

# Robust intra-model teleconnection patterns for extreme heatwaves

**George Miloshevich**<sup>1,2</sup>, **Philippe Rouby-Poizat**<sup>2</sup>, **Francesco Ragone**<sup>3,4</sup> and **Freddy Bouchet**<sup>2</sup>

<sup>1</sup>*Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay & IPSL, CE Saclay l'Orme des Merisiers, 91191 Gif-sur-Yvette, France*

<sup>2</sup>*ENSL, CNRS, Laboratoire de Physique, F-69342 Lyon, France*

<sup>3</sup>*Royal Meteorological Institute of Belgium, Brussels, Belgium*

<sup>4</sup>*Université catholique de Louvain, Louvain-la-Neuve, Belgium*

Correspondence\*:

Corresponding Author  
gmiloshe@lsce.ipsl.fr

## ABSTRACT

We investigate the statistics and dynamics of extreme heat waves over different areas of Europe. We find heatwaves over France and Scandinavia to be associated with recurrent wavenumber three teleconnection patterns in surface temperature and mid-tropospheric geopotential height. For heatwaves with return times of 4 years these teleconnection patterns and their dynamics are robustly represented in a hierarchy of models of different complexity and in reanalysis data. For longer return times, reanalysis records are too short to give statistically significant results, while models confirm the relevance of these large scale patterns for the most extreme heatwaves. A timeseries analysis shows that heatwave indices defined at synoptic scale are fairly well described by Gaussian stochastic processes, and that these Gaussian processes reproduce well return time plots even for very rare events. These results suggest that extreme heatwaves over different areas of Europe show recurrent typical behaviours in terms of long-range spatial correlations and subseasonal-scale temporal correlations. These properties are consistently represented among models of different complexity and observations, thus suggesting their relevance for a better understanding of the drivers and causes of the occurrence of extreme midlatitude heatwaves and their predictability.

**Keywords:** heatwave, extreme, climate, reanalysis, teleconnection, return time, Ornstein-Uhlenbeck, autocovariance

## 1 INTRODUCTION

Some of the most severe impacts of climate change are caused by rare and extreme events. Increased frequency and magnitude of extreme heatwaves is one of the most immediate and significant effects of global warming Van Oldenborgh et al. (2022). In the last decades a number of record breaking heatwaves have been observed (IPCC (2014); Seneviratne et al. (2021)). In the Northern Hemisphere mid-latitudes examples include the Western European heatwave of summer 2003, with a death toll of about 70,000 (García-Herrera et al. (2010)), the mega-heatwave over Russia of summer 2010 (Otto et al. (2012)) with the death toll of about 100,000, and more recently the heatwave 2021 Western North America heatwave (Philip et al. (2021)).

Recent studies have identified increased frequency of heatwaves occurring in mid-latitudes in Europe of duration longer than 6 consecutive days (Rousi et al., 2022). In fact, the reasons for the high impact of the 2003 Western European heatwave was not only the magnitude of the temperature anomalies, but also its long duration (two successive heat events along an overall period of one month). In the case of the 2010 Russian heatwave, there was the compounding effect of high temperature, long duration (one month), and related wildfires. 2021 Western North America heatwave lasted about 10 days with peculiarly strong peak on a 2-day scale. In this case the main cause of the impacts were short term extreme temperature fluctuations, that were unprecedented in historical record World Weather Attribution (2021); Schiermeier (2021).

Heatwaves typically occur when a stationary high pressure anomaly over a region leads to subsidence and increased incoming shortwave radiation fluxes, that increase surface and near surface temperatures (Horton et al. (2016); Perkins (2015)). Short-term events lasting a few days are typically understood as caused by the occurrence of an atmospheric blocking over the region impacted by the heatwave. The most extreme and persistent events, on the other hand, require the activation of more particular dynamics and of feedback processes that enhance the surface temperature response and the duration of the event. This involves processes acting on space and time scales that go beyond the regional spatial scales and the synoptic time scales characteristic of standard heatwaves. A comprehensive analysis of the properties of extreme midlatitude heatwaves however is hindered by their rarity, the inherent noisiness of midlatitude weather fluctuations, and thus the difficulty to sample a sufficient number of events to provide a reliable statistics.

From a spatial point of view, it has been observed in the literature in the past years that some of the most extreme and persistent heatwaves seem to be associated with large scale atmospheric teleconnection patterns over the entire Northern hemisphere, for example extreme heatwaves over the central USA (Teng et al. (2013)), Alberta (Petoukhov et al. (2018)), and Western Europe (Kornhuber et al. (2019)). Following a recent classification of compound extreme events (Zscheischler et al. (2020)), this would correspond to spatially compounding events. Some authors have interpreted the presence of these patterns in terms of amplification of quasi-stationary Rossby waves, which are claimed to be related to the genesis of extreme events in different regions of the world during the same season (Lau and Kim (2012); Petoukhov et al. (2013, 2016); Schubert et al. (2011)).

The research in this area has been directed towards comprehending the pivotal role of waves that exhibit a wavenumber ranging from 5 to 8. To this end, an array of diverse detection techniques has been utilized encompassing empirical orthogonal functions (EOFs) Teng et al. (2013), spectral analysis (Kornhuber et al. (2019); Petoukhov et al. (2018)), and indicators that rely on resonance models (Petoukhov et al. (2013, 2016)). Despite these strides, the precise identification of these teleconnection patterns is hindered by the limited amount of events available. Furthermore, we are increasingly facing the questions ranging from estimating risks for similar “black swan” events found in the tails of the distribution: from estimating their return times to attributing them to climate change Philip et al. (2020). This highlights the need to devise more efficient tools for extreme return time estimation.

To overcome the sampling issue, Ragone et al. (2018); Ragone and Bouchet (2020, 2021) have adopted rare event algorithm applied to simulations with climate models of different complexity. In this way they found that the warmest summers and most persistent heatwaves are robustly associated with teleconnection patterns with wavenumber 3-4, and were able to show that these patterns are indeed statistically significant (see in particular Ragone and Bouchet (2021)). However, a detailed analysis of the dynamical properties of

these peculiar atmospheric states, and the possible relation with the amplification of higher wavenumber Rossby waves suggested in the literature, have not been investigated yet.

From a temporal point of view, there are two factors that characterise the most extreme and persistent events, up to the seasonal scale. First, these events typically occur when a train of heatwaves leads to an extreme hot summer, as in western Europe in 2003 (Cassou et al. (2005)). This would correspond to temporal compounding mechanisms in Zscheischler et al. (2020). Possible reasons why a given season could be more likely to develop multiple heatwaves are not clear, but they would be of great interest for what concerns the seasonal to decadal predictability of temperature extremes. Second, it is known that pre-existing anomalously dry soil conditions can amplify the intensity of heatwaves occurring at a later time, a typical mechanism for example of certain classes of heatwaves in Europe (Stéfanon et al. (2012)). This would correspond to preconditioning mechanisms in Zscheischler et al. (2020). Preconditioning is typically studied with ad hoc numerical experiments initialized with dry or moist soil initial conditions, or in terms of correlations. A more detailed analysis of the properties of the decorrelation function of regional surface temperatures over different time scales seem to be absent from the literature.

The goal of this paper is to better understand the properties of long range correlations in space and time characterizing extreme heatwaves, and to compare their robustness among models of different complexity and reanalysis data. We adopt a very simple definition of a heatwave event that specifically aims at the properties of time persistence of the temperature anomalies from the subseasonal to seasonal scale (S2S). We analyse events over France and Scandinavia. In Section 2 we describe the data, the heatwave index, and the statistical tests and analysis adopted in this work. In Section 3 we describe the teleconnection patterns we find associated with the extreme events, we show the properties of the atmospheric activity associated with these events in terms of Hayashi spectra, and we discuss the statistical significance of the patterns. In Section 4 we analyse the temporal evolution of the events and the properties of the time autocovariance function of the regional temperatures, comparing them against a first order autoregressive process. Finally in Section 5 we discuss our conclusions.

## 2 METHODOLOGY

### 2.1 Data

We study extremely rare long lasting heatwaves over different areas of Europe during the June-July-August period (JJA). The data consists of daily averages of 2 meters temperature and 500 hPa geopotential height, from three different datasets: one intermediate complexity climate model (Plasim), one CMIP5-type Earth system model (CESM 1.2), and the ERA5 reanalysis.

#### 2.1.1 PlaSim

The Planet Simulator Fraedrich et al. (2005, 1998), also known as PlaSim, is a climate model of intermediate complexity that has been developed to simulate the Earth's climate. The model's dynamical core uses a spectral transform method to solve the primitive equations governing vorticity, divergence, temperature, and surface pressure. The model has a horizontal resolution of T42 in spectral space, which translates to a spatial resolution of 2.8 degrees by 2.8 degrees or to  $64 \times 128$  grid-points. Additionally, the model includes 10 vertical layers and incorporates simple parameterizations of the most important physical processes that influence the climate system, such as large-scale precipitation, clouds, moist and dry convection, boundary layer fluxes of latent and sensible heat, and vertical and horizontal diffusion and radiation. The atmospheric model is coupled to a simple bucket land surface scheme. To produce a

stationary state climate close to that of the 1990s, the model is driven by prescribed values of sea surface temperatures, sea ice cover, greenhouse gas concentration, and incoming solar radiation fluxes at the top of the atmosphere. We used the model to generate 1000 years of data in this stationary state, this dataset has also been utilized in previous studies conducted by Ragone et al. (2018) and Jacques-Dumas et al. (2022).

### 2.1.2 CESM

We use control runs from Ragone and Bouchet (2021) which were performed with version 1.2.2 of the Community Earth System Model (CESM) developed by Hurrell et al. (01 Sep. 2013) utilizing an atmosphere and land only setup, with the active components being version 4 of the Community Atmospheric Model (CAM4) and version 2 of the Community Land Model (CLM2). The model was run in a statistically stationary state, with the concentration of greenhouse gases, sea surface temperature (SST), and sea ice cover held at values corresponding to the climate of year 2000. The model features a horizontal resolution of 0.9 and 1.25 degrees in latitude and longitude, respectively, and includes 26 vertical layers in hybrid pressure coordinates.

### 2.1.3 ERA5 reanalysis

We consider data from the ERA5 reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) Hersbach et al. (2020); Muñoz Sabater et al. (2021), from 1950 to 2020. The data was downloaded from <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5> and re-gridded to a lower resolution of 241 by 480 while performing daily averaging over available 3 hour periods.

## 2.2 Definition of heatwaves and return times

To define heatwaves, several indices have been used in the literature, for different purposes (Perkins (2015)). Many meteorological criteria used in climate studies of temperature extremes focus on sub-daily fluctuations (see for instance IPCC (2014)). However, long-lasting heatwaves are the most detrimental to health and biodiversity (Barriopedro et al. (2011)). Moreover, many of the extreme heatwaves with the largest impact, for instance the Western European one in 2003 or the Russian one in 2010 lasted long, from two to five weeks. These long lasting heatwaves were often composed of several sub-events compatible with the classical, short-time based definitions (Perkins (2015)). The lack of comprehensive studies of the statistics of long-lasting events has actually been stressed in the last IPCC report Seneviratne et al. (2021). Moreover, many definitions that actually involve a measure related to the persistence of anomalous daily maximum temperature values with prescribed amplitude do not always carry a natural definition of a heatwave amplitude (Perkins (2015)). This prevents to study independently impact of amplitude and duration of the heatwave and calls for a complementary definition of heatwaves, that can quantify both their amplitude in terms of temperature and their duration, in an independent way.

