



Non-stationary time series attribution for heatwaves over Europe

Pascal Meurer, Sebastian Buschow, Svenja Szemkus, and Petra Friederichs

Institute of Geosciences, University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany

Correspondence: Pascal Meurer (pmeurer@uni-bonn.de) and Petra Friederichs (pfried@uni-bonn.de)

Received: 9 January 2026 – Revised: 21 September 2026 – Accepted: 30 September 2026 – Published: 7 October 2026

Abstract. The increasing occurrence of extreme weather events since the beginning of the 21st century has led to the development of new methods to attribute extreme events to anthropogenic climate change. The way in which the extreme event is defined has a major influence on the attribution result. A frequently overlooked aspect concerns the temporal dependence of extremes. This study presents an approach for attributing complete time series during extreme events to anthropogenic forcing. The approach is based on a non-stationary Markov process using bivariate extreme value theory to model the temporal dependence of the time series. We calculate the likelihood ratio of an observational time series from ERA5 given the distributions as estimated from CMIP6 simulations with historical natural-only and natural and anthropogenic forcing scenarios. The spatial fields are condensed by the extremal pattern index (EPI) as a compact description of spatial extremes. In addition, the study examines the extent to which attribution statements about the occurrence of extreme heat events change when the effect of the mean warming is eliminated. The resulting attribution statement provides very strong evidence for the scenario with anthropogenic drivers over Europe, especially since the beginning of the 21st century. For central and southern Europe, the influence of anthropogenic greenhouse gas emissions on heatwaves could already have been proven in the 1960s using today's knowledge. There is no reliable signal apart from a general shift in the temperature distribution, neither in terms of the temporal dependence of extreme heat days nor in terms of the shape of the extreme value distribution.

1 Introduction

The considerable increase in greenhouse gases resulting from the extensive use of fossil energy sources by humans, coupled with the substantial alterations in land use, has led to the emergence of an ongoing global climate change phenomenon, which began approximately a century ago (Gulev et al., 2021). In recent years, annual greenhouse gas emissions have reached unprecedented levels (IEA, 2021). This trend is mirrored by the concentrations of CO₂, CH₄ and N₂O, which have reached levels that have not been seen for at least 800 000 years (Gulev et al., 2021). The Earth's climate system is undergoing a transformation as a consequence of alterations in atmospheric composition. In the 1980s and 1990s, scientific interest therefore focused on detecting observed changes in the climate system (e.g. the increase in global mean temperature). These detection studies have already demonstrated the strong signal of anthropogenic cli-

mate change in observed changes in global mean surface temperature (Hegerl et al., 2006).

In addition to the increase in global mean temperature, the occurrence of extreme events such as heatwaves has increased since the beginning of the 21st century. This has led to a growing number of studies on the attribution of extreme weather events (Hulme, 2014; Seneviratne et al., 2021). One of the first studies of this kind was conducted by Stott et al. (2004). Such studies aim to investigate whether changes in extreme weather can be attributed to anthropogenic climate change. Several methods for the attribution of extreme events have been established that focus the attribution statement on a single extreme event, such as a heatwave or flood event (Perkins-Kirkpatrick et al., 2024). Typically, a data compression method (e.g., average over a region) is used to attribute the event based on univariate extreme value statistics (Philip et al., 2020). The attribution is then performed by calculating

the probability of the event given factual and counterfactual conditions, based on the so-called causal counterfactual theory (Hannart et al., 2016). These probabilities are then compared using a probability ratio, or in the case of a time series the likelihood ratio (Seong et al., 2022).

The temporal evolution and dependence are often ignored, or techniques are used to break down the time series into individual clusters of extreme values (Philip et al., 2020; Sippel et al., 2015; Wehner et al., 2016). A useful extension in extreme value statistics is the modelling of temporal dependence for extreme events (Fawcett and Walshaw, 2006). We use this approach to include temporal dependence in the attribution approach. The question of attribution for periods without extreme events is also relevant. For example, one may consider the attribution of a summer with cold spells and heat extremes. The European summer of 2025, for example, included cold spells as well as heat extremes (see Sect. 2.2.1). In this study, we present an approach for attributing complete time series that consist of extreme and non-extreme events. This allows us to extend the attribution not only to entire summer periods, but also to a series of summers. One can then assess whether non-extreme parts of the time series, in which a high threshold is not exceeded, contribute to evidence for the greenhouse gas scenario, and if not, how strong the opposing evidence is. In this sense, the likelihood ratio can be updated by adding additional observations to the attribution.

Since we want to process data with spatial and temporal dimensions and obtain reliable probabilities for extreme events, the first step is to reduce the dimensions in space using a suitable data compression method. We use a method presented by Szemkus and Friederichs (2024) based on a decomposition of extremal dependence. Their approach is closely related to principal component analysis, as used, for example, to identify teleconnections in the atmosphere, and follows Cooley and Thibaud (2019) and Jiang et al. (2020). Instead of correlation or covariance, a measure specifically designed for the dependence of extremes is used to identify spatially coherent patterns. The extremal pattern index (EPI) as introduced in Szemkus and Friederichs (2024) then provides a spatially aggregated measure of the strength of an extremal pattern of a meteorological variable.

We can use the standard peaks-over-threshold (POT) approach to model extreme values of the EPI above a high threshold. In this approach, threshold exceedances follow a generalized Pareto distribution (GPD). As multiple threshold exceedances can occur in succession, the times of their occurrence cannot be treated as independent events. Therefore, a common strategy is to decluster the time series and retain only the cluster maximum. Although this reduces serial dependence among exceedances, it introduces additional choices; for example, the declustering rule (Beirlant et al., 2004, Sect. 10.4). If a general warming trend is present, the number of threshold exceedances and, therefore, the clustering, will change. In this case, the clustering must also be

made time-dependent; that is to say, different cluster lengths in different years must be considered. Alternatively, we can model the time-dependence structure directly. This enables us to model the temporal dependence and consider a changing heatwave definition by increasing the threshold.

The temporal dependence of the daily EPI time series is considered using a Markov process. The Markov process can be described by an approximate likelihood based on bivariate extreme value theory (Smith et al., 1997; Beirlant et al., 2004). This likelihood is based on the censored likelihood model, which divides the two-dimensional plane into four regions according to whether or not the variables exceed the threshold value. The corresponding two-dimensional variable represents two consecutive observations in the Markov process. The resulting dependence is thus analogous to the autocorrelation in a time series. Declustering the extremes in a time series is therefore no longer necessary.

Here, we focus on a parametric model of temporal extremal dependence formulated within the class of asymptotically dependent bivariate extreme-value models. This assumption is consistent with the pre-asymptotic nature of the considered data, in which exceedances occur at high but finite thresholds. However, recent studies suggested models that can also represent asymptotic independence, in which extremal dependence vanishes in the limit but may still be present at finite levels (e.g., Ledford and Tawn, 1997; Ramos and Ledford, 2009; Wadsworth et al., 2017). We do not pursue this extension here, but we regard it as an important area for future research.

The approach is applied to different scenarios, represented by a multi-model ensemble of CMIP6 simulations, so that we can compare the likelihood of an observational time series given different scenarios. For each scenario and climate model, a set of model parameters can be estimated, and an attribution statement can be made using the respective likelihood ratio.

This study examines heatwaves, which are described by the maximum daily temperature near the surface. We would like to answer the following questions in particular.

1. Can [changes in] the temporal evolution of a heatwave be attributed to anthropogenic emissions?
2. Is there a climate change signal in the tail behaviour of heatwave extremes beyond a general shift in the temperature distribution?

For both questions, we adapt the Markov process model further to account for non-stationary conditions and ongoing changes in the climate system, following Hannart et al. (2016). They pointed out that the stationarity assumption is unrealistic because the mean temperature and thus the extremes have changed over time. Based on slowly varying covariates, possible changes in the parameters can be modelled.

The article is structured as follows. Section 2 introduces the data used in this study, as well as the compact descrip-

Table 1. Number and type of CMIP6 simulations with HIST (1940–2014), SSP2-4.5 (2015–2025), and HIST-NAT (1940–2020) for T2max used in this study.

Number and type of CMIP6 simulations			
Variable	T2max	T2max	T2max
Scenario	HIST	SSP2-4.5	HIST-NAT
ACCESS-ESM1-5	19	18	3
BCC-CSM2-MR	3	1	3
CanESM5	50	50	10
CNRM-CM6-1	30	1	3
HadGEM3-GC31-LL	4	4	4
IPSL-CM6A-LR	32	11	6
MIROC6	50	49	50
MPI-ESM1-2-LR	50	50	30
MRI-ESM2-0	7	4	5

tion of multivariate weather events in terms of EPI. The daily maximum surface temperatures in Europe in the summer of 2025 are used to illustrate the methods. In Sect. 3 we briefly introduce the attribution approach for extreme events, present the derivation of the approximate likelihood for the Markov process, and introduce non-stationarity. In Sect. 4, the workflow for the attribution is presented using the Markov process model. In Sect. 5, the results for heatwaves are presented. Lastly, Sect. 6 discusses these results and gives our concluding remarks.

2 Data and Indices

2.1 Data

We used the daily maximum 2 m temperature (T2max) of the ERA5 reanalysis (Hersbach et al., 2020) in Europe. ERA5 provides a spatially and temporally consistent description of the atmospheric state since 1940 based on data assimilation. We focus our study on daily time series during the northern hemisphere summer seasons June to August (JJA).

The climate simulations are taken from the CMIP6 and DAMIP v1.0 projects (Gillett et al., 2016). The scenarios we use are historical-natural (HIST-NAT) without anthropogenic forcing, the historical scenario (HIST) and SSP2-4.5 with anthropogenic forcing. The respective circulation model, the variable, and the number of members available in each ensemble are given in Table 1. As for the ERA5 dataset, we focus on JJA months. The study area consists of only land points. A grid cell is considered a land point when at least 50 % of the cell is occupied by land. We used the IPCC AR6 regions shown in Fig. 1 of northern, central, and southern Europe (Iturbide et al., 2020). The regions in North Africa are excluded from the analysis due to their different climates with extremely dry desert areas.

2.2 Extremal pattern index (EPI)

Our compact representation of a spatially extended multivariate weather event uses the EPI as suggested in Szemkus and Friederichs (2024). The approach is based on the decomposition of high-dimensional data based on a description of the tail dependence within the framework of regular variation (Cooley and Thibaud, 2019). The reader is referred to Szemkus and Friederichs (2024) for the details of the method.

We consider a data set containing daily values. The annual cycle is initially removed from the daily data set during JJA. For this purpose, the T2max values for each day of the JJA period and grid point are standardised using the respective mean value and standard deviation. Subsequently, the data at each grid point are transformed to a standard distribution, since the estimator of the extremal dependence is based on the assumption that the margins are Fréchet distributed. In our case, this is a Fréchet distribution with tail index 2. Transforming the data to Fréchet can be understood as giving little weight to small and negative temperature anomalies, and most weight to larger positive temperature anomalies. Figure 2 illustrates this effect.

The pairwise measure of dependence between extreme events at two locations is calculated for each pair of land grid points in Europe. The result is a matrix of extremal dependencies, referred to as the tail pairwise dependence matrix (TPDM), with dimensions $q \times q$, where q is the number of grid points over land. In the original paper that defined the principal component analysis (PCA) for extremes (Cooley and Thibaud, 2019), the same threshold was applied to all data pairs, which guarantees positive definiteness of the TPDM. We use another estimator as advised by Jiang et al. (2020), who suggest using an individual threshold for each pair of data. This estimator of the TPDM is not guaranteed to be positive definite; therefore, the positive definite matrix closest to the estimated one is used, which guarantees non-negative eigenvalues (Szemkus and Friederichs, 2024, Sect. 2.2). The TPDM's eigenvalue decomposition then yields the eigenvectors and eigenvalues. The eigendecomposition procedure is similar to that of the standard PCA, since the TPDM satisfies all necessary properties, namely it is symmetric, has non-negative entries, and is positive semi-definite. The eigenvectors of the TPDM represent spatial variation patterns, indicating regions where extremes occur together with greater frequency.

The first mode has a non-zero spatial mean and values that remain nearly constant across space. The higher-order eigenvectors represent large-scale patterns associated with typical dipole and multipole structures, much like in standard PCA. In this way, we utilise the properties of the PCA, which favours large-scale patterns, to identify heat events that exhibit a correspondingly large spatial extent. The eigenvalues λ_k , $k = 1, \dots, q$, indicate the proportion of pattern k within the total variability of the standardised extreme

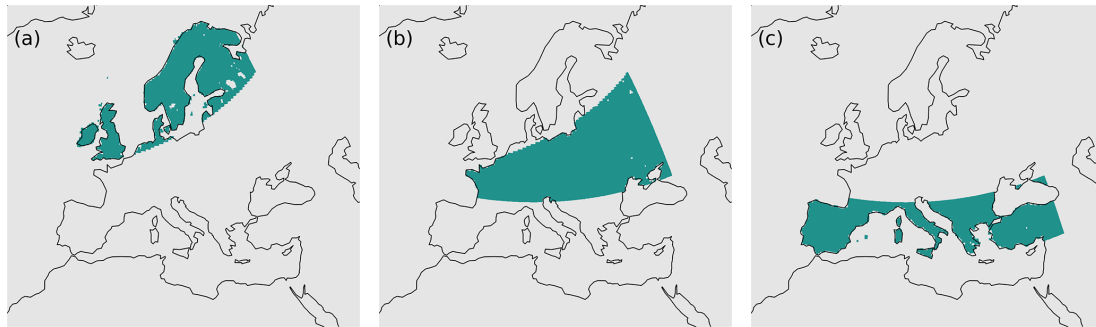


Figure 1. AR6 regions (Iturbide et al., 2020) of (a) northern, (b) central, and (c) southern Europe as used in this study with North Africa excluded and restricted to land points only. A grid cell is considered a land point when at least 50 % of the cell is occupied by land.

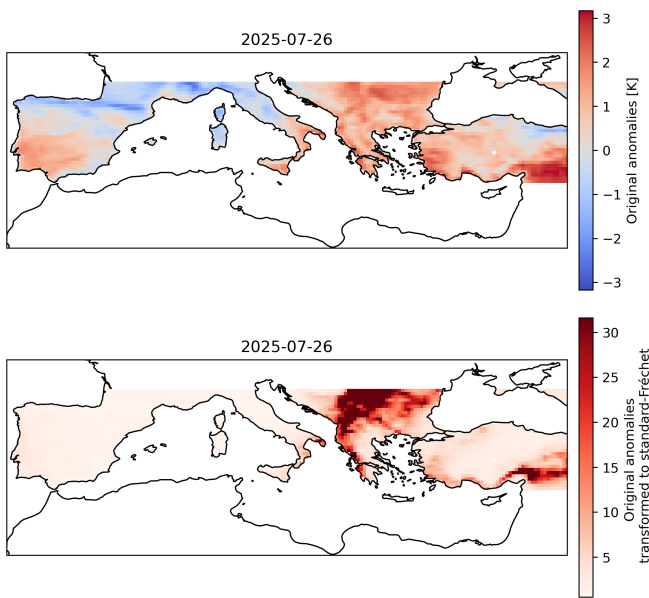


Figure 2. The effect of transforming temperature anomalies to a standard-Fréchet distribution using the ERA5 T2max on 26 July 2025.

events. The projection of the k th pattern onto the Fréchet-standardised data yields the principal component (PC) $\eta_{k,i,d}$, i.e. the strength of the pattern k on day d of the season of year i . Since the eigenvectors are projected onto the Fréchet-standardised data rather than the original anomalies, the construction of the TPDM and the projection steps are based on the same data.

The TPDM eigenvectors span an “extreme” subspace that, based on our experience, can be adequately covered with $n_p = 10$ extremal patterns. For example, these n_p extremal patterns of ERA5 T2max in the southern European region account for about 75 % of the total variance in the Fréchet-standardised data. If the temperature field does not exhibit any large-scale positive temperature extremes, it is projected onto this subspace only to a limited extent. Small-scale extremes are intentionally not captured. The total variability of

the first n_p extremal patterns is summarised in the extremal pattern index (EPI, Szemkus and Friederichs, 2024). The EPI on day d of the summer season of year i is defined as

$$\text{EPI}_{i,d} = \frac{\sqrt{\sum_{k=1}^{n_p} \eta_{k,i,d}^2}}{\sqrt{\sum_{k=1}^{n_p} \lambda_k}}. \quad (1)$$

The EPI represents a spatial aggregation of the extremal state of the variable in space (Szemkus and Friederichs, 2024). Normalisation in the definition of the EPI using the sum of the eigenvalues in Eq. (1) allows the comparison of data on different spatial grids or of EPIs computed using different numbers of patterns. The EPI allows multivariate extreme events with spatial extent to be described by a univariate index that has only a temporal dimension and accounts for spatial dependencies at extreme levels. In this study, the method is applied to different AR6 regions (Fig. 1), i.e., the TPDM is calculated and decomposed separately for each region. Throughout the remainder of this paper we write $y_{i,d} := \text{EPI}_{i,d}$; the EPI time series is thus the quantity to which the extreme value model of Sect. 3 is applied.

2.2.1 Example: Summer 2025 in Europe

The European summer of 2025 was characterised by a series of heatwave events across Europe. An early heatwave began in southwestern Europe in June, and the highest average surface air temperatures between 17 June and 2 July 2025 compared to the period 1979–2024 were recorded on the Iberian Peninsula and in large parts of France, extending to Germany and southern Great Britain, as reported in Copernicus Climate Change Service (2025). Figure 3 shows the EPI for the regions in Fig. 1 and the T2max anomalies for summer 2025 over Europe.

The early onset of the remarkable heatwave over southwestern Europe is visible at the end of the first week of June (Fig. 3). This event lasted until the beginning of July. As is common in this region, this heatwave was mainly caused by atmospheric ridges, as is evident from the large-scale circulation pattern during the heatwave (Faranda et al., 2025). Re-

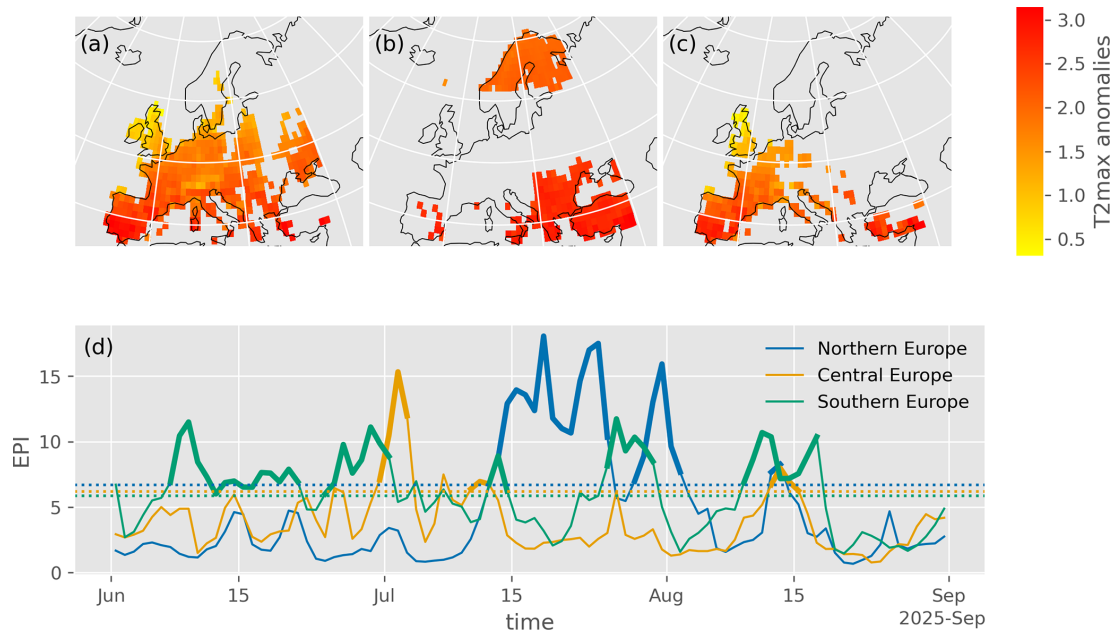


Figure 3. Mean standardised T2max anomalies in ERA5 between (a) 7 June to 5 July 2025, (b) 15 July to 1 August 2025, and (c) 7 August to 19 August 2025. Only grid points exceeding the 99 % quantile are shown. In (d) the EPI for T2max from June to August 2025 for northern (blue), central (orange) and southern (green) European regions from Fig. 1 is shown. The EPI is bold if it exceeds the corresponding 95 % quantile (1940–2025) of the region, with the corresponding quantile indicated by a dotted line.

garding the spatial extent, most parts of the southern and central European regions except Turkey were influenced by this event. Two additional strong heatwave events were observed in southern Europe, one in the second half of July and one in the first half of August. The event in July occurred over the eastern Mediterranean, while the event in August took place over the western Mediterranean, affecting Spain, Portugal, and France.

