

Inference for extreme spatial temperature events in a changing climate with application to Ireland

Dáire Healy¹, Jonathan Tawn², Peter Thorne³ and Andrew Parnell¹

¹*Hamilton Institute, Maynooth University, Ireland.* ²*Department of Mathematics and Statistics, Lancaster University, United Kingdom.* ³*ICARUS, Department of Geography, Maynooth University, Ireland.*

E-mail: 1daire.healy.2015@mumail.ie

To be read before The Royal Statistical Society at the second meeting on 'Statistical aspects of climate change' held online on Monday, 3 June 2024, Dr Castro-Camilo in the chair

Summary. We investigate the changing nature of the frequency, magnitude, and spatial extent of extreme temperatures in Ireland from 1942 to 2020. We develop an extreme value model that captures spatial and temporal non-stationarity in extreme daily maximum temperature data. We model the tails of the marginal variables using the generalised Pareto distribution and the spatial dependence of extreme events by a semi-parametric Brown-Resnick r -Pareto process, with parameters of each model allowed to change over time. We use weather station observations for modelling extreme events since data from climate models (not conditioned on observational data) can over-smooth these events and have trends determined by the specific climate model configuration. However, climate models do provide valuable information about the detailed physiography over Ireland and the associated climate response. We propose novel methods which exploit the climate model data to overcome issues linked to the sparse and biased sampling of the observations. Our analysis identifies a temporal change in the marginal behaviour of extreme temperature events over the study domain, which is much larger than the change in mean temperature levels over this time window. We illustrate how these characteristics result in increased spatial coverage of the events that exceed critical temperatures.

Keywords: Climate change, generalised Pareto distribution, extreme values, heatwaves, missing data, non-stationarity, Pareto processes, spatial extremes, temperatures.

1. Introduction

The Intergovernmental Panel on Climate Change, IPCC (2021, Chapter 11), reports an observable change in extreme weather and climate events since around 1950. Characterisation of extreme temperature events is crucial for societal development, for estimating risks, and to enable the mitigation of their effects for many sectors, e.g., healthcare, economic growth, agricultural disruption, and infrastructure. Brown et al. (2008) observed a warming of both maximum and minimum temperatures since 1950 for most regions indicating an increasing number of warm days and longer heatwaves.

In Ireland, changing extreme temperature behaviour has also been observed. McElwain & Sweeney (2007) found that a warming of both maximum and minimum temper-

ature observations occurred for all sites over 1961–2005. O’Sullivan et al. (2020) showed that the frequency of extreme temperature events for County Dublin has increased over the period 1981–2010. Both these approaches considered only the marginal behaviour of extremes. To the best of our knowledge, the only modelling of spatial extreme temperature events in Ireland is by Huser & Wadsworth (2022). They used a gridded Irish temperature data set, which has the potential to be over-smooth relative to the observed process, and fitted their model to these data under the assumption of stationarity over time and space. Under similar stationarity assumptions, Fuentes et al. (2013) and Cebrián et al. (2022) analyse extreme spatial temperatures for other locations.

We are interested in developing a model which captures the temporal evolution of spatial extreme temperature events over Ireland. This involves modelling how the marginal distributions vary over space, accounting for spatial dependence within extreme events, and modelling how these two elements vary over time. Our focus is on modelling extreme value data, however, for a spatial process, an extreme event can consist of abnormally high values in part of the region and typical values elsewhere (Davison et al. 2012).

Observational extreme event data are sparse and so they need to be used efficiently. The traditional statistical approach is to model these data with powerful probabilistic characterisations from extreme value theory. This theory provides a parsimonious asymptotic justification for extrapolation which enables us to describe the properties and behaviour of events which are more extreme than those previously observed. The theory by itself, however, will not provide information on how to spatially interpolate over heterogeneous geography or how to account for when the characteristics of complex spatial events change over time. Here we propose a novel approach to address these issues which exploits the physical knowledge of the climate processes from information given by fine-scale climate model data. We review existing extreme value methods for spatial and temporal processes and outline our strategies for using climate model data.

The theory of univariate extreme values for stationary processes (Leadbetter et al. 1983), and the associated statistical models (Coles 2001), fully determine a simple parametric distributional family, namely the generalised Pareto distribution (GPD), as the non-degenerate limit distribution for the normalised excesses of a threshold, as that threshold tends to the upper endpoint of the marginal distribution. The GPD has been very widely used in diverse applications since the exposition of Davison & Smith (1990). To deal with non-stationarity, the GPD parameters have been allowed to change smoothly with covariates, initially using fully parametric regression models and more recently with a range of different non-parametric smoothing methods (Chavez-Demoulin & Davison 2005, Youngman 2022). We model the upper tails of the marginal distribution of the temperature process using the GPD, with covariates selected from space, time, information from climate models, and established measures/causes of climate change.

