Simulating the West Pacific heatwave of 2021 with analog importance sampling

Flavio Pons, Pascal Yiou, Aglaé Jézéquel, Gabriele Messori

To cite this version:

Flavio Pons, Pascal Yiou, Aglaé Jézéquel, Gabriele Messori. Simulating the West Pacific heatwave of 2021 with analog importance sampling. 2023. hal-03982838v1

HAL Id: hal-03982838
https://hal.science/hal-03982838v1

Preprint submitted on 10 Feb 2023 (v1), last revised 21 Feb 2024 (v3)

Simulating the West Pacific Heatwave of 2021 with Analog Importance Samping

Flavio Pons, Pascal Yiou, Aglaé Jézéquel, Gabriele Messori

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 and lasted almost two months after the initial peak. The event caused severe impacts on human life and ecosystems, and was associated with the superposition of concurrent extreme drivers, whose effects were amplified by climate change. We evaluate whether this record-breaking heatwave could be 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 temperature sequences with realistic statistical properties 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.

Keywords— extreme events; North American heatwave; stochastic weather generator; climate change

1 Introduction

In June 2021, an intense heatwave affected the Pacific North West (PNW) of North America, particularly involving the states of Washington and Oregon in the U.S., and the Canadian province of British Columbia. This region experienced unprecedented temperature values, peaking at 49.6 °C in Lytton, British Columbia, on June 29 2021. This has been 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., 2022), 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 event in particular, 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 centered over West Canada, characterized by strong positive 500 hPa geopotential height (Z500) anomalies, shown in Fig. 1 for June 27 2021. 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 blocks over West 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 has been affected by such extreme temperature anomalies. The exceptional nature of the event has been likely caused by concurring and interacting extreme anomalies in common drivers of heatwaves over the PNW.

Although there is some agreement that the extreme temperatures developed due to subsidence inside the 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 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 heat 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 is 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 jetstream may have contributed to the intensification of a Rossby wave train crossing the Pacific and in phase-locking configuration with the PNW anticyclonic blocking. This wavetrain may have acted as an efficient guide for the teleconnection between 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 of 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, it is likely that nonlinear soil-atmosphere interactions have played an important role. Both the positive temperature anomaly and the negative soil moisture anomaly kept growing even after geopotential height peaked, suggesting that near-surface temperature may have been amplified by $\sim 40\%$ by nonlinear land-atmosphere interactions. In particular, dryness persisting since early June may have triggered a soil moisture feedback that affected the area at a monthly scale (Bartusek et al., 2022).

From the literature about the event, it emerges that many of the processes that have initiated and amplified the heat dome were driven by particular features in the atmospheric circulation, both over the region and at a larger scale. We question 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 heatwave on the PNW 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), recently back-extended to 1950. We start from hourly data of 500 hPa geopotential height (Z500), 2-meter 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 meters per year on the PNW (Fig. 10). To avoid that the results depend on long-term atmospheric trends, we subtract a 1950-2022 linear trend from geopotential height Yiou and Jézéquel (2020).

For the computation of circulation analogues, we rely on daily average 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). Anomalies are computed as the difference with respect to the 1950-2022 average for each day of the year, so that they are relative to the seasonal cycle.

For posterior analyses, 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 of T2M (TX) over the domain [44, 52N; 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 (TX1d), 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 daily average of TCWV over the region [0, 90N; 90, 270W] 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 vapor 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 midlatitudes.

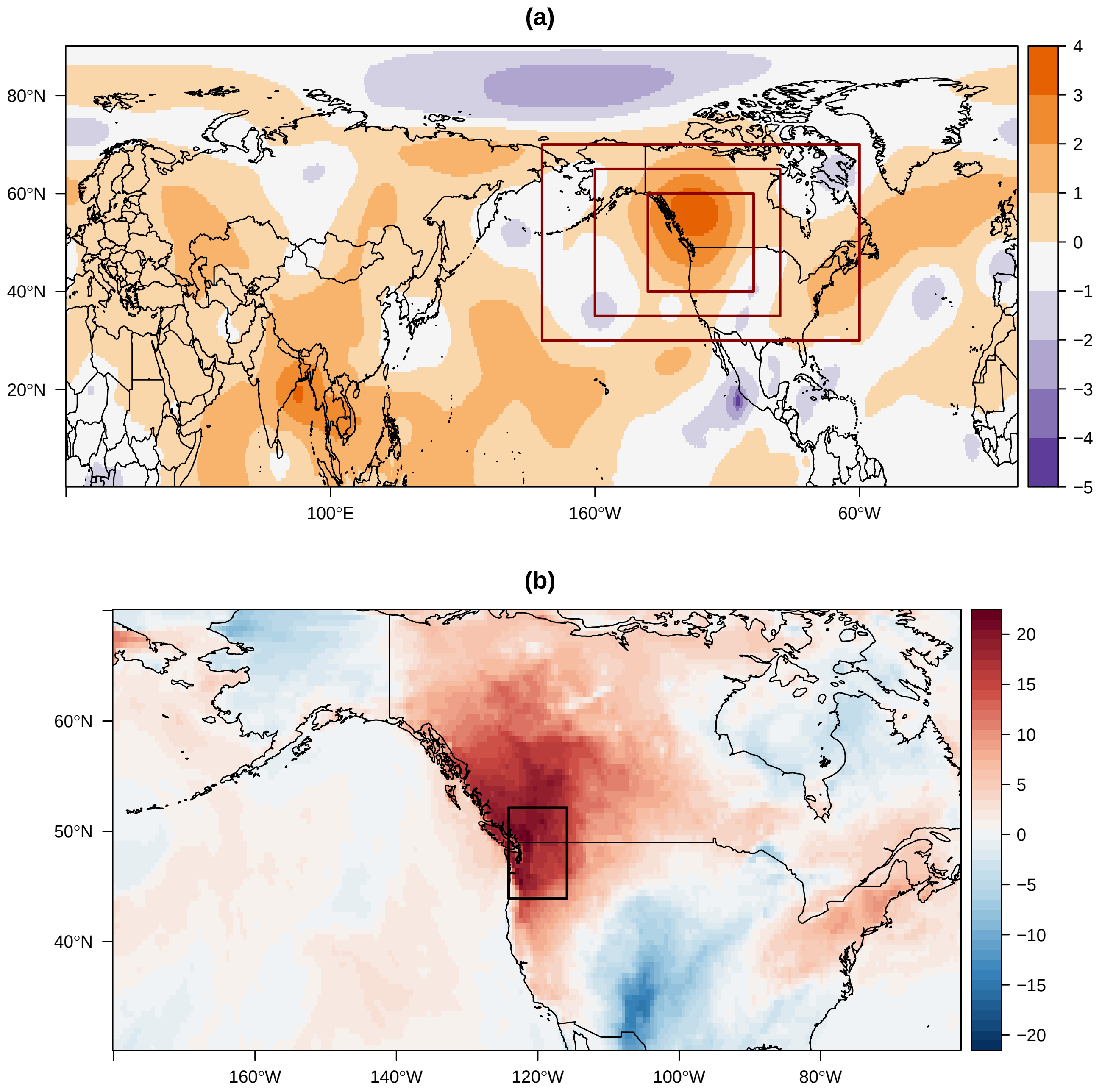


Figure 1: Upper panel: Z500 standardized anomalies with respect to the 1950-2022 climatology for June 27, 2021. The three boxes represent the three domains tested for the computation of circulation analogues. Lower panel: daily maximum 2-meter temperature anomaly (°C) with respect to the 1950-2022 climatology for June 27, 2021, at the peak of the heatwave. The inner box represents the area for the computation of the target time series.