Another remarkable heatwave in 2025 occurred over Scandinavia, from approximately mid-July to early August. This event received considerable media attention and was investigated by the World Weather Attribution in a report (Barnes et al., 2025). A detailed description of the synoptic situation and the resulting temperature extremes can be found in the report. A detailed scientific analysis of the European summer of 2025, including the characterisation of heat events, their impact, and the influence of climate warming, can be found in the ClimXtreme report by Lemburg et al. (2026).

3 Theory and Methods

3.1 Attribution of extreme events

The attribution of extreme events corresponds to the statistical answer to the question of whether or not a specific event was caused by anthropogenic climate change. The theoretical foundation of the attribution is presented in detail by Hannart and Naveau (2018) and is based on causal counterfactual theory.

We define m_0 as the scenario without anthropogenic forcing (HIST-NAT) and m_1 as the scenario with climate change (HIST). The probability ratio is then the ratio of the probability of an event under HIST to the probability of an event under HIST-NAT, given by

$$\text{PR} = \frac{p_1}{p_0}. \quad (2)$$

In this study, we model time series rather than single events. We therefore consider the joint density of a set of time series $\mathbf{Y} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)})$ under the two scenarios HIST and HIST-NAT, where the vector $\mathbf{y}^{(i)}$ denotes the daily evolution of the EPI during the season of year i . The probability ratio then becomes a likelihood ratio (LR):

$$\text{LR}(\mathbf{Y}) = \frac{l(\mathbf{Y} | m_1)}{l(\mathbf{Y} | m_0)}, \quad (3)$$

where $l(\mathbf{Y} | m_s)$ is the likelihood of observing $\mathbf{Y}$ under scenario m_s , $s \in \{0, 1\}$.

From a Bayesian perspective (Min et al., 2004; Seong et al., 2022), attribution can instead be stated as the ratio of the posterior odds for the scenarios (Kass and Raftery, 1995). Here, we treat m_0 and m_1 as two competing scenario-conditioned generative statistical models for the observed ERA5 EPI time series. We assume that $(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)})$ are independent vectors: the daily time series of one season is independent of the daily time series of the same season in the year before or after. Using Bayes' theorem, the posterior

Table 2. Scales of the Bayes factor and related evidence as given in Seong et al. (2022). Here and throughout, log denotes the natural logarithm.

log BF	BF	Evidence against m_0
0–1	1–3	Not worth more than a bare mention
1–2.5	3–12	Substantial
2.5–5	12–150	Strong
> 5	> 150	Decisive

odds can be written as

$$\begin{aligned} \frac{\mathbb{P}(m_1 | \mathbf{Y})}{\mathbb{P}(m_0 | \mathbf{Y})} &= \frac{l(\mathbf{Y} | m_1)}{l(\mathbf{Y} | m_0)} \cdot \frac{\mathbb{P}(m_1)}{\mathbb{P}(m_0)} \\ &= \prod_{i=1}^{n_y} \frac{l(\mathbf{y}^{(i)} | m_1)}{l(\mathbf{y}^{(i)} | m_0)} \cdot \frac{\mathbb{P}(m_1)}{\mathbb{P}(m_0)} \\ &= \underbrace{\prod_{i=1}^{n_y} \text{LR}(\mathbf{y}^{(i)})}_{=: \text{BF}} \cdot \frac{\mathbb{P}(m_1)}{\mathbb{P}(m_0)}. \end{aligned} \quad (4)$$

The likelihood ratio $\text{LR}(\mathbf{y}^{(i)}) = l(\mathbf{y}^{(i)} | m_1) / l(\mathbf{y}^{(i)} | m_0)$ represents the additional evidence contributed by the seasonal time series of the year i to the total evidence. Under uniform prior probabilities, the posterior ratio equals the product of the likelihood ratios, which corresponds to the Bayes factor (BF). By varying the most recent analysis year from 1940 to 2025, we can assess when in the past under today's knowledge would have provided sufficient evidence of climate change.

A statement of evidence based on the Bayes factor is given in Table 2, which differs slightly from that given in Kass and Raftery (1995). Given that interpretation is context-dependent, Table 2 provides a general indication of the level of evidence.

3.2 Likelihood formulation

3.2.1 Marginal formulation

Our aim is to model the extremes of the EPI time series, as illustrated in Fig. 3 by the bold sections of the respective time series that exceed the corresponding threshold. When modelling extremes over a high threshold, we use a common peaks-over-threshold (POT) approach, in which threshold exceedances follow a generalized Pareto distribution (GPD).

We write the full data set as

$$\mathbf{Y} = \{y_{i,d}; i = 1, \dots, n_y, d = 1, \dots, n\}, \quad (5)$$

where $y_{i,d}$ is the EPI on day d of the summer season of year i , n_y is the number of JJA seasons and $n = 92$ the number of days per season. The time series of a single season is $\mathbf{y}^{(i)} = (y_{i,1}, \dots, y_{i,n})$, so that $\mathbf{Y} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)})$ as in

Sect. 3.1. Season $i = 1$ corresponds to the summer of 1940 and season $i = n_y$ to the summer of 2025. An overview of the notation used throughout this paper is provided in Tables A1 and A2 in Appendix A5.

Let $u_{i,d}$ denote a sufficiently high threshold for observation d in year i . Conditional on exceeding $u_{i,d}$, the excess $y_{i,d} - u_{i,d}$ is modelled by a GPD with the scale parameter $\sigma_{i,d} > 0$ and the shape parameter $\xi_{i,d}$. Together with the exceedance probability $\phi_{i,d} = \mathbb{P}(y_{i,d} > u_{i,d})$, the marginal distribution function reads (Coles, 2001, Sect. 4.3)

$$F(y_{i,d}) \approx \begin{cases} 1 - \phi_{i,d}, & y_{i,d} \leq u_{i,d}, \\ 1 - \phi_{i,d} \left(1 + \xi_{i,d} \frac{y_{i,d} - u_{i,d}}{\sigma_{i,d}} \right)_+^{-1/\xi_{i,d}}, & y_{i,d} > u_{i,d}. \end{cases} \quad (6)$$

Thus $\phi_{i,d}$ is the probability that observation d in year i exceeds the threshold $u_{i,d}$; below the threshold the distribution is treated in censored form, so that only the probability of non-exceedance enters the model. Here F always denotes a univariate marginal distribution function; the joint distribution of two consecutive observations is denoted by G in Sect. 3.2.2. At this stage, all four quantities are allowed to depend on both the year and the day; the assumptions that reduce this dependence are stated explicitly in Sect. 3.3.

3.2.2 Dependence formulation

In practice, extreme heat often persists for several consecutive days. Consequently, several threshold exceedances may occur in sequence and cannot be treated as independent events. Since the temporal dependence is inherent in the data, we aim to model it explicitly, as the duration of a heatwave is a key factor in determining its impact. For alternative approaches to modelling the temporal dependence of a time series, we refer the reader to Beirlant et al. (2004).

We model the temporal dependence of the threshold exceedances as a first-order Markov process: the seasons are independent across the year index i , while within a season the values $y_{i,1}, \dots, y_{i,n}$ form a Markov process. The first-order Markov property implies that the conditional density of day d given all previous days reduces to the density conditioned on the previous day only, $l(y_{i,d} | y_{i,1}, \dots, y_{i,d-1}; \boldsymbol{\theta}) = l(y_{i,d} | y_{i,d-1}; \boldsymbol{\theta})$. Writing l_1 and l_2 for the univariate and bivariate densities implied by the model, the likelihood of the season i is

$$\begin{aligned} l(\mathbf{y}^{(i)}; \boldsymbol{\theta}) &= l_1(y_{i,1}; \boldsymbol{\theta}_1) \prod_{d=2}^n l(y_{i,d} | y_{i,d-1}; \boldsymbol{\theta}) \\ &= l_1(y_{i,1}; \boldsymbol{\theta}_1) \prod_{d=2}^n \frac{l_2(y_{i,d-1}, y_{i,d}; \boldsymbol{\theta})}{l_1(y_{i,d-1}; \boldsymbol{\theta}_1)} \\ &= \frac{\prod_{d=1}^{n-1} l_2(y_{i,d}, y_{i,d+1}; \boldsymbol{\theta})}{\prod_{d=2}^{n-1} l_1(y_{i,d}; \boldsymbol{\theta}_1)}, \end{aligned} \quad (7)$$

and the likelihood of the complete data set is

$$l(\mathbf{Y}; \boldsymbol{\theta}) = \prod_{i=1}^{n_y} l(y^{(i)}; \boldsymbol{\theta}). \quad (8)$$

Here $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_1$ the marginal parameters of Sect. 3.2.1 and $\boldsymbol{\theta}_2$ the possibly infinite set of parameters of the temporal extremal dependence. Due to the first-order assumption, the dependence is fully described by the joint distribution of pairs of consecutive days $(y_{i,d}, y_{i,d+1})$.

We use bivariate extreme value theory for the dependence formulation. Unlike the univariate case, multivariate extreme value theory does not provide a single, complete parametric model for the dependence structure. Instead, extremal dependence can be described either non-parametrically or by assuming a parametric family for the stable tail dependence function. In this study, we focus on a parametric model of temporal extremal dependence within the class of asymptotically dependent bivariate extreme-value models. This assumption is consistent with the pre-asymptotic nature of the data under consideration, in which exceedances occur above a high but finite threshold.

Consider two consecutive observations $(y_{i,d}, y_{i,d+1})$ with a joint distribution function G , and assume that G lies in the domain of attraction of a bivariate extreme value distribution. Before specifying the dependence structure, the margins are transformed to standard Fréchet.

The transformation to standard Fréchet is directly linked to the GPD formulation in Eq. (6), given by

$$v_{i,d} = \phi_{i,d} \left(1 + \xi_{i,d} \frac{y_{i,d} - u_{i,d}}{\sigma_{i,d}} \right)_+^{-1/\xi_{i,d}}, \quad (9)$$

so that $F(y_{i,d}) \approx 1 - v_{i,d}$ and $1/v_{i,d}$ is standard Fréchet distributed (Beirlant et al., 2004, Sect. 10.4). Note that $v_{i,d} = \phi_{i,d}$ whenever $y_{i,d} = u_{i,d}$, which will be used in Eq. (13) below. This separate modelling of the marginal distribution and extremal dependence corresponds to the approach used in deriving the TPDM in Sect. 2.2.

The joint distribution of two consecutive observations above high thresholds is then approximated by

$$G(y_{i,d}, y_{i,d+1}) \approx \exp \{ -V(v_{i,d}, v_{i,d+1}) \}, \quad (10)$$

where V denotes the stable tail dependence function. Since the marginal behaviour has already been absorbed into the transformation to $v_{i,d}$ and $v_{i,d+1}$, the function V describes the extremal dependence between two consecutive observations. Its significance lies in the fact that a bivariate extreme value distribution is completely determined by its margins and V . For the function V to be a stable tail dependence function in Eq. (10), it must fulfil certain necessary conditions. The stable tail dependence function restricted to the unit simplex is also known as the more popular Pickands dependence function, originally introduced by Pickands (1981), see Beirlant et al. (2004, Sect. 8.2.5) for details.

In practice, the function V can be expressed in non-parametric or parametric form. Our focus is on parametric families of extremal dependence, which can be conveniently specified through V . The logistic model introduced in the following is an example of this. The logistic model as the simplest parametric choice of the function V is given as (Beirlant et al., 2004, Sect. 10.4.5),

$$V(v_{i,d}, v_{i,d+1}) = \left(v_{i,d}^{1/\alpha_{i,d}} + v_{i,d+1}^{1/\alpha_{i,d}} \right)^{\alpha_{i,d}}, \quad v_{i,d}, v_{i,d+1} \geq 0, \quad (11)$$

where $\alpha_{i,d} \in (0, 1]$ is the single dependence parameter attached to the transition from day d to day $d+1$ in year i . The case $\alpha_{i,d} = 1$ corresponds to independence, whereas $\alpha_{i,d} \rightarrow 0$ corresponds to complete asymptotic dependence. With Eqs. (9) and (11), the bivariate model in Eq. (10) is fully determined.

More flexible parametric choices could also be used, such as asymmetric logistic or negative logistic models, as well as non-parametric descriptions of the stable tail dependence function. For alternative choices of V , the reader is referred to Beirlant et al. (2004, Sect. 9.2). In this study, the logistic model is used because it provides the most straightforward parametric description of the temporal extremal dependence, requiring only one dependence parameter.

Max-stable models, such as the logistic model, cannot capture asymptotic independence (Huser et al., 2021, Sect. 2.3). They can represent either complete independence or asymptotic dependence. Recent studies emphasise the importance of models that can represent both asymptotic dependence and independence (e.g., Ledford and Tawn, 1997; Bortot and Tawn, 1998; Ramos and Ledford, 2009, 2013; Wadsworth et al., 2017). We will not use this extension here, but we consider it an interesting avenue for future studies.

3.2.3 Likelihood inference

For the likelihood inference, we focus on the censored likelihood method. This model can handle all situations: when neither of the two consecutive days exceeds the threshold; when one of them exceeds the threshold; and when both exceed the threshold. An advantage of the censoring is that non-extreme observations do not affect the estimation of the extremal dependence. Depending on which of the two consecutive observations exceeds its threshold, the bivariate plane is divided into the four regions shown in Fig. 4, with the corresponding likelihood contributions

$$l_2(y_{i,d}, y_{i,d+1}; \boldsymbol{\theta}) \propto \begin{cases} G(u_{i,d}, u_{i,d+1}) & \text{if } y_{i,d} \leq u_{i,d}, \quad y_{i,d+1} \leq u_{i,d+1}, \\ \frac{\partial G}{\partial y_{i,d}}(y_{i,d}, u_{i,d+1}) & \text{if } y_{i,d} > u_{i,d}, \quad y_{i,d+1} \leq u_{i,d+1}, \\ \frac{\partial G}{\partial y_{i,d+1}}(u_{i,d}, y_{i,d+1}) & \text{if } y_{i,d} \leq u_{i,d}, \quad y_{i,d+1} > u_{i,d+1}, \\ \frac{\partial^2 G}{\partial y_{i,d} \partial y_{i,d+1}}(y_{i,d}, y_{i,d+1}) & \text{if } y_{i,d} > u_{i,d}, \quad y_{i,d+1} > u_{i,d+1}. \end{cases} \quad (12)$$

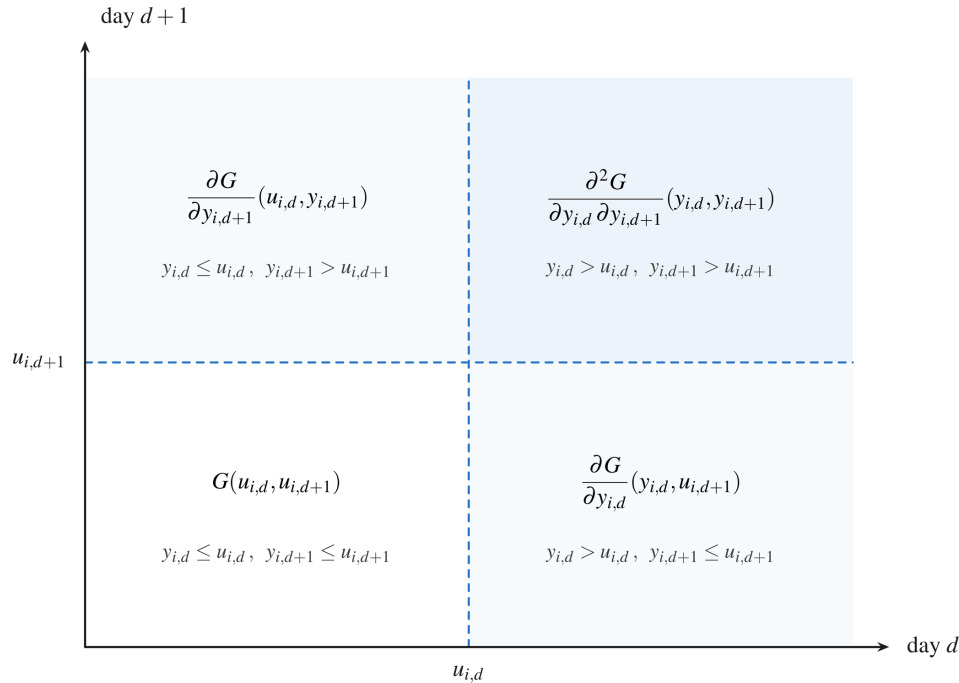


Figure 4. Regions of the censored likelihood model according to Eq. (12), where the likelihood contribution depends on which of the two variables exceeds its threshold.

Writing V_1 , V_2 and V_{12} for the partial derivatives of V with respect to its first, second and both arguments, the contributions follow from Eqs. (10) and (12) as (Beirlant et al., 2004)

$$l_2(y_{i,d}, y_{i,d+1}; \theta) \propto \begin{cases} \exp\{-V(\phi_{i,d}, \phi_{i,d+1})\} & \text{if } y_{i,d} \leq u_{i,d}, y_{i,d+1} \leq u_{i,d+1}, \\ -\frac{\partial v_{i,d}}{\partial y_{i,d}} V_1(v_{i,d}, \phi_{i,d+1}) \exp\{-V(v_{i,d}, \phi_{i,d+1})\} & \text{if } y_{i,d} > u_{i,d}, y_{i,d+1} \leq u_{i,d+1}, \\ -\frac{\partial v_{i,d+1}}{\partial y_{i,d+1}} V_2(\phi_{i,d}, v_{i,d+1}) \exp\{-V(\phi_{i,d}, v_{i,d+1})\} & \text{if } y_{i,d} \leq u_{i,d}, y_{i,d+1} > u_{i,d+1}, \\ \frac{\partial v_{i,d}}{\partial y_{i,d}} \frac{\partial v_{i,d+1}}{\partial y_{i,d+1}} [V_1 V_2 - V_{12}](v_{i,d}, v_{i,d+1}) \exp\{-V(v_{i,d}, v_{i,d+1})\} & \text{if } y_{i,d} > u_{i,d}, y_{i,d+1} > u_{i,d+1}, \end{cases} \quad (13)$$

where the first case uses $v_{i,d} = \phi_{i,d}$ when $y_{i,d} = u_{i,d}$.

Thus, every pair of consecutive time series points $l_2(y_{i,d}, y_{i,d+1}; \theta)$ must be evaluated. For the univariate likelihood $l_1(y_{i,d}; \theta_1)$, we can write

$$l_1(y_{i,d}; \theta_1) = \begin{cases} \frac{\phi_{i,d}}{\sigma_{i,d}} \left(1 + \xi_{i,d} \frac{y_{i,d} - u_{i,d}}{\sigma_{i,d}}\right)^{-1/\xi_{i,d}-1}, & \text{if } y_{i,d} > u_{i,d}, \\ 1 - \phi_{i,d}, & \text{if } y_{i,d} \leq u_{i,d}. \end{cases} \quad (14)$$

The full likelihood used for the fitting procedure is given by Eqs. (7) and (8), with the bivariate contributions of

Eq. (13) in the numerator and the univariate contributions of Eq. (14) in the denominator. Since the Markov process conditions on the previous day, the product of pairwise bivariate contributions has to be divided by the univariate ones. Because the bivariate density is required in the numerator, the distribution function must be differentiated with respect to those variables that exceed the threshold, following the censoring scheme of Fig. 4. The marginal and dependence parameters are finally obtained from

$$\hat{\theta} = \arg \min_{\theta} \{-\log l(\mathbf{Y}; \theta)\}. \quad (15)$$

3.2.4 Higher-order Markov processes

The first-order assumption of the Markov process may be questionable. To investigate the effect of this choice, we expand the analysis to include a second-order Markov process to capture higher-order dependencies on a longer timescale. To this end, we can rewrite Eq. (7) to

$$\begin{aligned} l(\mathbf{y}^{(i)}; \theta) &= l_2(y_{i,1}, y_{i,2}; \theta) \prod_{d=3}^n l(y_{i,d} | y_{i,d-1}, y_{i,d-2}; \theta) \\ &= l_2(y_{i,1}, y_{i,2}; \theta) \prod_{d=3}^n \frac{l_3(y_{i,d-2}, y_{i,d-1}, y_{i,d}; \theta)}{l_2(y_{i,d-2}, y_{i,d-1}; \theta)} \\ &= \frac{\prod_{d=1}^{n-2} l_3(y_{i,d}, y_{i,d+1}, y_{i,d+2}; \theta)}{\prod_{d=2}^{n-2} l_2(y_{i,d}, y_{i,d+1}; \theta)}, \end{aligned} \quad (16)$$

with l_3 the trivariate density. The likelihood ratio can then be calculated using this likelihood formulation.