The most established approach to spatial extreme modelling uses max-stable process models (de Haan & Ferreira 2006). These processes are the class of non-degenerate limiting distributions of linearly normalised site-wise maxima, typically fitted to annual maxima data observed at each site in a set of locations over years. They are a natural extension of univariate block maximum limit theory, and so have generalised extreme value distributions for their margins. Brown & Resnick (1977) introduced a widely used subclass of these models, derived from Gaussian random fields, known as Brown-Resnick

processes, with Davis et al. (2013) applying this model to spatio-temporal data.

The major problem with max-stable models is that they do not model, and so cannot capture, spatial patterns of observed extreme events. Inference using these models can lead to biased estimation of dependence (Huser & Wadsworth 2022). A recent development in the modelling of spatial threshold exceedances is the generalised r -Pareto process (Thibaud & Opitz 2015, de Fondeville & Davison 2018, 2022). Generalised r -Pareto processes, like max-stable processes, exhibit a strong form of dependence, known as asymptotic dependence (defined in Section 4.2) between all sites. This implies that for an event which is extreme at any location in space, there is a positive probability that this event will be extreme everywhere else in the spatial domain. For processes over spatial domains that are large relative to the scale of the spatial dependence of the process, this is an unrealistic assumption. More flexible spatial models, building from those in Wadsworth & Tawn (2022) are discussed in Healy et al. (2023, Section 7.1).

In Section 4 we define these extremal dependence properties precisely and provide evidence that generalised r -Pareto processes are suitable for daily maximum temperatures over Ireland. We identify extreme spatial fields, based on observations at d sites, as those which exceed a sufficiently high threshold for a risk function $r : \mathbb{R}^d \rightarrow \mathbb{R}_+$. We model these fields as realisations of a Brown-Resnick Pareto process, which is closed under marginalisation (Engelke & Hitz 2020), an important property given the time-varying level of missing temperature data in our application.

We have observational daily maximum data for a network of 182 Irish temperature stations, with only 38% of these having more than 30 years of data due to differential operational periods and quality controls, which is further compounded by spatial selection bias in the station locations. We also have a rich spatio-temporally complete data set generated from a climate model, giving daily maximum temperatures over 56 years on a fine grid over Ireland. These climate model data are not conditioned on the observed weather, so their values on any given day have no correlation to the observed data, but they have similar probability distributions to the observed data at the associated sites and time of the year. The climate model data have no missing values or location biases; they are on a dense regular grid and incorporate the impact of known geophysical structures on the temperature process.

Although it may be tempting to analyse the simpler climate model data than the observed station data, climate models involve some abstraction of the physical processes they model and so tend to under-predict extreme events in magnitude and to over-estimate dependence owing to the climate model's smoothness over space and time, see Sections 3 and 4. So direct analysis of the climate data is not ideal but clearly, they offer vital additional information to the observational data. Various attempts have been made to downscale the climate model data to produce a proxy for the observed data which gives a spatial and temporally complete data set, e.g., Maraun et al. (2017), with the focus to date being on assessing marginal features. We prefer to let the observational data stand for themselves, particularly in relation to the information they provide about temporal non-stationarity as the climate model data have trends determined by the climate model configuration that is an imperfect representation of the real-world processes.

The novelty of our paper is achieved through the use of state-of-the-art extreme value methods for marginal distributions, spatial dependence, and temporal non-stationarity

which collectively exploit knowledge from climate science and through the use of appropriate metrics for describing changes in spatial extreme events. Our use of climate science relies heavily on how our inference for the observational temperature data leverages core information from the climate model data, i.e., parameter estimates (within sample quantiles and GPD parameters) over space, and through our careful assessment of, and sensitivity to, the effects of the inclusion of various climate-based covariates.

The paper is organised as follows. Section 2 details the observational and climate model data used. Sections 3 and 4 describe the marginal and dependence modelling of the process respectively, in each case accounting for their changing behaviour over time. In Section 5 we use the model to explore how the properties of spatial extreme events have changed over time. Conclusions and a broader discussion are given in Sections 6 and 7. All our code and instructions on how to access the data are available on GitHub[†].

2. Data

We start with a note on nomenclature since we use multiple data sets which differ in their structure and use. We use the term ‘station data’ to refer to data taken directly from weather stations. These are irregularly located and suffer from missing values. The term ‘climate model data’ refers to physics-based simulations of the weather system which are run on high-resolution grids and do not aim to match individual weather events, rather they model the spatio-temporal dynamics of the weather system. Finally, ‘observation-based data products’ are gridded data sets which arise from some form of statistical or physical interpolation of station data.