We determine the statistical properties of the temperature variations in the PNW region in order to identify the important time scales for the event. Those properties are summarized in Figure 2. Figure 2(a) displays a comparison between TX, TX07, TX15, TX30 and the smoothed 1950-2022 seasonal cycle for the entire JJA 2021 period. During this period, TX stays consistently above the seasonal cycle for more than 60 days, with anomalies mostly larger than 5 °C. The main event, bringing the unprecedented heat peak at the end of June, lasts less than a week. However, two other significant peaks are observed at the end of July and around mid-August, with $> 10\,^{\circ}\mathrm{C}$ anomalies. Figure 2 (b) shows the 1950-2022 time series of yearly TJJA, TX30d, TX15d, TX07d and TX1d. 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 TJJA, 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 TX07d, TX15d and TX30d, excluding 2021 from the time series, in order to determine return values for those variables (Figure 2c–e). Since all the time series in panel (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 global average of the JJA 2-m temperature for year t, obtained from ERA5. Parameters are estimated using maximum likelihood, excluding 2021 from the time series. 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, $\tau(\mathrm{TX30d}) = 215$ years, while $\tau(\mathrm{TX1day})$ (not shown in figure), $\tau(\mathrm{TX07d})$ and $\tau(\mathrm{TX15d})$ are infinite. This follows from the fact that all the estimates of the shape parameter ξ are negative, which implies that the corresponding GEV distribution is upper bounded: the return time is infinite for events with a temperature value higher than the upper bound.

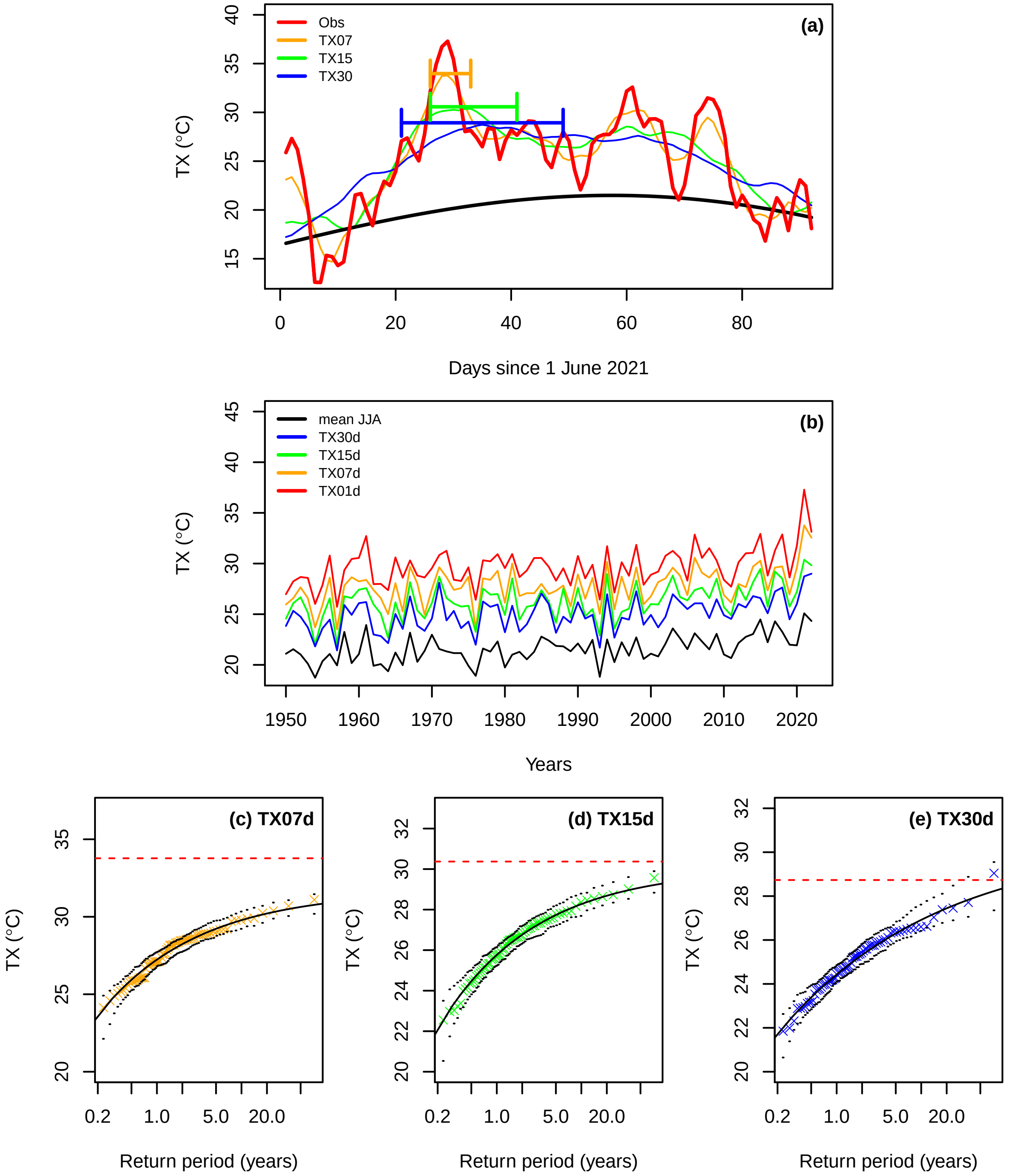


Figure 2: Panel (a): time series of daily maximum near-surface temperature for JJA 2021 (red line) compared to the seasonal cycle (black line). The orange, green and blue lines represent daily values of TX07, TX15 and TX30. The horizontal segments show the periods corresponding to TX07d, TX15d and TX30d. Panel (b): time series of TX1d (red), TX07d (orange), TX15d (green) and TX30d (blue), and TJJA (black). Panels (c)-(e): return level plots for TX07d, TX15d and TX30d. Units are in °C. Colored 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.

3 Methods

3.1 Analogues of circulation

We compute circulation analogues using average daily ERA5 Z500 fields on the region defined by largest rectangle drawn in Fig. 1 [30, 70 N; 60, 180 W].

For each day between 1 January 1950 and 3 September 2022, we compute the best 20 analogues based on the Euclidean distance between the Z500 fields, within 30 calendar days before or after the target date, excluding the year of the target date. The analogue computation is carried out 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, computed in the two periods 1950–1999 and 1971–2022. We consider the latter 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. We considered three possible domains centered over the PNW for the definition of the analogues, corresponding to the three red rectangles in Fig. 1(a). The inner domain [40, 60N; 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 blocking; the outer domain [30, 70N; 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 the SWG using analogues obtained with each domain (cf. SWG definition below). 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 event.

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 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 distribution of the number of analogues by day of the year, again divided in the two periods: there is no clear shift in the seasonal distribution of the analogues, as the difference between the two distributions is not statistically significant (χ^2-test p-value < 0.72).

