Simulating the West Pacific heatwave of 2021 with analog importance sampling

Flavio Pons, Pascal Yiou, Aglaé Jézéquel, Gabriele Messori

Weather and Climate Extremes 43 (2024) 100651

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Simulating the Western North America heatwave of 2021 with analogue importance sampling

Flavio Maria Emanuele Pons [a,*], Pascal Yiou [a], Aglaé Jézéquel [b,c], Gabriele Messori [d,e,f]

[a] Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay & IPSL, 91191 Gif-sur-Yvette, France
[b] Laboratoire de Météorologie Dynamique, UMR CNRS-ENSX, 24 rue Lhomond, Paris, 75005, France
[c] Ecole des Ponts Paristech, Champs-sur-Marne, 77420, France
[d] Department of Earth Sciences, Uppsala University, Uppsala, 752 36, Sweden
[e] Swedish Centre for Impacts of Climate Extremes (climes), Uppsala University, Uppsala, 752 36, Sweden
[f] Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, 106 91, Stockholm, Sweden

ARTICLE INFO

Keywords:
Extreme events
North American heatwave
Stochastic weather generator

ABSTRACT

During the summer of 2021, the North American Pacific Northwest was affected by an extreme heatwave that broke previous temperature records by several degrees. The event caused severe impacts on human life and ecosystems, and was associated with the superposition of concurrent drivers, whose effects were amplified by climate change. We evaluate whether this record-breaking heatwave could have been foreseen prior to its observation, and how climate change affects North American Pacific Northwest worst-case heatwave scenarios. To this purpose, we use a stochastic weather generator with empirical importance sampling. The generator simulates extreme temperature sequences using circulation analogues, chosen with an importance sampling based on the daily maximum temperature over the region that recorded the most extreme impacts. We show how some of the large-scale drivers of the event can be obtained form the circulation analogues, even if such information is not directly given to the stochastic weather generator.

1. Introduction

In June 2021, an intense heatwave affected the Pacific North West (PNW) of North America, notably the states of Washington and Oregon in the U.S.A., and the Canadian province of British Columbia. The region experienced unprecedented temperature values, peaking at 49.6 °C in Lytton, British Columbia, on June 29 2021. This makes the event one of the most extreme heatwaves ever recorded globally (Thompson et al., 2022).

The prolonged extreme heat had impacts on vegetation and related ecosystems due to hydraulic damage (Klein et al., 2022) and devastating forest fires (Overland, 2021), on marine life (White et al., 2023), and on human health, with a significant number of excess deaths (Romanello et al., 2021).

Extreme value and large deviation analyses have shown that human-induced global warming has increased both the likelihood and the intensity of this heatwave (Philip et al., 2021; Lucarini et al., 2023). However, anthropogenic forcing alone is not sufficient to explain this specific event, which remains a one-in-1000 years event in the present climate (Philip et al., 2021).

The heatwave was associated with an anticyclonic Omega blocking centred over Western Canada, characterized by strong positive 500 hPa geopotential height (Z500) anomalies (Fig. 1). Around June 17, a split of the Arctic Polar Vortex (PV) caused an area of low pressure to move over the Pacific, triggering the switching between three atmospheric patterns historically associated with anticyclonic blocking over the PNW (Wang et al., 2022). First, the polar jet stream was displaced to the South and intensified over the West Pacific, causing the formation of a blocking anticyclone over the East Pacific, a dipole configuration known as North Pacific pattern (NOAA, 2022). After June 24, the zonal dipole corresponding to the North Pacific pattern became a tripole, with another low-pressure area located over the Arctic, as the heatwave reached peak intensity over the PNW. Finally, at the end of June, the configuration switched to a meridional dipole (known as North American pattern), causing the high pressure dome to move eastward and eventually dissipate.

Omega blockings over Western Canada are historically associated with heatwave conditions over the PNW (Bumbaco et al., 2013). However, as observed by Bartusek et al. (2022), while other areas such as Central Eurasia and Northeastern Siberia experienced similar positive geopotential and negative soil moisture anomalies in the second half of June, none of these regions have been affected by temperature

* Corresponding author.
E-mail address: flavio.pons@lsce.ipsl.fr (F.M.E. Pons).

https://doi.org/10.1016/j.wace.2024.100651
Received 9 February 2023; Received in revised form 10 October 2023; Accepted 12 February 2024
Available online 17 February 2024

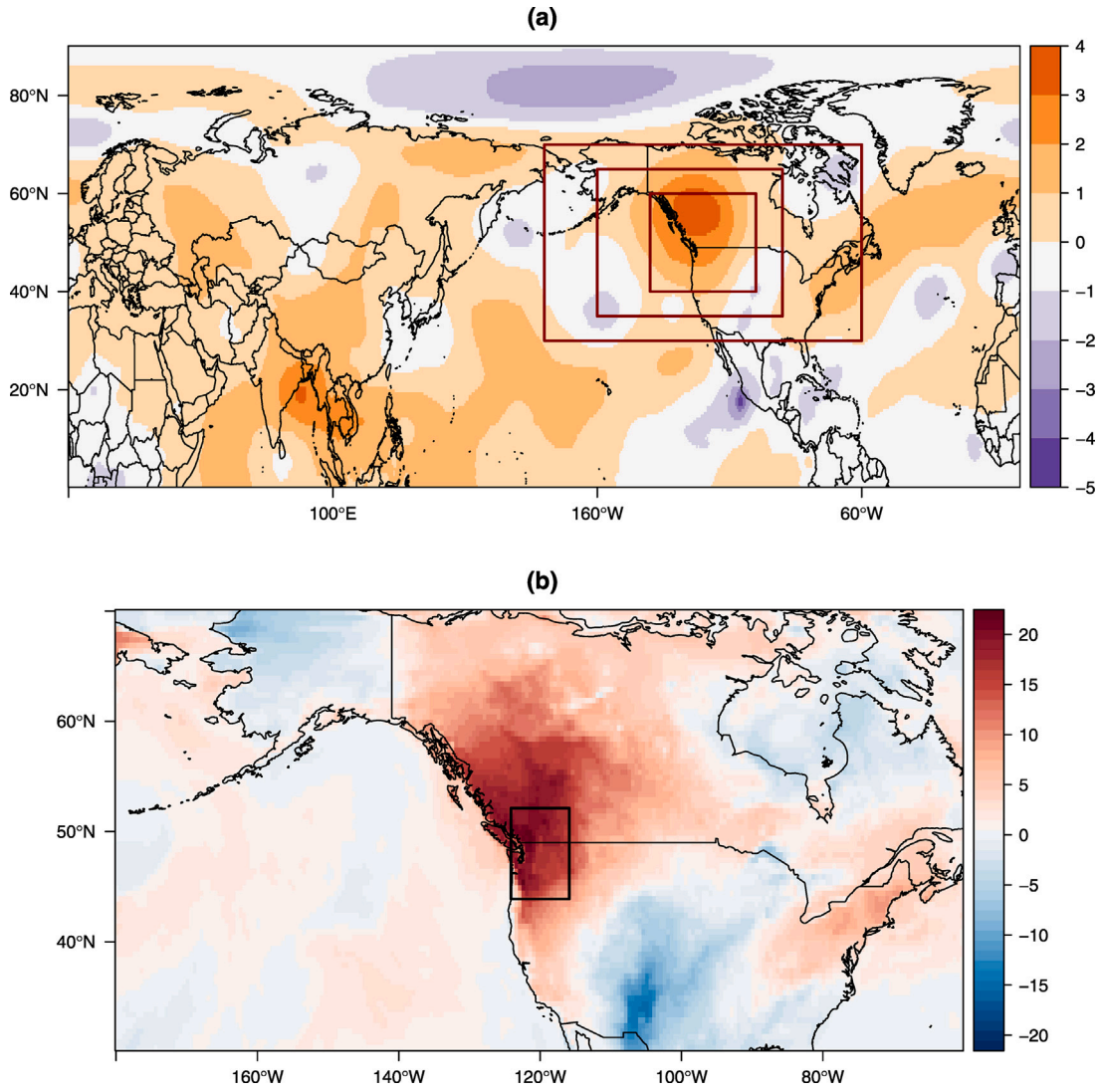

Fig. 1. (a) Z500 standardized anomalies with respect to the 1950–2022 climatology for June 27, 2021. The three boxes represent the three domains tested for the computation of circulation analogues. (b) daily maximum 2-metre temperature anomaly (°C) with respect to the 1950–2022 climatology for June 27, 2021, at the peak of the heatwave. The box represents the area for the computation of the target time series.

anomalies as extreme as those in the PNW. The exceptional nature of the PNW event has been likely caused by concurring and interacting anomalies in common drivers of heatwaves over the PNW.

Although there is some agreement that the extreme temperatures developed due to subsidence inside a high pressure dome (Philip et al., 2021; Neal et al., 2022; Qian et al., 2022; Wang et al., 2022) and were intensified by adiabatic heating downwind the Coast Mountains (Philip et al., 2021), teleconnections and diabatic processes linked to large-scale dynamics may have also been at play. Around June 25, the anticyclone developed an upper-level warm core, suggesting that heat was transported and injected into the high pressure dome from other regions. In particular, Neal et al. (2022) suggest that latent heating within the upstream cyclone in the North Pacific pattern produced an anomalous wave activity flux, with diabatic injection of heat inside the anticyclone. The authors also argue that this mechanism was likely enhanced by human-induced global warming, since the larger amount of water vapour in the atmosphere implies the intensification of the involved diabatic processes.