2.1. Station data

Our daily maximum temperature data comprise 182 Irish temperature stations compiled from two sources, the locations of which are shown in Figure 1. For the Republic of Ireland, data for 151 stations came from Met Éireann’s archive[‡]. Data for 31 Northern Ireland sites were obtained through the CEDA archive (Met Office 2012). Collectively, these data have many missing values, with availability of data decreasing further back in time. We have more than twice the data from the 1950s than the 1940s, with all stations pre-1950 (except one) being coastal. We have 56% of daily values observed in the last 30 years, and only 0.53% observed before 1942. No single day has data for every station. The average span of data for each station is about 30 years, with observations ranging from 1931 to 2022. The sites with the most data tend to be located near the coast reflecting historical and present-day observational priorities.

Our interest is in extreme warm temperatures in Ireland, so we restrict our analysis to data from the summer (JJA) and report findings for the years 1942–2020. This choice of season is supported by the finding that 93% of all the days with temperatures above the 99% site-wise marginal quantile occur in summer, with this proportion increasing with the level of marginal quantile. As exceedances are reasonably spread across the summer,

[†]github.com/dairer/Extreme-Irish-Summer-Temperatures.

[‡]Copyright Met Éireann. Source: met.ie/climate/available-data/historical-data.

we assume that within each summer the process is temporally stationary. Exploratory analysis supporting both of these choices is reported in Healy et al. (2023, Section 2.3).

2.2. *Climate model data*

Climate models are mathematical representations of the physical processes driving weather and climate and represent our best understanding of these natural phenomena (Giorgi 2019). Climate models are broadly run on two scales; large-scale global general circulation models (GCMs) and finer-scale regional climate models (RCMs). An RCM is informed by the GCM at its boundary. Crucially, climate model data do not correlate with the observed time-evolution of weather, rather they have probabilistic structures that reflect plausible weather sequences which could occur. They are typically designed to investigate the effect of potential external forcing on the climate system by, e.g., increases in greenhouse gas concentrations arising from anthropogenic activities. We identify and exploit physical and topographical features in the output of these models and use them to adjust for spatial and temporal bias in the observed data set.

We use RCMs for their detailed topographical information and their physical description of temperature processes. When relying only on climate models to understand extreme weather it is common to consider several GCM/RCM combinations, each with different initial conditions and future climate scenarios. This choice will have limited impact for us as we use station data to describe the magnitude and frequency of temperature events, and the climate model data only to inform non-temporal features. We use data from the CLMcom-CLM-CCLM4-8-17 RCM combined with the ICHEC-EC-EARTH GCM. Specifically, we have daily maximum temperatures over a 56-year period (created using the atmospheric climate drivers from 1950 to 2005) on a regular grid of 558 points over Ireland (corresponding to a 0.11^2 degree resolution). Figure 1 (right) shows the values for the day with the largest average temperature over Ireland in this data set. This plot illustrates two features which we exploit in Sections 3 and 4 respectively. Firstly, the RCM provides much greater spatial coverage in the interior of Ireland than the stations in Figure 1 (left). Secondly, extreme temperature events can be very widely spread across Ireland, since even the sites with the lowest values on this day have temperatures in their marginal distributions' upper tails.

2.3. *Covariates: Observation-based data products*

To model temporal non-stationarity of extreme temperature data it is common to use time as the sole covariate, although this will have severe limitations outside the range of the data as potential emission scenarios diverge. Instead, we use the time-varying covariates that climate scientists believe best represent changes in observed mean temperatures. These are predictable into the future under different emission scenarios. We use two covariates; smoothed monthly average temperature anomalies for the global average M_t^G and for the grid box over Ireland M_t^I , from the observation-based data product HadCRUT5. See Healy et al. (2023, Section 2.1) for details and plots of the covariates. Over 1942–2020, both covariates increase by approximately 1°C , with the change accelerating.

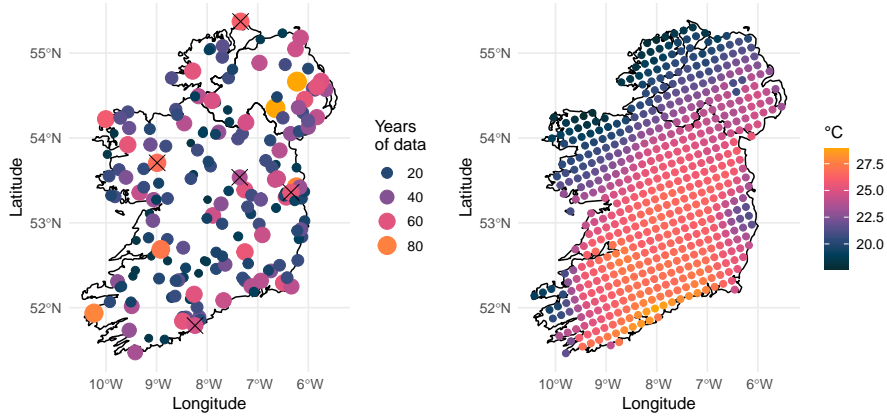


Fig. 1. Ireland data locations: (left) station data sites, with the amount of data indicated by colour and size. Sites marked with an 'X' correspond to Malin Head (North), Roches Point (South), Phoenix Park (East), Claremorris (West) and Mullingar (Centre); (right) climate model data from MOHC-HadREM3-GA7-05 showing a generated extreme temperature event.

