journal of Hydrometeorology* 19.2, pp. 375–392 (→ page 28).
Dole, R., M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, et al. (2011). "Was there a basis for anticipating the 2010 Russian heat wave?" In: *Geophysical research letters* 38.6 (→ page 15).
Domeisen, D. I., E. A. Eltahir, E. M. Fischer, R. Knutti, S. E. Perkins-Kirkpatrick, et al. (2023). "Prediction and projection of heatwaves". In: *Nature Reviews Earth & Environment* 4.1, pp. 36–50 (→ pages 15, 27, 29, 30, 35).
Dong, B., R. T. Sutton, and L. Shaffrey (2017). "Understanding the rapid summer warming and changes in temperature extremes since the mid-1990s over Western Europe". In: *Climate Dynamics* 48, pp. 1537–1554 (→ page 29).
Dorrington, J., K. Strommen, F. Fabiano, and F. Molteni (2022). "CMIP6 models trend toward less persistent European blocking regimes in a warming climate". In: *Geophysical Research Letters* 49.24, e2022GL100811 (→ page 34).
Douville, H., J. Colin, E. Krug, J. Cattiaux, and S. Thao (2016). "Midlatitude daily summer temperatures reshaped by soil moisture under climate change". In: *Geophysical Research Letters* 43.2, pp. 812–818 (→ page 30).
Douville, H. (2002). "Influence of soil moisture on the Asian and African monsoons. Part II: Interannual variability". In: *Journal of Climate* 15.7, pp. 701–720 (→ page 27).
Douville, H. and F. Chauvin (2000). "Relevance of soil moisture for seasonal climate predictions: A preliminary study". In: *Climate Dynamics* 16, pp. 719–736 (→ page 27).
Drótos, G., T. Bódai, and T. Tél (2015). "Probabilistic Concepts in a Changing Climate: A Snapshot Attractor Picture". EN. In: *Journal of Climate* 28.8. Publisher: American Meteorological Society Section: Journal of Climate, pp. 3275–3288. DOI: 10.1175/JCLI-D-14-00459.1 (→ page 9).
Duchez, A., E. Frajka-Williams, S. A. Josey, D. G. Evans, J. P. Grist, et al. (2016). "Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave". In: *Environmental Research Letters* 11.7, p. 074004 (→ page 29).
Dunn, R. J., L. V. Alexander, M. G. Donat, X. Zhang, M. Bador, et al. (2020). "Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3". In: *Journal of Geophysical Research: Atmospheres* 125.16, e2019JD032263 (→ page 31).
E, W., W. Ren, and E. Vanden-Eijnden (2002). "String method for the study of rare events". In: *Physical Review B* 66.5. Publisher: American Physical Society, p. 052301. DOI: 10.1103/PhysRevB.66.052301 (→ page 101).
— (2004). "Minimum action method for the study of rare events". In: *Communications on Pure and Applied Mathematics* 57.5. Publisher: John Wiley & Sons, Ltd, pp. 637–656. DOI: 10.1002/cpa.20005 (→ pages 119, 122).
Ebener, L., G. Margazoglou, J. Friedrich, L. Biferale, and R. Grauer (2019). "Instanton based importance sampling for rare events in stochastic PDEs". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.6, p. 063102 (→ page 120).
Eckmann, J.-P. and D. Ruelle (1985). "Ergodic theory of chaos and strange attractors". In: *Reviews of modern physics* 57.3, p. 617 (→ page 6).
Ek, M. and A. Holtslag (2004). "Influence of soil moisture on boundary layer cloud development". In: *Journal of hydrometeorology* 5.1, pp. 86–99 (→ page 27).
Fan, X., C. Miao, Q. Duan, C. Shen, and Y. Wu (2020). "The performance of CMIP6 versus CMIP5 in simulating temperature extremes over the global land surface". In: *Journal of Geophysical Research: Atmospheres* 125.18, e2020JD033031 (→ page 31).
Faranda, D., S. Bourdin, M. Ginesta, M. Krouma, R. Noyelle, et al. (2022). "A climate-change attribution retrospective of some impactful weather extremes of 2021". In: *Weather and Climate Dynamics* 3.4, pp. 1311–1340 (→ page 249).
Faranda, D., G. Messori, T. Alberti, C. Alvarez-Castro, T. Caby, et al. (2023a). "A statistical physics and dynamical systems perspective on geophysical extreme events". In: *arXiv preprint arXiv:2309.15393* (→ page 249).
Faranda, D., G. Messori, M. C. Alvarez-Castro, and P. Yiou (2017a). *Dynamical properties and extremes of Northern Hemisphere climate fields over the past 60 years*. en. preprint. Bifurcation, Chaos, Phase Transitions, Complex Systems/Climate, Atmosphere, Ocean, Hydrology, Cryosphere, Biosphere. DOI: 10.5194/npg-2017-36 (→ page 138).
Faranda, D., G. Messori, A. Jezequel, M. Vrac, and P. Yiou (2023b). "Atmospheric circulation compounds anthropogenic warming and impacts of climate extremes in Europe". In: *Proceedings of the National Academy of Sciences* 120.13, e2214525120 (→ pages 15, 32, 250).
Faranda, D., G. Messori, and P. Yiou (2017b). "Dynamical proxies of North Atlantic predictability and extremes". In: *Scientific reports* 7.1, pp. 1–10 (→ page 248).
Faranda, D., Y. Sato, G. Messori, N. R. Moloney, and P. Yiou (2019). "Minimal dynamical systems model of the Northern Hemisphere jet stream via embedding of climate data". en. In: *Earth System Dynamics* 10.3, pp. 555–567. DOI: 10.5194/esd-10-555-2019 (→ page 248).
Favre, A., B. Hewitson, M. Tadross, C. Lennard, and R. Cerezo-Mota (2012). "Relationships between cut-off lows and the semiannual and southern oscillations". In: *Climate dynamics* 38, pp. 1473–1487 (→ page 191).
Finkel, J., D. S. Abbot, and J. Weare (2020). "Path Properties of Atmospheric Transitions: Illustration with a Low-Order Sudden Stratospheric Warming Model". EN. In: *Journal of the Atmospheric Sciences* 77.7. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2327–2347. DOI: 10.1175/JAS-D-19-0278.1 (→ pages 103, 105, 118).
Finkel, J. and P. A. O'Gorman (2024). "Bringing statistics to storylines: rare event sampling for sudden, transient extreme events". In: *arXiv preprint arXiv:2402.01823* (→ pages 124, 138).
Finkel, J., R. J. Webber, E. P. Gerber, D. S. Abbot, and J. Weare (2021a). "Exploring stratospheric rare events with transition path theory and short simulations". In: *arXiv:2108.12727 [physics]*. arXiv: 2108.12727 (→ pages 103, 105, 118).
— (2021b). "Learning Forecasts of Rare Stratospheric Transitions from Short Simulations". EN. In: *Monthly Weather Review* 149.11. Publisher: American Meteorological Society Section: Monthly Weather Review, pp. 3647–3669. DOI: 10.1175/MWR-D-21-0024.1 (→ pages 103, 118).
Fischer, E., S. Sippel, and R. Knutti (2021). "Increasing probability of record-shattering climate extremes". In: *Nature Climate Change* 11.8, pp. 689–695 (→ page 30).

Fischer, E. M., U. Beyerle, L. Bloin-Wibe, C. Gessner, V. Humphrey, et al. (2023). "Storylines for unprecedented heatwaves based on ensemble boosting". In: *Nature Communications* 14.1, p. 4643 (→ pages 105, 124, 254).

Fischer, E. M. and R. Knutti (2015). "Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes". In: *Nature climate change* 5.6, pp. 560–564 (→ page 31).

Fischer, E. M., K. W. Oleson, and D. M. Lawrence (2012). "Contrasting urban and rural heat stress responses to climate change". In: *Geophysical research letters* 39.3 (→ page 35).

Fischer, E. M., J. Sedláček, E. Hawkins, and R. Knutti (2014). "Models agree on forced response pattern of precipitation and temperature extremes". In: *Geophysical Research Letters* 41.23, pp. 8554–8562 (→ page 31).

Fischer, E. M., S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär (2007). "Soil moisture–atmosphere interactions during the 2003 European summer heat wave". In: *Journal of Climate* 20.20, pp. 5081–5099 (→ pages 17, 27, 255).

Fragkoulidis, G., V. Wirth, P. Bossmann, and A. Fink (2018). "Linking Northern Hemisphere temperature extremes to Rossby wave packets". In: *Quarterly Journal of the Royal Meteorological Society* 144.711, pp. 553–566 (→ pages 23–25).

Francis, J. A. (2017). "Why are Arctic linkages to extreme weather still up in the air?" In: *Bulletin of the American Meteorological Society* 98.12, pp. 2551–2557 (→ page 32).

Freidlin, M. and A. Wentzell (1984). "Random Perturbations". In: *Random Perturbations of Dynamical Systems*. Springer, pp. 15–43 (→ page 110).

Gálfi, V. M., V. Lucarini, F. Ragone, and J. Wouters (2021). "Applications of large deviation theory in geophysical fluid dynamics and climate science". en. In: *La Rivista del Nuovo Cimento* 44.6, pp. 291–363. DOI: 10.1007/s40766-021-00020-z (→ pages 14, 197).

Gallavotti, G. and E. G. D. Cohen (1995). "Dynamical ensembles in stationary states". en. In: *Journal of Statistical Physics* 80.5, pp. 931–970. DOI: 10.1007/BF02179860 (→ page 7).

Gallavotti, G. (2014). *Nonequilibrium and irreversibility*. Springer (→ page 7).

Gallavotti, G. and D. Ruelle (1997). "SRB states and nonequilibrium statistical mechanics close to equilibrium". In: *Communications in Mathematical Physics* 190, pp. 279–285 (→ page 6).

Gao, Y., Y. Wu, X. Guo, W. Kou, S. Zhang, et al. (2023). "More Frequent and Persistent Heatwaves Due To Increased Temperature Skewness Projected by a High-Resolution Earth System Model". In: *Geophysical Research Letters* 50.18, e2023GL105840 (→ page 35).

García-Herrera, R., J. Díaz, R. M. Trigo, J. Luterbacher, and E. M. Fischer (2010). "A review of the European summer heat wave of 2003". In: *Critical Reviews in Environmental Science and Technology* 40.4, pp. 267–306 (→ page 15).

Gastineau, G. and C. Frankignoul (2015). "Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century". In: *Journal of Climate* 28.4, pp. 1396–1416 (→ page 28).

Gessner, C., E. M. Fischer, U. Beyerle, and R. Knutti (2021). "Very rare heat extremes: quantifying and understanding using ensemble reinitialization". In: *Journal of Climate* 34.16, pp. 6619–6634 (→ pages 104, 105, 124, 180, 269).

— (2022). "Multi-year drought storylines for Europe and North America from an iteratively perturbed global climate model". In: *Weather and Climate Extremes* 38, p. 100512 (→ pages 104, 105).

Gevaert, A., D. G. Miralles, R. A. de Jeu, J. Schellekens, and A. J. Dolman (2018). "Soil moisture-temperature coupling in a set of land surface models". In: *Journal of Geophysical Research: Atmospheres* 123.3, pp. 1481–1498 (→ page 26).

Ghil, M. (2001). "Hilbert problems for the geosciences in the 21st century". English. In: *Nonlinear Processes in Geophysics* 8.4/5. Publisher: Copernicus GmbH, pp. 211–211. DOI: 10.5194/npg-8-211-2001 (→ page 4).

Ghil, M., P. Yiou, S. Hallegatte, B. D. Malamud, P. Naveau, et al. (2011). "Extreme events: dynamics, statistics and prediction". English. In: *Nonlinear Processes in Geophysics* 18.3. Publisher: Copernicus GmbH, pp. 295–350. DOI: 10.5194/npg-18-295-2011 (→ page 10).

Ghil, M., M. D. Chekroun, and E. Simonnet (2008). "Climate dynamics and fluid mechanics: Natural variability and related uncertainties". en. In: *Physica D: Nonlinear Phenomena*. Euler Equations: 250 Years On 237.14, pp. 2111–2126. DOI: 10.1016/j.physd.2008.03.036 (→ page 9).

Ghil, M. and V. Lucarini (2020). "The physics of climate variability and climate change". In: *Reviews of Modern Physics* 92.3. Publisher: American Physical Society, p. 035002. DOI: 10.1103/RevModPhys.92.035002 (→ pages 3, 4).

Giardina, C., J. Kurchan, V. Lecomte, and J. Tailleur (2011). "Simulating Rare Events in Dynamical Processes". en. In: *Journal of Statistical Physics* 145.4, pp. 787–811. DOI: 10.1007/s10955-011-0350-4 (→ pages 128, 138).

Giardina', C., J. Kurchan, and L. Peliti (2006). "Direct evaluation of large-deviation functions". en. In: *arXiv:cond-mat/0511248*. arXiv: cond-mat/0511248. DOI: 10.1103/PhysRevLett.96.120603 (→ page 128).

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner's sons (→ page 5).

Gloege, L., K. Kornhuber, O. Skulovich, I. Pal, S. Zhou, et al. (2022). "Land-atmosphere cascade fueled the 2020 Siberian heatwave". In: *AGU Advances* 3.6, e2021AV000619 (→ page 17).

Grafke, T., R. Grauer, and T. Schäfer (2013). "Instanton filtering for the stochastic Burgers equation". en. In: *Journal of Physics A: Mathematical and Theoretical* 46.6. arXiv: 1209.0905, p. 062002. DOI: 10.1088/1751-8113/46/6/062002 (→ page 101).

— (2015). "The instanton method and its numerical implementation in fluid mechanics". en. In: *Journal of Physics A: Mathematical and Theoretical* 48.33. Publisher: IOP Publishing, p. 333001. DOI: 10.1088/1751-8113/48/33/333001 (→ pages 110, 116, 120).

Grafke, T., R. Grauer, T. Schäfer, and E. Vanden-Eijnden (2014). "Arclength parametrized Hamilton's equations for the calculation of instantons". en. In: *Multiscale Modeling & Simulation* 12.2. arXiv: 1309.5175, pp. 566–580. DOI: 10.1137/130939158 (→ page 120).

Grafke, T. and E. Vanden-Eijnden (2019). "Numerical computation of rare events via large deviation theory". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.6. Publisher: American Institute of Physics, p. 063118. DOI: 10.1063/1.5084025 (→ pages 101, 103, 105, 106, 111, 112, 119, 120).

Grassberger, P. (2002). "Go with the winners: a general Monte Carlo strategy". en. In: *Computer Physics Communications*. Proceedings of the Europhysics Conference on Computational Physics Computational Modeling and Simulation of Complex Systems 147.1, pp. 64–70. DOI: 10.1016/S0010-4655(02)00205-9 (→ page 123).

Grotjahn, R., R. Black, R. Leung, M. F. Wehner, M. Barlow, et al. (2016). "North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends". In: *Climate Dynamics* 46, pp. 1151–1184 (→ page 32).

Haarsma, R. J., F. Selten, B. v. Hurk, W. Hazeleger, and X. Wang (2009). "Drier Mediterranean soils due to greenhouse warming bring easterly winds over summertime central Europe". In: *Geophysical research letters* 36.4 (→ page 27).

Hartmann, C., O. Kebiri, L. Neureither, and L. Richter (2019). "Variational approach to rare event simulation using least-squares regression". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.6, p. 063107 (→ page 121).

Hasselmann, K. (1976). "Stochastic climate models Part I. Theory". In: *Tellus* 28.6. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/tellusa.v28i6.11316, pp. 473–485. DOI: 10.3402/tellusa.v28i6.11316 (→ pages 113, 255).

Hauser, M., R. Orth, and S. I. Seneviratne (2016). "Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia". In: *Geophysical Research Letters* 43.6, pp. 2819–2826 (→ page 17).

He, B., Z. Zhong, D. Chen, J. Liu, Y. Chen, et al. (2022). "Lengthening dry spells intensify summer heatwaves". In: *Geophysical Research Letters* 49.19, e2022GL099647 (→ page 26).

He, Y., X. Zhu, Z. Sheng, and M. He (2023). "Resonant Waves Play an Important Role in the Increasing Heat Waves in Northern Hemisphere Mid-Latitudes Under Global Warming". In: *Geophysical Research Letters* 50.14, e2023GL104839 (→ page 23).

Held, I. M. (1983). "Stationary and quasi-stationary eddies in the extratropical troposphere: Theory". In: *Large-scale dynamical processes in the atmosphere* 127, p. 168 (→ page 21).

Held, I. M., M. Ting, and H. Wang (2002). "Northern winter stationary waves: Theory and modeling". In: *Journal of climate* 15.16, pp. 2125–2144 (→ page 21).

Helfmann, L., E. Ribera Borrell, C. Schütte, and P. Koltai (2020). "Extending Transition Path Theory: Periodically Driven and Finite-Time Dynamics". en. In: *Journal of Nonlinear Science* 30.6, pp. 3321–3366. DOI: 10.1007/s00332-020-09652-7 (→ page 109).

Hermann, M., L. Papritz, and H. Wernli (2020). "A Lagrangian analysis of the dynamical and thermodynamic drivers of large-scale Greenland melt events during 1979–2017". In: *Weather and Climate Dynamics* 1.2, pp. 497–518 (→ page 17).

Heymann, M. and E. Vanden-Eijnden (2008). "The geometric minimum action method: A least action principle on the space of curves". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 61.8, pp. 1052–1117 (→ page 119).

Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, et al. (2011). "Observational evidence for soil-moisture impact on hot extremes in southeastern Europe". In: *Nature Geoscience* 4.1, pp. 17–21 (→ page 28).

Hoffman, R. N., J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehrkorn (2006a). "The Response of Damaging Winds of a Simulated Tropical Cyclone to Finite-Amplitude Perturbations of Different Variables". EN. In: *Journal of the Atmospheric Sciences* 63.7. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1924–1937. DOI: 10.1175/JAS3720.1 (→ page 104).

— (2006b). "Using 4d-VAR to Move a Simulated Tropical Cyclone in a Mesoscale Model". en. In: *Computers & Mathematics with Applications*. Variational Data Assimilation and Optimal Control 52.8, pp. 1193–1204. DOI: 10.1016/j.camwa.2006.11.013 (→ page 104).

Hoffmann, P. (2018). "Enhanced seasonal predictability of the summer mean temperature in Central Europe favored by new dominant weather patterns". In: *Climate dynamics* 50, pp. 2799–2812 (→ page 34).

Horowitz, R. L., K. A. McKinnon, and I. R. Simpson (2022). "Circulation and soil moisture contributions to heatwaves in the United States". In: *Journal of Climate* 35.24, pp. 8031–8048 (→ page 17).

Horton, D. E., N. C. Johnson, D. Singh, D. L. Swain, B. Rajaratnam, et al. (2015). "Contribution of changes in atmospheric circulation patterns to extreme temperature trends". In: *Nature* 522.7557, pp. 465–469 (→ pages 31, 32).

Horton, R. M., J. S. Mankin, C. Lesk, E. Coffel, and C. Raymond (2016). "A review of recent advances in research on extreme heat events". In: *Current Climate Change Reports* 2, pp. 242–259 (→ pages 15, 22, 33).

Hoskins, B. (1983). "Dynamical processes in the atmosphere and the use of models". In: *Quart. J. Roy. Meteor. Soc* 109.459, pp. 1–21 (→ page 15).

Hoskins, B. and T. Woollings (2015). "Persistent extratropical regimes and climate extremes". In: *Current Climate Change Reports* 1, pp. 115–124 (→ pages 19, 22, 32).

Hoskins, B. J. and T. Ambrizzi (1993). "Rossby wave propagation on a realistic longitudinally varying flow". In: *Journal of Atmospheric Sciences* 50.12, pp. 1661–1671 (→ page 21).

Hoskins, B. J., I. N. James, and G. H. White (1983). "The shape, propagation and mean-flow interaction of large-scale weather systems". In: *Journal of Atmospheric Sciences* 40.7, pp. 1595–1612 (→ page 189).

Hoskins, B. J. and D. J. Karoly (1981). "The steady linear response of a spherical atmosphere to thermal and orographic forcing". In: *Journal of the atmospheric sciences* 38.6, pp. 1179–1196 (→ page 21).

Hotz, B., L. Papritz, and M. Röthlisberger (2024). "Understanding the vertical temperature structure of recent record-shattering heatwaves". In: *Weather and Climate Dynamics* 5.1, pp. 323–343 (→ page 18).

Hoy, A., S. Haensel, P. Skalak, Z. Ustrnul, and O. Bochníček (2017). "The extreme European summer of 2015 in a long-term perspective". In: *International Journal of Climatology* 37.2, pp. 943–962 (→ page 17).

Huguenin, M. F., E. M. Fischer, S. Kotlarski, S. C. Scherrer, C. Schwierz, et al. (2020). "Lack of change in the projected frequency and persistence of atmospheric circulation types over Central Europe". In: *Geophysical Research Letters* 47.9, e2019GL086132 (→ pages 32, 34).

Huntingford, C., P. M. Cox, P. D. Ritchie, J. J. Clarke, I. M. Parry, et al. (2024). "Acceleration of daily land temperature extremes and correlations with surface energy fluxes". In: *npj Climate and Atmospheric Science* 7.1, p. 84 (→ pages 15, 30, 35).

Huntingford, C., D. Mitchell, K. Kornhuber, D. Coumou, S. Osprey, et al. (2019). "Assessing changes in risk of amplified planetary waves in a warming world". In: *Atmospheric Science Letters* 20.8, e929 (→ page 33).

IPCC (2023). "Annex VII: Glossary". In: *Glossary - Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, pp. 2215–2256 (→ pages 3, 7, 10, 15).

Jézéquel, A., J. Cattiaux, P. Naveau, S. Radanovics, A. Ribes, et al. (2018). "Trends of atmospheric circulation during singular hot days in Europe". en. In: *Environmental Research Letters* 13.5. Publisher: IOP Publishing, p. 054007. DOI: 10.1088/1748-9326/aab5da (→ page 32).

Jiménez-Esteve, B., K. Kornhuber, and D. Domeisen (2022). "Heat Extremes Driven by Amplification of Phase-Locked Circumglobal Waves Forced by Topography in an Idealized Atmospheric Model". In: *Geophysical Research Letters* 49.21, e2021GL096337 (→ pages 17, 25, 189, 190).

Kahn, H. and T. E. Harris (1951). "Estimation of particle transmission by random sampling". In: *National Bureau of Standards applied mathematics series* 12, pp. 27–30 (→ page 130).

Kala, J., M. G. De Kauwe, A. J. Pitman, B. E. Medlyn, Y.-P. Wang, et al. (2016). "Impact of the representation of stomatal conductance on model projections of heatwave intensity". In: *Scientific reports* 6.1, p. 23418 (→ page 28).

Katok, A. and B. Hasselblatt (1997). *Introduction to the modern theory of dynamical systems*. 54. Cambridge university press (→ page 139).

Kautz, L.-A., O. Martius, S. Pfahl, J. G. Pinto, A. M. Ramos, et al. (2022). "Atmospheric blocking and weather extremes over the Euro-Atlantic sector–a review". In: *Weather and climate dynamics* 3.1, pp. 305–336 (→ pages 19, 20, 25, 32).

Keller, H. B. (2018). *Numerical methods for two-point boundary-value problems*. Courier Dover Publications (→ page 119).

Kendall, M. G. et al. (1948). "The advanced theory of statistics. Vols. 1." In: *The advanced theory of statistics. Vols. 1.* 1.Ed. 4 (→ pages 265, 271).

Kim, Y.-H., S.-K. Min, X. Zhang, J. Sillmann, and M. Sandstad (2020). "Evaluation of the CMIP6 multi-model ensemble for climate extreme indices". In: *Weather and Climate Extremes* 29, p. 100269 (→ page 31).

Kim, Y.-H., S.-K. Min, X. Zhang, F. Zwiers, L. V. Alexander, et al. (2016). "Attribution of extreme temperature changes during 1951–2010". In: *Climate dynamics* 46, pp. 1769–1782 (→ page 31).

Kleiner, N., P. W. Chan, L. Wang, D. Ma, and Z. Kuang (2021). "Effects of climate model mean-state bias on blocking underestimation". In: *Geophysical Research Letters* 48.13, e2021GL094129 (→ page 21).

Kloeden, P. E. and M. Rasmussen (2011). *Nonautonomous dynamical systems*. 176. American Mathematical Soc. (→ page 9).

Kodra, E. and A. R. Ganguly (2014). "Asymmetry of projected increases in extreme temperature distributions". In: *Scientific reports* 4.1, p. 5884 (→ page 30).

Kolmogorov, A. N. (1954). "On conservation of conditionally periodic motions for a small change in Hamilton's function". In: *Dokl. akad. nauk Sssr*. Vol. 98, pp. 527–530 (→ page 5).

Kornhuber, K., D. Coumou, E. Vogel, C. Lesk, J. F. Donges, et al. (2020). "Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions". In: *Nature Climate Change* 10.1, pp. 48–53 (→ page 23).

Kornhuber, K., S. Osprey, D. Coumou, S. Petri, V. Petoukhov, et al. (2019). "Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern". In: *Environmental Research Letters* 14.5, p. 054002 (→ pages 23, 33).

Kornhuber, K. and T. Tamarin-Brodsky (2021). "Future changes in Northern Hemisphere summer weather persistence linked to projected Arctic warming". In: *Geophysical Research Letters* 48.4, e2020GL091603 (→ pages 33, 34).

Koster, R. D., Y. Chang, H. Wang, and S. D. Schubert (2016). "Impacts of local soil moisture anomalies on the atmospheric circulation and on remote surface meteorological fields during boreal summer: A comprehensive analysis over North America". In: *Journal of Climate* 29.20, pp. 7345–7364 (→ pages 27, 28).

Kovats, R. S. and S. Hajat (2008). "Heat stress and public health: a critical review". In: *Annu. Rev. Public Health* 29, pp. 41–55 (→ page 16).

Kuehn, C. et al. (2015). *Multiple time scale dynamics*. Vol. 191. Springer (→ page 9).

Laffargue, T. and J. Tailleur (2014). "Locating a regular needle in a chaotic haystack, and conversely, using Lyapunov Weighted Dynamics". In: *arXiv:1404.2600 [cond-mat, physics:nlin]*. arXiv: 1404.2600 (→ pages 128, 138).

Landau, L. D. and E. Lifshitz (1969). "Statistical physics. Pt. 1". In: *Course of theoretical physics-Pergamon International Library of Science* (→ page 5).

Laurie, J. and F. Bouchet (2015). "Computation of rare transitions in the barotropic quasi-geostrophic equations". en. In: 17.1. Publisher: IOP Publishing, p. 015009. DOI: 10.1088/1367-2630/17/1/015009 (→ pages 102, 105, 120).

Lee, Y.-Y. and R. Grotjahn (2016). "California Central Valley summer heat waves form two ways". In: *Journal of Climate* 29.3, pp. 1201–1217 (→ page 18).

Lehmann, J. and D. Coumou (2015). "The influence of mid-latitude storm tracks on hot, cold, dry and wet extremes". In: *Scientific reports* 5.1, p. 17491 (→ page 23).

Lehmann, J., D. Coumou, K. Frieler, A. V. Eliseev, and A. Levermann (2014). "Future changes in extratropical storm tracks and baroclinicity under climate change". In: *Environmental Research Letters* 9.8, p. 084002 (→ page 34).

Leith, C. E. (1975). "Climate Response and Fluctuation Dissipation". EN. In: *Journal of the Atmospheric Sciences* 32.10. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 2022–2026. DOI: 10.1175/1520-0469(1975)032<2022:CRAFD>2.0.CO;2 (→ page 9).

Lejeune, Q., E. L. Davin, L. Gudmundsson, J. Winckler, and S. I. Seneviratne (2018). "Historical deforestation locally increased the intensity of hot days in northern mid-latitudes". In: *Nature Climate Change* 8.5, pp. 386–390 (→ page 31).

Lemordant, L., P. Gentine, M. Stéfanon, P. Drobinski, and S. Fatichi (2016). "Modification of land-atmosphere interactions by CO2 effects: Implications for summer dryness and heat wave amplitude". In: *Geophysical Research Letters* 43.19, pp. 10–240 (→ pages 27, 35).

Lestang, T., F. Ragone, C.-E. Bréhier, C. Herbert, and F. Bouchet (2018). "Computing return times or return periods with rare event algorithms". en. In: 2018.4. Publisher: IOP Publishing, p. 043213. DOI: 10.1088/1742-5468/aab856 (→ pages 132, 138).

Levene, H. (1960). "Robust tests for equality of variances". In: *Contributions to probability and statistics*, pp. 278–292 (→ pages 220, 226).

Livi, R., S. Ruffo, and D. Shepelyansky (2003). "Kolmogorov pathways from integrability to chaos and beyond". In: *The Kolmogorov Legacy in Physics*, pp. 3–32 (→ page 5).

Lorenz, E. (1967). "The nature and theory of the general circulation of the atmosphere". In: *World meteorological organization* 161 (→ page 4).

Lorenz, E. N. (1955). "Available potential energy and the maintenance of the general circulation". In: *Tellus* 7.2, pp. 157–167 (→ page 4).

— (1963). "Deterministic Nonperiodic Flow". EN. In: *Journal of the Atmospheric Sciences* 20.2. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 130–141. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2 (→ pages 5, 6).

— (1996). *Predictability: A Problem Partly Solved* (→ pages 103, 104).

Lorenz, R., E. B. Jaeger, and S. I. Seneviratne (2010). "Persistence of heat waves and its link to soil moisture memory". In: *Geophysical Research Letters* 37.9 (→ page 25).

Lu, Y.-C. and D. M. Romps (2023). "Is a wet-bulb temperature of 35° C the correct threshold for human survivability?" In: *Environmental Research Letters* 18.9, p. 094021 (→ page 11).

Lucarini, V., R. Blender, C. Herbert, F. Ragone, S. Pascale, et al. (2014). "Mathematical and physical ideas for climate science". en. In: *Reviews of Geophysics* 52.4, pp. 809–859. DOI: 10.1002/2013RG000446 (→ pages 4, 7).

Lucarini, V., V. Melinda Galfi, J. Riboldi, and G. Messori (2023). "Typicality of the 2021 Western North America summer heatwave". In: *Environmental Research Letters* 18.1, p. 015004 (→ pages 12, 112, 150, 254).

Lucente, D., J. Rolland, C. Herbert, and F. Bouchet (2021). "Coupling rare event algorithms with data-based learned committor functions using the analogue Markov chain". en. In: *arXiv:2110.05050 [cond-mat, physics:physics]*. arXiv: 2110.05050 (→ pages 138, 139).

Luo, B., D. Luo, W. Zhuo, C. Xiao, A. Dai, et al. (2023). "Increased Summer European Heatwaves in Recent Decades: Contributions From Greenhouse Gases-Induced Warming and Atlantic Multidecadal Oscillation-Like Variations". In: *Earth's Future* 11.8, e2023EF003701 (→ pages 29, 31).

Luo, F., F. Selten, K. Wehrli, K. Kornhuber, P. Le Sager, et al. (2022). "Summertime Rossby waves in climate models: substantial biases in surface imprint associated with small biases in upper-level circulation". In: *Weather and Climate Dynamics* 3.3, pp. 905–935 (→ pages 28, 29).

Luo, M., S. Wu, G. N.-C. Lau, T. Pei, Z. Liu, et al. (2024). "Anthropogenic forcing has increased the risk of longer-traveling and slower-moving large contiguous heatwaves". In: *Science Advances* 10.13, eadl1598 (→ page 33).

Lyapunov, A. (1892). "The general problem of motion stability". In: *Annals of Mathematics Studies* 17 (→ page 5).

Maher, N., S. Milinski, L. Suarez-Gutierrez, M. Botzet, M. Dobrynin, et al. (2019). "The Max Planck Institute Grand Ensemble: enabling the exploration of climate system variability". In: *Journal of Advances in Modeling Earth Systems* 11.7, pp. 2050–2069 (→ page 10).

Mann, M. E., S. Rahmstorf, K. Kornhuber, B. A. Steinman, S. K. Miller, et al. (2017). "Influence of anthropogenic climate change on planetary wave resonance and extreme weather events". In: *Scientific reports* 7.1, pp. 1–12 (→ pages 23, 33).

Marshall, A., D. Hudson, M. Wheeler, O. Alves, H. Hendon, et al. (2014). "Intra-seasonal drivers of extreme heat over Australia in observations and POAMA-2". In: *Climate dynamics* 43, pp. 1915–1937 (→ page 17).

Martius, O., C. Schwierz, and H. Davies (2010). "Tropopause-level waveguides". In: *Journal of the atmospheric sciences* 67.3, pp. 866–879 (→ page 22).

Mattingly, K. S., J. V. Turton, J. D. Wille, B. Noël, X. Fettweis, et al. (2023). "Increasing extreme melt in northeast Greenland linked to foehn winds and atmospheric rivers". In: *Nature Communications* 14.1, p. 1743 (→ page 17).

Maxwell, J. C. (1860). "V. Illustrations of the dynamical theory of gases.—Part I. On the motions and collisions of perfectly elastic spheres". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 19.124, pp. 19–32 (→ page 5).

McHugh, C. E., T. L. Delworth, W. Cooke, and L. Jia (2023). "Using large ensembles to examine historical and projected changes in record-breaking summertime temperatures over the contiguous United States". In: *Earth's Future* 11.12, e2023EF003954 (→ page 14).

McKinnon, K. A., A. Rhines, M. P. Tingley, and P. Huybers (2016a). "The changing shape of Northern Hemisphere summer temperature distributions". In: *Journal of Geophysical Research: Atmospheres* 121.15, pp. 8849–8868 (→ page 30).

McKinnon, K. A., A. Rhines, M. Tingley, and P. Huybers (2016b). "Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures". In: *Nature Geoscience* 9.5, pp. 389–394 (→ page 29).

McPhillips, L. E., H. Chang, M. V. Chester, Y. Depietri, E. Friedman, et al. (2018). "Defining extreme events: A cross-disciplinary review". In: *Earth's Future* 6.3, pp. 441–455 (→ page 11).

Meehl, G. A. and C. Tebaldi (2004). "More intense, more frequent, and longer lasting heat waves in the 21st century". In: *Science* 305.5686, pp. 994–997 (→ page 17).

Merrifield, A. L., I. R. Simpson, K. A. McKinnon, S. Sippel, S.-P. Xie, et al. (2019). "Local and nonlocal land surface influence in European heatwave initial condition ensembles". In: *Geophysical Research Letters* 46.23, pp. 14082–14092 (→ page 28).

Metzner, P., C. Schütte, and E. Vanden-Eijnden (2006). "Illustration of transition path theory on a collection of simple examples". In: *The Journal of Chemical Physics* 125.8. Publisher: American Institute of Physics, p. 084110. DOI: 10.1063/1.2335447 (→ page 109).

Miloshevich, G., B. Cozian, P. Abry, P. Borgnat, and F. Bouchet (2023). "Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data". In: *Physical Review Fluids* 8.4, p. 040501 (→ page 138).

Miralles, D. G., P. Gentine, S. I. Seneviratne, and A. J. Teuling (2019). "Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges". In: *Annals of the New York Academy of Sciences* 1436.1, pp. 19–35 (→ pages 25, 27, 28, 35).

Miralles, D. G., A. J. Teuling, C. C. Van Heerwaarden, and J. Vilà-Guerau de Arellano (2014). "Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation". In: *Nature geoscience* 7.5, pp. 345–349 (→ pages 19, 27).

Moon, W., B.-M. Kim, G.-H. Yang, and J. S. Wettlaufer (2022). "Wavier jet streams driven by zonally asymmetric surface thermal forcing". In: *Proceedings of the National Academy of Sciences* 119.38, e2200890119 (→ page 35).

Mueller, B. and S. I. Seneviratne (2012). "Hot days induced by precipitation deficits at the global scale". In: *Proceedings of the national academy of sciences* 109.31, pp. 12398–12403 (→ page 26).

— (2014). "Systematic land climate and evapotranspiration biases in CMIP5 simulations". In: *Geophysical research letters* 41.1, pp. 128–134 (→ page 35).

Mueller, N. D., E. E. Butler, K. A. McKinnon, A. Rhines, M. Tingley, et al. (2016). "Cooling of US Midwest summer temperature extremes from cropland intensification". In: *Nature Climate Change* 6.3, pp. 317–322 (→ page 31).

Mukherjee, S., A. K. Mishra, J. Zscheischler, and D. Entekhabi (2023). "Interaction between dry and hot extremes at a global scale using a cascade modeling framework". In: *Nature Communications* 14.1, p. 277 (→ page 26).

Muñoz, C., D. Schultz, and G. Vaughan (2020). "A midlatitude climatology and interannual variability of 200-and 500-hPa cut-off lows". In: *Journal of Climate* 33.6, pp. 2201–2222 (→ page 191).

Murto, S., L. Papritz, G. Messori, R. Caballero, G. Svensson, et al. (2023). "Extreme surface energy budget anomalies in the high Arctic in winter". In: *Journal of Climate* 36.11, pp. 3591–3609 (→ page 17).

Nakamura, H., M. Nakamura, and J. L. Anderson (1997). "The role of high- and low-frequency dynamics in blocking formation". In: *Monthly Weather Review* 125.9, pp. 2074–2093 (→ page 18).

Neal, E., C. S. Huang, and N. Nakamura (2022). "The 2021 Pacific Northwest heat wave and associated blocking: meteorology and the role of an upstream cyclone as a diabatic source of wave activity". In: *Geophysical Research Letters* 49.8, e2021GL097699 (→ page 21).

Noé, F., C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl (2009). "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations". en. In: *Proceedings of the National Academy of Sciences* 106.45. Publisher: National Academy of Sciences Section: Biological Sciences, pp. 19011–19016. DOI: 10.1073/pnas.0905466106 (→ page 101).

Nyawo, P. T. and H. Touchette (2017). "A minimal model of dynamical phase transition". In: *Europhysics Letters* 116.5, p. 50009 (→ page 180).

O'Neill, B. C., E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, et al. (2014). "A new scenario framework for climate change research: the concept of shared socioeconomic pathways". In: *Climatic change* 122, pp. 387–400 (→ page 217).

O'Reilly, C. H., D. J. Befort, A. Weisheimer, T. Woollings, A. Ballinger, et al. (2021). "Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty". In: *Communications Earth & Environment* 2.1, p. 194 (→ page 31).

Ossó, A., R. P. Allan, E. Hawkins, L. Shaffrey, and D. Maraun (2022). "Emerging new climate extremes over Europe". en. In: *Climate Dynamics* 58.1, pp. 487–501. DOI: 10.1007/s00382-021-05917-3 (→ page 31).

Otto, F. E., N. Massey, G. J. van Oldenborgh, R. G. Jones, and M. R. Allen (2012). "Reconciling two approaches to attribution of the 2010 Russian heat wave". In: *Geophysical Research Letters* 39.4 (→ page 15).

Palmer, T. (2013). "Climate extremes and the role of dynamics". In: *Proceedings of the National Academy of Sciences* 110.14, pp. 5281–5282 (→ page 22).

— (2019). "Stochastic weather and climate models". In: *Nature Reviews Physics* 1.7, pp. 463–471 (→ page 105).

Patterson, M. (2023). "North-West Europe Hottest Days Are Warming Twice as Fast as Mean Summer Days". In: *Geophysical Research Letters* 50.10, e2023GL102757 (→ pages 15, 31, 32, 37, 208).

Peixoto, J. P. and A. H. Oort (1992). "Physics of climate". In: (→ pages 4, 9).

Peixóto, J. P. and A. H. Oort (1984). "Physics of climate". In: *Reviews of Modern Physics* 56.3, p. 365 (→ page 9).

Perkins, S. E. (2015). "A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale". In: *Atmospheric Research* 164, pp. 242–267 (→ pages 15–17, 26, 35).

Perkins, S. E. and L. V. Alexander (2013). "On the measurement of heat waves". In: *Journal of climate* 26.13, pp. 4500–4517 (→ page 16).

Perkins-Kirkpatrick, S. E. and P. B. Gibson (2017). "Changes in regional heatwave characteristics as a function of increasing global temperature". In: *Scientific Reports* 7.1, p. 12256 (→ page 31).

Perkins-Kirkpatrick, S. and S. Lewis (2020). "Increasing trends in regional heatwaves". In: *Nature communications* 11.1, p. 3357 (→ pages 15, 30).

Perron, O. (1929). "Über stabilität und asymptotisches verhalten der integrale von differentialgleichungssystemen". In: *Mathematische Zeitschrift* 29.1, pp. 129–160 (→ page 5).

Peters, O., K. Christensen, and J. D. Neelin (2012). "Rainfall and dragon-kings". In: *The European Physical Journal Special Topics* 205.1, pp. 147–158 (→ page 181).

Petoukhov, V., S. Petri, S. Rahmstorf, D. Coumou, K. Kornhuber, et al. (2016). "Role of quasiresonant planetary wave dynamics in recent boreal spring-to-autumn extreme events". In: *Proceedings of the National Academy of Sciences* 113.25, pp. 6862–6867 (→ page 23).

Petoukhov, V., S. Rahmstorf, S. Petri, and H. J. Schellnhuber (2013). "Quasiresonant amplification of planetary waves and recent Northern Hemisphere weather extremes". In: *Proceedings of the National Academy of Sciences* 110.14, pp. 5336–5341 (→ page 22).

Pfahl, S. and H. Wernli (2012). "Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-) daily time scales". In: *Geophysical Research Letters* 39.12 (→ page 17).

Pfahl, S., C. Schwierz, M. Croci-Maspoli, C. M. Grams, and H. Wernli (2015). "Importance of latent heat release in ascending air streams for atmospheric blocking". In: *Nature Geoscience* 8.8, pp. 610–614 (→ pages 18, 21).

Pfleiderer, P. and D. Coumou (2018). "Quantification of temperature persistence over the Northern Hemisphere land-area". In: *Climate Dynamics* 51.1, pp. 627–637 (→ pages 19, 34).

Pfleiderer, P., C.-F. Schleussner, K. Kornhuber, and D. Coumou (2019). "Summer weather becomes more persistent in a 2 °C world". en. In: *Nature Climate Change* 9.9. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Atmospheric dynamics;Climate-change impacts;Climate sciences;Projection and prediction Subject_term_id: atmospheric-dynamics;climate-change-impacts;climate-sciences;projection-and-prediction, pp. 666–671. DOI: 10.1038/s41558-019-0555-0 (→ pages 23, 34).

Philip, S. Y., S. F. Kew, G. J. Van Oldenborgh, F. S. Anslow, S. I. Seneviratne, et al. (2021). "Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021". In: *Earth System Dynamics Discussions* 2021, pp. 1–34 (→ pages 51, 53, 150).

Pinheiro, H. R., K. I. Hodges, M. A. Gan, and N. J. Ferreira (2017). "A new perspective of the climatological features of upper-level cut-off lows in the Southern Hemisphere". In: *Climate Dynamics* 48, pp. 541–559 (→ page 191).

Platzman, G. W. (1968). "The Rossby wave". In: *Quarterly Journal of the Royal Meteorological Society* 94.401, pp. 225–248 (→ page 22).

Plotkin, D. A., R. J. Webber, M. E. O'Neill, J. Weare, and D. S. Abbot (2019). "Maximizing Simulated Tropical Cyclone Intensity With Action Minimization". en. In: *Journal of Advances in Modeling Earth Systems* 11.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001419, pp. 863–891. DOI: 10.1029/2018MS001419 (→ pages 104, 105, 122, 123).

Poincaré, H. (1881). "Mémoire sur les courbes définies par une équation différentielle". In: *Journal de mathématiques pures et appliquées* 7, pp. 375–422 (→ page 5).

Power, S. B. and F. P. D. Delage (2019). "Setting and smashing extreme temperature records over the coming century". en. In: *Nature Climate Change* 9.7. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Projection and prediction Subject_term_id: climate-change-impacts;projection-and-prediction, pp. 529–534. DOI: 10.1038/s41558-019-0498-5 (→ page 30).

Previdi, M., K. L. Smith, and L. M. Polvani (2021). "Arctic amplification of climate change: a review of underlying mechanisms". In: *Environmental Research Letters* 16.9, p. 093003 (→ page 33).

Pyrina, M., W. Wicker, A. J. de Vries, G. Fragkoulidis, and D. I. Domeisen (2024). "Rossby wave packets driving concurrent and non-concurrent heatwaves in the Northern and Southern Hemisphere mid-latitudes". In: *EGUsphere* 2024, pp. 1–44 (→ page 25).

Quesada, B., R. Vautard, P. Yiou, M. Hirschi, and S. I. Seneviratne (2012). "Asymmetric European summer heat predictability from wet and dry southern winters and springs". In: *Nature Climate Change* 2.10, pp. 736–741 (→ page 27).

Quinting, J. F. and M. J. Reeder (2017). "Southeastern Australian heat waves from a trajectory viewpoint". In: *Monthly Weather Review* 145.10, pp. 4109–4125 (→ page 18).

Ragone, F. and F. Bouchet (2021). "Rare Event Algorithm Study of Extreme Warm Summers and Heatwaves Over Europe". en. In: *Geophysical Research Letters* 48.12. _eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL091197, e2020GL091197. DOI: 10.1029/2020GL091197 (→ pages 104, 105, 186, 213).
Ragone, F. and F. Bouchet (2020). "Computation of Extreme Values of Time Averaged Observables in Climate Models with Large Deviation Techniques". en. In: *Journal of Statistical Physics* 179.5, pp. 1637–1665. DOI: 10.1007/s10955-019-02429-7 (→ page 186).
Ragone, F., J. Wouters, and F. Bouchet (2018). "Computation of extreme heat waves in climate models using a large deviation algorithm". en. In: *Proceedings of the National Academy of Sciences* 115.1, pp. 24–29. DOI: 10.1073/pnas.1712645115 (→ pages 103, 105, 126, 128, 129, 183, 184, 186).
Randel, W. and I. Held (1991). "Phase speed spectra of transient eddy fluxes and critical layer absorption". In: *Journal of the atmospheric sciences* 48.5, pp. 688–697 (→ page 189).
Riboldi, J., E. Rousi, F. d'Andrea, G. Rivière, and F. Lott (2022). "Circumglobal Rossby wave patterns during boreal winter highlighted by space–time spectral analysis". In: *Weather and Climate Dynamics* 3.2, pp. 449–469 (→ page 189).
Rogers, C. D., K. Kornhuber, S. E. Perkins-Kirkpatrick, P. C. Loikith, and D. Singh (2022). "Sixfold increase in historical Northern Hemisphere concurrent large heatwaves driven by warming and changing atmospheric circulations". In: *Journal of Climate* 35.3, pp. 1063–1078 (→ pages 31, 33).
Rolland, J. (2021). "Collapse of transitional wall turbulence captured using a rare events algorithm". In: *arXiv:2103.16460 [physics]*. arXiv: 2103.16460 (→ page 130).
Rolland, J. and E. Simonnet (2015). "Statistical behaviour of adaptive multilevel splitting algorithms in simple models". In: *Journal of Computational Physics* 283, pp. 541–558 (→ page 138).
Rossby, C.-G. (1939). "Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action". In: *J. mar. Res.* 2, pp. 38–55 (→ page 21).
Rößler, A. (2010). "Runge–Kutta methods for the strong approximation of solutions of stochastic differential equations". In: *SIAM Journal on Numerical Analysis* 48.3, pp. 922–952 (→ pages 114, 132).
Röthlisberger, M., L. Frossard, L. F. Bosart, D. Keyser, and O. Martius (2019). "Recurrent synoptic-scale Rossby wave patterns and their effect on the persistence of cold and hot spells". In: *Journal of Climate* 32.11, pp. 3207–3226 (→ page 24).
Röthlisberger, M. and O. Martius (2019). "Quantifying the local effect of Northern Hemisphere atmospheric blocks on the persistence of summer hot and dry spells". In: *Geophysical Research Letters* 46.16, pp. 10101–10111 (→ page 20).
Röthlisberger, M. and L. Papritz (2023). "Quantifying the physical processes leading to atmospheric hot extremes at a global scale". In: *Nature Geoscience* 16.3, pp. 210–216 (→ page 18).
Röthlisberger, M., S. Pfahl, and O. Martius (2016). "Regional-scale jet waviness modulates the occurrence of midlatitude weather extremes". In: *Geophysical Research Letters* 43.20, pp. 10–989 (→ pages 23, 24).
Röthlisberger, M., M. Sprenger, E. Flaounas, U. Beyerle, and H. Wernli (2020). "The substructure of extremely hot summers in the Northern Hemisphere". In: *Weather and Climate Dynamics* 1.1, pp. 45–62 (→ page 193).
Rousi, E., A. H. Fink, L. S. Andersen, F. N. Becker, G. Beobide-Arsuaga, et al. (2023). "The extremely hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective". In: *Natural Hazards and Earth System Sciences* 23.5, pp. 1699–1718 (→ page 17).
Rousi, E., K. Kornhuber, G. Beobide-Arsuaga, F. Luo, and D. Coumou (2022). "Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia". In: *Nature communications* 13.1, p. 3851 (→ pages 15, 31, 33).
Rubino, G., B. Tuffin, et al. (2009). *Rare event simulation using Monte Carlo methods*. Vol. 73. Wiley Online Library (→ page 101).
Ruelle, D. (1999). "Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics". In: *Journal of Statistical Physics* 95, pp. 393–468 (→ page 7).
— (2009). "A review of linear response theory for general differentiable dynamical systems". en. In: *Nonlinearity* 22.4. Publisher: IOP Publishing, pp. 855–870. DOI: 10.1088/0951-7715/22/4/009 (→ page 6).
Ruelle, D. and F. Takens (1971). "On the Nature of Turbulence". In: *Commun. math. Phys* 20, pp. 167–192 (→ page 6).
Ruosteenoja, K. and K. Jylhä (2023). "Average and extreme heatwaves in Europe at 0.5–2.0 C global warming levels in CMIP6 model simulations". In: *Climate Dynamics* 61.9, pp. 4259–4281 (→ page 31).
Russo, E. and D. I. Domeisen (2023). "Increasing intensity of extreme heatwaves: the crucial role of metrics". In: *Geophysical Research Letters* 50.14, e2023GL103540 (→ pages 15, 16).
Russo, S., A. Dosio, R. G. Graversen, J. Sillmann, H. Carrao, et al. (2014). "Magnitude of extreme heat waves in present climate and their projection in a warming world". en. In: *Journal of Geophysical Research: Atmospheres* 119.22. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD022098, pp. 12, 500–12, 512. DOI: 10.1002/2014JD022098 (→ page 15).
Russo, S., J. Sillmann, and E. M. Fischer (2015). "Top ten European heatwaves since 1950 and their occurrence in the coming decades". In: *Environmental Research Letters* 10.12, p. 124003 (→ pages 16, 17).
Samset, B. H., M. Sand, C. J. Smith, S. E. Bauer, P. M. Forster, et al. (2018). "Climate impacts from a removal of anthropogenic aerosol emissions". In: *Geophysical Research Letters* 45.2, pp. 1020–1029 (→ page 29).
Sánchez-Benítez, A., R. García-Herrera, D. Barriopedro, P. M. Sousa, and R. M. Trigo (2018). "June 2017: the earliest European summer mega-heatwave of Reanalysis Period". In: *Geophysical Research Letters* 45.4, pp. 1955–1962 (→ page 17).
Sánchez-Benítez, A., H. Goessling, F. Pithan, T. Semmler, and T. Jung (2022). "The July 2019 European heat wave in a warmer climate: Storyline scenarios with a coupled model using spectral nudging". In: *Journal of Climate* 35.8, pp. 2373–2390 (→ page 17).
Sangelantoni, L., S. Sobolowski, T. Lorenz, Ø. Hodnebrog, R. Cardoso, et al. (2023). "Investigating the representation of heatwaves from an ensemble of km-scale regional climate simulations within CORDEX-FPS convection". In: *Climate Dynamics*, pp. 1–37 (→ page 19).
Santos, J. A., S. Pfahl, J. G. Pinto, and H. Wernli (2015). "Mechanisms underlying temperature extremes in Iberia: a Lagrangian perspective". In: *Tellus A: Dynamic Meteorology and Oceanography* 67.1, p. 26032 (→ page 18).
Sapsis, T. P. (2021). "Statistics of Extreme Events in Fluid Flows and Waves". In: *Annual Review of Fluid Mechanics* 53.1. _eprint: https://doi.org/10.1146/annurev-fluid-030420-032810, pp. 85–111. DOI: 10.1146/annurev-fluid-030420-032810 (→ page 144).
Sato, T. and T. Nakamura (2019). "Intensification of hot Eurasian summers by climate change and land–atmosphere interactions". In: *Scientific reports* 9.1, p. 10866 (→ pages 28, 35).
Sauer, J., J. Demaeyer, G. Zappa, F. Massonnet, and F. Ragone (2024). "Extremes of summer Arctic sea ice reduction investigated with a rare event algorithm". In: *Climate Dynamics*, pp. 1–19 (→ pages 104, 105).
Schaller, N., J. Sillmann, J. Anstey, E. M. Fischer, C. M. Grams, et al. (2018). "Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles". In: *Environmental Research Letters* 13.5, p. 054015 (→ page 17).
Schemm, S., G. Rivière, L. M. Ciasto, and C. Li (2018). "Extratropical cyclogenesis changes in connection with tropospheric ENSO teleconnections to the North Atlantic: Role of stationary and transient waves". In: *Journal of the Atmospheric Sciences* 75.11, pp. 3943–3964 (→ page 189).
Schielicke, L. and S. Pfahl (2022). "European heatwaves in present and future climate simulations: A Lagrangian analysis". In: *Weather and Climate Dynamics* 3.4, pp. 1439–1459 (→ page 18).
Schneider, T. (2006). "The general circulation of the atmosphere". In: *Annu. Rev. Earth Planet. Sci.* 34, pp. 655–688 (→ page 3).
Schneider, T., T. Bischoff, and H. Płotka (2015). "Physics of changes in synoptic midlatitude temperature variability". In: *Journal of Climate* 28.6, pp. 2312–2331 (→ page 21).
Schorlepp, T., T. Grafke, S. May, and R. Grauer (2022). "Spontaneous symmetry breaking for extreme vorticity and strain in the three-dimensional Navier–Stokes equations". In: *Philosophical Transactions of the Royal Society A* 380.2226, p. 20210051 (→ page 120).
Schubert, S., H. Wang, and M. Suarez (2011). "Warm season subseasonal variability and climate extremes in the Northern Hemisphere: The role of stationary Rossby waves". In: *Journal of Climate* 24.18, pp. 4773–4792 (→ page 22).
Schumacher, D., M. Hauser, and S. I. Seneviratne (2022a). "Drivers and mechanisms of the 2021 Pacific Northwest heatwave". In: *Earth's Future* 10.12, e2022EF002967 (→ page 21).
Schumacher, D. L., J. Keune, P. Dirmeyer, and D. G. Miralles (2022b). "Drought self-propagation in drylands due to land–atmosphere feedbacks". In: *Nature geoscience* 15.4, pp. 262–268 (→ page 27).
Schumacher, D. L., J. Singh, M. Hauser, E. M. Fischer, M. Wild, et al. (2024). "Exacerbated summer European warming not captured by climate models neglecting long-term aerosol changes". In: *Communications Earth & Environment* 5.1, p. 182 (→ pages 29, 51).
Schumacher, D. L., J. Keune, C. C. van Heerwaarden, J. Vilà-Guerau de Arellano, A. J. Teuling, et al. (2019). "Amplification of mega-heatwaves through heat torrents fuelled by upwind drought". en. In: *Nature Geoscience* 12.9. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Hydrology;Natural hazards Subject_term_id: climate-change-impacts;hydrology;natural-hazards, pp. 712–717. DOI: 10.1038/s41561-019-0431-6 (→ page 27).
Schwingshackl, C., M. Hirschi, and S. I. Seneviratne (2018). "Global contributions of incoming radiation and land surface conditions to maximum near-surface air temperature variability and trend". In: *Geophysical research letters* 45.10, pp. 5034–5044 (→ page 35).
Screen, J. A., C. Deser, D. M. Smith, X. Zhang, R. Blackport, et al. (2018). "Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models". In: *Nature Geoscience* 11.3, pp. 155–163 (→ page 33).
Screen, J. A. and I. Simmonds (2014). "Amplified mid-latitude planetary waves favour particular regional weather extremes". In: *Nature Climate Change* 4.8, pp. 704–709 (→ page 23).
Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, et al. (2010). "Investigating soil moisture–climate interactions in a changing climate: A review". In: *Earth-Science Reviews* 99.3-4, pp. 125–161 (→ pages 26, 28, 35, 51, 221).
Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär (2006). "Land–atmosphere coupling and climate change in Europe". In: *Nature* 443.7108, pp. 205–209 (→ page 35).
Seneviratne, S. I., M. Wilhelm, T. Stanelle, B. van den Hurk, S. Hagemann, et al. (2013). "Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment". In: *Geophysical Research Letters* 40.19, pp. 5212–5217 (→ page 26).
Seneviratne, S. I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, et al. (2021). "Weather and climate extreme events in a changing climate (Chapter 11)". In: (→ pages 10, 30).
Serrano-Notivoli, R., E. Tejedor, P. Sarricolea, O. Meseguer-Ruiz, M. de Luis, et al. (2023). "Unprecedented warmth: A look at Spain's exceptional summer of 2022". In: *Atmospheric Research*, p. 106931 (→ page 17).
Shaw, T. and A. Voigt (2015). "Tug of war on summertime circulation between radiative forcing and sea surface warming". In: *Nature Geoscience* 8.7, pp. 560–566 (→ page 35).
Shaw, T. A. (2019). "Mechanisms of future predicted changes in the zonal mean mid-latitude circulation". In: *Current Climate Change Reports* 5.4, pp. 345–357 (→ page 32).