There is also evidence that extreme heating may have been partly due to the interaction between the Omega blocking over the PNW and an atmospheric river (Mo et al., 2022; Lin et al., 2022; Bercos-Hickey et al., 2022) excited by the East Asian Summer Monsoon (EASM). Qian et al. (2022) and Bartusek et al. (2022) argue that subseasonal variations of the EASM and of the jet stream may have contributed to the intensification of a Rossby wave train crossing the Pacific in phase-locking with the PNW anticyclonic blocking. This wave train may have acted as an efficient guide for the teleconnection between the PNW and South East Asia, causing a subsidence anomaly to the South of the jet exit area. This concurred to the extreme heat conditions inside the anticyclonic dome (Qian et al., 2022), which may have amplified the hemispheric wavenumber-4 anomaly in which the blocking was embedded (Bartusek et al., 2022). Mo et al. (2022) propose two further mechanisms for the intensification of the heat dome: the direct injection into the anticyclonic dome of sensible heat transported from the tropics, and enhanced greenhouse effect due to large amounts of water vapour trapped inside the anticyclonic dome.

Finally, as noted by Bartusek et al. (2022), the increase in regional temperatures persisted throughout the event, even after the geopotential height had reached its peak, and subsidence was therefore decreasing. Following this observation, Bartusek et al. (2022) argue a possible involvement of shorter-term atmospheric dynamics, such as the upstream cyclogenesis leading to abrupt heat accumulation due to the blocking pattern (Neal et al., 2022), alongside a significant influence of bidirectional land–atmosphere feedback mechanisms that amplified and prolonged the heatwave in the PNW.

From the literature about the event, it emerges that many of the processes that have initiated and amplified the heat dome were driven by specific features in the atmospheric circulation, both over the region and at a larger scale. We investigate whether it is possible to simulate an extreme heatwave event – in this case, the most extreme in the observational data for the affected region – mainly based on information about the atmospheric circulation. We use and adapt the approach developed by Yiou and Jézéquel (2020), consisting of an empirical importance sampling with a stochastic weather generator (SWG) based on circulation analogues. The goal of the paper is to evaluate whether the 2021 PNW heatwave could be foreseen without having ever observed it, and how climate change affects PNW worst-case heatwave scenarios. We also discuss how the drivers of the event can be obtained from the circulation analogues.

The paper is structured as follows. Section 2 contains a description of the datasets and the pre-processing. Section 3 presents the methodology used to compute circulation analogues and a description of the SWG. The results are presented in Section 4 and Section 5 contains our conclusions.

2. Data and choice of climate variables

We use the ERA5 reanalysis dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), available on the Climate Data Store (CDS) of the Copernicus Climate Change Service (Hersbach et al., 2018). This data was recently back-extended to 1950. We use hourly data of 500 hPa geopotential height (Z500), 2-metre temperature (T2M) and total column water vapour (TCWV) for the May–September period between 1950 and 2022, at a horizontal resolution of 0.5 degrees over the Northern hemisphere. For all these variables, we consider anomalies with respect to the 1950–2022 seasonal cycle. It has been shown that Z500 has a positive global June–July–August (JJA) trend associated with global warming (Christidis and Stott, 2015). Indeed, we find 1950–2022 trends of about 0.4 m per year on the PNW (Fig. 10). To prevent the results from depending on long-term atmospheric trends, we subtract the 1950–2022 linear trend from Z500 (Yiou and Jézéquel, 2020).

For the computation of circulation analogues, we rely on daily averages of Z500 anomalies. As noted by Jézéquel et al. (2018), Z500 is a better variable than SLP to compute circulations analogues during heatwaves. The strong surface heating associated with persistent summer anticyclones causes the formation of a thermal low at the surface, which conceals the SLP signal associated with the positive Z500 anomalies (Rácz and Smith, 1999).

For posterior analysis, we consider composites over the entire hemisphere, to include larger-scale features connected to the onset of the heatwave.

For the thermal characterization of the heatwave, we use the daily maximum T2M (TX) over the domain [44–52 N; 116–124 W], marked by the black rectangle in Fig. 1(b). This region recorded the highest absolute temperatures and temperature anomalies in the PNW, and the largest number of affected people. To compare this heatwave to 1950–2022 temperature values, we compute several TX statistics: annual maximum daily temperature (TX01d), annual maximum of TX moving average over 7 days (TX07d), 15 days (TX15d) and 30 days (TX30d), and JJA average of TX (TJJA).

Finally, we consider the daily average TCWV over the region [0–90 N; 90–270 W] to track the atmospheric river that crossed the Pacific making landfall on the Alaskan panhandle during the last days of June, interacting with the already ongoing heatwave. TCWV is commonly used as a proxy of strong water vapour transport, to identify atmospheric rivers (Dacre et al., 2015). In particular, we look for filament-like regions with high values of TCWV stretching from the tropical regions towards the mid-latitudes.

We determine the statistical properties of the temperature variations in the PNW region in order to identify the relevant time scales for this event. These properties are summarized in Fig. 2. Fig. 2(a) displays a comparison between TX, running averages of TX with time windows of 7, 15 and 30 days (TX07, TX15, TX30) for the entire JJA 2021 season and the 1950–2022 and 1991–2020 JJA averages. During this period, TX stays mostly above the seasonal cycle for more than 60 days, with anomalies often larger or much larger than 5 °C. The main event, bringing the unprecedented heat peak at the end of June, lasts less than a week. Two other significant peaks are observed at the end of July and around mid-August, with anomalies up to 2 standard deviations from the 1991–2020 seasonal cycle. Fig. 2(b) shows the 1950–2022 time series of yearly JJA mean, TX30d, TX15d, TX07d and TX01d. With 37.3 °C, 2021 is by far the warmest year in terms of daily maximum temperature, followed by 2022 with 33.1 °C. It is also the warmest year for TX07d, TX15d and JJA mean, with 33.8 °C, 30.4 °C and 25.1 °C, all of them also followed by 2022. However, TX30d is the second warmest, with 28.7 °C, following 2022 at 29 °C. We fitted generalized extreme value distributions (GEV, Coles et al. (2001)) to TX01d, TX07d, TX15d and TX30d, excluding 2021 from the time series, in order to determine return values for those variables (Fig. 2c–e). Since all the time series in Fig. 2(b) have a significant increasing trend of about 0.3 °C per decade (p-values $< 10^{-3}$), we assume a non-stationary GEV specification. The non-stationary location parameter is given by $\mu_t = \mu_0 + \mu_1 \overline{T}_t^{JJA}$, where $\overline{T}_t^{JJA}$ is the JJA global mean surface temperature for year t, obtained from ERA5. Parameters are estimated using maximum likelihood. Estimated values and associated standard errors are listed in Table 1.

Since there is no close-form expression for the return period of an extreme value for a non-stationary GEV distribution, we can assume $\mu = \mu_0$ as a first approximation, and use the expression for the return period τ of a value x: $\tau(x) = 1/(1 - F_X(x))$, where $F_X(x)$ is the GEV probability distribution function evaluated at x. Under this assumption, τ(TX30d) = 215 years, while τ(TX01d) (not shown in figure), τ(TX07d) and τ(TX15d) result to be infinite. We also performed the GEV fit including 2021, finding τ(TX01d) = 14615 years, infinite τ(TX07d), τ(TX15d), and τ(TX30d) = 176 years. Considering the unprecedented nature of the 2021 PNW heatwave at time scales of up to 15 days, one might initially expect a significant reduction in the corresponding return period when incorporating data from 2021 into the analysis. However, our dataset encompasses the summer of 2022, during which the second-highest values of T07d and T15d, and the absolute T30d record in the time series were observed. In contrast, the 2021 record for TX01d was several degrees higher than the 2022 value, so that the inclusion of data from 2021 results in a finite, although very large, return period. It is an established fact in the literature about this event that a GEV analysis of the maximum temperatures produces infinite return time estimates (Bercos-Hickey et al., 2022; Philip et al., 2021; Zeder et al., 2023). Here we replicate the analysis to characterize the event and to have a reference for comparison with the SWG simulations.

Zeder et al. (2023) further observed that return periods obtained from a non-stationary GEV fit tend to be overestimated when the time series is relatively short. This systematic bias becomes particularly pronounced, especially in the context of recent temperature extremes, owing to the prominent global warming signal. Zeder et al. (2023) find infinite return times for analogues of the 2021 PNW heatwave simulated in a large ensemble of climate simulations, and observe that the issue can be only partially mitigated using a Bayesian approach rather than maximum likelihood estimation. In light of these findings, while the estimated infinite return times do indicate an extreme event occurring within the context of a background trend, they do not necessarily imply that the phenomenon was so rare as to be considered impossible before its occurrence, especially when dealing with relatively short historical data.