As described in Smith et al. (1997), the estimation of the model is similar to that for a first-order Markov process, but now also with respect to the newly introduced variable (trivariate problem). In this case, the construction of Fig. 4, which is not unique to the two-dimensional case, can be extended to more than two variables.

For the symmetric logistic model, an example of a corresponding tail dependence function is given by

$$V(v_{i,d}, v_{i,d+1}, v_{i,d+2}) = \left(v_{i,d}^{1/\alpha_{i,d}} + v_{i,d+1}^{1/\alpha_{i,d}} + v_{i,d+2}^{1/\alpha_{i,d}} \right)^{\alpha_{i,d}}, \quad v_{i,d}, v_{i,d+1}, v_{i,d+2} \geq 0. \quad (17)$$

The symmetric logistic model imposes the same dependence on all pairs among three consecutive days, although the dependence between days d and $d+2$ would be expected to be weaker than between adjacent days.

However, due to the more complex likelihood structure, the fitting procedure is more costly than for a first-order Markov process, and the likelihoods of different orders are not necessarily comparable (Beirlant et al., 2004, Sect. 10.4.6). The results for higher-order Markov processes are presented in Sect. A3.

3.3 Including non-stationarity

In the general formulation of Sect. 3.2 every parameter carries both indices, i.e. $u_{i,d}$, $\phi_{i,d}$, $\sigma_{i,d}$, $\xi_{i,d}$ and $\alpha_{i,d}$. In the main part of this study we assume that the parameters vary between seasons but are constant within a season,

$$u_{i,d} \equiv u_i, \quad \phi_{i,d} \equiv \phi_i, \quad \sigma_{i,d} \equiv \sigma_i, \quad \xi_{i,d} \equiv \xi_i, \quad \alpha_{i,d} \equiv \alpha_i, \quad (18)$$

so that the day index d can be dropped. The consequences of relaxing assumption (18), i.e. of allowing for a residual seasonal cycle, are examined in Appendix A2.

To account for non-stationarity due to the general warming trend, we include slow changes over time in the marginal and dependence parameters. Since the exceedance-probability and the threshold are time-dependent, one of the two is specified a priori. We investigate two modelling approaches:

1. $u = \text{const}$, ϕ_i time-varying: the threshold is fixed a priori as the 95 % quantile, and the exceedance probability is estimated as a time-varying parameter using logistic regression (Sect. 3.3.1).
2. $\phi = \text{const}$, u_i time-varying: the exceedance probability is fixed a priori at $\phi = 0.05$, and the warming trend is absorbed into a time-varying threshold u_i estimated by quantile regression (Sect. 3.3.2).

The two modelling approaches are visualised by the constant and time-varying thresholds in Fig. 5, which is based on the ERA5 EPI in the southern European region. Both modelling approaches become visible. When the threshold

is constant (blue line), the number of threshold exceedances has increased since 2000. When the threshold is time-varying (orange line), the number of threshold exceedances remains nearly constant. The slowly varying threshold shows the cooler decades between 1960 and 1990, as well as the warming trend since the 1990s. The reasoning behind the non-stationary modelling of exceedance probability and threshold can be found in Sect. A1. In this context, we refer to Eastoe and Tawn (2009), who discuss the two different modelling approaches in more detail. The time-varying parameters are modelled through slowly varying covariates. We use low-order Legendre polynomials (Min and Hense, 2006), which are well suited to model non-linear trends because they are orthogonal, data-independent, and smooth. Let

$$x_i = \frac{2(i-1)}{n_y-1} - 1 \in [-1, 1], \quad i = 1, \dots, n_y, \quad (19)$$

denote the season index rescaled to $[-1, 1]$, the interval on which the Legendre polynomials are orthogonal, and let $\mathbf{g}_i = (g_0(x_i), \dots, g_K(x_i))^T$ collect the polynomials of degree 0 to K for the year i . We set $K = 5$ throughout the article.

3.3.1 Can extremes be attributed, and is their frequency increasing over time? – Using a constant threshold

With a constant threshold u in a non-stationary climate, the probability of exceeding this threshold $\phi_i = \mathbb{P}(y_{i,d} > u)$, which under assumption (18) is the same for all days d of season i , varies over time and can therefore no longer be treated as a fixed model parameter. Since we assume that the exceedance probability is affected by the general warming trend, we model the non-stationarity using logistic regression with Legendre polynomials as covariates. Logistic regression is a special case of a generalised linear model (Nelder and Wedderburn, 1972). We define the logistic regression using a logit link function such that

$$\text{logit}(\phi_i) = \log\left(\frac{\phi_i}{1-\phi_i}\right) = \boldsymbol{\beta}_\phi^T \mathbf{g}_i, \quad i = 1, \dots, n_y, \quad (20)$$

with $\boldsymbol{\beta}_\phi = (\beta_{\phi,0}, \dots, \beta_{\phi,K})^T$. The coefficients $\boldsymbol{\beta}_\phi$ are jointly estimated from the daily binary exceedance indicators $\mathbf{1}\{y_{i,d} > u\}$ of all seasons. The scale and shape parameters are modelled as

$$\sigma_i = \exp(\boldsymbol{\beta}_\sigma^T \mathbf{g}_i), \quad \xi_i = \boldsymbol{\beta}_\xi^T \mathbf{g}_i, \quad i = 1, \dots, n_y, \quad (21)$$

with $\boldsymbol{\beta}_\sigma = (\beta_{\sigma,0}, \dots, \beta_{\sigma,K})^T$ and $\boldsymbol{\beta}_\xi = (\beta_{\xi,0}, \dots, \beta_{\xi,K})^T$ as the vectors of regression coefficients.

The dependence parameter α_i is also assumed to vary over time, assuming that the dependence structure between two consecutive days can change slowly over time. Using a sigmoid link to ensure $\alpha_i \in (0, 1)$, we can write

$$\alpha_i = [1 + \exp(-\boldsymbol{\beta}_\alpha^T \mathbf{g}_i)]^{-1}, \quad i = 1, \dots, n_y. \quad (22)$$

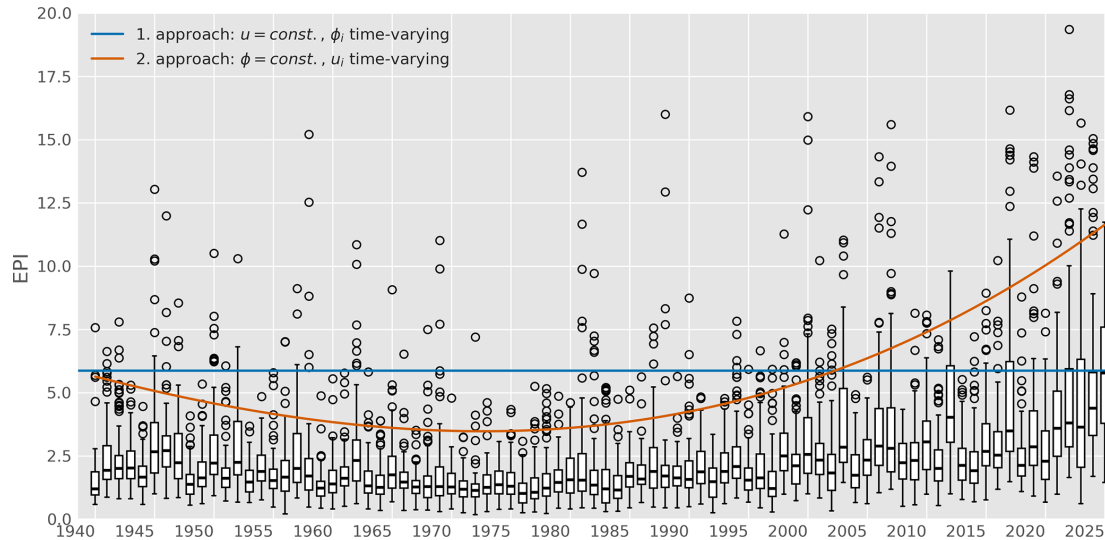


Figure 5. The ERA5 EPI for the southern European region is plotted as a box plot for each summer. In addition, the two thresholds from the two modelling approaches are shown. The first model's constant threshold is shown in blue and the second model's time-varying threshold is shown in orange, both for the 95 % quantile.

Note that the sigmoid link restricts α_i to the open interval $(0, 1)$; exact independence ($\alpha_i = 1$, cf. Eq. 11) is thus attained only in the limit $\beta_\alpha^T \mathbf{g}_i \rightarrow \infty$. The parameter vectors are therefore $\theta_1 = (\beta_\phi, \beta_\sigma, \beta_\xi)$ and $\theta_2 = \beta_\alpha$.

3.3.2 Is there a climate change signal in the tail behaviour of heatwave extremes beyond a general shift in the temperature distribution? – Using a variable threshold

Since large parts of the attribution statement are due to changes in the probability of exceeding the threshold u , we would like to exclude this effect and define the extremes on the original scale (Eastoe and Tawn, 2009). The attribution question that we ask in this case is as follows. Is there a change in the tail behaviour of heatwave extremes beyond a general shift in the temperature distribution? To model only distributional changes above a time-varying threshold u_i , we define u_i such that the non-exceedance probability $\mathbb{P}(y_{i,d} \leq u_i) = \tau$, which under assumption (18) is the same for all days d of season i , is constant, and hence $\phi = 1 - \tau$. We obtain u_i as the conditional τ -quantile by quantile regression (Koenker and Machado, 1999; Koenker, 2005), following Beirlant et al. (2004, Sect. 7.4.2). As Chavez-Demoulin and Davison (2012) point out, a time-dependent threshold is preferable for a more precise estimation of regression effects.

Quantile regression assumes a linear model for the conditional quantile,

$$u_i = \beta_\tau^T \mathbf{g}_i, \quad i = 1, \dots, n_y, \quad (23)$$

with coefficients β_τ estimated by minimising

$$\mathcal{R}_\tau = \frac{1}{n_y n} \sum_{i=1}^{n_y} \sum_{d=1}^n \rho_\tau(y_{i,d} - u_i), \quad (24)$$

where ρ_τ is the check function

$$\rho_\tau(z) = \begin{cases} \tau z & \text{if } z \geq 0, \\ (\tau - 1)z & \text{if } z < 0. \end{cases} \quad (25)$$

The scale and shape parameters are again modelled by Eq. (21) and the dependence parameter by Eq. (22), now based on the exceedances over the time-varying threshold. Since $\phi = 1 - \tau$ is constant, the parameter vectors are $\theta_1 = (\beta_\tau, \beta_\sigma, \beta_\xi)$ and $\theta_2 = \beta_\alpha$, again with $K = 5$.

3.4 Fitting procedure

The Markov process model is fitted according to Smith et al. (1997). We follow Beirlant et al. (2004) and first fit the marginal and dependence structure separately by introducing the covariates into the parameter estimation, as mentioned before. The optimal degree for the logistic regression (constant threshold) and the quantile regression (time-varying threshold) is then determined by the Bayesian information criterion (BIC). Then the GPD is fitted as non-stationary in order to find the optimal degree of scale and shape parameter using the BIC. With the optimal degree of the Legendre polynomials for each parameter determined by the BIC, we fit the model jointly using the previously estimated parameters separately as starting values for the parameters, as suggested by Beirlant et al. (2004). Here, “jointly” refers to joint optimisation of the GPD and dependence parameters. In practice,

Eq. (15) is thus solved over $(\beta_\sigma, \beta_\xi, \beta_\alpha)$, while β_ϕ (or β_τ , respectively) is held fixed at its separately estimated value and enters as a plug-in estimate. The reason for this is that the year-to-year probability of exceedance (first approach) or the year-to-year threshold value (second approach) is unlikely to change because of minor adjustments to the dependence function. Some advantages of the joint fit are mentioned by Beirlant et al. (2004), which are a better inference of the marginal parameters and a more reliable estimation of the dependence parameters.

4 Attribution statements for ERA5 using ensemble simulations

The complete workflow of the attribution process is summarised in Fig. 6.

4.1 Model estimation with constant threshold

To determine the likelihood, we need to estimate the parameter vectors for our scenarios m_s , where $s = 1$ refers to HIST and $s = 0$ to HIST-NAT as described in Sect. 2. We calculate the EPI for each realisation and both scenarios and pool the ensemble separately for each climate model and scenario, assuming that the ensemble members $r = 1, \dots, n_r$ are independent realisations of the non-stationary process under scenario m_s . We then estimate the parameter vectors $\beta_\phi^{(s)}$, $\beta_\sigma^{(s)}$, $\beta_\xi^{(s)}$ and $\beta_\alpha^{(s)}$, which yield the exceedance probability $\phi_i^{(s)}$ over the threshold $u^{(s)}$, the GPD parameters $\sigma_i^{(s)}$ and $\xi_i^{(s)}$, and the dependence parameter $\alpha_i^{(s)}$.

The parameter estimation process requires the implementation of a series of steps and the formulation of specific decisions.

1. First, the threshold u must be determined. It must be large enough to achieve asymptotic behaviour (Coles, 2001, Sect. 4.3.1), but small enough to allow sufficient data for reliable parameter estimation. In our application, we decided to use the 95 % quantile of the respective ensemble simulations for the period 1940 to 2020 (HIST-NAT) and 1940 to 2025 (HIST & SSP2-4.5). The threshold $u^{(s)}$ is therefore scenario-dependent. The advantage is that constant biases in the data are disregarded, and only the temporal evolution is investigated.
2. In the next step, we estimate the time-varying exceedance probability $\phi_i^{(s)}$ above the constant threshold $u^{(s)}$ using logistic regression (Sect. 3.3.1) for both scenarios. To determine the optimal maximum degree of the Legendre polynomials, we used the BIC. The fitting is performed using the statistical software package `statsmodels` (Seabold and Perktold, 2010). We model the non-stationarity of ϕ_i via its own regression formula because we expect this to be more robust than embedding it directly in the Markov likelihood.

3. The non-stationary GPD is estimated with `statsmodels` using maximum likelihood estimation. Again, the BIC is used to assess the optimal maximum degree of Legendre polynomials. Goodness-of-fit is assessed using quantile-quantile and probability-probability plots as described in Coles (2001, Sect. 6.2.3). All combinations of scale and shape parameters up to degree $K = 5$ are allowed, in which the degree of the shape parameter is not allowed to be larger than the degree of the scale parameter ($\deg(\sigma) \geq \deg(\xi)$). The reason for this condition is that the estimation of a covariate-dependent shape parameter introduces additional uncertainty (Friederichs et al., 2009); to account for this, the shape parameter is here restricted so as not to vary more flexibly than the scale parameter. This improves the stability of likelihood-based inference. Here, we use all threshold exceedances and do not apply a declustering scheme.

4. Next, the parametric form of V must be specified, in this example, the logistic model. For each scenario, the univariate estimates $\phi_i^{(s)}$ and the optimal degrees of $\sigma_i^{(s)}$ and $\xi_i^{(s)}$ are plugged into the censored likelihood approach. Then the optimal degree of the covariates for the dependence parameter is determined. Finally, using the estimated parameters as starting values, the model is fitted jointly and the resulting parameters are returned by the fit. The logistic-regression coefficients are not updated in this joint fit.

4.2 Likelihood of ERA5 with constant threshold

The likelihood ratios of Sect. 3.1 are evaluated for the ERA5 EPI time series. We write $\mathbf{Y}_{\text{ERA5}} = (\mathbf{y}_{\text{ERA5}}^{(1)}, \dots, \mathbf{y}_{\text{ERA5}}^{(n_y)})$ for the complete record and $\mathbf{y}_{\text{ERA5}}^{(i)}$ for the season of year i . The threshold u_{ERA5} is the 95 % quantile of the complete ERA5 EPI record over 1940 to 2025 (JJA), whereas the scenario parameters $\phi_i^{(s)}$, $\sigma_i^{(s)}$, $\xi_i^{(s)}$ and $\alpha_i^{(s)}$ vary from season to season. The scenario parameters describe the temporal evolution relative to the respective scenario threshold $u^{(s)}$; applying them to the ERA5 threshold thus disregards constant biases between the simulations and the reanalysis, consistent with the threshold choice in Sect. 4.1.

The evidence contributed by a single summer is then the seasonal likelihood ratio

$$\text{LR}(\mathbf{y}_{\text{ERA5}}^{(i)}) = \frac{l(\mathbf{y}_{\text{ERA5}}^{(i)}; u_{\text{ERA5}}, \phi_i^{(1)}, \sigma_i^{(1)}, \xi_i^{(1)}, \alpha_i^{(1)})}{l(\mathbf{y}_{\text{ERA5}}^{(i)}; u_{\text{ERA5}}, \phi_i^{(0)}, \sigma_i^{(0)}, \xi_i^{(0)}, \alpha_i^{(0)})}, \quad (26)$$

where the seasonal likelihood $l(\mathbf{y}^{(i)}; \cdot)$ is given by Eq. (7). Because the seasons are assumed independent, the likelihood ratio of the complete record follows as the product over seasons and coincides with the Bayes factor of Eq. (4),

$$\text{LR}(\mathbf{Y}_{\text{ERA5}}) = \prod_{i=1}^{n_y} \text{LR}(\mathbf{y}_{\text{ERA5}}^{(i)}) = \text{BF}. \quad (27)$$

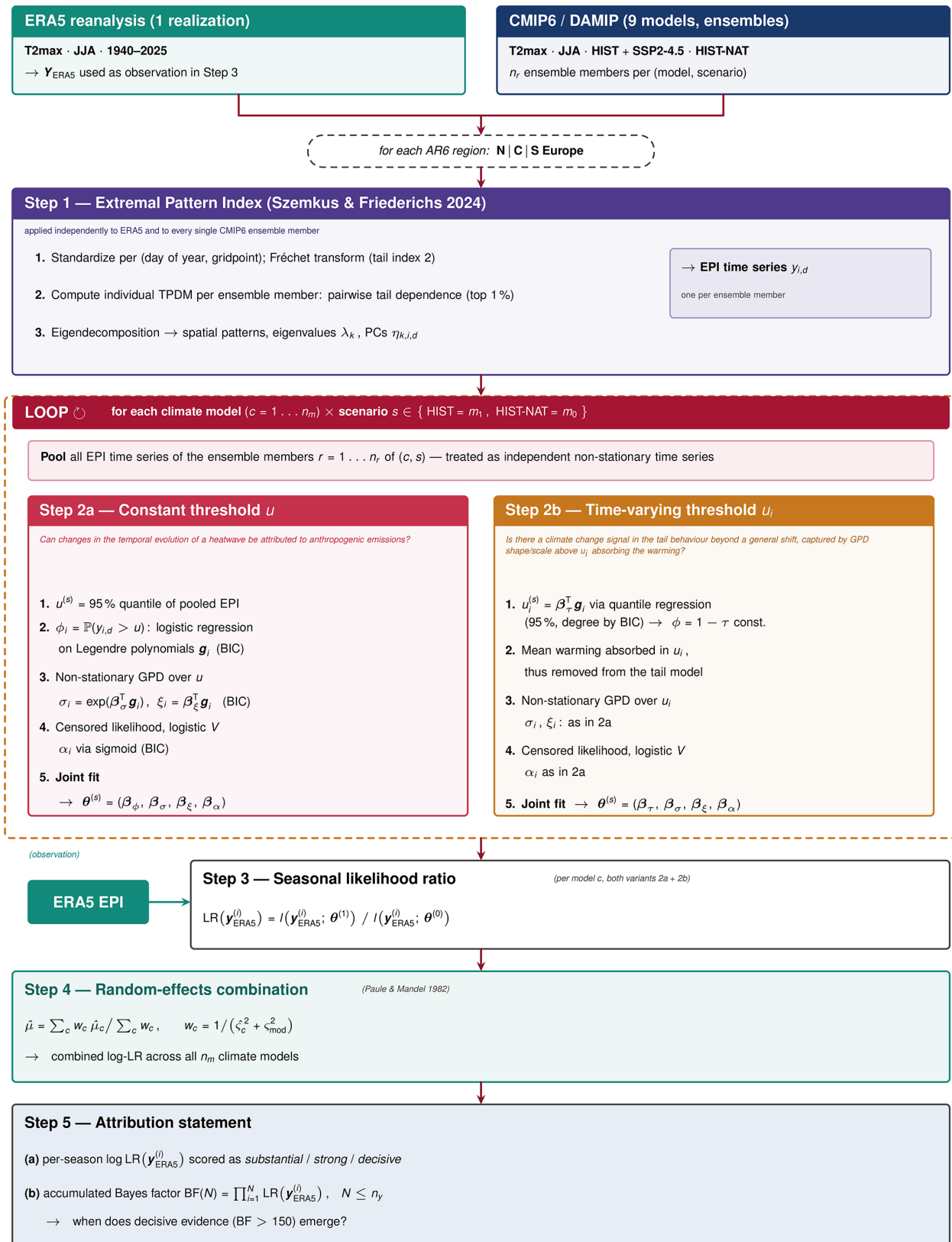


Figure 6. Flowchart of the attribution process, summarising Sect. 4 and describing all necessary steps in order to obtain the attribution result.