Shepherd, T. G. (2014). "Atmospheric circulation as a source of uncertainty in climate change projections". In: *Nature Geoscience* 7.10, pp. 703–708 (→ pages 31, 32).

Shepherd, T. G., E. Boyd, R. A. Calel, S. C. Chapman, S. Dessai, et al. (2018). "Storylines: an alternative approach to representing uncertainty in physical aspects of climate change". en. In: *Climatic Change* 151.3, pp. 555–571. DOI: 10.1007/s10584-018-2317-9 (→ pages 215, 253).

Simmons, A. J. (2022). "Trends in the tropospheric general circulation from 1979 to 2022". In: *Weather and Climate Dynamics* 3.3, pp. 777–809 (→ page 32).

Simolo, C. and S. Corti (2022). "Quantifying the role of variability in future intensification of heat extremes". In: *Nature Communications* 13.1, p. 7930 (→ page 30).

Simonnet, E., J. Rolland, and F. Bouchet (2021). "Multistability and Rare Spontaneous Transitions in Barotropic beta-Plane Turbulence". EN. In: *Journal of the Atmospheric Sciences* 78.6. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, pp. 1889–1911. DOI: 10.1175/JAS-D-20-0279.1 (→ pages 103, 105).

Sinai, Y. G. (1972). "Gibbs measures in ergodic theory". In: *Russian Mathematical Surveys* 27.4, p. 21 (→ page 6).

Singh, J., S. Sippel, and E. M. Fischer (2023). "Circulation dampened heat extremes intensification over the Midwest USA and amplified over Western Europe". In: *Communications Earth & Environment* 4.1, p. 432 (→ pages 32, 51).

Sippel, S., J. Zscheischler, M. Heimann, F. E. Otto, J. Peters, et al. (2015). "Quantifying changes in climate variability and extremes: Pitfalls and their overcoming". In: *Geophysical Research Letters* 42.22, pp. 9990–9998 (→ page 16).

Skrzyńska, M. and R. Twardosz (2023). "Long-term changes in the frequency of exceptionally cold and warm months in Europe (1831–2020)". In: *International Journal of Climatology* 43.5, pp. 2339–2351 (→ page 31).

Smale, S. (1967). "Differentiable dynamical systems". In: *Bulletin of the American mathematical Society* 73.6, pp. 747–817 (→ page 6).

Sobel, A. H., C.-Y. Lee, S. G. Bowen, S. J. Camargo, M. A. Cane, et al. (2023). "Near-term tropical cyclone risk and coupled Earth system model biases". In: *Proceedings of the National Academy of Sciences* 120.33, e2209631120 (→ page 248).

Sornette, D. and G. Ouillon (2012). "Dragon-kings: Mechanisms, statistical methods and empirical evidence". en. In: *The European Physical Journal Special Topics* 205.1, pp. 1–26. DOI: 10.1140/epjst/e2012-01559-5 (→ page 180).

Sousa, P. M., R. M. Trigo, D. Barriopedro, P. M. Soares, and J. A. Santos (2018). "European temperature responses to blocking and ridge regional patterns". In: *Climate Dynamics* 50, pp. 457–477 (→ page 17).

Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, et al. (2017). "Review of tropical-extratropical teleconnections on intraseasonal time scales". In: *Reviews of Geophysics* 55.4, pp. 902–937 (→ pages 21, 29).

Stefanon, M., F. D'Andrea, and P. Drobinski (2012). "Heatwave classification over Europe and the Mediterranean region". In: *Environmental Research Letters* 7.1, p. 014023 (→ pages 13, 17).

Stéfanon, M., P. Drobinski, F. d'Andrea, and N. de Noblet-Ducoudré (2012). "Effects of interactive vegetation phenology on the 2003 summer heat waves". In: *Journal of Geophysical Research: Atmospheres* 117.D24 (→ page 28).

Steinfeld, D. and S. Pfahl (2019). "The role of latent heating in atmospheric blocking dynamics: a global climatology". In: *Climate Dynamics* 53.9, pp. 6159–6180 (→ pages 18, 21, 25).

Stendel, M., J. Francis, R. White, P. D. Williams, and T. Woollings (2021). "The jet stream and climate change". In: *Climate change*. Elsevier, pp. 327–357 (→ page 32).

Stott, P. A., D. A. Stone, and M. R. Allen (2004). "Human contribution to the European heatwave of 2003". In: *Nature* 432.7017, pp. 610–614 (→ page 15).

Strahan, J., A. Antoszewski, C. Lorpaiboon, B. P. Vani, J. Weare, et al. (2021). "Long-time-scale predictions from short-trajectory data: A benchmark analysis of the trp-cage miniprotein". In: *Journal of chemical theory and computation* 17.5, pp. 2948–2963 (→ page 118).

Strandberg, G. and E. Kjellström (2019). "Climate impacts from afforestation and deforestation in Europe". In: *Earth Interactions* 23.1, pp. 1–27 (→ page 26).

Suarez-Gutierrez, L., W. A. Müller, and J. Marotzke (2023). "Extreme heat and drought typical of an end-of-century climate could occur over Europe soon and repeatedly". In: *Communications Earth & Environment* 4.1, p. 415 (→ page 14).

Sutton, R. T. (2019). "Climate Science Needs to Take Risk Assessment Much More Seriously". EN. In: *Bulletin of the American Meteorological Society* 100.9. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, pp. 1637–1642. DOI: 10.1175/BAMS-D-18-0280.1 (→ page 99).

Swanson, K. L., P. J. Kushner, and I. M. Held (1997). "Dynamics of barotropic storm tracks". In: *Journal of the atmospheric sciences* 54.7, pp. 791–810 (→ page 22).

Tailleur, J. and J. Kurchan (2007). "Probing rare physical trajectories with Lyapunov weighted dynamics". en. In: *Nature Physics* 3.3. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 3 Primary_atype: Research Publisher: Nature Publishing Group, pp. 203–207. DOI: 10.1038/nphys515 (→ pages 128, 138).

Taleb, N. N. (2010). *The Black Swan:: The Impact of the Highly Improbable*. Vol. 2. Random house trade paperbacks (→ page 99).

Tantet, A. J. J. et al. (2016). "Ergodic theory of climate: variability, stability and response". PhD thesis. Utrecht University (→ page 5).

Tél, T., T. Bódai, G. Drótos, T. Haszpra, M. Herein, et al. (2020). "The Theory of Parallel Climate Realizations". en. In: *Journal of Statistical Physics* 179.5, pp. 1496–1530. DOI: 10.1007/s10955-019-02445-7 (→ page 9).

Teng, H. and G. Branstator (2019). "Amplification of waveguide teleconnections in the boreal summer". In: *Current Climate Change Reports* 5, pp. 421–432 (→ pages 22, 23, 33).

Teng, H., G. Branstator, G. A. Meehl, and W. M. Washington (2016). "Projected intensification of subseasonal temperature variability and heat waves in the Great Plains". In: *Geophysical Research Letters* 43.5, pp. 2165–2173 (→ pages 35, 243).

Teng, H., G. Branstator, A. B. Tawfik, and P. Callaghan (2019). "Circumglobal response to prescribed soil moisture over North America". In: *Journal of Climate* 32.14, pp. 4525–4545 (→ page 27).

Teng, H., G. Branstator, H. Wang, G. A. Meehl, and W. M. Washington (2013). "Probability of US heat waves affected by a subseasonal planetary wave pattern". In: *Nature Geoscience* 6.12, pp. 1056–1061 (→ page 23).

Teuling, A. J., A. F. Van Loon, S. I. Seneviratne, I. Lehner, M. Aubinet, et al. (2013). "Evapotranspiration amplifies European summer drought". In: *Geophysical Research Letters* 40.10, pp. 2071–2075 (→ page 27).

Thiede, E. H., D. Giannakis, A. R. Dinner, and J. Weare (2019). "Galerkin approximation of dynamical quantities using trajectory data". In: *The Journal of Chemical Physics* 150.24. Publisher: American Institute of Physics, p. 244111. DOI: 10.1063/1.5063730 (→ page 118).

Thiery, W., E. L. Davin, D. M. Lawrence, A. L. Hirsch, M. Hauser, et al. (2017). "Present-day irrigation mitigates heat extremes". In: *Journal of Geophysical Research: Atmospheres* 122.3, pp. 1403–1422 (→ page 26).

Thiery, W., A. J. Visser, E. M. Fischer, M. Hauser, A. L. Hirsch, et al. (2020). "Warming of hot extremes alleviated by expanding irrigation". In: *Nature communications* 11.1, p. 290 (→ page 31).

Touchette, H. (2009). "The large deviation approach to statistical mechanics". en. In: *Physics Reports* 478.1, pp. 1–69. DOI: 10.1016/j.physrep.2009.05.002 (→ page 197).

Trenberth, K. E. (1986). "An assessment of the impact of transient eddies on the zonal flow during a blocking episode using localized Eliassen-Palm flux diagnostics". In: *Journal of the Atmospheric Sciences* 43.19, pp. 2070–2087 (→ page 189).

Trenberth, K. E. and J. T. Fasullo (2012). "Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010". In: *Journal of Geophysical Research: Atmospheres* 117.D17 (→ page 15).

Tripathy, K. P. and A. K. Mishra (2023). "How unusual is the 2022 European compound drought and heatwave event?" In: *Geophysical Research Letters* 50.15, e2023GL105453 (→ page 17).

Tuel, A. and O. Martius (2024). "Persistent warm and cold spells in the Northern Hemisphere extratropics: regionalisation, synoptic-scale dynamics and temperature budget". In: *Weather and Climate Dynamics* 5.1, pp. 263–292 (→ page 24).

Tuel, A., D. Steinfeld, S. M. Ali, M. Sprenger, and O. Martius (2022). "Large-Scale Drivers of Persistent Extreme Weather During Early Summer 2021 in Europe". In: *Geophysical Research Letters* 49.18, e2022GL099624 (→ page 17).

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press (→ pages 4, 21).

Van Loon, S. and D. W. Thompson (2023). "Comparing Local Versus Hemispheric Perspectives of Extreme Heat Events". In: *Geophysical Research Letters* 50.24, e2023GL105246 (→ pages 30, 35).

Vanden-Eijnden, E. (2006). "Transition path theory". In: *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*. Springer, pp. 453–493 (→ page 107).

Vanden-Eijnden, E. et al. (2010). "Transition-path theory and path-finding algorithms for the study of rare events." In: *Annual review of physical chemistry* 61, pp. 391–420 (→ pages 101, 107, 118).

Vanden-Eijnden, E. (2014). "Transition path theory". In: *An introduction to Markov state models and their application to long timescale molecular simulation*, pp. 91–100 (→ page 109).

Vautard, R., J. Cattiaux, T. Happé, J. Singh, R. Bonnet, et al. (2023). "Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends". In: *Nature Communications* 14.1, p. 6803 (→ pages 15, 31, 32, 37, 51, 208).

Vautard, R., A. Gobiet, D. Jacob, M. Belda, A. Colette, et al. (2013). "The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project". In: *Climate dynamics* 41, pp. 2555–2575 (→ page 19).

Vautard, R., P. Yiou, F. D'andrea, N. De Noblet, N. Viovy, et al. (2007). "Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit". In: *Geophysical Research Letters* 34.7 (→ page 27).

Ventura, S., J. R. Miró, J. C. Peña, and G. Villalba (2023). "Analysis of synoptic weather patterns of heatwave events". In: *Climate Dynamics* 61.9, pp. 4679–4702 (→ page 32).

Vogel, M. M., J. Zscheischler, E. M. Fischer, and S. I. Seneviratne (2020). "Development of future heatwaves for different hazard thresholds". In: *Journal of Geophysical Research: Atmospheres* 125.9, e2019JD032070 (→ page 31).

Vogel, M. M., J. Zscheischler, and S. I. Seneviratne (2018). "Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe". In: *Earth System Dynamics* 9.3, pp. 1107–1125 (→ pages 27, 28, 35).

Vogel, M. M., R. Orth, F. Cheruy, S. Hagemann, R. Lorenz, et al. (2017). "Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks". In: *Geophysical Research Letters* 44.3, pp. 1511–1519 (→ pages 26, 27).

Wallberg, L., L. Suarez-Gutierrez, D. Matei, and W. A. Müller (2024). "Extremely warm European summers preceded by sub-decadal North Atlantic ocean heat accumulation". In: *Earth System Dynamics* 15.1, pp. 1–14 (→ page 29).

Wang, H., S. D. Schubert, R. D. Koster, and Y. Chang (2019). "Phase locking of the boreal summer atmospheric response to dry land surface anoma-

lies in the Northern Hemisphere". In: *Journal of Climate* 32.4, pp. 1081–1099 (→ page 27).
Wang, P., Y. Yang, D. Xue, L. Ren, J. Tang, et al. (2023). "Aerosols overtake greenhouse gases causing a warmer climate and more weather extremes toward carbon neutrality". In: *Nature Communications* 14.1, p. 7257 (→ page 29).
Weare, J. (2009). "Particle filtering with path sampling and an application to a bimodal ocean current model". en. In: *Journal of Computational Physics* 228.12, pp. 4312–4331. DOI: 10.1016/j.jcp.2009.02.033 (→ pages 102, 105).
Webber, R. J. (2019). "Unifying sequential Monte Carlo with resampling matrices". In: *arXiv preprint arXiv:1903.12583* (→ page 126).
Webber, R. J., D. A. Plotkin, M. E. O'Neill, D. S. Abbot, and J. Weare (2019). "Practical rare event sampling for extreme mesoscale weather". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.5. Publisher: American Institute of Physics, p. 053109. DOI: 10.1063/1.5081461 (→ pages 104, 105, 124, 127, 138–140, 144).
Wehner, M., P. Gleckler, and J. Lee (2020). "Characterization of long period return values of extreme daily temperature and precipitation in the CMIP6 models: Part 1, model evaluation". In: *Weather and Climate Extremes* 30, p. 100283 (→ page 31).
Wehrli, K., B. P. Guillod, M. Hauser, M. Leclair, and S. I. Seneviratne (2019). "Identifying key driving processes of major recent heat waves". In: *Journal of Geophysical Research: Atmospheres* 124.22, pp. 11746–11765 (→ page 17).
Wehrli, K., M. Hauser, and S. I. Seneviratne (2020). "Storylines of the 2018 Northern Hemisphere heatwave at pre-industrial and higher global warming levels". In: *Earth System Dynamics* 11.4, pp. 855–873 (→ page 17).
Welch, B. L. (1947). "The generalization of 'STUDENT'S'problem when several different population varlances are involved". In: *Biometrika* 34.1-2, pp. 28–35 (→ page 226).
Wernli, H. and M. Sprenger (2007). "Identification and ERA-15 climatology of potential vorticity streamers and cutoffs near the extratropical tropopause". In: *Journal of the atmospheric sciences* 64.5, pp. 1569–1586 (→ page 191).
White, R. H., K. Kornhuber, O. Martius, and V. Wirth (2022). "From atmospheric waves to heatwaves: A waveguide perspective for understanding and predicting concurrent, persistent, and extreme extratropical weather". In: *Bulletin of the American Meteorological Society* 103.3, E923–E935 (→ pages 22, 31, 32).
White, R. H., J. M. Wallace, and D. Battisti (2021). "Revisiting the role of mountains in the Northern Hemisphere winter atmospheric circulation". In: *Journal of the Atmospheric Sciences* 78.7, pp. 2221–2235 (→ page 21).
Wille, J. D., S. P. Alexander, C. Amory, R. Baiman, L. Barthélemy, et al. (2024a). "The extraordinary March 2022 East Antarctica "heat" wave. Part I: observations and meteorological drivers". In: *Journal of Climate* 37.3, pp. 757–778 (→ page 17).
— (2024b). "The extraordinary March 2022 East Antarctica "heat" wave. Part II: impacts on the Antarctic ice sheet". In: *Journal of Climate* 37.3, pp. 779–799 (→ page 17).
Williams, E., C. Funk, P. Peterson, and C. Tuholske (2024). "High resolution climate change observations and projections for the evaluation of heat-related extremes". In: *Scientific Data* 11.1, p. 261 (→ page 19).
Wirth, V. (2020). "Waveguidability of idealized midlatitude jets and the limitations of ray tracing theory". In: *Weather and Climate Dynamics* 1.1, pp. 111–125 (→ page 22).
Wirth, V. and C. Polster (2021). "The problem of diagnosing jet waveguidability in the presence of large-amplitude eddies". In: *Journal of the Atmospheric Sciences* 78.10, pp. 3137–3151 (→ pages 22, 24).
Wirth, V., M. Riemer, E. K. Chang, and O. Martius (2018). "Rossby wave packets on the midlatitude waveguide—A review". In: *Monthly Weather Review* 146.7, pp. 1965–2001 (→ pages 22, 24).
WMO (2023). *Guidelines on the Definition and Characterization of Extreme Weather and Climate Events* (→ page 11).
Wolf, G., A. Czaja, D. Brayshaw, and N. Klingaman (2020). "Connection between sea surface anomalies and atmospheric quasi-stationary waves". In: *Journal of Climate* 33.1, pp. 201–212 (→ page 29).
Woollings, T., D. Barriopedro, J. Methven, S.-W. Son, O. Martius, et al. (2018). "Blocking and its response to climate change". In: *Current climate change reports* 4, pp. 287–300 (→ pages 20, 33).
Wouters, J. and F. Bouchet (2016). "Rare event computation in deterministic chaotic systems using genealogical particle analysis". en. In: 49.37. Publisher: IOP Publishing, p. 374002. DOI: 10.1088/1751-8113/49/37/374002 (→ pages 103, 105, 140, 146).
Wouters, J., R. K. Schiemann, and L. C. Shaffrey (2023). "Rare event simulation of extreme European winter rainfall in an intermediate complexity climate model". In: *Journal of Advances in Modeling Earth Systems* 15.4, e2022MS003537 (→ pages 104, 105).
Wunsch, C. and R. Ferrari (2004). "Vertical mixing, energy, and the general circulation of the oceans". In: *Annu. Rev. Fluid Mech.* 36, pp. 281–314 (→ page 3).
Xoplaki, E., J. F. González-Rouco, J. Luterbacher, and H. Wanner (2003). "Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs". In: *Climate dynamics* 20, pp. 723–739 (→ page 17).
Al-Yaari, A., Y. Zhao, F. Cheruy, and W. Thiery (2023). "Heatwave characteristics in the recent climate and at different global warming levels: A multimodel analysis at the global scale". In: *Earth's Future* 11.9, e2022EF003301 (→ pages 28, 31, 35).
Yiou, P., C. Cadiou, D. Faranda, A. Jézéquel, N. Malhomme, et al. (2023). "Ensembles of climate simulations to anticipate worst case heatwaves during the Paris 2024 Olympics". In: *npj climate and atmospheric science* 6.1, p. 188 (→ pages 104, 105, 249).
Yiou, P. and A. Jézéquel (2020). "Simulation of extreme heat waves with empirical importance sampling". en. In: *Geoscientific Model Development* 13.2, pp. 763–781. DOI: 10.5194/gmd-13-763-2020 (→ pages 104, 105, 121, 122).
Zampieri, M., F. D'andrea, R. Vautard, P. Ciais, N. de Noblet-Ducoudré, et al. (2009). "Hot European summers and the role of soil moisture in the propagation of Mediterranean drought". In: *Journal of Climate* 22.18, pp. 4747–4758 (→ page 27).
Zeder, J., S. Sippel, O. C. Pasche, S. Engelke, and E. M. Fischer (2023). "The effect of a short observational record on the statistics of temperature extremes". In: *Geophysical Research Letters* 50.16, e2023GL104090 (→ pages 53, 99, 253).
Zhang, X., T. Zhou, W. Zhang, L. Ren, J. Jiang, et al. (2023). "Increased impact of heat domes on 2021-like heat extremes in North America under global warming". In: *Nature Communications* 14.1, p. 1690 (→ page 32).
Zhang, Y. and W. R. Boos (2023). "An upper bound for extreme temperatures over midlatitude land". In: *Proceedings of the National Academy of Sciences* 120.12, e2215278120 (→ pages 37, 51, 52, 54, 99, 251).
Zhao, A., M. A. Bollasina, and D. S. Stevenson (2019). "Strong influence of aerosol reductions on future heatwaves". In: *Geophysical Research Letters* 46.9, pp. 4913–4923 (→ page 29).
Zimin, A. V., I. Szunyogh, D. Patil, B. R. Hunt, and E. Ott (2003). "Extracting envelopes of Rossby wave packets". In: *Monthly weather review* 131.5, pp. 1011–1017 (→ page 190).
Zschenderlein, P., A. H. Fink, S. Pfahl, and H. Wernli (2019). "Processes determining heat waves across different European climates". In: *Quarterly Journal of the Royal Meteorological Society* 145.724, pp. 2973–2989 (→ page 18).
Zschenderlein, P., G. Fragkoulidis, A. H. Fink, and V. Wirth (2018). "Large-scale Rossby wave and synoptic-scale dynamic analyses of the unusually late 2016 heatwave over Europe". In: *Weather* 73.9, pp. 275–283 (→ pages 18, 25).
Zschenderlein, P., S. Pfahl, H. Wernli, and A. H. Fink (2020). "A Lagrangian analysis of upper-tropospheric anticyclones associated with heat waves in Europe". In: *Weather and Climate Dynamics* 1.1, pp. 191–206 (→ pages 18, 19, 21, 25).

Appendix A

Supplementary materials

A.1 Supplementary materials for chapter 1

A.1.1 Relative error for the estimation of an observable in a rare set

In this section I show that the estimation the expectation of a general observable in a rare set is rendered difficult by the large variance associated — the so-called burden of rareness.

Let us note $(X_i)_{1\leq i\leq N}$ N independent samplings of the system of interest. Let us consider ψ an observable and A a set with a low probability p_A. I want to estimate the mean value of ψ in the set A. The naive estimator of $\mathbb{E}[\psi(X) \mid X \in A]$ is:

$$\hat{\psi} = \frac{\sum_{i=1}^{N} \psi(X_i)\mathbb{1}(X_i \in A)}{\sum_{i=1}^{N} \mathbb{1}(X_i \in A)}. \tag{A.1}$$

The numerator is noted as $\mathcal{N} := \sum_{i=1}^{N} \psi(X_i)\mathbb{1}(X_i \in A)$ and the denominator as $\mathcal{D} := \sum_{i=1}^{N} \mathbb{1}(X_i \in A)$. I also note $\psi_A = \psi(X)\mathbb{1}(X \in A)$. Then, using the independence of each term in the two sums:

$$\mathbb{E}[\mathcal{N}] = N\mathbb{E}[\psi_A] \tag{A.2}$$
$$\mathbb{E}[\mathcal{D}] = Np_A \tag{A.3}$$
$$\mathbb{V}[\mathcal{D}] = Np_A(1-p_A) \tag{A.4}$$
$$\mathbb{E}[\mathcal{N}\mathcal{D}] = N\mathbb{E}[\psi_A]((N-1)p_A + 1) \tag{A.5}$$
$$\mathrm{Cov}[\mathcal{N},\mathcal{D}] = \mathbb{E}[\mathcal{N}\mathcal{D}] - \mathbb{E}[\mathcal{N}]\mathbb{E}[\mathcal{D}] = N\mathbb{E}[\psi_A](1-p_A) \tag{A.6}$$

Therefore one can estimate the expectation of $\hat{\psi}$ (Kendall et al. 1948):

$$\mathbb{E}[\hat{\psi}] = \frac{\mathbb{E}[\mathcal{N}]}{\mathbb{E}[\mathcal{D}]}\left(1 - \frac{\mathrm{Cov}[\mathcal{N},\mathcal{D}]}{\mathbb{E}[\mathcal{N}]\mathbb{E}[\mathcal{D}]} + \frac{\mathbb{V}[\mathcal{D}]}{\mathbb{E}[\mathcal{D}]^2}\right) = \frac{\mathbb{E}[\psi_A]}{p_A} = \mathbb{E}[\psi(X) \mid X \in A]. \tag{A.7}$$

This shows that the naïve estimator is unbiased. I now want to estimate the variance of this estimator. One needs to compute the second moment of $\hat{\psi}$ the same way. For the second moment of $\mathcal{N}$ and $\mathcal{D}$ one shows easily that:

$$\mathbb{E}[\mathcal{N}^2] = N\left(\mathbb{E}[\psi_A^2] + (N-1)\mathbb{E}[\psi_A]^2\right) \tag{A.8}$$

$$\mathbb{E}[\mathcal{D}^2] = Np_A(1 + (N-1)p_A). \tag{A.9}$$

The term $\mathbb{E}[\mathcal{N}^2\mathcal{D}^2]$ is less straightforward. To begin, here is the definition of the product:

$$\mathcal{N}^2\mathcal{D}^2 = \sum_{i,j,k,m=1}^{N} \psi(X_i)\psi(X_j)\mathbb{1}(X_i \in A)\mathbb{1}(X_j \in A)\mathbb{1}(X_k \in A)\mathbb{1}(X_m \in A). \tag{A.10}$$

This sum has N^4 terms that can be decomposed into four cases for the quadruplet (i, j, k, m):

- two terms in the quadruplet are equal and the two others are different and different from one another, e.g. $i = j = 2$, $k = 3$ and $m = 4$. Among them:
  - there are $N(N-1)(N-2)$ such quadruplets for which $i = j$ and the associated expectation for each of them is:

$$\mathbb{E}[\psi(X_i)^2\mathbb{1}(X_i \in A)]\mathbb{E}[\mathbb{1}(X_k \in A)]^2 = \mathbb{E}[\psi_A^2]p_A^2 \tag{A.11}$$

  - there are $5N(N-1)(N-2)$ such quadruplets for which $i \neq j$ and the associated expectation for each of them is:

$$\mathbb{E}[\psi(X_i)\mathbb{1}(X_i \in A)]^2\mathbb{E}[\mathbb{1}(X_k \in A)] = \mathbb{E}[\psi_A]^2 p_A \tag{A.12}$$

- three terms in the quadruplet are equal and the last one is different from them, e.g. $i = j = k \neq m$. Among them:
  - there are $2N(N-1)$ such quadruplets for which $i = j$ and the associated expectation for each of them is:

$$\mathbb{E}[\psi(X_i)^2\mathbb{1}(X_i \in A)]\mathbb{E}[\mathbb{1}(X_k \in A)] = \mathbb{E}[\psi_A^2]p_A \tag{A.13}$$

  - there are $2N(N-1)$ such quadruplets for which $i \neq j$ and the associated expectation for each of them is:

$$\mathbb{E}[\psi(X_i)\mathbb{1}(X_i \in A)]^2 = \mathbb{E}[\psi_A]^2 \tag{A.14}$$

- the four terms in the quadruplet are equal: there are N such quadruplets and the associated expectation is:

$$\mathbb{E}[\psi(X_i)^2\mathbb{1}(X_i \in A)] = \mathbb{E}[\psi_A^2] \tag{A.15}$$

- the four terms in the quadruplet are different: there are $N^4 - 6N(N-1)(N-2) - 4N(N-1) - N$ such quadruplets and the associated expectation is:

$$\mathbb{E}[\psi(X_i)\mathbb{1}(X_i \in A)]^2\mathbb{E}[\mathbb{1}(X_k \in A)]^2 = \mathbb{E}[\psi_A]^2 p_A^2 \tag{A.16}$$

In the end:

$$\begin{aligned}\mathbb{E}[\mathcal{N}^2\mathcal{D}^2] = &\Big(N^4 - 6N(N-1)(N-2) - 4N(N-1) - N\Big)\,\mathbb{E}[\psi_A]^2 p_A^2 \\ &+ N(N-1)(N-2)\mathbb{E}[\psi_A^2]p_A^2 \\ &+ 5N(N-1)(N-2)\mathbb{E}[\psi_A]^2 p_A \\ &+ 2N(N-1)\mathbb{E}[\psi_A^2]p_A \\ &+ 2N(N-1)\mathbb{E}[\psi_A]^2 \\ &+ N\mathbb{E}[\psi_A^2].\end{aligned} \tag{A.17}$$

Applying this formula for $\psi = 1$, one retrieves the fourth moment of $\mathcal{D}$:

$$\begin{aligned}\mathbb{E}[\mathcal{D}^4] = &\Big(N^4 - 6N(N-1)(N-2) - 4N(N-1) - N\Big)\,p_A^4 \\ &+ 6N(N-1)(N-2)p_A^3 \\ &+ 4N(N-1)p_A^2 \\ &+ Np_A.\end{aligned} \tag{A.18}$$

The relative error in the estimator $\hat{\psi}$ should then be estimated. It is defined as:

$$\mathrm{RE} := \frac{\sqrt{\mathbb{V}[\hat{\psi}]}}{\mathbb{E}[\hat{\psi}]} = \sqrt{\frac{\mathbb{E}[\hat{\psi}^2]}{\mathbb{E}[\hat{\psi}]^2} - 1}. \tag{A.19}$$

The expectation of the second moment of $\hat{\psi}$ can be estimated by the same formula as above but it leads to unnecessary heavy computations and I now introduce the hypotheses which take into account the fact that the expectation is estimated over a set with low probability. The first hypothesis is that $p_A \ll 1$ and one can typically take $p_A \sim 10^{-3}$. The second hypothesis is that I am interested in events that are at the edge of what is available with the data set, i.e. $Np_A \sim 1$, therefore here $N \sim 10^3$. Using the conditional expectation of ψ, by definition:

$$\mathbb{E}[\psi_A] = p_A\mathbb{E}[\psi(X) \mid X \in A], \tag{A.20}$$

$$\mathbb{E}[\psi_A^2] = p_A\mathbb{E}[\psi(X)^2 \mid X \in A]. \tag{A.21}$$

The hypotheses therefore gives:

$$\mathbb{E}[\mathcal{N}^2] \simeq \mathbb{E}[\psi(X)^2 \mid X \in A] + \mathbb{E}[\psi(X) \mid X \in A]^2 \tag{A.22}$$

$$\mathbb{E}[\mathcal{D}^2] \simeq 2 \tag{A.23}$$

$$\mathbb{E}[\mathcal{D}^4] \simeq 12 \tag{A.24}$$

$$\mathbb{E}[\mathcal{N}^2\mathcal{D}^2] \simeq 7\mathbb{E}[\psi(X) \mid X \in A]^2 + 5\mathbb{E}[\psi(X)^2 \mid X \in A]. \tag{A.25}$$

Therefore, after introducing the ratio $r = \frac{\mathbb{E}[\psi(X)^2|X\in A]}{\mathbb{E}[\psi(X)|X\in A]^2} \geq 1$:

$$\mathbb{E}[\hat{\psi}^2] = \frac{\mathbb{E}[\mathcal{N}^2]}{\mathbb{E}[\mathcal{D}^2]}\left(1 - \frac{\mathbb{E}[\mathcal{N}^2\mathcal{D}^2]}{\mathbb{E}[\mathcal{N}^2]\mathbb{E}[\mathcal{D}^2]} + \frac{\mathbb{E}[\mathcal{D}^4]}{\mathbb{E}[\mathcal{D}^2]^2}\right) \tag{A.26}$$

$$= \mathbb{E}[\psi(X) \mid X \in A]^2 \frac{r+1}{2}\left(1 - \frac{5r+7}{2r+2} + \frac{12}{4}\right), \tag{A.27}$$

after which follows:

$$\mathrm{RE} = \frac{\sqrt{\mathbb{V}[\hat{\psi}]}}{\mathbb{E}[\hat{\psi}]} = \sqrt{\frac{r+1}{2}(4 - \frac{5r+7}{2r+2}) - 1}. \tag{A.28}$$

The function on the right hand side is a strictly increasing function of r. It is null — which corresponds to an infinitely precise estimate — for $r = 1$ but it increases strongly with r: for r as small as 1.1, the relative error is already of 27%. For $r = 2.35$, the error in the estimate of the observable is of the same order of magnitude as the estimate.

For an estimation that is precise at 10%, one needs $r \simeq 1.01$. If one assumes that the conditional distribution of ψ in the set A is Gaussian with a mean μ and a variance σ^2, this conditions imposes:

$$\frac{\mu^2 + \sigma^2}{\mu^2} \leq 1.01 \tag{A.29}$$

which gives:

$$\sigma \leq \frac{\mu}{10}. \tag{A.30}$$

In other words, the standard deviation of ψ in the set A has to be 10 times smaller than how much the average of ψ in the set A deviates from its climatological average. This condition is very stringent for most climate extremes.

A.2 Supplementary materials for chapter 4

A.2.1 Ensemble boosting estimator for extreme events

This section proposes a derivation of the estimator associated with the ensemble boosting method proposed by Gessner et al. (2021). It allows in particular to recover climatological probabilities of extreme events. The idea of the derivation originates from the work of V. Humphrey and L. Bloin-Wibe that we discussed during a visit at ETH Zürich. My contribution here is to propose a formalization of the derivation of the estimator and compute its variance.

a) Ensemble boosting estimator

Let us assume that one has access to an initial simulation which is N-days long. This would typically be the summer days of an initial run of a climate model. Let $(X_i)_{1\leq i\leq N}$ be the corresponding state vectors of the climate system (including atmosphere, land, eventually oceans etc) and $(T_i := T(X_i))_{1\leq i\leq N}$ the associated temperatures at a place of interest. Let us note T_{ref} a given temperature reference threshold that would correspond to a high quantile of the temperatures distribution. One can consider $T_{ext} \geq T_{ref}$ an extreme value above this reference threshold.

Let us note AC_t the ensemble of climate system conditions such that at at a lead time t, the temperature at the place of interest is equal or greater than the reference threshold:

$$AC_t := \{X_0 \mid T(X_t) \geq T_{ref}\}, \tag{A.31}$$

where X_t is the successor of the starting conditions X_0 under the dynamics of the system – i.e. if the model starts at X_0 the state vector will be equal to X_t at time t. Let us consider a slightly bigger ensemble of starting conditions to account for the perturbations added during the boosting procedure:

$$\tilde{AC}_t^{\epsilon} := \{X_0 \mid \exists X \in AC_t, \|X_0 - X\| \leq \epsilon\} \tag{A.32}$$

for some norm $\|.\|$ and some $\epsilon > 0$ which controls the size of the perturbations added at the initial step. Clearly, $AC_t \subset \tilde{AC}_t^{\epsilon}$.

Then by definition of the conditional probability:

$$\mathbb{P}\left((T(X_t) \geq T_{ref}) \cap \tilde{AC}_t^{\epsilon}\right) = \mathbb{P}\left(T(X_t) \geq T_{ref} \mid \tilde{AC}_t^{\epsilon}\right)\mathbb{P}(\tilde{AC}_t^{\epsilon}). \tag{A.33}$$

By definition of AC_t and because $AC_t \subset \tilde{AC}_t^{\epsilon}$: $\mathbb{P}\left((T(X_t) \geq T_{ref}) \cap \tilde{AC}_t^{\epsilon}\right) = \mathbb{P}(T(X_t) \geq T_{ref})$. Similarly:

$$\mathbb{P}\left((T(X_t) \geq T_{ext}) \cap \tilde{AC}_t^{\epsilon}\right) = \mathbb{P}\left(T(X_t) \geq T_{ext} \mid \tilde{AC}_t^{\epsilon}\right)\mathbb{P}(\tilde{AC}_t^{\epsilon}) \tag{A.34}$$

and because $T_{ext} \geq T_{ref}$: $\mathbb{P}((T(X_t) \geq T_{ext}) \cap \tilde{AC}_t^{\epsilon}) = \mathbb{P}(T(X_t) \geq T_{ext})$.

It should now be remarked that if the lead time t is not too big with respect to the time scale of the seasonal cycle, the probabilities $\mathbb{P}(T(X_t) \geq T_{ref})$ and $\mathbb{P}(T(X_t) \geq T_{ext})$ represent the climatological probabilities over the summer: $\mathbb{P}_{\mu}(T \geq T_{ref})$ and $\mathbb{P}_{\mu}(T \geq T_{ext})$.

From this, one therefore derives:

$$\mathbb{P}_\mu(T \geq T_{ext}) = \mathbb{P}_\mu(T \geq T_{ref}) \frac{\mathbb{P}_\mu(T \geq T_{ext} \mid \tilde{AC}_t^\epsilon)}{\mathbb{P}_\mu(T \geq T_{ref} \mid \tilde{AC}_t^\epsilon)}. \tag{A.35}$$

In the following, for simplicity, I drop the μ subscript indicating the climatological nature of the probability estimated. On the right hand side, the first term is the probability of the T_{ref} threshold, which can be estimated straightforwardly using the initial simulation:

$$\mathbb{P}(T \geq T_{ref}) \simeq \hat{p}_{T \geq T_{ref}} := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(T_i \geq T_{ref}). \tag{A.36}$$

The boosting ensemble will provide estimators for the ratio of conditional probabilities. Let us note now $(X_i^b)_{1 \leq i \leq N_b}$ the state vectors of the climate system obtained with the boosting procedure, starting from points in $\tilde{AC}_t^\epsilon$ and $(T_i^b := T(X_i^b))_{1 \leq i \leq N_b}$ the associated temperatures at the place of interest. Here all boosted ensemble members are pooled together, whatever their starting conditions. The estimators for the numerator and the denominator are then:

$$\mathbb{P}(T \geq T_{ext} \mid \tilde{AC}_t^\epsilon) \simeq \hat{p}^b_{T \geq T_{ext}} := \frac{1}{N_b} \sum_{i=1}^{N_b} \mathbb{1}(T_i^b \geq T_{ext}) \tag{A.37}$$

and

$$\mathbb{P}(T \geq T_{ref} \mid \tilde{AC}_t^\epsilon) \simeq \hat{p}^b_{T \geq T_{ref}} := \frac{1}{N_b} \sum_{i=1}^{N_b} \mathbb{1}(T_i^b \geq T_{ref}). \tag{A.38}$$

In other words, one computes simply the empirical frequency on which, in the boosted ensemble, the T_{ref} and T_{ext} thresholds are reached. Now the questions are on the properties of the boosting estimator

$$\hat{p}_{T \geq T_{ext}} := \hat{p}_{T \geq T_{ref}} \frac{\hat{p}^b_{T \geq T_{ext}}}{\hat{p}^b_{T \geq T_{ref}}} \tag{A.39}$$

of $\mathbb{P}(T \geq T_{ext})$: is this estimator unbiased ? Is it reducing the variance ?

b) Unbiasedness of the boosting estimator

To estimate the expectation of the boosting estimator, I make the assumption that on the right hand side, the estimator $\hat{p}_{T \geq T_{ref}}$ is independent from the ratio $\frac{\hat{p}^b_{T \geq T_{ext}}}{\hat{p}^b_{T \geq T_{ref}}}$. In other words, I assume that the probability to reach T_{ref} on the initial dataset is independent from how more likely it is to reach T_{ext} than to reach T_{ref} in the boosted ensemble. I make a second approximation, which is that in the boosted ensemble the $(T_i^b)_{1 \leq i \leq N_b}$ are independent one from another.

Both of these approximations may not be correct if the ensemble AC_t contains two few starting conditions — i.e. if one chooses a too high T_{ref} threshold —, or if the lead time t used for the ensemble boosting procedure is too low — i.e. if members in the boosted ensemble do not have time to diverge sufficiently and

become quasi-independent (which is guaranteed by the mixing property of a chaotic system). There are therefore clearly several trade-offs to consider: choosing T_{ref} high enough to sample extremes but not too high to avoid having too few starting conditions, and choosing t large enough to let members separate enough but not too large to still be able to sample extremes above T_{ref}.

Under these approximations, the expectation of the boosting estimator is:

$$\mathbb{E}[\hat{p}_{T\geq T_{ext}}] = \mathbb{E}[\hat{p}_{T\geq T_{ref}}]\mathbb{E}\left[\frac{\hat{p}^b_{T\geq T_{ext}}}{\hat{p}^b_{T\geq T_{ref}}}\right]. \tag{A.40}$$

The first term on the right hand side is easily estimated:

$$\mathbb{E}[\hat{p}_{T\geq T_{ref}}] = \mathbb{P}(T \geq T_{ref}). \tag{A.41}$$

Let us note $\mathcal{N} := \sum_{i=0}^{N_b} \mathbb{1}(T_i^b \geq T_{ext})$ and $\mathcal{D} := \sum_{i=0}^{N_b} \mathbb{1}(T_i^b \geq T_{ref})$ the numerator and denominator of the ratio on the right hand side.The expectation of the ratio is then (Kendall et al. (1948)):

$$\mathbb{E}[\frac{\mathcal{N}}{\mathcal{D}}] = \frac{\mathbb{E}[\mathcal{N}]}{\mathbb{E}[\mathcal{D}]}\left(1 - \frac{\mathrm{Cov}[\mathcal{N},\mathcal{D}]}{\mathbb{E}[\mathcal{N}]\mathbb{E}[\mathcal{D}]} + \frac{\mathbb{V}[\mathcal{D}]}{\mathbb{E}[\mathcal{D}]^2}\right). \tag{A.42}$$

The terms in this equation can be estimated independently:

$$\mathbb{E}[\mathcal{N}] := N_b\mathbb{E}[\hat{p}^b_{T\geq T_{ext}}] = N_b\mathbb{P}(T \geq T_{ext} \mid \tilde{AC}^\epsilon_t) = N_b * p^b_{T\geq T_{ext}} \tag{A.43}$$

$$\mathbb{E}[\mathcal{D}] = N_b\mathbb{E}[\hat{p}^b_{T\geq T_{ref}}] = N_b\mathbb{P}(T \geq T_{ref} \mid \tilde{AC}^\epsilon_t) = N_b * p^b_{T\geq T_{ref}} \tag{A.44}$$

$$\mathbb{V}[\mathcal{D}] = \sum_{i=0}^{N_b} \mathbb{V}[\mathbb{1}(T_i^b \geq T_{ref})] = N_b * p^b_{T\geq T_{ref}}(1 - p^b_{T\geq T_{ref}}) \tag{A.45}$$

$$\mathbb{E}[\mathcal{N}\mathcal{D}] = \sum_{i,j=0}^{N_b} \mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})\mathbb{1}(T_j^b \geq T_{ref})]. \tag{A.46}$$

For the last equation, one can separate the cases:

- there are N_b cases where $i = j$:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})\mathbb{1}(T_i^b \geq T_{ref})] = \mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})] = p^b_{T\geq T_{ext}} \tag{A.47}$$

  because $T_{ext} \geq T_{ref}$,

- there are $N_b(N_b - 1)$ cases where $i \neq j$, using the independence assumption:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})\mathbb{1}(T_j^b \geq T_{ref})] = \mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_j^b \geq T_{ref})] = p^b_{T\geq T_{ext}}p^b_{T\geq T_{ref}}. \tag{A.48}$$

In the end, this gives:

$$\mathbb{E}[\mathcal{N}\mathcal{D}] = N_b \cdot p^b_{T\geq T_{ext}} * (1 + (N_b - 1) \cdot p^b_{T\geq T_{ref}}). \tag{A.49}$$

Therefore:

$$\frac{\mathbb{V}[\mathcal{D}]}{\mathbb{E}[\mathcal{D}]^2} = \frac{1 - p^b_{T\geq T_{ref}}}{N_b \cdot p^b_{T\geq T_{ref}}} \tag{A.50}$$

and

$$\frac{\text{Cov}[\mathcal{N},\mathcal{D}]}{\mathbb{E}[\mathcal{N}]\mathbb{E}[\mathcal{D}]} = \frac{\mathbb{E}[\mathcal{N}\mathcal{D}]}{\mathbb{E}[\mathcal{N}]\mathbb{E}[\mathcal{D}]} - 1 = \frac{1 + (N_b - 1)\cdot p^b_{T\geq T_{ref}}}{N_b * p^b_{T\geq T_{ref}}} - 1 = \frac{1 - p^b_{T\geq T_{ref}}}{N_b \cdot p^b_{T\geq T_{ref}}}. \quad \text{(A.51)}$$

As a result:

$$\mathbb{E}[\frac{\mathcal{N}}{\mathcal{D}}] = \frac{\mathbb{E}[\mathcal{N}]}{\mathbb{E}[\mathcal{D}]}, \quad \text{(A.52)}$$

which shows that the boosting estimator is unbiased:

$$\mathbb{E}[\hat{p}_{T\geq T_{ext}}] = \mathbb{P}(T \geq T_{ext}). \quad \text{(A.53)}$$

c) Variance of the ensemble boosting estimator

For the boosted estimator to be useful to sample extremes, one needs to show that the relative error made when using this estimator is better than when using a naive estimator on the initial simulation. I now estimate the variance of the boosted estimator. With the same independence argument as previously:

$$\mathbb{E}[\hat{p}^2_{T\geq T_{ext}}] = \mathbb{E}[\hat{p}^2_{T\geq T_{ref}}]\mathbb{E}[\frac{\mathcal{N}^2}{\mathcal{D}^2}]. \quad \text{(A.54)}$$

The $\mathbb{E}[\hat{p}^2_{T\geq T_{ref}}]$ term can be computed the same way as for the $\mathbb{E}[\mathcal{N}\mathcal{D}]$ term previously, which gives:

$$\mathbb{E}[\hat{p}^2_{T\geq T_{ref}}] = \frac{p^b_{T\geq T_{ref}} * (1 + (N_b - 1) * p^b_{T\geq T_{ref}})}{N_b}. \quad \text{(A.55)}$$

The ratio can then be computed using the same formula as previously:

$$\mathbb{E}[\frac{\mathcal{N}^2}{\mathcal{D}^2}] = \frac{\mathbb{E}[\mathcal{N}^2]}{\mathbb{E}[\mathcal{D}^2]}\left(1 - \frac{\text{Cov}[\mathcal{N}^2,\mathcal{D}^2]}{\mathbb{E}[\mathcal{N}^2]\mathbb{E}[\mathcal{D}^2]} + \frac{\mathbb{V}[\mathcal{D}^2]}{\mathbb{E}[\mathcal{D}^2]^2}\right). \quad \text{(A.56)}$$

The terms $\mathbb{E}[\mathcal{N}^2]$ and $\mathbb{E}[\mathcal{D}^2]$ can be estimated as previously:

$$\mathbb{E}[\mathcal{N}^2] = N_b * p^b_{T\geq T_{ext}} * (1 + (N_b - 1) * p^b_{T\geq T_{ext}}) \quad \text{(A.57)}$$

$$\mathbb{E}[\mathcal{D}^2] = N_b * p^b_{T\geq T_{ref}} * (1 + (N_b - 1) * p^b_{T\geq T_{ref}}). \quad \text{(A.58)}$$

The term $\mathbb{E}[\mathcal{N}^2\mathcal{D}^2]$ is less straightforward. By the definition of the product:

$$\mathbb{E}[\mathcal{N}^2\mathcal{D}^2] = \sum_{i,j,k,m=1}^{N} \mathbb{E}[\mathbb{1}(T^b_i \geq T_{ext})\mathbb{1}(T^b_j \geq T_{ext})\mathbb{1}(T^b_k \geq T_{ref})\mathbb{1}(T^b_m \geq T_{ref})]. \quad \text{(A.59)}$$

This sum has N^4 terms that can be decomposed into four cases for the quadruplet (i, j, k, m):

- two terms in the quadruplet are equal and the two others are different and different from one another, e.g. $i = j = 2$, $k = 3$ and $m = 4$. Among them:

– there are $N(N-1)(N-2)$ such quadruplets for which $i = j$ and the associated expectation for each of them is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_k^b \geq T_{ref})]\mathbb{E}[\mathbb{1}(T_m^b \geq T_{ref})] = p^b_{T\geq T_{ext}}(p^b_{T\geq T_{ref}})^2 \tag{A.60}$$

– there are $5N(N-1)(N-2)$ such quadruplets for which $i \neq j$ and the associated expectation for each of them is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_j^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_k^b \geq T_{ref})] = (p^b_{T\geq T_{ext}})^2 p^b_{T\geq T_{ref}} \tag{A.61}$$

- three terms in the quadruplet are equal and the last one is different from them, e.g. $i = j = k \neq m$. Among them:

  – there are $2N(N-1)$ such quadruplets for which $i = j$ and the associated expectation for each of them is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_k^b \geq T_{ref})] = p^b_{T\geq T_{ext}} p^b_{T\geq T_{ref}} \tag{A.62}$$

  – there are $2N(N-1)$ such quadruplets for which $i \neq j$ and the associated expectation for each of them is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_j^b \geq T_{ext})] = (p^b_{T\geq T_{ext}})^2 \tag{A.63}$$

- the four terms in the quadruplet are equal: there are N such quadruplets and the associated expectation is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})] = p^b_{T\geq T_{ext}} \tag{A.64}$$

- the four terms in the quadruplet are different: there are $N^4 - 6N(N-1)(N-2) - 4N(N-1) - N$ such quadruplets and the associated expectation is:

$$\mathbb{E}[\mathbb{1}(T_i^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_j^b \geq T_{ext})]\mathbb{E}[\mathbb{1}(T_k^b \geq T_{ref})]\mathbb{E}[\mathbb{1}(T_m^b \geq T_{ref})] = (p^b_{T\geq T_{ext}} p^b_{T\geq T_{ref}})^2 \tag{A.65}$$

Thus, in the end:

$$\begin{aligned}\mathbb{E}[\mathcal{N}^2\mathcal{D}^2] = &\left(N^4 - 6N(N-1)(N-2) - 4N(N-1) - N\right)(p^b_{T\geq T_{ext}} p^b_{T\geq T_{ref}})^2 \\ &+ N(N-1)(N-2)p^b_{T\geq T_{ext}}(p^b_{T\geq T_{ref}})^2 \\ &+ 5N(N-1)(N-2)(p^b_{T\geq T_{ext}})^2 p^b_{T\geq T_{ref}} \\ &+ 2N(N-1)p^b_{T\geq T_{ext}} p^b_{T\geq T_{ref}} \\ &+ 2N(N-1)(p^b_{T\geq T_{ext}})^2 \\ &+ N p^b_{T\geq T_{ext}}\end{aligned} \tag{A.66}$$

With a similar manner, one retrieves the fourth moment of $\mathcal{D}$ (to compute the variance of $\mathcal{D}^2$):

$$\begin{aligned}\mathbb{E}[\mathcal{D}^4] = &\left(N^4 - 6N(N-1)(N-2) - 2N(N-1) - N\right)(p^b_{T\geq T_{ref}})^4 \\ &+ 6N(N-1)(N-2)(p^b_{T\geq T_{ref}})^3 \\ &+ 4N(N-1)(p^b_{T\geq T_{ref}})^2 \\ &+ Np^b_{T\geq T_{ref}}.\end{aligned} \tag{A.67}$$

Using the formulas above, one can then give an expression for the variance of the boosted estimator and the relative error (of which I do not give a closed form here):

$$\mathbb{RE} := \frac{\sqrt{\mathbb{V}[\hat{p}_{T\geq T_{ext}}]}}{\mathbb{E}[\hat{p}_{T\geq T_{ext}}]} = \sqrt{\frac{\mathbb{E}[\hat{p}^2_{T\geq T_{ext}}]}{\mathbb{E}[\hat{p}_{T\geq T_{ext}}]^2} - 1}. \tag{A.68}$$

The relative error of the boosted estimator therefore depends on:

- the size N of the initial simulation from which the starting conditions are boosted,
- the probability $\mathbb{P}(T \geq T_{ref})$ of the reference threshold T_{ref} in the initial ensemble (this probability would be typically close to the value of the quantile chosen on the initial simulation),
- the number N_b of members in the boosted ensemble,
- the probability $\mathbb{P}(T \geq T_{ref} \mid \tilde{AC}^\epsilon_t)$ to be above the reference threshold in the boosted ensemble (which may depends only marginally on the value of the threshold itself),
- the probability $\mathbb{P}(T \geq T_{ext} \mid \tilde{AC}^\epsilon_t)$ to be above an extreme value T_{ext} in the boosted ensemble.