Our exploratory analysis, using spline-based models, identified that the shortest distance to the coast, for each site, was a potential descriptor of the change of temporal trends. We define this covariate by $C(\mathbf{s})$, for each site \mathbf{s} . We consider the covariate of the annual CO_2 emissions ($\text{CO}_{2,t}$) for Ireland; see Healy et al. (2023, Section 2.2). Since there is strong collinearity present in the collective covariates $\mathbf{z}_t := (t, M_t^G, M_t^I, \text{CO}_{2,t}, C(\mathbf{s}))$ we only use one of the time-varying covariates in each model.

3. Marginal models

3.1. Overview and Strategy

Let $X_o(t, \mathbf{s})$ denote the observed station data comprising summer maximum daily temperature at time t and site \mathbf{s} , and let $X_c(t, \mathbf{s})$ be the equivalent process from the climate model data. We assume temporal stationarity within each year for each site and each process. Here $t \in \mathbb{N}$ indexes summer days within and across years and $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^2$, where \mathcal{S} denotes Ireland, with \mathbf{s} corresponding to the vector of latitude and longitude. We have data on the two processes at $\mathcal{S}_o \subset \mathcal{S}$ and $\mathcal{S}_c \subset \mathcal{S}$ and at times \mathcal{T}_o and \mathcal{T}_c respectively. For \mathcal{T}_o we also have missing data for some of the stations as discussed in Section 2. Throughout we use the subscripts to identify the type of process, though the indexing is dropped when discussing methods which apply similarly to both processes.

In Section 3.2, we propose a spatial and temporal quantile regression model for the data to derive an estimate of the distribution function of $X(t, \mathbf{s})$. As the tails of this distribution are particularly important to model well, we introduce a threshold $u(\mathbf{s})$ which is fixed over time but varies in space, above which we replace the quantile model with the generalised Pareto distribution (GPD) parametric model with temporal and spatial covariates. The justification for our choice of a constant threshold over time is discussed in Section 7.2. Novelty in our approach comes from using estimates from the

X_c process to infer features of the X_o process, which is appealing as $|\mathcal{S}_o| \ll |\mathcal{S}_c|$ and $|\mathcal{T}_o| \ll |\mathcal{T}_c|$ for most sites. Given all these considerations, we need to estimate thresholds, the temporally varying marginal distributions over \mathcal{S} for below the thresholds, the GPD parameters for above the thresholds, and to do this for both the X_c and X_o processes.

When analysing the climate model data we need to account for the following issues. First, our use of these data is to improve our spatial mapping and to overcome issues of missing data in the analysis of the observational data. Second, climate model data can show different time dynamics from that of the observed process since they are based on incomplete physics and forcing detail. We want our analysis to be robust to temporal non-stationary aspects of the climate model data, so we assume that $X_c(t, \mathbf{s})$ are temporally stationary in our analyses. As the trend in the climate model data is 4% of the variation in the data at each site, this is not too restrictive an assumption.

3.2. Modelling the body of the distribution

Given the issues raised in Section 3.1 about $X_c(t, \mathbf{s})$, we take the following simple approach for the inference of its distribution. For site \mathbf{s} we estimate the τ th quantile of X_c by using the empirical sample quantile for the climate model data at that site alone; we denote this estimator by $q_c^{(\tau)}(\mathbf{s})$. We use this approach for all τ over the range $[0.01, 0.99]$. This estimator is reliable as we have sufficient data (5152 days with none missing) and, due to the climate model data being numerical model output, their spatial variation is very smooth, so statistical spatial smoothing methods are more likely to induce bias than to enhance the analysis through information sharing.

For the analysis of $X_o(t, \mathbf{s})$, the issues of spatial sparsity of stations, limited data, varying periods of records of stations, and the need to account for temporal variations, lead to a different approach than for $X_c(t, \mathbf{s})$. We follow the approaches of Yu & Moyeed (2001), Youngman (2019) and that of the R package `evgam` (Youngman 2022) by using the asymmetric Laplacian distribution (ALD) for quantile regression to estimate a range of spatially and temporally varying τ th quantiles, $q_o^{(\tau)}(t, \mathbf{s})$, for a grid of $\tau \in [0.01, 0.99]$, for all $t \in \mathcal{T}$ and $\mathbf{s} \in \mathcal{S}$. The density function of the ALD_τ is

$$f_{\text{ALD}_\tau}(y; q, \psi) = \tau(1 - \tau)\psi^{-1} \exp\{-\rho_\tau(y - q)\psi^{-1}\}, \quad y \in \mathbb{R}, \quad (1)$$

where $\rho_\tau(z) = \{\tau - I(z < 0)\}z$ is the check function, $q \in \mathbb{R}$ is a location parameter, corresponding to the τ th quantile of interest, and $\psi > 0$ is a scale parameter. We assume that q and ψ vary smoothly over \mathcal{S}_o and \mathcal{T}_o .