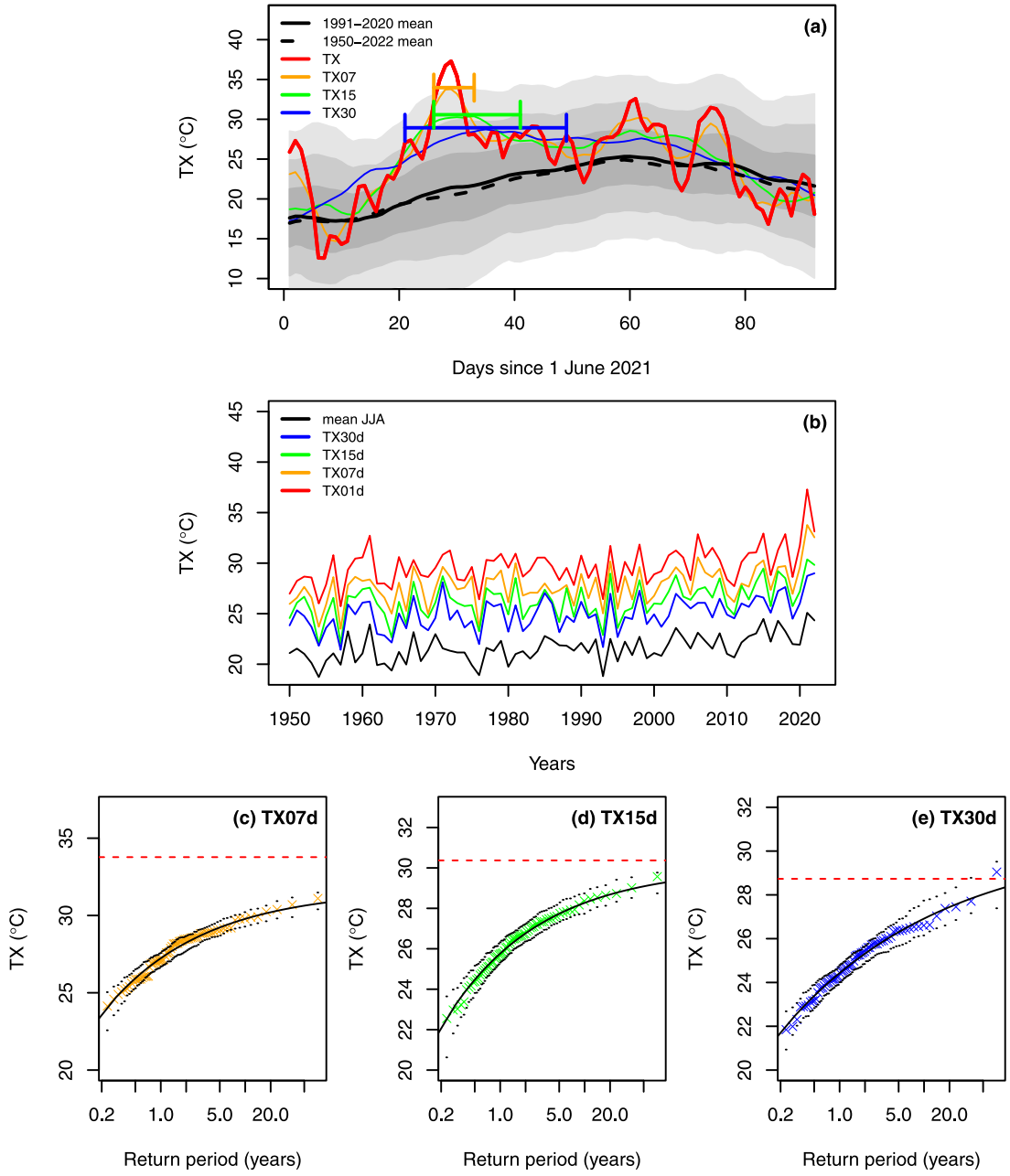

Fig. 2. (a): time series of daily maximum near-surface temperature for JJA 2021 (red line) compared to the 1991–2020 seasonal cycle smoothed with a 3-day running mean (black line). The three grey shaded areas represent 1, 2, and 3 standard deviations from the mean. The orange, green and blue lines represent daily values of TX07, TX15 and TX30. The horizontal segments show the periods corresponding to TX07d, TX15d and TX30d. (b): time series of TX01d (red), TX07d (orange), TX15d (green) and TX30d (blue), and JJA mean (black). (c)–(e): return level plots for TX07d, TX15d and TX30d. Units are in °C. Coloured crosses represent observed values, full lines the return levels computed from a GEV fit for the period 1950–2022 excluding 2021. The dashed lines are the 95% confidence intervals of the return levels. Horizontal dashed red lines are the corresponding 2021 values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Methods

3.1. Analogues of circulation

We compute circulation analogues using daily average ERA5 Z500 fields on the region [30–70 N; 60–180 W], corresponding to the largest rectangle drawn in Fig. 1(a) . For each day between 1 January 1950 and 3 September 2022, we obtain the best 20 analogues based on the Euclidean distance between Z500 fields, within 30 calendar days before or after the target date, excluding the year of the target date. The analogue search is performed using the open source software CASTf90 (Circulation Analogue Simulation Tool in fortran90), available online at https://a2c2.lsce.ipsl.fr/index.php/licences/file/castf90?id=3.

Fig. 3 shows statistics for Z500 analogues for the 30 days between June 21 and July 21 2021, found in the two periods 1950–1999 and 1971–2022. We consider the latter as representative of the current climate and we call it *factual* period, while the former will be considered as *counterfactual* with respect to the current climate.

Fig. 3(a) shows the distribution of the analogue quality in the two periods, measured by the Euclidean distance between each Z500 daily field and its analogues. The two distributions are statistically indistinguishable (two-sided Kolmogorov–Smirnov test p-value = 0.46), suggesting that the quality of the analogues of the 2021 PNW heatwave

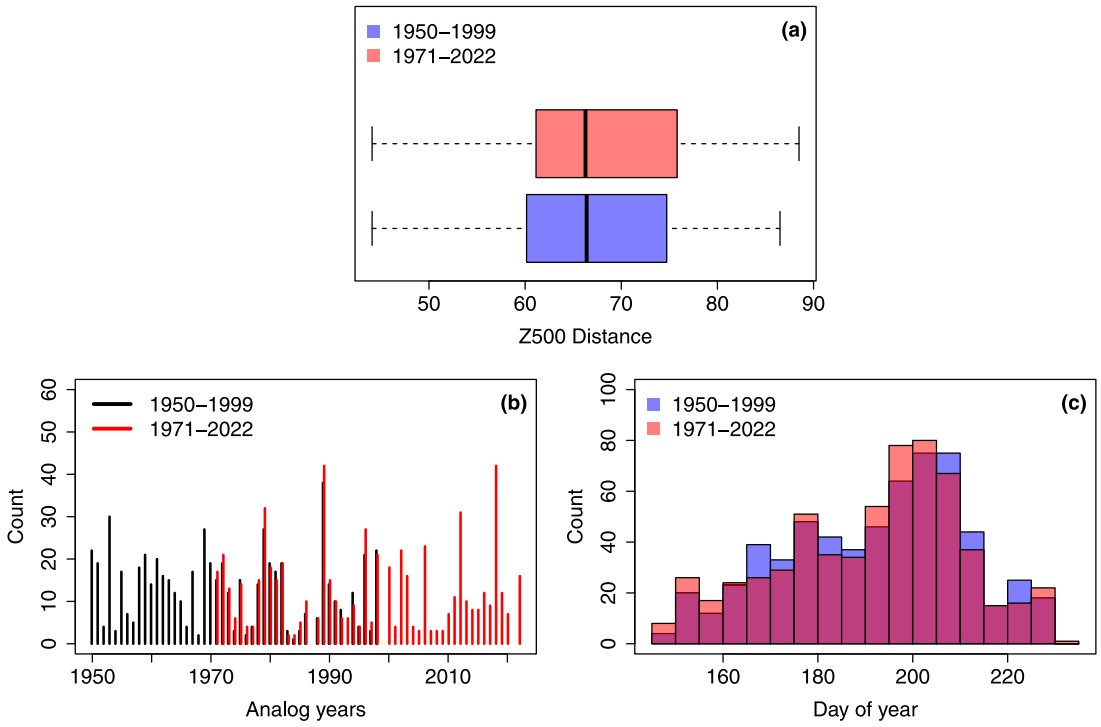

Fig. 3. (a): empirical probability distribution of distances of the best 20 analogues of $Z500$ between June 21 and July 21, 2021; analogues are constrained to be searched within a 30-days window around the target date. The distances are computed in the counterfactual (1950–1999, blue) and factual (1971–2022, red) period, respectively. (b): distribution of yearly number of analogues of $Z500$ for days between June 21 and July 21, 2021, for analogues chosen in the counterfactual (black) and factual (red) periods. (c): histograms representing the number of analogues binned according to day of year in the counterfactual (blue) and factual (red) periods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Point estimates and associated standard errors (in parentheses) of the GEV parameters fitted to TX01d, TX07d, TX15d and TX30d, excluding and including 2021 values.

Excluding 2021				
	Location	Loc. trend	Scale	Shape
TX01d	29.12 (0.20)	0.66 (0.18)	1.59 (0.15)	−0.35 (0.07)
TX07d	27.23 (0.21)	0.77 (0.16)	1.64 (0.15)	−0.37 (0.08)
TX15d	25.77 (0.21)	0.63 (0.19)	1.64 (0.15)	−0.38 (0.08)
TX30d	24.43 (0.18)	0.68 (0.18)	1.39 (0.12)	−0.23 (0.07)
Including 2021				
	Location	Loc. trend	Scale	Shape
TX01d	29.01 (0.20)	0.76 (0.20)	1.60 (0.14)	−0.14 (0.06)
TX07d	27.18 (0.21)	0.88 (0.20)	1.65 (0.15)	−0.26 (0.07)
TX15d	25.79 (0.21)	0.73 (0.17)	1.65 (0.15)	−0.37 (0.07)
TX30d	24.45 (0.18)	0.74 (0.18)	1.40 (0.13)	−0.23 (0.07)

has not significantly changed between the counterfactual and the factual climate. Fig. 3(b) shows the number of analogues found each year: no detectable trend of the analogue number against time is present either in the counterfactual (p-value = 0.44) or in the factual period (p-value = 0.98). Fig. 3(c) compares the distributions of the number of analogues by day of the year in the two periods: since the difference between the two distributions is not statistically significant (χ^2-test p-value = 0.72), we conclude that there is no clear shift in the seasonal distribution of the analogues.

3.2. Stochastic weather generator

Stochastic Weather Generators (SWGs) are tools designed to simulate ensembles of trajectories of the variable of interest (in our case, daily maximum temperature) based on statistical techniques rather than running a full climate model. Here, a trajectory is a time series of the simulated variable, from a prescribed initial condition. We consider the SWG introduced by Yiou (2014), which produces ensembles of Z500 trajectories based on resampling the analogues of the event of interest. Each sequence of Z500 fields corresponds to a time series of TX over the region of interest.