Of these parameters, only the three first ones can be fixed by the experimenter, but the values of the last two ones can be estimated to give an idea of the error made using the boosted estimator.

d) Numerical application

Let us do an example with realistic values of these parameters:

- I consider a 1000y-long simulation and I am interested in summer days: $N = 1000 \times 100 = 10^5$,
- I consider two quantiles of the initial empirical distribution: the 99th ($\mathbb{P}(T \geq T_{ref}) = 10^{-2}$) and the 99.99th ($\mathbb{P}(T \geq T_{ref}) = 10^{-4}$). The 99th quantile is very well estimated with $N = 10^5$ whereas the 99.99th quantile is at the top limit of what can be estimated precisely with this amount of data (only 10 days are above this value),

- I take the number of boosted members to vary from $N_b = 100$ and $N_b = 500$ to $N_b = 1000$,
- the value of the probability $\mathbb{P}(T \geq T_{ref} \mid \tilde{AC}_t^{\epsilon})$ may have to be estimated in practice, here I take for the example $\mathbb{P}(T \geq T_{ref} \mid \tilde{AC}_t^{\epsilon}) = 0.25$,
- $\mathbb{P}(T \geq T_{ext} \mid \tilde{AC}_t^{\epsilon})$ will be the varying parameter (it will define how low $\mathbb{P}(T \geq T_{ext})$ one can estimate).

Figure A.1 shows the results of the computation of the relative error for these parameters. For not so extreme values of the quantile chosen to define T_{ref}, the naive estimator performs better than the boosted ensemble. However, when one considers higher values of the quantile (Fig. A.1b), the boosted estimator performs better. As one could expect, the greater the number of members in the boosted ensemble, the more precise is the estimation. With 1000 members, one can estimate precisely probabilities one to two order of magnitudes less likely than what is possible with the naive estimator using the initial long simulation. Finally, the relative error of the boosted ensemble is bounded downward by the precision in the estimation of the quantiles chosen for T_{ref}, which shows the importance to consider a not so extreme reference threshold to choose starting conditions for the boosting.

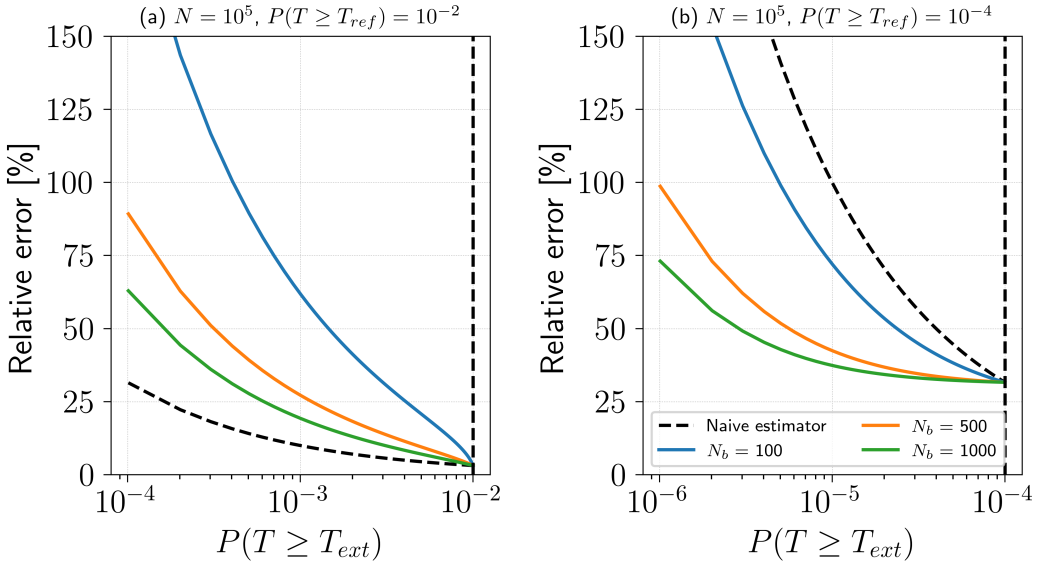


Figure A.1: **Relative error of the ensemble boosting estimator**. (a) $\mathbb{P}(T \geq T_{ref}) = 10^{-2}$ and (b) $\mathbb{P}(T \geq T_{ref}) = 10^{-4}$. The variance of the naive estimator (i.e. using only the initial long simulation) is shown in dashed line. The vertical dashed lines show $\mathbb{P}(T \geq T_{ref})$.

A.3 Supplementary materials for chapter 6

A.3.1 Additional figures

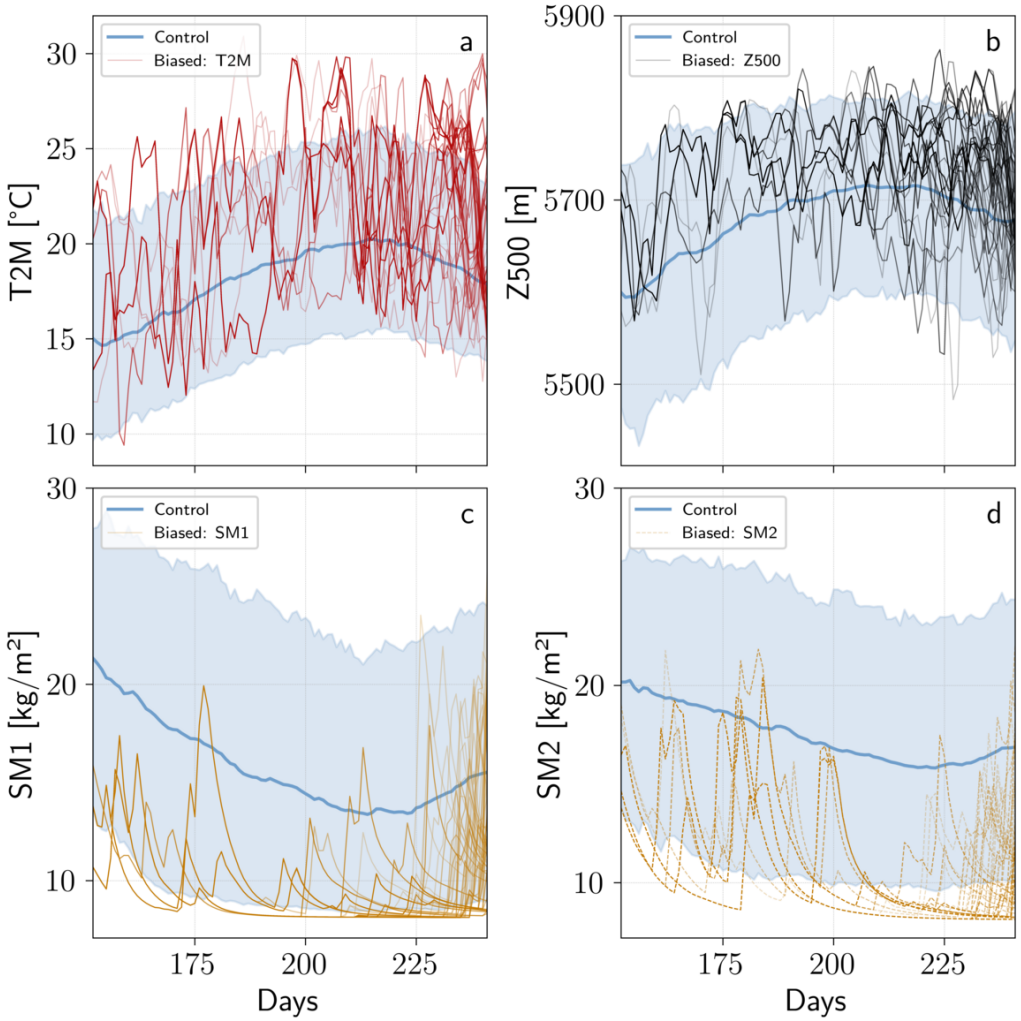


Figure A.2: **Example of individual time series of grid point observables.** All the plots are drawn for one biased simulation with $N = 100$ members. (a) 2-m air temperature, (b) geopotential height at 500hPa and (c) upper-level soil moisture at grid point 1 and (d) upper-level soil moisture at grid point 2. For all the plots the plain blue line shows the ensemble mean and the shading the 5 and 95% climatological quantiles of the variable displayed.

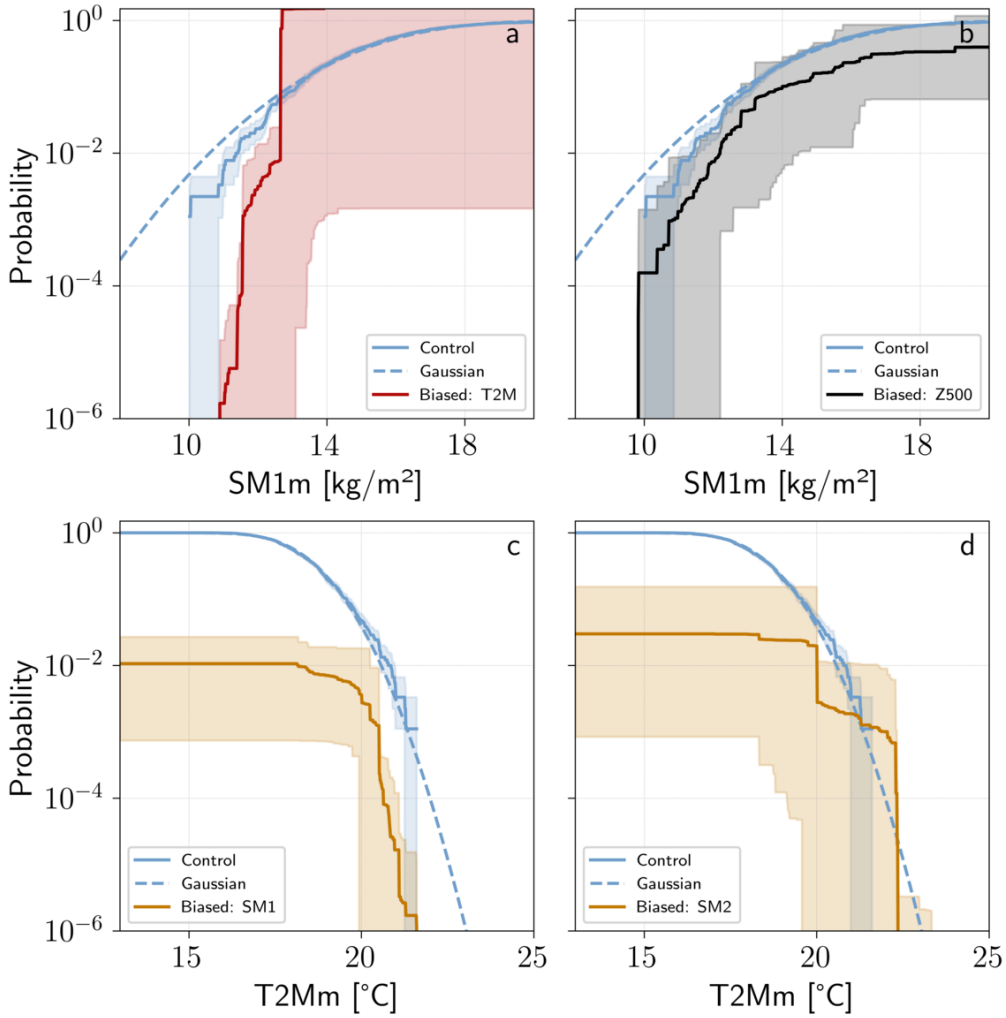


Figure A.3: **Probabilities of an extreme summer according to the naive estimator (control simulation) and the rare event estimator.** Same as Figure 6.5 with the following observables. Probability of the summer averaged upper-level soil moisture at grid point 1 to be below a certain threshold for (a) the biased T2M simulation and (b) the biased Z500 simulation. Probability of the summer averaged 2-m air temperature at grid point 1 to be above a certain threshold for (c) the biased SM1 simulation and (d) the biased SM2 simulation. The blue dashed lines represent the corresponding probabilities for a Gaussian distribution fitted on the control simulation by the method of moments. The shadings for the control simulation show the 5–95% quantiles of the estimated probability obtained using bootstrap on the 900 members. For the biased simulations the shadings show the minimum and maximum of the estimated probability over the $N_s = 9$ simulations.

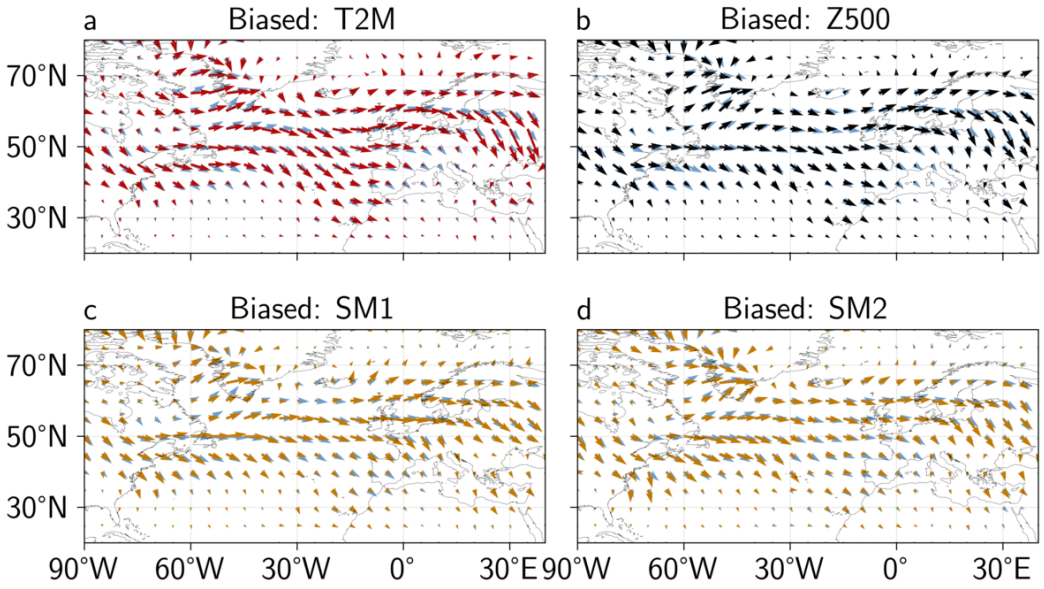


Figure A.4: **Summer averaged E-vector at 500hPa for centennial events.** In each panel the blue arrows show the climatological **E**-vector computed on the control simulation.

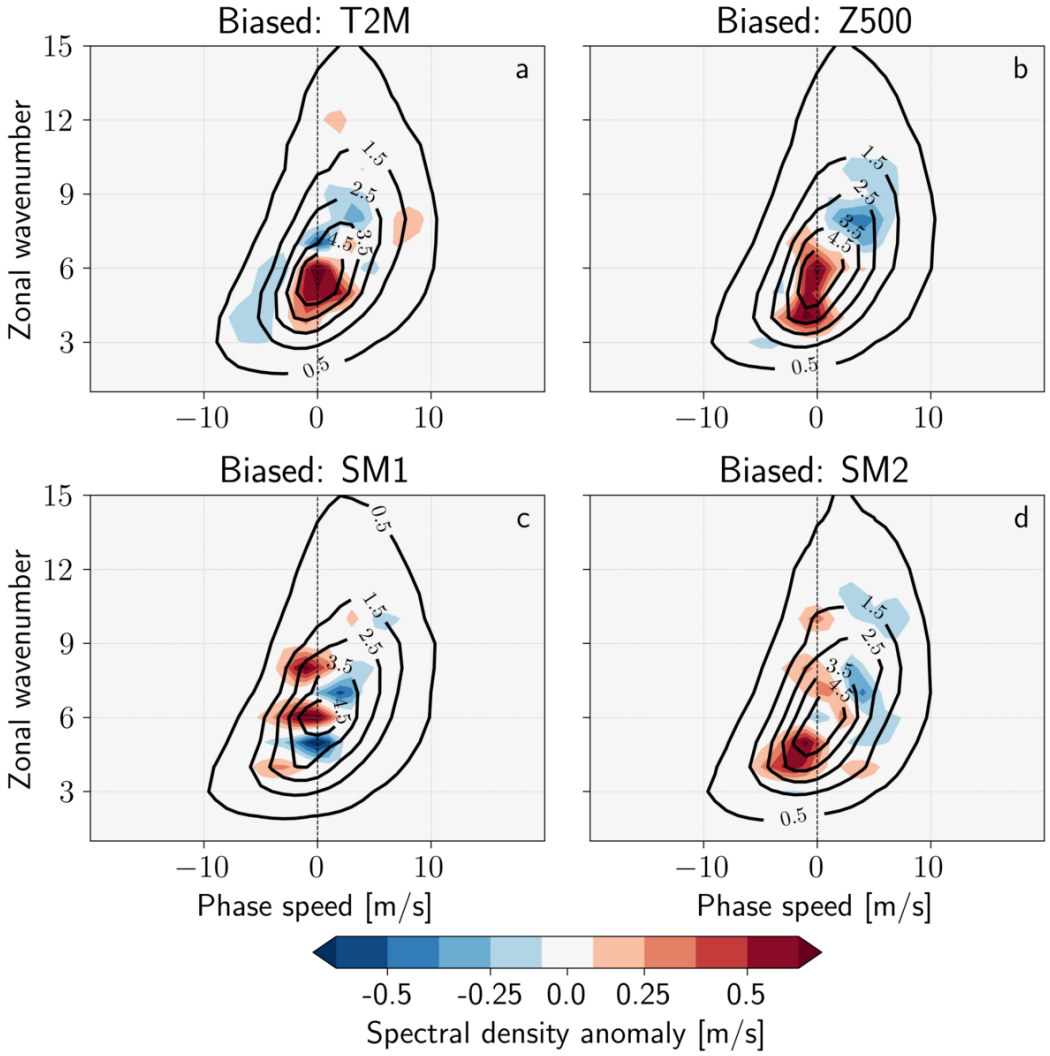


Figure A.5: **Wavenumber-phase speed spectra of meridional wind anomalies at 500hPa for centennial events.** Raw spectrum (contours) and difference with the climatological spectrum (colors).

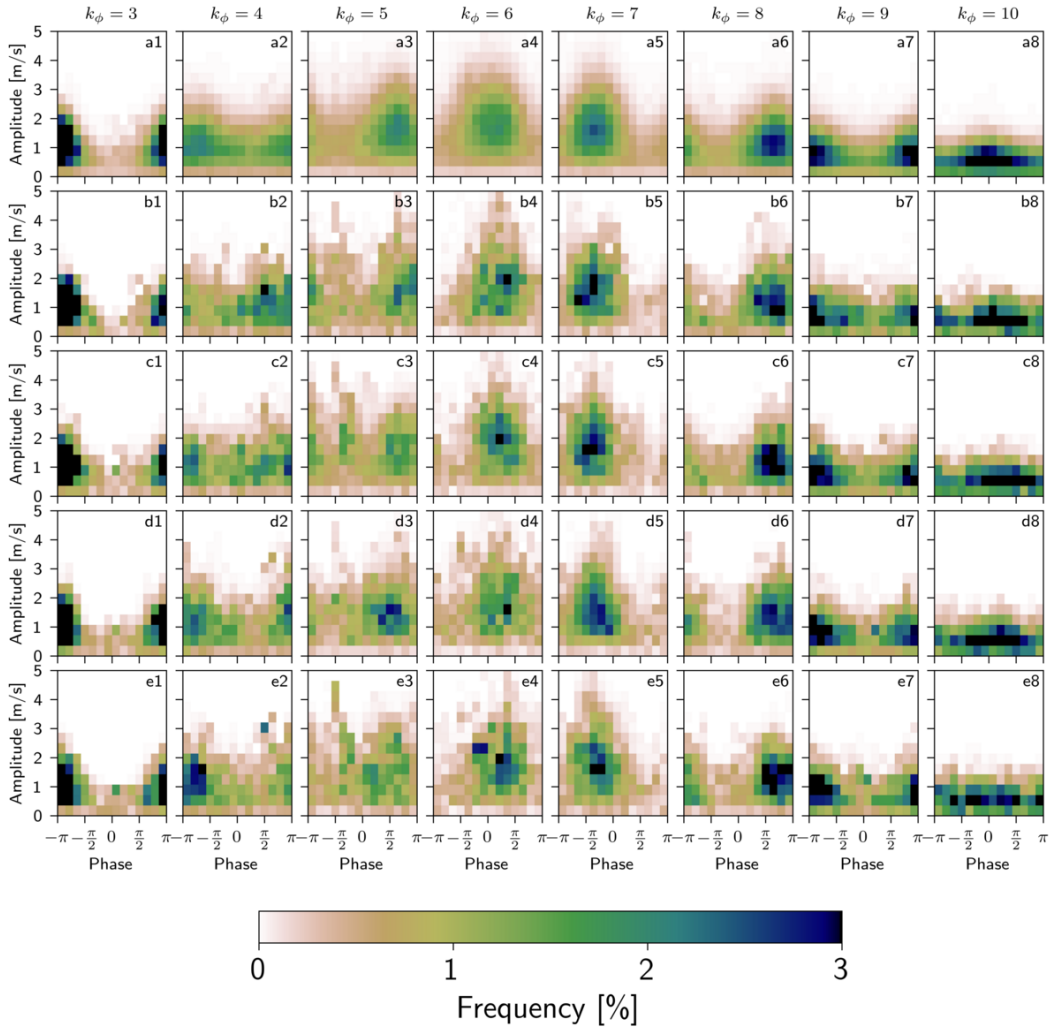


Figure A.6: **Amplitude-phase histograms of meridional wind speed at 200hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) control simulation, (b) biased T2M simulation, (c) biased Z500 simulation, (d) biased SM1 simulation and (e) biased biased SM2 simulation.

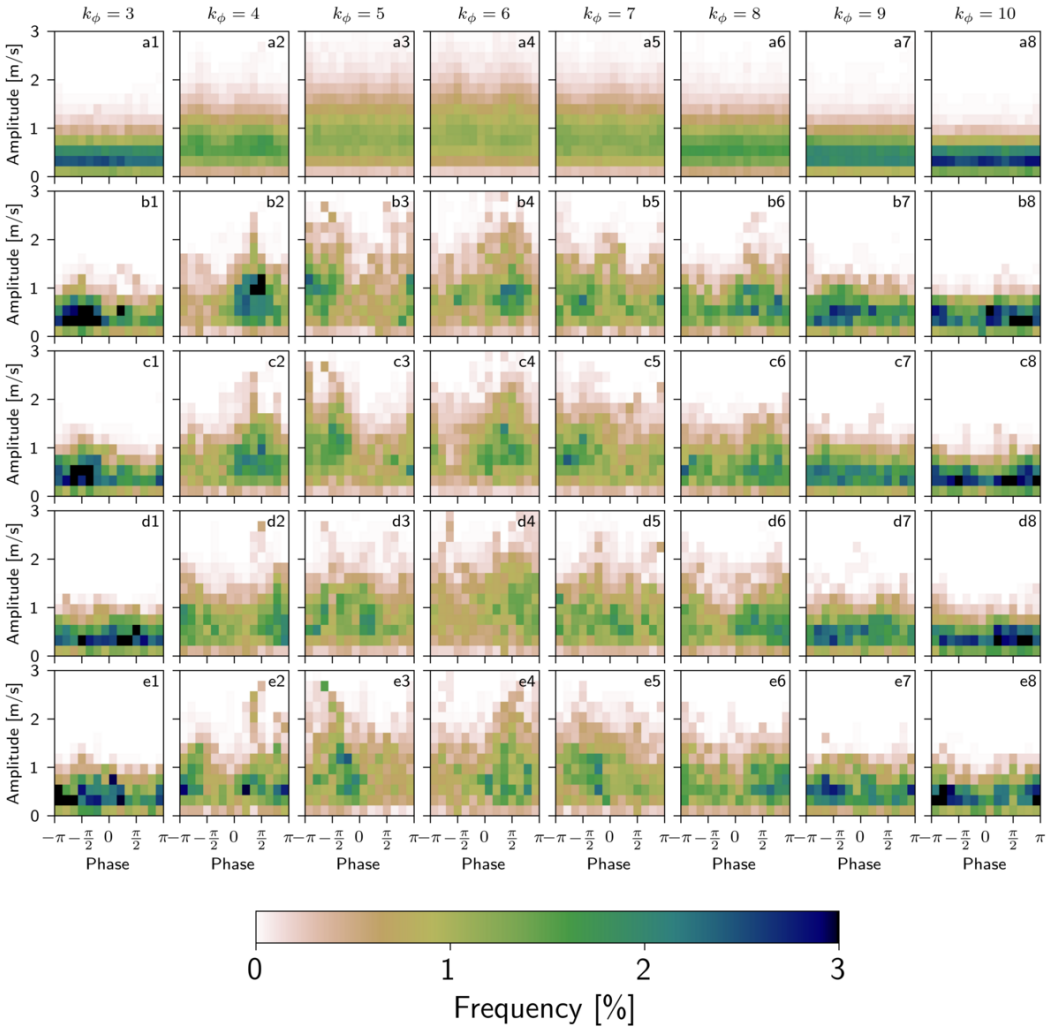


Figure A.7: **Amplitude-phase histograms of meridional wind speed anomalies at 500hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) control simulation, (b) biased T2M simulation, (c) biased Z500 simulation, (d) biased SM1 simulation and (e) biased biased SM2 simulation.

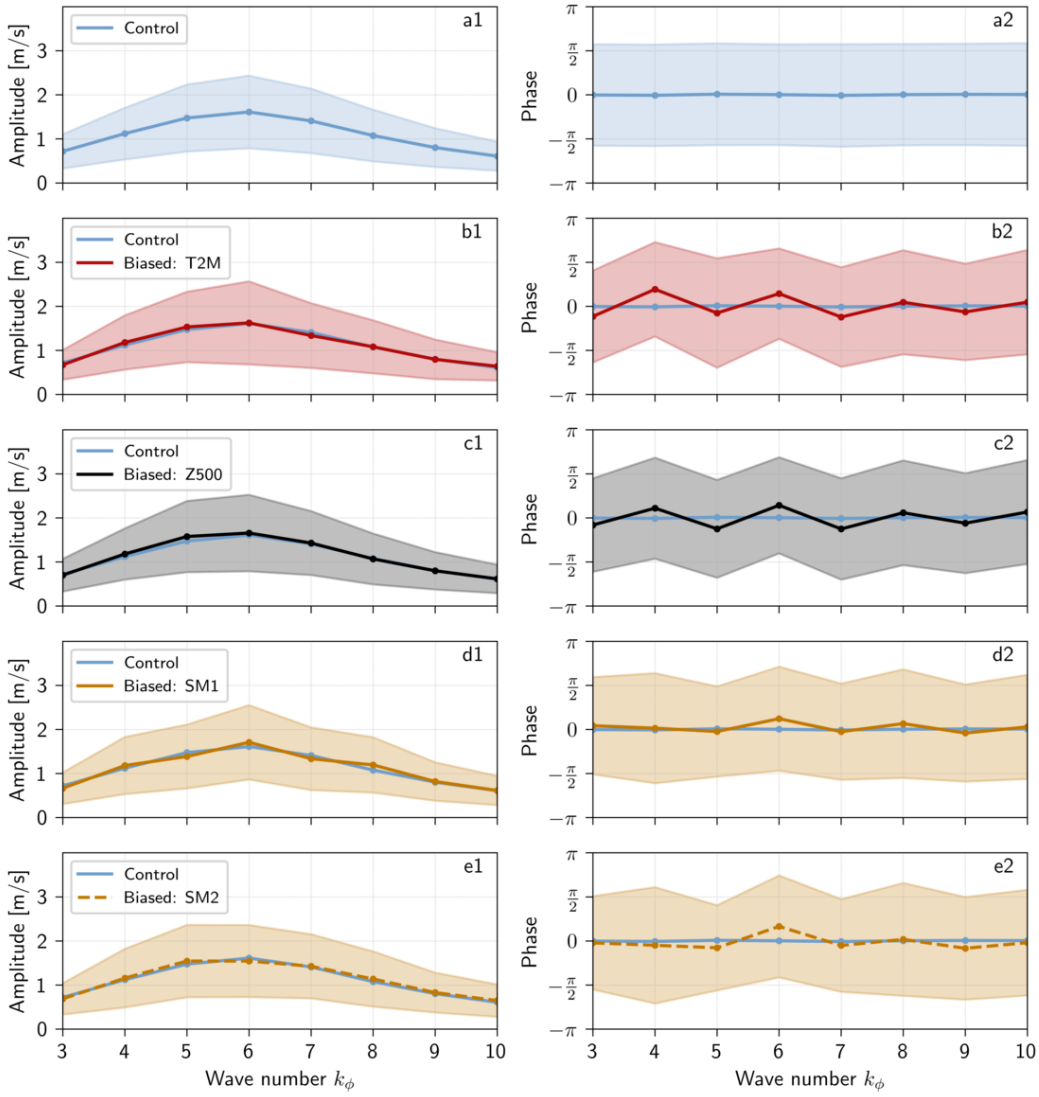


Figure A.8: **Spectrum of meridional wind speed anomalies at 200hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

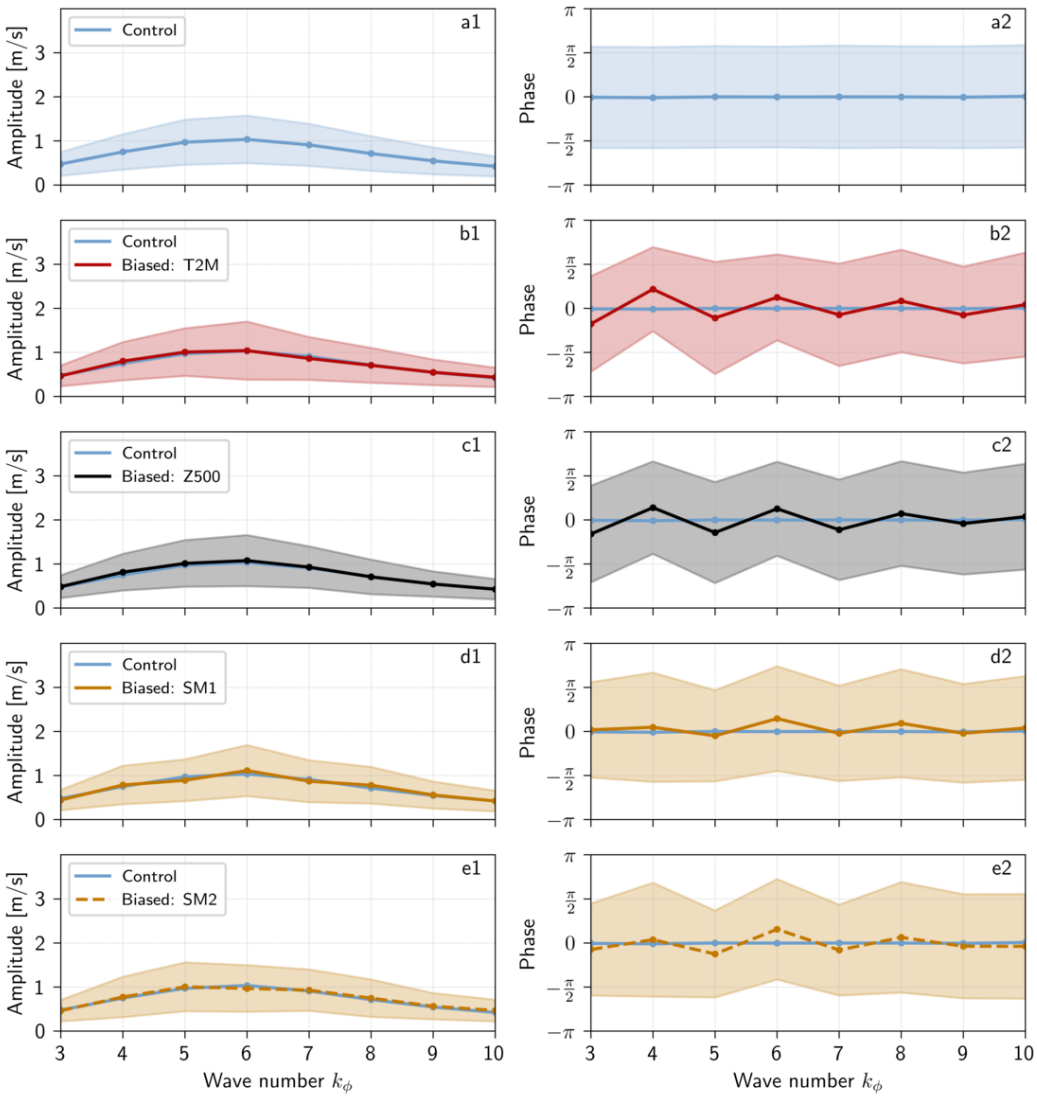


Figure A.9: **Spectrum of meridional wind speed anomalies at 500hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

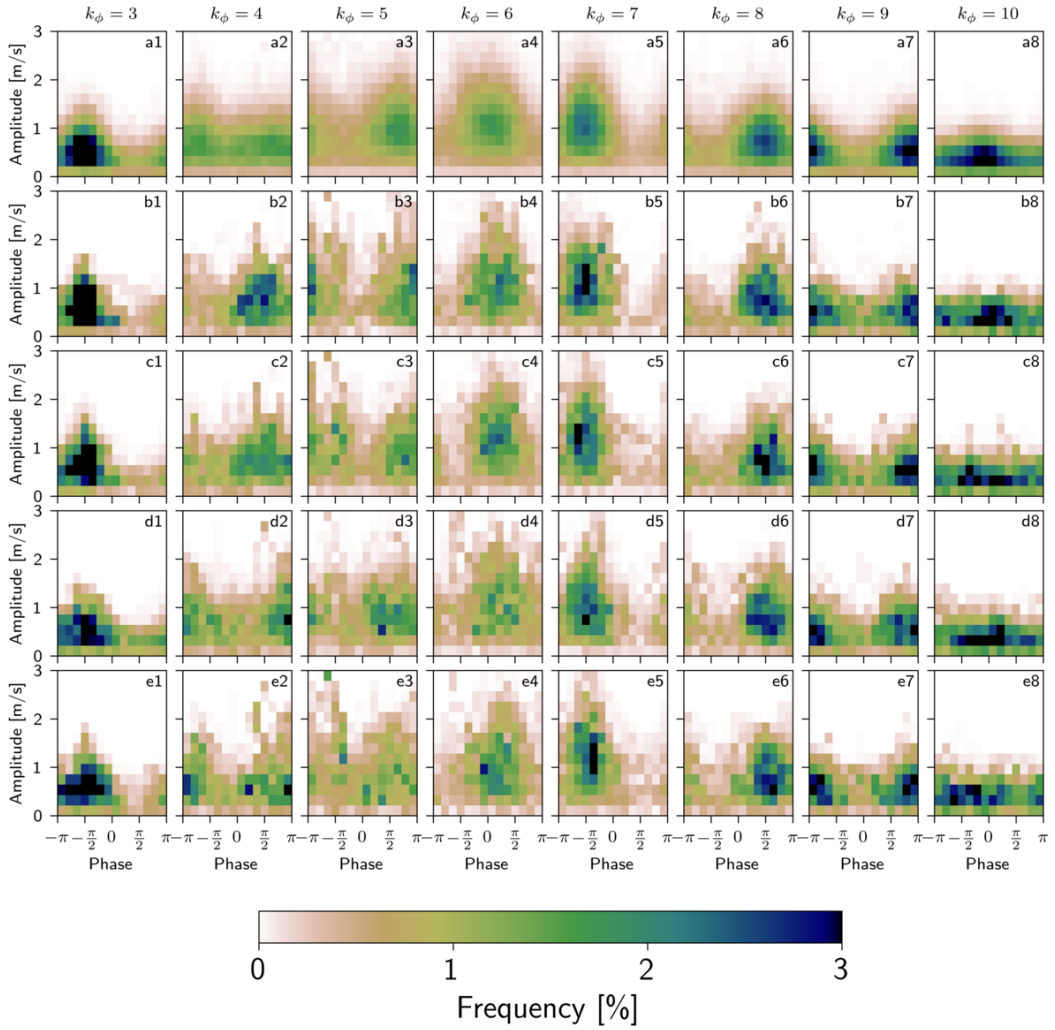


Figure A.10: **Amplitude-phase histograms of meridional wind speed at 500hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) control simulation, (b) biased T2M simulation, (c) biased Z500 simulation, (d) biased SM1 simulation and (e) biased biased SM2 simulation.

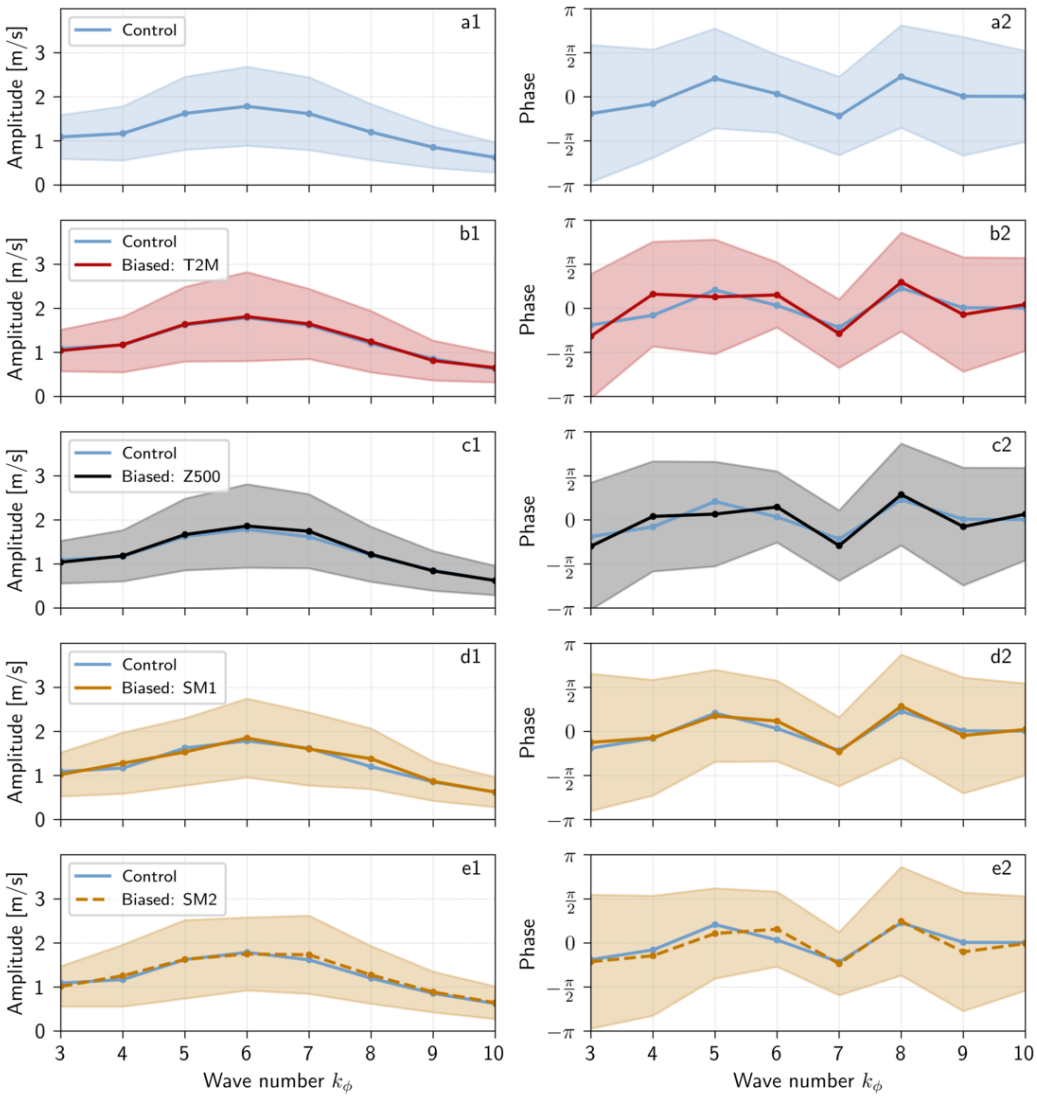


Figure A.11: **Spectrum of meridional wind speed at 200hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

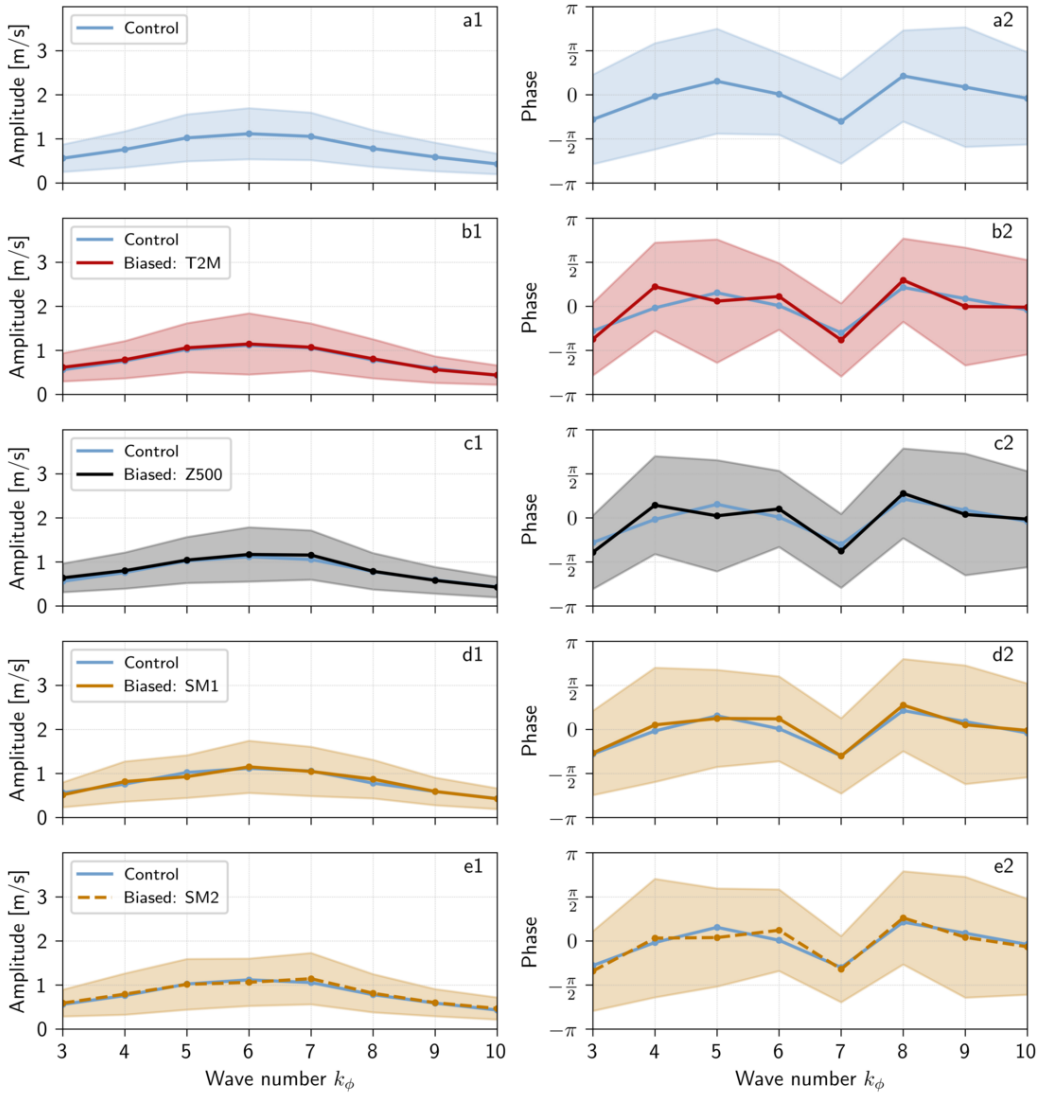


Figure A.12: **Spectrum of meridional wind speed at 500hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

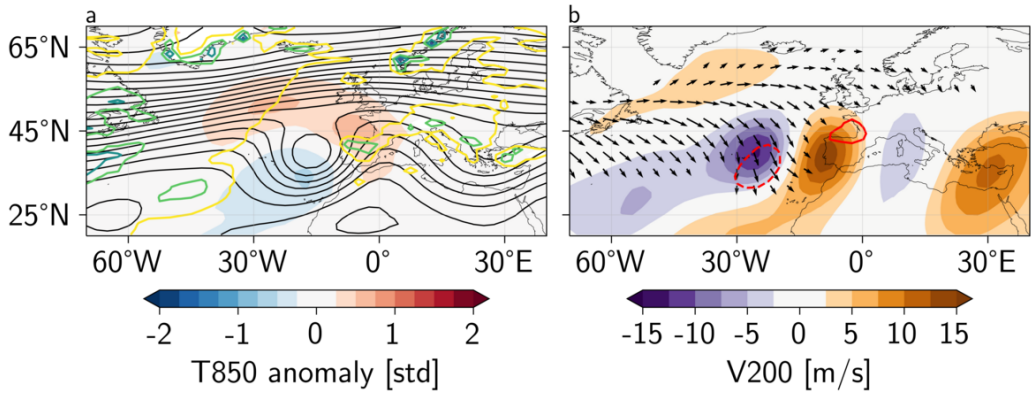


Figure A.13: **Atmospheric dynamics associated to cut-off lows west of the Iberian peninsula.** The composites are made for all events ($n = 1124$) in the control simulation for which a cut-off low is detected in the 33.5°N-42.5°N - 22.5°W-10°W area. (a) Geopotential height at 500hPa (black contours), 3-day average precipitations (colored contours) and 3-day average air temperature at 850hPa anomaly (colors). The geopotential height contours are drawn every 20 meters starting at 5200m. The precipitations contours are drawn every 2mm starting at 2mm. (b) 3-day average 2-m air temperature anomaly (contours), **E**-vector (arrows) and meridional wind speed at 200hPa (colors). The **E**-vector is drawn for norms above 10 m^2/s^2. The anomalies are computed with respect to the daily ensemble mean and standard deviation of the control simulation. For panel (b), the anomalies are drawn every 0.5 standard deviation, starting at +/- 0.5 std. Plain (dashed) lines represent positive (negative) values of the standard deviation.

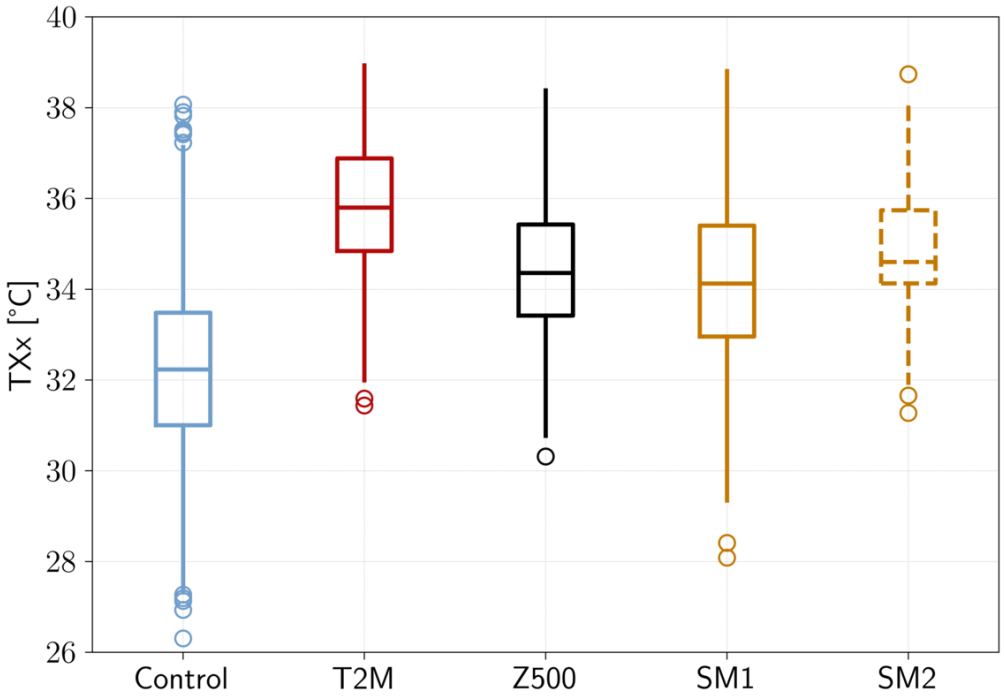


Figure A.14: **Comparison of the distribution of summer maxima of maximum daily temperatures (TXx) at grid point 1.** The boxplots show the 25 and 75% quantiles, the median and the outliers.