For estimating $q_o^{(\tau)}(t, \mathbf{s})$ and $\log\{\psi_o^{(\tau)}\}$ we consider not just t and \mathbf{s} as covariates but also incorporate as potential covariates the associated quantile from the climate model data $q_c^{(\tau)}(\mathbf{s})$ and each of the climate-based covariates of Section 2.3. The former provides richer spatial information that is not captured in the observational data set, and the latter gives a causal set of time-varying covariates. Details of the analysis using these models are presented in Section 3.5.

To provide estimates for all τ , we fit this model separately for a grid of τ values and use a cubic interpolation spline for each \mathbf{s} to give a continuous estimate over $0.01 \leq \tau \leq 0.99$ for $\hat{q}_o^{(\tau)}(t, \mathbf{s})$. We kept the grid of τ values relatively coarse to avoid issues of quantile

estimates crossing. This gives us an estimate of the distribution function of $X_o(t, \mathbf{s})$ as

$$F_{o,t,\mathbf{s}}(\hat{q}_o^{(\tau)}(t, \mathbf{s})) := \Pr\{X_o(t, \mathbf{s}) < \hat{q}_o^{(\tau)}(t, \mathbf{s})\} \approx \tau \text{ for all } \tau, \text{ for all } \mathbf{s} \in \mathcal{S}. \quad (2)$$

This model provides estimates for all quantiles for any $\mathbf{s} \in \mathcal{S}_c$, not just \mathcal{S}_o , and at all times where we have the covariates, e.g., not just for $t \in \mathcal{T}_o$. At each site \mathbf{s} , below the threshold $u(\mathbf{s})$ (defined in Section 3.3) we use this distributional model $F_{o,t,\mathbf{s}}$.

3.3. Modelling the tails of the distribution

It is well known that quantile regression, and hence the ALD model, is unreliable for estimating quantiles in the tails of the distribution and provides no means to extrapolate beyond observed data. As the upper extremes of the distribution of $X_o(t, \mathbf{s})$ are important to us, we chose an alternative model based on extreme value methods. This enabled us to produce a model that is continuous over all t and \mathbf{s} , with the extreme value model being used above a high threshold, and the ALD model describing data below.

One option is to have a threshold, $u(t, \mathbf{s})$, that varies over time and space computed for a given quantile, e.g., $u(t, \mathbf{s}) = \hat{q}_o^{(\tau)}(t, \mathbf{s})$, for a choice of τ . However, in extreme value inference, it is well-known that it is difficult to objectively select a threshold or to account for the uncertainty in that choice (Northrop et al. 2017). Here it is the temporal change in extreme events which is of most interest, and this trend is small relative to other sources of variations in the data. We do not select a time-varying threshold using information from the body of the distribution, as this may bias results for the extremes. Instead, we choose the threshold to be $u(\mathbf{s})$, i.e., only varying over space.

To reduce subjectivity, for each site and for both X_c and X_o processes, we use a common exceedance probability for the fixed-over-time threshold. Based on the use of standard extreme value threshold selection methods for stationary processes (Coles 2001) which we applied at each site/process separately, we identified that the 90% quantile was suitable. For the reasons discussed in Sections 3.1 and 3.2, we use the site-specific 90% empirical sample quantile for $u_c(\mathbf{s})$ but a model-based estimate for $u_o(\mathbf{s})$. Specifically, we fit the model for density (1) with $\tau = 0.9$ with the location parameter structured as $u_o(\mathbf{s}) := q_o^{(0.9)}(\mathbf{s}) = \beta_0 + \beta_1 u_c(\mathbf{s})$, with (β_0, β_1) parameters. Thus the climate model data provides a means by which the spatially varying threshold u_o for the observed data X_o can be estimated. This routine aims to overcome the data quality limitations and to provide estimates for all $\mathbf{s} \in \mathcal{S}_c$.