We simulate an event starting at time t_0 and ending at time $t_0 + L$, for $L \in \{7, 15, 30\}$ days. We denote $\mathrm{Z500}_t$ and TX_t the Z500 field and the corresponding maximum daily temperature at time t. For each day $t_0 \le t \le t_0 + L$, we select the best $K = 20$ Z500 analogues falling within 30 days before or after t. Day $\tilde{t}$ is selected among the ensemble of $K + 1$ days containing t and its K best analogues via a random sampling with weights $w^{(k)} = w^{(k)}_{opt} \cdot w^{(k)}_{cal}$. Here $w^{(k)}_{opt}$ are weights based on an optimization observable, and $w^{(k)}_{cal}$ are calendar weights inversely proportional to the time lag between t and $\tilde{t}$ in calendar days, $|t - \tilde{t}|$. In particular, $w^{(k)}_{cal} \propto \exp(\alpha_{cal}|t - \tilde{t}|)$, where $\alpha_{cal} \ge 0$ weighs the importance given to seasonality: larger values of α_{cal} privilege analogues that are closer to the target date in the seasonal cycle. In general, introducing calendar weights ensures that time in the simulations moves forward, as the resampling will not get stuck on periods characterized by optimal values of the optimization observable.

The next day in the simulation is taken to be $t' = \tilde{t} + 1$, where $\tilde{t}$ is the analogue selected with the importance sampling. This resampling of the analogues is repeated for L steps, until a complete sequence of Z500 fields and corresponding values of TX has been obtained. Then, the L-day mean TXLd can be obtained and compared to reanalysis or observed values. The entire procedure is repeated S times to obtain surrogate ensembles of analogue trajectories. The SWG can be viewed as a Markov chain that simulates ensembles of TXLd trajectories, sampling the right tail of the temperature distribution, thanks to the importance sampling of the analogues in the intermediate step. The analogue sequences yield a form of constrained randomness that makes them more noisy than observed trajectories. Therefore, temporal autocorrelation values are smaller than for observations. A schematic representation of the algorithm for the SWG is shown in Fig. 4.

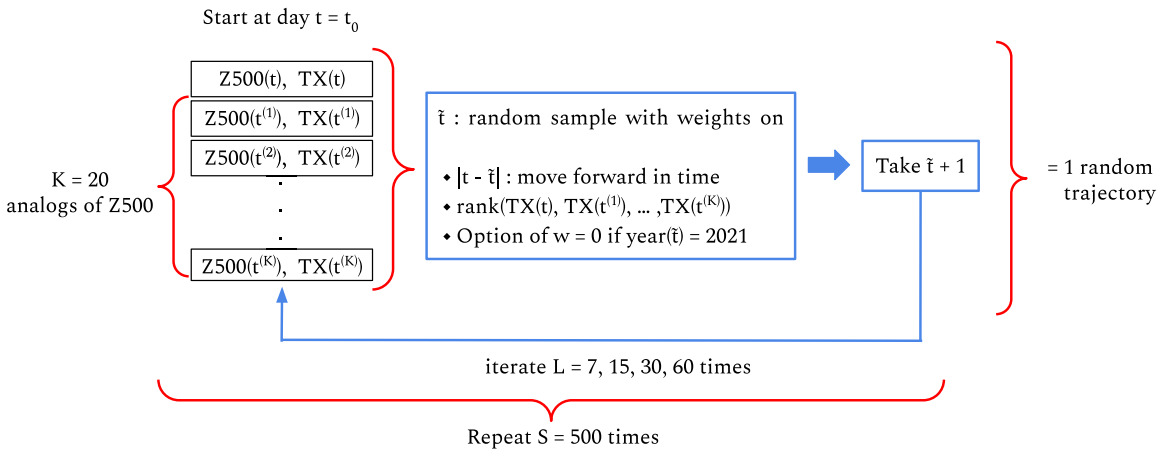

Fig. 4. Schematic of the dynamic SWG.

The idea of using an optimization observable to simulate rare and extreme events was introduced by Ragone et al. (2018), who proposed a large deviation algorithm based on importance sampling. Trajectories that do not optimize the observable are suppressed and replaced by perturbations of more optimal trajectories. On the contrary, here we use the optimization observable to nudge the trajectories in the desired direction. The choice of the optimization observable and the definition of the weights depend on the type of event under consideration. For example, to give high importance to the atmospheric circulation, one can sort the K best analogues of time t according to the spatial correlation between each $Z500$ field and $Z500_t$: in this way, among analogues that are optimal in terms of Euclidean distance, the ones with the most spatially similar $Z500$ patterns are favoured. This rule strongly favours the choice of date t, since by definition $Z500_t$ has zero distance from and unit correlation with itself, which is why the day itself is excluded from the analogue selection.

In case the event of interest is a heatwave or a cold spell, the optimization observable of choice can be the daily average, minimum or maximum temperature, spatially averaged over a region of interest. In our case, we choose maximum daily temperature averaged over [44 N– 52 N; 116–124 W] as the optimization observable. Then, we sort the values of TX in the $K+1$ candidate dates (K best analogues plus day t) in decreasing order, and denote their rank R_k with $k = 1, \dots K+1$. The optimization weights are defined as $w_{obs}^{(k)} = Ae^{-\alpha_{TN} R_k}$, where $A = e^{-\alpha_{TN}}(1-e^{-K\alpha_{TN}})(1-e^{-\alpha_{TN}})^{-1}$ is a normalization constant, and $\alpha_{TN} \geq 0$ controls how flexible the SWG is at selecting analogues characterized by lower temperatures (or, in general, a less optimal value of the observable).

To tune the value of α_{TM}, we initially ran simulations using the largest domain described in Section 3.1 and fixed $\alpha_{cal} = 4$. The probability distributions of TX07, TX15 and TX30 from sets of 500 trajectories are shown in Fig. 11 in Appendix A as a function of α_{TN}. We choose $\alpha_{TN} = 1$ for all three sets of simulations, since this is the value for which the increase in the simulated value as a function of α_{TN} starts to show saturation.

We initially considered three possible domains centred over the PNW for the definition of the analogues, corresponding to the three red rectangles in Fig. 1(a). The inner domain [40–60 N; 100–140 W] includes the core of the anticyclone; the intermediate [35–65 N; 90–160 W] domain includes most of the anticyclone and part of the two low pressure areas to the SE and SW, defining the Omega block; the outer domain [30–70 N; 60–180 W] includes the entire Omega blocking configuration and covers most of North America, except for Greenland and the extreme North of Canada. We ran simulations with $\alpha_{cal} = 4$ and $\alpha_{TN} = 1$ using analogues obtained with each domain.

We found that the number of simulations reaching or exceeding 2021 temperature values is consistently lower for the two smaller domains, compared to the largest. Thus, we concluded that the largest domain is better at sampling analogues that favour extreme heat condition over the region because it includes the key large-scale features of the event. This domain includes the entire blocking structure, thus selecting analogues that have a similar large-scale configuration, plausibly linked to the drivers of the PNW heatwave.

We perform simulations using analogues from the counterfactual (1950–1999) and the factual (1971–2022) periods; the factual simulations are run both including and excluding the 2021 event from the possible $K+1$ resampling dates.

In the following, we will denote the counterfactual and factual periods as C and F respectively; simulations including information from the 2021 event will be labelled as E ("Event"), and simulation excluding 2021 will be labelled nE ("no Event"). Thus, simulations considering analogues from the counterfactual period will be denoted C.nE, while simulations from the factual period including or excluding 2021 information will be denoted F.E and F.nE, respectively.

4. Results

We first evaluate how climate change (counterfactual C vs. factual F simulations) affects the probability of reaching or exceeding the 2021 temperature values, for the three identified time scales. We also assess whether the event (for the three time scales) could have been anticipated from prior observations, i.e. excluding information from 2021, apart from the initial conditions.

We perform $S = 500$ simulations of the 2021 PNW heatwave using the SWG with simulation lengths of 7, 15 and 30 days. Each simulation is initialized at the beginning of the warmest period of corresponding duration, i.e. t_0 =June 26 for TX07 and TX15 and June 21 2021 for TX30.

Fig. 5 summarizes the results of the three SWG configurations described above for TX07 (a), TX15 (b) and TX30 (c). For all SWG configurations, the simulations produce mean TX values that are larger relative to the 2021 TX values for longer trajectories. No trajectory warmer than 2021 can be simulated with analogues from the counterfactual period with lengths of 7 days, while 25 warmer trajectories are obtained for both the 15 and 30 day simulations. For TX30, 23 of the 25 trajectories warmer than the 28.73 °C observed in 2021, are also warmer than the 29.0 °C characterizing the warmest 30 days in 2022. This is possible because the 2021 value of TX30 is less anomalous than TX07 and TX15 with respect to the temperature distributions of the counterfactual period. However, even at this time scale, the event remains very difficult to simulate using 1950–1999 data, considering that only 5% of the trajectories reach the 2021 value.

The fact that the SWG struggles or fails at reproducing the 2021 event using analogues from the counterfactual period (1950–1999) shows that this event was extremely unlikely in a climate less affected by global warming due to anthropogenic forcing than in recent years.

The TX07 of 2021 is never reached with SWG simulations using analogues from the factual period excluding information from 2021 (F.nE simulations). This shows that, at a short time scale around the peak temperature (≈ 7 days), this event is an outlier even in the present climate and could hardly be anticipated from previous information.

The TX15 case is the most interesting, as the 2021 heatwave at this time scale is characterized by an infinite return period estimated from the GEV fit, but it can be reproduced by the SWG in a way similar to TX30. In particular, all the trajectories are warmer than 2021 for F.E and 41% for F.nE. The fact that some of trajectories can be warmer than 2021 in F.nE shows that the SWG approach can simulate events that are possible (because they have been observed), yet are outside of the range that can be predicted through GEV estimates.

Our simulations excluding 2021 display larger temperature differences between the counterfactual and factual periods, compared to the background warming in TX over the considered region (+0.69 °C), with +1.62 °C for TX07d and TX15d, and +1.02 °C for TX30d.