A.4 Supplementary materials for chapter 7

A.4.1 Additional figures

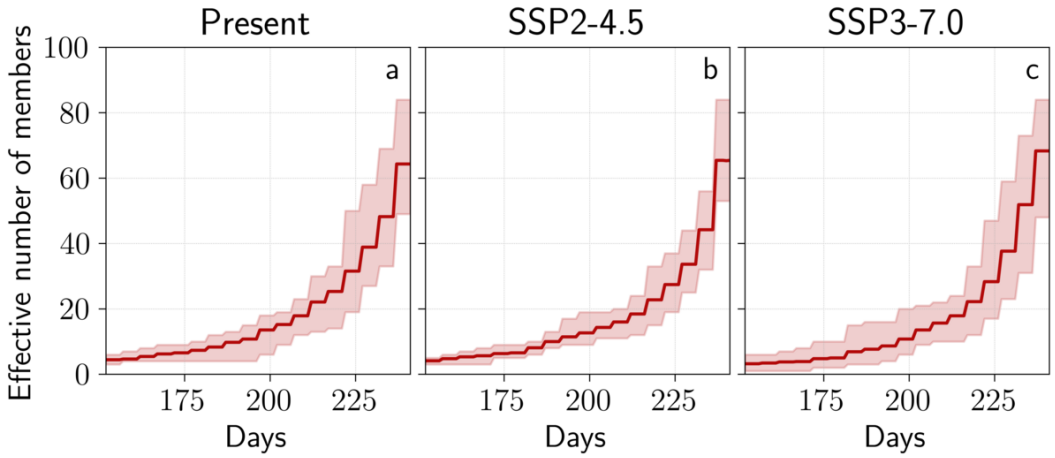


Figure A.15: **Evolution of the number of effectively different members during the simulation.** The effective number of members represents the number of members in each 100-member simulation which are truly unique at time t. The computation is done after the simulation is run and the members are reconstructed from the 1st of June, hence the increasing function. The jumps show the moments when the resamplings occur. The shadings show the minimum and maximum of the effective number of members over the $N_s = 9$ simulations.

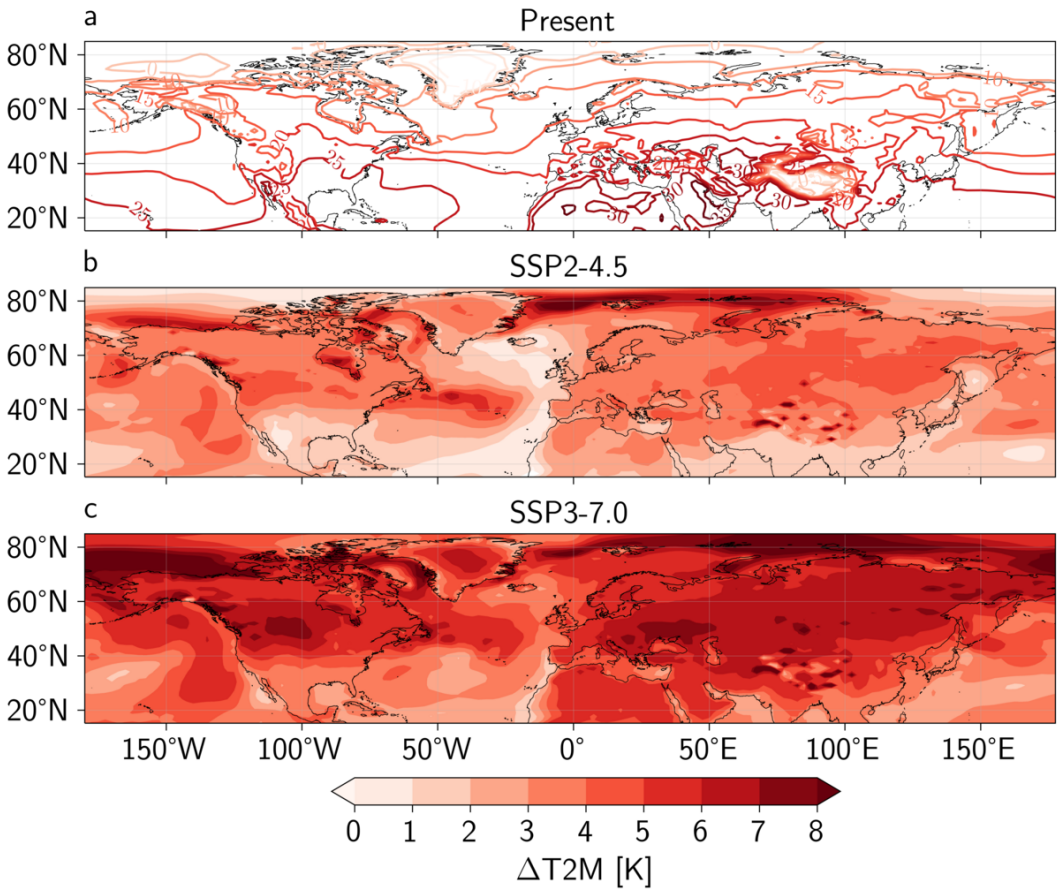


Figure A.16: **Summer averaged (a) 2-m air temperature in the present simulation (°C, contours) and 2-m air temperature anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

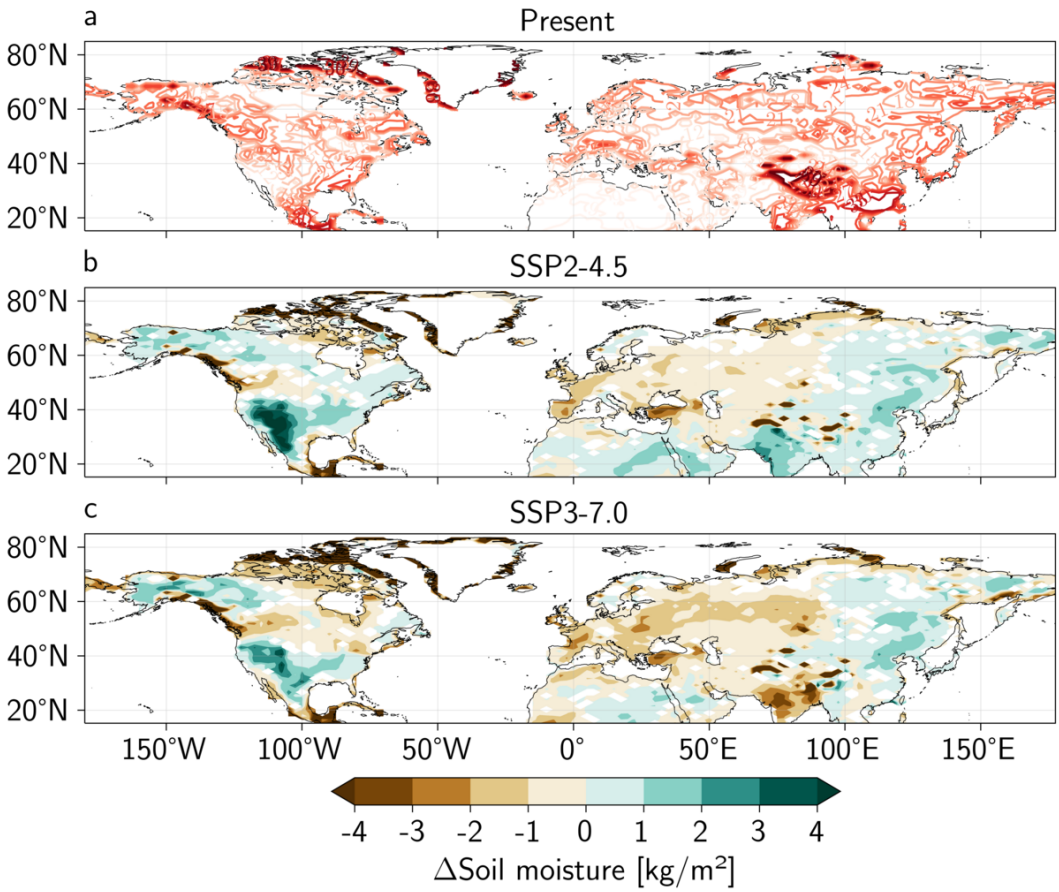


Figure A.17: **Summer averaged (a) upper-level soil moisture in the present simulation (kg/m², contours) and upper-level soil moisture anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

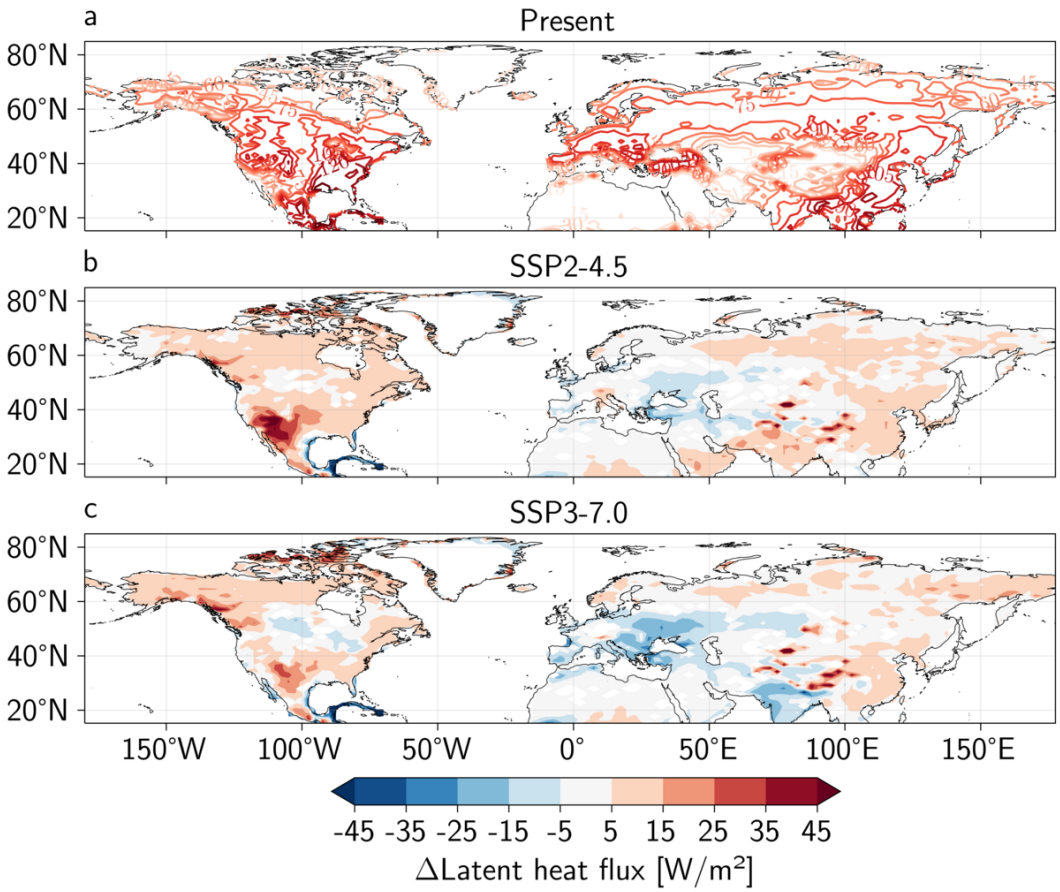


Figure A.18: **Summer averaged (a) latent heat flux at the ground (W/m², contours) and latent heat flux anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

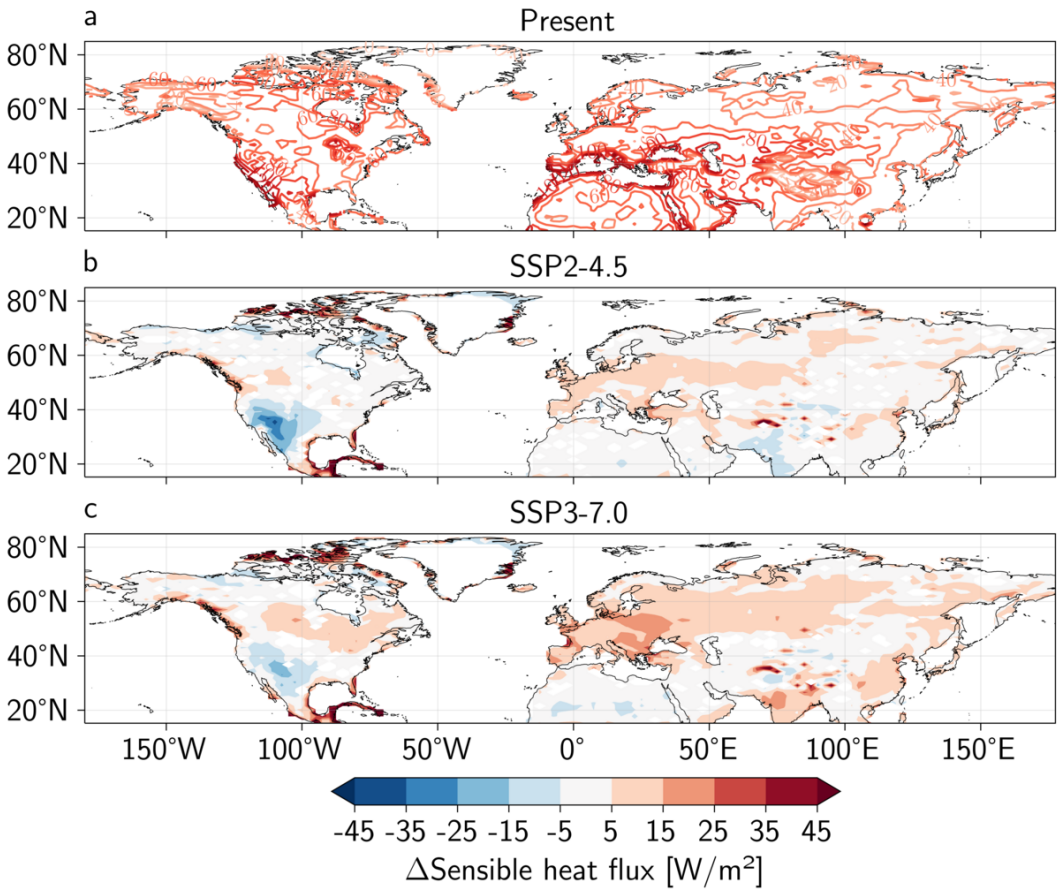


Figure A.19: **Summer averaged (a) sensible heat flux at the ground (W/m², contours) and sensible heat flux anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

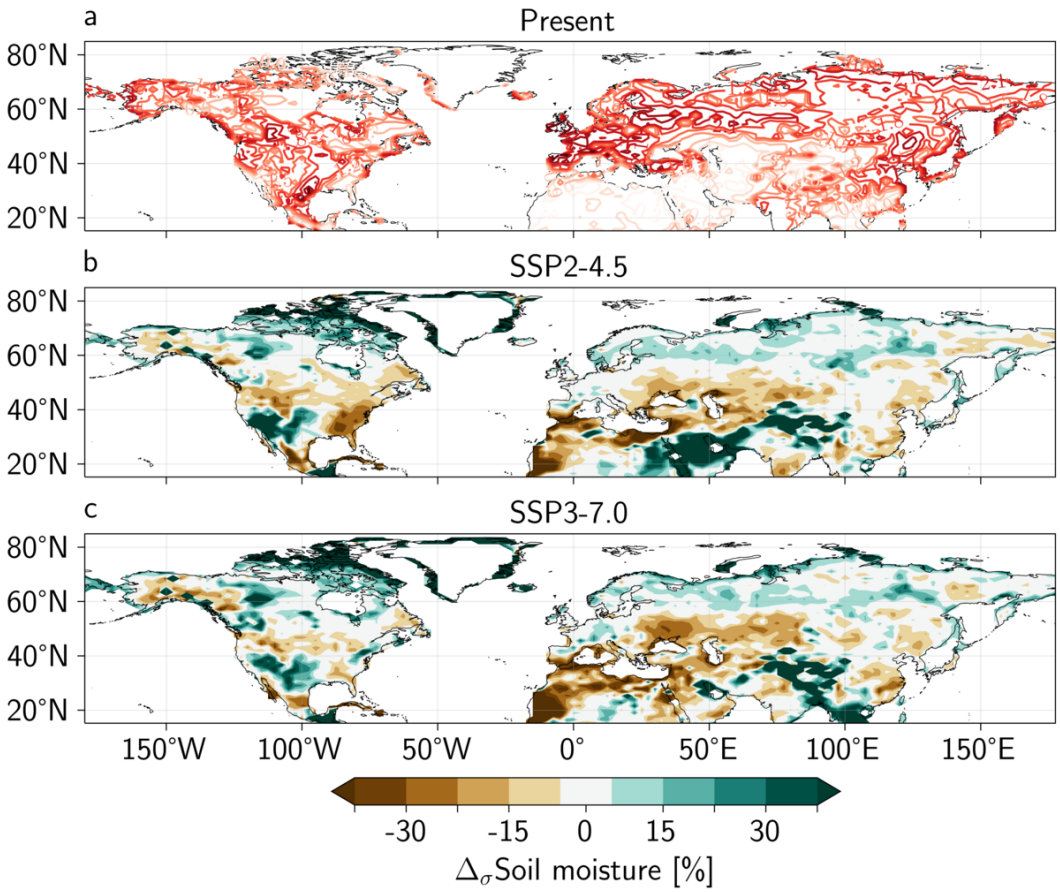


Figure A.20: **Standard deviation of the summer averaged upper-level soil moisture in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

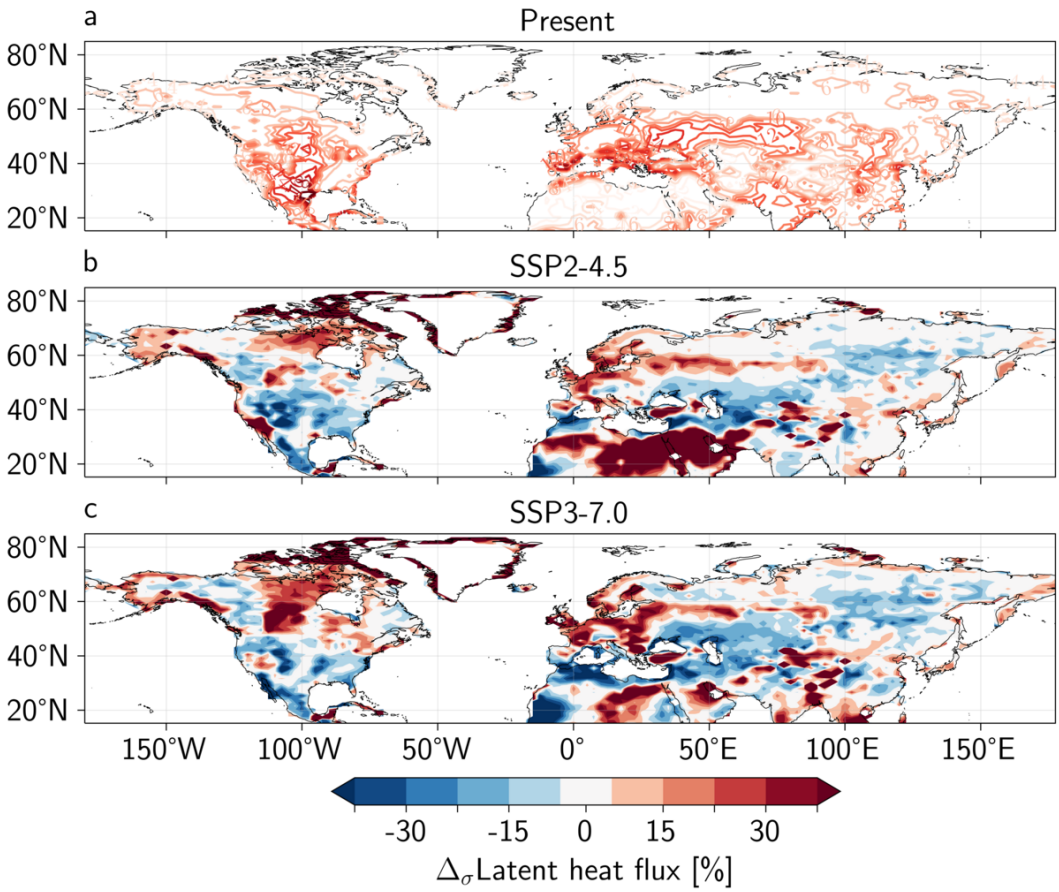


Figure A.21: **Standard deviation of the summer averaged latent heat flux in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

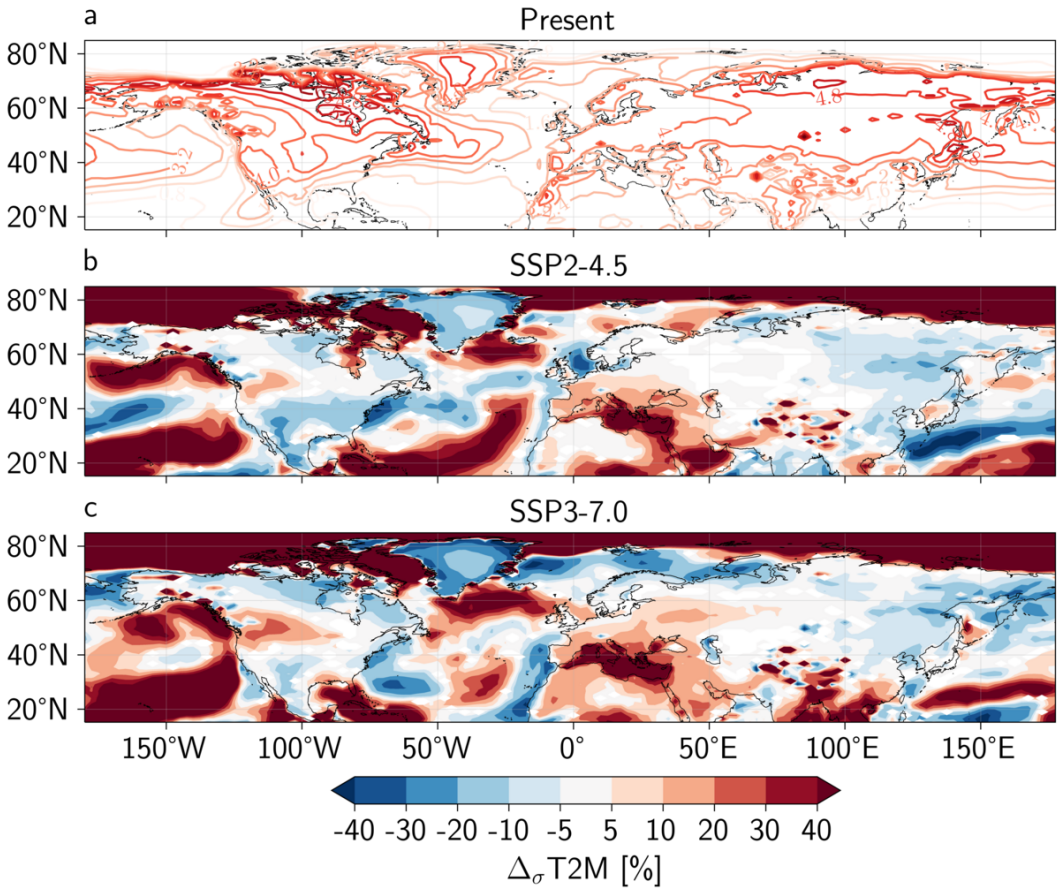


Figure A.22: **Standard deviation of the summer distribution of daily 2-m air temperature in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

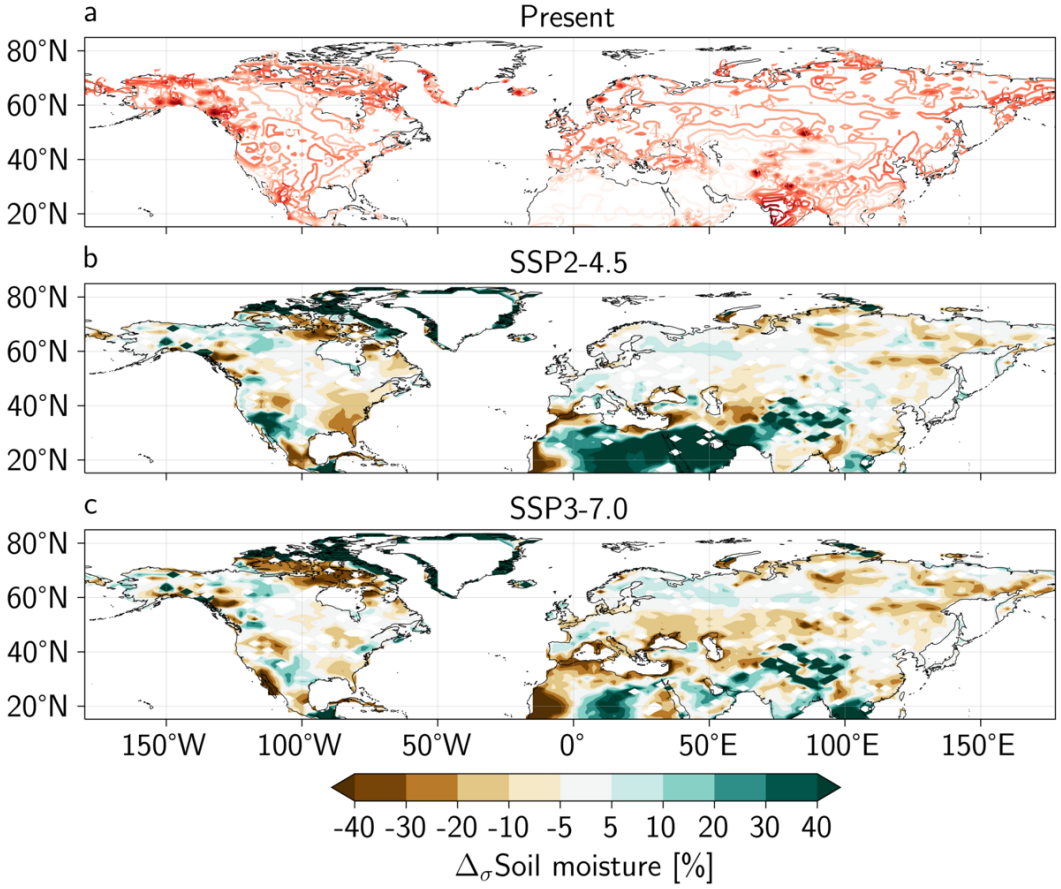


Figure A.23: **Standard deviation of the summer distribution of daily upper-level soil moisture in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

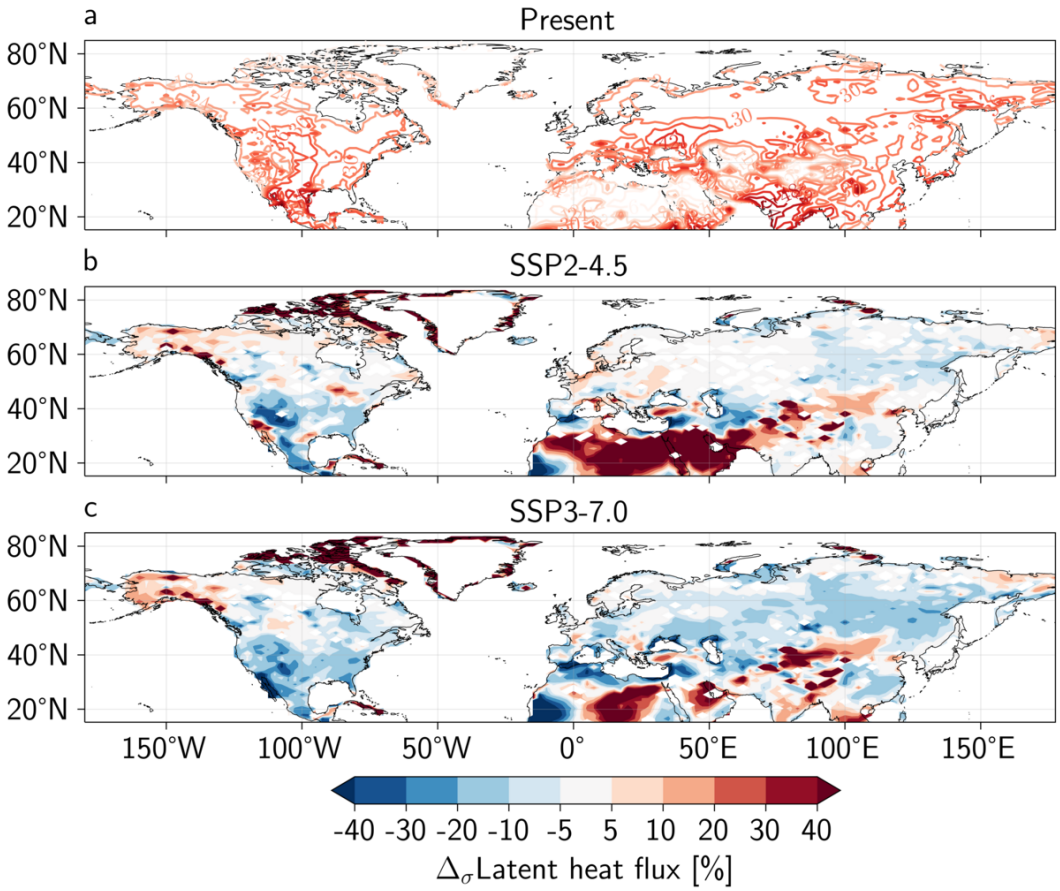


Figure A.24: **Standard deviation of the summer distribution of daily latent heat flux in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

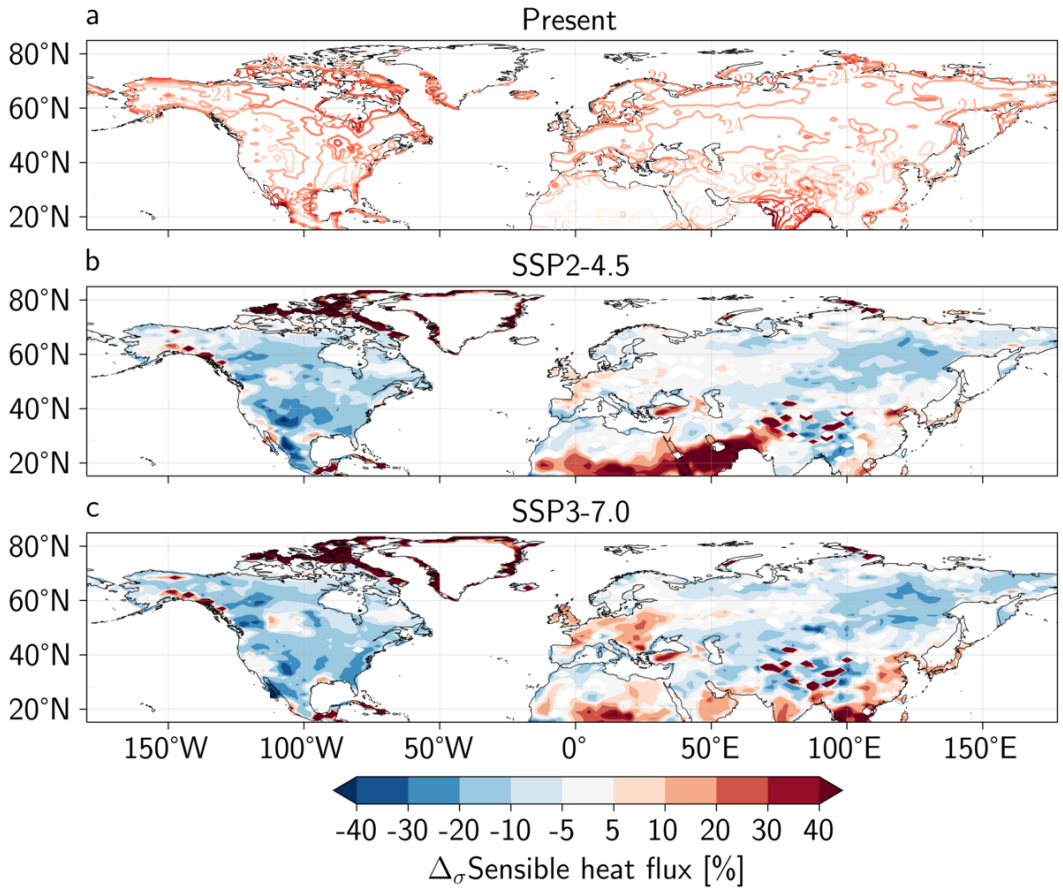


Figure A.25: **Standard deviation of the summer distribution of daily sensible heat flux in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

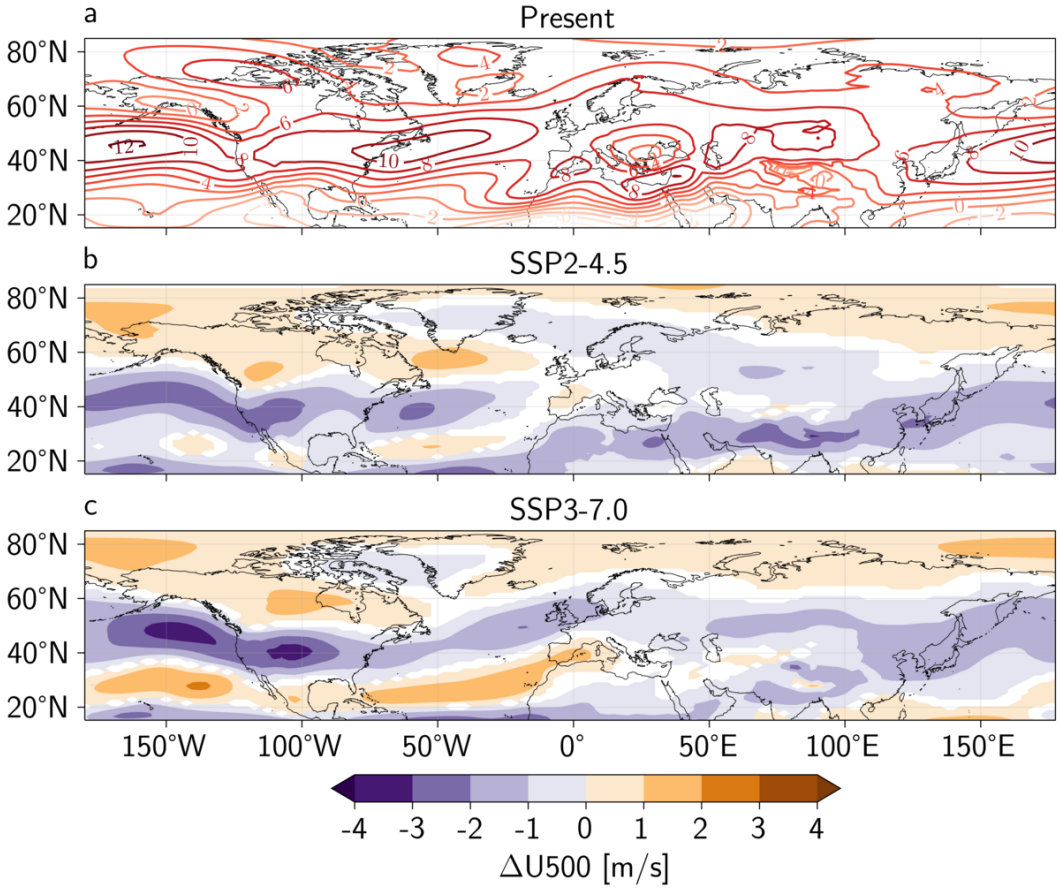


Figure A.26: **Summer averaged (a) 500hPa zonal wind in the present simulation (m/s, contours) and 500hPa zonal wind anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

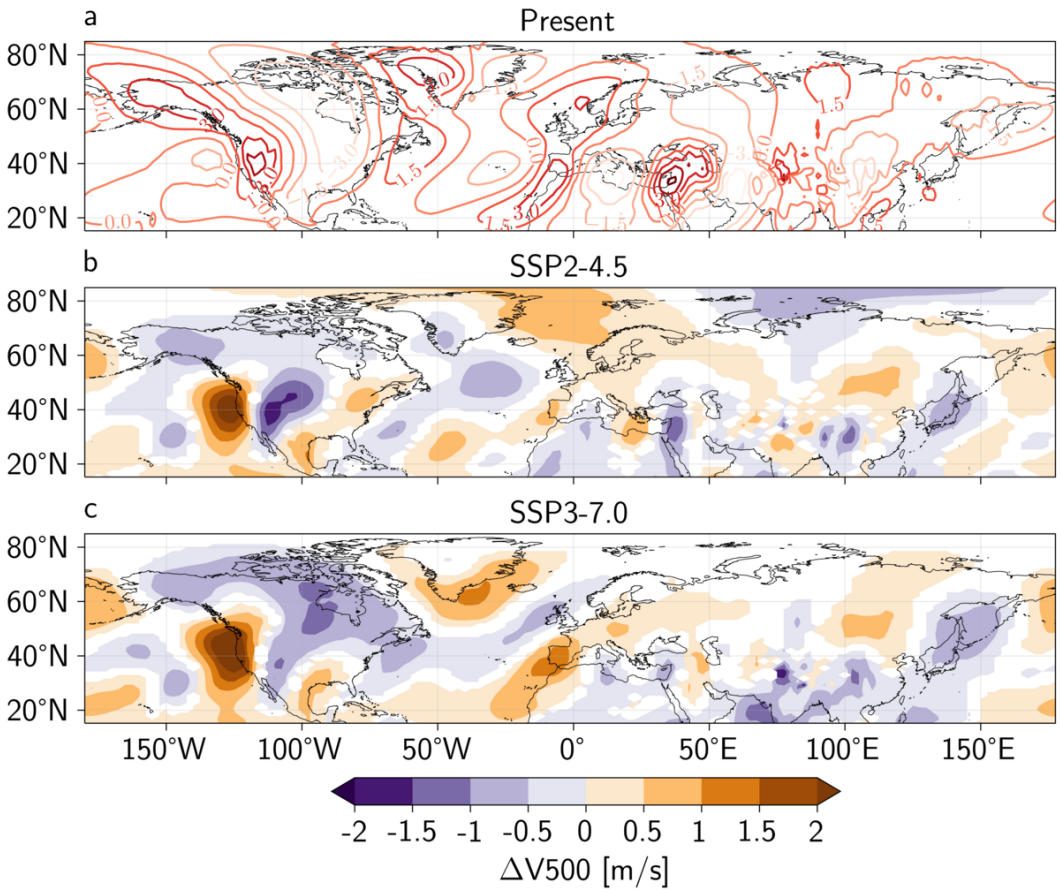


Figure A.27: **Summer averaged (a) 500hPa meridional wind in the present simulation (m/s, contours) and 500hPa meridional wind anomaly in in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

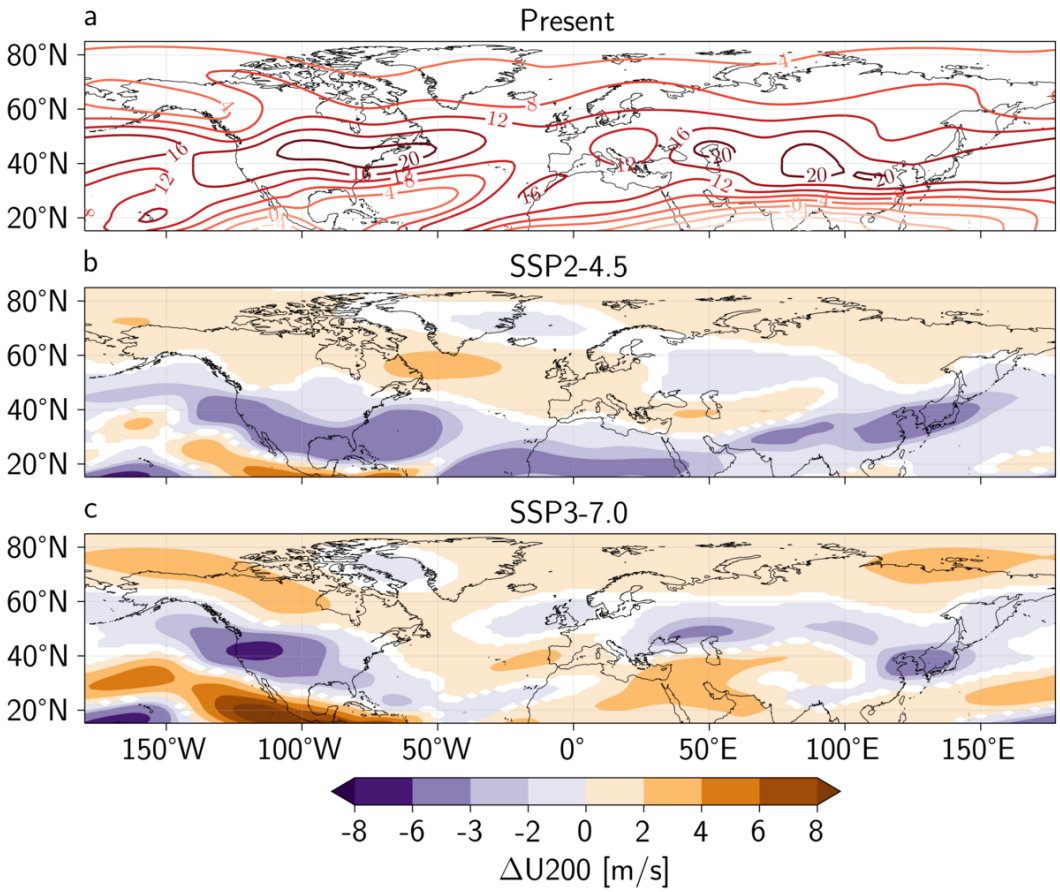


Figure A.28: **Summer averaged (a) 200hPa zonal wind in the present simulation (m/s, contours) and 200hPa zonal wind anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

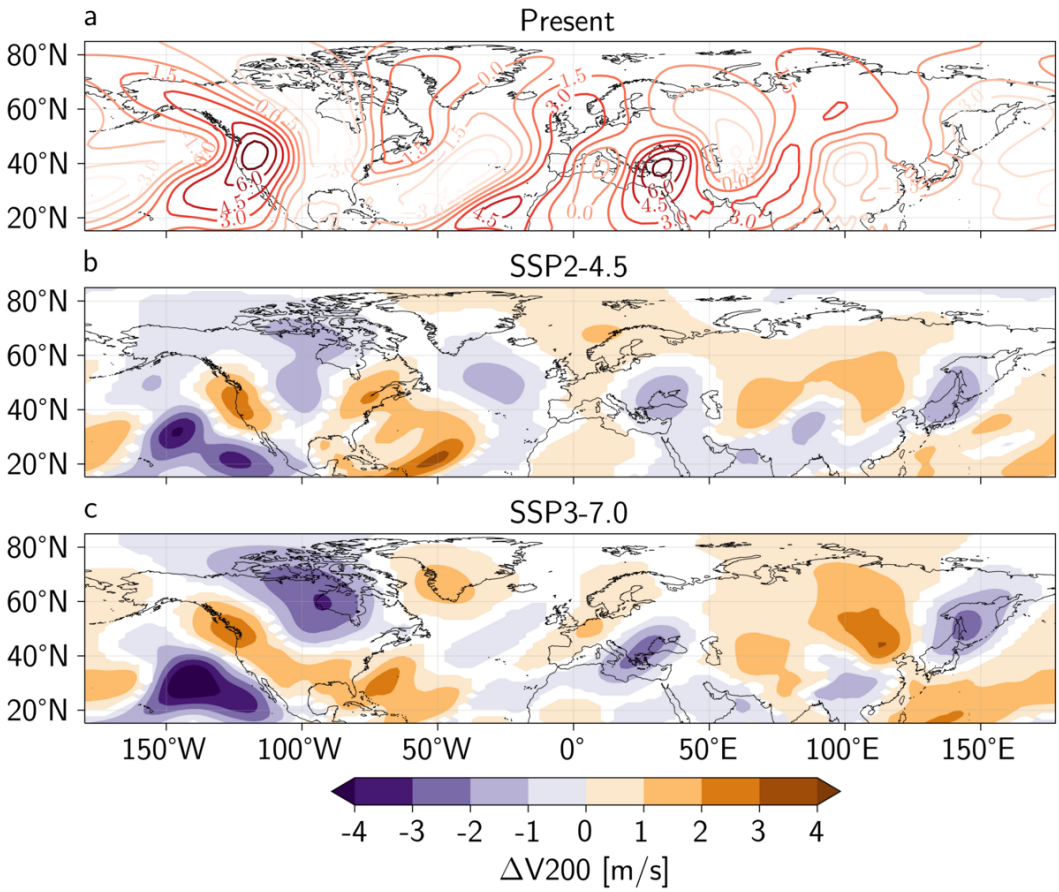


Figure A.29: **Summer averaged (a) 200hPa meridional wind in the present simulation (m/s, contours) and 200hPa meridional wind anomaly in in (b) SSP2-4.5 and (c) SSP3-7.0 (°C, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 0.25 std in absolute value are colored in white.

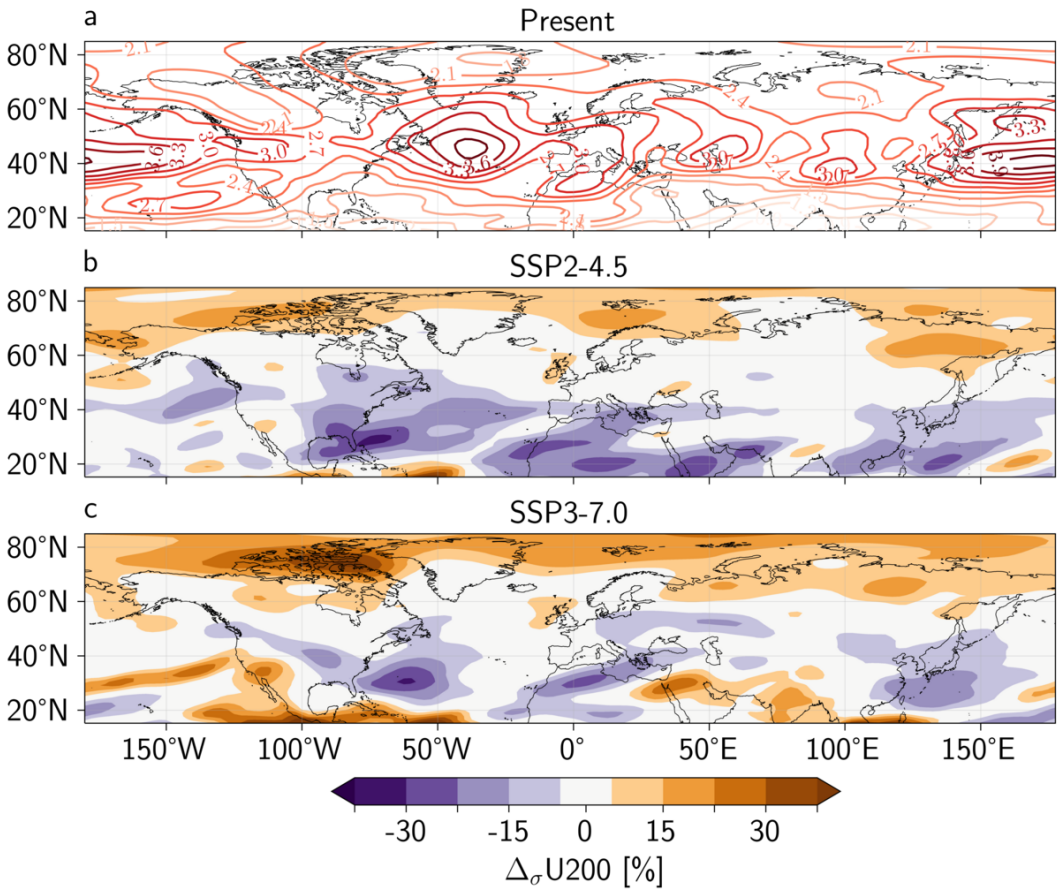


Figure A.30: **Standard deviation of the summer averaged 500hPa zonal wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

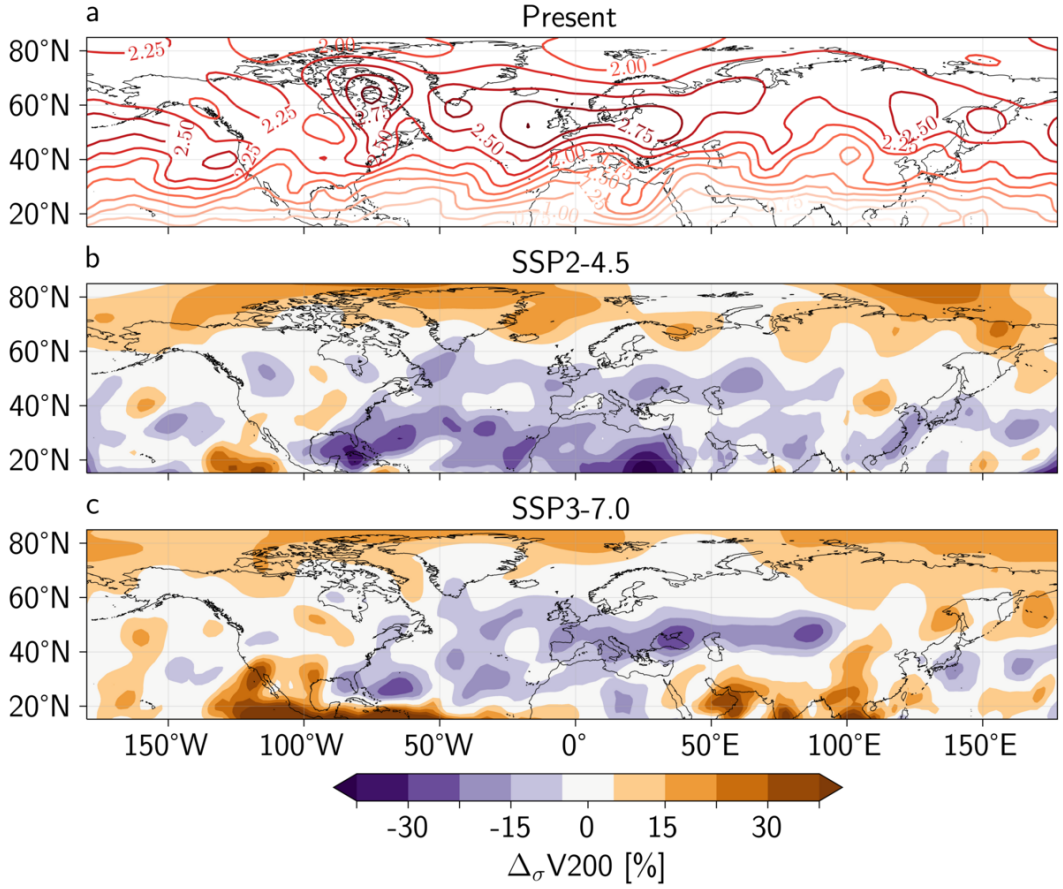


Figure A.31: **Standard deviation of the summer averaged 500hPa meridional wind in (a) the present simulation (°C, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 15% in absolute value are colored in white.