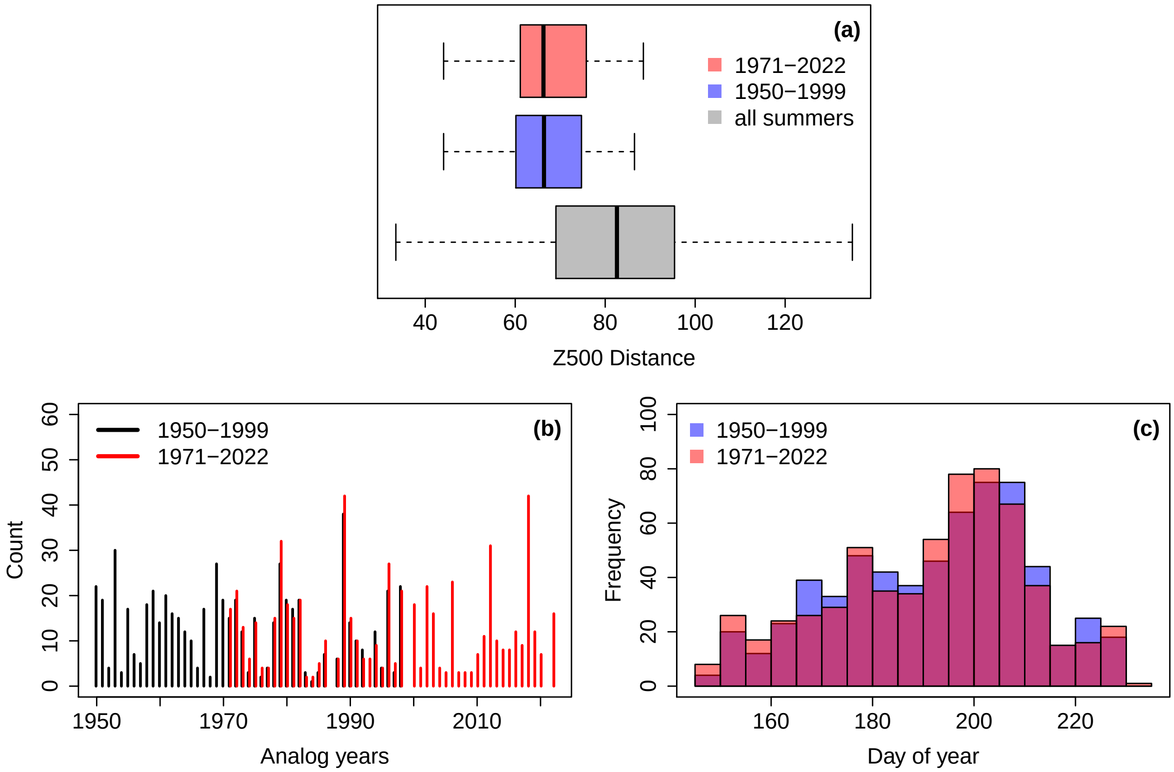


Figure 3: Panel (a): empirical probability distribution of distances of the best 20 analogs of $Z500$ between June 21 and July 21, 2021; analogues are constrained to be searched within a 30-days window around the target date. The distances are computed in the counterfactual (1950–1999, blue) and factual (1971–2022, red) period, respectively. The grey boxplot shows the distribution of analogues for the same period of all summers. Panel (b): distribution of years of analogues of $Z500$ between June 1 and August 31, 2021, for analogs chosen in the counterfactual (black) and factual (red) period. Panel (c): histograms representing the frequency of each day of the year in the chosen analogues in the counterfactual (blue) and factual (red) periods.

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 Z500_t and TX_t the Z500 field and the maximum daily temperature at time t. For each day $t_0 \leq t \leq t_0+L$, 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 $\alpha_{cal} \geq 0$ 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, in our case high temperatures at the apex of the seasonal cycle.

Once the analogue $\tilde{t}$ has been selected, the next day in the simulation is taken to be $t' = \tilde{t} + 1$. This re-sampling of the analogues is repeated for L steps, until a complete sequence of Z500 fields and corresponding values of TX has been obtained. The entire procedure is repeated S times to obtain surrogate ensembles of analogue trajectories. A schematic representation of the algorithm for the dynamic SWG is shown in Fig. 4

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 correlation between each $Z500$ field and $Z500_t$: this way, among analogues that are optimal in terms of Euclidean distance, the ones with the most similar $Z500$ pattern are favoured. Notice that this rule strongly favours the choice of date t, since by definition Z500_t has zero distance from and unit correlation with itself.

In case the event of interest is a heatwave or a cold spell, the observable of choice can be the daily average, minimum or maximum temperature, spatially averaged over a region of interest. In our case, we

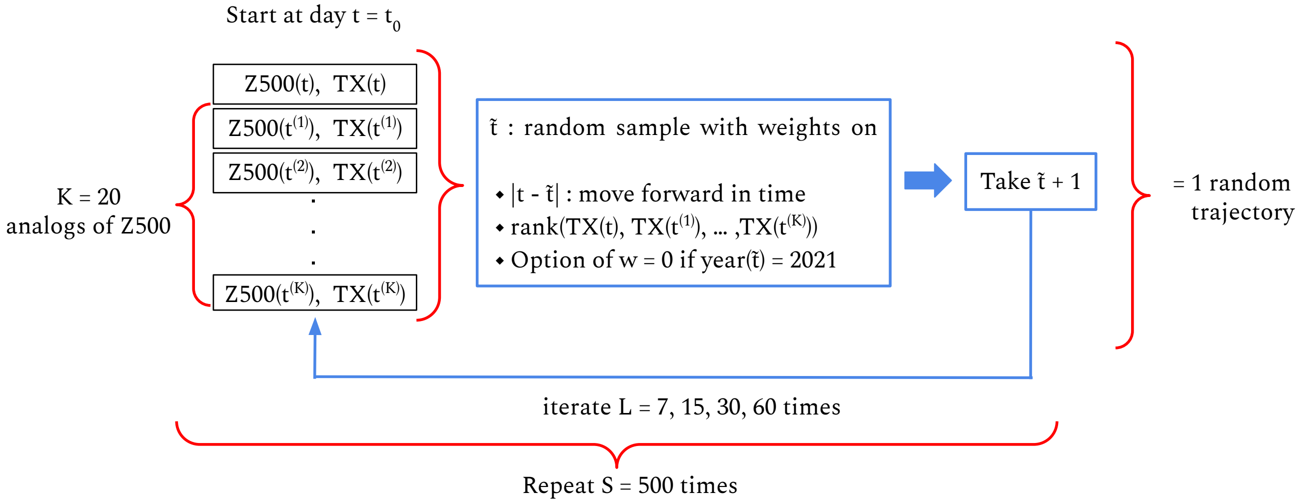


Figure 4: Schematic of the dynamic SWG.

choose maximum daily temperature averaged over [44N, 52N; 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 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). The probability distributions of TX30d from sets of 500 trajectories for the period between June 21 and July 21 2021, as a function of α_{TN}, is shown in Fig. 11 with fixed $\alpha_{cal} = 4$. We choose $\alpha_{TN} = 0.5$, since it is the value for which the dynamic simulations are closest to the observed PNA 2021 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. We fix $K = 20$, $\alpha_{cal} = 4$, $\alpha_{TN} = 0.5$.

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 record temperature values, for the three identified time scales. We also assess whether that 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 PNA heatwave using the dynamic 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. 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 exceed the values of 2021 more easily for longer trajectories. No trajectory warmer than 2021 can be simulated with analogues from the counterfactual period with lengths of 7 and 15 days, and only two warmer trajectories are obtained for the 30 day simulations. TX30 reaches 28.80 °C and 28.82 °C for these two trajectories, slightly higher than the 28.73 °C observed in 2021, but lower than the 29.0 °C of 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, considering that only 0.4% of the trajectories reach the 2021 value.