In this study we choose a specific criterion for selecting heatwaves (following the definitions of Ragone et al. (2018); Gálfy et al. (2019); Ragone and Bouchet (2020); Gálfy and Lucarini (2021); Ragone and Bouchet (2021); Gálfy et al. (2021)), which consists of extremes of time averaged surface temperature fluctuations (anomalies) defined as

$$A(t) := \frac{1}{T} \int_t^{t+T} \frac{1}{|\mathcal{D}|} \int_D (T_s - \mathbb{E}(T_s)) (\vec{r}, u) d\vec{r} du \quad (1)$$

where  $\mathbb{E}(T_s)(\vec{r}, t)$  is the mean surface temperature at each point of the grid, with a seasonal and spatial variation. In the following the symbol  $\mathbb{E}$  refers to the statistical empirical average over all available years, which corresponds to the duration of the datasets (1000 years for PlaSim and CESM and 71 years for ERA5). In this study the heatwave area  $\mathcal{D}$  correspond either to France or Scandinavia and is depicted via shaded boxes on Figure 9. France is defined as a land enclosed between 43 N - 51 N and 4 W - 6 E, while Scandinavia is defined as a land enclosed between 5 E - 39 E and 57 N - 71 N. The heatwave duration is chosen depending on the impact of interest. In this study we select  $T = \{5, 14, 30, 90\}$  days. In order to remove the climate change signal from ERA5 reanalysis the fields were linearly detrended for each grid point and anomalies were defined with respect to such detrended mean.

The choice to study the extremes of temperature anomalies rather absolute temperature is made for the following reason. From a dynamical perspective, the summer months are similar, and provide more statistics than individual months. This definition emphasizes dynamical characteristics, which are thought to be described reasonably well on a seasonal time scale, rather than physical impacts which are often related to the physical temperature. Moreover, the emphasis on anomalies allows to collect richer statistics and to concentrate on properties that are better captured by the models, as opposed to the absolute magnitudes.

We are interested in summer statistics June July August (JJA) so that duration of heatwave is constrained to JJA and cannot occur outside this period. The yearly block-maxima extremes  $a_i$  occurring at calendar day  $t_i$  are defined

$$a_i := A(t_i) = \max_{t \in \text{summer}(i)} \{A(t)\}, \quad (2)$$

Since temperatures are correlated on a daily timescales one has to be careful about the selection of the maxima which occur on the first or the last summer day. Thus, for such events we check whether they correspond to local maxima. If they do not, we look for greatest local maximum in the summer time series for that year, thus avoiding registering extrema of May or September.<sup>1</sup>

To compute the return times, the yearly summer extremes  $a_i$  are ranked in decreasing order, as usual (see Section 4.2.1). Next, we compute return times for heatwaves using *modified block maximum estimator*, following the method described in Lestang et al. (2018), assuming i.i.d. Poisson statistics  $P(t) = \lambda \exp(-\lambda t)$ , which reads in our notation as

$$r = \frac{1}{\log(1 - \text{rank}(a)/M)}, \quad (3)$$

where  $M$  corresponds to the total number of years and  $\text{rank}(a)$  corresponds to the index  $i$  of the extreme found in the sequence  $\{a_i\}$ . In a regime of large extremes  $a \rightarrow \infty$  this expression asymptotes to the more familiar  $r \rightarrow M/\text{rank}(a)$ .

## 2.3 Statistical analysis

To identify long lasting heatwaves we will consider time averages of temperature and geopotential height on different time scales. In particular, given a variable  $X(\mathbf{x}, t)$  we consider averages conditioned on the extreme, also called composite maps. The composite map of  $X$  at the lead time  $\tau$  of an extreme heatwave

<sup>1</sup> This is because with the original definition any extreme which actually occurs in May or September can superficially inflate the summer extremes for the first day or the last day of the summer because of temporal correlations.

with a return time  $r$  is

$$X^r(\mathbf{x}, \tau) = \frac{1}{N_r} \sum_{\{i | a_i \geq a^r\}} X(\mathbf{x}, t_i - \tau). \quad (4)$$

where  $N_r$  is a number of events for a given threshold  $a_r$  and  $t_i$  is the time of the onset of the heatwave. This definition gives informations on how a typical heatwave above a given threshold develops and disappears.

We test statistical significance using the standard Student  $t$  test. For instance, let us discuss the estimation of the average of a random variable  $Y$ , using  $N$  independent samples  $\{Y_n\}$  through the empirical average

$$\mathbb{E}_e(Y) = \frac{1}{N} \sum_{n=1}^N Y_n \quad (5)$$

If our aim is to test whether this sample average is significant, in the sense that the probability that the actual average is 0 is very low, we will compute the t-value

$$t = \frac{\sqrt{N} \mathbb{E}_e(Y)}{\sigma_e(Y)}, \quad (6)$$

where  $\sigma_e(Y)$  is the empirical standard deviation

$$\sigma_e(Y) = \sqrt{\frac{1}{N-1} \sum_{n=1}^N [Y_n - \mathbb{E}_e(Y)]^2}. \quad (7)$$

Then we will say that the average is statistically significant with  $|t| \geq 1$  or with  $|t| \geq 2$ , or any desired value.

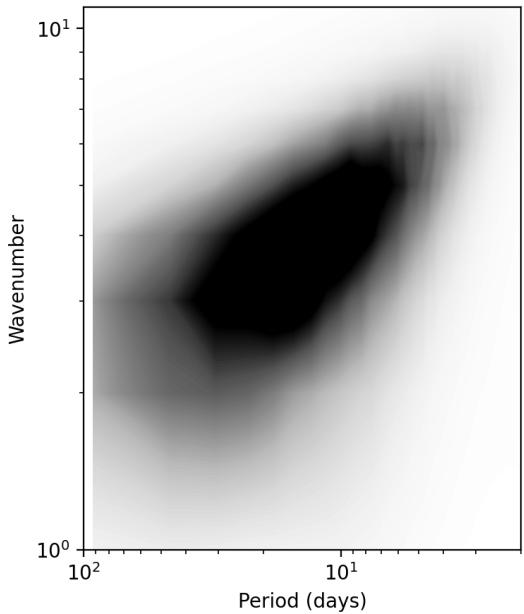
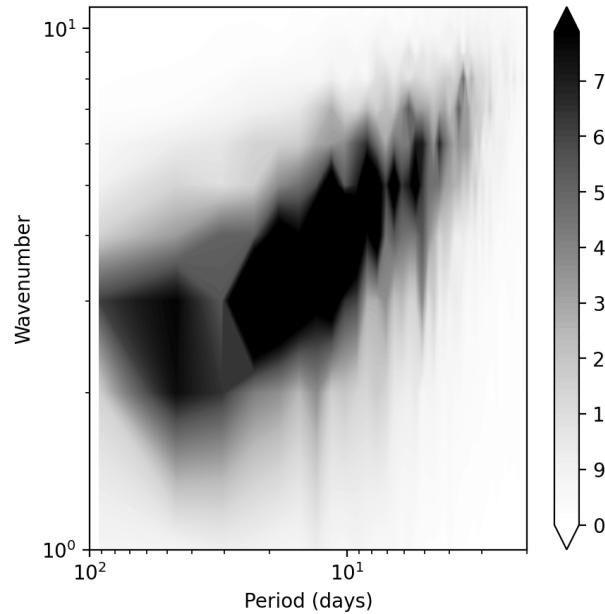
The method of estimating uncertainties known as *bootstrapping* will be used. This involves breaking the total temporal sequence into 10 subsets and computing the relevant quantities correspondingly. Afterwards, mean  $\langle x \rangle$  serves as the best estimate, while standard deviation  $\sigma_x$  as uncertainty.

### 3 HEMISPHERIC TELECONNECTION PATTERNS

#### 3.1 Hayashi spectra of midlatitude atmospheric waves

The extreme teleconnection patterns found by Ragone et al. (2018); Ragone and Bouchet (2020, 2021) in simulations with climate models of different complexity, suggest that the local atmospheric blockings responsible for heatwaves occurring in subregions of Europe could be associated with stationary or very slowly moving atmospheric waves with wavenumber between 3 and 4. The body of literature of Petoukhov et al. (2013, 2016, 2018); Kornhuber et al. (2019, 2020) identifies the amplification of quasi-stationary Rossby waves with a different range of wavenumbers, between 5 and 8 as responsible for the occurrence of several observed midlatitude extreme events, including heatwaves.

Since we are seeking to detect global quasi-stationary Rossby wave patterns we use classical space-time Fourier decomposition into eastward/westward and stationary Rossby waves Hayashi (1971). This analysis is performed on 500 hPa geopotential height averaged in the meridional direction over a latitudinal band centered over the region affected by the heatwaves. The procedure is done via an assumption about the nature of the waves, such as attributing incoherent part of the spectrum to real travelling waves Pratt (1976);

**Figure 1a.**  $H_E(k, \omega)$ ,  $\text{th} = \infty$ .**Figure 1b.**  $H_E(k, \omega)$ ,  $\text{th} = -4.5$ .

**Figure 1.** Comparison of Hayashi spectra for **(A)** 1000 years of CESM climatology and **(B)** a subset of 1000 years of CESM with Scandinavian heatwaves above the threshold of 4.5 degrees of  $T = 30$  day heatwaves. Boundaries of the latitude band over which the meridional average is taken are 55 and 75.

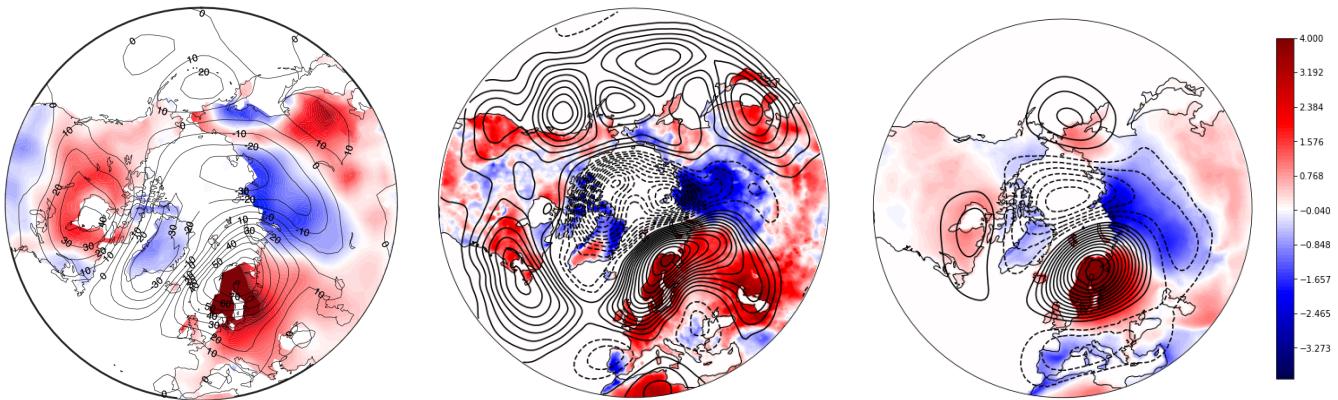
Hayashi (1979). The low frequency - low wavenumber domain corresponds to standing and westward propagating waves, while the high frequency-high wavenumber to synoptic disturbances. For the details also see the appendix of Dell'Aquila et al. (2005). We chose the latitudinal belt 55–75 N as a region relevant for Scandinavia over which the averaging was performed. This belt is part of the larger belt 30–75 N where most of the baroclinic and the low frequency wave activity occurs. Following the procedures in the literature, the meridional integration is an average, not a sum, so the total power does not scale with the size of the latitude band. Additionally, in order to compensate for the non-constant density of points on a log-log plot the spectra are multiplied by  $k \cdot \omega / 2\pi$ .