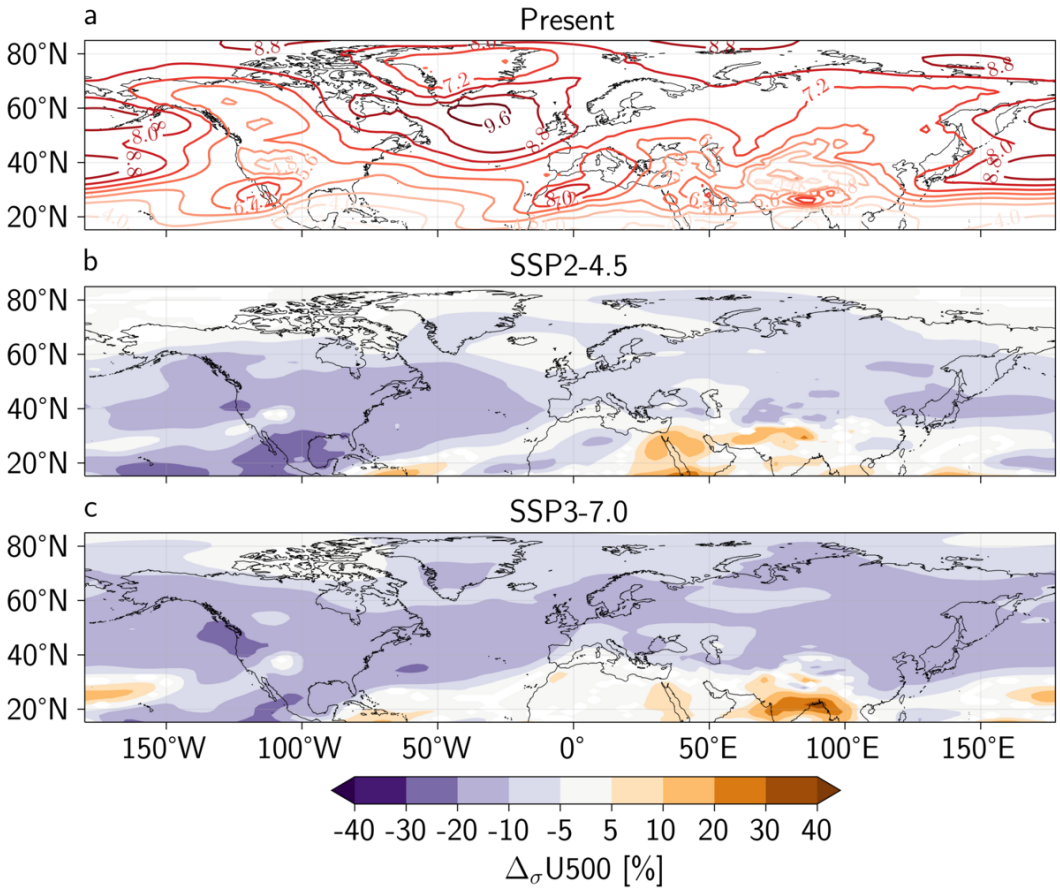


Figure A.32: **Standard deviation of the summer distribution of daily 500hPa zonal wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

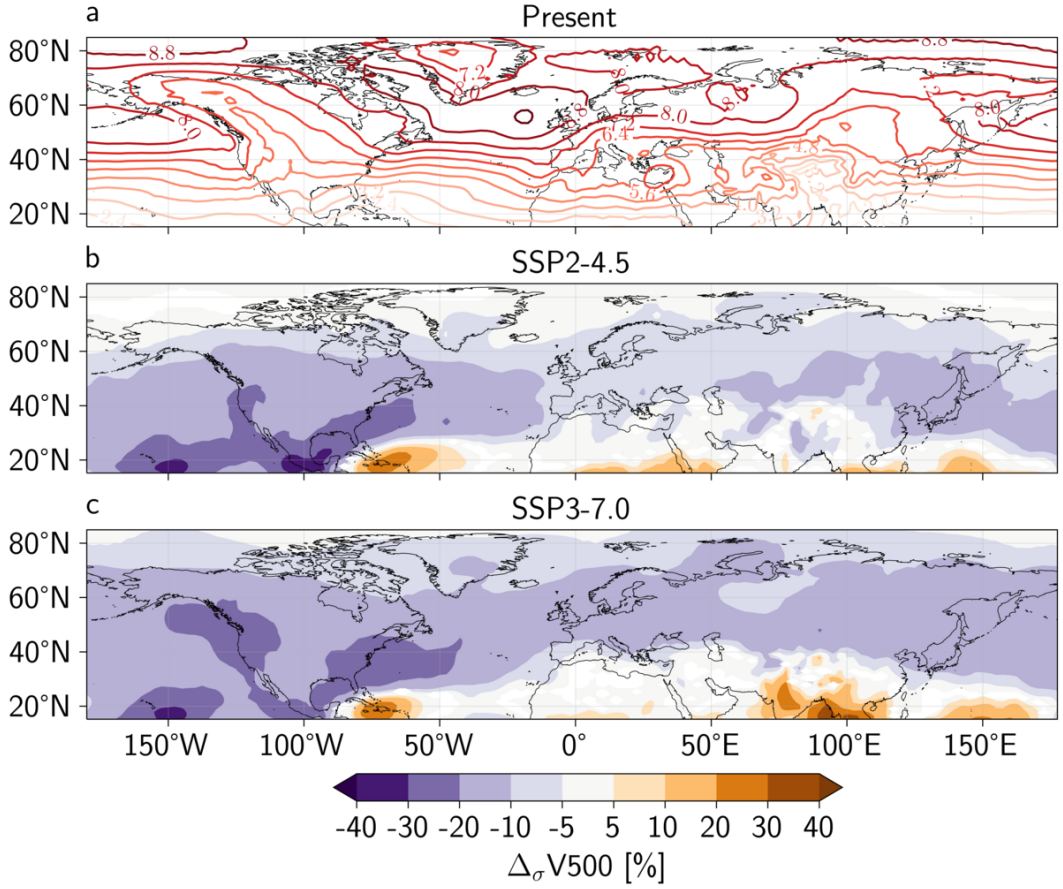


Figure A.33: **Standard deviation of the summer distribution of daily 500hPa meridional wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

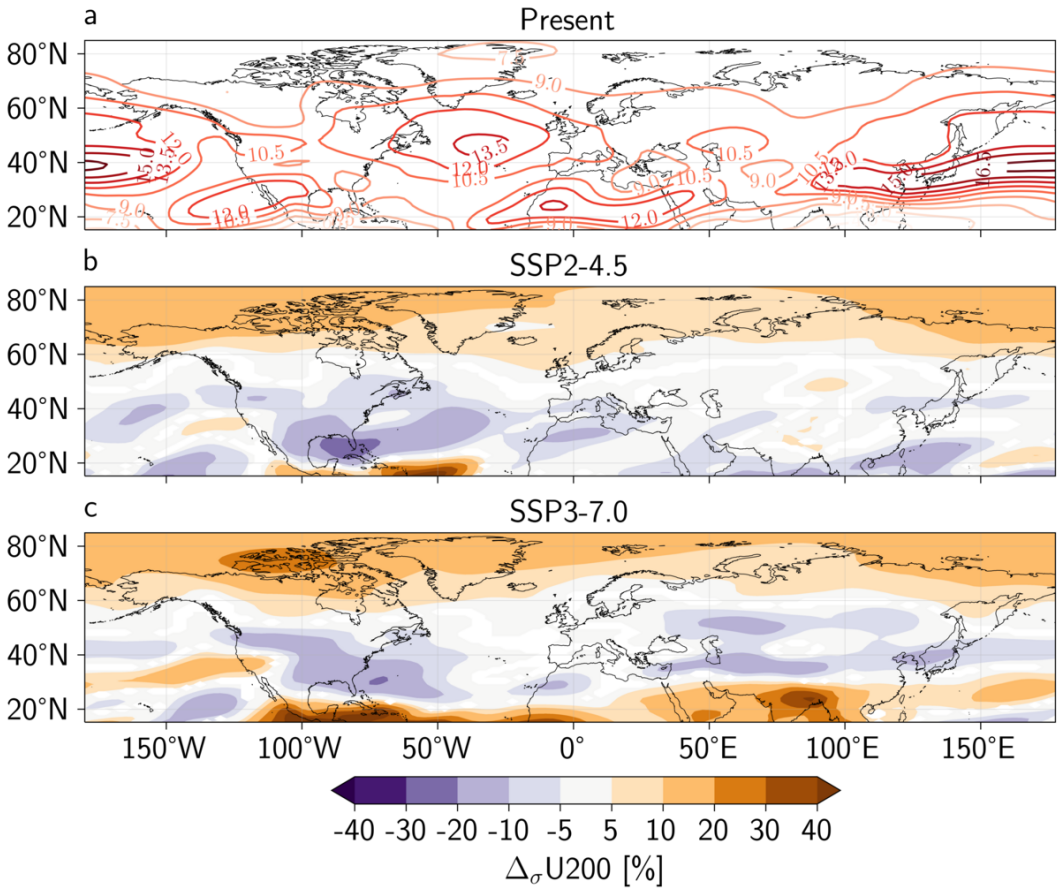


Figure A.34: **Standard deviation of the summer distribution of daily 500hPa zonal wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

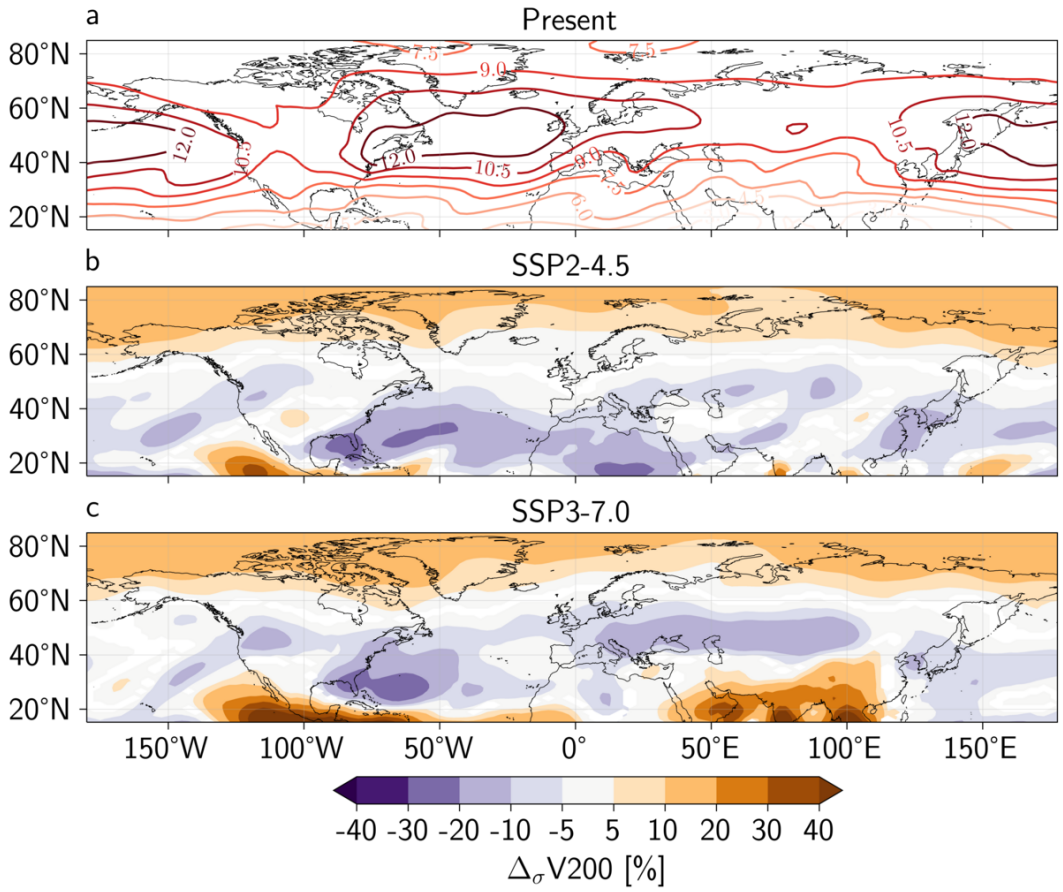


Figure A.35: **Standard deviation of the summer distribution of daily 200hPa meridional wind in (a) the present simulation (m/s, contours) and standard deviation anomaly in (b) SSP2-4.5 and (c) SSP3-7.0 (%, colors).** For panels b and c, the anomalies are computed with respect to the present simulations. The grid point where the normalized anomaly with respect to the present simulation is below 5% in absolute value are colored in white.

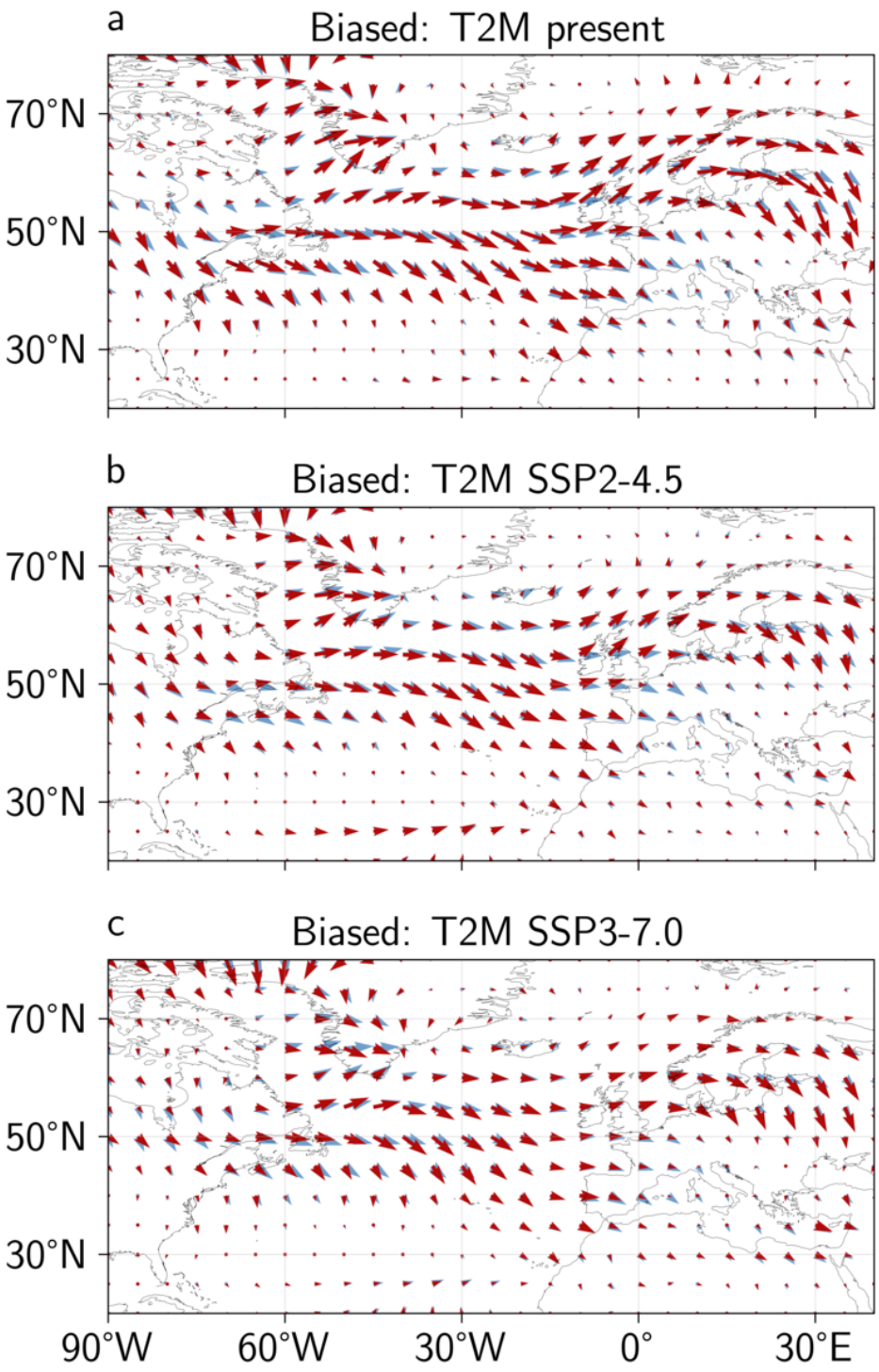


Figure A.36: **Summer averaged E-vector at 500hPa for centennial events.** In each panel the blue arrows show the climatological **E**-vector computed on the control simulation.

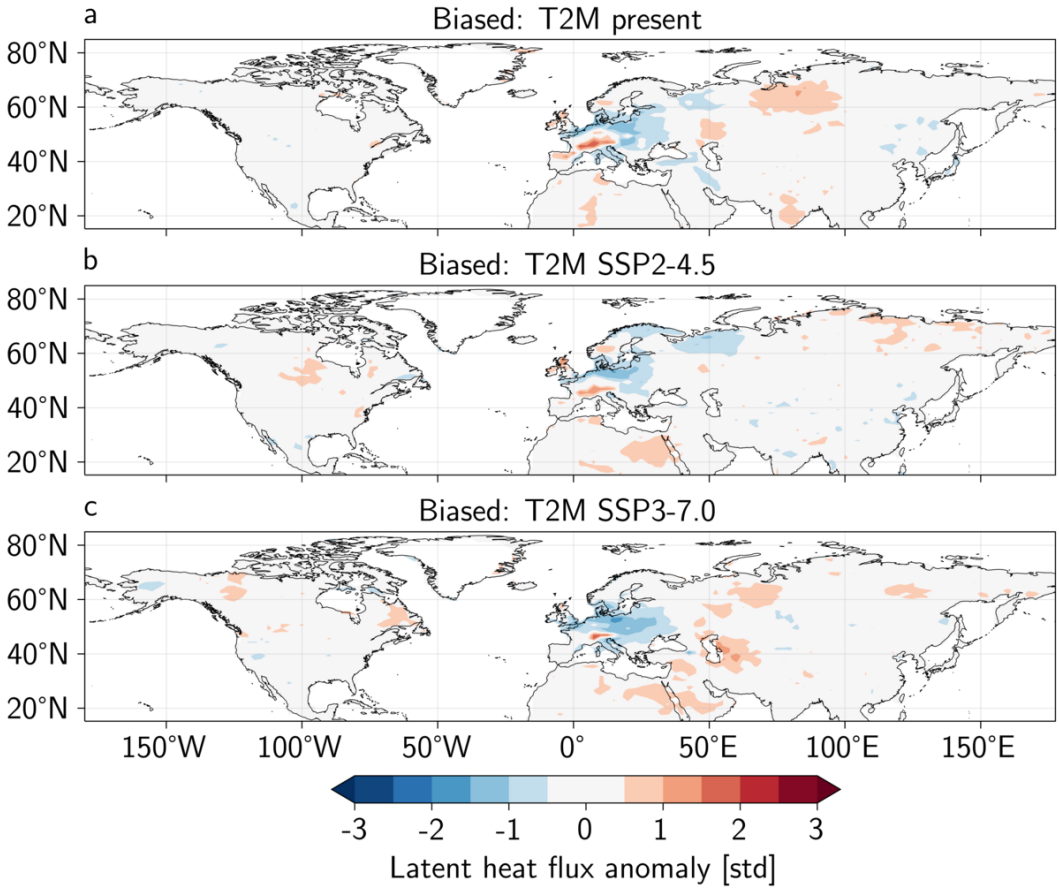


Figure A.37: **Summer averaged normalized anomalies of latent heat flux at the ground for centennial events.** Normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation.

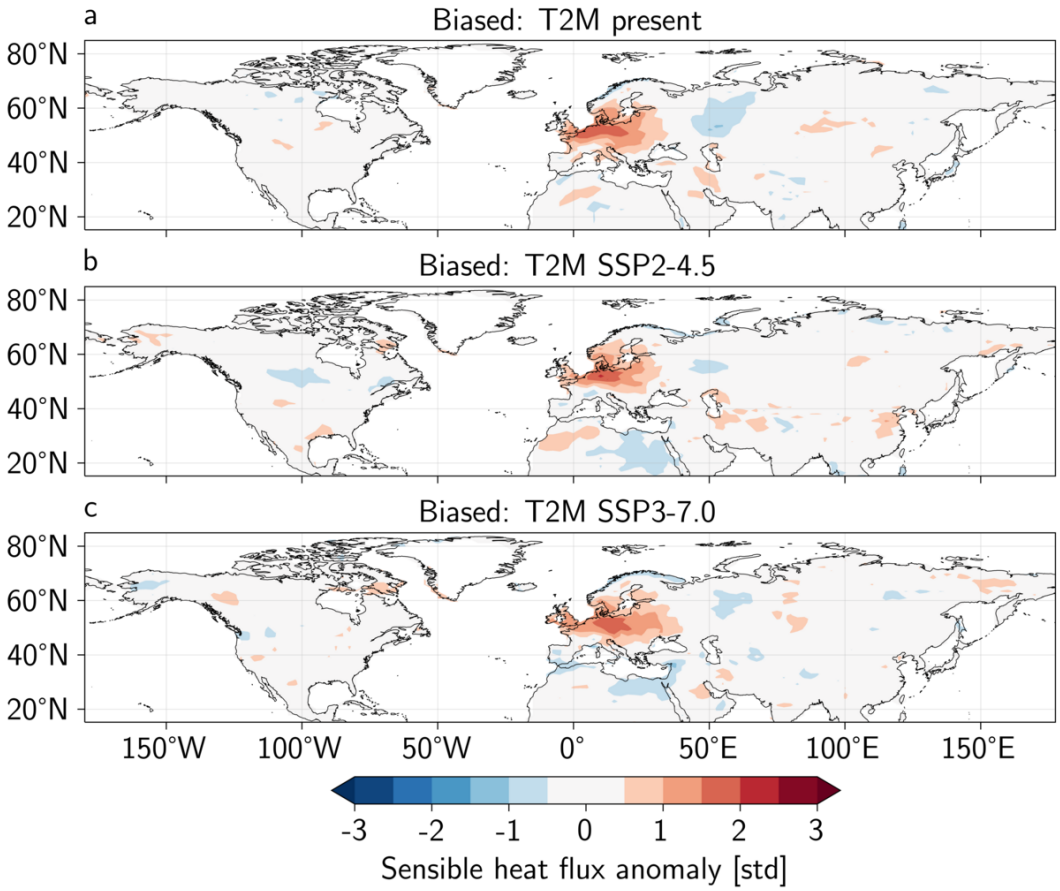


Figure A.38: **Summer averaged normalized anomalies of sensible heat flux at the ground for centennial events.** Normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation.

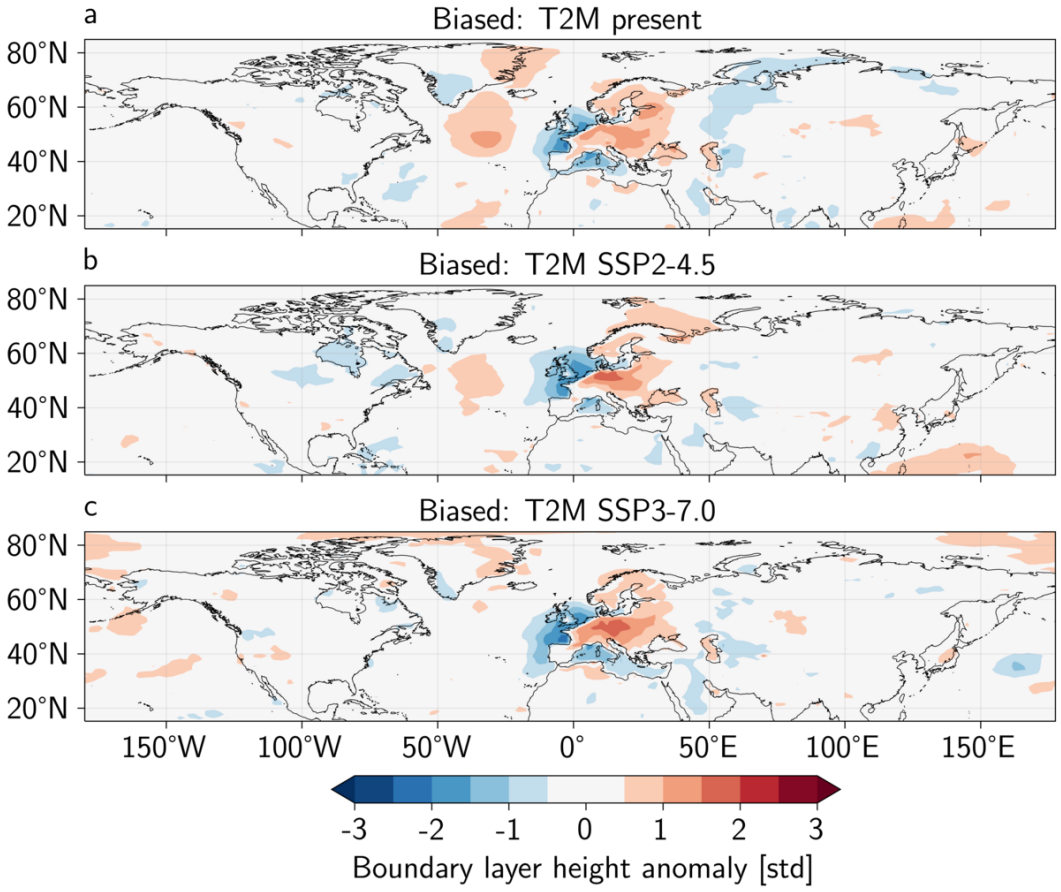


Figure A.39: **Summer averaged normalized anomalies of boundary layer height for centennial events.** Normalized anomalies are computed by removing at each grid point the mean and dividing by the standard deviation computed on the control simulation.

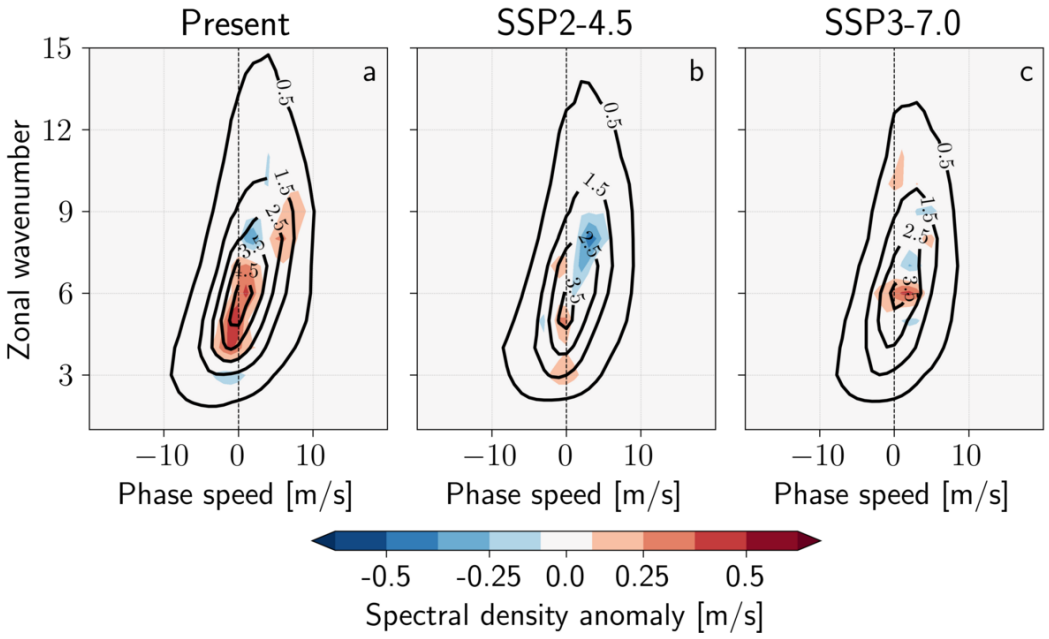


Figure A.40: **Wavenumber-phase speed spectra of meridional wind anomalies at 500hPa for centennial events.** Raw spectrum (contours) and difference with the climatological spectrum (colors).

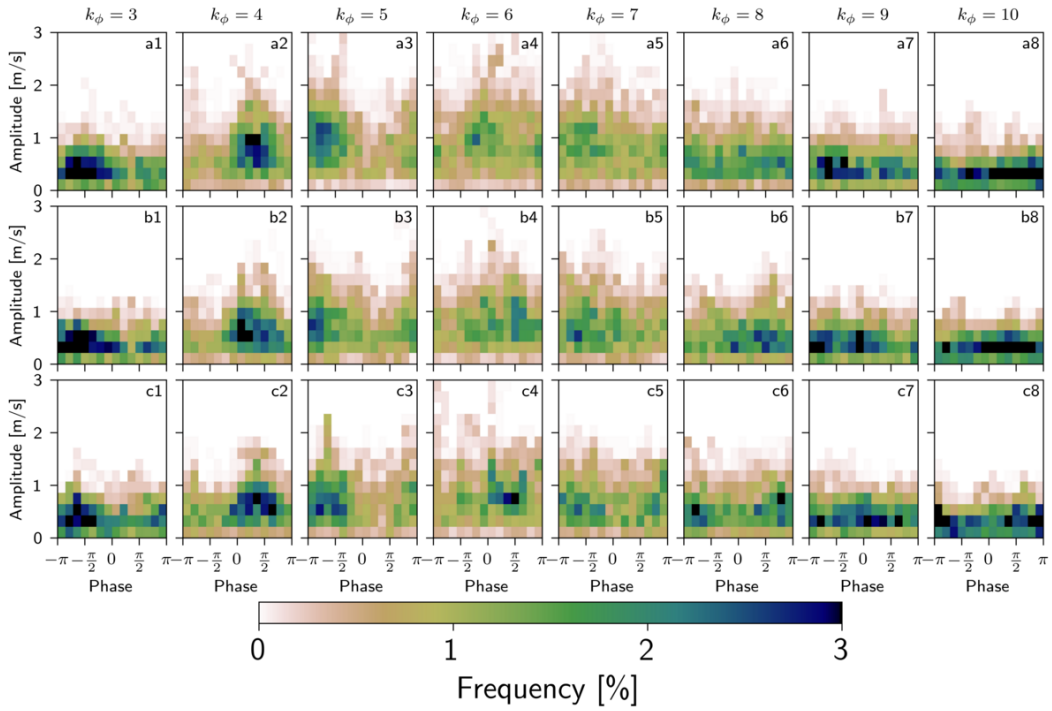


Figure A.41: **Amplitude-phase histograms of meridional wind speed anomalies at 500hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

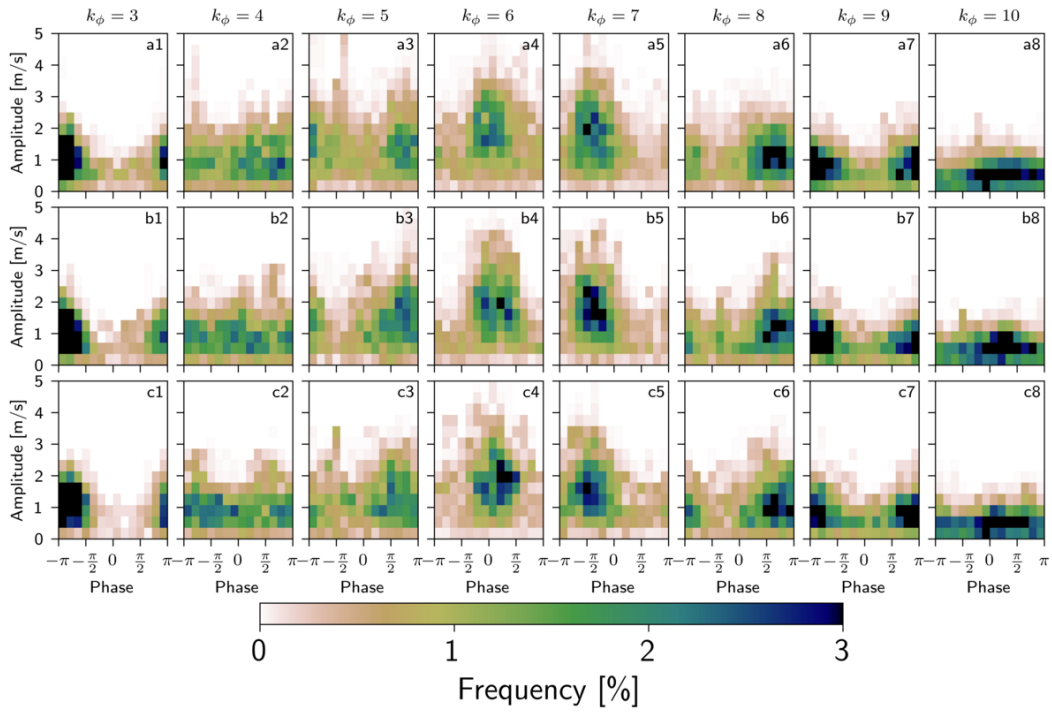


Figure A.42: **Amplitude-phase histograms of meridional wind speed at 200hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

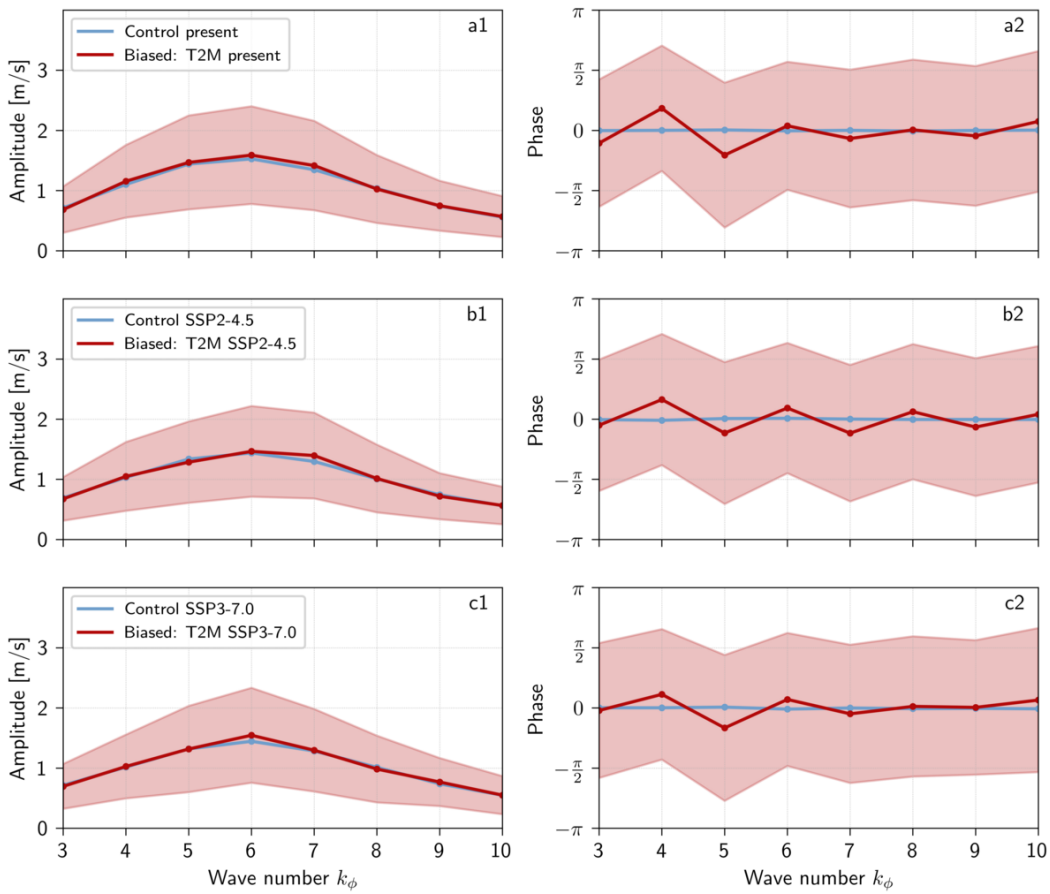


Figure A.43: **Spectrum of meridional wind speed anomalies at 200hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

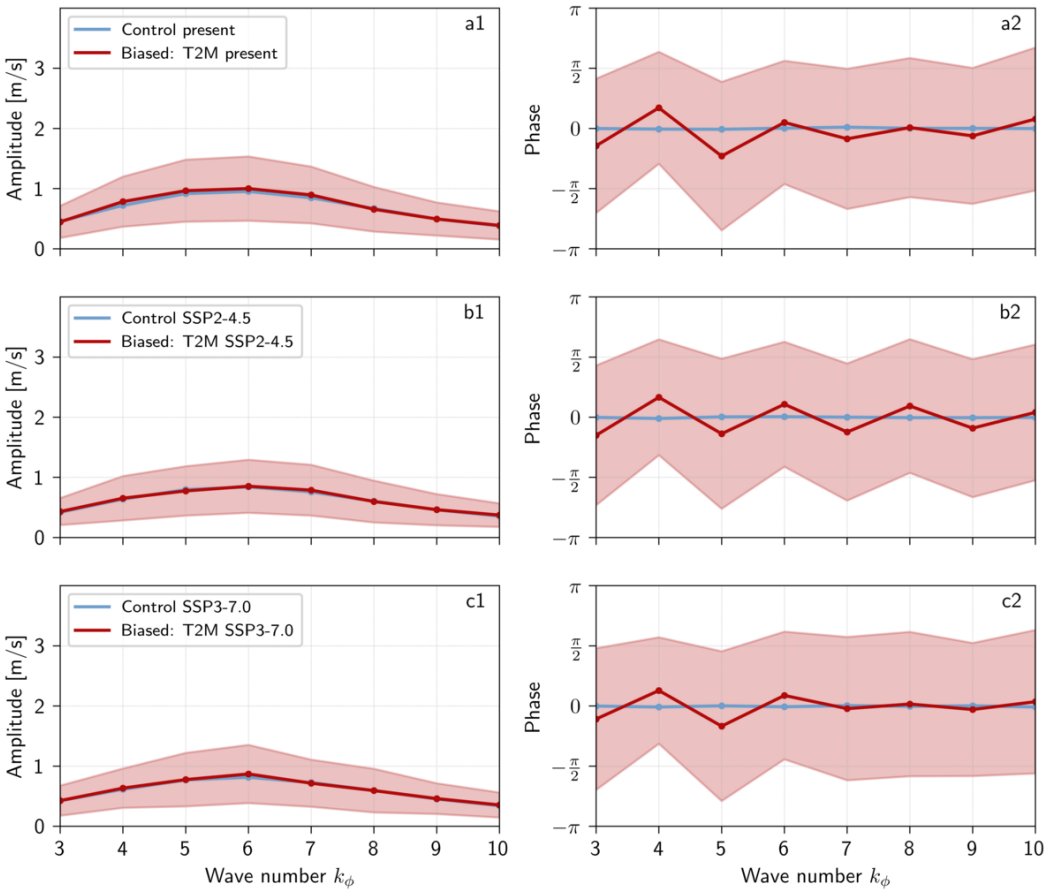


Figure A.44: **Spectrum of meridional wind speed anomalies at 500hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

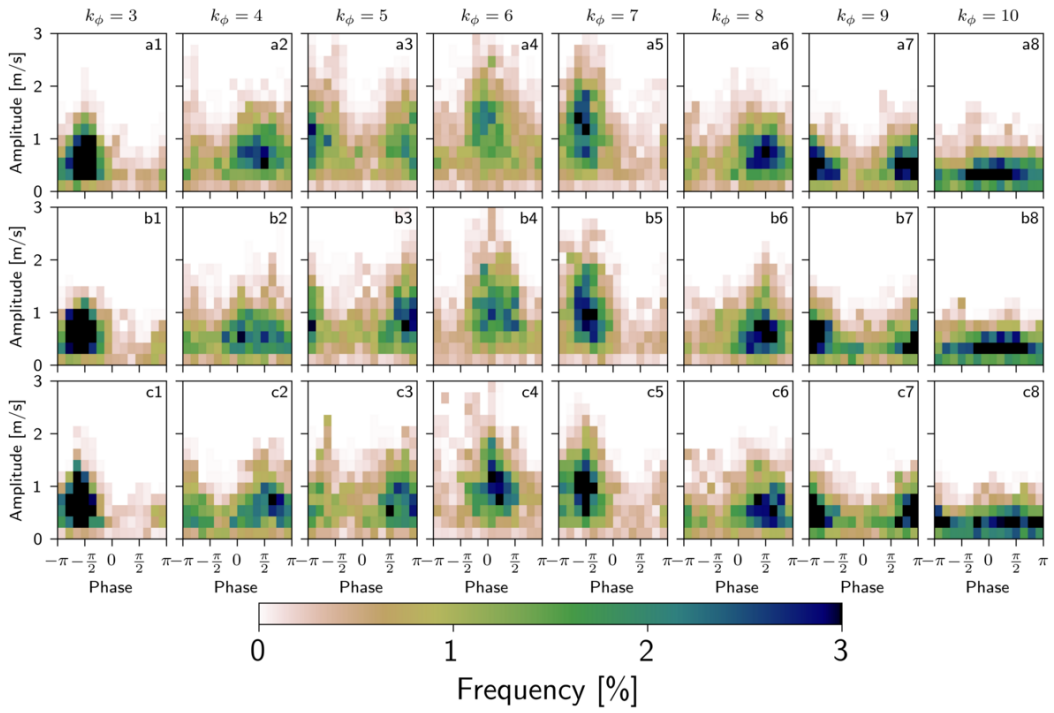


Figure A.45: **Amplitude-phase histograms of meridional wind speed at 500hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

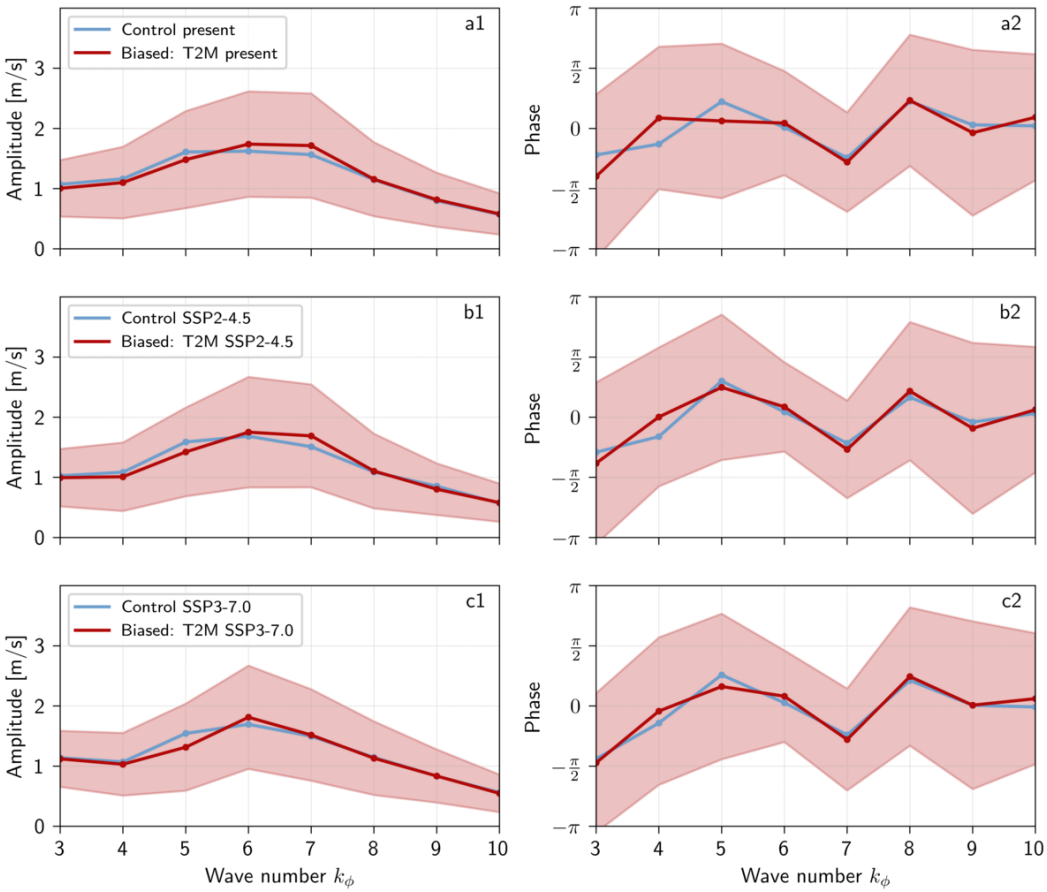


Figure A.46: **Spectrum of meridional wind speed at 200hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

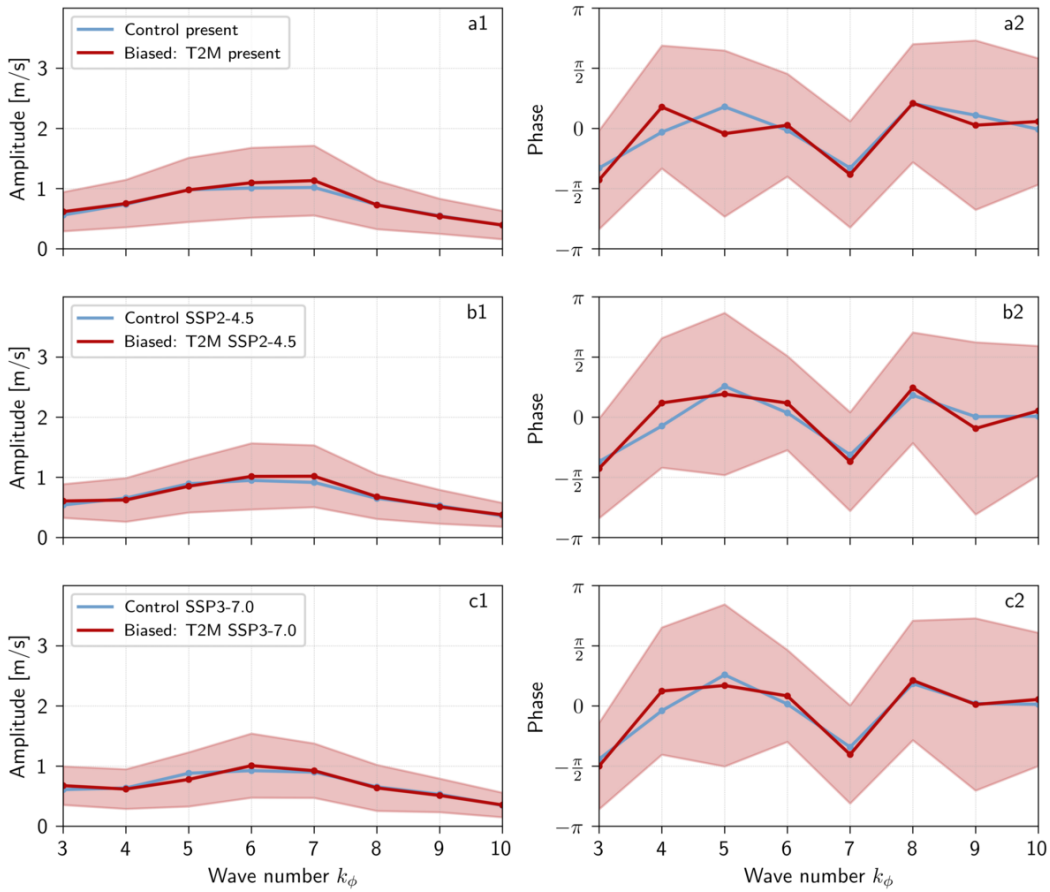


Figure A.47: **Spectrum of meridional wind speed at 500hPa for centennial events.** The first column shows the average amplitude per zonal wave number. The second column shows the average phase per zonal wave number. The shadings show a +/- 1 std range around the average.

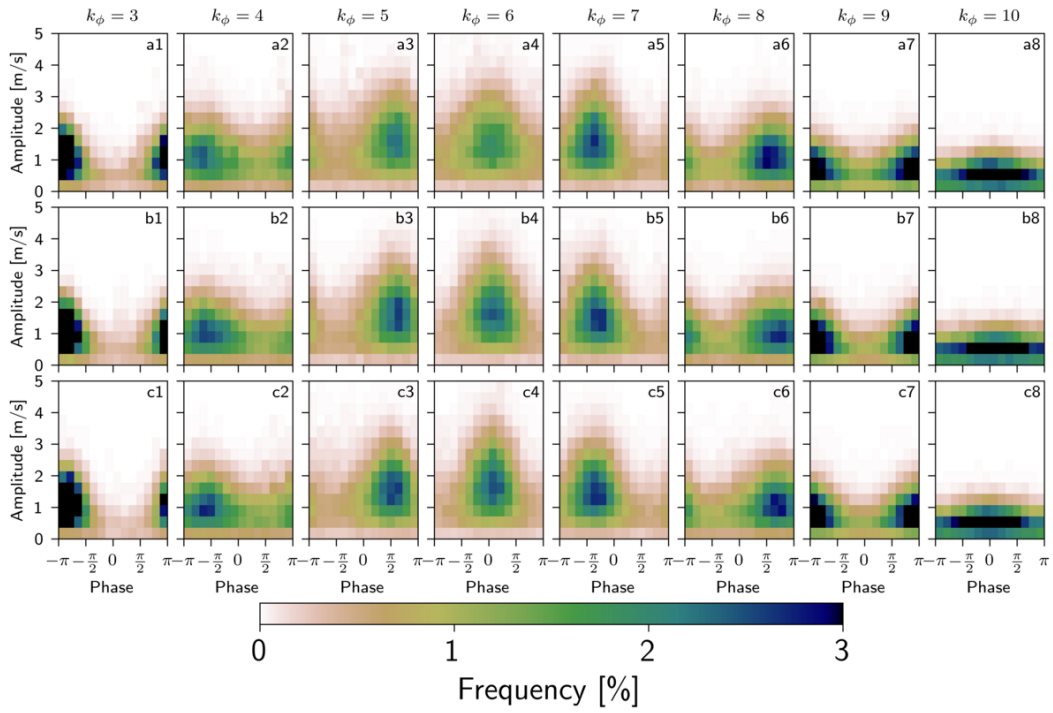


Figure A.48: **Climatological amplitude-phase histograms of meridional wind speed at 200hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

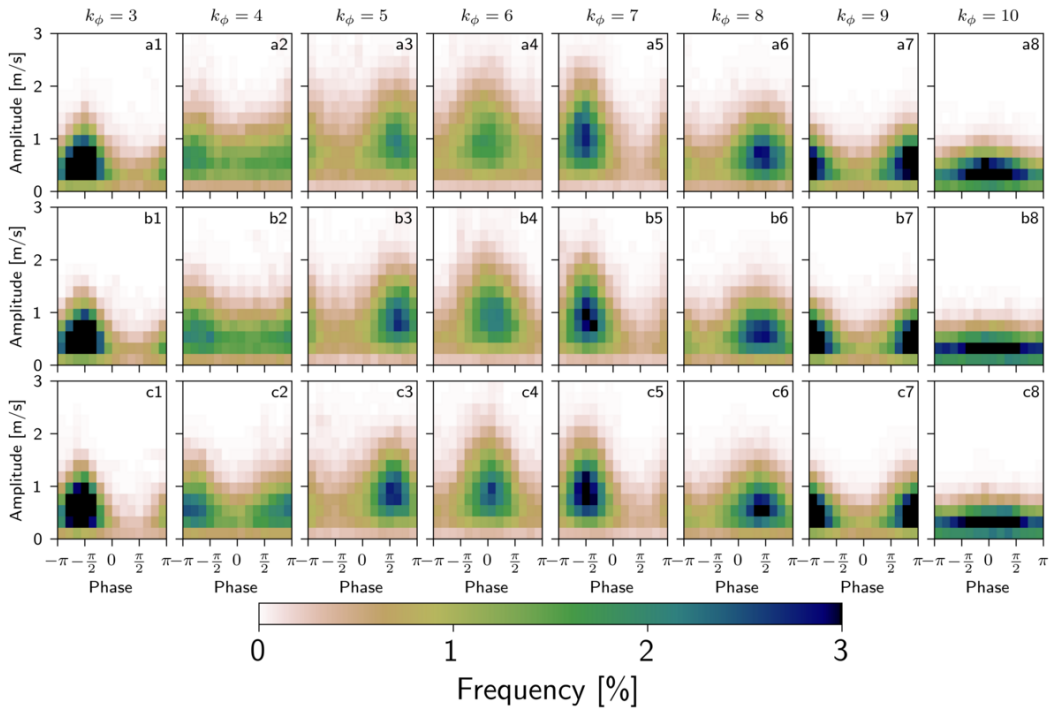


Figure A.49: **Climatological amplitude-phase histograms of meridional wind speed at 500hPa for centennial events.** Wave numbers $k_\phi = 3$ to $k_\phi = 10$ in columns and biased simulations in rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0.