These results overall show that the PNW heatwave has become more likely in the recent decades, an expected fact due to the effect of global warming. They also suggest that, even in the current climate, this heatwave was an exceptional event, especially at time scales of one week or less. The fact that the analogue quality and frequency have not increased and their seasonality has not significantly shifted between counterfactual and factual periods (Fig. 3) suggests that the increased likelihood is not linked to a long-term change in the atmospheric circulation, but rather to the combination of co-occurring drivers and warming due anthropogenic forcing.

Fig. 6 shows the ensembles of trajectories warmer than the 2021 heatwave for TX15 for F.E and F.nE, compared to 2021 observations and to the 1991–2020 seasonal cycle, smoothed with a 3-day running mean. During the period between June 26 and July 1, 2021 values can only be reached in the F.E simulations, and never exceeded, since these are the highest TX values in the entire time series. During the following 10 days, 2021 values are exceeded by the majority of the trajectories, despite observations being between 5 °C and 7 °C above the seasonal cycle.

Next, we analyse the atmospheric patterns that prevail during the simulated heatwaves. We compare the composites of the analogues sampled by the SWG to the average of observations between June 26 and July 3, corresponding to the period covered by the 7-day simulations including the initial condition. Fig. 7 considers standardized Z500 anomalies over the entire Northern hemisphere. The composite of the 2021 event shows the drivers detected by previous studies (Wang et al., 2022; Bartusek et al., 2022; Qian et al., 2022): the wavenumber-4 hemispheric disturbance, with positive Z500 anomalies over Eastern Europe, Eastern Asia, PNW and the North Atlantic; the Rossby wave train across the Pacific, in phase-locking with the PNW anticyclone; and the negative anomaly over the Arctic. The yearly distribution of analogues for the peak phase of the heatwave reveals that F.E simulations sample almost all analogues from 2021, while F.nE are dominated by 2015, 1979, 2022 and 2002. The general pattern is well reproduced in simulations that incorporate data from 2021, with an amplified anomaly over the PNW. This is expected, given that most of the analogues are drawn from 2021. Conversely, factual simulations that exclude 2021 struggle to replicate both the broader hemispheric pattern and the depression observed over the Arctic. Nevertheless, they do manage to generate a positive Z500 anomaly over the PNW, closely resembling the observed conditions. They also reflect the Pacific Rossby wave train in phase locking with the PNW anticyclone, a configuration often associated with blocking anticyclones (Tyrlis and Hoskins, 2008). This could be a recurring factor in summer heatwaves over the PNW, contributing to directing atmospheric rivers towards the region (Lin et al., 2022), and reinforcing local subsidence anomalies (Qian et al., 2022).

Fig. 8 shows composites of TX standardized anomalies over the PNW region. From the histogram of the day of year of the analogues (Fig. 8(b)) it becomes evident that F.E simulations use almost only analogues sampled from the peak of the event itself. This is expected, since these observations are the highest values in the TX time series, and many simulated trajectories match such values (see Fig. 6(a) in Appendix A). F.nE simulations sample analogues from the last 10 days of July, as observed in Fig. 12. The analogue composites show the difficulty of the F.nE to reach 2021 values, especially inland, compared to F.E simulations.

Finally, Fig. 9 shows composites of TCWV standardized anomalies. During the selected dates, the atmospheric river conveyed from the Western Pacific by the Rossby wave train had already made landfall, bringing a large amount of water vapour to Canada. Further transport across the Pacific is noticeable by the positive TCWV anomaly between Hawaii and the PNW, while a dry patch is associated with the high pressure located halfway between Hawaii and Japan. Naturally, F.E simulations closely reproduce this pattern, as they resample days from the observation period. However, a similar pattern – even though with anomalies of smaller amplitude – can also be noticed in composites from F.nE simulations, with a dry area NW of Hawaii and the transport of water vapour towards the PNW. The SWG thus reflects this driver of extreme heatwaves in the PNW, selecting analogues characterized by a Rossby wave train across the Pacific, that in the 2021 event led to the landfall of an atmospheric river inside the anticyclonic dome.

5. Conclusions

The 2021 PNW heatwave was a record-shattering event, with extreme impacts on people and ecosystems. We have used a stochastic weather generator with importance sampling and ERA5 reanalysis data to simulate heatwaves that match or surpass multi-day average temperatures recorded during this event.

Our results confirm the role of global warming already found in other studies: while it is impossible for the SWG to produce heatwaves reaching 2021 values by sampling analogues from the counterfactual climate, analogues from the factual period enable to reproduce the 2021 heatwave even when excluding the event itself, at least for time scales of 15 and 30 days. At a time scale of 7 days, it is impossible to produce heatwaves as warm as or warmer than 2021 without including the event itself, and even in the latter case it is very difficult. This shows that, at such a short time scale, the 2021 PNW heatwave was very rare and unlikely to be matched even in the current climate. On the other hand, at a 30-day time scale, 2021 is surpassed by many stochastically generated trajectories even when excluding analogues from the event. Indeed, an even stronger 30-day heatwave affected the region in 2022. In the intermediate 15-day case, 4% of the trajectories simulated excluding 2021 from the resampling, are warmer than the actual event. This shows that events of comparable or larger magnitude at this time scale are rare but not impossible in the current climate. Trajectories including 2021 analogues lead to a much higher probability of reaching or surpassing the event. This largely supports the very recent findings of Lucarini et al. (2023) based on a large deviation analysis. When considering a 15-day temperature average, the authors argued that the 2021 event is an unlikely but possible manifestation of climate variability, whose probability of occurrence is greatly amplified by the ongoing climate change. When considering a 7-day average, they found that a similar conclusion holds for some locations within the heatwave region, while at one of the selected grid points the magnitude of the heatwave was such as to make it impossible to draw robust conclusions.

While the role of global warming in exacerbating heatwaves at mid-latitudes is established, the 2021 PNW heatwave was also the result of the combination and co-occurrence of drivers and nonlinear land–atmosphere interactions. These drivers include an Omega blocking

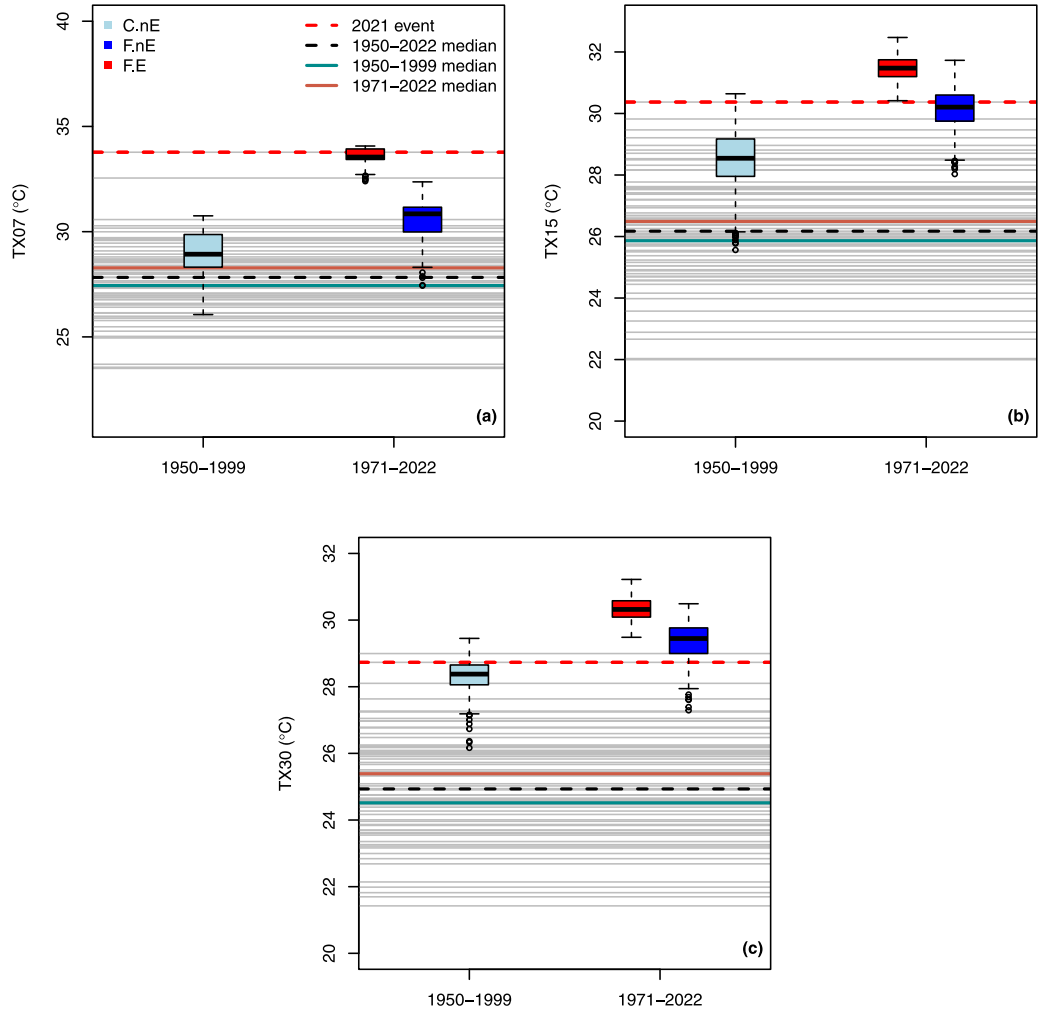

Fig. 5. Empirical probability distributions of means of simulated TX computed over 7 days (a), 15 days (b) and 30 days (c). The grey horizontal lines represent the observed values between 1950 and 2022. The dashed red line is the value for 2021, the dashed black line is the median value. The green and orange lines show median values for the counterfactual and factual periods, respectively. Red boxes: simulations that include information from 2021. Light blue and blue boxes: simulations that exclude all information from 2021 using analogues from the counterfactual and factual periods, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