The fact that the SWG is overall incapable of reproducing the 2021 event using analogs from the counterfactual period (1950–1999) shows that this event was extremely unlikely in a climate where global warming due to anthropogenic forcing was weaker than in the current climate.

The TX07 of 2021 is never reached with SWG simulations using analogues from the factual period excluding information in 2021 (F.nE simulations). Only $\approx 4\%$ of SWG simulations reach or exceed the TX07 value of 2021 when using information on 2021 (Fig. 5a). 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, 88% of the trajectories are warmer than 2021 for F.E and 4% for F.nE.This means that the SWG approach can simulate events that are possible (because they have been observed), yet are outside of the range predicted by GEV estimates.

These results overall suggest that this heatwave has become much more likely in the recent decades, and that even in the current climate it is an exceptional event, especially at short time scales. 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 trend in atmospheric circulation, but rather to the combination of a peculiar superposition of extreme drivers and warming due anthropogenic forcing.

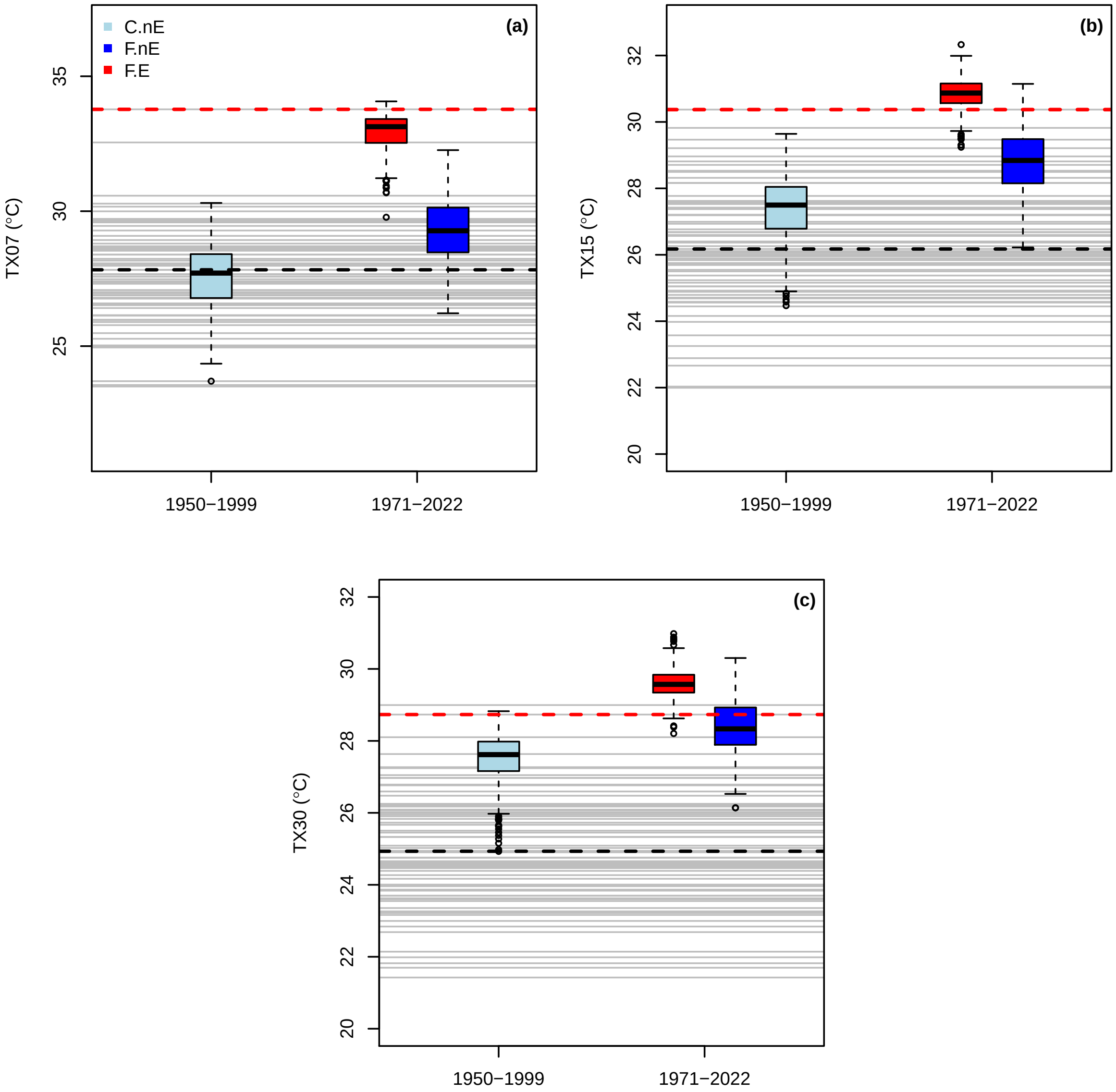


Figure 5: Empirical probability distribution of means of simulated TX computed over 7 days (a), 15 days (b) and 30 days (c). The grey horizontal lines represent the observed values between 1950 and 2022. The dashed red line is the value for 2021. Red boxes: simulations that include information from 2021. Blue boxes: simulations that exclude all information from 2021.

Figure 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 1950-2021 seasonal cycle, smoothed with cubic splines. 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.

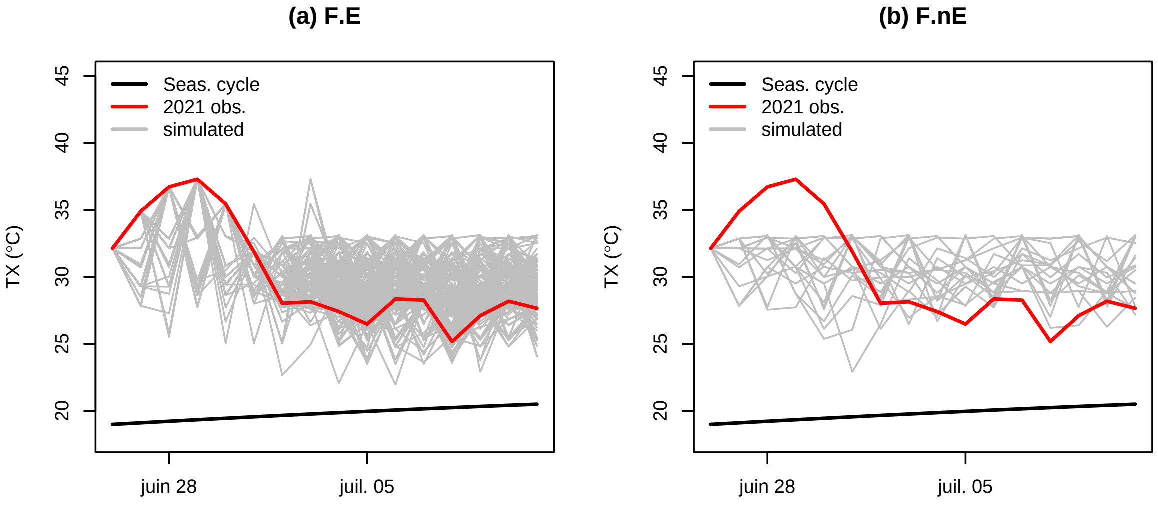


Figure 6: Trajectories of TX for 15 days simulations initialized on June 26, 2021. Simulations use analogues from the factual period including (a) and excluding (b) 2021. The red lines show 2021 observations, the black line the 1950-2022 seasonal cycle.