For a given threshold there are two remaining elements required to model the extremes, namely the threshold exceedance probability $\lambda_o(t, \mathbf{s})$ and the distribution H of the excesses of the threshold (Chavez-Demoulin & Davison 2005). We consider these in turn. We estimate $\lambda_o(t, \mathbf{s})$ from the model for the body of the distribution, using the set of estimated distribution functions (2). Specifically,

$$\lambda_o(t, \mathbf{s}) = 1 - \tau_{u_o}(t, \mathbf{s}),$$

where $\tau_{u_o}(t, \mathbf{s})$ is the value of τ , at time t , which makes $\hat{q}_o^{(\tau)}(t, \mathbf{s}) = u_o(\mathbf{s})$. If there is no temporal non-stationarity in $X_o(t, \mathbf{s})$ then by construction of the threshold $u_o(\mathbf{s})$, if that was correct, we would have $\lambda_o(t, \mathbf{s}) = 1 - \tau = 0.1$ across $t \in \mathcal{T}_o$ and $\mathbf{s} \in \mathcal{S}_o$.

For each site \mathbf{s} we assume that excesses of the threshold $u(\mathbf{s})$ follow a generalised Pareto distribution (GPD); see Pickands (1975) and Davison & Smith (1990) for the probabilistic justification and properties. The GPD(σ, ξ) has distribution function

$$H(y; \sigma, \xi) = 1 - (1 + \xi y / \sigma)_+^{-1/\xi} \quad (3)$$

for $y > 0$, with a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\sigma > 0$, with the notation $x_+ = \max(x, 0)$, and $\xi = 0$ is obtained by taking the limit as $\xi \rightarrow 0$. When $X(t, \mathbf{s}) > u(\mathbf{s})$ the threshold excess, $Y(t, \mathbf{s}) = X(t, \mathbf{s}) - u(\mathbf{s})$, is taken to be distributed as

$$Y(t, \mathbf{s}) \sim \text{GPD}(\sigma(t, \mathbf{s}), \xi), \quad (4)$$

where we discuss the choice of models for $\sigma(t, \mathbf{s})$ below, and where the shape parameter is taken to be constant over time and space. This choice of homogeneity for the shape parameter for both X_o and X_c (i.e., values ξ_o and ξ_c respectively) is supported by exploratory analysis in Healy et al. (2023, Section 3.2), but it is typical in GPD modelling as there is limited evidence against this in almost all applications, and for the pragmatic reason that even a homogeneous value is difficult to estimate well. Furthermore, this choice reduces the risk of parameter identifiability problems (Davison et al. 2012).

Combining $\lambda_o(t, \mathbf{s})$ with the model for H gives our overall marginal distributional model $F_{o;t,\mathbf{s}}$ for the upper tail of $X_o(t, \mathbf{s})$. Specifically for $y > u_o(\mathbf{s})$ we have

$$F_{o;t,\mathbf{s}}(y) = 1 - \lambda_o(t, \mathbf{s})[1 - H(y - u_o(\mathbf{s}); \sigma_o(t, \mathbf{s}), \xi_o)]. \quad (5)$$

As with the estimation of the quantiles below $u_o(\mathbf{s})$ for X_o , we use information from X_c to provide a spatial covariate for $\sigma_o(t, \mathbf{s})$. As discussed in Section 3.1, we aim to learn about temporal non-stationarity exclusively from the observational data, so only information about the spatial variation of the marginal tail distribution is taken from X_c . We fit a model of the form

$$Y_c(t, \mathbf{s}) \sim \text{GPD}\{\sigma_c(\mathbf{s}), \xi_c\}, \quad (6)$$

for the excesses of $u_c(\mathbf{s}) = q_c^{(0.9)}(\mathbf{s})$ and we keep σ_c constant over t . When modelling the climate model data we believe we have sufficient observations and spatial consistency, from their generation, to treat $\sigma_c(\mathbf{s})$ as site-specific, i.e., not imposing any spatial smoothness on the GPD scale parameters over $\mathbf{s} \in \mathcal{S}$. Clearly, it would be wrong to smooth well-estimated parameters spatially if we want to capture the relevant geophysical features of the climate system.

Full likelihood inference is not possible as any realistic model for the station data is likely to be highly complex, requiring spatial and temporal dependence of the data to be modelled. Instead, we use a pseudo-log-likelihood

$$p\ell_c(\sigma_c(\mathbf{s}) : \mathbf{s} \in \mathcal{S}_c; \xi_c) = \sum_{\mathbf{s} \in \mathcal{S}_c} \left[\sum_{t \in \mathcal{T}_c} \log h(y_{c,t,\mathbf{s}}; \sigma_c(\mathbf{s}), \xi_c) \right], \quad (7)$$

constructed under the false assumption of spatial and temporal independence, with h being the density function of the GPD. This inference approach is commonly used, e.g.,

by Davison et al. (2012). The maximisation of this function can be broken down into a series of 1-dimensional optimisations by alternating the maximisation over ξ_c with the scale parameters fixed, and then exploiting the partition of the function $p\ell_c$ with respect to \mathbf{s} when maximising over each $\sigma_c(\mathbf{s})$ in turn whilst treating ξ_c as constant. Iterating in this way until convergence is achieved gives estimated values $\{\hat{\sigma}_c(\mathbf{s}); \mathbf{s} \in \mathcal{S}_c\}$ and $\hat{\xi}_c$.