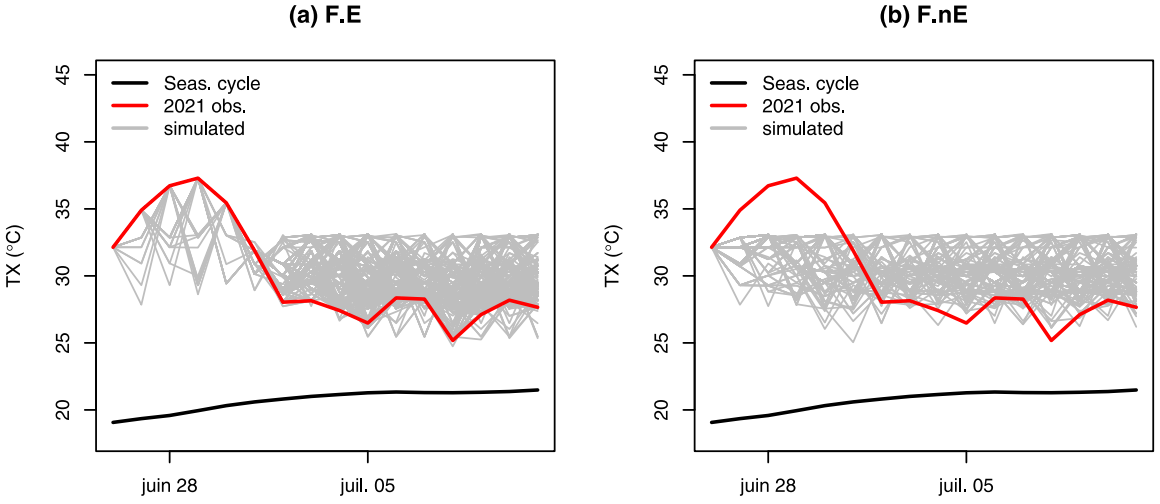

Fig. 6. Trajectories of TX for 15-day simulations initialized on June 26, 2021. Simulations use analogues from the factual period including (a) and excluding (b) 2021. The red lines show 2021 observations, the black line the 1991–2020 seasonal cycle.

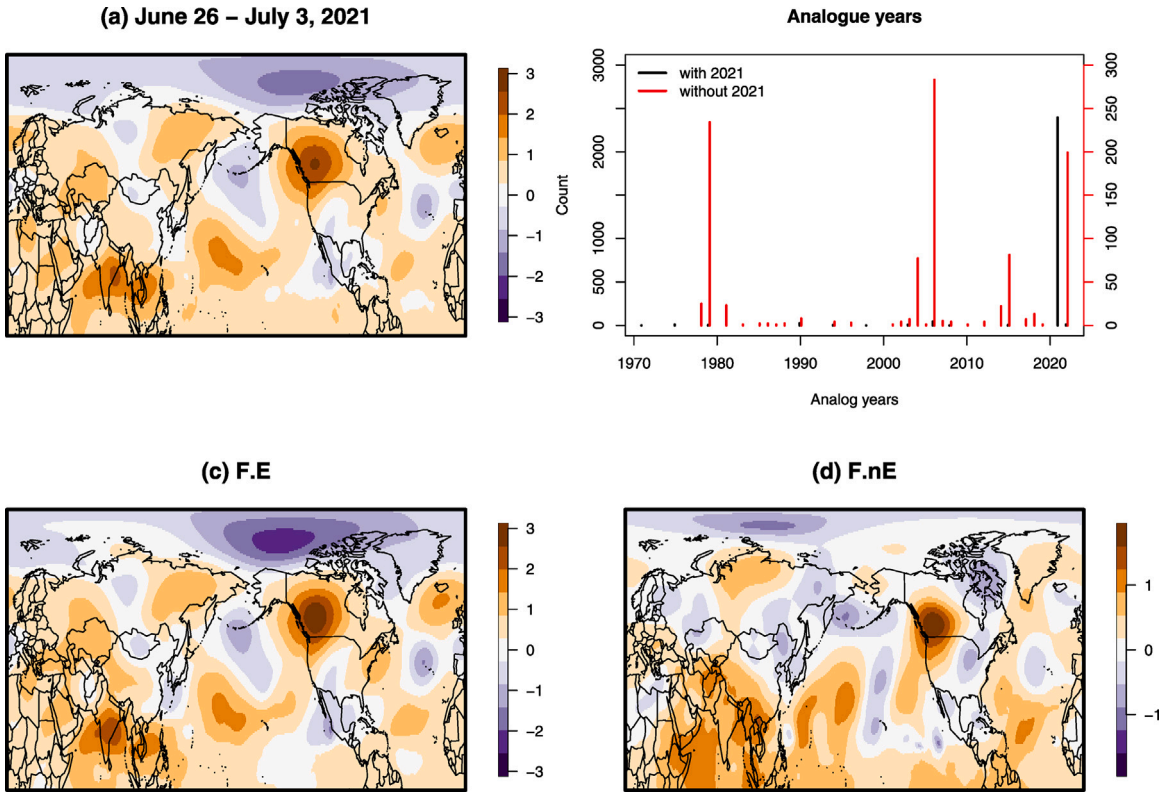

Fig. 7. Summary of Z500 standardized anomaly composites for the analogues used in the simulation for the period June 26–July 3, corresponding to the peak of the heatwave over the PNW. (a): standardized Z500 anomaly composite from ERA5 data; (b): number of analogues per year; (c) and (d): standardized Z500 anomaly composites of analogues for the simulations including and excluding 2021, respectively.

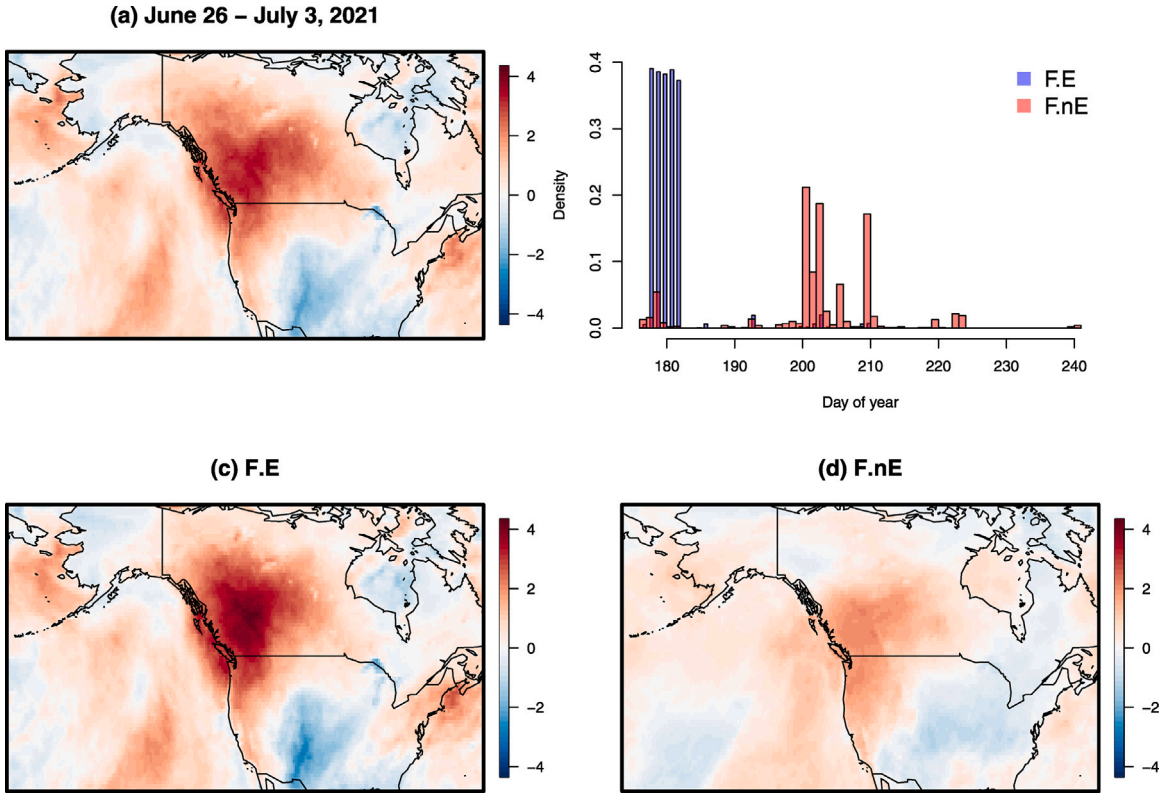

Fig. 8. Summary of the maximum temperature standardized anomaly composites for the analogues used in the simulation for the period June 26–July 3, corresponding to the peak of the heatwave over the PNW. (a): composite maximum temperature standardized anomalies from ERA5 data; (b): distribution of day of year of analogues. (c) and (d): composite maximum temperature standardized anomalies of analogues for the simulations including and excluding 2021, respectively.