In what follows we work with conditional averages (condition is to keep only the years where  $A(t) > \text{th}$ , equation (1), for a certain threshold  $\text{th}$ , so that  $\text{th} = -\infty$  corresponds to climatology). We show Hayashi spectra for different thresholds on Figure 1. On the left panel we plot climatology threshold  $\text{th} = -\infty$ , which displays the classical spectrum of Rossby waves. On the right panel we plot Hayashi spectra for the summers with 16 most extreme  $T = 30$  day heatwaves, which corresponds to the threshold  $\text{th} = 4.5K$ . We observe a structure at a period of order 50 days and wavenumber 3 for eastward propagating waves. We note that this pattern in the Hayashi spectrum is isolated from the rest of the Rossby wave branch and it suggests that quasi-stationary processes take place when conditioned to long-lasting heatwaves.

### 3.2 Extreme teleconnection pattern

Hayashi spectra associated with heatwaves in Scandinavia, computed with the CESM model, show a clear local, low-frequency, low wave-number peak (See Figure 1). There are two questions that arise immediately:

- What is the real-space pattern that corresponds to this peak?

**Figure 2a.** PlaSim rare event**Figure 2b.** ERA5 reanalysis**Figure 2c.** CESM 1000 years

**Figure 2.** (A) Northern Hemisphere surface temperature anomaly (colors) and 500-hPa geopotential height anomaly (contours), conditional on the occurrence of European heatwaves with  $T = 90$  days and the threshold of  $a = 2K$ , estimated from the large deviation algorithm, courtesy to Ragone et al. (2018). (B) ERA5 reanalysis averaged over July 2018 and detrended relative to the 1950–2020 linear trend, we follow the same conventions for representing the geopotential height and temperature anomalies, except that negative geopotential anomalies are presented via dashed lines. (C) Composite map of CESM drawn from 1000 year long run and conditioned to  $T = 30$  day heatwaves above the threshold of 3.5 degrees. We follow the same conventions for representing the geopotential height and temperature anomalies. The temperature colormap is shared among the three plots.

- Does the teleconnection pattern have a counterpart in the observational record

The chief quantities of interest are 500 hPa geopotential height (500GPH) and the 2 meter temperature (T2M). The 500GPH is a convenient quantity to analyse the large scale circulation in the troposphere where the effects of the topography are minor (Blackmon (1976)).

To answer the first question we compute composite maps (see equation (4)) of 500GPH and T2M on Figure 2 conditioned to the  $T = 30$  day long heatwaves of Scandinavia of threshold  $a = 3.5K$  (containing 65 events). We observe a tripole-like structure, which consists of an Arctic cyclonic anomaly surrounded by three anti-cyclonic anomalies, over Scandinavia, Chukotka and North-East Canada. The cyclonic anomaly extends to Greenland and Siberia, and it is associated with negative temperature anomalies. This teleconnection pattern indeed corresponds to the local maximum in the low-frequency low-wavenumber branch of the Hayashi spectrum (Figure 1b, 1).

Motivated by the second question we select the July 2018 event from ERA5 reanalysis. The details concerning the data processing of ERA5 are discussed in the Section 2.1.3. The July 2018 event consists of a prolonged heatwave in Scandinavia, whose temperature anomaly was above  $3 - 3.5$  degrees for about a month. Its composite map (Figure 2b) displays various similarities with the CESM teleconnection (Figure 2c). Most striking resemblance are the negative Arctic 500GPH anomaly, as well as positive anomalies in Scandinavia and North-East Canada. Reanalysis favors a positive 500GPHA shifted eastwards towards the Atlantic. In the Pacific ocean we see a train of 500GPHA instead of a single positive anomaly. In general, one finds that geopotential anomalies are stronger in reanalysis; for instance, 500GPHA over Scandinavia is about  $140m$ , whereas in CESM we find values on the order of  $70m$ . The stronger anomalies can be attributed to the fact that in reanalysis we were averaging over the single event, while in CESM the averaging is performed over multiple heatwaves.

	France	Scandinavia
PlaSim	3.23K	1.74K
CESM	3.47K	3.79K
ERA5	3.45K	3.25K

**Table 1.** Values of a  $r = 4$  year return time threshold  $a_4$  computed for France and Scandinavian heatwaves (columns) over PlaSim, CESM and ERA5 datasets (rows).

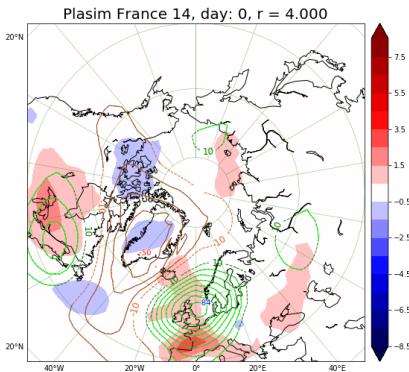
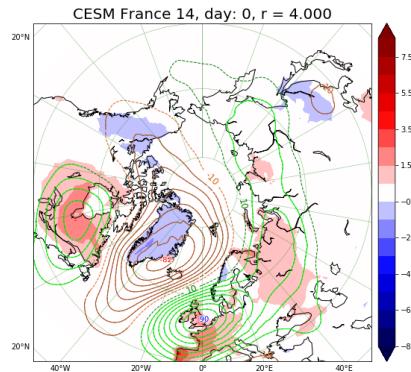
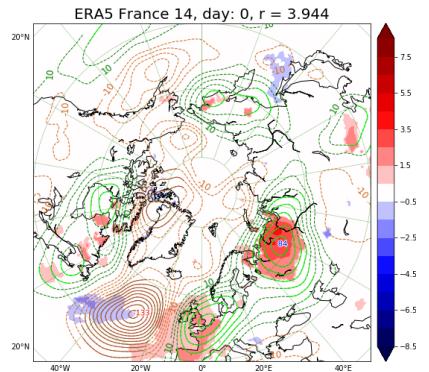
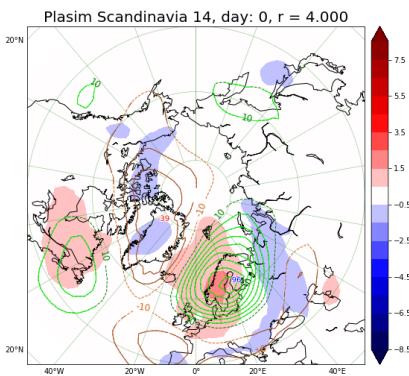
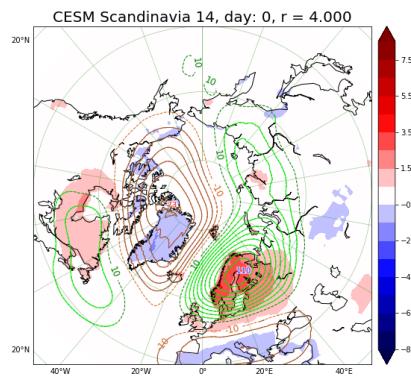
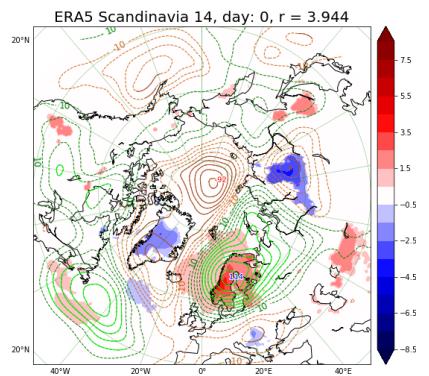
We are now in a position to also comment on the relative similarity between these two teleconnection patterns with the one presented in Ragone et al. (2018). For illustration purposes we display the copy of the relevant teleconnection on Figure 2a. It was obtained using the rare event algorithm for PlaSim conditioned to  $T = 90$  day heatwaves of threshold  $a = 2K$  occurring in Europe (to be more precise each grid point is averaged over 90 days). It turns out that the main contribution is due to heatwaves in Scandinavia, of threshold that can be estimated as  $a \sim 3.5 - 4K$ . It is interesting that the similarities between the panels exist despite the different complexity of the two models, PlaSim and CESM. This gives us indication that such teleconnection patterns could be robust feature of the large scale dynamics. However, the specific conditions are not identical, and the similarity to the July 2018 event could be accidental. In order to better test this, we perform a more extensive intra-model comparison in the next Section.

### 3.3 Teleconnections in models versus reanalysis

We aim to provide systematic comparisons between the models. We concentrate on two areas in Europe, corresponding to France and Scandinavia. The areas have been chosen to correspond closely to the heatwave clusters in Western Europe and Scandinavia identified by Stéfanon et al. (2012).

Since the ERA5 dataset contains only 71 years we limit the study to return times not larger than 4 years, in order to have enough data (at least 18 heatwaves). For consistency we select heatwaves with the same return time of 4 years in CESM and PlaSim (250 heatwave events in both models) which allows for stronger statistical inference. The corresponding thresholds  $a$  are not the same across the models, especially in Scandinavia, where the threshold for PlaSim is much lower than the other two datasets (see table 1). This is further discussed in Section 4.2.1.