Next, we model the extreme observational data excess above the threshold, $u_o(\mathbf{s})$, denoted by $Y_o(t, \mathbf{s})$. The generic form of each of the models we consider is

$$Y_o(t, \mathbf{s}) \sim \text{GPD}(\sigma_o\{\mathbf{z}_t, \hat{\sigma}_c(\mathbf{s})\}, \xi_o), \quad (8)$$

where we model $\log \sigma_o\{\mathbf{z}_t, \hat{\sigma}_c(\mathbf{s})\}$ as either a parametric linear model of $\log \hat{\sigma}_c(\mathbf{s})$ and the covariates \mathbf{z}_t (defined in Section 2.3) or via a GAM formulation. We denote the parameters of σ_o by $\boldsymbol{\theta}$. As with the inference for the climate model data we have to use a pseudo-log-likelihood, constructed under the false assumption of spatial and temporal independence. For the fully parametric model, the pseudo-log-likelihood is

$$p\ell_o(\boldsymbol{\theta}; \xi_o) = \sum_{\mathbf{s} \in \mathcal{S}_o} \left[\sum_{t \in \mathcal{T}_o} \log h(y_{o,t,\mathbf{s}}; \sigma_o, \xi_o) \right], \quad (9)$$

whereas in the GAM setting $p\ell_o$ is adapted by incorporating an additive spline smoothing penalty term (Wood 2006). Given the use of a pseudo-penalised likelihood, we cannot use standard methods for the evaluation of parameter uncertainty and model selection. Instead, the approaches we use are discussed in Section 3.4, with our marginal tail inference for the data being presented in Section 3.5.

3.4. Model uncertainty quantification and selection

In cases where a pseudo-likelihood is used, as in Section 3.3, the most widely adopted method for model selection is to adapt standard information criteria to account for model/likelihood mis-specification to greater penalise complexity relative to a better pseudo-likelihood fit. For spatial extremes, the composite likelihood information criterion (CLIC, Davison et al. 2012) is used, which includes a first-order asymptotically motivated additive adjustment factor. Despite being used in many pseudo-likelihood approaches, we have chosen not to use CLIC for model selection. This is because the likelihoods for extremes are far from the asymptotic elliptical forms around the mode, yet CLIC relies on such asymptotic theory; CLIC measures only goodness of fit in the sample yet we have rich enough data to exploit out-of-sample model assessment; and for determining the parameter uncertainty we are not relying on asymptotic theory. Below we outline the alternative approaches we use.

3.4.1. Bootstrap methods

For both model selection and parameter uncertainty evaluation we generate bootstrapped samples X_o^* of $\{X_o(t, \mathbf{s}) : t \in \mathcal{T}_o, \mathbf{s} \in \mathcal{S}_o\}$ for a given marginal distribution model. These bootstrap samples need to preserve all spatial dependencies, short-range temporal dependence consistent with the passage of weather systems, missing data patterns of the observational data, and to exhibit the temporal non-stationarity of the fitted model.

For a given marginal model, the bootstrap takes the set of transformed observed data $\{x_o^U(t, \mathbf{s}) = F_{o,t,\mathbf{s}}(x_o(t, \mathbf{s})) : t \in \mathcal{T}_o, \mathbf{s} \in \mathcal{S}_o\}$ where $F_{o,t,\mathbf{s}}$ is given by the two model components of Sections 3.2 and 3.3. Assuming $F_{o,t,\mathbf{s}}$ is the correct marginal model, then the $x_o^U(t, \mathbf{s})$ values are realisations of $\text{Uniform}(0, 1)$ random variables that are identically distributed over time for each $\mathbf{s} \in \mathcal{S}_o$, but with the temporal and spatial dependence structure of the $X_o(t, \mathbf{s})$ process retained. To these data, we apply a vector temporal block bootstrap, with details of block structure and adaptations to account for the missing data described in Healy et al. (2023, Section 3.4). For each bootstrapped data set $X^{U*} = \{X_o^{U*}(t, \mathbf{s}) : t \in \mathcal{T}_o, \mathbf{s} \in \mathcal{S}_o\}$ we use the inverse of the distribution function $F_{o,t,\mathbf{s}}$ to create the bootstrapped sample X_o^* with elements

$$X_o^*(t, \mathbf{s}) = F_{o,t,\mathbf{s}}^{-1}\{X_o^{U*}(t, \mathbf{s})\}. \quad (10)$$