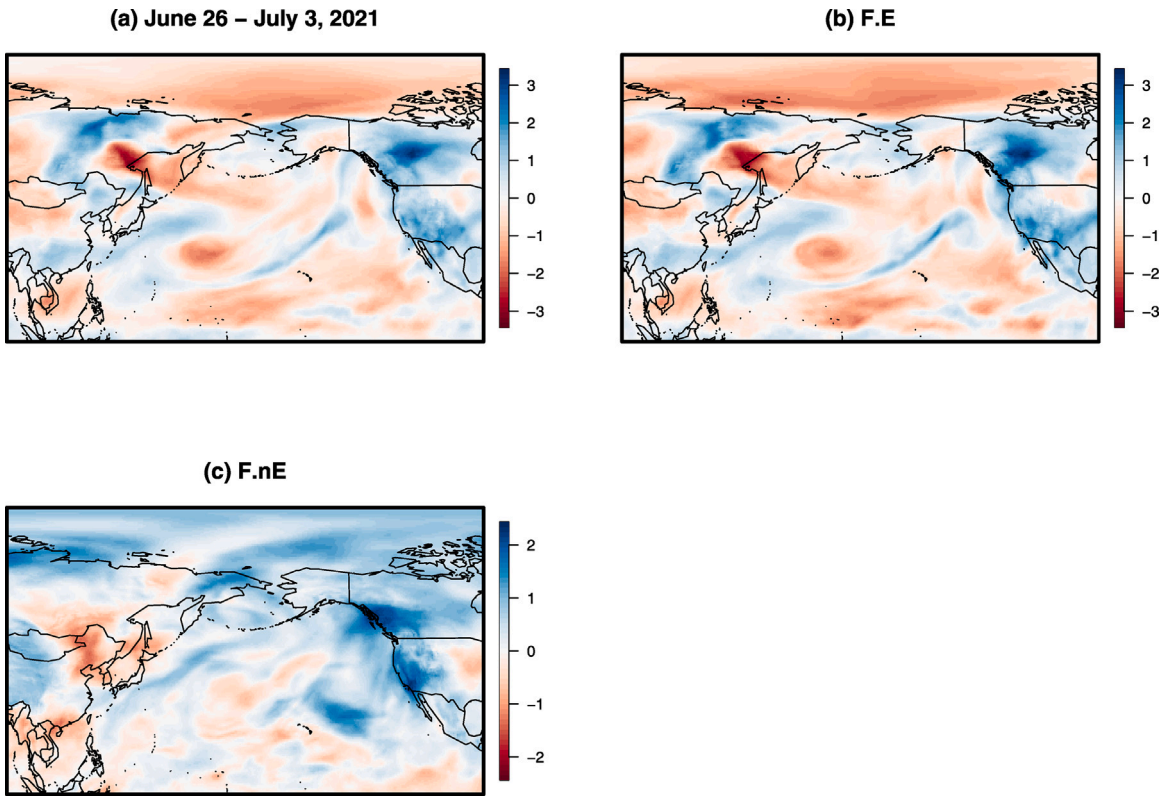

Fig. 9. Summary of the total column water vapour (TCWV) standardized anomaly composites for the analogues used in the simulation for the period June 26–July 3, corresponding to the peak of the heatwave over the PNW. (a): TCWV standardized anomaly composite from ERA5 data. (b) and (c): TCWV standardized anomaly composites of analogues for the simulations including and excluding 2021, respectively.

anticyclone on the PNW embedded in a hemispheric wavenumber 4 configuration, a split of the polar vortex triggering a sequence of quick weather pattern changes over North America, and a Rossby wave train in phase locking with the Omega structure, driving an atmospheric river towards the region.

Analogue composites for the peak of the simulated heatwaves excluding analogues from 2021 show some similarities to the event itself: the Omega blocking over the PNW, the negative geopotential anomaly over the Azores, and the Rossby wave train conveying large amounts of water vapour across the Pacific are still visible. However, important large scale differences can also be observed, e.g., the hemispheric disturbance is not overall in phase with the 2021 event, especially over Eurasia, and the deep Arctic negative geopotential anomaly is not present.

The SWG allowed us to simulate an extremely rare event at reduced computational cost, and using a relatively short reanalysis time series compared to the event's very long return period. This advantage is also balanced by some shortcomings: the SWG is a purely statistical method, therefore our results produce reliable statistics of TX, but some unrealistic properties in the simulated trajectories, e.g. temporal autocorrelation. Moreover, we only assess drivers of the event in a posterior analysis, while it would be interesting to incorporate some of the involved large-scale physics in the simulation. This could be achieved by adding a further weight in the stochastic sampling to measure the similarity of specific large-scale circulation features. To this end, one could leverage machine learning techniques capable of decomposing atmospheric fields and projecting them onto low-dimension time series (Fery et al., 2022).

Our results could be extended by using climate simulations from CMIP6. While reanalysis allowed us to analyse the 2021 NWP heatwave in the context of the present and recent climate, climate models would make it possible to evaluate its likelihood and the magnitude of worst case scenario heatwaves under future global warming scenarios. Moreover, the historical period of CMIP6 models traces back to 1850, providing a counterfactual period closer to actual pre-industrial conditions than the one we selected for the reanalysis data.

The PNW 2021 heatwave broke historical records at sub-monthly time scales, producing temperatures that would have been previously considered unattainable over the PNW. Using simulations from a stochastic weather generator, we found that global warming has amplified the heatwave's magnitude and that many large-scale features of the atmospheric circulation in the Northern Hemisphere during this event are recurrent – if not typical – PNW heatwave drivers. Their co-occurrence and interaction alone thus cannot explain how extreme this event was. The 2021 heatwave remains a worst-case scenario for extreme heat periods up to 15 days, while the present climate could foster more extreme events at the monthly and seasonal scales.

CRediT authorship contribution statement

Flavio Maria Emanuele Pons: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Pascal Yiou:** Conceptualization, Methodology, Software, Writing – review & editing. **Aglaé Jézéquel:** Methodology, Writing – review & editing. **Gabriele Messori:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: none.

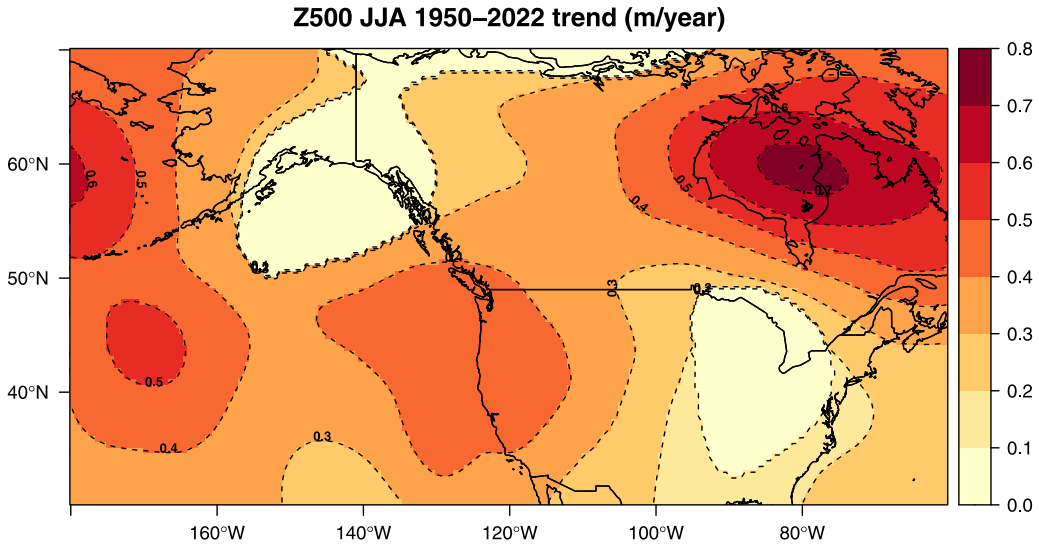

Fig. 10. Observed 1950–2022 trend in mean JJA geopotential height in ERA5. Units are metres per year.

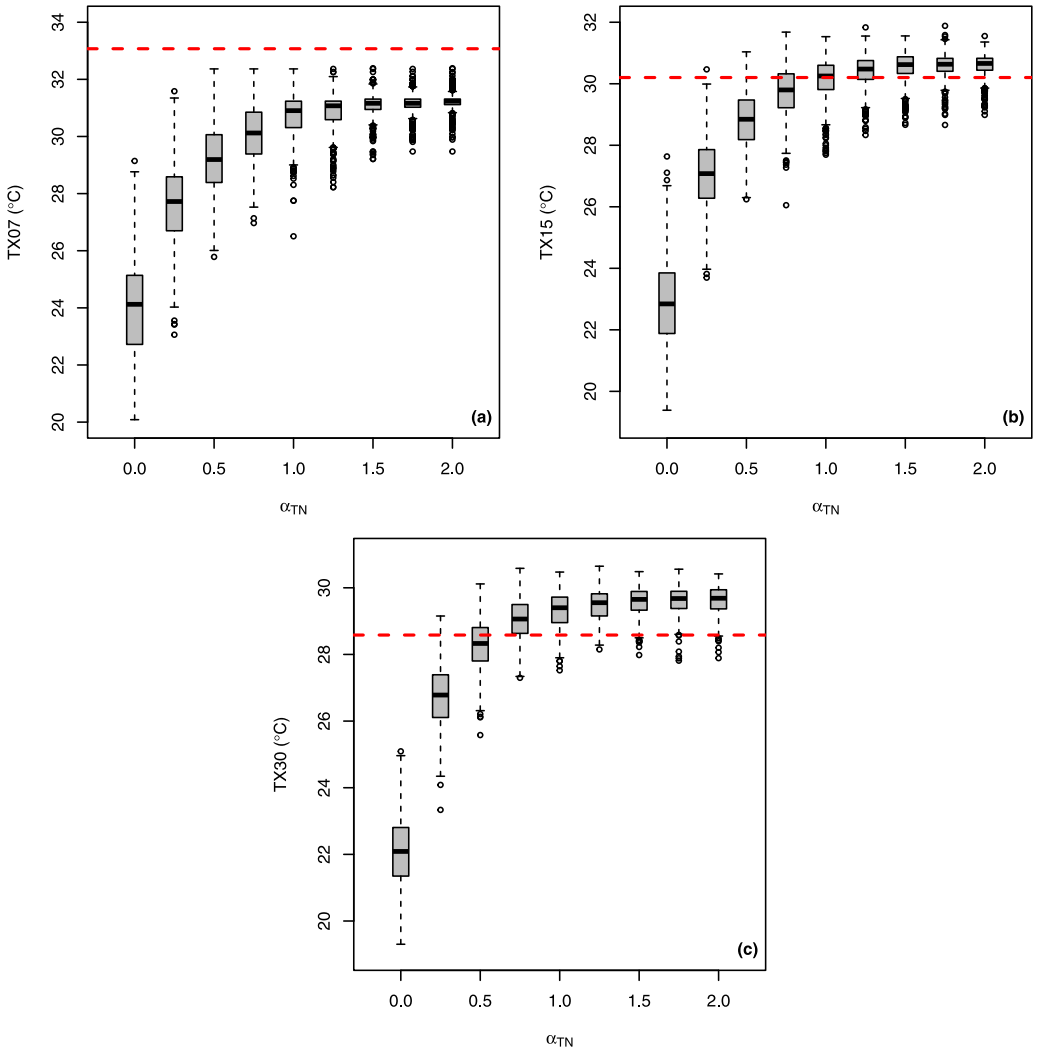

Fig. 11. Boxplots of simulated temperatures for the hottest 7 and 15 days periods starting on June 26, and the 30-day period starting on June 21, as a function of the parameter α_{TN}, with $\alpha_{cal} = 4$. The red dashed line is the average TX between June 21 and July 21, 2021.