Equation (26) therefore yields an attribution statement for a single summer, and Eq. (27) for any period of consecutive summers up to a season to be chosen.

4.3 Model estimation with variable threshold

The attribution statement with a variable threshold u_i is different, since here we only consider the distribution above u_i . The time-varying threshold is assumed to contain a large part of the anthropogenic climate change signal, which is thus removed. Following Sect. 4.1, the likelihood of $\mathbf{Y}$ under scenario m_s is determined by the parameter vectors $\beta_\tau^{(s)}$, $\beta_\sigma^{(s)}$, $\beta_\xi^{(s)}$ and $\beta_\alpha^{(s)}$, which yield the time-varying threshold $u_i^{(s)}$, the marginal parameters $\sigma_i^{(s)}$ and $\xi_i^{(s)}$, and the dependence parameter $\alpha_i^{(s)}$. Again, we pool the ensemble separately for each climate model and scenario and describe the necessary steps and decisions for the model estimation.

1. The first step is to perform the quantile regression of the ensemble simulations for a sufficiently high quantile, in our case the 95 % quantile, as selected in Sect. 4.1, leading to the time-varying thresholds $u_i^{(s)}$. Again, we make use of the BIC to derive the optimal maximum degree of the Legendre polynomials. The fitting is performed once again using the statistical software package `statsmodels`.
2. In principle, the steps are similar to those described in Sect. 4.1, but now use time-varying thresholds to account for the effect of the mean increase in the time series. For fitting the non-stationary GPD, once again the package `statsmodels` can be used, and the model evaluation is the same as mentioned in Sect. 4.1. As for the logistic regression, the quantile regression is estimated using its own regression formula.
3. Again, the parametric form of V must be specified, in this example, the logistic model. For each scenario, the thresholds $u_i^{(s)}$ and the optimal degrees of $\sigma_i^{(s)}$ and $\xi_i^{(s)}$ are plugged into the censored likelihood approach to determine the optimal degree of the covariates for the dependence parameter. Finally, the model is fitted jointly with the parameters estimated separately as starting values. The quantile-regression coefficients are kept fixed and are not re-estimated in the joint fit.

4.4 Likelihood of ERA5 with variable threshold

For the model with a variable threshold, the ERA5 threshold $u_{i,\text{ERA5}}$ is obtained from the 95 % quantile regression of the complete ERA5 EPI record over 1940 to 2025 (JJA) and therefore varies from season to season. Since the exceedance probability is constant by construction, it does not enter the

ratio. The seasonal likelihood ratio reads

$$\text{LR}(y_{\text{ERA5}}^{(i)}) = \frac{l(y_{\text{ERA5}}^{(i)}; u_{i,\text{ERA5}}, \sigma_i^{(1)}, \xi_i^{(1)}, \alpha_i^{(1)})}{l(y_{\text{ERA5}}^{(i)}; u_{i,\text{ERA5}}, \sigma_i^{(0)}, \xi_i^{(0)}, \alpha_i^{(0)})}, \quad (28)$$

and the corresponding Bayes factor over any period of consecutive summers again follows from Eq. (27).

4.5 Uncertainty assessment

According to Paciorek et al. (2018), uncertainties arise from different sources. Uncertainty with respect to internal climate variability is generally assessed using large climate model ensembles, whereas ensembles with different climate models as used in this study further assess model uncertainty. Additional uncertainty arises from the observations of the climate system, which, however, is generally small compared to other sources of uncertainty.

To account for sampling uncertainty, two approaches are often used: the delta-method and bootstrapping (Jeon et al., 2016). In this study, we use the bootstrap method. For the bootstrap method, we proceed as follows for each scenario (repeating this for both scenarios of each climate model).

1. Split the data set into annual parts.
2. For each year, sample n_r indices with replacement from $\{1, \dots, n_r\}$, where n_r is the number of ensemble members of the respective climate model and scenario.
3. For each year, the ensemble members corresponding to the selected indices are used.
4. All years are combined.

Using this bootstrap method, the attribution is then repeated n_B times for each of the climate models; in this study, we use $n_B = 10$.

4.6 Combining different likelihood ratios

Due to the use of different climate models, we obtain a set of different likelihood ratios for an event or a time series. One may ask which likelihood ratio should be favoured or how a combined likelihood ratio can be calculated, since when communicating attribution results, the interest often lies in a specific value rather than a range of possible values. We follow an approach presented by Otto et al. (2024) that aims to combine different pieces of evidence (i.e., logarithmic probability ratios in their study) using an approach also used in meta-analyses and known as the random-effects model. This model is based on a paper published by Paule and Mandel (1982).

The main idea is to decompose the total variability into a contribution from natural variability ς_{nat} and a model uncertainty ς_{mod} , resulting in $\varsigma_{\text{tot}}^2 = \varsigma_{\text{nat}}^2 + \varsigma_{\text{mod}}^2$. For each climate model $c = 1, \dots, n_m$, we have a best estimate $\hat{\mu}_c$ of the logarithmic likelihood ratio and an estimate $\hat{\varsigma}_c$ of its standard

deviation from the bootstrap approach discussed in Sect. 4.5. The focus now lies on estimating the weights w_c for each climate model c , which represent the belief or confidence in the specific model, given by

$$w_c(\varsigma_{\text{mod}}) = \frac{1}{\hat{\varsigma}_c^2 + \varsigma_{\text{mod}}^2}. \quad (29)$$

The best estimate of the logarithmic likelihood ratio over the n_m climate models is then given by a weighted average

$$\hat{\mu}(\varsigma_{\text{mod}}) = \frac{\sum_{c=1}^{n_m} w_c(\varsigma_{\text{mod}}) \hat{\mu}_c}{\sum_{c=1}^{n_m} w_c(\varsigma_{\text{mod}})}, \quad (30)$$

and is only a function of the model uncertainty or the representation error term ς_{mod} . For the combined uncertainty, we can write

$$\varsigma_{\text{tot}}^2 = \frac{1}{\sum_{c=1}^{n_m} w_c} \sum_{c=1}^{n_m} w_c (\hat{\varsigma}_c^2 + \varsigma_{\text{mod}}^2). \quad (31)$$

The estimator of ς_{mod}^2 can be calculated using the methodology based on Paule and Mandel (1982). The main idea behind this approach is that the variable Q , which is the sum of the ratios of the two different estimates of variability, is distributed according to a chi-squared distribution with $n_m - 1$ degrees of freedom,

$$Q(\varsigma_{\text{mod}}) := \sum_{c=1}^{n_m} \frac{(\hat{\mu}_c - \hat{\mu}(\varsigma_{\text{mod}}))^2}{\hat{\varsigma}_c^2 + \varsigma_{\text{mod}}^2} \sim \chi_{n_m-1}^2. \quad (32)$$

The expected value of this sum is given by $\mathbb{E}[Q] = n_m - 1$, from which we can estimate ς_{mod} using the formula

$$\hat{\varsigma}_{\text{mod}} = \begin{cases} 0, & \text{for } Q(0) \leq n_m - 1, \\ \text{Root}(Q - n_m + 1), & \text{for } Q(0) > n_m - 1. \end{cases} \quad (33)$$

Thus, the estimation of ς_{mod} amounts to a root-solving problem if $Q(0) > n_m - 1$.

The random-effects model of Eq. (30) treats the climate models as a priori independent. Given the documented dependence among CMIP6 models, which share, e.g., components or parameterisations, this assumption is likely to be violated to some degree; the resulting ς_{tot} should therefore be regarded as a lower bound on the combined uncertainty.

5 Results

5.1 Model parameter estimation

According to Sect. 4, both models with constant and time-varying thresholds are fitted to the data. For each climate model and scenario, we obtain a set of parameter coefficients. Figure 7 shows the coefficient estimates of the model with a constant threshold (Sect. 4.1) for the EPI of the southern European region. The same model was also fitted to the EPI

in ERA5 for comparison. The first coefficient in each panel always represents the intercept, the second is a linear trend over time, and the third is a quadratic trend. For the HIST-NAT scenario simulations, we do not expect tendencies, and indeed the intermodel variability of the trend parameters generally includes the zero line.

For the simulations of the HIST scenario, the exceedance probability shows significant positive trends, particularly in the linear and quadratic Legendre polynomials. For some CMIP6 models, the higher-order polynomials improve the BIC, but the signals are not consistent between the models. The parameter estimates for the threshold exceedance parameters in ERA5 are close to the range spanned by the estimates of the CMIP6 models. The uncertainty in the ERA5 parameter estimation is significantly larger than in the CMIP6 models, as the latter are estimated using a simulation ensemble and not, as in ERA5, a single realisation. Consequently, the ERA5 estimates may lie outside the uncertainty range given by the models.

The scale parameter estimates in the HIST scenario show a consistent linear increase, with a less consistent quadratic component. Again, ERA5 shows a similar behaviour in the non-stationarity of the scale parameter. However, the positive tendency in the scale parameter is counteracted by a negative linear trend in the shape parameter. This negative trend in the shape parameter is visible in some models in the HIST scenario, but not in ERA5. The overall positive shape parameter is expected in all scenarios, all simulations, and in ERA5, since the input data are the EPI, and the EPI in turn relies on Fréchet-transformed T2max values. For the dependence parameter, the linear and quadratic coefficients are relevant, while in ERA5 only the linear term is non-zero. Similar tendencies in the parameters are observed for the northern and central European regions as shown in the Appendix (Sect. A4) in Figs. A6 and A7, with a similarly good agreement between the CMIP6 simulations and ERA5.

The resulting temporal evolution of the parameters is shown in Fig. 8. The increasing threshold exceedance probability in the HIST scenario is clearly visible, with a strong increase starting in the 1980s as a consequence of the second-order Legendre polynomial. A similar trend, with an even stronger increase during the last decade, is visible for the ERA5 threshold exceedance probability. The scale parameter increases in the HIST scenario. Although this trend is stronger and starts at a lower level, it is consistent with the trend in ERA5. The trends in the shape parameter show large uncertainties between the different CMIP6 models. The ERA5 shape parameter is at the upper uncertainty level of the models and could counteract the effect of the comparatively low scale parameter.

The increasing dependence (i.e., decreasing α_i) is visible for the HIST scenario and ERA5. The dependence in ERA5 is weaker at the beginning, and the trend towards increasing dependence is significantly stronger compared to the HIST scenario. The increasing dependence reflects the greater fre-

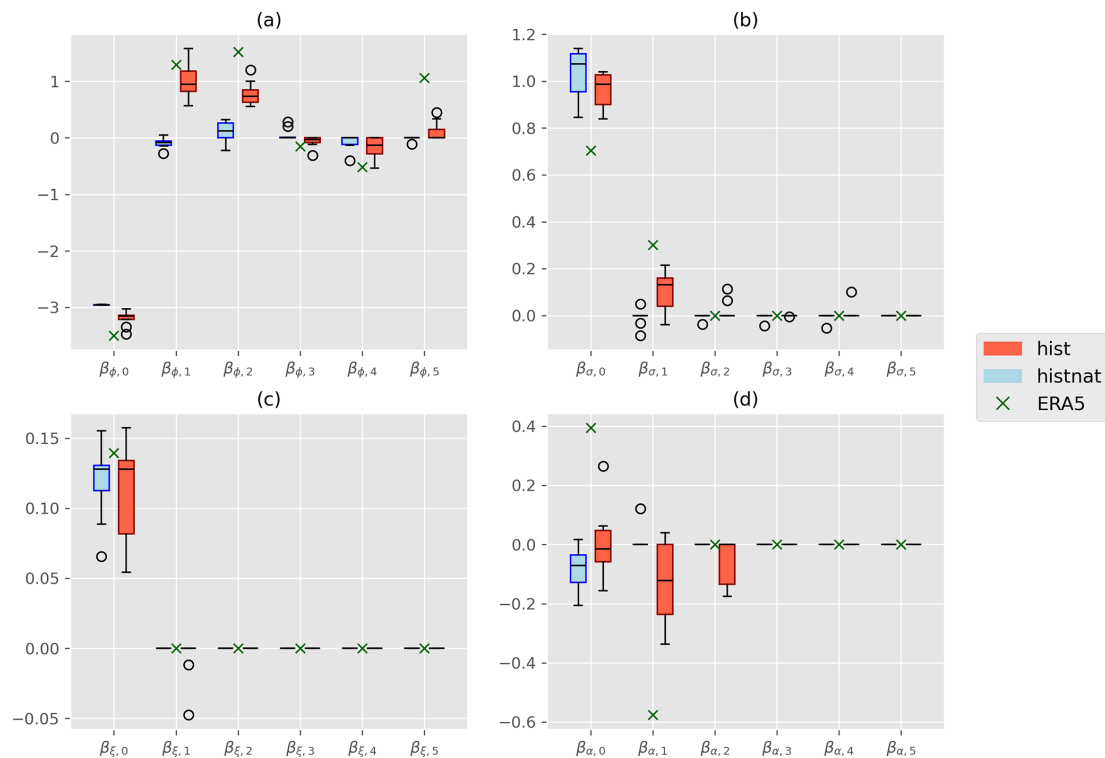


Figure 7. Estimated coefficients for the model with constant threshold (Sect. 4.1) in the southern European region. The coefficients (x -axis) are shown for (a) the threshold exceedance ϕ_i , (b) the scale σ_i , (c) the shape ξ_i , and (d) the dependence α_i . The estimates for the different climate models are represented as box-whiskers, those for ERA5 as green crosses. The whiskers and fliers cover the whole range of data.

quency and duration of heatwaves, which leads to higher persistence of hot extremes, and therefore, more consecutive hot days.

Figure 9 shows the coefficient estimates of the model with time-varying threshold (Sect. 4.3) for the EPI of the southern European region. In the model with a time-varying threshold, there are no coefficients for the threshold exceedance probability, since this is constant by definition. Instead, we display the quantile regression parameters, which contain a large part of the non-stationarity but which are not part of the likelihood model.

For the HIST-NAT scenario, the model variability of the trend parameters includes the zero line, similar to the model with a constant threshold. For the HIST scenario, the linear and quadratic coefficients of the quantile regression are significantly different from zero, and thus show a positive trend. ERA5 exhibits behaviour similar to that of the HIST scenario.

For the scale and shape parameters, the linear and quadratic coefficients become relevant. Although the coefficients indicate a positive trend in the scale parameter in the HIST scenario, the trend for the shape parameter is negative. Higher-order coefficients are only relevant for a subset of the climate models in the HIST scenario. In contrast, for ERA5 only the first-order Legendre polynomial becomes relevant for the scale parameter, while the shape parameter is station-

ary. In terms of dependence, trend coefficients are only relevant for a small subset of models. This indicates that stationary dependence is the norm for both scenarios, as is the case for ERA5.

The resulting temporal evolution of the parameter estimates (Fig. 10) of the HIST-NAT scenario does not show significant trends. An increase in the threshold is visible for the HIST scenario and in ERA5, which is a consequence of the general warming trend. In this HIST scenario, the scale parameter also increases with time. In contrast, the shape parameter initially increases slightly, then decreases, and becomes negative for most of the climate models after 2010. This behaviour is not reflected by ERA5. Regarding the dependence, the temporal evolution is nearly constant for both scenarios and ERA5.

The negative shape parameter ($\xi_i < 0$) in the HIST scenario means that the exceedances are bounded above, with an upper limit of $u_i - \sigma_i/\xi_i$ (Coles, 2001, Sect. 4.2.1). We therefore observe that the distribution of exceedances of the time-dependent 95 % quantile becomes light-tailed over time. This effect is partially offset by the increases in the scale parameter. It can thus be concluded that although the variance of the exceedances increases, the probability of large outliers decreases. However, the tendency towards a negative shape parameter may also indicate that the Fréchet transformation required to determine the EPI is no longer adequate and that

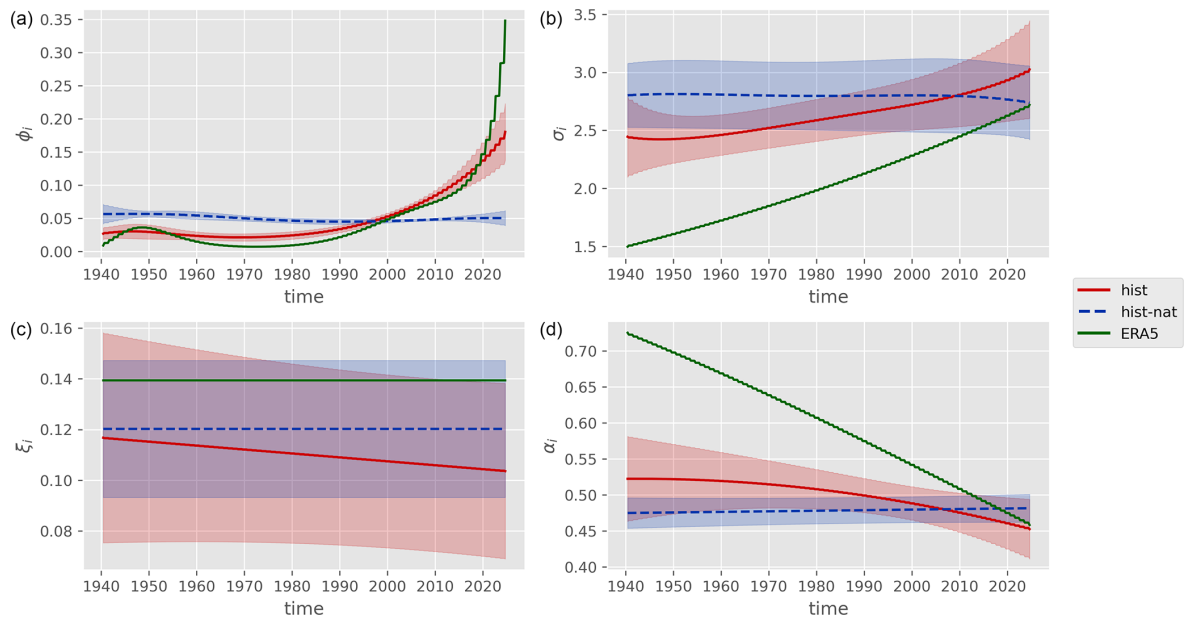


Figure 8. Temporal evolution of (a) threshold exceedance probability, (b) scale, (c) shape, and (d) dependence parameters for the model with constant threshold in the southern European region. Mean (solid line) and standard deviation (shading) for the different climate models are plotted based on the estimates in Fig. 7.