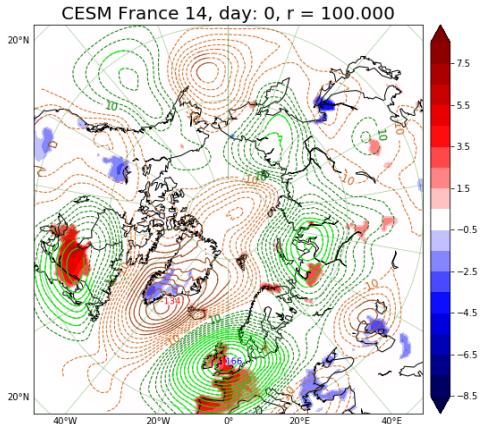
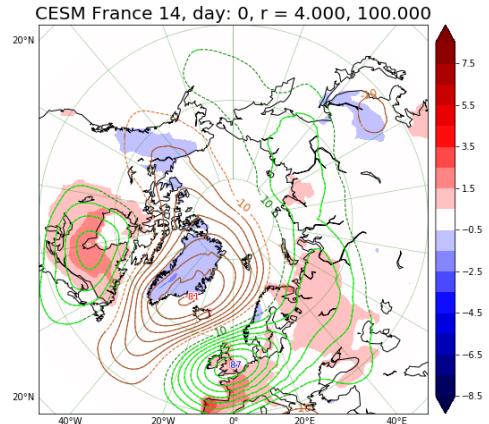
One wonders how similar are the teleconnection patterns drawn from models at the onset of individual heatwave events, i.e. at  $\tau = 0$  days. We take duration  $T = 14$  days, and plot the corresponding composites on Figure 3, following the method described in Section 2, for heatwaves occurring in France (top) and Scandinavia (bottom). We clearly see the presence of a tripole-like structure in all the images, but there are many additional details. PlaSim and CESM share more features, although qualitative similarities can also be seen when comparing with ERA5. In particular, one observes statistically significant positive 500GPHA in the North sea of order 84 m in PlaSim, 90 m in CESM and 50 m in ERA5. The second largest in magnitude feature is the negative Arctic 500GPHA of order -50 m in PlaSim and -82 m in CESM which extends to the Atlantic. The extension of this pattern seems to be consistent with what one observes in ERA5. However, in ERA5 it is not statistically significant. In contrast, we have a statistically significant negative 500GPHA in the Atlantic ocean of -133 m, which is not so strong in PlaSim or CESM. The third order feature is the positive 500GPHA in East Canada which reaches 30 m in CESM and PlaSim and approximately 40 m in ERA5. Finally, in PlaSim we have 10 m in North-East Asia anomalies, which are stronger in CESM and ERA5. Most features are statistically significant for CESM and PlaSim but many are

**Figure 3a.** PlaSim France**Figure 3b.** CESM France**Figure 3c.** ERA5 France**Figure 3d.** PlaSim Scandinavia**Figure 3e.** CESM Scandinavia**Figure 3f.** ERA5 Scandinavia

**Figure 3.** Composites of  $r = 4$  year return time in different models (left to right), where 2m temperature and 500 mbar geopotential anomalies are plotted conditioned to different areas (top-bottom): (A) PlaSim France. (B) CESM France. (C) ERA5 France. (D) PlaSim Scandinavia. (E) CESM Scandinavia. (F) ERA5 Scandinavia. at  $\tau = 0$  lag (onset of the heatwave). The colormap for the temperature is indicated next to the figures. Only  $|t| > 2$  statistically significant anomalies of temperature are displayed (see the colormap in K), whereas all level sets of geopotential height anomalies graded by 10m are shown (Blue - positive, Orange - Negative). The values of geopotential that are statistically significant  $|t| > 2$  are plotted with solid lines and otherwise with dashed lines (see Section 2.3 on statistical significance). The global maximum and minimum of geopotential is shown using blue and red text tooltips respectively. All subsequent figures displaying composites will follow this general style.

not for ERA5 (see Section 2.3 on statistical significance test), which is due to scarcity of the observational record.

The panels at the bottom of Figure 3 corresponding to teleconnection patterns conditioned to Scandinavian heatwaves also show three features: a *primary* one consisting of positive 500GPHA in Scandinavia: 96 m in PlaSim, 110 m in Scandinavia and 114 m in ERA5; a *secondary* one consisting of negative 500GPHA in the Arctic with -39 m in PlaSim, -73 m in CESM, and in ERA5 of order -90m. While this peak is larger than expected, we find that overall shape and magnitude of 500GPHA in the Arctic are qualitatively consistent between the models and reanalysis. The *tertiary* pattern consists of positive 500GPHA in Eastern Canada which is approximately 20 m in both PlaSim and CESM and of order 60 m

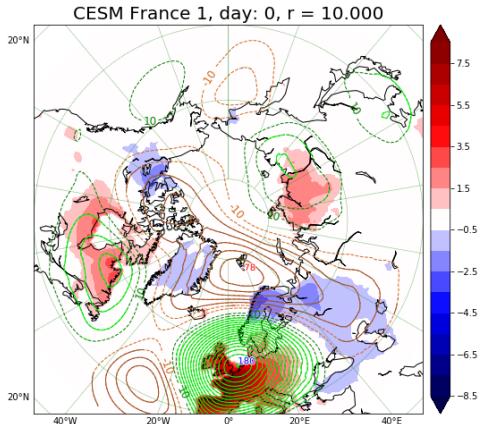
**Figure 4a.**  $r > 100$  years**Figure 4b.**  $4 < r < 100$  years**Figure 4.** Comparison between (A)  $r > 100$  and (B)  $4 < r < 100$  for France heatwaves in CESM dataset. See the caption of Figure 3 for plotting conventions.

in ERA5 with a easterly shift. Finally there is also a pattern which seems to be generally consistent across the models and reanalysis: the Mediterranean depression of order -20 m. However, in reanalysis it is, again, not statistically significant.

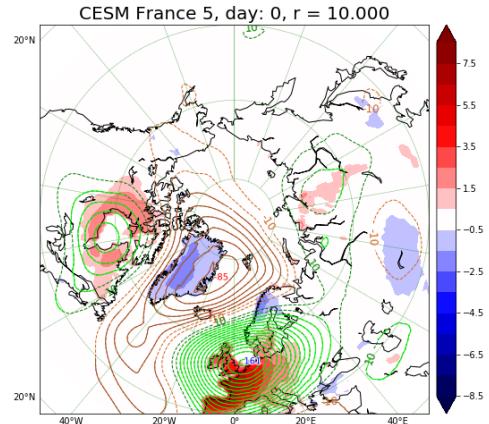
To summarize, we find qualitative agreement between Figures 3a, 3b and 3c as well as agreement between 3d, 3e and 3f. CESM seems to be capturing the teleconnection patterns closer to reanalysis. For instance, in case of heatwaves conditioned to France we see that CESM captures better the distribution of geopotential anomalies over Canada and Northern Russia when compared to ERA5. Nevertheless, the similarities displayed in the teleconnection patterns suggest that they are model independent and have observational counterparts that are rather robust.

Another important question is how these patterns change when conditioning for more extreme heatwaves. In Figure 4 we compare extreme heatwaves with return time larger than 100 years, with the ones with return time in the range  $4 < r < 100$ . In the latter case we have actually removed the 10 most extreme events to show that they do not contribute noticeably compared to the remaining 240 events. Like in the previous analysis we identify the three most important features as positive GPHA over North sea, Greenland and East Canada. In Figure 4a, for the events that are more extreme, they take values 166 m, -134 m and approximately 90 m, while on Figure 4b, for the events that are milder, they take the values 87 m, -81m and approximately 40 m. Furthermore, Figure 4a has variety of other mostly statistically insignificant cyclonic/anticyclonic anomalies. The conclusion is that when going from less extreme to more extreme heatwaves, the most significant features remain geographically fixed (500GPHA naturally deepen), while we obtain a few more anomalies that do not pass  $|t| > 2$  test.

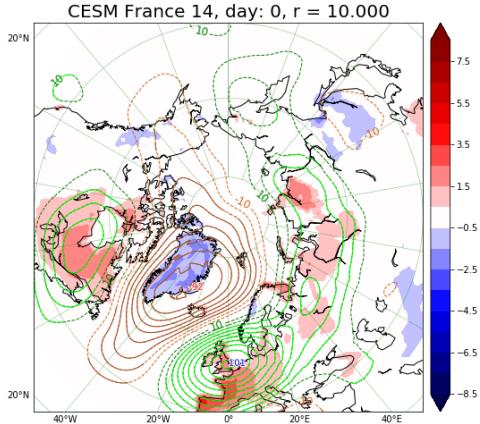
Finally, we assess how the teleconnection patterns depend on the choice of the heatwave duration  $T$  (see equation (1)). We test 4 periods,  $T = \{1, 5, 14, 30\}$  days, that are shown in Fig. 5 for the CESM dataset conditioned to heatwaves in France. Note that Figure 3b is repeated here as Figure 5c for convenience. The patterns discussed before, including the nodes of the tripole structure, such as positive GPHA on the North sea, Canada and East Siberia can be recovered also from Figure 5b and Figure 5d with mostly the same intensities and some minor shifts. Most different of all is Figure 5a which is actually a daily



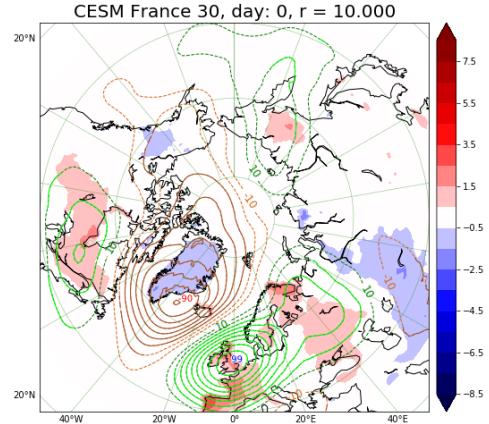
**Figure 5a.**  $T = 1$  days



**Figure 5b.**  $T = 5$  days



**Figure 5c.**  $T = 14$  days



**Figure 5d.**  $T = 30$  days

**Figure 5.** Composites of 2m temperature and 500 mbar geopotential conditioned on  $r \geq 4$  year return time heatwaves over France at  $\tau = 0$  lag. The reference model is CESM. Among the different panels we vary the length of heatwaves: **(A)**  $T = 1$ , **(B)**  $T = 5$ , **(C)**  $T = 14$  and **(D)**  $T = 30$  days. See the caption of Figure 3 for further plotting conventions.

average. In this case the Arctic anomaly descends towards Scandinavia/western Russia region and we have an isolated negative GPHA in the Atlantic. The differences are to be expected since on daily time scale we may observe many of the synoptic features. We conclude that the features we have described in CESM at  $\tau = 0$  appear to be also robust across different time-averaging intervals, namely at  $T = \{5, 14, 30\}$  days. However, we observe that for shorter periods such as 1-5 days negative anomaly in the Arctic develops tongues that reach the midlatitudes. These are not present for large values of  $T$  as the faster synoptic type perturbations are averaged over.

## 4 TEMPORAL EVOLUTION OF THE EXTREMES

The goal of this Section is to compare the temporal statistics between the reanalysis and the models. The time evolution of composite statistics will only be considered for CESM, whereas other types of statistical properties will be compared across the models.

### 4.1 The dynamical evolution of composites

Here we address the dynamical evolution leading up to and during the heatwave. We limit this study to the CESM model because we lack data in ERA5 reanalysis and CESM is a higher fidelity model than PlaSim. In Figure 6 we show the dynamical evolution of heatwave composites, conditioned to  $r \geq 4$ , in France, and varying the lag time  $\tau$ . This shows how the teleconnection patterns change as a function of lead time. We see that 15 days prior to the onset of heatwaves (Figure 6a) we already have significant negative GPHA of order -27 m in the Arctic region that slowly broadens and deepens (-35m at  $\tau = -6$  days, Figure 6b, and -67 m at  $\tau = -3$  days, Figure 6c). Meanwhile, an anticyclonic block develops at the midlatitudes, reminiscent of a similar pattern in Figure 5c. Five days after the onset (Figure 6d) we can distinguish the tripole-structure (see Section 3.2) with strong significant 133 m positive GPHA in the North sea and negative GPHA that attains -76 m in the North Pole. This pattern starts to disappear at  $\tau = 15$  days (Figure 6f), corresponding to the duration of the heatwaves.