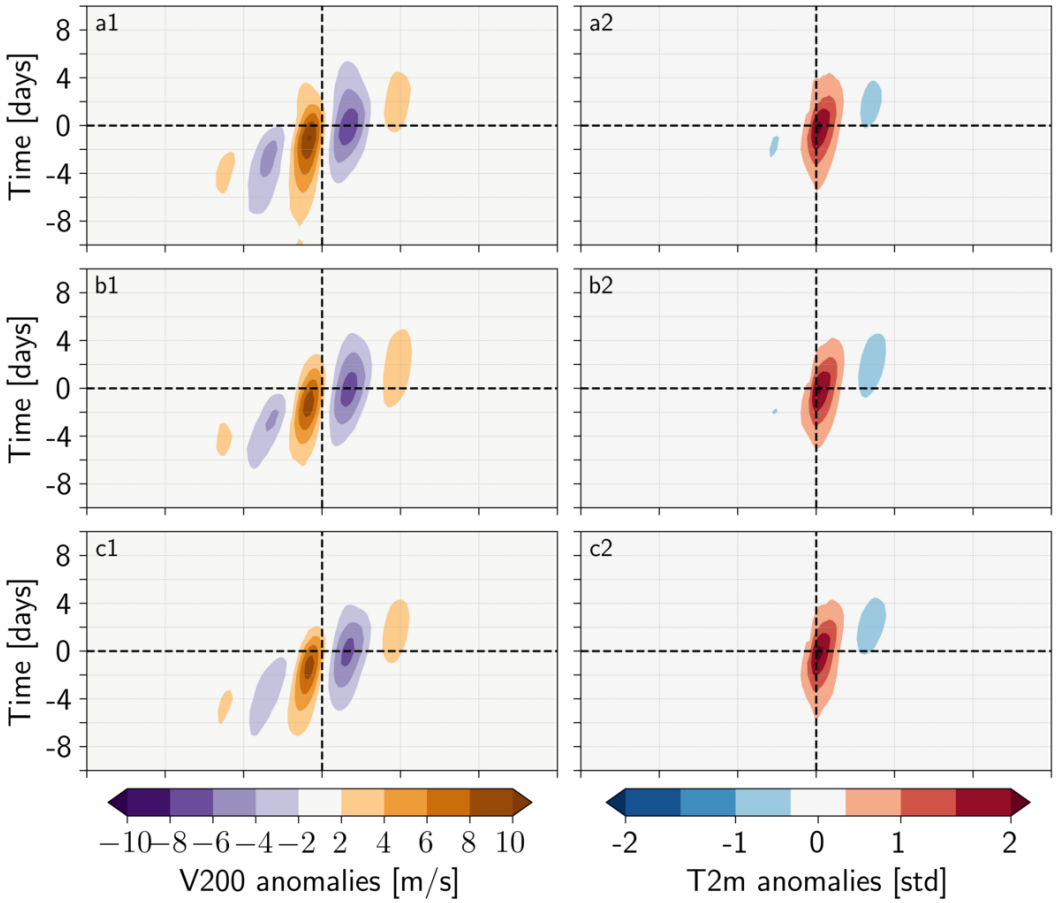


Figure A.50: **Composite Hovmöller plot of the atmospheric and surface dynamics during heatwave events for the control simulations.** First column: meridional wind speed at 200hPa anomalies averaged between 35°N and 65°N. Second column: 2-m air surface temperature normalized anomalies averaged between 45°N and 55°N. The time is expressed relative to the hottest day of each heatwave event. Rows: (a) present, (b) SSP2-4.5 and (c) SSP3-7.0. Number of events: (a) $n = 1089$, (b) $n = 1091$ and (c) $n = 1085$.

Appendix B

Article published in the *Asia-Pacific Journal of Atmospheric Sciences*

Asia-Pacific Journal of Atmospheric Sciences
https://doi.org/10.1007/s13143-022-00305-1

Online ISSN 1976-7951
Print ISSN 1976-7633

ORIGINAL ARTICLE

Challenges in Attributing the 2022 Australian Rain Bomb to Climate Change

Camille Cadiou[1] · Robin Noyelle[1] · Nemo Malhomme[1] · Davide Faranda[1,2,3]

Received: 13 July 2022 / Revised: 25 October 2022 / Accepted: 15 November 2022

Abstract

In February and March 2022, the eastern coast of Australia recorded an unprecedented amount of precipitation with extended floods and damages to properties amounting at least to AUD 2.3 billions. In this paper we use both reanalysis and observations to perform a statistical and dynamical attribution of this precipitation event to climate change. We define 1948-1977 as the counterfactual period and 1990-2019 as the factual one. The statistical attribution is based on fitting the generalized extreme value distribution for 3-days averaged precipitation annual maxima for the two periods, while the dynamical attribution aims at looking at the recurrence properties of sea-level pressure and geopotential height patterns in both periods. We find that the dynamics of the event consists in an unprecedented combination of several factors: a tropical atmospheric river, the presence of the Coral low pressure system and a blocking anticyclone offshore Eastern Australia. Our main finding is that no clear attribution statements can be made, both because of the unprecedented nature of this event, the lack of long high quality available data and the dependence of the results on the La Nina phase of El Nino Southern Oscillation.

Keywords Climate change · Extreme precipitation · Attribution · Australia

1 Introduction

In the latest IPCC report (Masson-Delmotte et al. 2021), researchers confirm the role that human activity plays in the climate upheavals of recent decades. The scientists also confirm the role of anthropogenic greenhouse gases emissions in modifying frequency and intensity of extreme weather events that we experience today. The IPCC scientists warn in particular about the effects that climate change is having on the water cycle. Studies have shown that rising temperatures are leading to more intense rainfall, flooding but also to more severe droughts (Cook et al. 2018) in some regions of the world. The distribution, frequency and intensity of precipitations are changing significantly across the globe, especially in subtropical regions accustomed to monsoons (Douville et al. 2021). In the 21st century, coastal areas are already experiencing and will increasingly experience flooding due to heavy precipitations and sea level rise (Wilby and Keenan 2012). Flooding from rising seas, which was only occasional a few decades ago, could occur every year by the end of the century (Hirabayashi et al. 2013).

Australian average land temperatures have risen by 1.44° since 1910, according to the recent study by Grainger et al. (2022). The oceans surrounding the continent have also warmed by an average of one degree over the same period,

The original online version of this article was revised: In this article the history line has been updated

Communicated by:June-Yi Lee

Camille Cadiou and Robin Noyelle contributed equally to this work.

✉ Davide Faranda
davide.faranda@cea.fr

Camille Cadiou
camille.cadiou@lsce.ipsl.fr

Robin Noyelle
robin.noyelle@lsce.ipsl.fr

Nemo Malhomme
nemo.malhomme@lsce.ipsl.fr

[1] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, 91191, France

[2] London Mathematical Laboratory, 8 Margravine Gardens London, W6 8RH, London, UK

[3] Laboratoire de Météorologie Dynamique/IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, Paris, France

leading to more frequent marine heat waves. Half of the corals in Australia's Great Barrier Reef have died since 1995 because of this rise in water temperature in conjunction with increased acidification (Wolff et al. 2018; Holland et al. 2020). There is a wide consensus on the fact that sea levels will rise in future climates, affecting Australian coasts (Church et al. 2006; Woodruff et al. 2013; Church et al. 2017). As in other regions, this will likely increase the hazards linked to storm surges associated with tropical cyclones. Other authors have also pointed out that rainfall has decreased in southwestern Australia as well as in the fire-ravaged southeast, while increasing in the north, which has been hit by major floods and destructive cyclones in recent years (Dey et al. 2019; Abram et al. 2021).

If climate change projections for Australia are worrying, the country is already experiencing severe extreme weather events. Indeed, climate change has already taken its toll on Australia (Head et al. 2014; Abram et al. 2021; Canadell et al. 2021), which has been hit by fires, drought and intense cyclones. In 2019-2020 the country was already ravaged by wildfires of exceptional proportions after experiencing the hottest and driest year on record (Borchers Arriagada et al. 2020; Jetten et al. 2021). The fires destroyed an area almost the size of the United Kingdom, killed several people and killed or displaced nearly three billion animals, costing the Australian economy about 7 billion AUD (Filkov et al. 2020). In the current 2022 year, both high and low temperature records have been broken in Australia. On one hand, on the 13/01, the coastal town of Onslow reached an unprecedented 50.7°, the hottest temperature ever recorded in Australia for sixty-two years (The Guardian 2022). On the other hand, on the 02/06, the temperature at the Thredbo Top weather station reached −6.9° and historic snowfalls for this time of year (The New Daily 2022). Future projections of climate-related disasters for the coastal areas as well as the vast island-continent report an increasing risk of extreme events capable of heavily damaging natural ecosystems and impact health and agriculture (Hobday et al. 2018; Dey et al. 2019; Ukkola et al. 2020).

Here we will focus on the record-breaking extreme precipitation event which occurred from the last week of February to mid-March 2022, in the eastern coast of Australia. Heavy precipitations - 345mm of rain were recorded at Brisbane on the 28th of February - met soils already close to saturation from the recent rains in the end of 2021. This lead to severe and enduring flooding in March. This event caused 22 recorded casualties, rendered uninhabitable 25,000 homes and businesses, forced 200,000 people to evacuate and produced massive power and transports outages for a total of ∼2.32 billion AUD of damage (Foreign Policy 2022).

In order to explain this rain bomb event, we have identified two main ingredients: i) an atmospheric river, formed north of the continent around the 22nd of February, in the upper atmosphere – between eight and 10 kilometres up – moved from the ocean to the south reaching the continent and producing a continuous flux of precipitation originating from the tropics, and ii) a dipolar structure consisting of a low pressure system in the Coral Sea and a blocking anticyclone further east. The combination of these elements allowed to transport moist air from the tropics to the Coral Sea, and then condense the moisture that has then fallen as rain thanks to the low pressure, while winds were pushing rain over Queensland's south east. Precipitation was further enhanced by orographic effects. At an even larger scale, the phenomena were embedded in a strong La Nina phase, which induces additional air moisture in the south west Pacific and wind circulation patterns that could have played a role in the intensity of the event.

Preliminary results of the University of Melbourne reported by national and international newspapers (see, e.g. (The Guardian 2022)) showed that on the 26th and 27th of February, the two days with heaviest rain, 16 times the water held by Sydney harbour (500 bn liters) flowed in the atmospheric river above Queensland. Indeed, during those days rainfall of over 400 millimetres was recorded across the greater Brisbane area. This would add up on the 28th February to a total of 676.8 millimetres of rainfall, the largest three, and seven, day total ever recorded in Brisbane. Another record breaking amount of rainfall was recorded in Mount Glorious which received rainfall in excess of 1770 millimetres in the week until 28th of February. More than 30 locations across the south-east recorded rainfall in excess of 1000 millimetres, resulting in the floods being more intense than that of 1974. The unprecedented flooding that raised many rivers above record highs moved south, leaving towns underwater. The area of Sydney was more affected on the 3rd and the 8th of March with the city's chief reservoir, spilling at a rate in excess of 70 gigalitres a day on March 3. Residents in parts of western Sydney were told to evacuate for the second year in a row as the city's Warragamba Dam overflowed. But the system stalled before it passed over Sydney.

Damage from floods is only partially estimated at this stage, but is expected to reach almost AUD 2.3 billion (Foreign Policy 2022), an estimate that exceeds that of the Insurance Council of Australia, which also summed-up the cost of claims from the disaster to AUD 1.45 billion (ClimateCouncilAustralia 2022). However, this is only a lower bound, as the estimates are still expected to grow.

Figure 1(a) presents the synoptic situation averaged over the 26th, 27th and 28th of February over the region. We see the dipole structure with an upper level low pressure system centered above the Brisbane region and a high pressure system over the Coral sea. Figure 1(b) presents the

standardized anomaly of accumulated precipitations for a one month period (20/02-20/03) with respect to the 1990-2019 period. The standardized anomaly was computed by considering the series of accumulated precipitations for the 20/02-20/03 period for each year over 1990-2019, removing the mean and dividing by the standard deviation. The Eastern coast of Australia saw extremely strong precipitations, with many locations exceeding 5 standard deviation anomalies with respect to the climatology.

We study whether this event can be attributed to climate change using dynamical systems theory to target the concurrent atmospheric circulation patterns and search for pattern recurrences in the far (1948-1977) and recent past (1990-2019). Our working hypothesis is that the far past acts as a counterfactual world where the Earth climate was not influenced by anthropogenic forcing when compared to the recent past (the factual world). Additionally, we assume that the 30-year period is long enough to average out the interannual variability of the atmospheric motions (as that caused, for example, by El-Niño - Southern Oscillation). Finally we verify that these events produce similar impacts on the targeted regions.

The paper is organized as follows: Section 2 describes the data and methods used in this study. This is followed by the results (Section 3) providing both a statistical a dynamical attribution analysis. We conclude with a discussion and perspectives in Section 4.

2 Data and Methods

2.1 Data

In order to detect significant changes in the circulation associated with Australian Rain bombs, we use daily sea level pressure (slp), 500hPa geopotential height ($z500$) and 2m air temperature ($t2m$) data from the NCEP/NCAR reanalysis (Kalnay et al. 1996) over the period 01/01/1948 – 31/03/2022. The data have a horizontal resolution of 2.5° × 2.5°.

The precipitation data are taken from the high resolution daily rainfall gridded datasets of the Bureau of Meteorology of Australia. The data have a horizontal resolution of 0.05° × 0.05°. Daily precipitation data from the Alderley station, close to Brisbane, the Cairns Aero and the Sidney Botanic gardens stations are also used. The gridded data set will be used for the dynamical attribution (Section 3.2) and the station data for the statistical attribution (Section 3.1). For the ENSO index, we used the Nino3.4 detrended index as in Van Oldenborgh et al. (2021).

2.2 Methods

Extreme Value Theory (EVT) has been introduced in the study of atmospheirc flows a decade ago (Freitas et al. 2008) and has gained a considerable amount of attention in both the applied mathematics and the climate science communities (Lucarini 2012; Lucarini et al. 2016; Faranda et al. 2019). It can be used to compute recurrence times statistics for meteoroloical observables, but also gives access to important information on the stability and the predictability of a particular climatic state. We can apply this framework to the study of how weather extremes are influenced by anthropogenic-driven climate change conditioned to the occurrence of a specific atmospheric circulation pattern (Faranda et al. 2020). Indeed weather extremes are associated with synoptic objects, i.e. a cyclone,

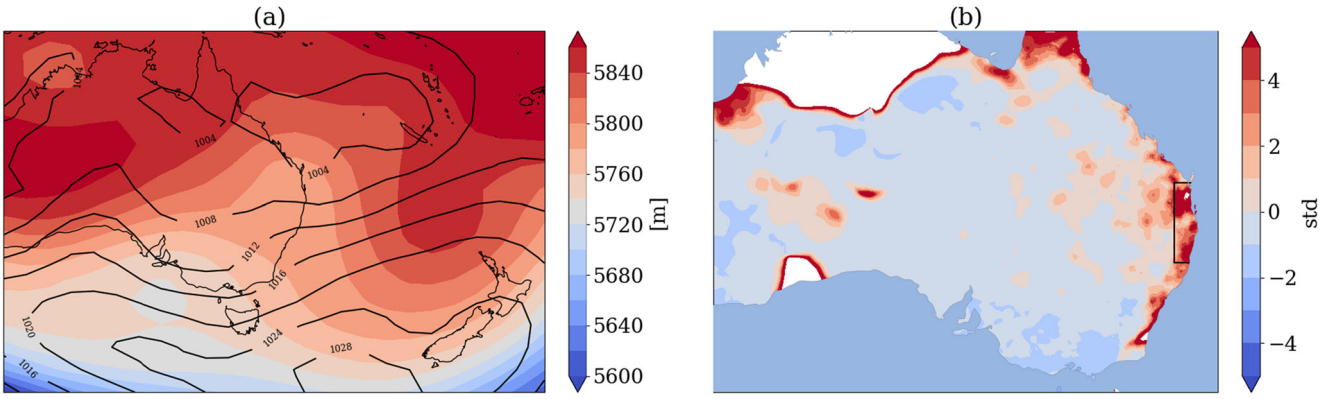


Fig. 1 Description of the rain bomb event. (a) Geopotential height at 500hPa (colors) and sea level pressure (contours) averaged over the 26th, 27th and 28th of February 2022. (b) Standardized anomaly of accumulated precipitations for the period 20th February - 20th March with respect to the 1990-2019 period. Regions in white correspond to standardized anomaly above 15 standard deviations

an anticyclone, or a couple of these structures. During this identification, we select a region (a longitude/latitude box) that fits the synoptic objects that we want to examine. Once fixed the day of interest ζ, we scan all the 3-days rolling averaged sea-level pressure (slp) and 500hPa geopotential ($z500$) maps of the selected region in two periods: 1948-1977 (counterfactual world) and 1990-2019 (factual world). For both periods, we select the best 35 analogs as the slp and $z500$ maps minimizing the pointwise euclidean distance with the target day ζ. We verified that the results are not sensitive to the choice of the number of analogs provided that we extract between 25 and 50 analog maps. 30 years long factual and counterfactual periods ensure a large enough statistical sample of slp and $z500$ maps to be able to select sufficiently authentic analogs. Furthermore, it is short enough to satisfy the assumption of climate stationarity, with respect to anthropogenic climate change, while still long enough for the interannual and interdecadal natural variability of the atmospheric circulation to be averaged over.

Once obtained the two sets of analogs for the factual and counterfactual periods, we average them out to search for significant differences Δslp and $\Delta z500$. To determine significant changes, we apply a two-tails Welch's t-test (Welch 1947) with different variance at each grid point. We mark as significant only grid point changes for which the p-value of the test is below 0.05. Conditioning to the dates determined for slp or $z500$ maps, we also study the associated 2 meters temperatures $t2m$ and total precipitation maps tp. The Welch's t-test procedure is repeated on these ensembles to identify significant changes.

Using the dynamical systems framework described above, we then introduce a few metrics to determine the recurrence properties of the map in the counterfactual and factual worlds. These metrics are computed twice, once with the set of analogs obtained in each period. The formulas for computing these metrics are given in Appendix A. For more information, please refer to Faranda et al. (2022).

- Analogs Quality Q: We can study the typicality of ζ by comparing the euclidean distance of ζ from its analogs with the distances of the analogs from their own analogs. If the value of Q belongs to the same distribution as the values of Q analogs, then the typicality of the event is ensured and attribution of ζ can be performed. If instead the Q for the chosen day is larger than that of the analog days, then this is an unprecedented slp or $z500$ configuration and the results for attribution must be taken more carefully. A difference in Q between the counterfactual and factual periods indicate a change in typicality of the event due to climate change.
- Predictability Index D: Also called instantaneous dimension. For an atmospheric state ζ, $D(\zeta)$ measures the density of similar configurations. We can compute $D(\zeta)$ in both the periods to detect a climate-change induced shift in predictability. Indeed, D is a proxy for the number of degrees of freedom of the map, meaning that the higher the dimension the more unpredictable the next slp or $z500$ maps will be. If the dimension D of the chosen day is higher or lower than that of the analogs, then the day will be less or more predictable than its closest dynamical situations.
- Persistence Index Θ: Another information derived from the dynamical systems theory is the persistence of a given configuration. The persistence counts for how many days we are likely to observe a map that will resemble ζ. Once again we will compute Θ for the two different periods and using the analogs as well to detect shifts from the factual to the counterfactual worlds.
- Seasonality of Analogs: We can also simply count the number of analogs in each month to detect whether a given circulation shifts towards earlier or later months. This can have strong thermodynamic implications.

We also provide results of the attribution conditional not on the past and present climate, but on the El-Niño - Southern Oscillation (ENSO) mode. ENSO is known to be a major factor in the variability of tropical and subtropical regions around the Pacific and likely played a role in the February 2022 event for which the mode was strongly negative. Therefore, we decided to run the attribution analysis conditional on the Nino3.4 index being above the 75% percentile (ENSO+) and below the 25% percentile (ENSO-) for the period 1948-2019. Analogs are then computed in those two pools of data using either 3-days rolling averaged slp or $z500$ maps.

3 Results

3.1 Statistical Attribution

In order to assess the possible changes in the probability of occurrence for the event due to climate change, we use the rapid statistical attribution method described in Section 2. We fit the 3-days averaged precipitation yearly maxima to a Generalized Extreme Value (GEV) law for the factual and counterfactual periods for three stations, namely the Alderley station (Brisbane) at the epicentre of the 26th-28th February event, the Sydney Botanic garden station and the Cairns Aero station. We repeat the same analysis over the region $152° - 154°$W, $25.5° - 31°$S centered on Brisbane

where the event was the most extreme (see, e.g. Fig. 1b) .

Figure 2 shows the resulting GEV fits for Brisbane (a), Brisbane region (b), Sydney (c) and Cairns (d). In both the factual and counterfactual worlds the event of April 2022 in Brisbane was exceptional, with return periods estimated of more than a thousand years. The event was also a 100-year precipitation event in the Brisbane region in both periods. The 26th-28th February event was less extreme in Sydney and Cairns, with return periods of less than a decade in both stations for the factual and counterfactual periods. In Alderley station return levels are slightly higher in the counterfactual period. We can explain this by the occurrence of an extreme precipitation events in January 1974, that also produced floods in Brisbane. However, the 95% confidence intervals are too wide to conclude to any detectable effect of climate change with this method.

Figure 3 shows the same analysis but conditioned to the phases of ENSO. In particular, we condition using Nino3.4 index, sampling ENSO+ and ENSO- phases, defined respectively as the first and last quartile of the index. Results show a stronger effect of the la Niña phase (ENSO-) on precipitation rates at both Brisbane stations (Fig. 3a). The 2022 26th to 28th February event is very rare during ENSO+ phases with a return period of over 400 years. Instead, it has a return period of 120 years in ENSO- phases. This is in agreement with the influence of the strong la Niña phase currently ongoing on the intensity of the 2022 event suggested in Section 1. This effect is less clear if we consider the Brisbane region (Fig. 3b): in this case the confidence intervals are so wide that it

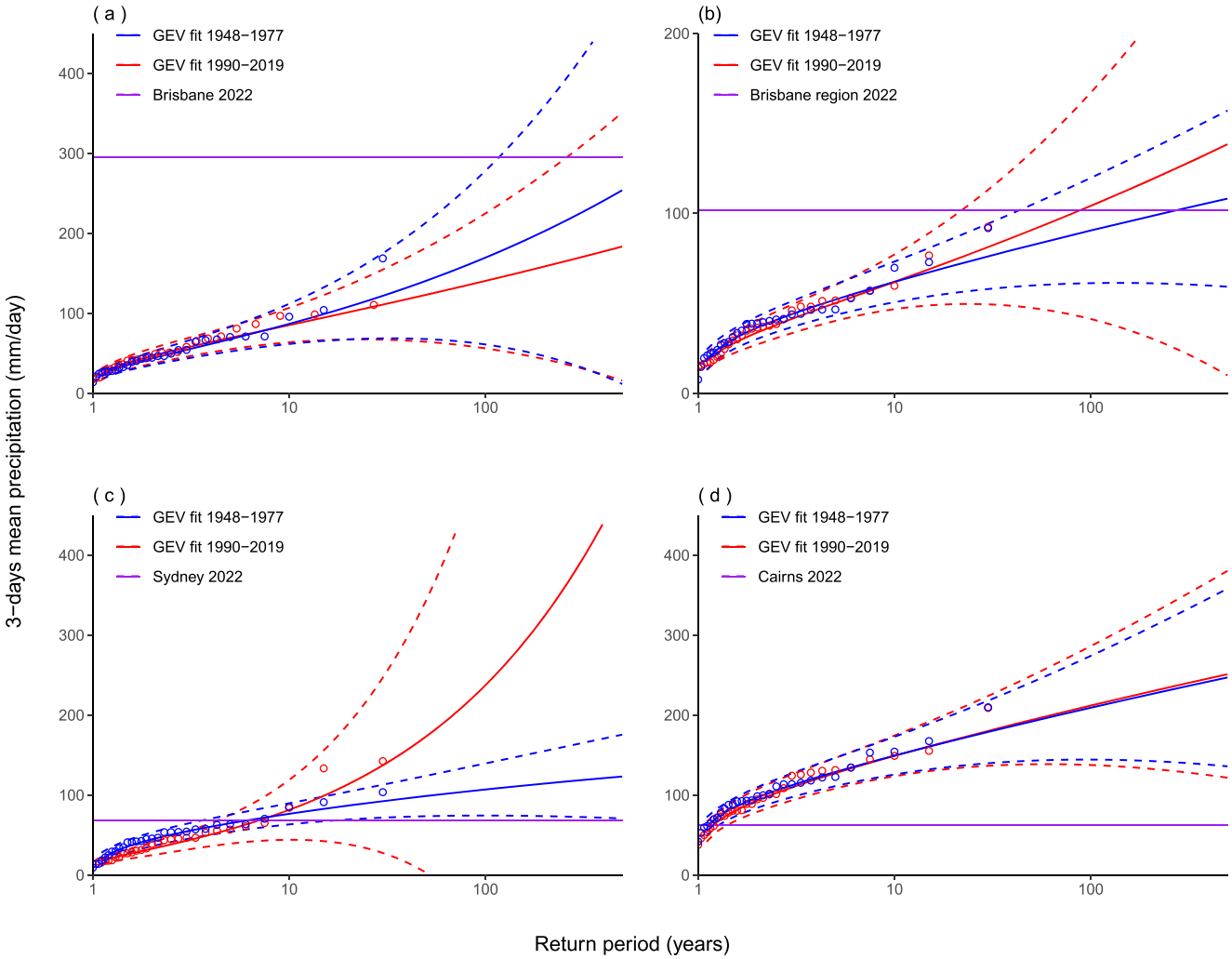


Fig. 2 Fit of the annual maximum of 3-days averaged precipitation running means at Brisbane (Alderley station) (a), the Brisbane region (152° − 154°W, 25.5° − 31°S) (b), Sydney (Sydney Botanic gardens station) (c) and Cairns (Cairns Aero station) (d) to a GEV in the factual and counterfactual periods. Gumbel plot of the GEV fit over 1948-1977 (blue lines with 95% confidence intervals), and 1992-2021 (red lines) using the maximum likelihood estimation. The purple line shows the value of intensity of the observed event at the station or area of interest in 2022

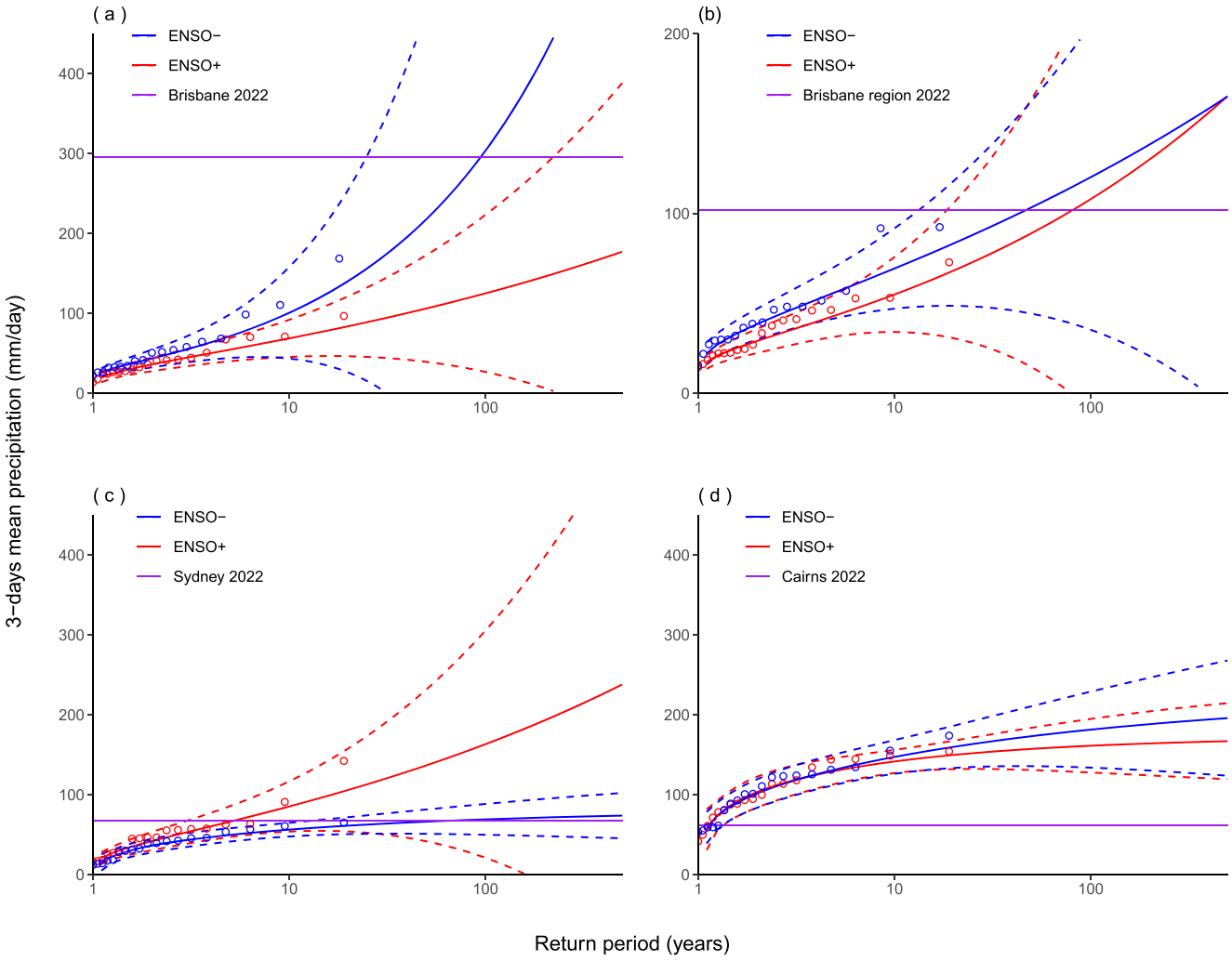


Fig. 3 Fit of the annual maximum of 3-days averaged precipitation running means at Alderley station (a) and over the Brisbane region (152° − 154°W, 25.5° − 31°S) (b) to a GEV, depending on El Niño 3.4 detrended index. Gumbel plot of the GEV fit over years in the 1st quartile (red lines) and 4th quartile (blue lines) of NINO3.4 index average over JFM from 1948 to 2021, with 95% confidence intervals. The purple line shows the value of intensity of the observed event at Alderley station in 2022

is difficult to draw any conclusion. The same procedure applied to Sydney (Fig. 3c) and Cairns (Fig. 3d) station data does not allow drawing any conclusion because of the wide confidence intervals, wich prevent to obtain statistical significant differences, at least at the 95% level chosen in this study.

3.2 Dynamical Attribution

To identify the dynamic and thermodynamic factors which may have played a role in the intensity of the event, we ran the dynamical attribution framework presented previously.

Figures 4 and 5 present the results using respectively slp and $z500$ analogs. These figures display the $slp/z500$ (a–d), $t2m$ (e–h) and $prate$ (i–l) maps and, from left to right, the maps of the event (a,e,i), the composites of counterfactual (b,f,j) and factual (c,g,k) analogs and the difference between the factual and counterfactual analogs (d,h,l) where hashed-filled areas show significant differences. Distributions of additional metrics comparing the counterfactual and factual periods, namely analogs quality, predictability, persistence and distribution by season are shown in panels m–p, respectively.

For the slp analogs, we firstly notice that the analogs quality (Q) is not good and therefore results must be interpreted carefully (panel (m)). In the factual period, there is a signal of intensified high-pressures over New Zealand, which may be related to the stronger advection of moist air from the Coral sea. When we look at the precipitation difference map (Fig. 4 (l)), there are indeed stronger precipitations on the Eastern coast of Australia in the factual period but only few regions are significantly different than

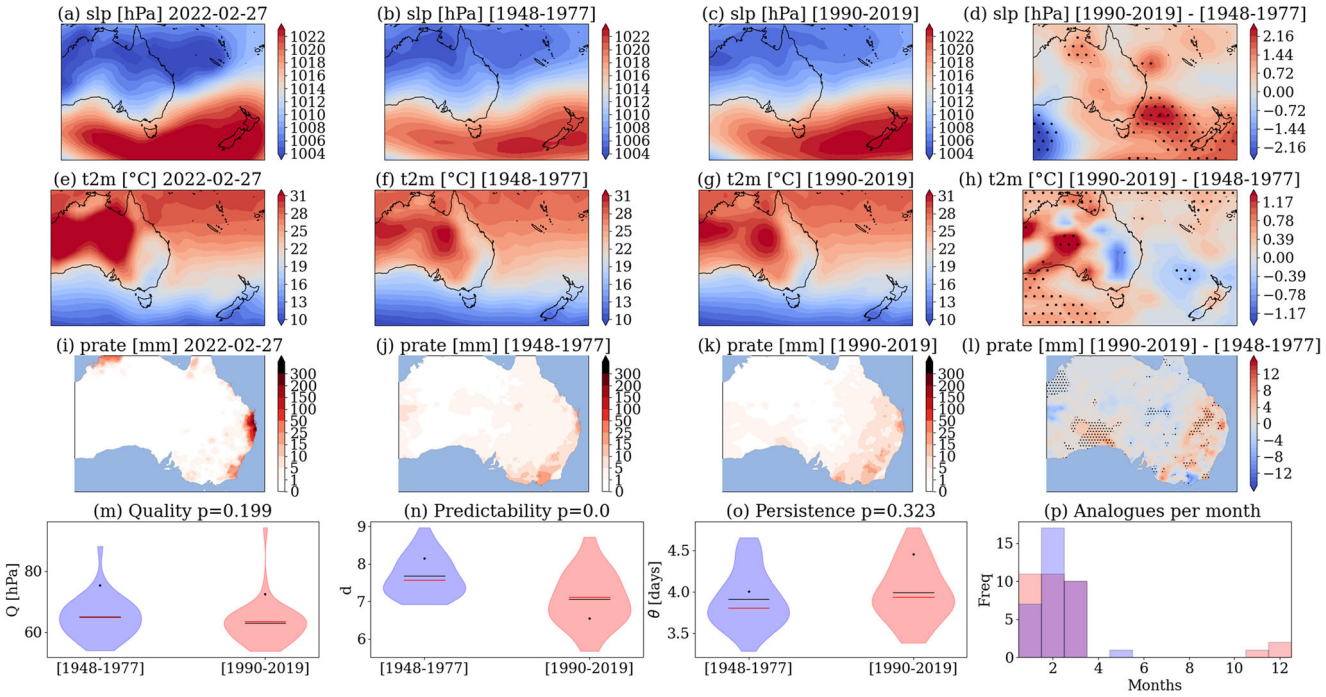


Fig. 4 Dynamical attribution using slp analogs for the Australian rain bomb on 27-02-2022. 3-days rolling mean sea-level pressure slp (a), 2-meter temperatures $t2m$ (e) and total precipitation $prate$ (i) centered on the 27-02-2022. Average of the 35 sea-level pressure analogs found for the counterfactual [1948-1977] (b) and factual [1990-2019] (c) periods and corresponding 2-meter temperatures (f,g) and daily precipitation rate (j,k). Δslp (d), $\Delta t2m$ (h) and $\Delta prate$ (i) between factual and counterfactual periods: hashed-filled areas show significant differences. Violin plots for counterfactual (blue) and factual (red) periods for the Analogs Quality Q (m) the Predictability index D (n), the Persistence index θ (o) and the distribution of analogs in each month (p). Values for the selected day are marked by a black dot. Black (resp. red) lines represent the empirical mean (resp. median) of the distribution

the counterfactual period and do not fully correspond to the regions where the precipitations are maximum during the February 2022 event. There is moreover no strong temperature signal as depicted in panel (h) using 2m air temperature which does not display significant differences between the factual and counterfactual periods on the Eastern coast and in the Coral sea. However, panel (o) and (p) show that the predictably is lower and the persistence is higher in the factual period. The event is particularly persistent compared to its analogs in the factual period which may explain the intensity of the event.

These results are confirmed when using the $z500$ analogs (Fig. 5), which are better as shown in panel (m). The factual period displays stronger high $z500$ over the northern region but also almost no significant temperature signal (panel (h)). The low predictability D (panel (n)) but high persistence Θ (panel (p)) of the event is also noticed using $z500$ analogs. When it comes to precipitations, even though they are stronger during the factual period (panel (l)), this difference is not significant.

These results are coherent with the statistical attribution in so far as the temperature signal associated with climate change is weak. There is however a noticeable dynamical signal in the two periods as represented in the slp and $z500$ analogs difference maps. One may indeed notice that ENSO, which is the major driver of variability in the region, was in a strong and persistent La Nina phase during the event, which is not fully represented using analogs (see Fig. 6). We therefore decided to run the same dynamical attribution analysis using the Nino3.4 index to define ENSO- phases and factual ENSO+ phases.

The results of this analysis are presented in Figs. 7 and 8 using respectively slp and $z500$ analogs. On the dynamical side, there is a significant stronger upper level low during ENSO- phases (Fig. 8 panel (d)) analogs. Using both analogs, the temperature signal is much stronger during ENSO+ phases, which may explain the absence of signal when doing the analysis during past and present periods: the natural variability is stronger than the climate change signal on this region. When it comes to precipitations, there is a strong positive signal on the Eastern coast and Brisbane regions during La Nina phases using slp analogs but the quality of analogs is low and these results are not significant using $z500$ analogs.

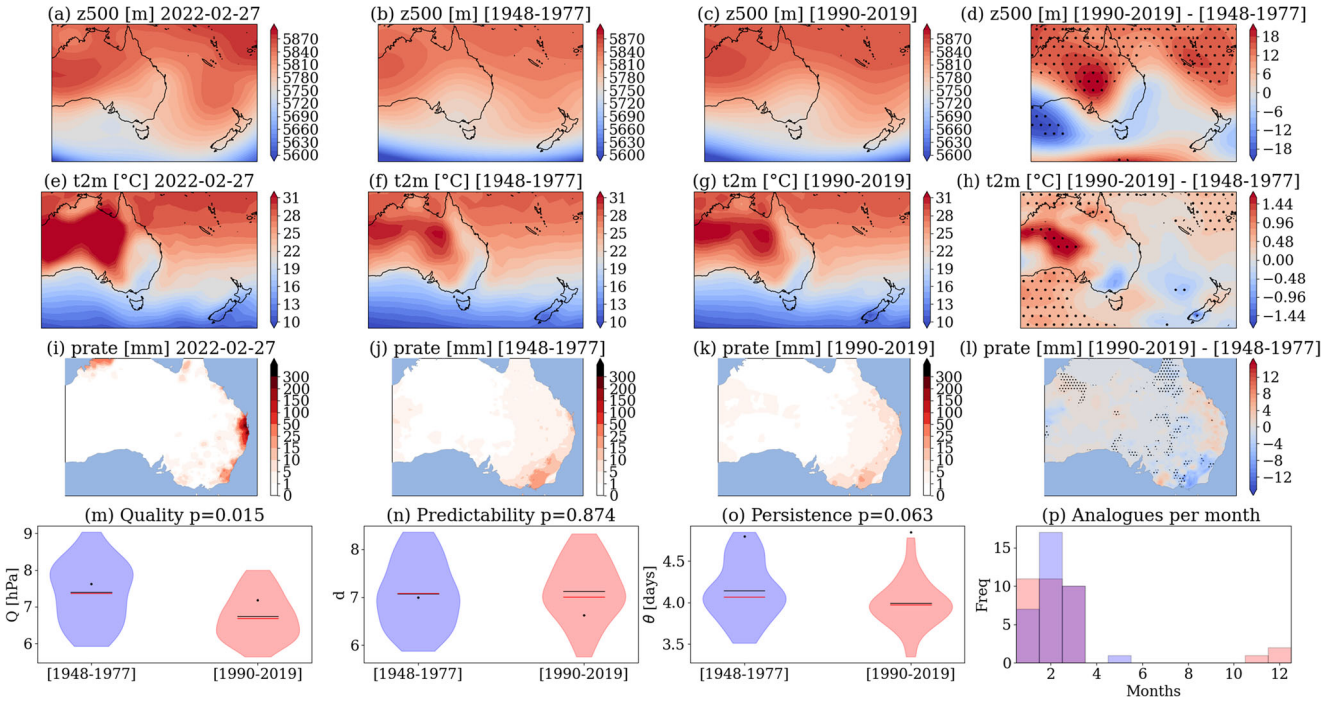


Fig. 5 Dynamical attribution using $z500$ analogs for the Australian rain bomb on 27-02-2022. 3-days rolling mean 500hPa geopotential height $z500$ (a), 2-meter temperatures $t2m$ (e) and total precipitation $prate$ (i) centered on the 27-02-2022. Average of the 35 500hPa geopotential height analogs found for the counterfactual [1948-1977] (b) and factual [1990-2019] (c) periods and corresponding 2-meter temperatures (f,g) and daily precipitation rate (j,k). $\Delta z500$ (d), $\Delta t2m$ (h) and $\Delta prate$ (i) between factual and counterfactual periods: hashed-filled areas show significant differences. Violin plots for counterfactual (blue) and factual (red) periods for the Analogs Quality Q (m) the Predictability index D (n), the Persistence index θ (o) and the distribution of analogs in each month (p). Values for the selected day are marked by a black dot. Black (resp. red) lines represent the empirical mean (resp. median) of the distribution

4 Conclusions

In this paper we have performed a statistical and dynamical attribution of the Australian rain bomb event of February 2022. We have used both reanalyses and historical records of daily precipitations in the past (1948-1977, counterfactual) and present (1990-2019, factual) periods. From a statistical point of view, this extreme precipitation event was unprecedented in the Brisbane region and very intense with respects to previous historical records in the broader Eastern coast of Australia. The statistical attribution suggests that this event has a low probability of happening both in the past and present climate (less than one in a century). There is no clear signal of a climate change

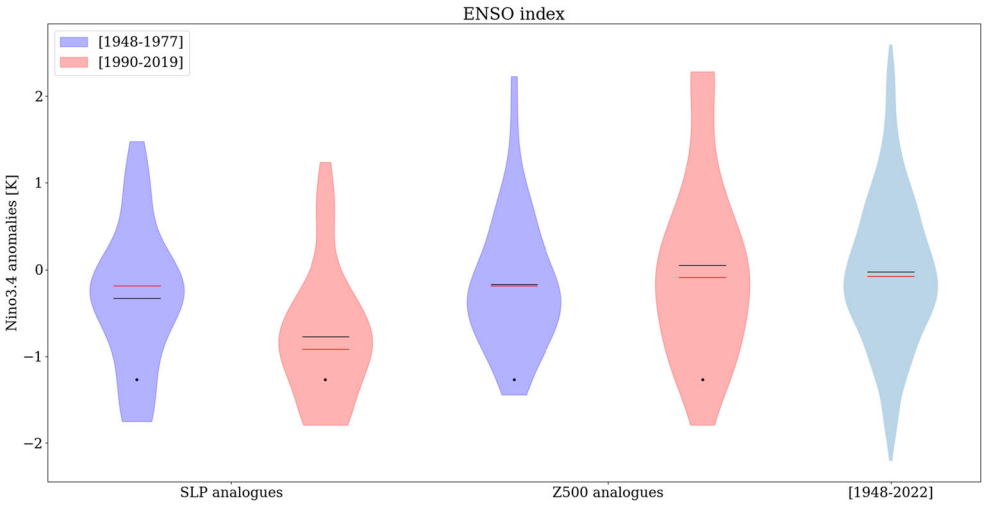


Fig. 6 Distribution of ENSO index between factual and counterfactual periods for the two types of analogs. Blue (resp. red) violin plots represent the distribution of the ENSO index for the analogs in the counterfactual (resp. factual) period. The red (resp. black) line depicts the median (resp. mean) of the distribution. The black dot corresponds to the value of the ENSO index during the February 2022 event. The last violin plot depicts the distribution of the ENSO index for all dates between 1948 and 2019

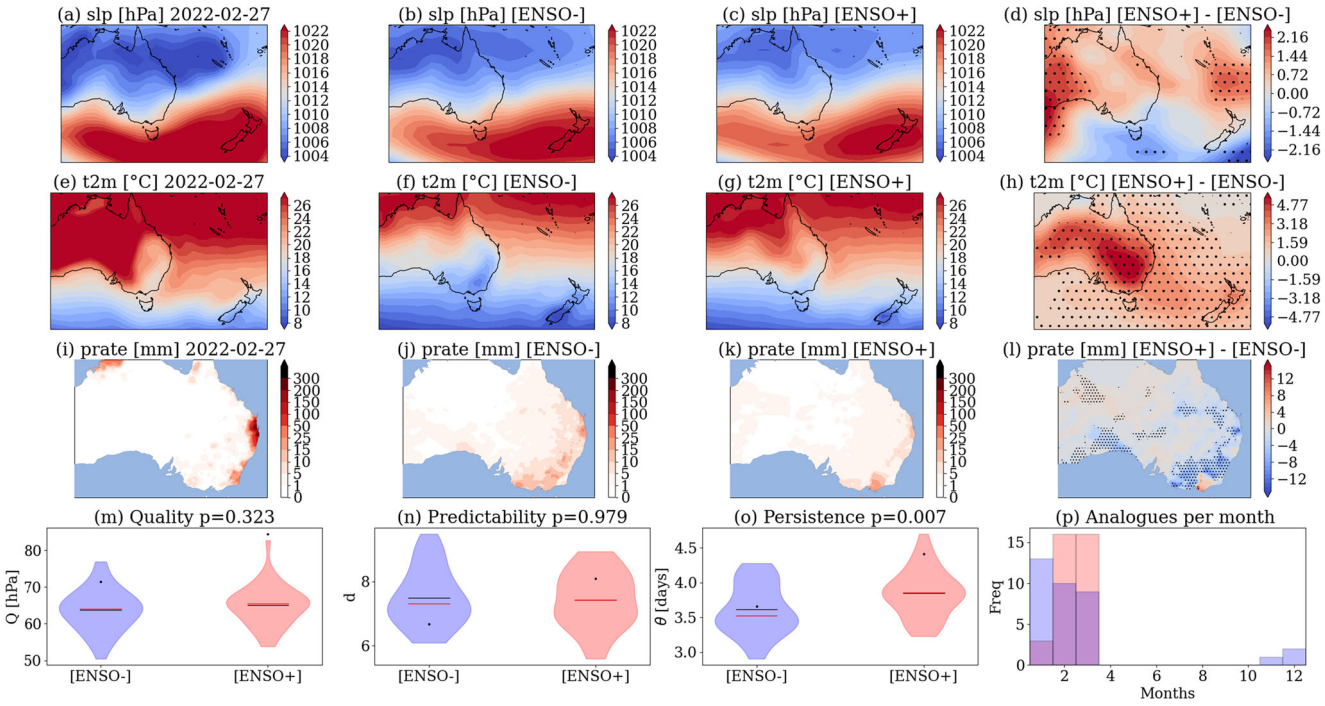


Fig. 7 Dynamical attribution using *slp* analogs for the Australian rain bomb on 27-02-2022 relative to the ENSO index. 3-days rolling mean 500hPa geopotential height $z500$ (a), 2-meter temperatures $t2m$ (e) and total precipitation $prate$ (i) centered on the 27-02-2022. Average of the 35 sea-level pressure analogs found for the counterfactual [ENSO-] (b) and factual [ENSO+] (c) periods and corresponding 2-meter temperatures (f,g) and daily precipitation rate (j,k). $\Delta z500$ (d), $\Delta t2m$ (h) and $\Delta prate$ (i) between factual and counterfactual periods: hashed-filled areas show significant differences. Violin plots for counterfactual (blue) and factual (red) periods for the Analogs Quality Q (m) the Predictability index D (n), the Persistence index θ (o) and the distribution of analogs in each month (p). Values for the selected day are marked by a black dot. Black (resp. red) lines represent the empirical mean (resp. median) of the distribution

influence (high uncertainty) which translates in a low evidence for the attribution of this event to climate change.

The dynamical attribution, performed using the method of analogs circulation patterns proposed in Faranda et al. (2022) to those observed during the events suggest the exceptional characteristics of the event which appears to be unprecedented in both the factual and counterfactual distributions of weather patterns. It has arisen as the combination of ingredients: the tropical moisture was deflected to the subtropics, collected and lifted by a low pressure system which stationed over Eastern Australia blocked by an high pressure offshore the Coral sea. This combination created an atmospheric river capable of transporting several gigaliters of water towards the Queensland and the South West of the continent. Finally, La Niña phase of the ENSO likely played a significant role in the intensity of the event.

The main limitation of this study is that we do not use climate models: the rational for doing so comes from the evidence that both large, regional and local phenomena contributed to this event. In order to perform an attribution study based on models, we would need a large ensemble of convection permitting models resolving at least the region shown in Fig. 1, with specific runs capable to evaluate also the contribution of la Nina to the event. Even with such models, the detailed physics of the precipitation could still depend on the microphysics introduced in the model as shown by Ban et al. (2021). A further limitation of this study is the use of factual and counterfactual periods consisting of only 30 years. While we could in principle consider the whole period twice, as in the protocol described by Philip et al. (2020), we cannot here use the implicit assumption of stationarity of the variability of the rainfall extremes that is underlying the protocol. While this is a limitation on the available data, it is a fair way to account for the variability of extremes, allowing for a larger uncertainty and a conservative estimates on the role of climate change to triggering this event.

To frame our results in a more general framework, we observe that they are in line with what the IPCC report AR6, WG1, Chapter 11.4, states about rainfall over the region, namely that "Available evidence has not shown an increase or a decrease in heavy precipitation over Australasia as a whole (medium confidence), but heavy precipitation tends to increase over Northern Australia (particularly the north-west) and decrease over the eastern and southern

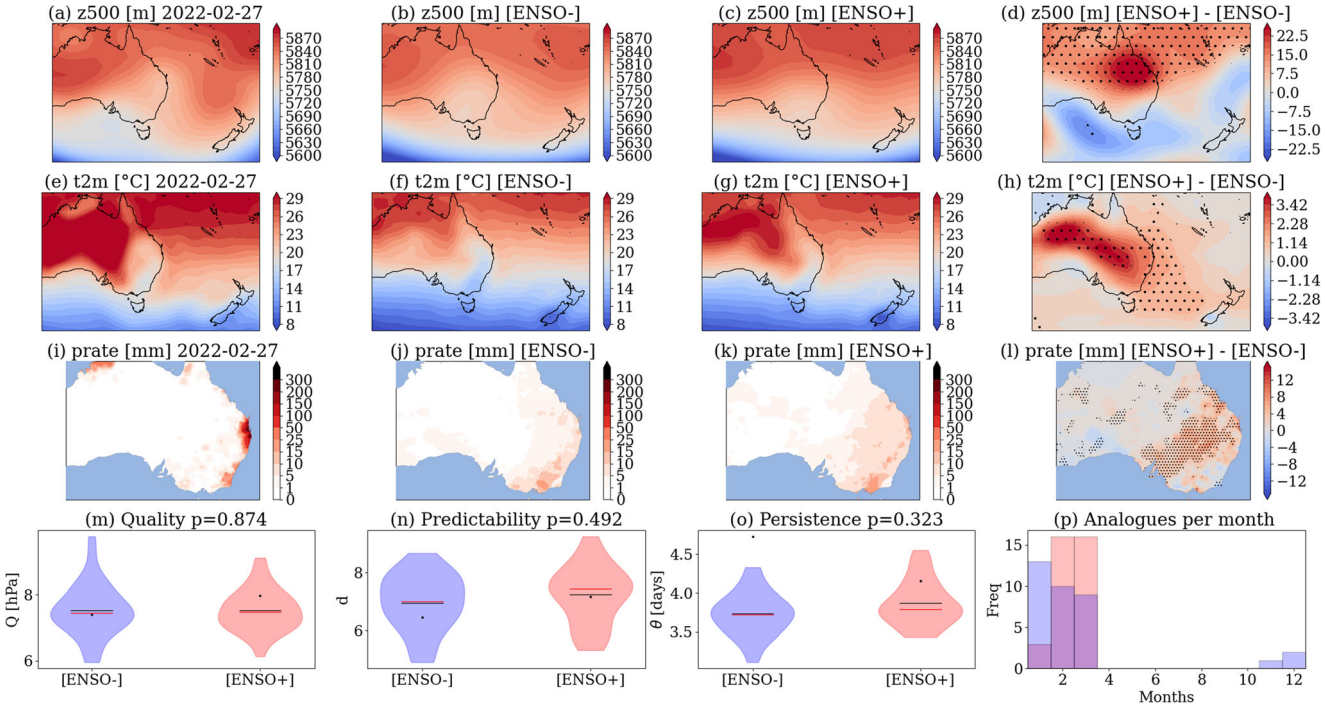


Fig. 8 Dynamical attribution using $z500$ analogs for the Australian rain bomb on 27-02-2022 relative to the ENSO index. 3-days rolling mean 500hPa geopotential height $z500$ (a), 2-meter temperatures $t2m$ (e) and total precipitation $prate$ (i) centered on the 27-02-2022. Average of the 35 500hPa geopotential https://www.overleaf.com/project/627ba03eceee706667e9e0ca height analogs found for the counterfactual [ENSO-] (b) and factual [ENSO+] (c) periods and corresponding 2-meter temperatures (f,g) and daily precipitation rate (j,k). $\Delta z500$ (d), $\Delta t2m$ (h) and $\Delta prate$ (i) between factual and counterfactual periods: hashed-filled areas show significant differences. Violin plots for counterfactual (blue) and factual (red) periods for the Analogs Quality Q (m) the Predictability index D (n), the Persistence index θ (o) and the distribution of analogs in each month (p). Values for the selected day are marked by a black dot. Black (resp. red) lines represent the empirical mean (resp. median) of the distribution

regions." (Seneviratne et al. 2021). We however observe that the compounding dynamical elements driving this event are also observed in a series of unprecedented extreme weather events occurred in the last few years, including the 2021 Canada heat dome, the Antarctica atmospheric rivers, the 2022 Indian, Chinese, Western European and North American heatwaves. Such series of record breaking events raise the question of the emergence of new phenomena linked to global warming, a field to explore statistically, e.g. by using the concept of time of emergence introduced in Hawkins and Sutton (2012) and dynamically, e.g. by looking the possible presence of tipping elements in the climate system (Lenton et al. 2008).