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 27 and July 1, corresponding to the peak of the heatwave. Figure 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 wavetrain across the Pacific, in phase-locking with the PNW anticyclone; and the negative anomaly over the Arctic. The year distribution of analogues for the peak phase of the heatwave reveals that F.E simulations sample almost all analogues from 2021, and F.nE are dominated by 2015, 1979, 2022 and 2002. The general pattern is well reproduced by simulations including 2021, with an amplified anomaly over the PNW. This is expected, since most of the analogues are sampled from 2021 among those that maximize the PNW heatwave. Factual simulations excluding 2021, on the contrary, fail to reproduce the general hemispheric pattern and the depression over the Arctic. However, they are capable to produce a positive Z500 anomaly over the PNW similar to the observations and, interestingly, a Pacific Rossby wavetrain in phase locking with the PNW anticyclone. This could then be a recurring factor in summer heatwaves over the region, contributing by directing atmospheric rivers towards the PNW (Lin et al., 2022), and reinforcing large-scale patterns (Bartusek et al., 2022) and local subsidence anomalies (Qian et al., 2022).

Figure 8 has a similar structure for the daily maximum temperature standardized anomalies, with composites shown over the PNW. From the distribution of the day-of-year of the analogues it is clear that F.E simulations use almost only analogues sampled from the peak of the event itself, which is expected, since these observations are the highest values in the TX time series, and many trajectories match such values (see panel (a) of Fig. 6). F.nE simulations sample analogues from the last decade 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, Figure 9 shows composites for TCWV. During the selected dates, the atmospheric river conveyed from the Western Pacific by the Rossby wavetrain had already made landfall, bringing a high 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 to 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 vapor towards the PNW. This is in agreement with the capability of the SWG to catch this driver of extreme heatwaves on the PNW, selecting analogues characterized by a Rossby wavetrain across the Pacific, that led to the landfall of an atmospheric river inside the anticyclonic dome.

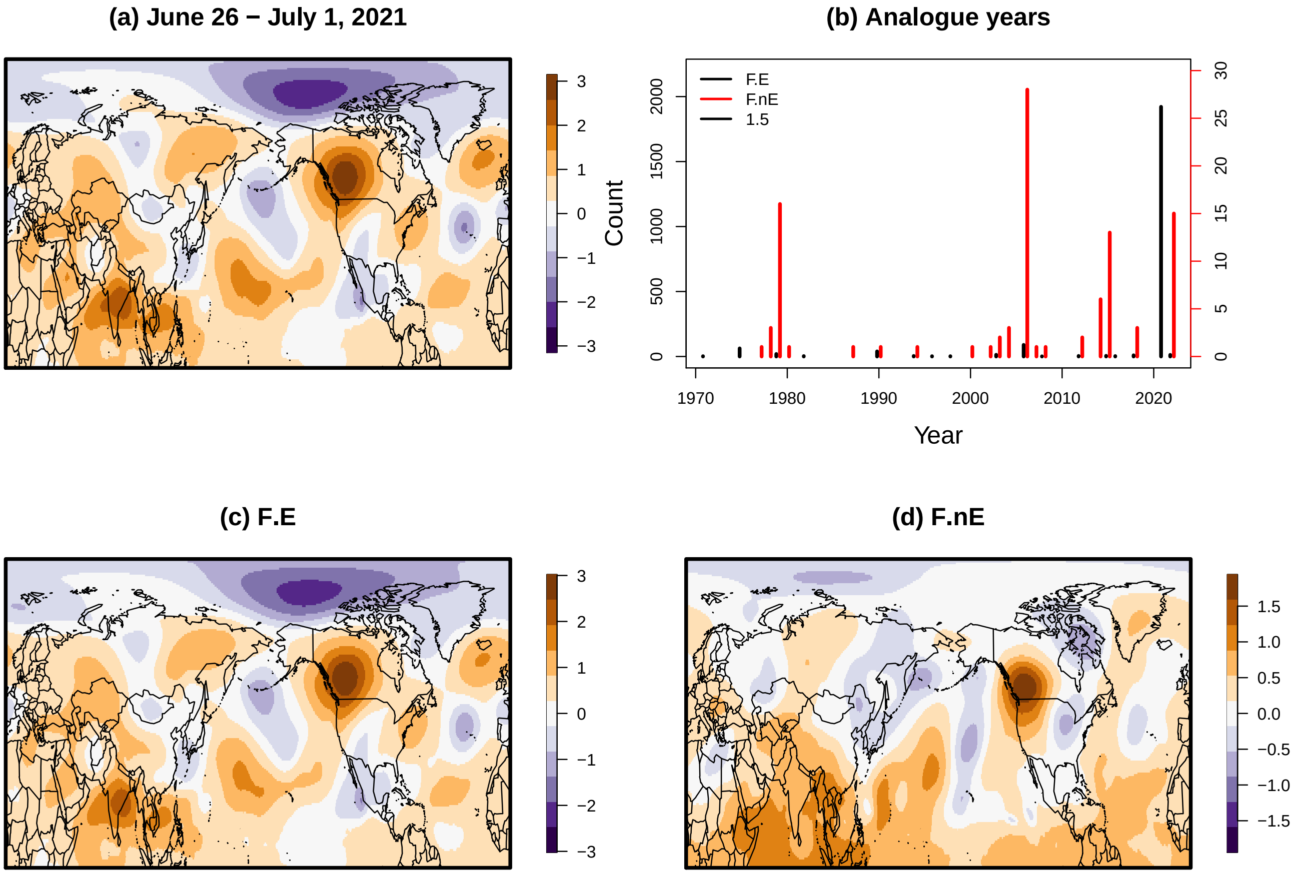


Figure 7: Summary of Z500 standardized anomalies composites for the analogues used in the simulation for the period June 27 - July 1, corresponding to the peak of the heatwave over the PNW. Panel (a): distribution of analogue years; panel (b): ERA5 data. Panels (c) and (d): composites of analogues for the simulations including and excluding 2021.