### 4.2 Time series analysis

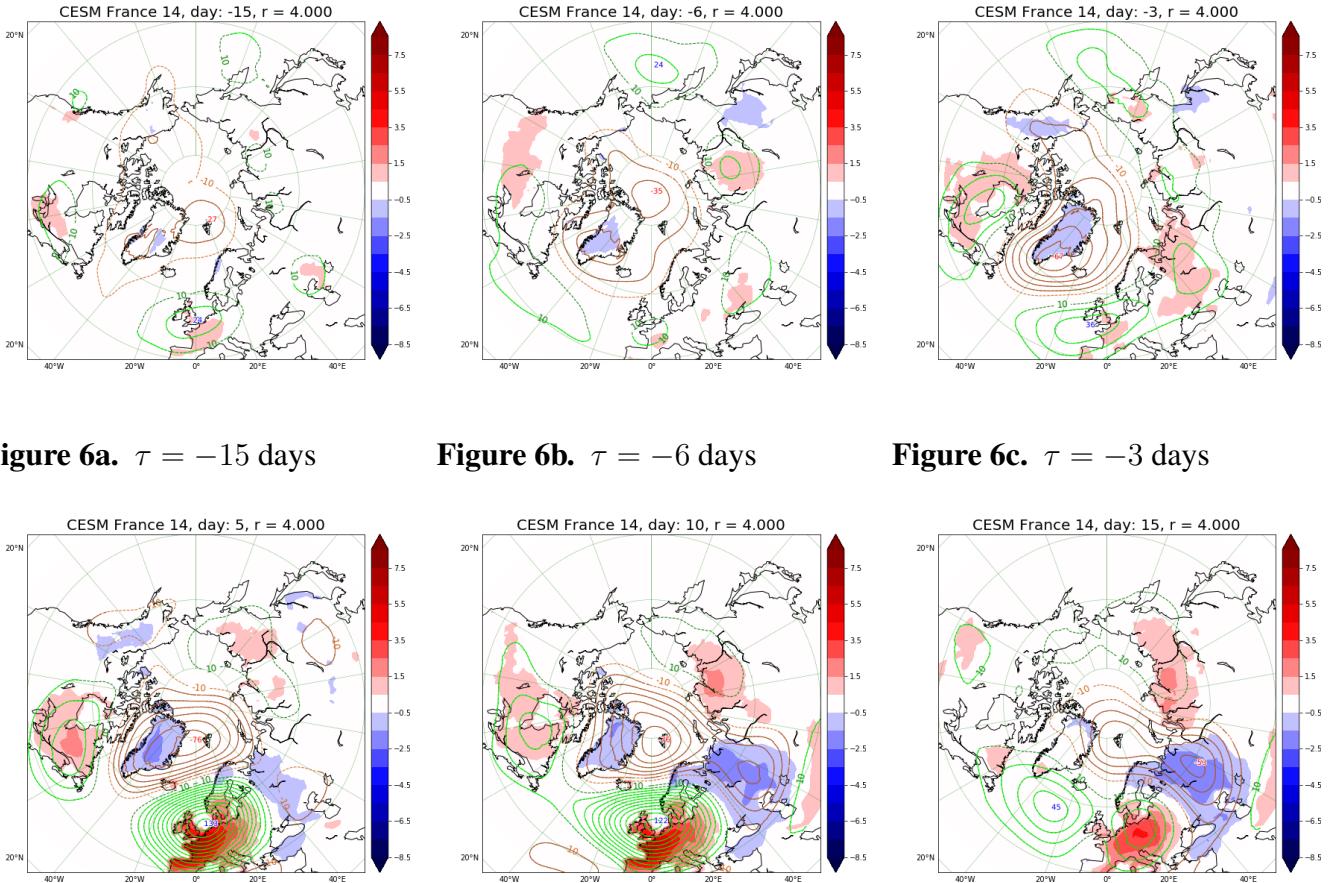
We discuss the properties of the time series of the area averaged temperature, over France and Scandinavia separately. The idea is to show how closely the models can resemble the real time series and, further, whether we can reduce the dynamics to a simple stochastic process.

#### 4.2.1 Return time plots

We show the probability distributions of the 14 day running mean of two meter temperature anomaly (T2MA) over France and Scandinavia on Figure 7. Both areas have Probability Distribution Functions (PDFs) that are Gaussian in the bulk, with fatter tails, particularly for PlaSim (see in Section 4.2.3 for more details on this). In general, reanalysis datasets are characterized by a noticeable skewness but relatively small kurtosis, which is captured well by CESM in both areas. PlaSim, on the other hand, does not represent well PDF of ERA5 in Scandinavia. Nevertheless, Gaussian statistics seems a reasonable initial assumption as we shall see in Section 4.2.3. This is consistent with the fact that we perform long-time large-area averages: time-averaged statistics that are long enough should asymptote to Gaussian distribution, assuming they are not too correlated.

Figure 8 depicts return time plots computed using equation (3). The left panels correspond to return times for France heatwaves. In particular, Figure 8a shows a remarkable agreement between PlaSim and reanalysis, despite PlaSim being a climate model of intermediate complexity. Long-lasting 90 day heatwaves are fitted even better by CESM (Figure 8c). If we look at the most extreme 14 day events (orange dots) in ERA5 reanalysis there are two apparent outliers (the most extreme being the European heatwave 2003) compared to the climate models. However, since for these two events we are in the rightmost tail of the distribution where we don't have enough data one could argue that the climate model fits we described cannot be rejected solely based on these two events.

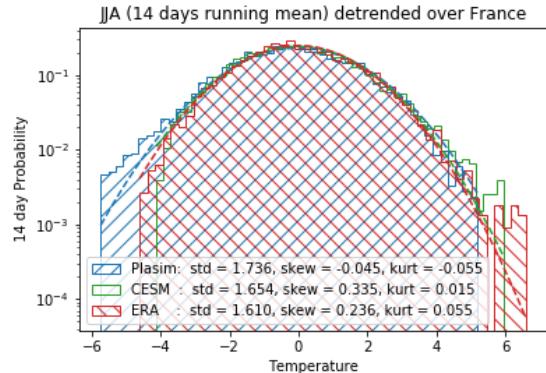
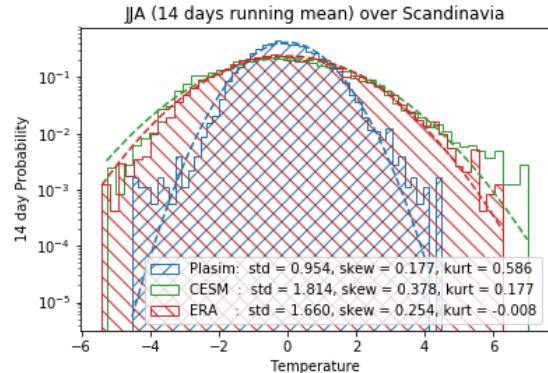
For Scandinavia the models do not compare so well with ERA5 reanalysis. For instance, Figure 8b demonstrates that PlaSim systematically underestimates the intensity of the extreme events by a large margin and across different values of  $T$ , i.e. not only the few extreme outliers. As a result, the 4 year

**Figure 6a.**  $\tau = -15$  days**Figure 6b.**  $\tau = -6$  days**Figure 6c.**  $\tau = -3$  days**Figure 6d.**  $\tau = 5$  days**Figure 6e.**  $\tau = 10$  days**Figure 6f.**  $\tau = 15$  days

**Figure 6.** Composites of 2m temperature and 500 mbar geopotential conditioned on  $r \geq 4$  year return time heatwaves in France at  $\tau = 0$  lag. The reference model is CESM and  $T = 14$  days is chosen. Control parameter for this figure is  $\tau$  with (A)  $\tau = -15$ , (B)  $\tau = -6$ , (C)  $\tau = -3$ , (D)  $\tau = 5$ , (E)  $\tau = 10$  and (F)  $\tau = 15$  days, thus time evolution of CESM composite heatwave in France is displayed. See the caption of Figure 3 for plotting conventions.

return time has a threshold  $a_4 = 1.74K$  in PlaSim and  $a_4 = 3.25K$  in CESM consistent with the Table 1. Figure 8d shows that CESM performs much better for Scandinavia. The return times of the events are overestimated, at least for  $T < 30$  days. In other words, relationship between thresholds and return times is not the same. Nevertheless, it is much closer to reanalysis, for instance for 4 year return time events the predicted threshold is  $a_4 = 3.79K$  (Table 1).

One may wonder why models have difficulty accurately describing Scandinavia heatwaves. From the theory of single time scale Gaussian stochastic fluctuations, return times depend on the variance and decorrelation time scale of the time series ( Lestang et al. (2018)). The picture is more complicated when dealing with time averages of signals with multiple time scales. However, generally speaking, return times are expected to scale with the variance of the underlying probability distribution (See Section 4.2.2). Figure 9 shows standard deviation of the grid-point 2 meter temperature anomaly (T2MA) PDF taken over JJA period in PlaSim CESM and ERA5. It demonstrates that the models have trouble capturing precisely the variance over the region of north Russia/ Scandinavia, where PlaSim has a strong negative bias while

**Figure 7a.** France**Figure 7b.** Scandinavia

**Figure 7.** 14 day running mean distributions of daily two meter temperature anomaly (T2MA) integrated over the area of (A) France, (B) Scandinavia, in 3 different datasets: (blue) JJA 1000 year dataset of PlaSim, (green) JJA 1000 year dataset of CESM, (red) JJA 1950-2020 ERA5 reanalysis dataset. We see that France data follows roughly Gaussian statistics, while in Scandinavia PlaSim and to some extent CESM favor non-Gaussianity. The values of standard deviation (std), skewness (skew) and kurtosis (kurt) are given in the inset for each color-coded distribution.

CESM a small positive bias. This goes hand-in-hand with how ERA5 return time plots are underestimated in Figure 8b and overestimated in Figure 8d. In general, according to Figure 9, PlaSim captures poorly standard deviation in higher latitudes. If we consider other parts of the European continent differences are small and, broadly speaking, CESM compares more favorably to ERA5.