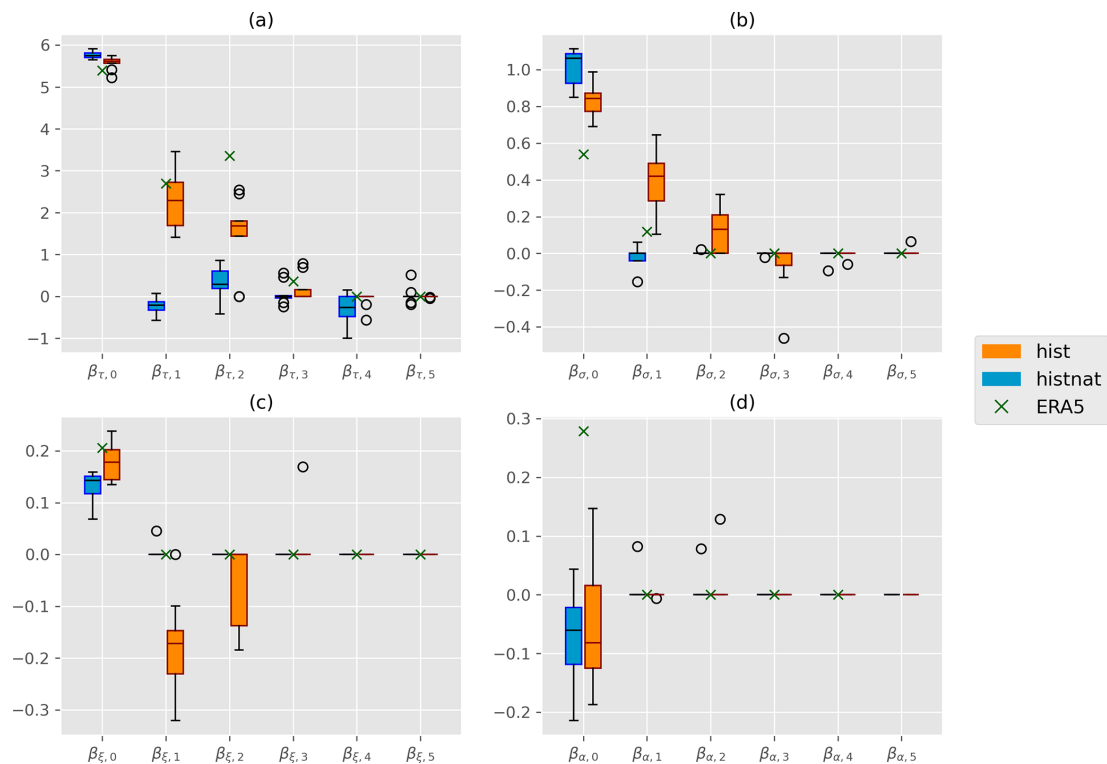


Figure 9. Same as Fig. 7 but for the model with time-varying threshold (Sect. 4.3) in the southern European region, except that (a) shows the quantile regression parameters of the time-varying threshold.

the strong non-stationarity must therefore be taken into account when defining the spatial patterns.

What we can show is that the sign change of the shape parameter in this model shows a significant improvement in BIC compared to a model that either keeps the shape parameter constant or restricts it to the positive value range. Very similar tendencies, particularly the negative trend in the shape parameter, are also observed for the other regions, as shown and described in Appendix A4 in Figs. A8 and A9.

Regarding potential biases of the GCMs, we can assume that first- and second-order biases are removed by the standardisation to Fréchet. This ensures a common scale for all the different climate models. Another bias is removed by standardising the EPI. A good indicator for this is that the ERA5 parameter estimates generally lie within the inter-model spread of the HIST scenario. Due to the smaller sample size, the sampling uncertainty of ERA5 and of the individual models is much larger, but it is not displayed in the figures. This might explain some of the discrepancies where the ERA5 coefficients fall outside the uncertainty range of the HIST coefficients. If the estimates from ERA5 and the climate models differ significantly, this may indicate differences in the physical processes represented in the models. However, this investigation is beyond the scope of this work. Better climate models with higher resolution may be required. It could also point to inconsistent changes in ERA5. The advantage of our modelling approach is that such differences can be made visible.

5.2 Attribution results

Given the parameter estimates of the non-stationary Markov models, we can calculate the seasonal likelihood ratios in Eqs. (26) and (28) according to Sect. 4.6. Figure 11 shows the logarithmic likelihood ratios for both models with constant and variable thresholds, and the respective standard deviations for summers in the northern European region since 2000. The first event that provided substantial evidence for the HIST scenario was in 2010, and two events in 2018 and 2022 even showed strong evidence for a single northern European summer time series. All events are documented as extreme heat events. The likelihood ratio for the summer of 2025 is not among the three highest likelihood ratios. This might be surprising, since the July 2025 heatwave was exceptional, as documented by Barnes et al. (2025). Our attribution refers to the entire summer period with a relatively cold June, which reduces the corresponding likelihood ratio (Fig. 12), while the attribution in Barnes et al. (2025) only considers the hot phase in July. Compared to summer 2022, which has a similar number of threshold exceedances, the resulting likelihood ratio of summer 2022 is higher due to the different types of heatwave. In 2022, there were three heat events, whereas in 2025 there was only one (see Fig. 12), which affects the attribution result, since the Markov transitions are different. In general, we can conclude that the attribution result is affected

not only by the number of threshold exceedances (as shown in Fig. 11), but also by the type (short- or long-term events) of the event, the number of events as well as the strength of the EPI.

For the model with a time-varying threshold, in the northern European region, most summers have a logarithmic likelihood ratio of approximately zero (i.e., a likelihood ratio of one). This means that no other effect can be detected in the northern European region apart from a general increase in the threshold value. The significant difference in the likelihood ratios between 2021 and 2022 is due to the much higher EPI values observed during the heat events in 2021 compared to 2022.

For the central European region, Fig. 13 shows more years with a logarithmic likelihood ratio that indicates strong evidence than for the northern European region. The summer of 2010 with the heatwave over eastern Europe and Russia is still outstanding in terms of threshold exceedances. The highest likelihood ratios are observed for the summers in 2010, 2012, 2015, 2019, 2022, and 2025. All of these years fall into the *strong* category and are considered to be summers with heatwaves. The compound heat and drought event in 2018 (Xoplaki et al., 2025) has a much smaller likelihood ratio, also with a lower number of threshold exceedances. In terms of maximum temperature, the summer of 2018 in central Europe was not as extreme as in the other years. The severe impacts in this summer were mainly due to the combination of high temperatures and drought. The variable threshold model again provides no evidence beyond a trend in the threshold, except perhaps for the summer of 2019.

In the southern European region (Fig. 14), the year 2025 was outstanding compared to the other years. On almost half of the summer days, the EPI for the southern European region exceeded the 95 % quantile, whereas the summer was similar to the summer of 2024 in terms of average maximum summer temperatures. High likelihood ratios can be observed primarily since 2021, with the summers of 2012, 2021, 2023, and 2025 falling into the *decisive* category. The southern European region clearly shows the strongest evidence of climate change among the three regions.

When the effect of the variable threshold is included, the logarithmic likelihood ratio fluctuates around zero in all regions. However, uncertainty appears to increase with higher threshold values, which can cause both downward and upward excursions over a period of several years. The negative shape parameter of the variable threshold model observed in the anthropogenic-driven simulations after 2010 is not observed in ERA5 and could therefore be an artefact of the numerical models. Alternatively, this could indicate possible changes in processes that are not yet apparent in the ERA5 data but occur in climate simulations under the driving scenario. Therefore, it is also possible that the modelled trends in the CMIP6 simulations are not yet observable in ERA5.

Years with a high likelihood ratio in the time-dependent threshold model are usually years with many event/non-event

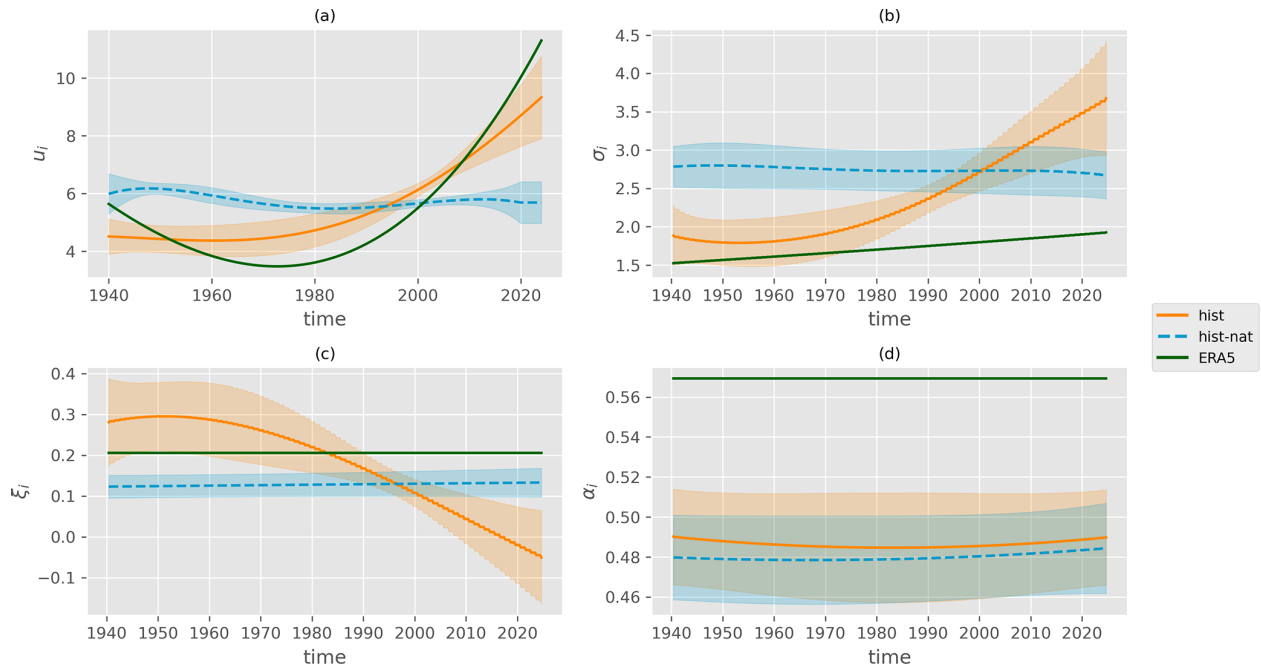


Figure 10. Same as Fig. 8 but for the model with time-varying threshold in the southern European region.

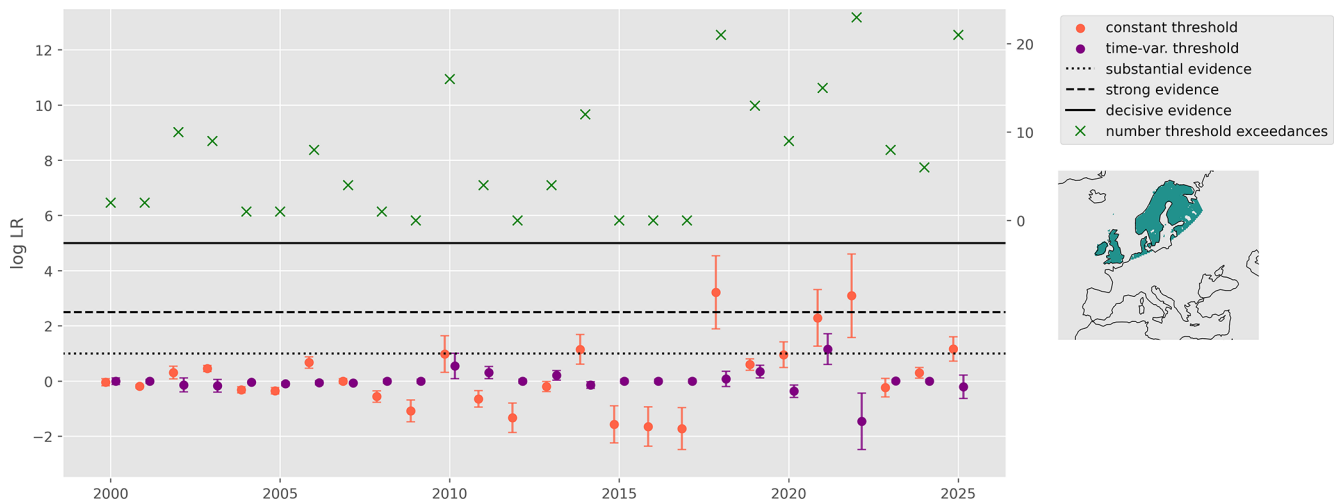


Figure 11. Logarithmic seasonal likelihood ratio $\log \text{LR}(\mathbf{y}_{\text{ERA5}}^{(i)})$ including $\pm$ standard deviation (y-axis left) for the northern European region, estimated according to Sect. 4.6. Green crosses indicate the number of threshold exceedances of the ERA5 EPI over the 95 % quantile in the respective summer (y-axis right).

transitions, where the EPI exceeds the threshold on only one of two consecutive days. Such events become more likely in the HIST scenario, especially due to the change in the shape parameter. This applies to all three regions, as the parameters and their trends are similar for the different regions.

Finally, we are interested in the question of when in the past we would have had sufficient evidence for climate change using the means available to us today. To answer this question, we accumulate the seasonal likelihood ratios of Eq. (26) over consecutive summers. Under uniform prior

probabilities, Eq. (4) gives

$$\frac{\mathbb{P}(m_1 | \mathbf{y}_{\text{ERA5}}^{(1)}, \dots, \mathbf{y}_{\text{ERA5}}^{(N)})}{\mathbb{P}(m_0 | \mathbf{y}_{\text{ERA5}}^{(1)}, \dots, \mathbf{y}_{\text{ERA5}}^{(N)})} = \prod_{i=1}^N \text{LR}(\mathbf{y}_{\text{ERA5}}^{(i)}) = \text{BF}(N), \quad (34)$$

where $N \leq n_y$ denotes the most recent season included, and $\text{BF}(n_y) = \text{BF}$ as defined in Eq. (4). Varying N from the first to the last season allows us to follow the accumulation of evidence from 1940 to 2025. If $\text{BF}(N)$ exceeds the value of 150 (see Table 2), we conclude that there is decisive evidence

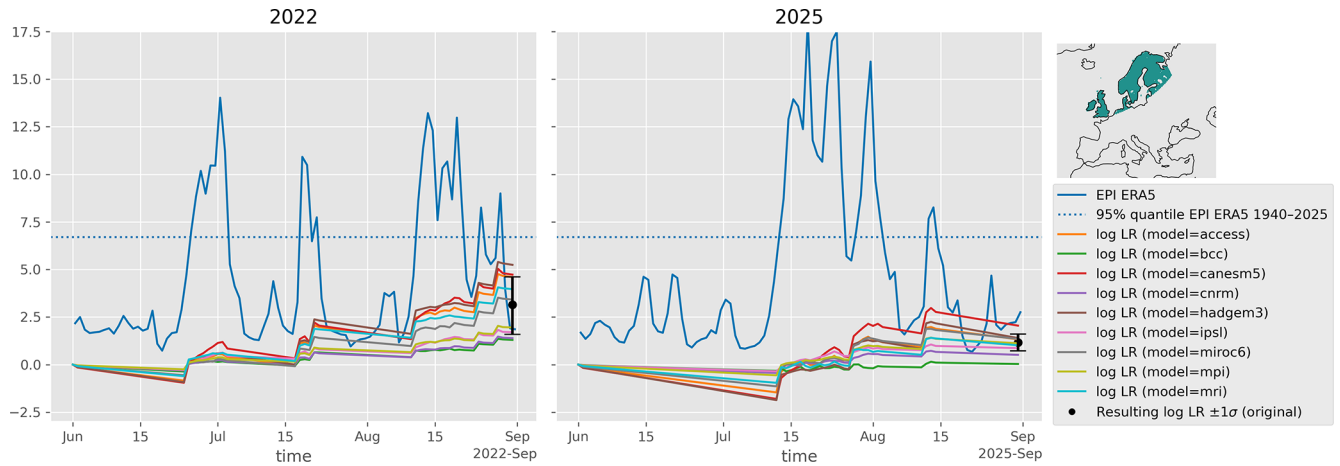


Figure 12. EPI of ERA5 and the logarithmic seasonal likelihood ratio for the summers 2022 and 2025 in the northern European region and the constant threshold model. The estimate shown in Fig. 11 is marked in black.

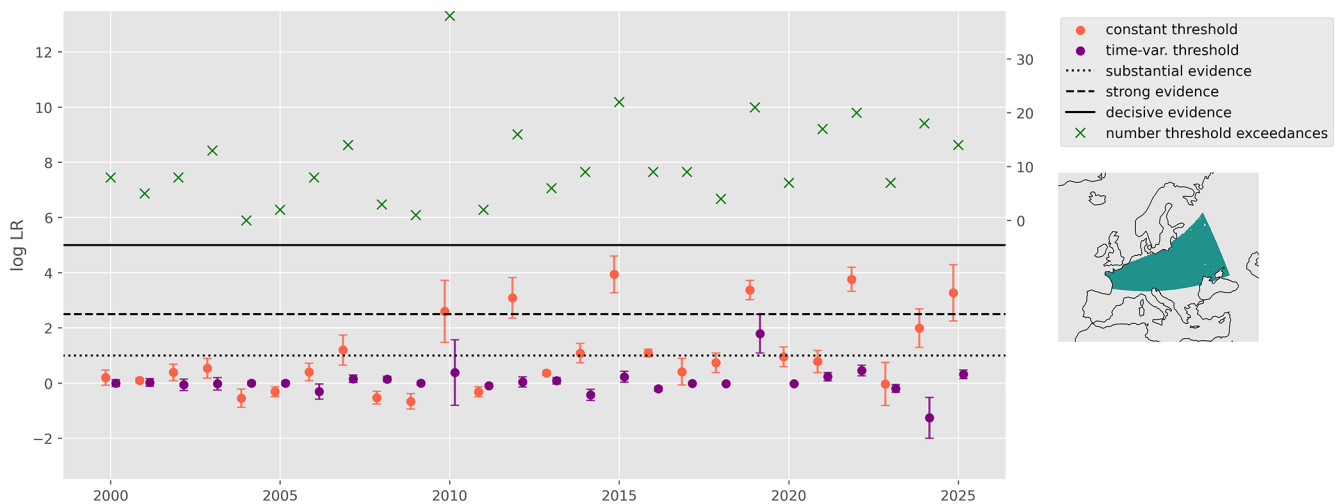


Figure 13. Same as Fig. 11 but for the central European region.

against the HIST-NAT scenario after the summer of season N .

For the model with constant threshold, the level of decisive evidence is reached in 1962 in the central European region and in 1964 in the southern European region. For the northern European region, the Bayes factor has repeatedly fallen below and risen above this value since 1940. However, since 2021, it has remained consistently above 150. For the central and southern European regions, the Bayes factor since 1940 has an increasing trend throughout the time period. For example, after the summer of 2025, we find a value of

$$\text{BF}(n_y) = \prod_{i=1}^{n_y} \text{LR}(y_{\text{ERA5}}^{(i)}) \approx 2.25 \times 10^{26}, \quad (35)$$

for the southern European region, i.e. accumulated over the summers from 1940 to 2025. In other words, the probability of the scenario with anthropogenic emissions given the

summers from 1940 to 2025 is 10^{26} times higher than the probability of the scenario without anthropogenic emissions given the summers from 1940 to 2025. For the model with a time-varying threshold, we can still attribute the entire period from 1940 onwards to the HIST scenario. As can be seen in Fig. 15, after an overall increase in the Bayes factor from 1940 to 1980, it remains almost constant thereafter. This implies that summers after 1980 do not provide additional evidence for the HIST scenario in the model with a time-varying threshold, and all support for the HIST scenario is due to the period in which we do not expect an anthropogenic climate change signal.