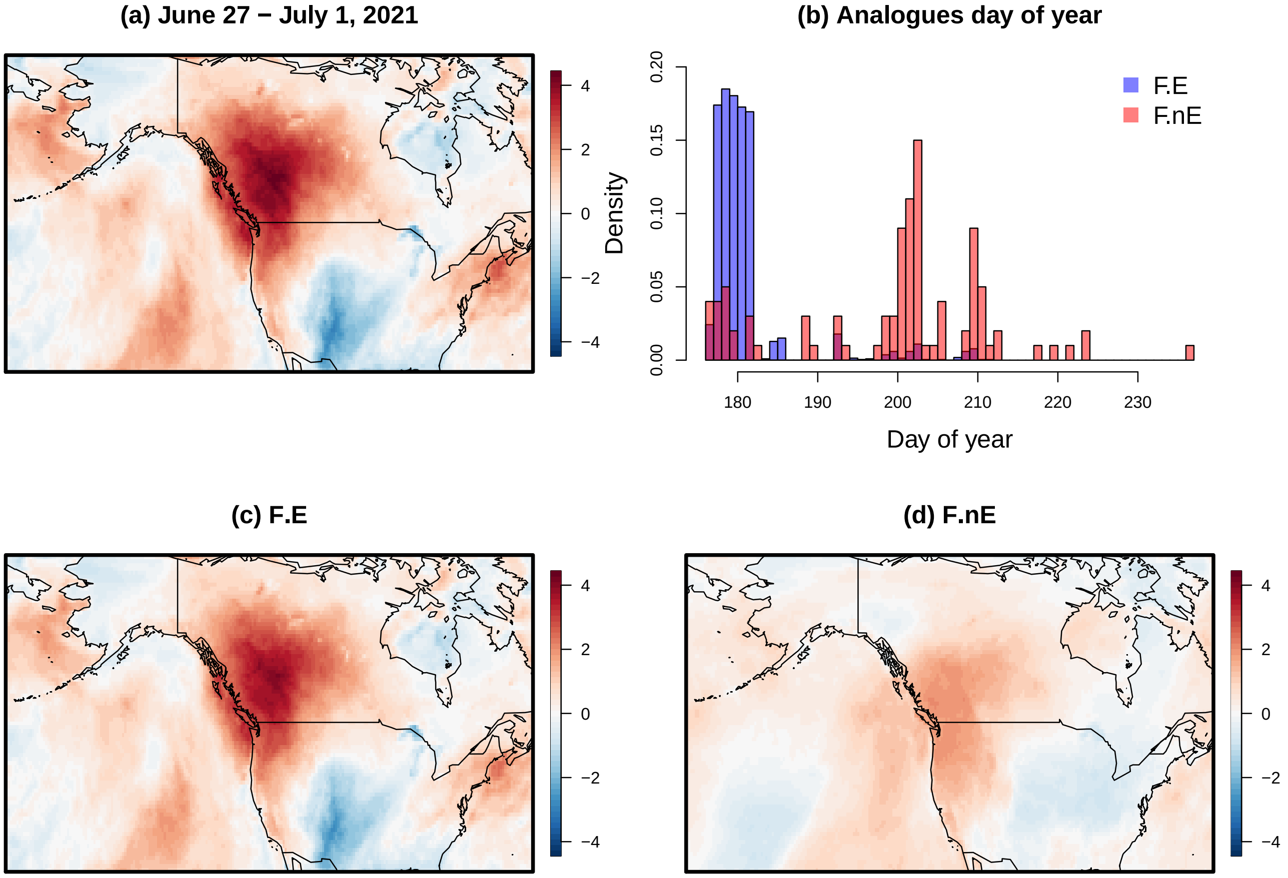
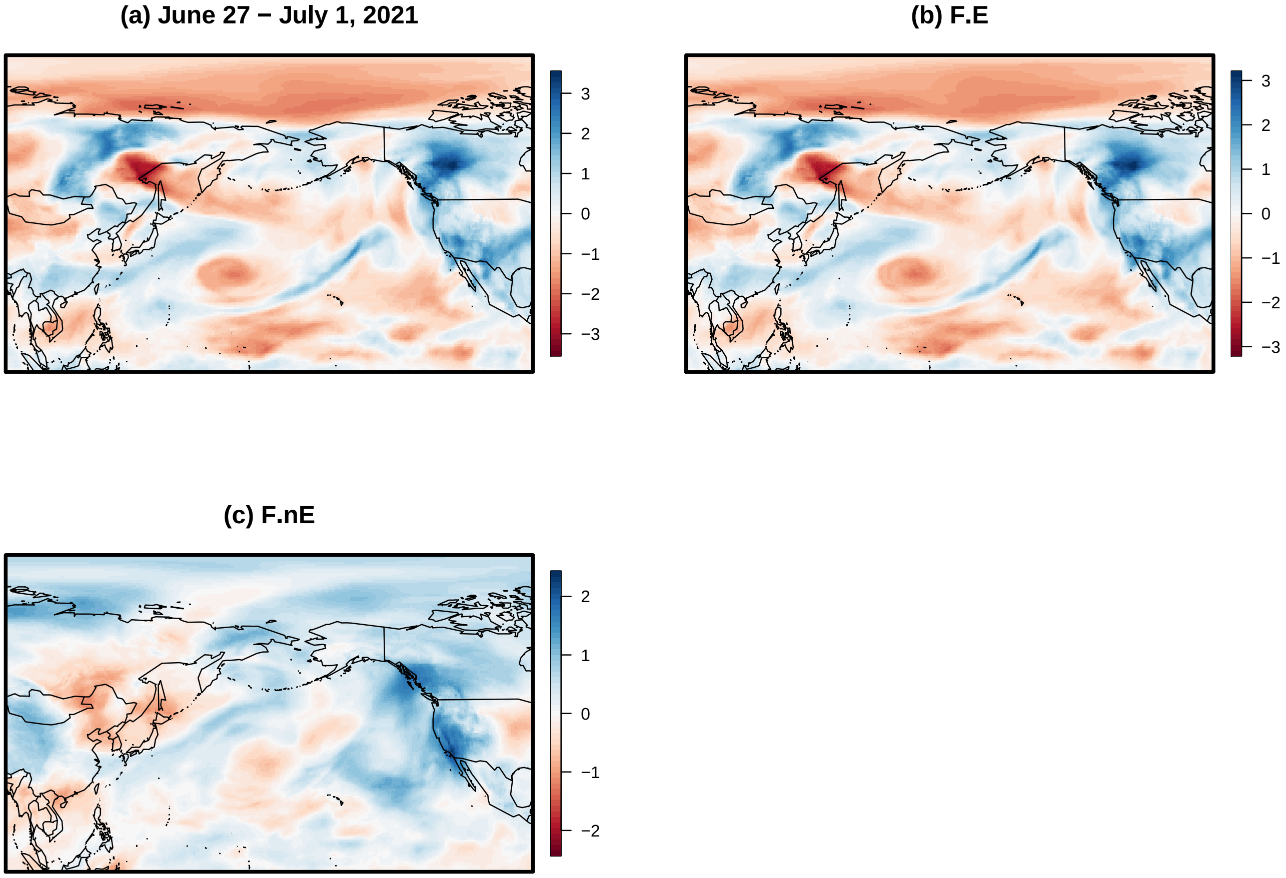


Figure 8: Summary of the maximum temperature standardized anomalies composites for the analogues used in the simulation for the period June 27 - July 1, corresponding to the peak of the heatwave over the PNW. Panel (a): distribution of analogue day-of-year; panel (b): ERA5 data. Panels (c) and (d): composites of analogues for the simulations including and excluding 2021.


(a) June 27 − July 1, 2021

(b) F.E

(c) F.nE

Figure 9: Summary of the total column water vapour standardized anomalies composites for the analogues used in the simulation for the period June 27 - July 1, corresponding to the peak of the heatwave over the PNW. Panel (a): ERA5 data. Panels (b) and (c): composites of analogues for the simulations including and excluding 2021.

5 Conclusions

The 2021 PNW heatwave was a record-shattering event on time scales ranging from one day to the entire summer, 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 daily maximum temperatures recorded during this event.