Appendix A: Predictability and Persistence Indices

The attractor of a dynamical system is a geometric object defined in the space hosting all the possible states of the system (phase-space). Each point ζ on the attractor can be characterized by two dynamical indicators: the local dimension D, which indicates the number of degrees of freedom active locally around ζ, and the persistence Θ, a measure of the mean residence time of the system around ζ (Faranda et al. 2017). To determine D, we exploit recent results from the application of extreme value theory to Poincaré recurrences in dynamical systems. This approach considers long trajectories of a system — in our case successions of daily SLP latitude–longitude maps — corresponding to a sequence of states on the attractor. For a given point ζ in phase space (e.g., a given SLP map), we compute the probability that the system returns within a ball of radius ϵ centered on the point ζ. The (Freitas et al. 2010) theorem, modified by Lucarini (2012), states that logarithmic returns:

$$g(x(t)) = -\log(\mathrm{dist}(x(t), \zeta)) \tag{1}$$

yield a probability distribution such that:

$$\Pr(z > s(q)) \simeq \exp\left[-\vartheta(\zeta)\left(\frac{z-\mu(\zeta)}{\sigma(\zeta)}\right)\right] \tag{2}$$

where $z = g(x(t))$ and s is a high threshold associated to a quantile q of the series $g(x(t))$. Requiring that the orbit falls within a ball of radius ϵ around the point ζ is equivalent

to asking that the series $g(x(t))$ is over the threshold s; therefore, the ball radius ϵ is simply $e^{-s(q)}$. The resulting distribution is the exponential member of the Generalized Pareto Distribution family. The parameters μ and σ, namely the location and the scale parameter of the distribution, depend on the point ζ in phase space. $\mu(\zeta)$ corresponds to the threshold $s(q)$ while the local dimension $D(\zeta)$ can be obtained via the relation $\sigma = 1/D(\zeta)$. This is the metric of predictability introduced in Section 2.2.

When $x(t)$ contains all the variables of the system, the estimation of D based on extreme value theory has a number of advantages over traditional methods (e.g. the box counting algorithm (Liebovitch and Toth 1989; Sarkar and Chaudhuri 1994)). First, it does not require estimating the volume of different sets in scale-space: the selection of $s(q)$ based on the quantile provides a selection of different scales s which depends on the recurrence rate around the point ζ. Moreover, it does not require the a priori selection of the maximum embedding dimension as the observable g is always a univariate time-series.

The persistence of the state ζ is measured via the extremal index $0 < \vartheta(\zeta) < 1$, an adimensional parameter, from which we extract $\Theta(\zeta) = \Delta t/\vartheta(\zeta)$. Here, Δt is the timestep of the dataset being analysed. $\Theta(\zeta)$ is therefore the average residence time of trajectories around ζ, namely the metric of persistence introduced in Section 2.2, and it has unit of a time (in this study days). If ζ is a fixed point of the attractor, then $\Theta(\zeta) = \infty$. For a trajectory that leaves the neighborhood of ζ at the next time iteration, $\Theta = 1$. To estimate ϑ, we adopt the Süveges estimator (Süveges 2007). For further details on the the extremal index, see Moloney et al. (2019).

Acknowledgements We acknowledge the Bureau of Meteorology of Australia for their help in obtaining the high resolution daily rainfall gridded datasets. The authors acknowledge the support of the INSU-CNRS-LEFE-MANU grant (project DINCLIC), as well as the gran ANR-19-ERC7-0003 (BOREAS). This work has received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101003469 (XAIDA) and the Marie Sklodowska-Curie grant agreement No. 956396 (EDIPI).

Author Contributions CC, NM and RN performed the analysis. DF designed the analyses. All authors participated to the manuscript preparation and writing.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement N° 956396 under grant agreement No. 101003469 (XAIDA).

Availability of data and materials NCEP data are available on the NOAA website at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html. Rainfall data over Australia are available at the National Computational Infrastructure (NCI) THREDDS server at https://dapds00.nci.org.au/thredds/catalog/zv2/agcd/v1/precip/total/r005/01day/catalog.html.

Code Availability The main results of this work were obtained using Python. The scripts are available upon request.

Declarations

Conflict of Interests The authors declare no conflict of interest.

References

Abram, N.J., Henley, B.J., Sen Gupta, A., et al.: Connections of climate change and variability to large and extreme forest fires in southeast australia. Commun. Earth & Environ. **2**(1), 1–17 (2021)

Ban, N., Caillaud, C., Coppola, E., et al.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part i: evaluation of precipitation. Clim Dyn **57**(1), 275–302 (2021)

Borchers Arriagada, N., Palmer, A.J., Bowman, D.M., et al.: Unprecedented smoke-related health burden associated with the 2019–20 bushfires in eastern australia. Med. J. Aust. **213**(6), 282–283 (2020)

Canadell, J.G., Meyer, C., Cook, G.D., et al.: Multi-decadal increase of forest burned area in Australia is linked to climate change. Nat. Commun. **12**(1), 1–11 (2021)

Church, J.A., Hunter, J.R., McInnes, K.L., et al.: Sea-level rise around the australian coastline and the changing frequency of extreme sea-level events. Aust. Meteorol. Mag. **55**(4), 253–260 (2006)

Church, J.A., McInnes, K.L., Monselesan, D., et al.: Sea level rise and allowances for coastal councils around australia–guidance material (2017)

ClimateCouncilAustralia: A supercharged climate: rain bombs, flash flooding and destruction. Available at https://www.climatecouncil.org.au/resources/supercharged-climate-rain-bombs-flash-flooding-destruction Accessed: 08/07/2022 (2022)

Cook, B.I., Mankin, J.S., Anchukaitis, K.J.: Climate change and drought: from past to future. Curr. Clim. Chang. Rep. **4**(2), 164–179 (2018)

Dey, R., Lewis, S.C., Arblaster, J.M., et al.: A review of past and projected changes in australia's rainfall. Wiley Interdiscip. Rev. Clim. Chang. **10**(3), e577 (2019)

Douville, H., Raghavan, K., Renwick, J., et al.: Water cycle changes. In: Climate Change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (2021)

Faranda, D., Messori, G., Yiou, P.: Dynamical proxies of north atlantic predictability and extremes. Sci. Reports **7**, 41–278 (2017)

Faranda, D., Alvarez-Castro, M.C., Messori, G., et al.: The hammam effect or how a warm ocean enhances large scale atmospheric predictability. Nat. Commun. **10**(1), 1–7 (2019)

Faranda, D., Vrac, M., Yiou, P., et al.: Changes in future synoptic circulation patterns: consequences for extreme event attribution. Geophys. Res. Lett. **47**(15), e2020GL088,002 (2020)

Faranda, D., Bourdin, S., Ginesta, M., et al.: A climate-change attribution retrospective of some impactful weather extremes of 2021. Weather Clim. Dynam. **3**, 1311–1340 (2022). https://doi.org/10.5194/wcd-3-1311-2022

Filkov, A.I., Ngo, T., Matthews, S., et al.: Impact of australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends. J. Saf. Sci. Resilience **1**(1), 44–56 (2020)

Foreign Policy: The 'rain bomb' that could shape the australian election. Available at https://foreignpolicy.com/2022/03/11/australia-flood-election-climate-change/ Accessed 08 July 2022 (2022)

Freitas, A.C.M., Freitas, J.M., Todd, M. (2008)

Freitas, A.C.M., Freitas, J.M., Todd, M.: Hitting time statistics and extreme value theory. Probab. Theory Relat. Fields **147**(3-4), 675–710 (2010)
Grainger, S., Fawcett, R., Trewin, B., et al.: Estimating the uncertainty of australian area-average temperature anomalies. Int. J. Climatol. **42**(5), 2815–2834 (2022)
Hawkins, E., Sutton, R.: Time of emergence of climate signals. Geophys. Res. Lett. 39(1) (2012)
Head, L., Adams, M., McGregor, H.V., et al.: Climate change and australia. Wiley Interdiscip. Rev. Clim. Chang. **5**(2), 175–197 (2014)
Hirabayashi, Y., Mahendran, R., Koirala, S., et al.: Global flood risk under climate change. Nat. Clim. Chang. **3**(9), 816–821 (2013)
Hobday, A.J., Pecl, G.T., Fulton, B., et al.: Climate change impacts, vulnerabilities and adaptations. Australian marine fisheries (2018)
Holland, M.M., Smith, J.A., Everett, J.D., et al.: Latitudinal patterns in trophic structure of temperate reef-associated fishes and predicted consequences of climate change. Fish Fisheries **21**(6), 1092–1108 (2020)
Jetten, J., Fielding, K.S., Crimston, C.R., et al.: Responding to climate change disaster: the case of the 2019/2020 bushfires in australia. Eur. Psychol. **26**(3), 161 (2021)
Kalnay, E., Kanamitsu, M., Kistler, R., et al.: The ncep/ncar 40-year reanalysis project. Bull. Am. Meteorol. Soc. **77**(3), 437–472 (1996)
Lenton, T.M., Held, H., Kriegler, E., et al.: Tipping elements in the earth's climate system. Proc. Nat. Acad. Sci. **105**(6), 1786–1793 (2008)
Liebovitch, L.S., Toth, T.: A fast algorithm to determine fractal dimensions by box counting. Phys. Lett. A. **141**(8-9), 386–390 (1989). ISBN: 0375-9601 Publisher: Elsevier
Lucarini, V.: Universal behaviour of extreme value statistics for selected observables of dynamical systems, Wouters, J. (2012)
Lucarini, V., Faranda, D., Freitas, A.C.M., et al.: Extremes and recurrence in dynamical systems. 1605.07006 (2016)
Masson-Delmotte, V., Zhai, P., Pirani, A., et al.: Ipcc The physical science basis. Contribution of working group I to the 6th Assessment Report of the Intergovernmental Panel on Climate Change (2021)
Moloney, N.R., Faranda, D., Sato, Y.: An overview of the extremal index. Chaos: Interdiscip. J. Nonlinear Sci. **29**(2), 022,101 (2019)
Philip, S., Kew, S., van Oldenborgh, G.J., et al.: A protocol for probabilistic extreme event attribution analyses. Adv. Stat. Climatol. Meteorol. Oceanogr. **6**(2), 177–203 (2020)
Sarkar, N., Chaudhuri, B.B.: An efficient differential box-counting approach to compute fractal dimension of image. IEEE Trans. Syst., Man, Cybern. **24**(1), 115–120 (1994)
Seneviratne, S., Zhang, X., Adnan, M., et al.: Weather and climate extreme events in a changing climate. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (2021)
Süveges, M.: Likelihood estimation of the extremal index. Extremes **10**(1-2), 41–55 (2007)
The Guardian: 'This is an emergency': Australia's extreme weather crises spark anger at climate inaction. Available at https://www.theguardian.com/australia-news/2022/apr/09/this-is-an-emergency-australias-extreme-weather-crises-spark-anger-at-climate-inaction. Accessed 08 July 2022 (2022)
The New Daily: 'Very unusual': Australia's east coast cities are feeling their coldest start to winter in decades. Available at https://thenewdaily.com.au/news/national/2022/06/09/cold-winter-weather-australia. Accessed 08 July 2022 (2022)
Ukkola, A.M., De Kauwe, M.G., Roderick, M.L., et al.: Robust future changes in meteorological drought in cmip6 projections despite uncertainty in precipitation. Geophys. Res. Lett. **47**(11), e2020GL087,820 (2020)
Van Oldenborgh, G.J., Hendon, H., Stockdale, T., et al.: Defining el niño indices in a warming climate. Environ. Res. Lett. **16**(4), 044,003 (2021)
Welch, B.L.: The generalization of 'student's'problem when several different population varlances are involved. Biometrika **34**(1-2), 28–35 (1947)
Wilby, R.L., Keenan, R.: Adapting to flood risk under climate change. Prog. Phys. Geogr. **36**(3), 348–378 (2012)
Wolff, N.H., Mumby, P.J., Devlin, M., et al.: Vulnerability of the great barrier reef to climate change and local pressures. Glob. Chang. Biol. **24**(5), 1978–1991 (2018)
Woodruff, J.D., Irish, J.L., Camargo, S.J.: Coastal flooding by tropical cyclones and sea-level rise. Nature **504**(7478), 44–52 (2013)

Appendix C

Article published in *Climate Dynamics*

Climate Dynamics
https://doi.org/10.1007/s00382-023-07022-z

ORIGINAL ARTICLE

Decrease of the spatial variability and local dimension of the Euro-Atlantic eddy-driven jet stream with global warming

Robin Noyelle[1] · Vivien Guette[1] · Akim Viennet[1] · Bénédicte Colnet[4] · Davide Faranda[1,2,3] · Andreia N. S. Hisi[1] · Pascal Yiou[1]

Received: 8 November 2022 / Accepted: 7 November 2023

Abstract
The atmospheric eddy-driven jet stream is one of the main features of the mid-latitude circulation. Although mostly zonal in climatological mean, the jet stream meanders at meteorological time scales. The jet and its variability have been under great scrutiny in the past years for their role in the triggering of extreme events (e.g. heat or cold waves) in mid-latitudes regions. Because of the large variability of the jet, the impact of climate change remains elusive. Here we study the eddy-driven jet stream over the Euro-Atlantic sector and assess its dynamical properties in ERA5 and ERA20C reanalysis data set using indicators from dynamical system theory. We control for global modes of variability and aerosols emissions to disentangle the impact of global warming from the impact of natural variability of the climate system on the jet. We find that over the period 1900–2010, global warming decreased the local dimension and spatial variability of the jet. This decrease in variability is connected to an increase in jet persistence and speed. We additionally observe a poleward shift of the jet. Our results suggest a zonalisation of the jet under global warming. This evolution is more pronounced in summer than in winter.

Keywords Jet stream · Global warming · Inter-decadal variability

1 Introduction

Jet streams are narrow, fast-flowing westerly air currents in the troposphere. They are a major feature of the large-scale atmospheric circulation and modulate the frequency, severity and persistence of weather events across the extratropics (Charney 1947; Holton 1973; Hurrell and Deser 2010). Two types of atmospheric jets can be identified: thermally driven subtropical jets associated with the eastward deflection of the upper branch of the Hadley cell (Held and Hou 1980), and eddy-driven jets caused by the transfer of energy from baroclinic eddies to the mean flow at the polar front (Held 1975; Rhines 1975). Real jets may arise from a combination of these mechanisms and thermally and eddy-driven jets are

Robin Noyelle, Vivien Guette and Akim Viennet contributed equally to this work.

✉ Robin Noyelle
robin.noyelle@lsce.ipsl.fr

Vivien Guette
vivien.guette@lsce.ipsl.fr

Akim Viennet
akim.viennet@lsce.ipsl.fr

Bénédicte Colnet
benedicte.colnet@inria.fr

Davide Faranda
davide.faranda@lsce.ipsl.fr

Andreia N. S. Hisi
andreia.hisi@lsce.ipsl.fr

Pascal Yiou
pascal.yiou@lsce.ipsl.fr

[1] Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL & Université Paris Saclay, 91191 Gif-sur-Yvette, France

[2] London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK

[3] Laboratoire de Météorologie Dynamique, IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, 75005 Paris, France

[4] INRIA, Soda Project-Team, Premedical Project-Team, Saclay, France

Published online: 06 December 2023

actually two limits in a continuous spectrum (Lee and Kim 2003; Spensberger and Spengler 2020; Messori et al. 2021).

Even though the climatological eddy-driven jet is mostly zonal flowing to the east, on a daily basis it can present large meanders. In these cases, the local flow becomes predominantly meridional or can even split or break. Those meanders have a typical spatial and temporal variability of a few thousand kilometers and of 10 days (Röthlisberger et al. 2016). The meanders allow air masses coming from the south or the north to persist around mid-latitude regions, potentially triggering temperature or precipitation extremes (Kautz et al. 2022). For its role in the triggering of extreme events in mid-latitude regions, the eddy-driven jet has been under great scrutiny in the past years.

The variability of the eddy-driven jet stream is an example of the large spontaneous variability of the climate system. The inter-decadal jet variability is important, with decades of strong and steady jet being interspersed with decades of a weak and more variable jet (Woollings et al. 2018; Simpson et al. 2019; Osman et al. 2021). Therefore, identifying the impact of global warming on the jet has remained elusive and controversial (Barnes and Screen 2015).

Even though there is a broad agreement on the poleward shift of the tropospheric subtropical jets (Pena-Ortiz et al. 2013; Gulev et al. 2021), the impact of global warming on the Euro-Atlantic eddy-driven jet mean latitudinal position and other characteristics is still a matter of debate (Woollings et al. 2014; Oudar et al. 2020; Lee et al. 2021; Stendel et al. 2021). Indeed, the jet has been caught in the 'tug-of-war' (Held 1993) between two competing phenomena: the Artic amplification (AA) and the tropical upper-tropospheric warming.

As the Artic is warming more rapidly than the rest of the world (Cohen et al. 2018)—reducing the Arctic-to-mid-latitude temperature gradient—it has been argued that this could lead to changes in the configuration of the jet stream (Francis and Vavrus 2015). The reduced equator-to-pole temperature gradient could weaken the predominant westerly winds, which, in turn, could cause larger-amplitude waves in the midlatitude circulation. However, global warming also leads to tropical upper-tropospheric warming, which would in contrary act to increase the equator to pole temperature gradient, reinforcing the jet (Stendel et al. 2021). Another mechanism could increase the waviness of the jet: the increased land-sea gradients (Portal et al. 2022), as supported by the recent theoretical work of Moon et al. (2022).

Changes in the latitudinal or longitudinal temperature gradients are however not the only mechanisms through which the dynamics of the jet could have changed during the twentieth century. Antropogenic aerosols emissions and internal variability of the climate system are two other competing factors. Several studies have demonstrated the role of antropogenic aerosols in changing the dynamics of the North Atlantic atmosphere (e.g. Pausata et al. (2015); Diao and Xu (2022); Murakami (2022); Robson et al. (2022)) through anomalous heating or cooling in the mid-latitudes. Low frequency variability of the ocean is another confounding phenomenon. The Atlantic Multi-decadal Oscillation (AMO), a 60–80 year basinwide quasi-oscillation in North Atlantic sea surface temperatures (Kerr 2000), could be invoked to explain the inter-decadal changes in the dynamics of the jet. The variations in the coupled El-Nino–Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) could also influence the variability of the jet through the generation of Rossby wave trains (Ding et al. 2017; Jiménez-Esteve and Domeisen 2018; Mezzina et al. 2020).

Studies on reanalysis data and climate models projections have shown conflicting results on the evolution of jets under global warming with some of them concluding to a weakening of the North Atlantic jet (Francis and Vavrus 2015; Coumou et al. 2015; Harvey et al. 2020), while others concluded to a stronger jet under global warming (Iqbal et al. 2018; Tenenbaum et al. 2022; Hallam et al. 2022). Some studies have also targeted directly the measurement of the "waviness" of the mid-latitude circulation, with various metrics (Cattiaux et al. 2016; Peings et al. 2018; Blackport and Screen 2020). A common theme of those studies is that the natural variability of the jet stream may be a sufficient explanation to the recent observed increases of its waviness (Osman et al. 2021; Blackport and Screen 2020). Therefore, the recent observed covariability between waviness and temperature gradients on interannual to decadal time scales may not represent a forced response.

Here, we address the question of quantifying the changes in the characteristics of the Euro-Atlantic eddy-driven jet stream due solely to global warming over the twentieth century.

Several approaches have been developed to detect the eddy-driven jet stream position from wind and pressure maps, each with advantages but also limitations. Some algorithms were developed to capture the 3D (Limbach et al. 2012) or 2D structure (Molnos et al. 2017; Spensberger et al. 2017) of the jet. A large part of the literature investigating the jet variability however reduces the jet to a single point characterized by a latitude, the so-called Jet Latitude Index (JLI), and a wind speed by finding the point where the zonally averaged low-level zonal wind is maximum (Woollings et al. 2010). This method provides useful insights and is very helpful to make time-series statistics using only two indexes, nonetheless it misses key geometric features of the jet such as the omega-shaped pattern associated with blocking events. Here we adopt a more geometric view by considering the latitudinal position of the jet at each longitude (Faranda et al. 2019b).

We analyze this representation of the jet with recently developed indicators (Lucarini et al. 2016; Faranda et al. 2017; Messori et al. 2021) based on dynamical systems theory. We expand the results of Faranda et al. (2019a) and Rodrigues et al. (2018) who have shown the interest of using time series of these indicators to detect changes in the large scale circulation. Here we target specifically the eddy-driven jet. We finally propose a linear model where we control for the influence of confounding factors (Kretschmer et al. 2021)—internal variability and anthropogenic aerosols emissions—and quantify the sole effect of global warming on various indicators of the jet variability.

This paper is organized as follows: in Sect. 2, we introduce the data used and the methods developed to detect the jet stream position and quantify its variability. In Sect. 3, we assess the ability of our indicators to characterize the jet variability by investigating how they are related to other jet characteristics. We also show the relevance of targeting specifically the jet as an atmospheric feature rather than studying sea-level pressure or geopotential maps. In Sect. 4, we investigate the inter-decadal variability of the eddy-driven jet over the twentieth century and relate it to classical indices of natural variability. Finally, the discussion of the results and the conclusions drawn from our analysis are presented in Sect. 5.

2 Data and methods

2.1 Data

The analyses proposed here are based on the ERA5 reanalysis data of the European Centre for Medium range Weather Forecasts (Hersbach et al. 2020). We use daily averaged fields with a 0.25° horizontal resolution over the 1950–2021 period for the Euro-Atlantic region from 22.5°N to 70°N in latitude and from 80°W to 50°E in longitude. For disentangling the role of natural and forced response on the low-frequency evolution of the jet variability, we use the ERA20C reanalysis dataset, which is the twentieth century reanalysis of the ECMWF (Poli et al. 2016), over the 1900–2010 period with the same spatial extension. This dataset has a resolution of 1.125°. For both data sets, the variables considered are the daily-averaged geopotential height at 500 hPa (Z500), sea-level pressure (SLP) and horizontal wind speed between 850 and 700 hPa.

For quantifying the link between the variability of the jet position, classical indices of natural variability and global warming, we use monthly indices downloaded from the Climate Explorer web tool (https://climexp.knmi.nl/). For the AMO, we use the AMO index of the Met Office Hadley Centre/Climatic Research Unit (van Oldenborgh et al. 2009). This index corresponds to the average of monthly Sea Surface Temperature (SST) anomalies with respect to the ensemble mean of the reanalysis over the North Atlantic. For the PDO, we use the PDO index of the Hadley Center based on an EOF decomposition of Pacific SSTs. For the ENSO we take the Nino 3.4 index, which is the area averaged SST from 5°S–5°N and 170–120°W (Van Oldenborgh et al. 2021). For quantifying global warming, we use the monthly global mean Earth surface temperature anomalies (relative to the 1961–1990 period) provided by the Hadley Centre (HadCRUT5 data set, Morice et al. (2021)). For the AMO, PDO and ENSO indices a linear trend has been removed to account for the global warming signal.

We further consider the impact of aerosols forcing. We use the ambient aerosol absorption optical thickness at 550 nm provided at a monthly time scale by the IPSL model (Dufresne et al. 2013) under the CMIP5 historical configuration for the 1900–2005 period and under the RCP4.5 scenario for the period 2006–2021. The optical thickness is provided at each grid point and we average the field over an extended sector encompassing most of North America (125° W–50° E and 22.5° N–70° N) as aerosols emitted there could influence the jet by cooling the land and increasing land-sea contrast (Robson et al. 2022). The ERA5 and ERA20C reanalyses both use the CMIP5 forcing files for aerosols concentration (Poli et al. 2016; Hersbach et al. 2020), which consists of monthly 10 year-averaged files.

2.2 Detecting the jet position

For detecting the jet position, we first average over 850–700 hPa pressure levels the horizontal wind speed. Contrary to Woollings et al. (2010), we investigate the variability of the jet not only over the North Atlantic ocean but also over continental Europe (80° W–50° E). In order to avoid boundary layer effects over continental Europe we prefer to begin the averaging process at a higher pressure level: 850 hPa rather than 925 hPa. We checked that the detection of the jet position is not sensitive to averaging between 850–700 hPa or 925–700 hPa over the Euro-Atlantic sector. We then apply a 10 days low-pass Lanczos filter with a window of 61 days to remove the influence of transient eddies (Duchon 1979). At this point we do not apply a zonal mean but take a two-step approach close to the procedure used by Faranda et al. (2019b). The first step consists in finding, for each longitude, the latitude at which the wind horizontal kinetic energy $E = \frac{1}{2}\vec{u}_H^2$ is maximum. The second step is to apply a longitudinal rolling median with a 25° window to the previous positions. This rolling median is applied to avoid a nonphysical detection of breaks in the jet. With a 0.25° horizontal resolution for ERA5 and considering the low-level jet, the algorithm sometimes detects high-wind speeds in the lee of mountains. 25° of longitudes approximately corresponds

to 2000 km at 45° N, which is also the typical size of mid-latitude baroclinic disturbances (Hoskins and James 2014) so that we consider that this rolling median has a physical basis. An example of the jet position found with this method is given in Fig. 1.

Our approach considers the jet position as a vector of positions indexed by longitudes. To study time series, we however need lower-dimensional objects. The following sections present the indicators we used.

2.3 Dynamical indicators

In physics, dynamical systems can be defined as objects whose states vary with time. In atmospheric physics in particular, fields such as sea-level pressure, temperature or precipitation can be considered as observables of a dynamical system, namely the atmospheric flow (Lucarini et al. 2016; Faranda et al. 2017). In our case, the observable we are studying is the jet position at each longitude, going through different daily states, noted ζ. A state of our system can be described as a vector of dimension *nlon* where *nlon* represents the number of grid points along the longitudinal axis. The yth position in this vector contains the latitudinal value of the jet position for the yth longitude. The ensemble of states of the jet at all time approximate the dynamics of the atmospheric flow and should retain some of the properties of the full, high dimensional, attractor of mid-latitude atmospheric motions.

To fully characterize the properties of this attractor, one needs to know how often each state ζ occurs over a certain time interval and how long the dynamics "sticks" to ζ before leaving its neighborhood. Those two quantities are characterized by the local dimension d and the persistence θ^{-1}. The distribution of instantaneous dimensions of the attractor of a dynamical system is a proxy for the predictability of observed states because it is related to the Lyapunov exponents (Young 1982). The persistence around one state gives instantaneous measures of the tendency of the dynamics to stay around one state. Therefore, estimating the distributions of these metrics in the phase space defined by the jet positions helps characterizing the overall dynamics of this system.

d is a proxy for the system's active number of degrees of freedom when reaching a region of phase space. Thus, even when considering a system with a large number—possibly infinite—of dimensions, d provides the local number of dimensions that the system can be summarized to. Therefore a state 1 with a local dimension d_1 greater than the local dimension d_2 of another state 2 means that the behavior of the system around state 1 has more dimensions on which to evolve and is therefore less predictable than around state 2 (Messori et al. 2017; Hochman et al. 2019). Additionally, Pons et al. (2020) showed that d can be used as a measure of co-dependance: a high degree of synchronization between the variables defining the system is associated with a low value of d. The second dynamical system indicator we use is the persistence θ^{-1} of a given state ζ, which is equivalent to the mean residence time of the trajectories when they enter the neighborhood of ζ. This metric corresponds to the inverse of a well defined statistical quantity introduced in extreme value statistics, namely the extremal index θ. Note that in the framework of dynamical systems, we find $\theta = 0$ at stable fixed points of the dynamics (the trajectory resides an infinite amount of time in the neighborhood of this state), with an infinite number of infinitely time resolved trajectories. Conversely, $\theta = 1$ is found at non persistent states of the dynamics (see Moloney et al. (2019) for more details). In general, for time-continuous systems sampled at a given resolution dt, $\theta^{-1} > 1$. For daily sea-level pressure fields over the North Atlantic, Faranda et al. (2017) found θ^{-1} values varying between 2 and 3 days. One may note that these values depend on the size and the time step of the data set used, and on the chosen percentile q. Therefore the local persistence θ^{-1} is to be used to compare different states within the same data set.

Both of these indicators are computed using the fact that the probability for a recurrence of a system configuration (a state) can be linked to the generalized Pareto distribution (Pickands 1975). To compute this probability from data, we

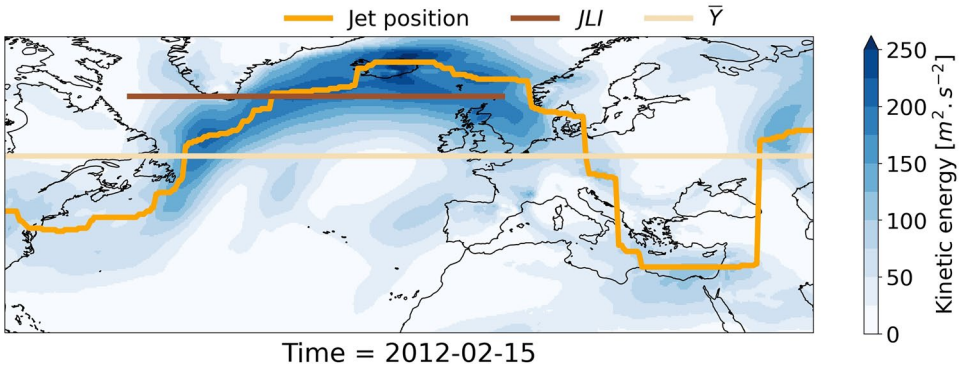


Fig. 1 Snapshot of the horizontal wind kinetic energy and jet position for one example day. Wind horizontal kinetic energy $E = \frac{1}{2}\vec{u}_H^2$ (colors) and jet position found by the algorithm (orange line). The yellow and brown lines represent respectively the jet latitude index *JLI* and the mean position of the jet $\bar{Y}$ found with our method

compute the series of distances $\text{dist}(x(t), \zeta))$ between a state of the system ζ and all other points $x(t)$ on the trajectory of the system. This time series of distances is then transformed into: $g(t) = -\log(\text{dist}(x(t), \zeta))$ so that being close to state ζ is equivalent to exceeding a threshold $s(q)$ where q is a percentile of the series $g(t)$. We use the 98% percentile of all values of $g(t)$, which ensures to have enough data while keeping only the extremes. For the calculation of the distances between states, we use the Euclidian distance. It can be shown that the probability distribution of $g(t)$ when it exceeds $s(q)$ converges to a Pareto distribution (Lucarini et al. 2016) with scale parameter σ, and a shape parameter $\xi = 0$. The local dimension d is practically estimated as the inverse of the scale parameter of the generalized Pareto distribution fitted on the data which satisfies $g(t) > s(q)$. The persistence is here estimated using the (Süveges 2007) estimator on the time series $g(t)$.

2.4 Other variability indicators

To the local dimension d and persistence θ^{-1} indicators we add three other indicators of the jet stream state for analysing its variability. For each day we compute:

- the jet mean position $\bar{Y}$, defined as the zonal average of the jet positions found with the detection algorithm. As illustrated in Fig. 1 one should note that this indicator is not equivalent to the JLI for two reasons: (1) our indicator is computed on the full Euro-Atlantic sector and not only over the North Atlantic and (2) the mean position is computed without applying a zonal average to zonal wind speed;
- the jet mean speed $\bar{U}$, defined as the zonal average of the norm of the horizontal wind vector $\vec{u}_J$ at the jet position: $\bar{U} = \frac{1}{nlon} \sum_{lon} \sqrt{u_J^2 + v_J^2}$. For each longitude, the horizontal wind vector $\vec{u}_J$ is computed as the mean of horizontal wind vectors within a 2.5° latitudinal extent around the jet position. Again, this indicator is not equivalent to the zonal jet speed as defined in Woollings et al. (2010);
- the jet waviness W, defined as the longitudinal standard deviation of the jet position. Before computing this parameter, we remove a linear longitudinal trend to the jet position as the jet has a tendency to present a northward tilt over the North Atlantic. This indicator is always positive and indicates a jet without (with) meanders when taking low (high) values.

In addition to our indicator $\bar{Y}$ and $\bar{U}$, we also use the more classical jet latitude index *JLI* and zonally-averaged zonal wind speed U_{JLI}. We note that an increase in the indicator $\bar{U}$ can happen even if the maximum speeds of the jet overall do not change. This can be interpreted as an increase in the extension of the jet inside continental Europe, contrary to an increase in U_{JLI}. Nevertheless, we also note that U_{JLI} is computed using zonal wind only, contrary to $\bar{U}$, therefore if the jet is more zonal U_{JLI} will increase and this may not reflect an increase in the maximum speeds either.

There is no universally accepted metric to measure the "waviness" of mid-latitude circulation and therefore several measures have been used in the literature (Blackport and Screen 2020). Our measure of waviness W is straightforward to compute, intuitive and gives easily interpretable results. We note that our approach is close to the sinuosity metric proposed by Cattiaux et al. (2016).

2.5 Causal networks theory

Using only pairwise correlations between jet metrics and global mean temperature to quantify the past changes in the jet characteristics due to global warming bears the risk of missing to control for some important variables and therefore leading to a wrong estimation of its effect (Runge et al. 2014; Kretschmer et al. 2016). Here we want to estimate what is the response of the mean values of key metrics characterizing the jet for a given state of the Earth system defined by its global mean temperature T. However, T and natural oscillations—ENSO for instance—are affecting the jet and are simultaneously changing. Only looking at the effect of T without controlling for ENSO would over or under estimate its effect on the jet. To quantify the evolution of jet metrics over the last century, we formalize our hypotheses using so-called causal inference theory (Kretschmer et al. 2021).

In causal statistics (Pearl 2009; Pearl and Mackenzie 2018), a causal inference from a process X to another process Y means that intervening in X while keeping everything else fixed changes the probability distribution of Y (Glymour et al. 2016). This "interventional" probability distribution may be different from the conditional probability distribution of Y on X in general. Intervening in the climate system is generally not possible, but quantifying causal effects to predict the effects of interventions based on observational data is possible considering that naturally occurring "interventions" in X that led to changes in Y are present in the data but are confounded by other processes that affect both X and Y. To isolate the causal effect from X to Y, one therefore needs to take into account the influence of such confounders.

Causal inference theory is a process-based framework which can formalize hypotheses about physical mechanisms in the form of a causal network (e.g. Hannart et al. 2016; Jiménez-Esteve and Domeisen 2018; Lee et al. 2019; Hirt et al. 2020; Monnin et al. 2022). Causal networks are Directed Acyclic Graphs (DAG) consisting of

physical variables involved in the process (nodes) and indications of the presence and direction of assumed causal relationships between these variables (links). These networks represent plausible models of the data-generating processes. It is not feasible to represent all the physical processes involved in the climate system in a causal network, therefore one can use only a causal network limited to the processes of importance in the case studied. The key point is to include all common drivers of X and Y, which could confound the causal effect of X on Y, and omit variables being consequences of X (mediators) or Y (colliders).

In a second step, one chooses how to estimate the proposed network. Here we use simply multi-linear regression but other methods are possible (Kretschmer et al. 2021). We stress that while the use of non-zero lags may in some cases be helpful, these are not an essential part of the causal network framework: as explained above, the key point is to represent the effect of confounding variables, which can be done in a model with no lags.

We emphasize that the "causal effect" quantified with this analysis is causal only if the proposed data-generating process formalized in the DAG is correct, which is always unclear. For example, it is hard to guarantee that all relevant confounders have been taken into account, and omission of any one would make any strict claim of causality impossible. In the following we make use of available climate literature to choose one reasonable data-generating process. We emphasize that the quantification that follows from this choice is "causal" only in the sense described here, and with these provisos in mind.

3 Diagnosing the jet variability

Faranda et al. (2017) made the case for the use of dynamical indicators to investigate the weather variability over the Euro-Atlantic sector with the tools presented in 2.3 using SLP as a state vector. Here we specifically target the jet position (Faranda et al. 2019b; Messori et al. 2021). Figure 2 presents the cross distribution of d and θ^{-1} computed on SLP, Z500 and jet positions reanalysis data over the Euro-Atlantic sector. The dynamical indicators for SLP and Z500 are correlated (Pearson correlation coefficient $r = 0.45$ for d and $r = 0.71$ for θ^{-1}), which was expected in so far as those fields carry similar information about the synoptic meteorological state of the atmosphere. Conversely, almost no correlation is observed between the dynamical indicators computed on the jet position and the one computed on the SLP or the Z500. According to this analysis the persistence/predictability of a jet position is unrelated to the persistence/predictability of the corresponding SLP/Z500 pattern. The reason is that even though the jet is one of the main structures of the Euro-Atlantic circulation, other structures are present in the SLP/Z500 pattern which influence its persistence and its predictability (Dorrington and Strommen 2020).

The absolute values of the indicators can be compared from one data set to another in so far as they have a similar

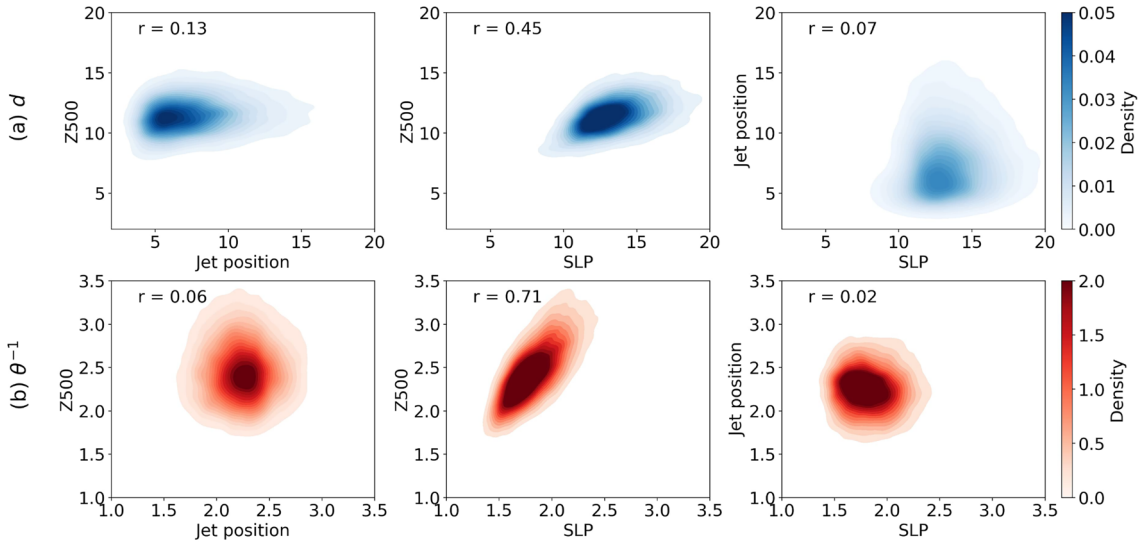


Fig. 2 Cross distributions of dynamical indicators for Z500, SLP and jet positions (ERA5). **a** Cross distributions of the local dimension d between the Z500 field and jet position, the Z500 field and the SLP field, and jet position and the SLP field. **b** Cross distributions of the local persistence θ^{-1} between the Z500 field and jet position, the Z500 field and the SLP field, and jet position and the SLP field. For computational reasons, the dynamical indicators on the Z500 and the SLP fields have been computed with a spatial resolution of 0.5° instead of 0.25°

sampling frequency. The mean local dimension found for the ERA5 data set is 7.6. With a different spatial resolution (0.25° vs 1.125°), we find a mean value of 7.5 for the ERA20C data set. These values must be compared with the values of 11.4 and 13.5 for the mean local dimension of the Z500 and SLP fields on the ERA5 data set. For the persistence, we find values evolving in the same range for the three variables (between 2 and 3 days).

Figure 3c displays the cross distribution of points for the dynamical indicators d and θ^{-1} computed on the jet position vectors. The marginal distributions are shown in Fig. 3a and d. The points are colored with respect to which tercile of the distribution of the waviness indicator W they belong to (Fig. 3b). The less wavy jets are associated with high values of θ^{-1} and low values of d while the most wavy jets are found to have lower values of θ^{-1} and higher values of d. These results are consistent with what Messori et al. (2021) found using an idealised quasi-geostrophic model and computing indicators on the wind field. Figures A1 and A2 in Online appendix present the same analysis for the $\bar{U}$ and $\bar{Y}$ indicators. Contrary to Fig. 3, the d and θ^{-1} indicator have difficulties discriminating the jet dynamical behavior when using the mean speed and mean position. This analysis suggests that the spatial variability in the jet pattern is not well captured by the indicators $\bar{U}$ and $\bar{Y}$ contrary to d and θ^{-1}.

As in Faranda et al. (2017) and Messori et al. (2017), to better understand what kind of dynamical information the indicators reveal, we analyze the extreme cases of the d and θ^{-1} indicators. We take the composite maps over the days belonging to the top 2% in term of one of the indicator and for which the other one is not extreme (in the sense that it does not belong to the top 2% neither to the bottom 2% quantiles). The results shown are not sensitive to moderate changes in the value of this quantile. We define four situations: low local dimension ($N = 377$), high local dimension

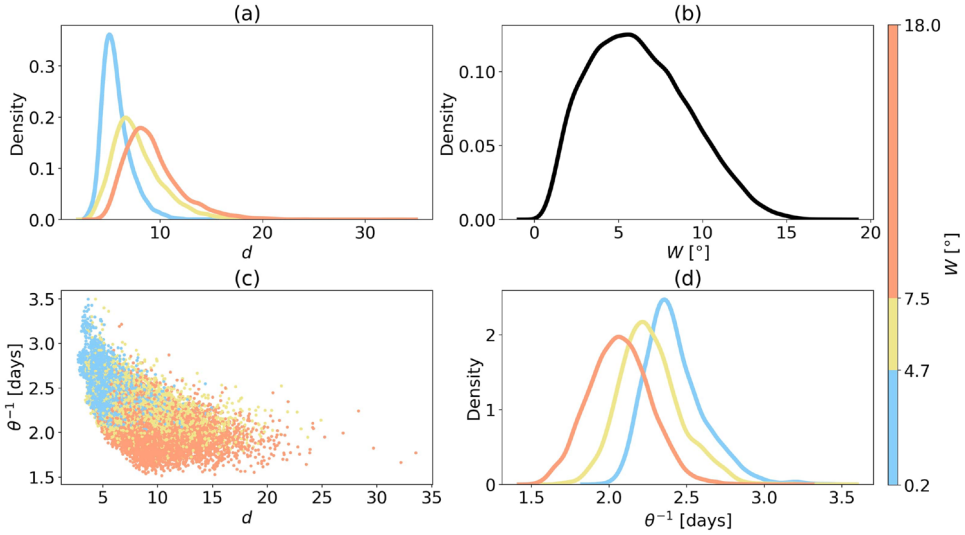


Fig. 3 Link between the dynamical indicators and the jet waviness (ERA5). **a** Distribution of local dimension d for the three terciles of the waviness W. **b** Distribution of W. **c** Cross distribution of d and θ^{-1} colored by the tercile of the W indicator. **d** Distribution of local persistence θ^{-1} for the three terciles of the waviness W

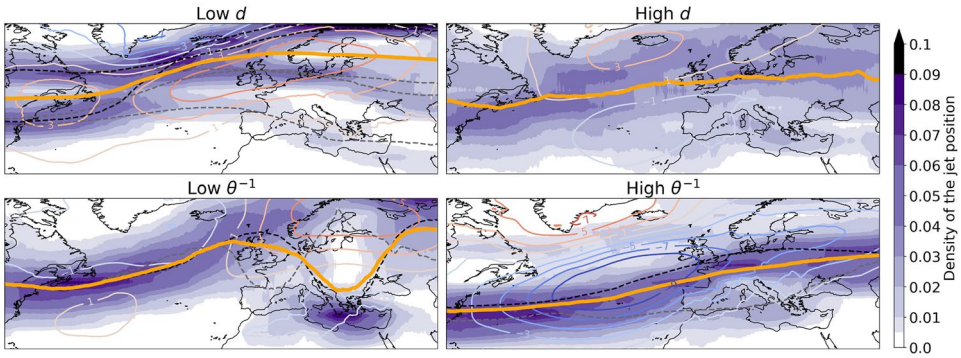


Fig. 4 Extreme d and θ^{-1} situations (ERA5). Composite maps of SLP anomalies in hPa (contours), density of jet positions (colors) and associated mean position (orange lines) for the days with extreme values of the local dimension d and the local persistence θ^{-1}. The black dashed lines are the main modes (more than 30% of the trajectories) and the gray dashed lines are the minor modes (less than 30% of the trajectories)

($N = 458$), low local persistence ($N = 457$) and high local persistence ($N = 357$), which are presented in Fig. 4.

In the high local dimension case, the synoptic meteorological situation shows no clear pattern, and the density of the jet position is widespread. This behavior is consistent with the concept of local dimension: for points with a very high local dimension, the system has many degrees of freedom and can correspond to many dynamical situations. In the low local dimension case, the density of the jet stream position shows a bi-modality over the eastern part of the American continent and a tri-modality over the European continent, leading to five distinct clusters of trajectories. The two main modes represented by black dashed lines count for approximately 60% of all the trajectories. The synoptic situation displays a large anticyclone over North-Western Europe reminiscent of the NAO+ weather regime but is associated with several possible patterns of the jet.

In the low local persistence case, the anomalies of SLP show a pattern, close to the Scandinavian Blocking situation. The averaged jet pattern is peculiar but nonetheless well defined—in so far as there is a low dispersion of the jet position density—and has a pronounced dip over the Mediterranean Sea in its main mode (80% of the trajectories). This very wavy situation of the jet is reminiscent of recent synoptic situation corresponding to temperature extremes over Western Europe (e.g. Mitchell et al. (2019)). Finally, the high local persistence case is reminiscent of the NAO- situations with a positive SLP anomaly over Greenland and a negative SLP anomlay over the Azores (Michelangeli et al. 1995).

We now turn to using these indicators to study the recent past evolution of the Euro-Atlantic eddy-driven jet stream and its relation to usual modes of variability of the climate. In the following, we mainly display results based on the ERA20C data set because it spans a longer time period. The corresponding results for the ERA5 data set are presented in Online appendix.

4 Inter-decadal variations of the eddy-driven jet stream

Figure 5a and b present the temporal evolution of the 1-year rolling mean of the d and θ^{-1} indicators computed on the jet position vectors for the ERA20C reanalysis data set over the 1900–2010 period. A Gaussian filter with a cut-off frequency of 5 years was applied on the raw time series for obtaining the smoothed time series. The 5 years cut-off frequency was chosen to average out most inter-annual variability. Both indicators display substantial inter-decadal variability, up to ~30% (6.5–8.5) with respect to its mean for local dimension but only ~10% with respect to its mean for local persistence (3.5–3.8). When computing the power spectrum on the raw time series of our indicators (panel (c)) one sees a strong peak at the 1 year period, which corresponds to the annual cycle. For higher periods, the log-log plot shows that the indicators behave as red noises (straight line), except around the 50 years period where a dominant peak emerges. We ran a similar spectrum analysis over the 1950–2020 period for the ERA5 data set and found a similar peak around the 50 years period (Fig. A3). This peak is also seen when using other indicators (Fig. 6). Nonetheless, a 50 year period in a 110 (or 70) years data set is near the limit of detectability, therefore we cannot assert that this is a relevant feature of the jet variability.

Figure 5a displays a decrease of local dimension from 1970 to 1990 and then an increase up to 2000. This inter-decadal variation is large even after applying a 5-year low pass filter (~10% of the mean value). A well documented phenomenon occurred in the North Atlantic ocean during the period 1970–2000 (Sutton and Dong 2012; Robson et al.

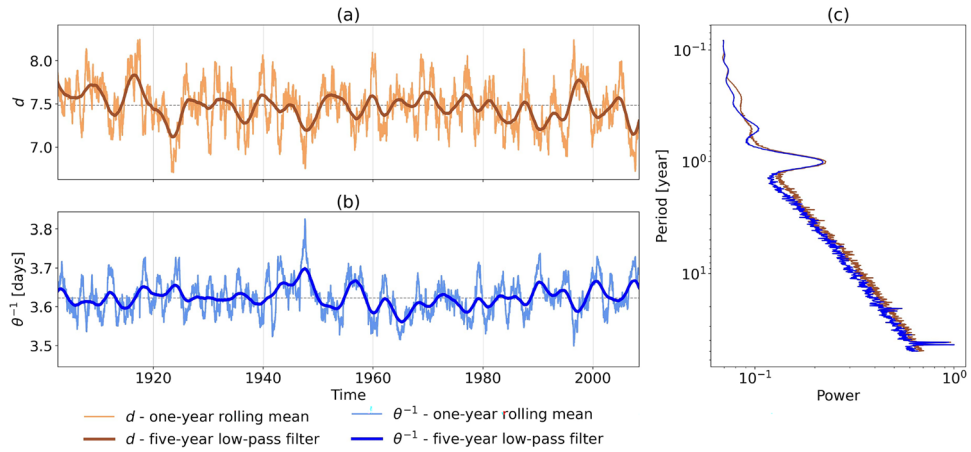


Fig. 5 Time series and power spectrum of d and θ^{-1} (ERA20C). **a** Time series of d with a 1-year rolling mean and a 5-year low-pass filter. **b** Time series of θ^{-1} with a 1-year rolling mean and a 5-year low-pass filter. **c** Power spectrum of d and θ^{-1}

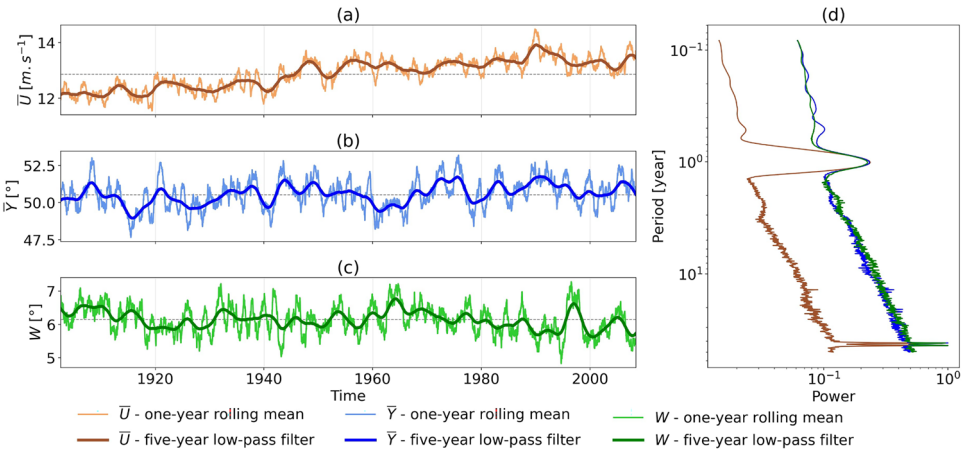


Fig. 6 Time series and power spectrum of $\bar{U}$, $\bar{Y}$ and W (ERA20C). **a** Time series of $\bar{U}$ with a 1-year rolling mean and a 5-year low-pass filter. **b** Time series of $\bar{Y}$ with a 1-year rolling mean and a 5-year low-pass filter. **c** Time series of W with a 1-year rolling mean and a 5-year low-pass filter. **d** Power spectrum of $\bar{U}$, $\bar{Y}$ and W

2016; Jackson et al. 2022). The European climate experienced substantial changes, with anomalously mild and wet summers in Northern Europe, and hot and dry summers in Southern Europe. This coincided with a major warming of the North Atlantic Ocean due to a strengthening of the Atlantic Meridional Overturning Circulation (AMOC). On Fig. A3 we also see an increase of inter-decadal variability between the periods 1950–1980 and 1990–2020.