#### 4.2.2 Autocovariance function

To have a more complete understanding of the statistics of the tails of the distribution, the standard deviation is not sufficient and the decorrelation properties of the signal must be investigated. We compute the autocovariance function of the T2M time series over France and Scandinavia in PlaSim and CESM and compare them to the corresponding time series in ERA5. The autocovariance function is given by <sup>2</sup>

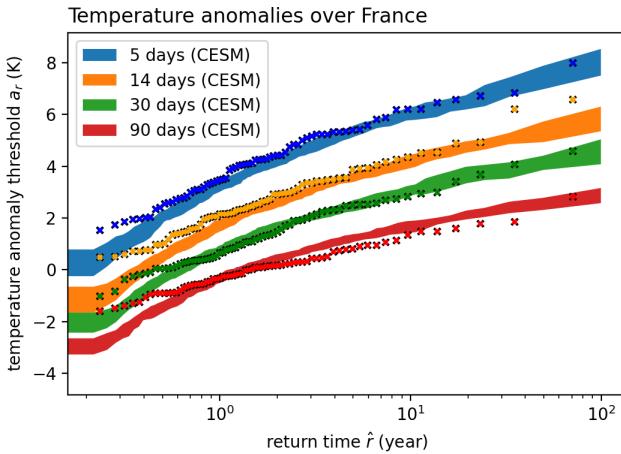
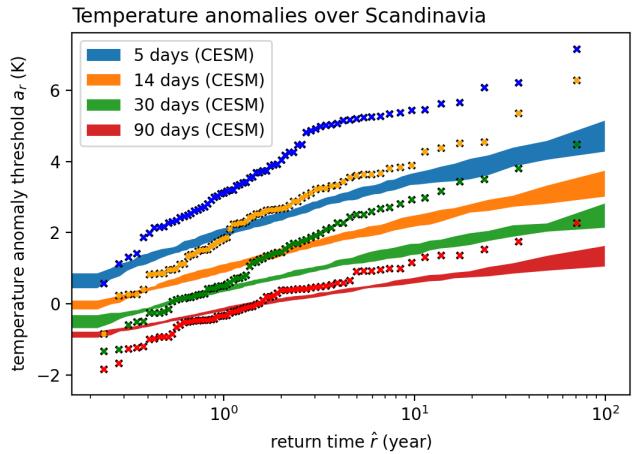
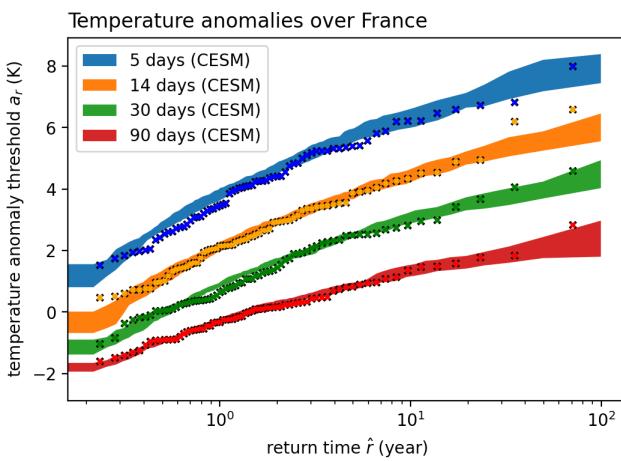
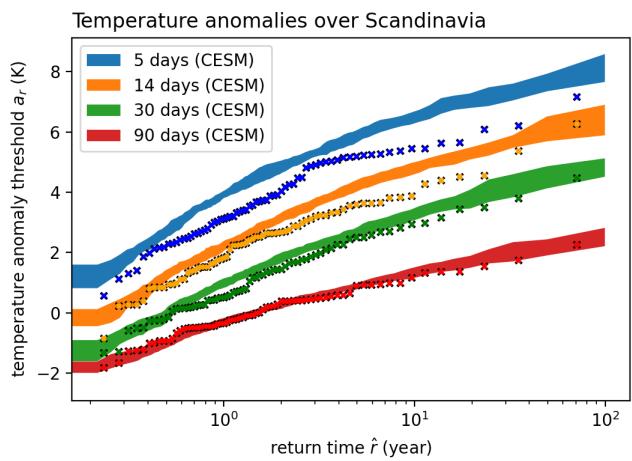
$$C(t) := \mathbb{E}[A(X(t))A(X(0))], \quad (8)$$

When the process can be described by a single time scale one expects exponential scaling and the corresponding time scale can be extracted simply as

$$\tau_c := \int_0^{+\infty} \frac{C(t)}{\mathbb{E}[A^2(X(0))]} dt \quad (9)$$

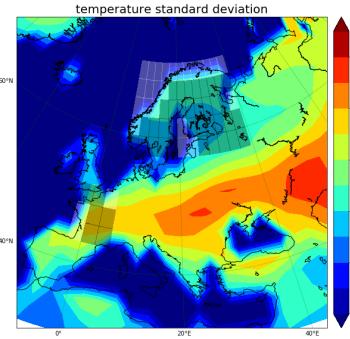
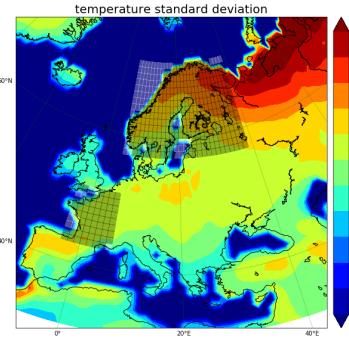
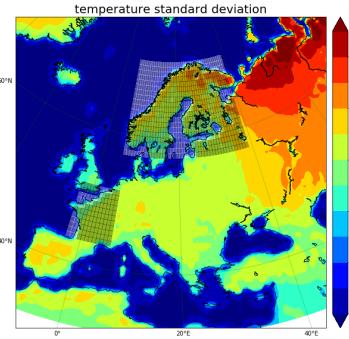
A similar analysis has been performed in Ragone and Bouchet (2020) on 1000 years of data generated with PlaSim in perpetual summer conditions. In that case it was found that time series of area averaged surface temperatures show two time-scales, the *fast* one of order of 4 days related to synoptic variations and the *slow* one of the order of 30 days, which the authors relate to heat capacity of the soil and therefore the soil moisture content.

<sup>2</sup> We give un-normalized definition of autocovariance, which in some disciplines is referred to as autocovariance, whereas autocovariance is reserved for normalizing it by variance. In some disciplines both terms are used interchangeably

**Figure 8a.** PlaSim France**Figure 8b.** PlaSim Scandinavia**Figure 8c.** CESM France**Figure 8d.** CESM Scandinavia

**Figure 8.** Return time plots for heatwaves, summer extremes of two meter temperature anomaly (T2MA), in different regions (left-right) and different models (bottom-top). **(A)** PlaSim France, **(B)** PlaSim Scandinavia, **(C)** CESM France and **(D)** CESM Scandinavia. The dots correspond to ERA5 data, while the shaded regions to bootstrapping on climate model data, i.e. the data is split into 10 subsets and return times are computed using the approach described in Section 2.2. Then, mean and standard deviation are computed so that the shaded region corresponds to mean plus or minus one standard deviation. The colors and the corresponding duration of heatwaves  $T$  are indicated on the legend.

As stated above, the properties of the autocovariance function provide a contribution to the return time plot. As is known from the theory of stochastic processes Gardiner et al. (1985) a Gaussian process can be generated knowing the parameters of the autocovariance function. We compute it using the PlaSim and CESM time series of T2M integrated over the area of France and Scandinavia and plot the results on Figure 10. The plots show that the processes involved indeed are well represented by two time scales, which means they can be fitted by two exponential functions. The fit parameters do not precisely match the ones provided by Ragone and Bouchet (2020) which is not surprising since in that study a heatwave over much larger European region was considered in the perpetual summer regime. Another important parameter to be inferred from Figure 10 is the *cross-over scale*, which corresponds to the time at which the slow time scale starts to dominate the decay.

**Figure 9a.** PlaSim**Figure 9b.** CESM**Figure 9c.** ERA5

**Figure 9.** T2M standard deviation over JJA in different data sets: **(A)** PlaSim, **(B)** CESM, **(C)** Detrended ERA5 reanalysis. Relevant areas (France and Scandinavia) that are used below for regional analysis are shaded.

	France	$A_1$	$A_2$	$\tau_1$	$\tau_2$		Scandinavia	$A_1$	$A_2$	$\tau_1$	$\tau_2$
PlaSim	6.0	1.9	2.1	108		PlaSim	2.1	0.13	2.7	82	
CESM	6.5	0.69	3.4	38.6		CESM	7.1	0.28	4.6	80	
ERA5	5.9	0.45	3.4	108		ERA5	5.0	0.38	4.5	62	

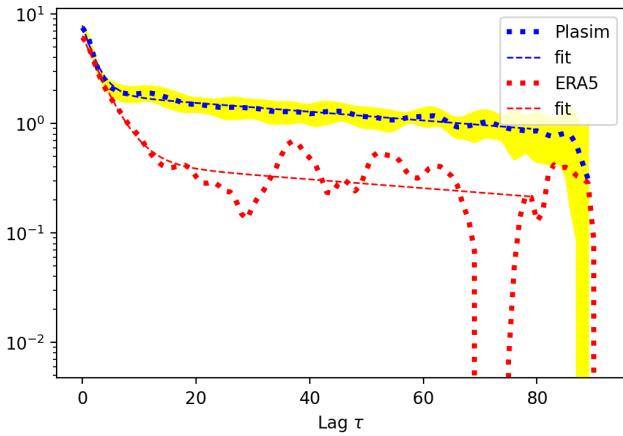
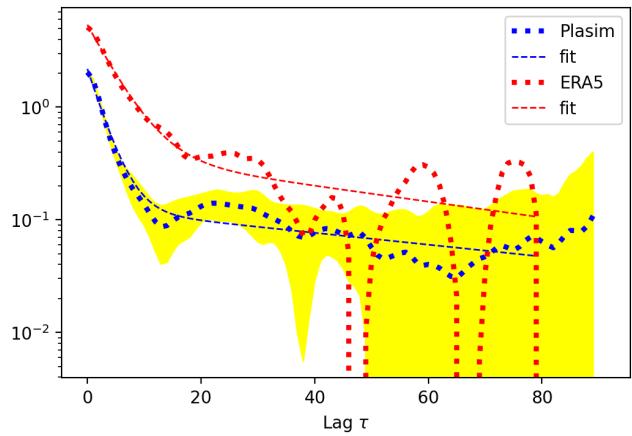
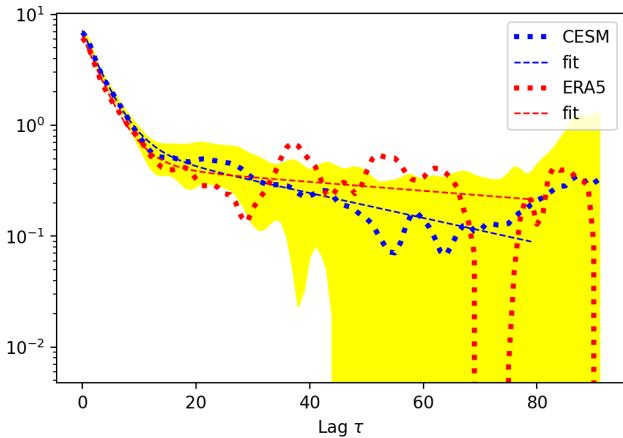
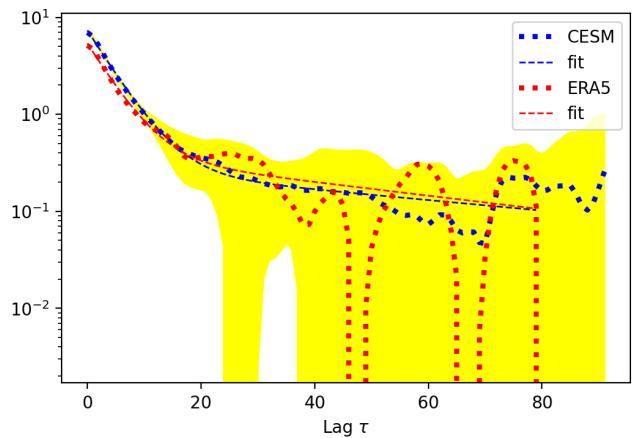
**Table 2.** Exponential parameters for equation (10) fit. The coefficients are given inside the table for each model (rows) and both areas: France on the left and Scandinavia on the right.

The presence of two time scales motivates performing exponential fit, using the following nomenclature for its parameters

$$f = A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right), \quad (10)$$

with  $\tau_1 < \tau_2$  the *fast* and *slow* time-scales respectively. The importance of the slow time scale depends on the ratio  $A_1/A_2$ . The fit is applied to PlaSim/CESM/ERA5 for both France and Scandinavia T2M series and the parameters of the fit are displayed in Table 2. According to the table (as well as the corresponding Figure 10a) one finds that PlaSim is biased especially in Scandinavia, where it fails to capture both the variance (also inferred from the value  $C(0)$ , i.e. at lag  $\tau = 0$ ) and the fast time scale: both  $A_1$  and  $\tau_1$  are off by approximately factor 2. These discrepancies translate directly into mismatch between return time plots of Plasim and ERA5 as displayed in Figure 8b already discussed earlier.