Our results confirm the role of global warming already found in other studies, as it is impossible for the SWG to produce heatwaves reaching 2021 values using analogues from the counterfactual climate. At a time scale of 7 days, it is also impossible to produce heatwaves as warm as or warmer than 2021 without including the event itself, and it is very difficult even including 2021 analogues. This shows that, at such a short time scale, the event is very rare and unlikely to be matched in the current climate. On the other hand, at a 30-day time scale, 2021 is surpassed by many trajectories even excluding analogues from the event, since an even stronger 30-day heatwave affected the region in 2022. In the intermediate 15-day case, trajectories including 2021 analogues lead to a much higher probability of reaching or surpassing the event, but 4% of the trajectories are higher even with analogues taken from other years, showing that events of comparable or larger magnitude at this time scale are rare but not impossible in the current climate. 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 other locations 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, this particular event was also the result of the combination and co-occurrence of extreme drivers and nonlinear land-atmosphere interactions. These drivers include an Omega blocking anticyclone on the PNW embedded in a hemispheric wavenumber 4 configuration, a split of the polar vortex triggering a sequence of quick changes of weather patterns over North America, and a Rossby wavetrain in phase locking with the omega structure, driving an atmospheric river towards the region.

Analogue composites for the peak of the heatwaves in the case excluding 2021 show some similarities to the event itself: the Omega blocking over the PNW, the negative geopotential anomaly over the Azores, and the Rossby wavetrain conveying large amounts of water vapor 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, especially over Eurasia, and the deep Arctic negative geopotential anomaly is not present.

The SWG allowed us to simulate an extremely rare event with very small computation time, and using a short time series despite its 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. time autocorrelation. Moreover, we only assess drivers of the event in a posterior analysis, while it may be interesting to incorporate some of the involved large-scale physics in the simulation. This could be achieved, by adding a further weight measuring the similarity of some large-scale circulation features, for example leveraging on machine learning techniques capable of decomposing atmospheric fields and projecting them on low-dimension time series (Fery et al., 2022).

Our results could be extended by using climate simulations from CMIP6. While reanalysis allowed us to analyze this event 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 realistic global warming scenarios. Moreover, the historical period of CMIP6 models trace back to 1850, constituting a counterfactual period closer to actual pre-industrial conditions.

The PNW 2021 heatwave broke historical records at all sub-seasonal time scales, producing temperatures that would have been previously considered unattainable over the PNW. Using simulations from a SWG, we found that global warming has amplified its magnitude at all averaging times 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, and their co-occurrence and interaction alone cannot explain how extreme this event was. This heatwave remains a worst-case scenario for extreme heat periods below 15 days, while the present climate could foster more extreme events at the monthly and seasonal scales.

6 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) and ERC grant agreement No. 948309 (CENÆ: GM), and the grant ANR-20-CE01-0008-01 (SAMPRACE: PY).

References

Bartusek, S., Kornhuber, K., and Ting, M. (2022). 2021 north american heatwave amplified by climate change-driven nonlinear interactions. *Nature Climate Change*, 12(12):1143–1150.

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

Bumbaco, K. A., Dello, K. D., and Bond, N. A. (2013). History of pacific northwest heat waves: Synoptic pattern and trends. *Journal of applied meteorology and climatology*, 52(7):1618–1631.

Christidis, N. and Stott, P. A. (2015). Changes in the geopotential height at 500 hpa under the influence of external climatic forcings. *Geophysical Research Letters*, 42(24):10–798.

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

Dacre, H. F., Clark, P. A., Martinez-Alvarado, O., Stringer, M. A., and Lavers, D. A. (2015). How do atmospheric rivers form? *Bulletin of the American Meteorological Society*, 96(8):1243–1255.

Fery, L., Dubrulle, B., Podvin, B., Pons, F., and Faranda, D. (2022). Learning a weather dictionary of atmospheric patterns using latent dirichlet allocation. *Geophysical Research Letters*, 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., and Thépaut, J.-N. (2018). Era5 hourly data on single levels from 1959 to present.

Jézéquel, A., Yiou, P., and Radanovics, S. (2018). Role of circulation in european heatwaves using flow analogues. *Climate dynamics*, 50(3):1145–1159.

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

Lin, H., Mo, R., and Vitart, F. (2022). The 2021 western north american heatwave and its subseasonal predictions. *Geophysical Research Letters*, 49(6):e2021GL097036.

Lucarini, V., Galfi, V. M., Riboldi, J., and Messori, G. (2023). Typicality of the 2021 western north america summer heatwave. *Environmental Research Letters*, 18(1):015004.

Mo, R., Lin, H., and Vitart, F. (2022). An anomalous warm-season trans-pacific atmospheric river linked to the 2021 western north america heatwave. *Communications Earth & Environment*, 3(1):1–12.

Neal, E., Huang, C. S., and 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. *Geophysical Research Letters*, 49(8):e2021GL097699.

NOAA (last access: 02/09/2022). https://www.cpc.ncep.noaa.gov/data/teledoc/np.html.

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., and Otto, F. E. L. (2021). Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the us and canada june 2021. *Earth System Dynamics Discussions*, pages 1–34.

Qian, Y., Hsu, P.-C., Yuan, J., Zhu, Z., Wang, H., and Duan, M. (2022). Effects of subseasonal variation in the east asian monsoon system on the summertime heat wave in western north america in 2021. *Geophysical Research Letters*, 49(8):e2021GL097659.

Rácz, Z. and Smith, R. K. (1999). The dynamics of heat lows. *Quarterly Journal of the Royal Meteorological Society*, 125(553):225–252.

Ragone, F., Wouters, J., and Bouchet, F. (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.

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

Wang, C., Zheng, J., Lin, W., and Wang, Y. (2022). Unprecedented heatwave in western north america during late june of 2021: Roles of atmospheric circulation and global warming. *Advances in Atmospheric Sciences*, pages 1–15.

White, R., Anderson, S., Booth, J., Braich, G., Draeger, C., Fei, C., Harley, C. D. G., Henderson, S. B., Jakob, M., Lau, C.-A., et al. (2022). The unprecedented pacific northwest heatwave of june 2021.

Yiou, P. (2014). Anawege: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543.

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

7 Supplementary Material

7.1 Z500 trends

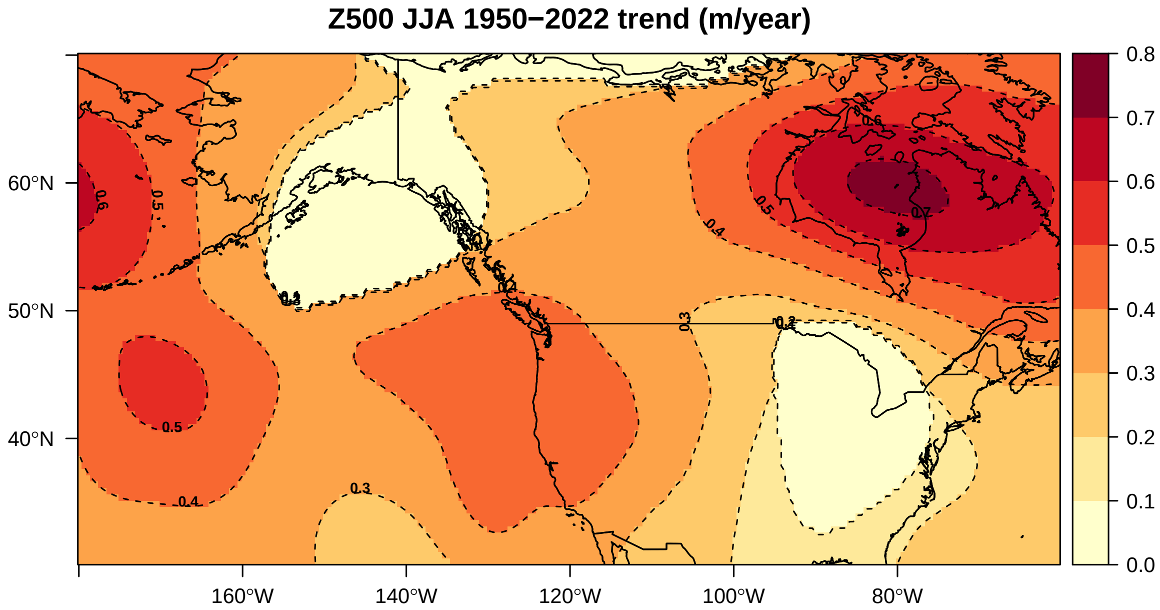


Figure 10: Observed 1950-2022 trend in mean JJA geopotential height in ERA5. Units are meters per year