6 Discussion and conclusion

The Markov process model, which is based on bivariate extreme value theory, has been shown to be suitable for mod-

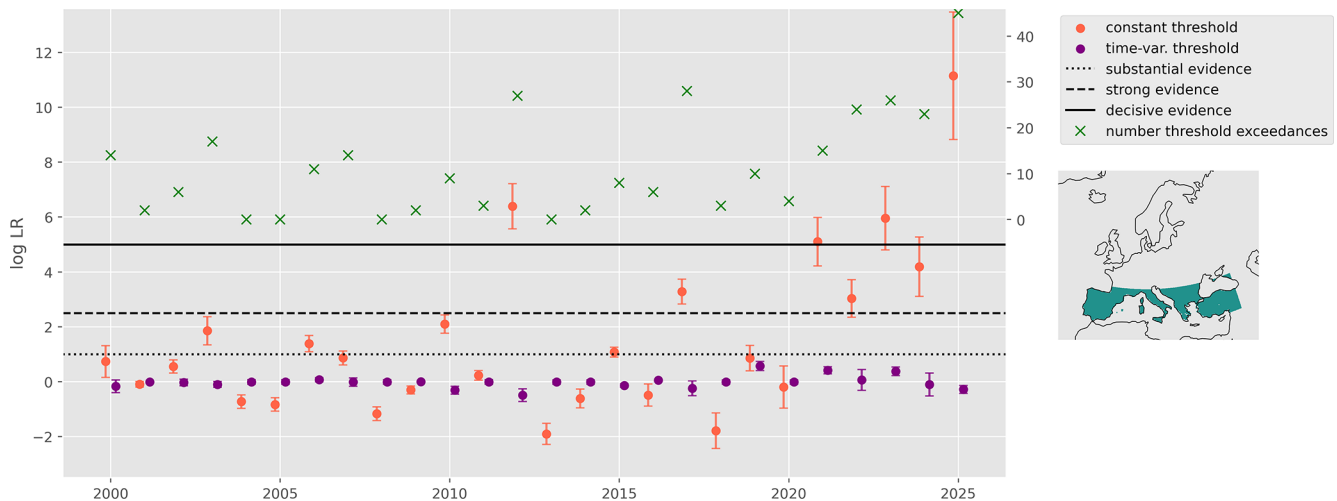


Figure 14. Same as Fig. 11 but for the southern European region.

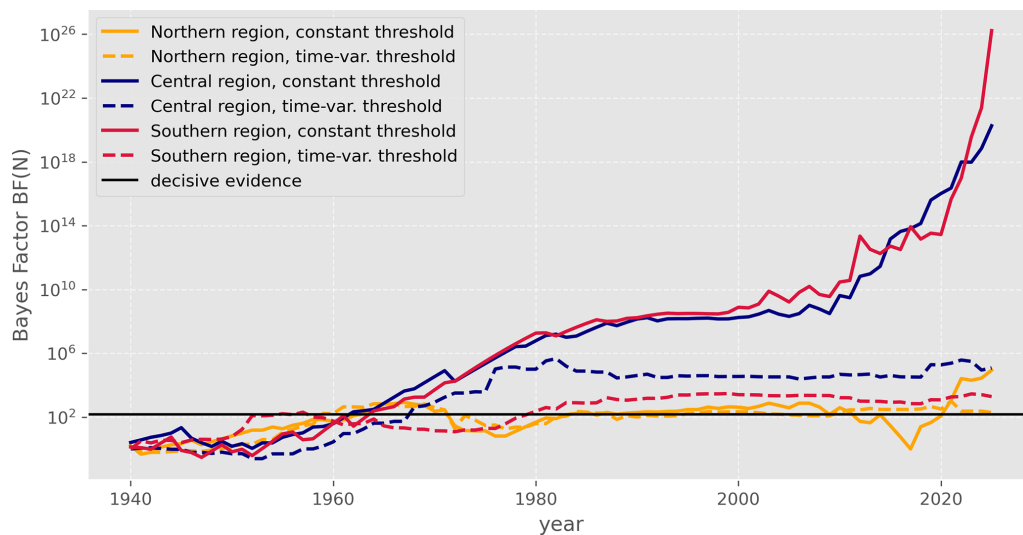


Figure 15. Accumulated Bayes factor $BF(N)$ of Eq. (34) for all summers since 1940 up to the most recent season N included (x -axis, labelled by calendar year), assuming equal prior probabilities for both scenarios.

elling the temporal dependence of the EPI, and thus deriving the likelihood of the observed EPI in ERA5 given different scenarios. Due to the Markov assumption, time periods ranging from two days to several years can be considered for attribution. The non-stationarity of the Markov process allows us to examine the temporal changes in the time series due to climate change. To our knowledge, the adaptation of the censored threshold model to the non-stationary case is new. With a second model featuring a constant exceedance probability and a corresponding time-dependent threshold value, we can also answer the question of whether climate signals exist in the behaviour of extremes that go beyond a mean increase in the threshold value.

Our approach enables us to statistically model the temporal development of extreme events and thus also to conduct

a corresponding attribution study. For heatwaves in Europe, there is a clear answer to the question of whether their temporal evolution can be attributed to anthropogenic emissions: The entire summer time series of the EPI from ERA5 can be assigned to the HIST scenario for each region with decisive evidence. Since 2000 in particular, heatwave events have increased in frequency, and the number of summers that can be attributed to the HIST scenario with strong or decisive evidence has risen. In the central and southern European regions, decisive evidence in favour of the scenario with anthropogenic emissions has been available since the 1960s, when considering all summers since 1940.

There is no clear answer to the second question, whether there is a climate change signal in the tail behaviour of heatwave extremes beyond a general shift in the temperature dis-

tribution. In the first years of the time series, an increasing shape parameter in the HIST scenario supports this hypothesis. However, the estimated non-stationarities over the last three decades significantly differ between the HIST scenario simulations and ERA5 in the variable threshold model. The trend towards a negative shape parameter in the HIST scenario suggests an upper endpoint for extremes, albeit counteracted by an increased scale parameter. In addition, the shape parameter is only slightly negative, so the upper bound applies only to very extreme observations. Neither is reproduced in ERA5. Particular summers with many event/non-event transitions are attributed to the HIST scenario, since the respective CMIP6 model shape parameter estimates make these events more likely. All support for the HIST scenario is due to the period in which we do not expect an anthropogenic climate change signal.

What we can summarise is that heatwave extremes are becoming more extreme in terms of intensity – as evidenced by higher threshold exceedances over time. However, once the general warming trend is absorbed into a time-varying threshold, the picture changes: in the HIST scenario, the GPD scale parameter increases, but there is no evidence of a corresponding increase in the shape parameter. In other words, relative to the shifting temperature distribution, the tail behaviour of exceedances has not changed in a detectable and consistent way across CMIP6 models and ERA5 – and in this specific sense, extremes are not becoming more extreme. It is important to emphasise that this statement differs fundamentally from the first. The increasing frequency and intensity of heatwaves is a robust and decisive finding. In contrast, the absence of a detectable change in tail behaviour is a statement about the scale and shape of the distribution conditional on the warming trend having already been accounted for.

In principle, we can evaluate climate models using a similar approach as the World Weather Attribution. Unlike their analysis, we could take into account more than just linear trends (Philip et al., 2020), since the methodology for testing trends could also be applied to higher-order trend coefficients. Our results show that the models in the HIST scenario essentially reproduce the ERA5 trends. Despite the relatively complex statistical model, comparing the coefficients of the non-stationary model allows us to assess how well the past is represented in the climate models. However, this can only be done by taking into account the uncertainty in the ERA5 coefficient estimates. If non-stationarities in the historical datasets are well captured, this could also provide an indication of how well climate models can handle transient climate conditions.

In the Appendix, we further examine the influence of a seasonal cycle that may still be present in the data on the attribution (Appendix A2) and an extension to a second-order Markov process in Appendix A3. Since both extensions have only a minor influence on the attribution statement, the results are not included in the main body of the article.

To summarise, the main contribution of this paper lies in the extension of the censored likelihood model for extremes to non-stationary time series, and its application to the attribution of the evolution of heatwaves over different regions in Europe. The significant non-stationarities in the temperature data are an issue for any extreme value model. A constant threshold model needs to be handled with care in the presence of strong temporal trends. In particular, the strong non-stationarity affects the estimated dependence structure in the constant threshold model.

Our second approach uses a time-varying threshold value to eliminate its effect and evaluate changes beyond it. However, the peaks-over-non-stationary threshold model of Friederichs (2010) can also be adapted so that it can be used for extreme dependencies in time series and thus also, for example, for attribution in Eq. (28). The main difference from Friederichs (2010) in this article is that the time-varying threshold model uses a threshold estimated on ERA5 data to remove the ERA5 warming trends. The significantly stronger non-stationarities in future climate projections require sophisticated statistical extreme value models that adequately account for non-stationarities. This work is therefore an important step towards the statistical evaluation of non-stationary time series for extremes in the past and future. In this sense, our work can supplement existing analyses, such as those carried out by the World Weather Attribution, by offering a perspective that is less dependent on event definitions.

In terms of future work, we could also use models that combine asymptotic dependence and independence (e.g., Ramos and Ledford, 2009) to attribute time series using the Markov process approximation.

Appendix A

A1 Reasoning behind the non-stationary modelling of exceedance probability and threshold

In this subsection, we explain why we use non-stationary modelling for exceedance probability and threshold if the scale and shape parameters of the GPD are non-stationary. Throughout this subsection, we drop the day index d , in line with assumption (18), and write $z \geq u$ for a generic level of interest; the season index i is dropped as well wherever the process is stationary.

Allowing the distributional parameters to depend on covariates already induces non-stationary changes in the distribution – but only with respect to the scale and shape parameters of the GPD. Given a stationary process, we have, for $z \geq u$,

$$\mathbb{P}(y > z \mid y > u) = \left(1 + \xi \frac{z - u}{\sigma}\right)_+^{-1/\xi}. \quad (\text{A1})$$

This is the conditional survivor function of observing a large value z , given that the threshold is exceeded (Coles, 2001,

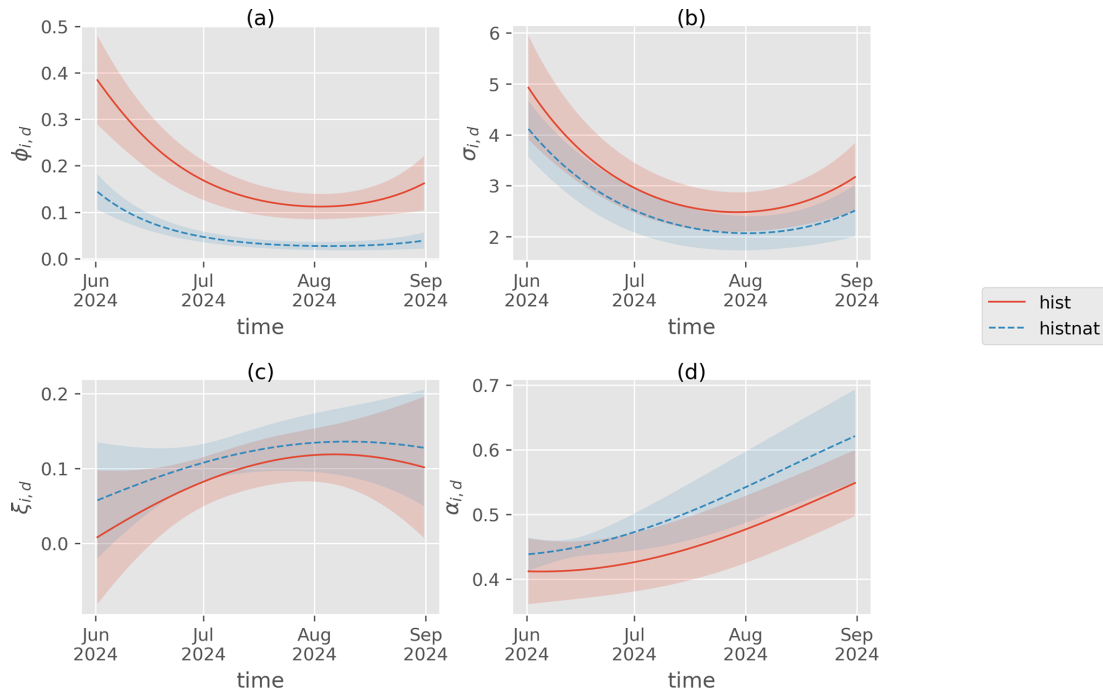


Figure A1. The seasonal cycle of (a) threshold exceedance probability, (b) scale, (c) shape, and (d) dependence parameters for the southern European region is shown for the year 2024, with mean and standard deviation calculated based on the different climate models.

Sect. 4.3.3). Only the scale and shape parameters of the GPD enter here. An additional quantity that comes into play when considering the unconditional distribution $\mathbb{P}(y > z)$ is the threshold exceedance probability $\phi = \mathbb{P}(y > u)$. For the censored threshold model, we require this unconditional distribution (see Eq. 6), given by

$$\begin{aligned} \mathbb{P}(y > z) &= \mathbb{P}(y > u) \mathbb{P}(y > z \mid y > u) \\ &= \phi \left(1 + \xi \frac{z - u}{\sigma} \right)_+^{-1/\xi}, \quad z \geq u. \end{aligned} \quad (\text{A2})$$

Equation (A2) is precisely the transformed margin $v_{i,d}$ of Eq. (9); the exceedance probability therefore enters the model already at the stage of the transformation to standard Fréchet margins.

We now assume that the process is non-stationary and that a sequence of covariates $\mathbf{g}_i$ is available; here these are Legendre polynomials. Equation (A1) is then generalised to

$$\mathbb{P}(y_i > z \mid y_i > u, \mathbf{g}_i) = \left(1 + \xi_i \frac{z - u}{\sigma_i} \right)_+^{-1/\xi_i}, \quad z \geq u, \quad (\text{A3})$$

with $\sigma_i = \exp(\boldsymbol{\beta}_\sigma^T \mathbf{g}_i)$ and $\xi_i = \boldsymbol{\beta}_\xi^T \mathbf{g}_i$, as defined in Eq. (21) of Sect. 3.3, following the covariate-dependent GPD framework of Eastoe and Tawn (2009). As in the stationary case, however, the censored likelihood requires the *unconditional* distribution of an exceedance, i.e.

$$\begin{aligned} \mathbb{P}(y_i > z \mid \mathbf{g}_i) &= \mathbb{P}(y_i > u \mid \mathbf{g}_i) \mathbb{P}(y_i > z \mid y_i > u, \mathbf{g}_i), \\ &z \geq u. \end{aligned} \quad (\text{A4})$$

Because of the non-stationarity, the conditioning on $\mathbf{g}_i$ applies to *all* components of the model, including the exceedance probability, even though this conditioning is often suppressed in the notation. We can therefore not use the stationary quantity $\mathbb{P}(y > u)$, but must instead use the non-stationary threshold exceedance probability $\phi_i = \mathbb{P}(y_i > u \mid \mathbf{g}_i)$, which we model as in Eq. (20),

$$\text{logit}(\phi_i) = \boldsymbol{\beta}_\phi^T \mathbf{g}_i, \quad (\text{A5})$$

using a logistic regression, again in line with Eastoe and Tawn (2009).

The stationary quantity $\phi = \mathbb{P}(y > u)$ cannot simply be substituted for $\mathbb{P}(y_i > u \mid \mathbf{g}_i)$ in Eq. (A4). The reason for this is that Eq. (A4) is not a modelling choice but an identity. Inserting a constant ϕ in place of $\phi_i = \mathbb{P}(y_i > u \mid \mathbf{g}_i)$ would violate this identity unless $\mathbb{P}(y_i > u \mid \mathbf{g}_i)$ happened to be constant in time – an assumption that is not justified here, in particular given the underlying warming trend.

In other words, the two components are not two alternative ways of introducing non-stationarity; they are two parts of the same non-stationary marginal model and cannot be specified independently of one another. When the threshold u is held fixed and the parameters σ_i and ξ_i vary with $\mathbf{g}_i$, the probability mass above u varies as well. Assuming a constant ϕ alongside time-varying σ_i and ξ_i would render the marginal model internally inconsistent. Thus, the logistic regression in Eq. (A5) is not an additional mechanism for generating non-stationarity, but rather the component that maintains the coherence of the fixed-threshold formulation. A nat-

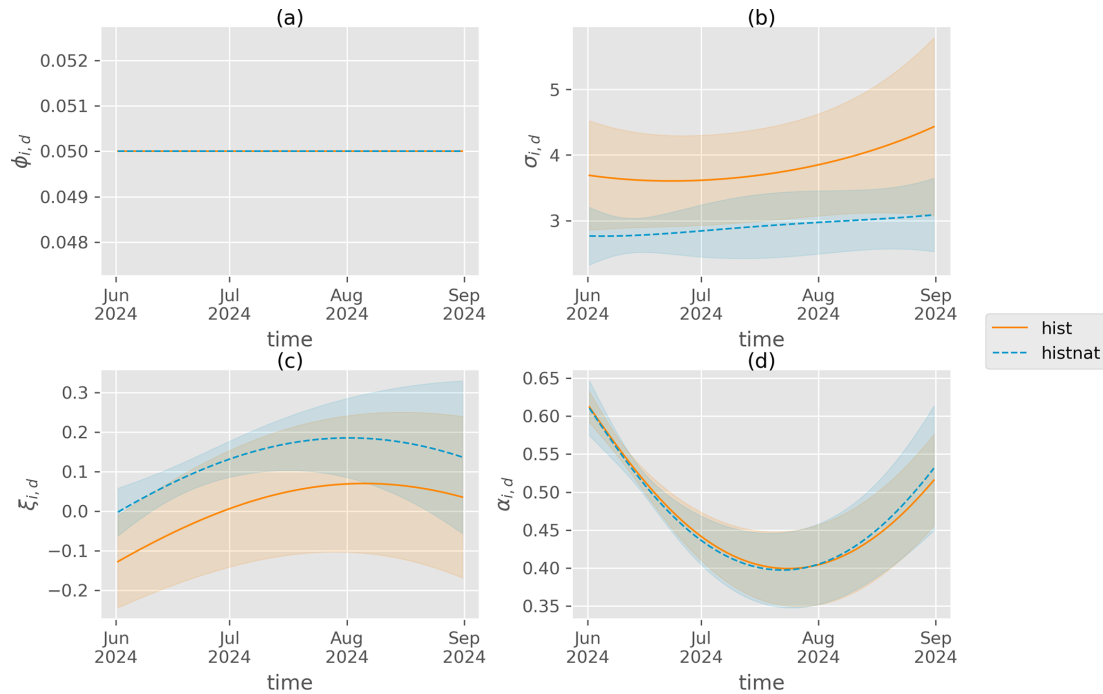


Figure A2. Same as Fig. A1 but for the model with time-varying threshold.

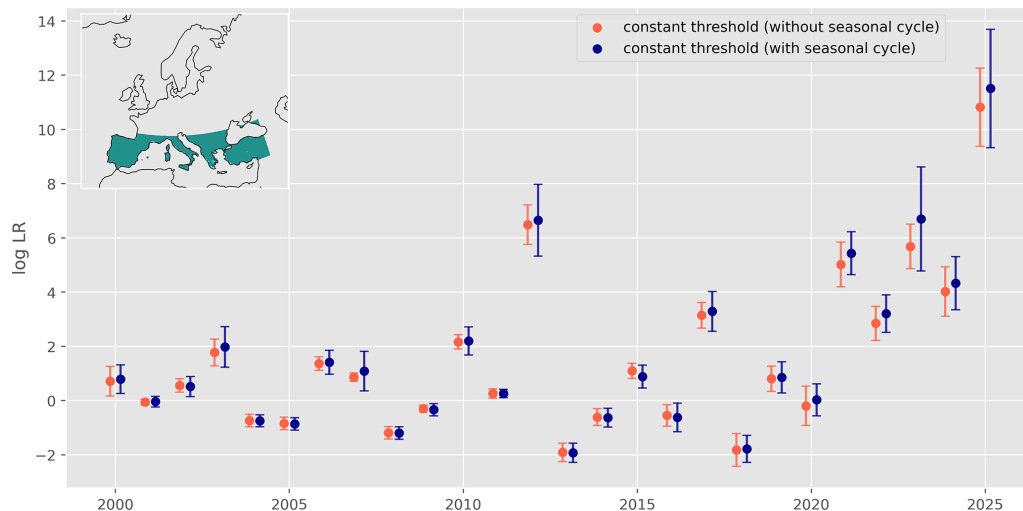


Figure A3. Logarithmic seasonal likelihood ratio including $\pm$ standard deviation for the southern European region, estimated according to Sect. 4.6, for the model with a constant threshold with and without modelling the seasonal cycle. For simplicity, no random effect model is used; instead, only the mean and standard deviation for the various models are displayed.

ural alternative is to allow the threshold to vary, for example as a time-varying quantile u_i obtained by quantile regression (Sect. 3.3.2). This would hold the exceedance probability constant by construction, shifting the non-stationarity into the threshold rather than removing it.