CESM fits ERA5 quite well in both France and Scandinavia, although in the latter case there is an over-estimation of variance discussed earlier in Section 4.2.1. Since ERA5 time series is relatively short we cannot make robust comparisons for the second time scale, which tends to be of the order of 30 to 80 days and, as already mentioned, is likely associated with soil-atmosphere interactions. We note that PlaSim overestimates the importance of this second time-scale, which on Figure 10a leads to scaling which resembles slowly decaying plateau. This quasi-plateau can be identified also in CESM and ERA5, however to a lesser degree. This means that France heatwaves in PlaSim have a bias towards stronger temporal persistence than in observations. Earlier we concluded that CESM fits return time for France heatwaves somewhat better than Plasim (Figures 8a and 8c). Thus one may wonder to what extend fitting the second

**Figure 10a.** PlaSim France**Figure 10b.** PlaSim Scandinavia**Figure 10c.** CESM France**Figure 10d.** CESM Scandinavia

**Figure 10.** autocovariance function accross two regions (left-right) and two models (top-bottom): **(A)** PlaSim France, **(B)** PlaSim Scandinavia, **(C)** CESM France and **(D)** CESM Scandinavia. The blue dotted curves correspond to the 1000 year-long time series, while the blue dashed lines represent the fit given by equation 2. To provide an estimate of uncertainty we have performed bootstrapping (splitting the time series into 10 trajectories) indicated by the yellow region which represents mean plus or minus one standard deviation. The ERA5 autocovariance function over 71 years is displayed via red dashed line and the underlying time series by red dotted line.

time scale is indispensable for providing a better fit. What if  $A_2 = 0$ ? We will address this question in the next section.

#### 4.2.3 Gaussian process

In this Section we compare the statistics of the time series presented above to the synthetic ones generated from Ornstein-Uhlenbeck (OU) process

$$dX_t^{\alpha,\epsilon} = -\alpha X_t dt + \sqrt{2\epsilon} dW_t \quad (11)$$

which can be interpreted as reformulation of a Langevin equation with correlation time  $\tau_c = \alpha^{-1}$  Gardiner et al. (1985). From the corresponding Fokker-Planck equation one finds a stationary probability density

$P_s(x)$  that is a Gaussian distribution with standard deviation  $\sigma = \sqrt{\epsilon/\alpha}$ .

$$P_s(x) = \sqrt{\frac{\alpha}{\pi\epsilon}} \exp\left(-\frac{\alpha x^2}{\epsilon}\right) \quad (12)$$

$$\langle X(t)X(s) \rangle_s = \frac{\epsilon}{2\alpha} \exp(-\alpha|t-s|) \quad (13)$$

The first passage time conditioned on the stationary measure Lestang et al. (2018) can be computed analytically for this model.

However, it is not so simple to compute analytically the autocovariance of the time averaged series. We will therefore adopt an empirical approach following Herbert and Bouchet (2017). We simulate the OU process based on the parameters obtained from the exponential fit on Figure 10 and prepare our Gaussian process by a direct sum of the two time series obtained from the OU integrator.

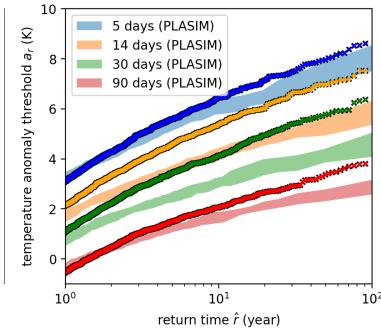
$$X_t^{\alpha,\epsilon} = X_t^{\alpha_1,\epsilon_1} + X_t^{\alpha_2,\epsilon_2} \quad (14)$$

The parameters are chosen as  $\alpha_i = 1/\tau_i$  and  $\sqrt{2\epsilon_i} = A_i\alpha_i$  (see Table 2). The synthetic time series of length  $10^5$  (“days”) are generated which we then separate into 10 instances of 100 year-long sequences, each of the length 100 “days” in order to mimic the summer sequences from the models. The mean and standard deviation is computed based on these 10 instances. The resulting return times (equation (3)) are plotted on Figure 11 against the ones from General Circulation Models (GCMs) that were shown on Figure 8. We observe that the match between the Gaussian process and GCMs is relatively good, especially in case of PlaSim over France (Figure 11b): the model points are within the error bars of the bootstrapped OU simulation. This is consistent with the fact that the corresponding fluctuations in the upper tail of the distribution are rather Gaussian (see Figure 7). Both short and long duration heatwaves are captured relatively well (within the error bars), although in some cases the Gaussian process tends to produce less extreme events.

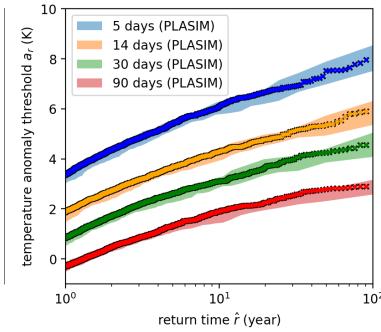
Now we are in a position to asses the importance of the second time scale. This is particularly interesting over France where the corresponding coefficient is larger (see Table 2). To answer this question we only retained the first term in equation (10) and chose the coefficient  $A_1 \rightarrow A_1 + A_2$  and  $A_2 \rightarrow 0$  to match the true variance and the integral time scale for  $\tau_1 \rightarrow \tau_c$ ,  $\tau_2 \rightarrow 0$ , equation (9). The mismatch is quite pronounced for PlaSim (Figure 11a). In the case of CESM (Figure 11d), on the other hand, the differences are modest but are still noticeable as the extremes start to diverge outside of the error bars. Thus we conclude that the second time scale is less relevant in CESM yet it helps to more accurately capture the return time plot.

## 5 CONCLUSION

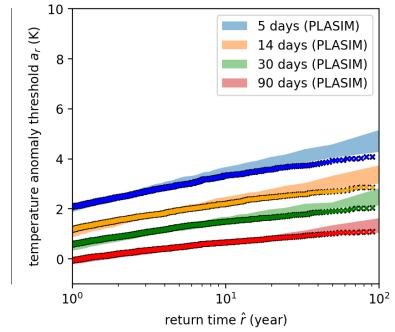
In this paper we presented a spatio-temporal analysis of persistent heatwaves over France and Scandinavia. One of the main findings of this study is that two-week long summer heatwaves are associated with robust teleconnection patterns that involve a strong positive 500 mbar geopotential anomaly (500GPHA) in the region of interest, negative 500GPHA near Greenland and positive 500GPHA in North-East Canada. The patterns are qualitatively consistent across the ERA5 reanalysis and two models of different level of complexity, PlaSim and CESM.



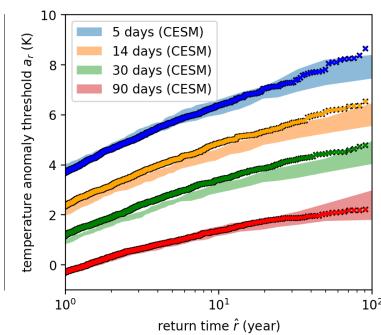
**Figure 11a.** PlaSim France, **Figure 11b.** PlaSim France Single Scale



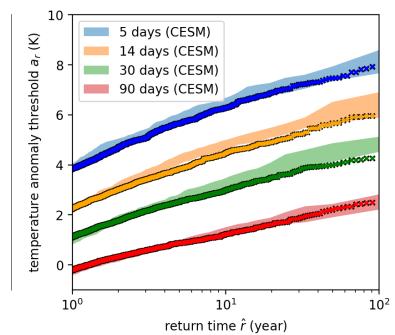
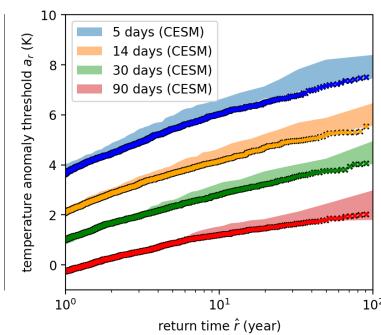
**Figure 11b.** PlaSim France



**Figure 11c.** PlaSim Scandinavia



**Figure 11d.** CESM France, **Figure 11e.** CESM France single time scale



**Figure 11f.** CESM Scandinavia

**Figure 11.** Return time plots for heatwaves in both regions for  $T = \{5, 14, 30, 30\}$  days. **(A)** PlaSim France single time scale, **(B)** PlaSim France two time scales, **(C)** PlaSim Scandinavia two time scales, **(D)** CESM France single time scale, **(E)** CESM France two time scales and **(F)** CESM Scandinavia two time scales. The dots correspond to the synthetic time series obtained from OU process (11). (left) single time scale OU process with (top) Fit parameters:  $A = 8.101, \tau = 18.6$  days, where the former corresponds to the total variance of the PlaSim data and the latter to the corresponding integral time scale and using the same approach (bottom) Fit parameters:  $A = 7.168, \tau = 7$  days. (center, right) two time scale OU process obtained by a direct sum , equation (14), with the parameters taken from the Figure 10. This figure illustrates that we need two time scales to reproduce well the return time plots. The approach works slightly worse in PlaSim Scandinavia, where likely this is due to the strong kurtosis in the model (See Figure 7).

This quasi-stationary structure appears in Hayashi spectra as a wave-number 3 peak isolated from the general eastward-propagating Rossby wave spectrum. We note that the same wave-number 3 pattern is identified when a neural network is trained to predict heatwaves using a longer dataset generated with PlaSim Miloshevich et al. (2023) (but with diurnal cycle). Similar patterns and spectral features can be found during the July 2018 heatwave over Scandinavia. In the literature there has been a discussion Petoukhov et al. (2013, 2016, 2018); Kornhuber et al. (2019, 2020) around the amplification of quasi-stationary Rossby waves with a different range of wavenumbers, between 5 and 8.

The teleconnection patterns are robust to changes in a period (length) of heatwave as long as it is sufficiently large (larger than several days). The main features (wave-number 3 pattern) are consistent across less extreme (with few year return times) and more extreme (from 4 to 100 year return times).

We have also shown that the statistical properties of area averaged surface temperature in the two considered regions can be modelled using a two time-scale Ornstein-Uhlenbeck process. The length of the model runs (1000 years) allows us to draw conclusions on the quality of the fits (estimating uncertainty) of extreme heatwaves with returns as large as 100 years. This allows us to respond positively to the question on whether correctly capturing variance and autocovariance function of the process allows to properly infer the return time curves. The return times of climate models are compared with reanalysis which reveals quite good agreement for heatwaves in France. For the heatwaves in Scandinavia return times in CESM and ERA5 diverge for periods of order 5 to 14 days but are generally well captured for longer events.