7.2 Choice of importance sampling parameter

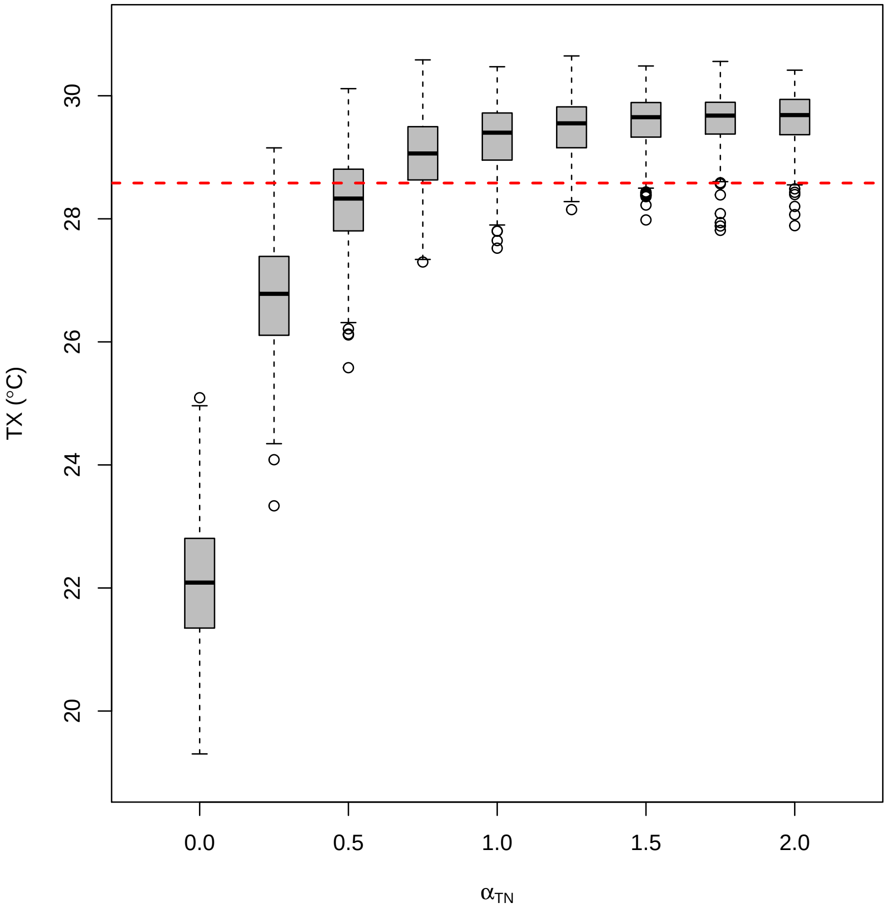


Figure 11: Boxplots of simulated temperatures for the hottest 30 days period, between June 21 and July 21, as a function of the parameter α_{TN}, with $\alpha_{cal} = 4$. The red dashed line is the average TX between June 21 and July 21, 2021.

7.3 Analogue statistics

The number of single analogues used in the simulation is 300 to simulate the 6570 F.E days and 80 to simulate the 300 F.nE days. The number of unique analogues used to produce the trajectories is 4.6% of the total for F.E (300 unique days out of 6570) and 26.7% of the total for F.nE (80 unique days out of 300). This indicates a reduced variability in the analogue sampling for the F.E case, where each analogue is repeated on average 22 times, versus the 3.75 average repetitions in the F.nE case. Figure 12 shows the temporal distribution of the analogues used to obtain the simulations warmer that 2021, taking into account the number of time each one has been chosen. First, we show the distribution of the sampled analogues over the years. For the F.E configuration, almost half of analogues is sampled from 2021, while for F.nE almost all analogues are accounted for considering an ensemble of 9 different years. The 10 years providing most analogues for both sets of simulations are summarized in Table 1. Concerning the day-of-year distribution, for F.nE simulations the peak of the distribution is in the last decade of July, at the peak of the seasonal cycle, while 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.

Including 2021										
year	2021	2002	2018	1975	2022	2006	2015	2012	1998	2003
# analogues	3035	449	401	355	328	265	254	203	196	101
Excluding 2021										
year	1979	2022	2006	2004	2015	1978	2014	1994	2007	1990
# analogues	52	46	43	26	25	23	18	11	10	5

Table 1: Number of analogues, counted with repetition, from the 10 years providing most analogues for simulations including and excluding 2021.

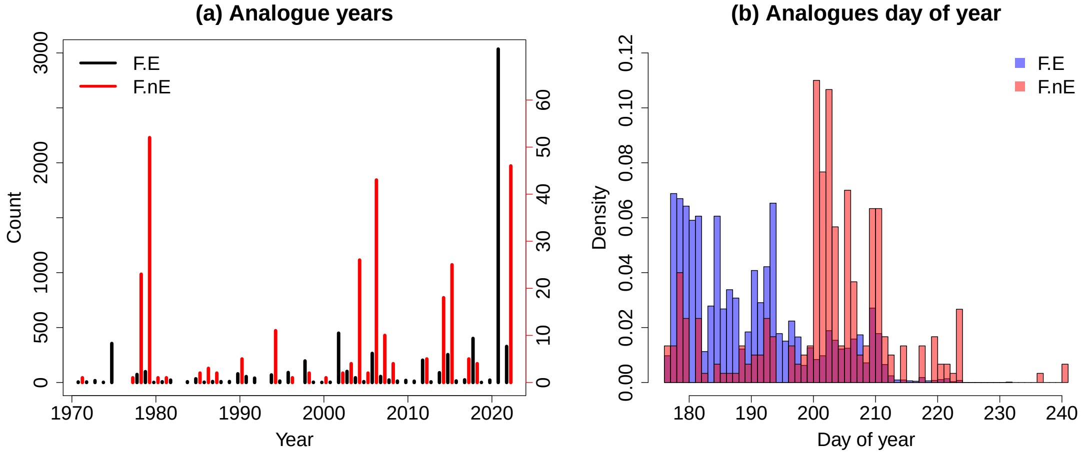


Figure 12: Time distribution of the analogues used in the 15 days simulations with TX15 warmer than 2021. Panel (a): distribution of analogues in the years; black lines are simulations including the event, red lines are simulations excluding 2021. Panel (b): distribution of analogues per day-of-year. Blue bars are simulations including the event, red bars are simulations excluding 2021.