A2 Including seasonal cycle in parameter estimation

Although the seasonal cycle is removed when calculating the TPDM, a residual seasonal cycle remains in the EPIs, as can be demonstrated. One way to account for this is to model this residual seasonal cycle when fitting the parameters.

To this end, we relax the assumption of Eq. (18) and allow the parameters to vary within a season, so that the day index d reappears. We assume that the residual seasonal cycle can

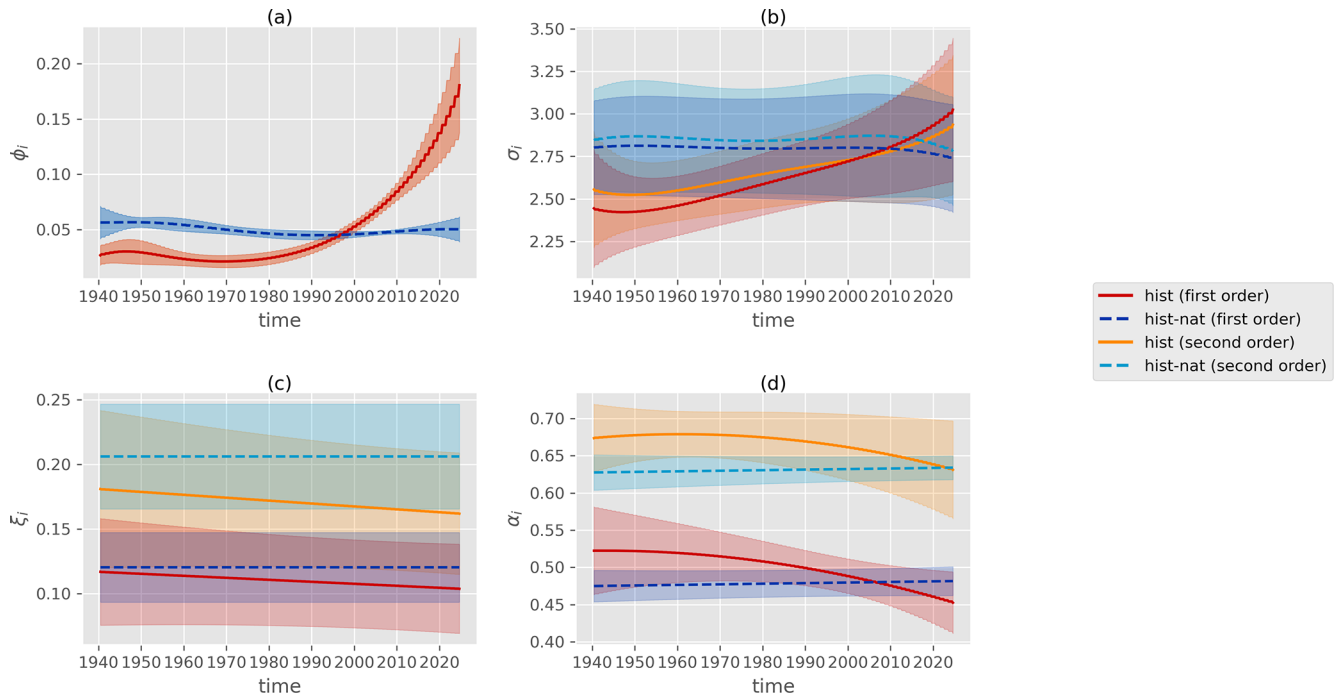


Figure A4. Same as Fig. 8 but now including the parameter estimates for the second-order Markov process.

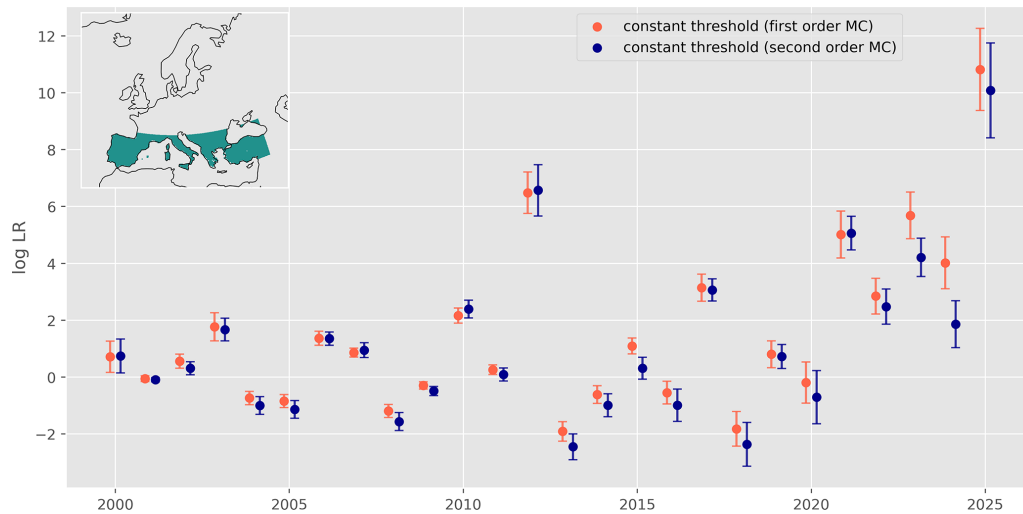


Figure A5. Effect of using a second-order Markov process: the mean and $\pm$ standard deviation of the seasonal likelihood ratios for the different climate models using a first- and a second-order Markov process. For the sake of simplicity, we did not apply the random effect model to combine the likelihood ratios of the different models.

be described by sine and cosine functions and write

$$\mathbf{h}_d = (\sin(2\pi \text{doy}_d/T_{\text{yr}}), \cos(2\pi \text{doy}_d/T_{\text{yr}}))^T, \quad (A6)$$

$$T_{\text{yr}} = 365.25,$$

where doy_d is the day of the year corresponding to day d of the season. While the Legendre polynomials $\mathbf{g}_i$ of Sect. 3.3 describe the variation between seasons, $\mathbf{h}_d$ describes the variation within a season. For the model with a constant thresh-

old (Sect. 4.1), the parameters then read

$$\text{logit}(\phi_{i,d}) = \beta_\phi^T \mathbf{g}_i + \delta_\phi^T \mathbf{h}_d, \quad (A7)$$

$$\sigma_{i,d} = \exp(\beta_\sigma^T \mathbf{g}_i + \delta_\sigma^T \mathbf{h}_d), \quad (A8)$$

$$\xi_{i,d} = \beta_\xi^T \mathbf{g}_i + \delta_\xi^T \mathbf{h}_d, \quad (A9)$$

$$\alpha_{i,d} = [1 + \exp(-\beta_\alpha^T \mathbf{g}_i - \delta_\alpha^T \mathbf{h}_d)]^{-1}, \quad (A10)$$

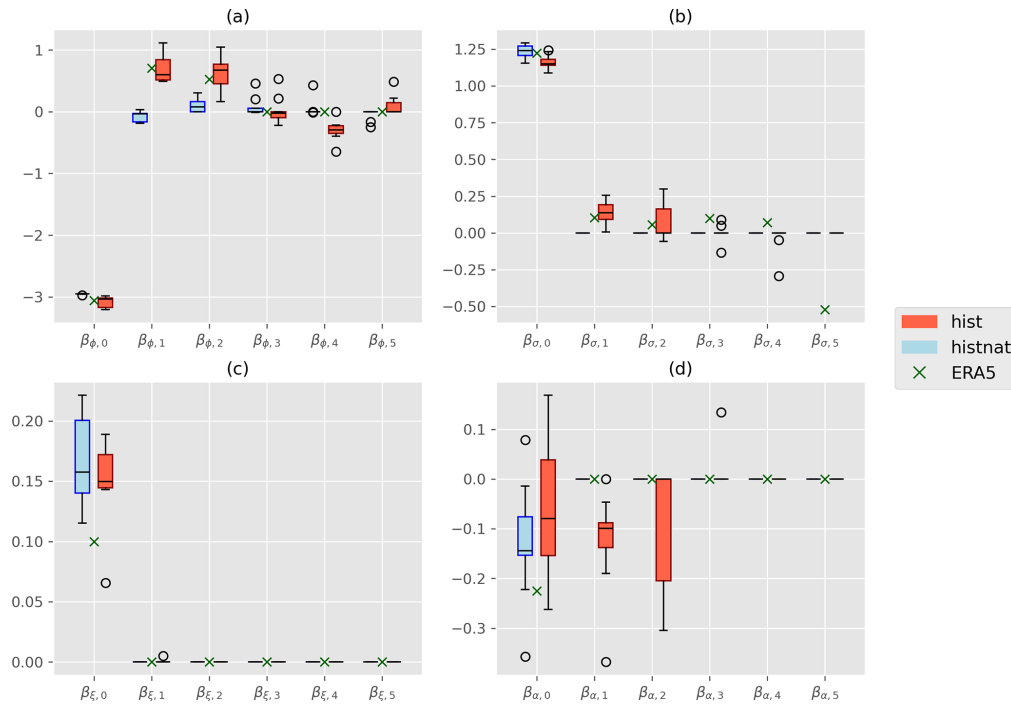


Figure A6. Same as Fig. 7 but for the northern European region.

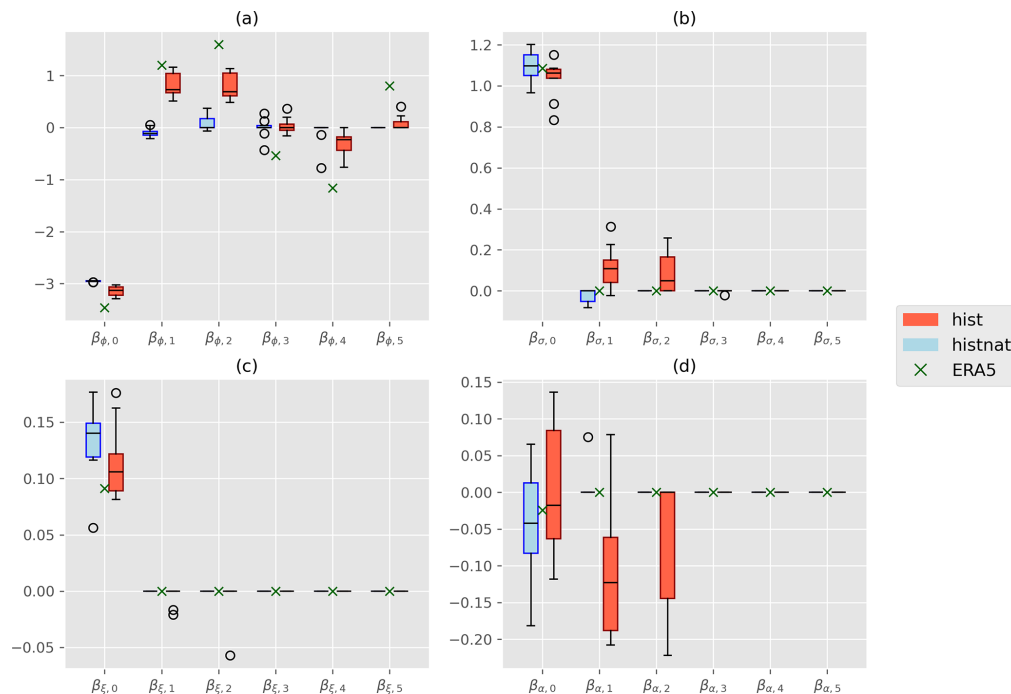


Figure A7. Same as Fig. 7 but for the central European region.

with $\delta_\phi, \delta_\sigma, \delta_\xi$ and $\delta_\alpha \in \mathbb{R}^2$ the additional coefficients to be estimated. For the model with a time-varying threshold, the procedure is analogous. For the sake of simplicity, we do not introduce a seasonal cycle into the quantile regression, i.e. $u_{i,d} \equiv u_i$ is retained.

The workflow for fitting the models is then the same as described in Sect. 4. Based on the maximum temperature example in the southern European region, the modelled seasonal cycle is plotted for models with constant and time-varying thresholds in Figs. A1 and A2.

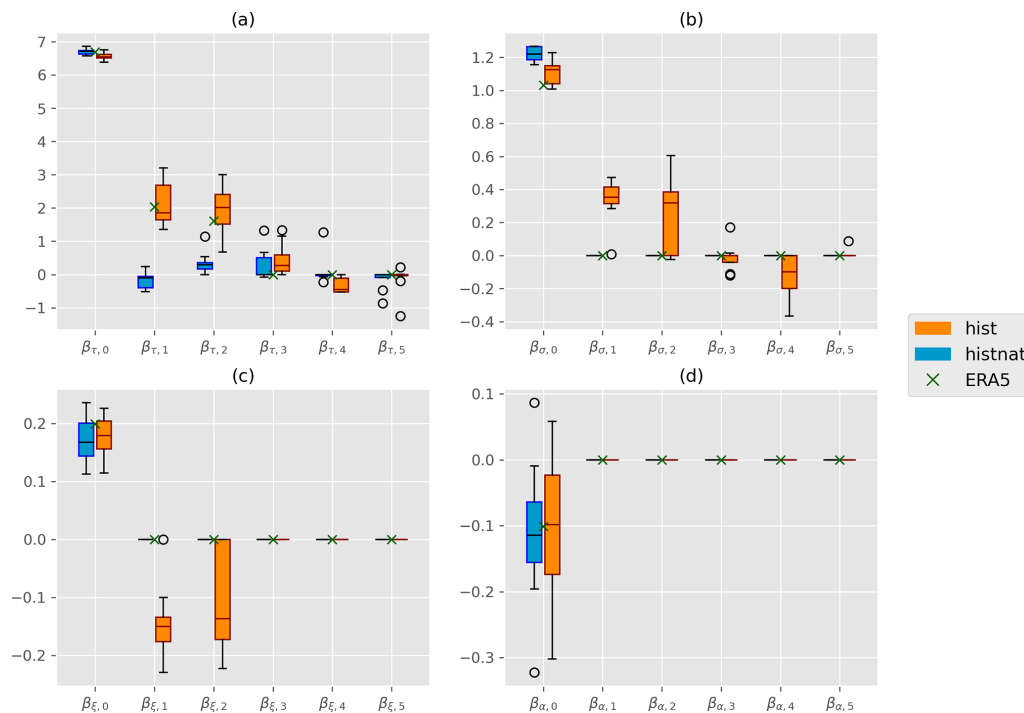


Figure A8. Same as Fig. 9 but for the northern European region.

The seasonal cycle is similar for both models and scenarios. This suggests that the seasonal cycle may not shift significantly in response to climate change. However, the seasonal cycle shown here should be interpreted with caution, as it is a residual cycle remaining after the seasonal cycle has been removed from the data to calculate the TPDM. To evaluate the effect of modelling the seasonal cycle on the estimation of the likelihood ratio, we calculate the seasonal likelihood ratio for each summer both with and without the seasonal cycle (as before).

The resulting likelihood ratios (Fig. A3) are similar for many years, with slightly higher likelihood ratios (and greater uncertainty) observed for the last five years when modelling the seasonal cycle. However, the effect of modelling the seasonal cycle is limited.

A3 Second-order Markov process

The effect of choosing a second-order Markov process will be shown based on maximum temperature and the southern European region, using a fitting procedure analogous to that of Sect. 4 and the logistic model as tail dependence function, i.e. Eq. (17) together with the likelihood of Eq. (16).

As can be seen in Fig. A4, the resulting scale parameters are in a similar range for the first- and second-order Markov process. The shape parameter estimates are slightly higher for the second-order Markov process in both scenarios, indicating a tendency towards greater variability in the data. As expected, the main difference between using a first- and

second-order Markov process lies in the dependence parameter. Using the second-order Markov process results in weaker dependence. The reason is that, for higher time lags, the variability of the atmosphere reduces the dependence.

When calculating the likelihood ratios for the second-order Markov process model (Fig. A5), a tendency towards smaller values can be seen for the second-order Markov process model over a large number of years. This may indicate the presence of dependencies with a lag greater than one day, such that the resulting likelihoods differ due to the dependence assumption of the first-order Markov process.

In order to assess whether the second-order Markov process gives a statistical improvement, the BIC (which is only one criterion pointed out by Smith et al. (1997) for the model comparison) is compared. For both scenarios and all climate models, the second-order Markov process model gives an improvement in terms of the BIC. However, it is not straightforward and not necessarily possible to compare the Markov processes of different orders, due to the use of the censored likelihood (Beirlant et al., 2004, Sect. 10.4.6). In terms of attribution, we can conclude that choosing between a first- and second-order Markov process has only a small effect on the resulting likelihood ratio, and strong attribution to the HIST scenario is also present when using a second-order Markov process.

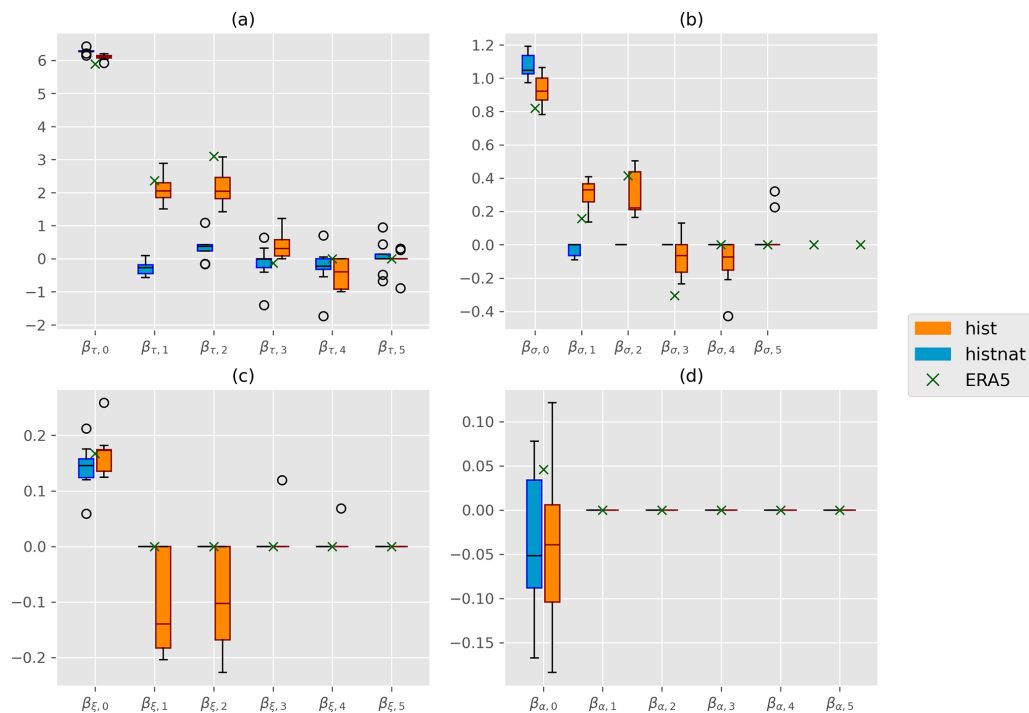


Figure A9. Same as Fig. 9 but for the central European region.

A4 Parameter estimates for the northern and central European regions

The parameters for the northern and central European regions are estimated in a similar way to those for the southern European region, as shown in Sect. 5. For the model with a constant threshold, the estimated coefficients are quite similar across different regions (Figs. A6 and A7).

As in the model with a constant threshold, the coefficients show high similarity for the time-varying threshold model (Figs. A8 and A9).

A5 Notation overview

Tables A1 and A2 summarise the notation used throughout this paper. The last column refers to the equation or section in which the respective symbol is introduced.

Table A1. Notation used in this study: indices, sample sizes, data, and distributions.