Table 2
Number of analogues, counted with repetition, from the 10 years providing most analogues for simulations including and excluding 2021.

Including 2021										
year	2021	2022	2018	2006	2002	2003	2012	1998	1975	1979
# analogues	3303	640	546	414	407	285	222	220	199	166
Excluding 2021										
year	1979	2004	2006	2022	2014	2015	1978	2017	1990	2018
# analogues	758	561	397	359	211	136	111	97	74	48

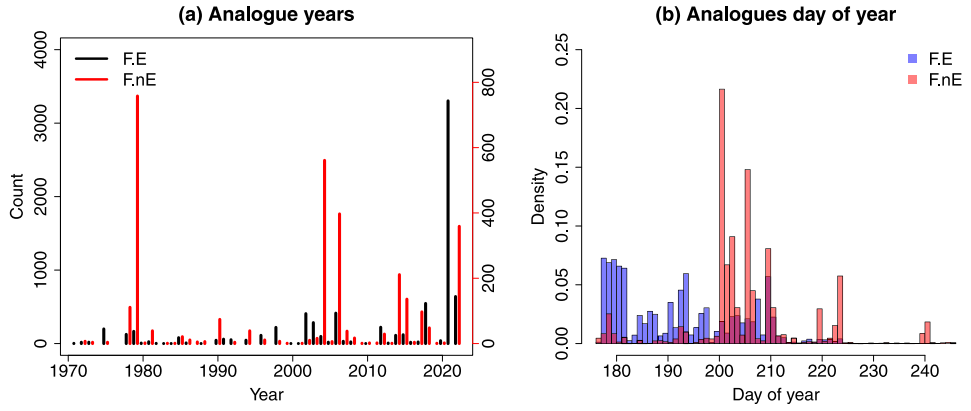

Fig. 12. Time distributions of the analogues used in the 15 days simulations with TX15 warmer than 2021. (a): distribution of analogues across years; black lines are simulations including the event in 2021, red lines are simulations excluding 2021. (b): distribution of analogues per day of year. Blue bars are simulations including the event, red bars are simulations excluding 2021. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Data availability

ERA5 data is available from the Climate Data Store (CDS) of the Copernicus Climate Change Service (Hersbach et al., 2018).

Acknowledgements

The authors acknowledge Rodrigo Caballero and Davide Faranda for useful comments and discussions. This work has received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101003469 (XAIDA: PY, FP, AJ), ERC grant agreement No. 948309 (CENÆ: GM), and the grant ANR-20-CE01-0008-01 (SAMPRACE: PY). ANR, Agence Nationale de la Recherche, i.e. the National French Research Agency.

Appendix A

A.1. Z500 trends

See Fig. 10.

A.2. Choice of the importance sampling parameter

See Fig. 11.

A.3. Analogue statistics

Here we report some statistics about the analogues sampled by the SWG from the factual period to simulate the TX15 trajectories warmer than the observed TX15d value of 2021. All 500 F.E simulated trajectories and 203 out of 500 F.nE trajectories surpass the 2021 value, for a total of 7500 and 3045 sampled analogues, respectively. The number of unique analogues (i.e. counted without repetition) used in the simulation is 214 to simulate the 7500 F.E days and 143 to simulate the 3045 F.nE days. Fig. 12 shows the temporal distribution of the analogues used to obtain the simulations warmer that 2021, counted with repetition over the years and according to the day of year. For the F.E simulations, almost half of analogues is sampled from 2021, while for F.nE the analogue count is dominated by a few of the years in the sample. The 10 years providing most analogues for both sets of simulations are summarized in Table 2. Concerning the day of year distribution, for F.nE simulations the peak of the distribution is in the last 10-day period of July, at the peak of the temperature seasonal cycle. For F.E simulations the distribution is skewed towards the end of June and the beginning of July, due to the large number of analogues sampled from the 2021 event itself.

References

Bartusek, S., Kornhuber, K., Ting, M., 2022. 2021 North American heatwave amplified by climate change-driven nonlinear interactions. Nature Clim. Change 12 (12), 1143–1150.

Bercos-Hickey, E., O'Brien, T.A., Wehner, M.F., Zhang, L., Patricola, C.M., Huang, H., Risser, M., 2022. Anthropogenic contributions to the 2021 Pacific northwest heatwave.

Bumbaco, K.A., Dello, K.D., Bond, N.A., 2013. History of Pacific northwest heat waves: Synoptic pattern and trends. J. Appl. Meteorol. Climatol. 52 (7), 1618–1631.

Christidis, N., Stott, P.A., 2015. Changes in the geopotential height at 500 hPa under the influence of external climatic forcings. Geophys. Res. Lett. 42 (24), 10–798.

Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values, vol. 208, Springer.

Dacre, H.F., Clark, P.A., Martinez-Alvarado, O., Stringer, M.A., Lavers, D.A., 2015. How do atmospheric rivers form? Bull. Am. Meteorol. Soc. 96 (8), 1243–1255.

Fery, L., Dubrulle, B., Podvin, B., Pons, F., Faranda, D., 2022. Learning a weather dictionary of atmospheric patterns using latent Dirichlet allocation. Geophys. Res. Lett. 49 (9), e2021GL096184.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2018. ERA5 hourly data on single levels from 1959 to present.. URL https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

Jézéquel, A., Yiou, P., Radanovics, S., 2018. Role of circulation in European heatwaves using flow analogues. Clim. Dyn. 50 (3), 1145–1159.

Klein, T., Torres-Ruiz, J.M., Albers, J.J., 2022. Conifer desiccation in the 2021 NW heatwave confirms the role of hydraulic damage. Tree Physiol. 42 (4), 722–726.

Lin, H., Mo, R., Vitart, F., 2022. The 2021 western North American heatwave and its subseasonal predictions. Geophys. Res. Lett. 49 (6), e2021GL097036.

Lucarini, V., Galfi, V.M., Riboldi, J., Messori, G., 2023. Typicality of the 2021 Western North America summer heatwave. Environ. Res. Lett. 18 (1), 015004.

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.

Neal, E., Huang, C.S., 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), e2021GL097699.

NOAA, 2022. https://www.cpc.ncep.noaa.gov/data/teledoc/np.html. (Last Access 2 September 2022).

Overland, J.E., 2021. Causes of the record-breaking pacific northwest heatwave, late june 2021. Atmosphere 12 (11), 1434.

Philip, S.Y., Kew, S.F., van Oldenborgh, G.J., Anslow, F.S., Seneviratne, S.I., Vautard, R., Coumou, D., Ebi, K.L., Arrighi, J., Singh, R., van Aalst, M., Pereira Mafgidan, C., Wehner, M., Yang, W., Li, S., Schumacher, D.L., Hauser, M., Bonnet, R., Luu, L.N., Lehner, F., Gillet, N., Tradowsky, J., Vecchi, G.A., Rodell, C., Stull, R.B., Howard, R., Otto, F.E.L., 2021. Rapid attribution analysis of the extraordinary heatwave on the Pacific coast of the US and Canada June 2021. Earth Syst. Dyn. Discuss. 1–34.

Qian, Y., Hsu, P.-C., Yuan, J., Zhu, Z., Wang, H., Duan, M., 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), e2021GL097659.

Rácz, Z., Smith, R.K., 1999. The dynamics of heat lows. Q. J. R. Meteorol. Soc. 125 (553), 225–252.

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.

Romanello, M., McGushin, A., Di Napoli, C., Drummond, P., Hughes, N., Jamart, L., Kennard, H., Lampard, P., Rodriguez, B.S., Arnell, N., et al., 2021. The 2021 report of the lancet countdown on health and climate change: code red for a healthy future. Lancet 398 (10311), 1619–1662.

Thompson, V., Kennedy-Asser, A.T., Vosper, E., Lo, Y.E., Huntingford, C., Andrews, O., Collins, M., Hegerl, G.C., Mitchell, D., 2022. The 2021 western North America heat wave among the most extreme events ever recorded globally. Sci. Adv. 8 (18), eabm6860.

Tyrlis, E., Hoskins, B., 2008. The morphology of northern hemisphere blocking. J. Atmos. Sci. 65 (5), 1653–1665.

Wang, C., Zheng, J., Lin, W., Wang, Y., 2022. Unprecedented heatwave in western North America during late june of 2021: Roles of atmospheric circulation and global warming. Adv. Atmos. Sci. 1–15.

White, R.H., Anderson, S., Booth, J.F., Braich, G., Draeger, C., Fei, C., Harley, C.D., Henderson, S.B., Jakob, M., Lau, C.-A., et al., 2023. The unprecedented Pacific Northwest heatwave of June 2021. Nature Commun. 14 (1), 727.

Yiou, P., 2014. Anawege: A weather generator based on analogues of atmospheric circulation. Geosci. Model Dev. 7 (2), 531–543.

Yiou, P., Jézéquel, A., 2020. Simulation of extreme heat waves with empirical importance sampling. Geosci. Model Dev. 13 (2), 763–781, Publisher: Copernicus GmbH.

Zeder, J., Sippel, S., Pasche, O.C., Engelke, S., Fischer, E.M., 2023. The effect of a short observational record on the statistics of temperature extremes. Geophys. Res. Lett. 50 (16), e2023GL104090.