These results demonstrate two things. First, for the regions involved in this study, numerical climate models even of moderate complexity can reproduce the main dynamical and statistical features of persistent heatwaves at synoptic scale. In other words, long-lasting, time averaged phenomena are not only more important from the point of view of impacts, but also to a good extent easier to study with numerical models than shorter phenomena. Second, the presence of recurrent large scale teleconnection patterns during long lasting heatwaves over these regions indicates that these events are characterized by a high degree of typicality. This could be potentially very useful in order to identify precursors for the risk of extreme persistent heatwaves, thus improving their predictability, if the dynamical reasons for the formation of these patterns could be properly identified. Recent studies have suggested that typicality of heatwave teleconnection patterns found in reanalysis and CMIP6 models may be compatible with the concept of instanton in large deviation theory (Gálfy et al., 2021; Gálfy and Lucarini, 2021). While this is a suggestive proposition, we stress that comprehensive quantitative analysis, beyond visual similarity of two-dimensional maps of selected fields, is necessary to support this type of hypothesis.

These results motivate future studies which could be performed with CMIP6 ensemble data, and would help to pin down the physical mechanisms behind the teleconnection patterns described in this study, and how they interact with other modes of variability, such as ENSO or NAO. Moreover, it would be interesting to investigate how the teleconnection patterns change under different levels of greenhouse gas emissions and how this affects the frequency, intensity, and duration of heatwaves.

## CODE AVAILABILITY STATEMENT

The coding resources for this work, such as the python and jupyter notebook files, are available on a GitHub page <https://github.com/georgemilosh/Climate-Learning> and is part of a larger project at LSCE/IPSL ENS de Lyon with multiple collaborators working on rare event algorithms.

## ACKNOWLEDGEMENT

This work was supported by the ANR grant SAMPRACE, project ANR-20-CE01-0008-01 (F. Bouchet). This work has received funding through the ACADEMICS grant of the IDEXLYON, project of the Université de Lyon, PIA operated by ANR-16-IDEX-0005. We acknowledge CBP IT test platform (ENS de Lyon, France) for ML facilities and GPU devices, operating the SIDUS solution E. Quemener (2014). This work was granted access to the HPC resources of CINES under the DARI allocations A0050110575, A0070110575, A0090110575 and A0110110575 made by GENCI. We acknowledge the help of Alessandro Lovo in maintaining the GitHub page.

## REFERENCES

- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R. (2011). The hot summer of 2010: redrawing the temperature record map of europe. *Science* 332, 220–224. doi:10.1126/science.1201224
- Blackmon, M. L. (1976). A Climatological Spectral Study of the 500 mb Geopotential Height of the Northern Hemisphere. *Journal of Atmospheric Sciences* 33, 1607–1623. doi:10.1175/1520-0469(1976)033<1607:ACSSOT>2.0.CO;2
- Cassou, C., Terray, L., and Phillips, A. S. (2005). Tropical atlantic influence on european heat waves. *Journal of Climate* 18, 2805 – 2811. doi:10.1175/JCLI3506.1
- Dell'Aquila, A., Lucarini, V., Ruti, P. M., and Calmanti, S. (2005). Hayashi spectra of the northern hemisphere mid-latitude atmospheric variability in the NCEP–NCAR and ECMWF reanalyses. *Climate Dynamics* 25, 639–652. doi:10.1007/s00382-005-0048-x
- E. Quemener, M. (2014). "SIDUS", the solution for extreme deduplication of an operating system. *The Linux Journal*
- Fraedrich, K., Jansen, H., Kirk, E., Luksch, U., and Lunkeit, F. (2005). The planet simulator: Towards a user friendly model. *Meteorologische Zeitschrift* 14, 299–304. doi:10.1127/0941-2948/2005/0043
- Fraedrich, K., Kirk, E., and Lunkeit, F. (1998). Puma: Portable university model of the atmosphere. *Deutsches Klimarechenzentrum*, 38
- Gálfy, V. M. and Lucarini, V. (2021). Fingerprinting heatwaves and cold spells and assessing their response to climate change using large deviation theory. *Phys. Rev. Lett.* 127, 058701. doi:10.1103/PhysRevLett.127.058701
- Gálfy, V. M., Lucarini, V., Ragone, F., and Wouters, J. (2021). Applications of large deviation theory in geophysical fluid dynamics and climate science. *La Rivista del Nuovo Cimento* 44, 291–363
- Gálfy, V. M., Lucarini, V., and Wouters, J. (2019). A large deviation theory-based analysis of heat waves and cold spells in a simplified model of the general circulation of the atmosphere. *Journal of Statistical Mechanics: Theory and Experiment* 2019, 033404. doi:10.1088/1742-5468/ab02e8
- García-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J., and Fischer, E. M. (2010). A review of the european summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology* 40, 267–306. doi:10.1080/10643380802238137
- Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, vol. 3 (springer Berlin)
- Hayashi, Y. (1971). A generalized method of resolving disturbances into progressive and retrogressive waves by space fourier and time cross-spectral analyses. *Journal of the Meteorological Society of Japan. Ser. II* 49, 125–128. doi:10.2151/jmsj1965.49.2\_125
- Hayashi, Y. (1979). A generalized method of resolving transient disturbances into standing and traveling waves by space-time spectral analysis. *Journal of Atmospheric Sciences* 36, 1017 – 1029. doi:10.1175/1520-0469
- Herbert, C. and Bouchet, F. (2017). Predictability of escape for a stochastic saddle-node bifurcation: when rare events are typical. *Phys. Rev. E* 96, 030201(R). doi:10.1103/BF01106788
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049. doi:<https://doi.org/10.1002/qj.3803>
- Horton, R. M., Mankin, J. S., Lesk, C., Coffel, E., and Raymond, C. (2016). A review of recent advances in research on extreme heat events. *Current Climate Change Reports* 2, 242–259. doi:10.1007/s40641-016-0042-x

- Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (01 Sep. 2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society* 94, 1339 – 1360. doi:10.1175/BAMS-D-12-00121.1
- IPCC (2014). *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press). doi:10.1017/CBO9781107415324
- Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., and Bouchet, F. (2022). Deep learning-based extreme heatwave forecast. *Frontiers in Climate* 4. doi:10.3389/fclim.2022.789641
- Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J. F., Lehmann, J., et al. (2020). Amplified rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nature Climate Change* 10, 48–53. doi:10.1038/s41558-019-0637-z
- Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S., et al. (2019). Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environmental Research Letters* 14, 054002. doi:10.1088/1748-9326/ab13bf
- Lau, W. K. M. and Kim, K.-M. (2012). The 2010 pakistan flood and russian heat wave: Teleconnection of hydrometeorological extremes. *Journal of Hydrometeorology* 13, 392 – 403. doi:10.1175/JHM-D-11-016.1
- Lestang, T., Ragone, F., Bréhier, C.-E., Herbert, C., and Bouchet, F. (2018). Computing return times or return periods with rare event algorithms. *Journal of Statistical Mechanics: Theory and Experiment* 2018, 043213. doi:10.1088/1742-5468/aab856
- Miloshevich, G., Cozian, B., Abry, P., Borgnat, P., and Bouchet, F. (2023). Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. *Phys. Rev. Fluids* 8, 040501. doi:10.1103/PhysRevFluids.8.040501
- Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). Era5-land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* 13, 4349–4383. doi:10.5194/essd-13-4349-2021
- Otto, F., Massey, N., Van Oldenborgh, G. J., Jones, R., and Allen, M. (2012). Reconciling two approaches to attribution of the 2010 russian heat wave. *Geophysical Research Letters* 39, 4702–. doi:10.1029/2011GL050422
- Perkins, S. E. (2015). A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research* 164-165, 242–267. doi:<https://doi.org/10.1016/j.atmosres.2015.05.014>
- Petoukhov, V., Petri, S., Kornhuber, K., Thonicke, K., Coumou, D., and Schellnhuber, H. J. (2018). Alberta wildfire 2016: Apt contribution from anomalous planetary wave dynamics. *Scientific Reports* 8, 12375. doi:10.1038/s41598-018-30812-z
- Petoukhov, V., Petri, S., Rahmstorf, S., Coumou, D., Kornhuber, K., and Schellnhuber, H. J. (2016). Role of quasiresonant planetary wave dynamics in recent boreal spring-to-autumn extreme events. *Proceedings of the National Academy of Sciences* 113, 6862–6867. doi:10.1073/pnas.1606300113
- Petoukhov, V., Rahmstorf, S., Petri, S., and Schellnhuber, H. J. (2013). Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. *Proceedings of the National Academy of Sciences* 110, 5336–5341. doi:10.1073/pnas.1222000110
- Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography* 6, 177–203. doi:10.5194/ascmo-6-177-2020. Publisher: Copernicus GmbH

- Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vautard, R., et al. (2021). Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the us and canada june 2021. *Earth System Dynamics Discussions* 2021, 1–34. doi:10.5194/esd-2021-90
- Pratt, R. W. (1976). The interpretation of space-time spectral quantities. *Journal of Atmospheric Sciences* 33, 1060–1066. doi:[https://doi.org/10.1175/1520-0469\(1976\)033<1060:TIOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<1060:TIOTS>2.0.CO;2)
- Ragone, F. and Bouchet, F. (2020). Computation of extreme values of time averaged observables in climate models with large deviation techniques. *Journal of Statistical Physics* 179, 1637–1665. doi:10.1007/s10955-019-02429-7
- Ragone, F. and Bouchet, F. (2021). Rare event algorithm study of extreme warm summers and heatwaves over europe. *Geophysical Research Letters* 48, e2020GL091197. doi:<https://doi.org/10.1029/2020GL091197> E2020GL091197 2020GL091197
- 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, 24–29. doi:10.1073/pnas.1712645115
- Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F., and Coumou, D. (2022). Accelerated western european heatwave trends linked to more-persistent double jets over eurasia. *Nature Communications* 13, 3851. doi:10.1038/s41467-022-31432-y
- Schiermeier, Q. (2021). Climate change made north america's deadly heatwave 150 times more likely. *Nature* doi:10.1038/d41586-021-01869-0. Bandiera\_abtest: a Cg\_type: News Publisher: Nature Publishing Group Subject\_term: Climate change, Climate sciences
- Schubert, S., Wang, H., and Suarez, M. (2011). Warm season subseasonal variability and climate extremes in the northern hemisphere: The role of stationary rossby waves. *Journal of Climate* 24, 4773 – 4792. doi:10.1175/JCLI-D-10-05035.1
- Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., et al. (2021). *Weather and Climate Extreme Events in a Changing Climate* (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press). 1513–1766. doi:10.1017/9781009157896.013
- Stéfanon, M., D'Andrea, F., and Drobinski, P. (2012). Heatwave classification over europe and the mediterranean region. *Environmental Research Letters* 7, 014023. doi:10.1088/1748-9326/7/1/014023
- Teng, H., Branstator, G., Wang, H., Meehl, G., and Washington, W. (2013). Probability of us heat waves affected by a subseasonal planetary wave pattern. *Nature Geoscience* 6, 1056–1061. doi:10.1038/ngeo1988
- Van Oldenborgh, G. J., Wehner, M. F., Vautard, R., Otto, F. E. L., Seneviratne, S. I., Stott, P. A., et al. (2022). Attributing and projecting heatwaves is hard: We can do better. *Earth's Future* 10, e2021EF002271. doi:10.1029/2021EF002271. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002271>
- [Dataset] World Weather Attribution (2021). Western north american extreme heat virtually impossible without human-caused climate change
- Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., et al. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment* 1, 333–347. doi:10.1038/s43017-020-0060-z