Figure 6 shows the same plot for the $\bar{U}$, $\bar{Y}$ and W indicators. The most striking feature of this figure is the increase by almost 2 m/s of the mean speed of the jet $\bar{U}$ between 1900 and 2010. We come back to this observation below. The mean speed and position indicators show no peculiar evolution during the 1970–2000 period. The W indicator however seems to display a similar evolution as d. It should be noted that these changes in the observed behavior of the North-Atlantic eddy-driven jet stream between the 1950–1980 and 1990–2010/2020 periods may be due to different data collection processes. It is indeed well-known that the so-called satellite era beginning in the 1980s increased by a large factor the quantity of available data. The large scale structures of the atmosphere are usually well resolved in reanalysis data (e.g. Slivinski et al. (2021)), nonetheless we cannot completely rule out the possibility that the observed changes are not physically relevant features of the jet (see Rodrigues et al. (2018) for a similar discussion). The purpose of this paper is not to evaluate whether these changes are artifacts of the data collection process, but one should be cautious when attributing these results to a real physical behavior of the jet position.

We now wish to estimate the effect of global warming on the characteristics of the jet using our indicators. Isolating the contribution of global warming to the observed changes in the behavior of the jet is not straightforward in so far as the jet displays a large variability and any signal of forced change may therefore be obscured by the influence of other factors. Controlling for these factors would allow to isolate the global warming signal, but one should take care of controlling only for the relevant factors. It is necessary to control only for factors that are confounders of the link between global warming and the variability of the jet (Kretschmer et al. 2021) as explained in Sect. 2.5. Based on existing literature (Newman et al. 2016; Levine et al. 2017; Lin and Qian 2022), we propose the DAG depicted in Fig. 7. This graph summarizes our hypotheses for quantifying the impact of global warming on the jet on inter-decadal timescales. It does not include all possible phenomenon influencing the jet but rather those that we hypothesise are crucial for confounding the influence of the global warming index T.

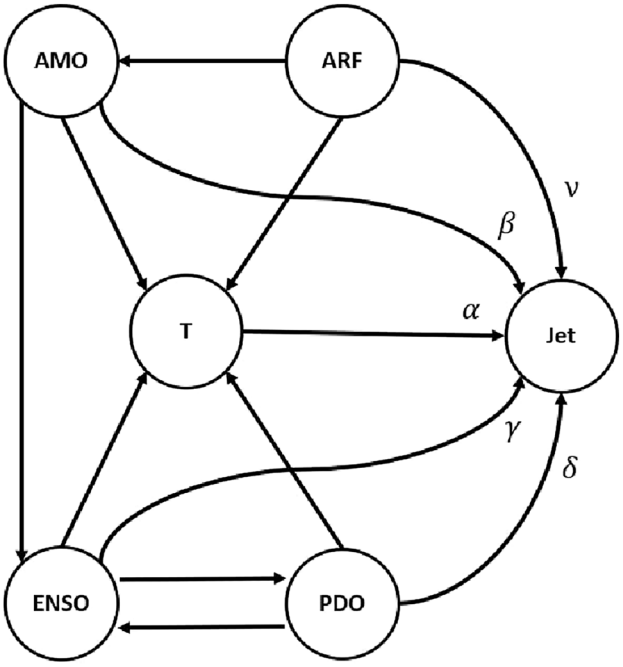


Fig. 7 Estimated DAG of the influence of global warming on the jet. AMO stands for Atlantic Multi-decadal Oscillation, ENSO for El-Nino Southern Oscillation, PDO for Pacific Decadal Oscillation, ARF for Aerosols Radiative Forcing and T for global mean temperature of the Earth

Potential confounders of the variability of the North-Atlantic eddy-driven jet stream and global warming are supposed to be the AMO, the ENSO, the PDO and Aerosols Radiative Forcing (ARF). The variability of the jet is strongly linked to anomalies of SSTs in the North Atlantic (Simpson et al. 2018), therefore it is natural to posit an impact of the AMO on the jet, even though here we do not precise what is the exact physical phenomenon by which the AMO influences the jet. We note that whether the AMO is a true oceanic oscillation is still a subject of scientific debate (Mann et al. 2021), but we assume it is the case here. The impact of the ENSO and the PDO on the Euro-Atlantic climate are also well documented (Ding et al. 2017; Jiménez-Esteve and Domeisen 2018; Mezzina et al. 2020), and these two modes influence the inter-decadal variability of the global temperature of the Earth (Foster and Rahmstorf 2011). To the natural variability of the ocean presented by the AMO, the ENSO and the PDO, we add the ARF over the Euro-Atlantic sector in so far as the changes in the emissions of aerosols can alter the global circulation (Pausata et al. 2015; Diao and Xu 2022; Murakami 2022). To measure this quantity, we average over an extended spatial region (125° W–50° E and 22.5° N–70° N) the aerosols optical thickness at 550nm computed using the IPSL-CM5 model (Dufresne et al. 2013) which uses the same forcing files as the ERA20C and ERA5 reanalysis. As documented by Qin et al. (2020), ARF impacted the AMO over the twentieth century, but this mode still possesses its own variability. We stress that in the DAG proposed in Fig. 7, even if the influence of the previous confounders on T is small or even nonexistent, including them in the regression would not bias the estimation of the impact of T on the jet and in the contrary would help to decrease its variance.

One may note that the model we are estimating does not show any feedback loop between global warming and the AMO, the PDO and the ENSO. The SSTs of the oceanic regions on which those indicators are computed did change during the twentieth century in response to the radiative forcing caused by anthropogenic emissions, but here we use the detrended time series of these indicators. We therefore explicitly assume that the global increase of temperature on the Earth had no impacts on the natural evolution of the AMO, PDO and ENSO indicators over the period studied that is not taken into account when removing a linear trend (Trenberth and Shea 2006). For small increases of the global temperature, this hypothesis is reasonable. However, when extrapolating our results in the future with a much larger warming level, this hypothesis may prove to be wrong.

We use monthly-averaged normalized time series and we apply low-pass Gaussian filters with 2, 5 and 10-year cut-off frequencies to ensure robustness with respect to the time-filtering procedure. The rationale for using low-pass filters with cut-off frequencies greater than 2 years is to remove any variations that are irrelevant at the inter-decadal time scale, the main one being the annual cycle. For the filters with 5 and 10-year cut-off frequencies, it may be possible that the ENSO signal is filtered out and therefore we will mainly interpret the results found using the 2-year filter. For simplicity reasons, we choose to estimate a linear model. The model is the following:

$$\text{Jet} = \alpha \text{T} + \beta \text{AMO} + \gamma \text{ENSO} + \delta \text{PDO} + \nu \text{ARF} + \epsilon. \quad (1)$$

In Eq. (1), the "Jet" variable represents the different indicators on which we compute the regression and T is the global warming index. The ϵ term represents a noise term supposed to be normally distributed. Even though controlling for the AMO, the PDO, the ENSO and the ARF is essential to estimate the response to an increase of T, we stress that their associated coefficients cannot be interpreted as a total causal effect of their respective phenomena on the variability of the jet. We are estimating only a limited part of the DAG, therefore we only interpret the α coefficient associated with the T parameter.

Figure 8 displays the results of the estimation of the α coefficient in Eq. (1) for the local dimension d, the local persistence θ^{-1}, the mean speed $\bar{U}$, the mean position $\bar{Y}$, the waviness W, the jet latitude index JLI and the zonal wind speed at the jet latitude index U_{JLI} using the ERA20C data set over the period 1900–2010. All estimated α coefficients are significant at the 0.1% level ($p < 0.001$) when estimating confidence intervals with a maximum likelihood estimator (see Table A1). Results are stable when applying different cut-off frequencies for the low-pass filter. For illustrative purposes, Fig. 8 also shows in red the regression without controlling for the potential confounders.

We find that global warming is associated with a decrease of the local dimension d and the waviness W of the jet position and with a slight increase of its persistence. We also find that global warming is associated with an increase in the mean speed $\bar{U}$ and the zonal wind speed U_{JLI} at the JLI. The increase in $\bar{U}$ may reflect both an increase in the maximum jet speed and an increase in the extension of the jet inside continental Europe. The increase in U_{JLI} in contrast is not related to the extension of the jet but also incorporates a change due to the zonalisation of the jet, because it can increase even if the total kinetic energy of the jet does not change. We do not disentangle these effects here but it is likely that they all play a role (Oudar et al. 2020).

Finally, global warming is associated with a poleward shift of the position of the eddy-driven jet, measured using either the mean position $\bar{Y}$ or the JLI. When going back to dimensionalized units, we estimate that a 1K increase

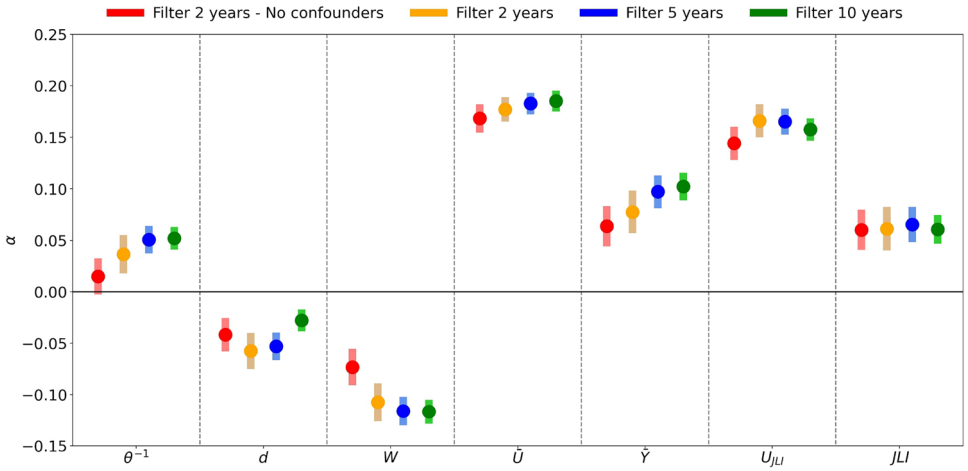


Fig. 8 Estimation of the impact of global warming on indicators of the jet variability (ERA20C). The plot represents the estimated α coefficient for the local dimension d, the local persistence θ^{-1}, the mean speed of the jet $\bar{U}$, the mean position of the jet $\bar{Y}$, the waviness of the jet W, the jet latitude index JLI and the zonal wind speed at the jet latitude index U_{JLI}. The dots represent the estimated coefficient and the shaded vertical bars the associated 95% confidence interval. The orange (resp. blue and green) estimation is found using the time series after applying a 2-year (resp. 5-year and 10-year) low-pass Gaussian filter. The red estimation is found using the time series with the 2-year filter with only T as a regressor. All time series are monthly averages

of the global temperature of the Earth leads to a decrease of 2.4% [1.8, 3.0] (−0.18/K [−0.22, −0.14]) of the local dimension and 7.9% [6.8, 9.0] (−0.48°/K [−0.55, −0.42]) of the waviness of the jet position with respect to their mean over the 1900–2010 period. It also leads to an increase of the local persistence by 0.5% [0.3, 0.7] (0.02 d/K [0.01, 0.03]), of the mean speed by 9.7% [9.3, 10.2] (1.25 m s^{-1}/K [1.20, 1.31]), of the zonal wind speed at the JLI by 10.9% [10.1, 11.7] (1.26 m s^{-1}/K [1.17, 1.35]), of the mean position by 1.5% [1.2, 1.8] (0.75°/K [0.59, 0.91]) and of the JLI by 1.8% [1.3, 2.3] (0.86°/K [0.62, 1.11]).

The results of this analysis are dependent on whether the proposed DAG correctly contains all relevant variables. In particular, our results would be sensitive to the exclusion of any confounder of T and Jet. To investigate this question, we apply the sensitivity analysis proposed by Cinelli and Hazlett (2020) on the 2-year filtered time series. This analysis investigates how strong the influence of an omitted confounder (or a group of confounders) has to be to change our conclusions and how robust are our results to all omitted confounders acting together (possibly non-linearly). We especially test for the sign of the α coefficient. For every confounder C included in our regression (C = AMO, PDO, ENSO and ARF), the test investigates whether an omitted variable which would have up to three times the impact of C on both T and Jet would change the sign of the α coefficient (given a confidence level of 5%). The results of this analysis are reported in Online appendix (Figs. A4, A5). All regressions are robust to the presence of an omitted confounder, except for the ARF confounder in the regression on θ^{-1}. For the latter, the sensitivity analysis shows that an omitted confounder which would have twice the impact that ARF has on T and on the persistence θ^{-1} of the jet could change the sign of our results (from an increase to a decrease of the persistence). This is coherent with the small increase of θ^{-1} determined by our analysis. We nonetheless emphasize that it is not straightforward to imagine what would be a confounder whose influence on both the global mean temperature of the Earth and the dynamics of the Euro-Atlantic eddy-driven jet is twice as important as the one of anthropogenic aerosols emissions over the twentieth century. Therefore we are still confident with respect to this result.

Previous work on the evolution of the jet have often focused on specific seasons (e.g. Coumou et al. 2018; Harvey et al. 2020). The physical changes associated with global warming can indeed be different from one season to another. We therefore perform our analysis for the meteorological winter (DJF) and summer (JJA) seasons and the results are presented in Fig. 9. The estimation of the α coefficient displays strong differences from one season to another for some indicators. There is no evolution in the persistence indicator for winter, whereas it increases in summer. The changes in local dimension and waviness are close to zero in winter, but strongly negative in summer. The increase in the jet speed, measured either with $\bar{U}$ or U_{JLI} is preserved in both seasons. There is a small but insignificant at the 5% level increase in the jet mean position ($\bar{Y}$ and JLI) in winter, but strangely enough the results for $\bar{Y}$ and JLI in summer diverge: $\bar{Y}$ increases strongly whereas the change in JLI is null. The discrepancy between those indicators in summer may be

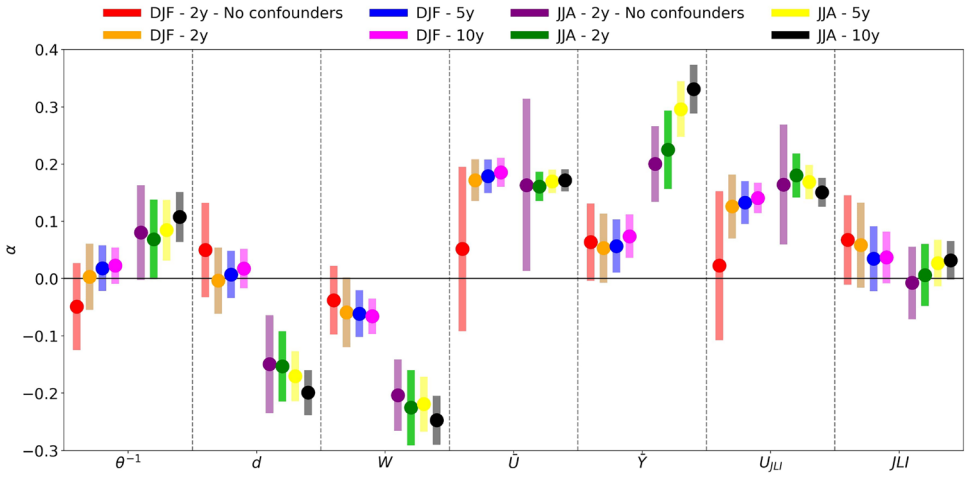


Fig. 9 Estimation of the impact of global warming on indicators of the jet variability for the DJF and JJA seasons (ERA20C). The plot represents the estimated α coefficient for the local dimension d, the local persistence θ^{-1}, the mean speed of the jet $\bar{U}$, the mean position of the jet $\bar{Y}$, the waviness of the jet W, the jet latitude index JLI and the zonal wind speed at the jet latitude index U_{JLI}. The dots represent the estimated coefficient and the shaded vertical bars the associated 95% confidence interval

caused by increased high pressure systems over continental Europe which systematically deviate the jet towards the North, an effect which would be small over the ocean because of the differential warming rate enhanced in summer in both regions.

Figures A6 and A7 in Online appendix present the same analysis using the ERA5 data set over the 1950–2020 period. Broadly speaking, the estimations of the α coefficients are much less precise and many coefficients are not statistically different from zero. The decrease in waviness is still present but not significant and the results by season are opposed to the results in the ERA20C data set. This is also the case for local dimension. The result on local persistence are not distinguishable from zero. The mean speed $\bar{U}$ does not seem to be affected by global warming over the year but the results points toward an acceleration in winter. For U_{JLI}, the results depend on the filter used but the winter acceleration seems to also be present. The shift in position measured either with $\bar{Y}$ or JLI, points towards a southward shift, especially in summer. These results using the ERA5 data set are much less precise and inconsistent from one specification of the filter to another, in contrast with the stability of the results for the ERA20C data set.

5 Discussion and conclusions

We studied the variability of the North Atlantic eddy-driven jet stream described by its instantaneous latitudinal position at each longitude. We used indicators from dynamical system theory to characterize the underlying attractor on which the jet is evolving. We showed that this representation is more relevant to characterize the jet variability than using the more classical Z500 and SLP fields and associated weather regimes. Z500 and SLP fields carry more information than needed and targeting the jet position per se as we did allows to focus on the sole dynamics of the jet.

We then studied the inter-decadal variability of the jet using the dynamical indicators. After controlling for potential confounders between the global temperature of the Earth and the jet, we estimate the impact of global warming on key indicators of the jet variability. We showed that, over the 1900–2010 period, global warming is associated to a decrease of the local dimension and waviness of the jet position and an increase of its persistence. Global warming is also linked to an increase in the wind speed of the jet and a northward shift in its mean position. These results are stronger in summer than in winter, the latter showing small or insignificant evolution of local dimension, persistence and waviness. Our results are robust to the presence of an omitted confounder.

Our results on the evolution of dynamical indicators (d and θ^{-1}) are coherent with previous studies. Faranda et al. (2019a) and Rodrigues et al. (2018) also showed a decrease in local dimension over the North-Atlantic in future warming scenarios using the SLP field as a state vector. Faranda et al. (2019a) attribute this decrease to the warming of the ocean. Our results suggest a similar mechanism may apply to the eddy-driven jet. We note that we find an increase in the variability of the jet over the 1990–2020 period with respect to the 1950–1980 period, but this observation is not consistent with the general trend over the twentieth century. This may explain the contradicting results found in the literature on the recent changes of the variability of the jet stream (Francis and Vavrus 2015; Coumou et al. 2015; Harvey et al. 2020).

The validity of our results strongly relies on the quality of the reanalysis data sets, especially with regards to their capacity to reproduce correctly the behavior of the North-Atlantic eddy-driven jet stream in the past. Even though the large structures of the atmosphere are probably the features best resolved by reanalysis (Slivinski et al.

2021), there is no doubt that the reanalysis over the period 1980–2020, the so-called satellite era, provide much more accurate results than in the previous periods. It is not clear whether we should have more confidence on the results found on the ERA5 data set over the period 1950–2020 or on the ERA20C data set over the period 1900–2010. The properties of the jet are probably closer to reality in the first data set but the natural variability of the climate system is better sampled in the second. Therefore, even though we control the influence of confounders from the influence of global warming of the Earth on the jet, we are cautious in asserting that our results give the causal impact of global warming on the variability of the jet. Note that if the proposed estimated impact is correct, the pathways are expected to be through Arctic amplification, tropical upper-troposphere warming and increased land-sea gradient as mentioned in the Introduction. How much each of these pathways contributed to the change in jet characteristics is an interesting avenue for future research.

Finally, our results demonstrate the interest of using tools from dynamical system theory to target specific patterns of the large scale atmospheric circulation and quantify their natural and forced variability. We also think that these tools could be used to evaluate whether climate models are able to represent correctly the natural variability of the jet position and we are currently investigating this question.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00382-023-07022-z.

Acknowledgements The authors wish to thank S. Charbit, G. Messori, J. Riboldi, Y. Robin and T. Caby for their fruitful inputs. The authors also thank Y. Balkanski and P. Poli for their help with aerosols data.

Author contributions RN designed the analysis. Material preparation and data collection were performed by VG, AV, ANSH and RN. Data analysis was performed by VG, AV and RN. BC advised the causal network analysis and ran the sensitivity tests. All authors contributed to the writing of the paper.

Funding This work has received support from the European Union's Horizon 2020 research and innovation programme under Grant agreement no. 101003469 (XAIDA), by the Integrated Training Network (ITN) under the Marie Sklodowska-Curie Grant agreement no. 956396 (EDIPI).

Availability of data and materials The ERA5 and ERA20C are available on the Copernicus web site (https://cds.climate.copernicus.eu/). The results of the detection procedure and the time series of indicators are available on demand.

Code availability The main results of this work were obtained using Python. The scripts are available upon request.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Barnes EA, Screen JA (2015) The impact of arctic warming on the midlatitude jet-stream: can it? has it? Will it? Wiley Interdiscip Rev Clim Change 6(3):277–286

Blackport R, Screen JA (2020) Insignificant effect of arctic amplification on the amplitude of midlatitude atmospheric waves. Sci Adv 6(8):eaay2880

Cattiaux J, Peings Y, Saint-Martin D et al (2016) Sinuosity of midlatitude atmospheric flow in a warming world. Geophys Res Lett 43(15):8259–8268

Charney JG (1947) The dynamics of long waves in a baroclinic westerly current. J Atmos Sci 4(5):136–162

Cinelli C, Hazlett C (2020) Making sense of sensitivity: extending omitted variable bias. J R Stat Soc Ser B (Stat Methodol) 82(1):39–67

Cohen J, Zhang X, Francis J et al (2018) Arctic change and possible influence on mid-latitude climate and weather: a us clivar white paper. US CLIVAR reports

Coumou D, Lehmann J, Beckmann J (2015) The weakening summer circulation in the northern hemisphere mid-latitudes. Science 348(6232):324–327

Coumou D, Di Capua G, Vavrus S et al (2018) The influence of arctic amplification on mid-latitude summer circulation. Nat Commun 9(1):1–12

Diao C, Xu Y (2022) Reassessing the relative role of anthropogenic aerosols and natural decadal variability in driving the mid-twentieth century global "cooling'': a focus on the latitudinal gradient of tropospheric temperature. Clim Dyn 59:1–27

Ding S, Chen W, Feng J et al (2017) Combined impacts of pdo and two types of la niña on climate anomalies in Europe. J Clim 30(9):3253–3278

Dorrington J, Strommen K (2020) Jet speed variability obscures euro-atlantic regime structure. Geophys Res Lett 47(15):e2020GL087907

Duchon CE (1979) Lanczos filtering in one and two dimensions. J Appl Meteorol Climatol 18(8):1016–1022

Dufresne JL, Foujols MA, Denvil S et al (2013) Climate change projections using the ipsl-cm5 earth system model: from cmip3 to cmip5. Clim Dyn 40(9):2123–2165

Faranda D, Messori G, Yiou P (2017) Dynamical proxies of north atlantic predictability and extremes. Sci Rep 7(1):1–10

Faranda D, Alvarez-Castro MC, Messori G et al (2019a) The hammam effect or how a warm ocean enhances large scale atmospheric predictability. Nat Commun 10:1316

Faranda D, Sato Y, Messori G et al (2019b) Minimal dynamical systems model of the northern hemisphere jet stream via embedding of climate data. Earth Syst Dyn 10(3):555–567

Foster G, Rahmstorf S (2011) Global temperature evolution 1979–2010. Environ Res Lett 6(4):022–044

Francis JA, Vavrus SJ (2015) Evidence for a wavier jet stream in response to rapid arctic warming. Environ Res Lett 10(1):005–014

Glymour M, Pearl J, Jewell NP (2016) Causal inference in statistics: a primer. Wiley, New York

Gulev SK, Thorne PW, Ahn J et al (2021) Changing state of the climate system. Cambridge University Press, Cambridge

Hallam S, Josey SA, McCarthy GD et al (2022) A regional (land–ocean) comparison of the seasonal to decadal variability of the northern hemisphere jet stream 1871–2011. Clim Dyn 59:2655–2681

Hannart A, Pearl J, Otto F et al (2016) Causal counterfactual theory for the attribution of weather and climate-related events. Bull Am Meteorol Soc 97(1):99–110

Harvey B, Cook P, Shaffrey L et al (2020) The response of the northern hemisphere storm tracks and jet streams to climate change in the cmip3, cmip5, and cmip6 climate models. J Geophys Res Atmos 125(23):e2020JD032701
Held IM (1975) Momentum transport by quasi-geostrophic eddies. J Atmos Sci 32(7):1494–1497
Held IM (1993) Large-scale dynamics and global warming. Bull Am Meteorol Soc 74(2):228–242
Held IM, Hou AY (1980) Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. J Atmos Sci 37(3):515–533
Hersbach H, Bell B, Berrisford P et al (2020) The era5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049
Hirt M, Craig GC, Schäfer SA et al (2020) Cold-pool-driven convective initiation: using causal graph analysis to determine what convection-permitting models are missing. Q J R Meteorol Soc 146(730):2205–2227
Hochman A, Alpert P, Harpaz T et al (2019) A new dynamical systems perspective on atmospheric predictability: Eastern mediterranean weather regimes as a case study. Sci Adv 5(6):eaau0936
Holton JR (1973) An introduction to dynamic meteorology. Am J Phys 41(5):752–754
Hoskins BJ, James IN (2014) Fluid dynamics of the mid-latitude atmosphere. Wiley, New York
Hurrell JW, Deser C (2010) North Atlantic climate variability: the role of the north Atlantic oscillation. J Mar Syst 79(3–4):231–244
Iqbal W, Leung WN, Hannachi A (2018) Analysis of the variability of the north Atlantic eddy-driven jet stream in cmip5. Clim Dyn 51(1):235–247
Jackson LC, Biastoch A, Buckley MW et al (2022) The evolution of the North Atlantic meridional overturning circulation since 1980. Nat Rev Earth Environ 3:241–254
Jiménez-Esteve B, Domeisen DI (2018) The tropospheric pathway of the enso-north Atlantic teleconnection. J Clim 31(11):4563–4584
Kautz LA, Martius O, Pfahl S et al (2022) Atmospheric blocking and weather extremes over the euro-Atlantic sector—a review. Weather Clim Dyn 3(1):305–336
Kerr RA (2000) A north Atlantic climate pacemaker for the centuries. Science 288(5473):1984–1985
Kretschmer M, Coumou D, Donges JF et al (2016) Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. J Clim 29(11):4069–4081
Kretschmer M, Adams SV, Arribas A et al (2021) Quantifying causal pathways of teleconnections. Bull Am Meteorol Soc 102(12):E2247–E2263
Lee S, Kim H (2003) The dynamical relationship between subtropical and eddy-driven jets. J Atmos Sci 60(12):1490–1503
Lee RW, Woolnough SJ, Charlton-Perez AJ et al (2019) Enso modulation of mjo teleconnections to the north Atlantic and Europe. Geophys Res Lett 46(22):13535–13545
Lee JY, Marotzke J, Bala G et al (2021) Future global climate: scenario-based projections and near-term information. In: Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press
Levine AF, McPhaden MJ, Frierson DM (2017) The impact of the amo on multidecadal enso variability. Geophys Res Lett 44(8):3877–3886
Limbach S, Schömer E, Wernli H (2012) Detection, tracking and event localization of jet stream features in 4-d atmospheric data. Geosci Model Dev 5(2):457–470
Lin J, Qian T (2022) The Atlantic multi-decadal oscillation. Atmos Ocean 60(3–4):307–337
Lucarini V, Faranda D, de Freitas JMM et al (2016) Extremes and recurrence in dynamical systems. Wiley, New York
Mann ME, Steinman BA, Brouillette DJ et al (2021) Multidecadal climate oscillations during the past millennium driven by volcanic forcing. Science 371(6533):1014–1019
Messori G, Caballero R, Faranda D (2017) A dynamical systems approach to studying midlatitude weather extremes. Geophys Res Lett 44(7):3346–3354
Messori G, Harnik N, Madonna E et al (2021) A dynamical systems characterization of atmospheric jet regimes. Earth Syst Dyn 12(1):233–251
Mezzina B, García-Serrano J, Bladé I et al (2020) Dynamics of the enso teleconnection and nao variability in the north Atlantic-European late winter. J Clim 33(3):907–923
Michelangeli PA, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi stationarity. J Atmos Sci 52(8):1237–1256
Mitchell D, Kornhuber K, Huntingford C et al (2019) The day the 2003 European heatwave record was broken. Lancet Planet Health 3(7):e290–e292
Molnos S, Mamdouh T, Petri S et al (2017) A network-based detection scheme for the jet stream core. Earth Syst Dyn 8(1):75–89
Moloney NR, Faranda D, Sato Y (2019) An overview of the extremal index. Chaos Interdiscip J Nonlinear Sci 29(2):022101
Monnin E, Kretschmer M, Polichtchouk I (2022) The role of the timing of sudden stratospheric warmings for precipitation and temperature anomalies in Europe. Int J Climatol 42(6):3448–3462
Moon W, Kim BM, Yang GH et al (2022) Wavier jet streams driven by zonally asymmetric surface thermal forcing. Proc Natl Acad Sci 119(38):e2200890119
Morice CP, Kennedy JJ, Rayner NA et al (2021) An updated assessment of near-surface temperature change from 1850: the hadcrut5 data set. J Geophys Res Atmos 126(3):e2019JD032361
Murakami H (2022) Substantial global influence of anthropogenic aerosols on tropical cyclones over the past 40 years. Sci Adv 8(19):eabn9493
Newman M, Alexander MA, Ault TR et al (2016) The pacific decadal oscillation, revisited. J Clim 29(12):4399–4427
Osman MB, Coats S, Das SB et al (2021) North Atlantic jet stream projections in the context of the past 1250 years. Proc Natl Acad Sci 118(38):e2104105118
Oudar T, Cattiaux J, Douville H (2020) Drivers of the northern extratropical eddy-driven jet change in cmip5 and cmip6 models. Geophys Res Lett 47(8):e2019GL086695
Pausata FSR, Gaetani M, Messori G et al (2015) The role of aerosol in altering north Atlantic atmospheric circulation in winter and its impact on air quality. Atmos Chem Phys 15(4):1725–1743
Pearl J (2009) Causal inference in statistics: an overview. Stat Surv 3:96–146
Pearl J, Mackenzie D (2018) The book of why: the new science of cause and effect. Basic books, New York
Peings Y, Cattiaux J, Vavrus SJ et al (2018) Projected squeezing of the wintertime north-Atlantic jet. Environ Res Lett 13(7):016–074
Pena-Ortiz C, Gallego D, Ribera P et al (2013) Observed trends in the global jet stream characteristics during the second half of the twentieth century. J Geophys Res Atmos 118(7):2702–2713
Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3(1):119–131
Poli P, Hersbach H, Dee DP et al (2016) Era-20c: an atmospheric reanalysis of the twentieth century. J Clim 29(11):4083–4097
Pons FME, Messori G, Alvarez-Castro MC et al (2020) Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions. J Stat Phys 179(5):1698–1717
Portal A, Pasquero C, D'andrea F et al (2022) Influence of reduced winter land-sea contrast on the midlatitude atmospheric circulation. J Clim 35(19):2637–2651

Qin M, Dai A, Hua W (2020) Quantifying contributions of internal variability and external forcing to Atlantic multidecadal variability since 1870. Geophys Res Lett 47(22):e2020GL089504

Rhines PB (1975) Waves and turbulence on a beta-plane. J Fluid Mech 69(3):417–443

Robson J, Ortega P, Sutton R (2016) A reversal of climatic trends in the north Atlantic since 2005. Nat Geosci 9:513–517

Robson J, Menary MB, Sutton RT et al (2022) The role of anthropogenic aerosol forcing in the 1850–1985 strengthening of the amoc in cmip6 historical simulations. J Clim 35(20):3243–3263

Rodrigues D, Alvarez-Castro MC, Messori G et al (2018) Dynamical properties of the north Atlantic atmospheric circulation in the past 150 years in cmip5 models and the 20crv2c reanalysis. J Clim 31(15):6097–6111

Röthlisberger M, Pfahl S, Martius O (2016) Regional-scale jet waviness modulates the occurrence of midlatitude weather extremes. Geophys Res Lett 43(20):10–989

Runge J, Petoukhov V, Kurths J (2014) Quantifying the strength and delay of climatic interactions: the ambiguities of cross correlation and a novel measure based on graphical models. J Clim 27(2):720–739

Simpson IR, Deser C, McKinnon KA et al (2018) Modeled and observed multidecadal variability in the north Atlantic jet stream and its connection to sea surface temperatures. J Clim 31(20):8313–8338

Simpson IR, Yeager SG, McKinnon KA et al (2019) Decadal predictability of late winter precipitation in western Europe through an ocean-jet stream connection. Nat Geosci 12(8):613–619

Slivinski L, Compo G, Sardeshmukh P et al (2021) An evaluation of the performance of the twentieth century reanalysis version 3. J Clim 34(4):1417–1438

Spensberger C, Spengler T (2020) Feature-based jet variability in the upper troposphere. J Clim 33(16):6849–6871

Spensberger C, Spengler T, Li C (2017) Upper-tropospheric jet axis detection and application to the boreal winter 2013/14. Mon Weather Rev 145(6):2363–2374

Stendel M, Francis J, White R et al (2021) Chapter 15 - The jet stream and climate change. In: Climate change. Ed. Trevor M. Letcher. Elsevier, pp 327–357

Sutton R, Dong B (2012) Atlantic ocean influence on a shift in European climate in the 1990s. Nat Geosci 5:788–792

Süveges M (2007) Likelihood estimation of the extremal index. Extremes 10(1):41–55

Tenenbaum J, Williams PD, Turp D et al (2022) Aircraft observations and reanalysis depictions of trends in the North Atlantic winter jet stream wind speeds and turbulence. Q J R Meteorol Soc 148:2927–2941

Trenberth KE, Shea DJ (2006) Atlantic hurricanes and natural variability in 2005. Geophys Res Lett 33(12):L12704

van Oldenborgh GJ, te Raa LA, Dijkstra HA et al (2009) Frequency-or amplitude-dependent effects of the Atlantic meridional overturning on the tropical pacific ocean. Ocean Sci 5(3):293–301

Van Oldenborgh GJ, Hendon H, Stockdale T et al (2021) Defining el niño indices in a warming climate. Environ Res Lett 16(4):003–044

Woollings T, Hannachi A, Hoskins B (2010) Variability of the north Atlantic eddy-driven jet stream. Q J R Meteorol Soc 136(649):856–868

Woollings T, Czuchnicki C, Franzke C (2014) Twentieth century north Atlantic jet variability. Q J R Meteorol Soc 140(680):783–791

Woollings T, Barnes E, Hoskins B et al (2018) Daily to decadal modulation of jet variability. J Clim 31(4):1297–1314

Young LS (1982) Dimension, entropy and lyapunov exponents. Ergodic Theory Dyn Syst 2(1):109–124

Appendix A Supporting information

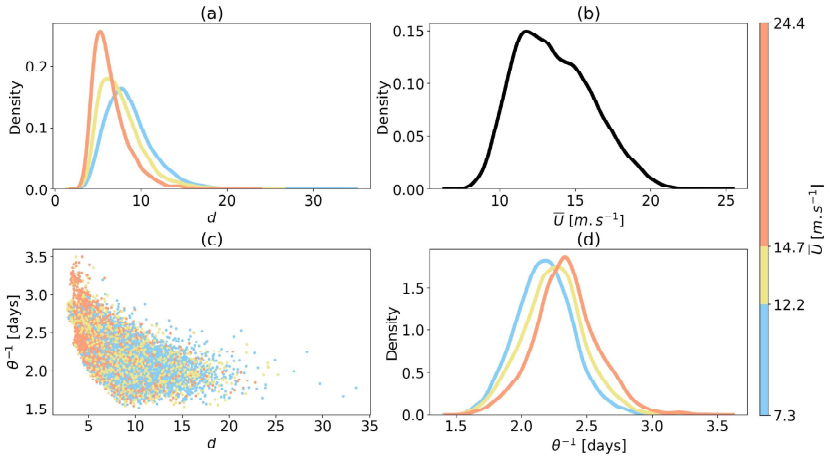


Fig. A1: **Link between the dynamical indicators and the jet mean speed.** (a) Distribution of local dimension d for the three terciles of the mean speed $\bar{U}$. (b) Distribution of $\bar{U}$. (c) Cross distribution of d and θ^{-1} colored by the tercile of the $\bar{U}$ indicator. (d) Distribution of local persistence θ^{-1} for the three terciles of the mean speed $\bar{U}$.

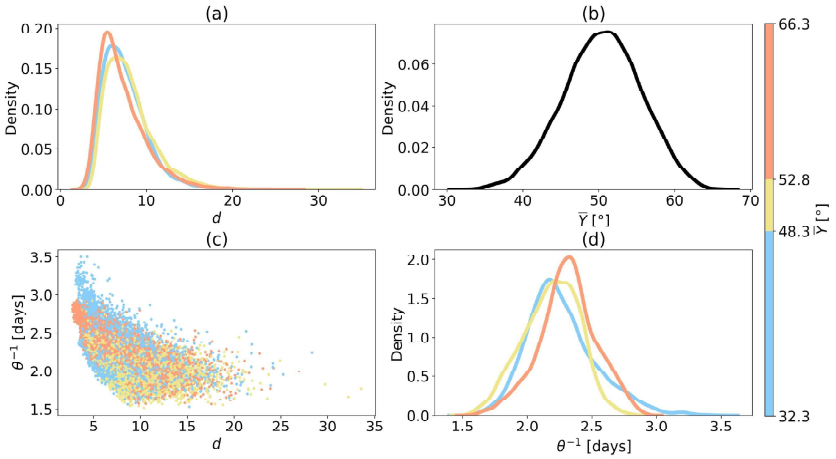


Fig. A2: **Link between the dynamical indicators and the jet mean position.** (a) Distribution of local dimension d for the three terciles of the mean position $\bar{Y}$. (b) Distribution of $\bar{Y}$. (c) Cross distribution of d and θ^{-1} colored by the tercile of the $\bar{Y}$ indicator. (d) Distribution of local persistence θ^{-1} for the three terciles of the mean position $\bar{Y}$.

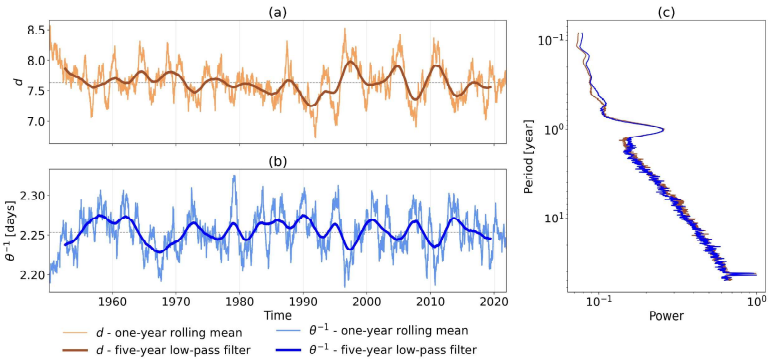
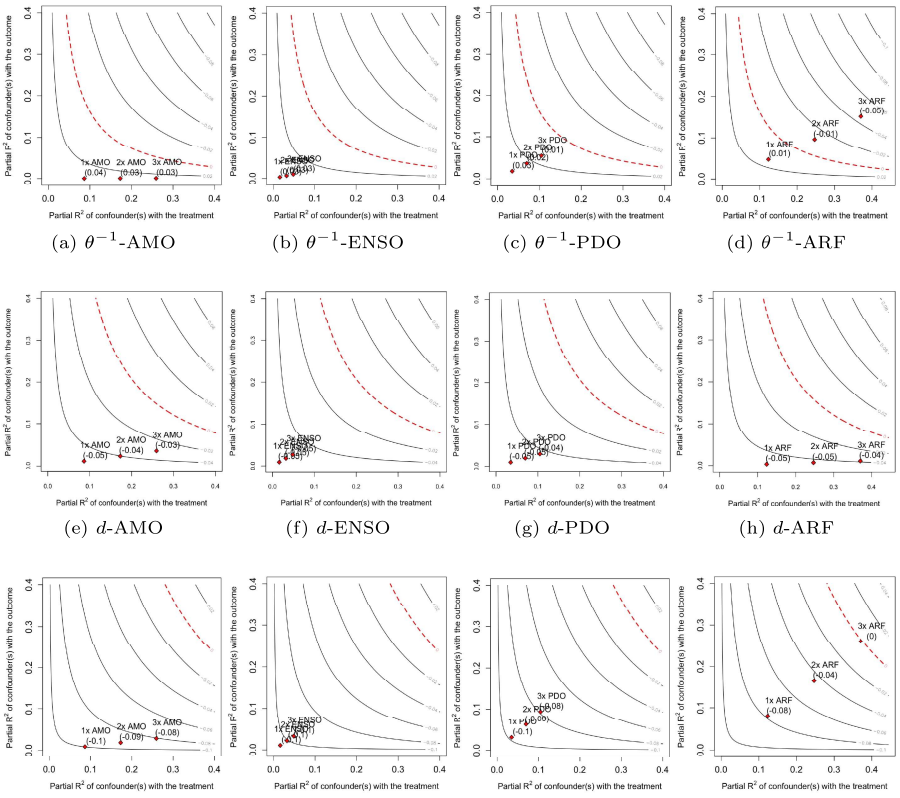


Fig. A3: **Time series and power spectrum of d and θ^{-1} (ERA5).** (a) Time series of d with a one-year rolling mean and a five-year low-pass filter. (b) Time series of θ^{-1} with a one-year rolling mean and a five-year low-pass filter. (c) Power spectrum of d and θ^{-1}.

Jet indices	AMO	ENSO	PDO	ARF	T	R^2
θ^{-1}	0.005	-0.023*	-0.052***	0.593***	0.037***	0.065
	[−0.013, 0.023]	[−0.045, −0.002]	[−0.074, −0.030]	[0.422, 0.764]	[0.022, 0.052]	
d	0.029***	0.034***	0.035***	-0.141	-0.058***	0.064
	[0.013, 0.046	[0.014, 0.055]	[0.015, 0.056]	[−0.299, 0.017]	[−0.071, −0.044]	
W	0.028**	0.041***	0.068***	-0.772***	-0.107***	0.163
	[0.011, 0.046]	[0.019, 0.063]	[0.046, 0.090]	[−0.942, −0.603]	[−0.122, −0.093]	
$\bar{U}$	-0.089***	-0.079***	-0.048***	-0.869***	0.177***	0.717
	[−0.098, −0.079]	[−0.091, −0.067]	[−0.060, −0.036]	[−0.962, −0.777]	[0.169, 0.185]	
$\bar{Y}$	-0.017	-0.101***	-0.013	-0.129	0.078***	0.117
	[−0.037, 0.003]	[−0.126, −0.076]	[−0.038, 0.012]	[−0.324, 0.065]	[0.061, 0.095]	
U_{JLI}	-0.111***	-0.041***	-0.065***	-0.304***	0.166***	0.439
	[−0.126, −0.096]	[−0.059, −0.023]	[−0.083, −0.047]	[−0.444, −0.164]	[0.154, 0.178]	
JLI	0.024*	-0.052***	0.004	-0.073	0.061***	0.066
	[0.003, 0.044]	[−0.078, −0.027]	[−0.022, 0.030]	[−0.273, 0.127]	[0.044, 0.079]	

Table A1: **Estimated coefficients of the linear model (ERA20C - 1900–2010) with the 2-year filtered data.** Values in brackets correspond to the 95% confidence interval. Coefficients with one star correspond to $p < 0.05$, two stars to $p < 0.01$ and three stars to $p < 0.001$.

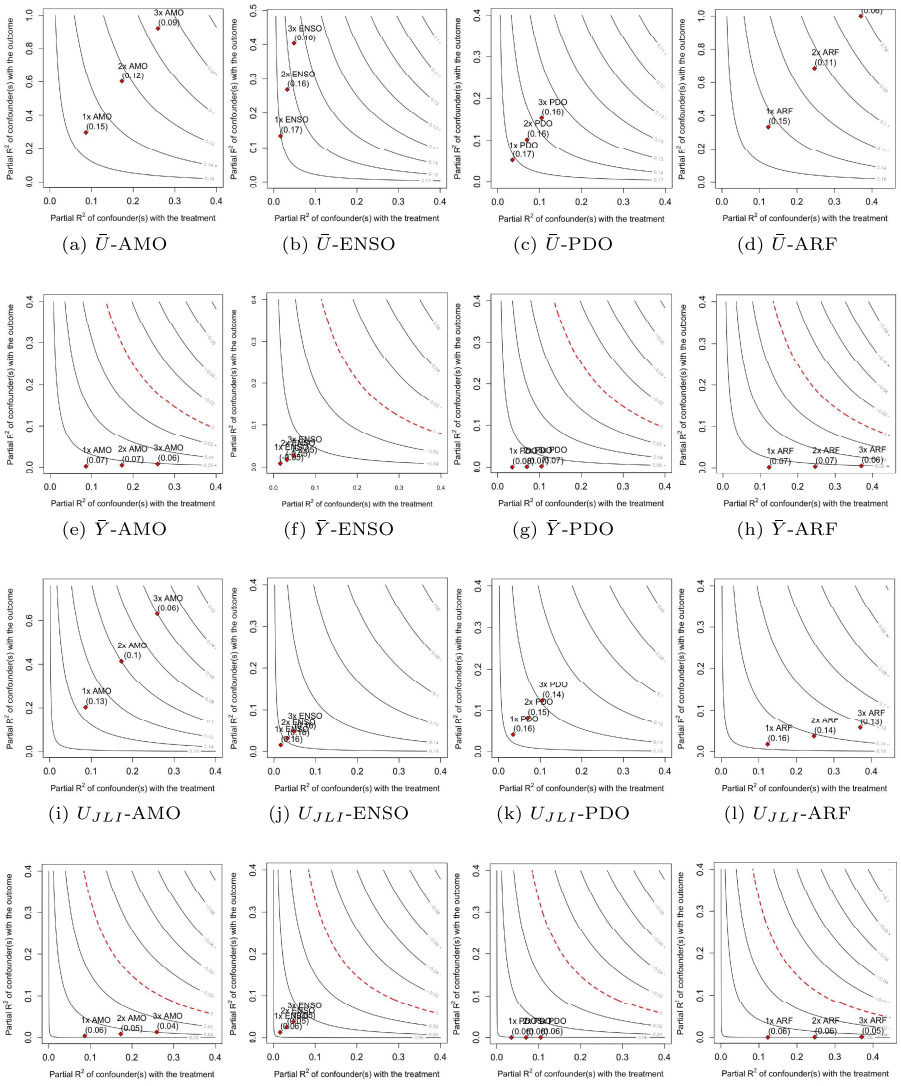


(a) θ^{-1}-AMO (b) θ^{-1}-ENSO (c) θ^{-1}-PDO (d) θ^{-1}-ARF

(e) d-AMO (f) d-ENSO (g) d-PDO (h) d-ARF

(i) W-AMO (j) W-ENSO (k) W-PDO (l) W-ARF

Fig. A4: **Sensitivity analysis to the omission of a confounder in the estimation of the effect of T on θ^{-1}, d and W.** The plots display what would be the coefficient α in the regression if a confounder which would influence both T and Jet one to three times as much as the confounder considered had been forgotten. We consider the result robust as soon as the estimated coefficient does not change sign, i.e. does not cross the red line.



(a) $\bar{U}$-AMO (b) $\bar{U}$-ENSO (c) $\bar{U}$-PDO (d) $\bar{U}$-ARF

(e) $\bar{Y}$-AMO (f) $\bar{Y}$-ENSO (g) $\bar{Y}$-PDO (h) $\bar{Y}$-ARF

(i) U_{JLI}-AMO (j) U_{JLI}-ENSO (k) U_{JLI}-PDO (l) U_{JLI}-ARF

(m) JLI-AMO (n) JLI-ENSO (o) JLI-PDO (p) JLI-ARF

Fig. A5: **Sensitivity analysis to the omission of a confounder in the estimation of the effect of T on $\bar{U}$, $\bar{Y}$, U_{JLI} and JLI.** The plots display what would be the coefficient α in the regression if a confounder which would influence both T and Jet one to three times as much as the confounder considered had been forgotten. We consider the result robust as soon as the estimated coefficient does not change sign, i.e. does not cross the red line.

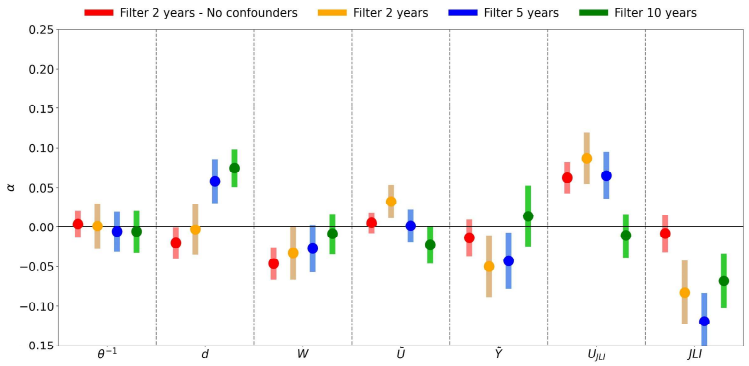


Fig. A6: **Estimation of the impact of global warming on indicators of the jet variability (ERA5).** The plot represents the estimated α coefficient for the local dimension d, the local persistence θ^{-1}, the mean speed of the jet $\bar{U}$, the mean position of the jet $\bar{Y}$, the waviness of the jet W, the jet latitude index JLI and the zonal wind speed at the jet latitude index U_{JLI}. The dots represent the estimated coefficient and the shaded vertical bars the associated 95% confidence interval. The orange (resp. blue and green) estimation is found using the time series after applying a 2-year (resp. 5-year and 10-year) low-pass Gaussian filter. All time series are monthly averages.

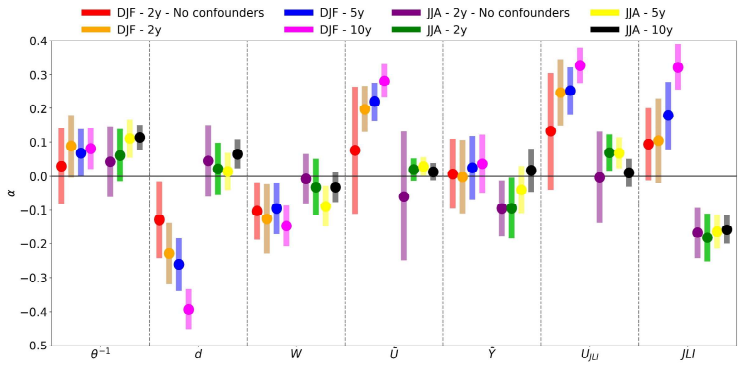


Fig. A7: Estimation of the impact of global warming on indicators of the jet variability for the DJF and JJA seasons (ERA5). The plot represents the estimated α coefficient for the local dimension d, the local persistence θ^{-1}, the mean speed of the jet $\bar{U}$, the mean position of the jet $\bar{Y}$, the waviness of the jet W, the jet latitude index JLI and the zonal wind speed at the jet latitude index U_{JLI}. The dots represent the estimated coefficient and the shaded vertical bars the associated 95% confidence interval.