Symbol	Meaning	Introduced in
Indices		
i	Season (year) index; $i = 1$ corresponds to the summer of 1940	Eq. (5)
d	Day within a season	Eq. (5)
k	Extremal pattern and eigenvalue index (mode)	Sect. 2.2
s	Scenario index, $s \in \{0, 1\}$	Sect. 3.1
r	Ensemble member index	Sect. 4.1
c	Climate model index	Sect. 4.6
Sample sizes and model dimensions		
n_y	Number of seasons (summers 1940 to 2025)	Eq. (5)
n	Number of days per season ($n = 92$, JJA)	Eq. (5)
q	Number of land grid points entering the TPDM	Sect. 2.2
n_p	Number of leading extremal patterns retained ($n_p = 10$)	Eq. (1)
n_m	Number of climate models	Sect. 4.6
n_r	Number of ensemble members of a climate model and scenario	Sect. 4.1
n_B	Number of bootstrap repetitions	Sect. 4.5
K	Maximum degree of the Legendre polynomials ($K = 5$)	Sect. 3.3
N	Most recent season included in the accumulated Bayes factor	Eq. (34)
Data		
λ_k	k th eigenvalue of the TPDM	Sect. 2.2
$\eta_{k,i,d}$	Principal component of pattern k on day d of season i	Sect. 2.2
$y_{i,d}$	EPI on day d of season i ; $y_{i,d} := \text{EPI}_{i,d}$	Eq. (1)
$\mathbf{y}^{(i)}$	Time series of season i , $(y_{i,1}, \dots, y_{i,n})$	Eq. (5)
$\mathbf{Y}$	Complete data set, $(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)})$	Eq. (5)
$\mathbf{Y}_{\text{ERA5}}, \mathbf{y}_{\text{ERA5}}^{(i)}$	Corresponding quantities for the ERA5 reanalysis	Eq. (26)
Distributions and likelihoods		
F	Univariate marginal distribution function	Eq. (6)
G	Joint distribution function of consecutive observations	Eq. (10)
l_1, l_2, l_3	Univariate, bivariate, and trivariate densities of the model	Eqs. (7), (16)
$l(\cdot; \boldsymbol{\theta})$	Likelihood of a season or of the complete data set	Eqs. (7), (8)
V	Stable tail dependence function	Eq. (10)
V_1, V_2, V_{12}	Partial derivatives of V with respect to its arguments	Eq. (13)
$v_{i,d}$	Transformed margin; $1/v_{i,d}$ is standard Fréchet distributed	Eq. (9)

Table A2. Notation used in this study: model parameters, covariates, and attribution quantities. Parameters are given in their general form with both indices; the assumption under which the day index d is dropped is stated in Eq. (18).

Symbol	Meaning	Introduced in
Model parameters		
$u_{i,d}$	Threshold; constant u in the first model, time-varying u_i in the second	Eq. (6)
$\phi_{i,d}$	Probability of exceeding the threshold; ϕ_i in the first model, $\phi = 1 - \tau$ in the second	Eq. (6)
$\sigma_{i,d}, \xi_{i,d}$	Scale and shape parameters of the generalized Pareto distribution	Eq. (6)
$\alpha_{i,d}$	Dependence parameter of the logistic model, attached to the transition from day d to day $d + 1$	Eq. (11)
τ	Quantile level of the quantile regression ($\tau = 0.95$)	Sect. 3.3.2
$u_{\text{ERA5}}, u_{i,\text{ERA5}}$	Constant and time-varying threshold estimated from the ERA5 record	Eqs. (26), (28)
θ_1, θ_2	Marginal and dependence parameters; $\theta = (\theta_1, \theta_2)$	Eq. (7)
Covariates and regression coefficients		
x_i	Season index rescaled to $[-1, 1]$	Eq. (19)
g_i	Legendre polynomials of degree 0 to K , evaluated at x_i	Eq. (19)
h_d	Sine and cosine of the day of the year (residual seasonal cycle)	Eq. (A6)
β_ϕ, β_τ	Coefficients of the logistic and quantile regression	Eqs. (20), (23)
$\beta_\sigma, \beta_\xi, \beta_\alpha$	Coefficients of the non-stationary scale, shape, and dependence parameters	Eqs. (21), (22)
$\delta_\phi, \delta_\sigma, \delta_\xi, \delta_\alpha$	Coefficients of the residual seasonal cycle	Appendix A2
ρ_τ	Check function of the quantile regression	Eq. (25)
$\mathcal{R}_\tau$	Cost function minimised in the quantile regression	Eq. (24)
Attribution and combination of climate models		
m_0, m_1	Scenario without (HIST-NAT) and with (HIST) anthropogenic forcing	Sect. 3.1
p_0, p_1	Probability of an event under the two scenarios	Sect. 3.1
PR, LR	Probability ratio and likelihood ratio	Sect. 3.1
BF, BF(N)	Bayes factor, product of the seasonal likelihood ratios up to season n_y (resp. N)	Eqs. (4), (34)
$\hat{\mu}_c, \hat{\zeta}_c$	Estimate of the logarithmic likelihood ratio of climate model c and its bootstrap standard deviation	Sect. 4.6
w_c	Weight assigned to climate model c	Sect. 4.6
$\varsigma_{\text{nat}}, \varsigma_{\text{mod}}, \varsigma_{\text{tot}}$	Natural, model, and total variability of the combined estimate	Sect. 4.6
Q	Paule–Mandel statistic used to estimate ς_{mod}	Sect. 4.6

Code and data availability. This study used a selection of ERA5 and CMIP6 data stored at DKRZ. ERA5 is freely available via the Copernicus data store, CMIP6 data can be retrieved from the ESGF data nodes. An implementation of the extremal pattern index is available via the ExtrPatt R-package (<https://CRAN.R-project.org/package=ExtrPatt>, last access: 22 May 2026). The Python model code is provided, along with a Jupyter Notebook for the analysis and a Jupyter Notebook for visualisation. The extremal pattern index (EPI) output from the CMIP6 simulations and ERA5 reanalysis is available together with the code from Zenodo at <https://doi.org/10.5281/zenodo.20084245> (Meurer et al., 2026).

Interactive computing environment. The interactive computing environment is available at <https://doi.org/10.5281/zenodo.20084245> (Meurer et al., 2026). The necessary explanations to run the model code and visualise the results can be found in the README file. No further downloads are needed.

Author contributions. All of the authors contributed to developing the idea for this study (conceptualisation). PF secured the funding of this work. PF and SB supervised the work. SvS provided support for the TPDM calculation. Data analysis and visualisation were performed by PM. PM also led the writing of the original draft. PF contributed to the original draft. All authors contributed to reviewing and editing the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors thank Jonas Schröter and Erik Haufs for providing detailed background on the random effects model. We also thank the three anonymous reviewers for their valuable feedback, which has helped significantly improve the manuscript. The flowchart and notation tables were generated using the Claude AI assistant (Anthropic). The code used in this study was written entirely by the authors, without the use of AI tools. The version deposited at Zenodo differs from the original in that it has been revised to comply with the PEP 8 Python style guidelines. This reformatting was assisted by the Claude AI assistant (Anthropic).

Financial support. This research was funded within the Bundesministerium für Forschung und Technologie project ClimX-treme II – Module B under grant number FKZ 01LP2323A. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bm1159.

Review statement. This paper was edited by Seung-Ki Min and reviewed by three anonymous referees.

References

- Barnes, C., Clarke, B., Rantanen, M., Skålevåg, A., Ødemark, K., Kjellström, E., Vahlberg, M., Singh, R., Otto, F., Zachariah, M., Kew, S., Bergin, C., Vrkic, D., Hansson, L. B., Vikström, T., Hodgson, P., Norberg, L., Holten, T., Virkkunen, S., Kusterä, K., Aura, S., Haro, P., Drobina, D., and Sjölund, H.: Intense two-week heatwave in Fennoscandia hotter and more likely due to climate change, Centre for Environmental Policy, <https://doi.org/10.25560/122924>, 2025.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J.: Statistics of Extremes: Theory and Applications, Wiley, 522 pp., ISBN 0471976474, 2004.
- Bortot, P. and Tawn, J. A.: Models for the extremes of Markov chains, *Biometrika*, 85, 851–867, <https://doi.org/10.1093/biomet/85.4.851>, 1998.
- Chavez-Demoulin, V. and Davison, A.: Modelling Time Series Extremes, *REVSTAT-Stat. J.*, 10, 109–133, <https://doi.org/10.57805/revstat.v10i1.113>, 2012.
- Coles, S.: An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer-Verlag, London, ISBN 1-85233-459-2, 2001.
- Cooley, D. and Thibaud, E.: Decompositions of dependence for high-dimensional extremes, *Biometrika*, 106, 587–604, <https://doi.org/10.1093/biomet/asz028>, 2019.
- Copernicus Climate Change Service: Heatwaves contribute to the warmest June on record in western Europe, <https://climate.copernicus.eu/heatwaves-contribute-warmest-june-record-western-europe> (last access: 8 October 2025), 2025.
- Eastoe, E. F. and Tawn, J. A.: Modelling Non-Stationary Extremes with Application to Surface Level Ozone, *J. Roy. Stat. Soc. Ser. C*, 58, 25–45, <https://doi.org/10.1111/j.1467-9876.2008.00638.x>, 2009.
- Faranda, D., Guinaldo, T., Pastor, J. F., Alberti, T., and Khodayar, S.: Attribution of the 2025 Mediterranean Marine Heatwave to Climate Change Using Analogues, HAL open science [preprint], <https://hal.science/hal-05289765> (last access: 9 January 2026), 2025.
- Fawcett, L. and Walshaw, D.: Markov chain models for extreme wind speeds, *Environmetrics*, 17, 795–809, <https://doi.org/10.1002/env.794>, 2006.
- Friederichs, P.: Statistical downscaling of extreme precipitation events using extreme value theory, *Extremes*, 13, 109–132, <https://doi.org/10.1007/s10687-010-0107-5>, 2010.
- Friederichs, P., Göber, M., Bentzien, S., Lenz, A., and Krampitz, R.: A probabilistic analysis of wind gusts using extreme value statistics, *Meteorol. Z.*, 18, 615–629, <https://doi.org/10.1127/0941-2948/2009/0413>, 2009.
- Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6, *Geosci. Model Dev.*, 9, 3685–3697, <https://doi.org/10.5194/gmd-9-3685-2016>, 2016.
- Gulev, S., Thorne, P., Ahn, J., Dentener, F., Domingues, C., Gerland, S., Gong, D., Kaufman, D., Nnamchi, H., Quaas, J., Rivera, J., Sathyendranath, S., Smith, S., Trewin, B., von Schuckmann, K., and Vose, R.: Changing State of the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.004>, 2021.
- Hannart, A. and Naveau, P.: Probabilities of Causation of Climate Changes, *J. Climate*, 31, 5507–5524, <https://doi.org/10.1175/JCLI-D-17-0304.1>, 2018.
- Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., and Ghil, M.: Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events, *B. Am. Meteorol. Soc.*, 97, 99–110, <https://doi.org/10.1175/BAMS-D-14-00034.1>, 2016.
- Hegerl, G. C., Karl, T. R., Allen, M., Bindoff, N. L., Gillett, N., Karoly, D., Zhang, X., and Zwiers, F.: Climate Change Detection and Attribution: Beyond Mean Temperature Signals, *J. Climate*, 19, 5058–5077, <https://doi.org/10.1175/JCLI3900.1>, 2006.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hulme, M.: Attributing weather extremes to “climate change”: A review, *Prog. Phys. Geogr.*, 38, 499–511, <https://doi.org/10.1177/0309133314538644>, 2014.

- Huser, R., Opitz, T., and Thibaud, E.: Max-infinitely divisible models and inference for spatial extremes, *Scand. J. Stat.*, 48, 321–348, <https://doi.org/10.1111/sjos.12491>, 2021.
- IEA: Global Energy Review: CO₂ Emissions in 2021 Global emissions rebound sharply to highest ever level, <https://www.iea.org/reports/global-energy-review-co2-emissions-in-2021-2> (last access: 30 July 2025), 2021.
- Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cima-devilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanar, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets, *Earth Syst. Sci. Data*, 12, 2959–2970, <https://doi.org/10.5194/essd-12-2959-2020>, 2020.
- Jeon, S., Paciorek, C. J., and Wehner, M. F.: Quantile-based bias correction and uncertainty quantification of extreme event attribution statements, *Weather Clim. Extremes*, 12, 24–32, <https://doi.org/10.1016/j.wace.2016.02.001>, 2016.
- Jiang, Y., Cooley, D., and Wehner, M. F.: Principal Component Analysis for Extremes and Application to U.S. Precipitation, *J. Climate*, 33, 6441–6451, <https://doi.org/10.1175/JCLI-D-19-0413.1>, 2020.
- Kass, R. E. and Raftery, A. E.: Bayes Factors, *J. Am. Stat. Assoc.*, 90, 773–795, <https://doi.org/10.1080/01621459.1995.10476572>, 1995.
- Koenker, R.: Quantile regression, vol. 38 of *Econometric Society Monographs*, Cambridge University Press, ISBN 978-0-521-84573-1, 2005.
- Koenker, R. and Machado, J. A. F.: Goodness of fit and related inference processes for quantile regression, *J. Am. Stat. Assoc.*, 94, 1296–1310, 1999.
- Ledford, A. W. and Tawn, J. A.: Modelling Dependence within Joint Tail Regions, *J. Roy. Stat. Soc. Ser. B*, 59, 475–499, <https://doi.org/10.1111/1467-9868.00080>, 1997.
- Lemburg, A., Buschow, S., Dietz, V., Dillerup, I., Fischer-Frenzel, P., Friederichs, P., Grieger, J., Kraulich, F., Pfeiderer, P., Pinto, J. G., Reuter, L., Schröter, J., Szemkus, S., Ulbrich, U., and Vlachopoulos, O.: Analyse des europäischen Sommers 2025 mit Fokus auf Hitzeereignisse, Bericht des Forschungsnetzwerkes ClimXtreme, <https://doi.org/10.17169/refubium-51330>, 2026.
- Meurer, P., Buschow, S., Szemkus, S., and Friederichs, P.: Non-stationary time series attribution for heatwaves over Europe, Zenodo [code and data set], <https://doi.org/10.5281/zenodo.20084246>, 2026.
- Min, S.-K. and Hense, A.: A Bayesian Assessment of Climate Change Using Multimodel Ensembles. Part I: Global Mean Surface Temperature, *J. Climate*, 19, 3237–3256, <https://doi.org/10.1175/JCLI3784.1>, 2006.
- Min, S.-K., Hense, A., Paeth, H., and Kwon, W. T.: A Bayesian decision method for climate change signal analysis, *Meteorol. Z.*, 13, 421–436, <https://doi.org/10.1127/0941-2948/2004/0013-0421>, 2004.
- Nelder, J. A. and Wedderburn, R. W.: Generalized linear models, *J. Roy. Stat. Soc. Ser. A*, 135, 370–384, 1972.
- Otto, F. E. L., Barnes, C., Philip, S., Kew, S., van Oldenborgh, G. J., and Vautard, R.: Formally combining different lines of evidence in extreme-event attribution, *Adv. Stat. Clim. Meteorol. Oceanogr.*, 10, 159–171, <https://doi.org/10.5194/ascmo-10-159-2024>, 2024.
- Paciorek, C. J., Stone, D. A., and Wehner, M. F.: Quantifying statistical uncertainty in the attribution of human influence on severe weather, *Weather Clim. Extremes*, 20, 69–80, <https://doi.org/10.1016/j.wace.2018.01.002>, 2018.
- Paule, R. C. and Mandel, J.: Consensus Values and Weighting Factors, *J. Res. Nat. Bur. Stand.*, 87, 377–385, <https://doi.org/10.6028/jres.087.022>, 1982.
- Perkins-Kirkpatrick, S., Alexander, L., King, A., Kew, S., Philip, S., Barnes, C., Maraun, D., Stuart-Smith, R., Jézéquel, A., Bevacqua, E., Burgess, S., Fischer, E., Hegerl, G., Kimutai, J., Koren, G., Lawal, K., Min, S.-K., New, M., Odoulami, R., Patricola, C., Pinto, I., Ribes, A., Shaw, T., Thiery, W., Trewin, B., Vautard, R., Wehner, M., and Zscheischler, J.: Frontiers in attributing climate extremes and associated impacts, *Front. Clim.*, 6, <https://doi.org/10.3389/fclim.2024.1455023>, 2024.
- Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., and van Aalst, M.: A protocol for probabilistic extreme event attribution analyses, *Adv. Stat. Clim. Meteorol. Oceanogr.*, 6, 177–203, <https://doi.org/10.5194/ascmo-6-177-2020>, 2020.
- Pickands, J.: Multivariate extreme value distributions, in: *Proceedings of the 43rd Session*, vol. 49 of *Bulletin of the International Statistical Institute*, Buenos Aires, 859–879, 1981.
- Ramos, A. and Ledford, A.: A new class of models for bivariate joint tails, *J. Roy. Stat. Soc. Ser. B*, 71, 219–241, <https://doi.org/10.1111/j.1467-9868.2008.00684.x>, 2009.
- Ramos, A. and Ledford, A.: Estimation of the Extremal Index Function in Case of Asymptotically Independent Markov Chains and Its Application to Stock Market Indices, Springer Berlin Heidelberg, Berlin, Heidelberg, 89–96, ISBN 978-3-642-32419-2, https://doi.org/10.1007/978-3-642-32419-2_10, 2013.
- Seabold, S. and Perktold, J.: statsmodels: Econometric and statistical modeling with python, in: *9th Python in Science Conference*, <https://doi.org/10.25080/Majora-92bf1922-011>, 2010.
- Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dercozynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, 1513–1765, <https://doi.org/10.1017/9781009157896.013>, 2021.
- Seong, M.-G., Min, S.-K., and Zhang, X.: A Bayesian Attribution Analysis of Extreme Temperature Changes at Global and Regional Scales, *J. Climate*, 35, 8189–8203, <https://doi.org/10.1175/JCLI-D-22-0104.1>, 2022.
- Sippel, S., Mitchell, D., Black, M. T., Dittus, A. J., Harrington, L., Schaller, N., and Otto, F. E.: Combining large model ensembles with extreme value statistics to improve attribution

- statements of rare events, *Weather Clim. Extremes*, 9, 25–35, <https://doi.org/10.1016/j.wace.2015.06.004>, 2015.
- Smith, R. L., Tawn, J. A., and Coles, S. G.: Markov chain models for threshold exceedances, *Biometrika*, 84, 249–268, <https://doi.org/10.1093/biomet/84.2.249>, 1997.
- Stott, P. A., Stone, D. A., and Allen, M. R.: Human contribution to the European heatwave of 2003, *Nature*, 432, 610–614, <https://doi.org/10.1038/nature03089>, 2004.
- Szemkus, S. and Friederichs, P.: Spatial patterns and indices for heat waves and droughts over Europe using a decomposition of extremal dependency, *Adv. Stat. Clim. Meteorol. Oceanogr.*, 10, 29–49, <https://doi.org/10.5194/ascmo-10-29-2024>, 2024.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M.: Modelling Across Extremal Dependence Classes, *J. Roy. Stat. Soc. Ser. B*, 79, 149–175, <https://doi.org/10.1111/rssb.12157>, 2017.
- Wehner, M., Stone, D., Krishnan, H., AchutaRao, K., and Castillo, F.: The Deadly Combination of Heat and Humidity in India and Pakistan in Summer 2015, *B. Am. Meteorol. Soc.*, 97, S81–S86, <https://doi.org/10.1175/BAMS-D-16-0145.1>, 2016.
- Xoplaki, E., Ellsäßer, F., Grieger, J., Nissen, K. M., Pinto, J. G., Augenstein, M., Chen, T.-C., Feldmann, H., Friederichs, P., Gliksmann, D., Goulier, L., Haustein, K., Heinke, J., Jach, L., Knutzen, F., Kollet, S., Luterbacher, J., Luther, N., Mohr, S., Mudersbach, C., Müller, C., Rousi, E., Simon, F., Suarez-Gutierrez, L., Szemkus, S., Vallejo-Bernal, S. M., Vlachopoulos, O., and Wolf, F.: Compound events in Germany in 2018: drivers and case studies, *Nat. Hazards Earth Syst. Sci.*, 25, 541–564, <https://doi.org/10.5194/nhess-25-541-2025>, 2025.