Applying this raw bootstrap method induces bias in parameter estimates, and hence in sampling distribution estimates. The bias stems from ties in the extreme bootstrapped data that this method produces. As the very largest observations in a dataset are known to be the most influential on the GPD model fit (Davison & Smith 1990) this is particularly problematic. There is negative bias in the estimate of the shape parameter of the GPD. Since the shape and scale parameters are negatively correlated, there is also a positive bias in the scale parameter estimator. To adjust for these biases we use a bootstrap error correction, which we validate via assessing the coverage properties of the resulting confidence intervals in a simulation study, see Healy et al. (2023, Sections 3.5-3.6). The bias-correction is a two-step procedure with an additive adjustment to the bootstrapped shape parameter estimate, then the scale parameter is re-estimated after fixing the adjusted shape parameter.

3.4.2. Cross validation

For model fit diagnostics, we use two types of cross-validation (CV) to evaluate the performance of our models on out-of-sample data (Hastie et al. 2001, Ch 7.). We use standard 15-fold CV (15-CV) so that the data are divided into 15 groups (folds), where each fold is removed in turn and the model is fitted to the remaining folds. Since standard CV can perform poorly when the data have spatial or temporal correlation (Roberts et al. 2017), we also use a spatio-temporal CV (ST-CV) with 15 folds, corresponding to 5 spatial clusters of station data (i.e., divided spatially into 5 contiguous groups) and 3 temporal folds. Each temporal fold consists of every third week in the summer months, preserving long-term temporal non-stationarity. We choose 5 spatial partitions of our 182 sites as being low enough to help account for, and reduce, bias introduced via spatial auto-correlations as well as being sufficiently large that it reduces variance in our performance metrics across folds (Schratz et al. 2019). We define the 15 ST-CV folds as all combinations of spatial and temporal clusters, taking the intersection as a fold. We also investigated higher number of spatial clusters (up to 30) and higher number of random folds (up to 90) and found equivalent model rankings as those presented here.

For each left out fold, we compute two different goodness-of-fit measures to evaluate out-of-sample performance, the root mean square error (RMSE) and the continuous ranked probability score (CRPS, Gneiting & Katzfuss 2014). The RMSE evaluates the

Table 1. Cross validation (RMSE) on the quantile regression analysis for the body of the distribution.

Model structure for $\hat{q}_o^{(\tau)}(t, \mathbf{s})$	ST-CV	15-CV
$\beta_0^{(\tau)}$	1.442	1.455
$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\mathbf{s})$	1.350	1.347
$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\mathbf{s}) + \beta_2^{(\tau)} M^I(t)$	1.322	1.319

general closeness between the empirically estimated and predicted quantiles, whilst the CRPS aims to match both the calibration and the sharpness of these extremes quantiles (Zamo & Naveau 2018). Here the empirical quantile, $x_o^{(\tau)}(t, \mathbf{s})$ is evaluated using the ordered data at site \mathbf{s} and the year which contains time t , whereas the predicted quantiles are estimated as $\hat{x}_o^{(\tau)}(t, \mathbf{s}) = F_{o,t,\mathbf{s}}^{-1}(\tau)$ for quantile τ from the appropriate model. The comparisons between $x_o^{(\tau)}$ and $\hat{x}_o^{(\tau)}$, for the same t , \mathbf{s} , and τ , are averaged across the folds. Lower values of RMSE and CPRS generally indicate a superior fit.

3.5. Marginal Data Analysis

3.5.1. Body of distribution

Following exploratory analysis, we identified three potential models for the body of the distribution, which we present in Table 1 along with their CV RMSE. The first model serves as a base, in which the location parameter is constant over space and time for each quantile τ . In the second model we allow the quantile regression to vary spatially by using the corresponding climate model data quantiles $q_c^{(\tau)}(\mathbf{s})$ as a covariate. The third model also includes the temporal Irish mean temperature covariate $M^I(t)$. The inclusion of the climate model covariate reduces the RMSE for both types of CV, whereas $M^I(t)$ improves the CV scores further, though not as much. We fitted a number of other covariate combinations for \mathbf{z}_t , as well as using the principal components of \mathbf{z}_t to avoid issues of collinearity. Overall, we found the third model provides the best balance of simplicity and fit, so use this for subsequent analysis.

Healy et al. (2023, Section 3.1) provides estimates of $\beta_2^{(\tau)}$, which show a slight decrease with τ although the confidence intervals widen. For all τ , $\beta_2^{(\tau)} = 1$ appears consistent with the data, indicating that mean summer temperatures in Ireland are a good representation of the temporal change for all the body of the distribution. The estimates of $\beta_1^{(\tau)}$ (not plotted) decrease, approximately linearly, from 0.75 to around 0.65 with $0 < \tau < 1$, showing that the climate model is not giving identical descriptions to the station data, as the estimates differ from 1 significantly and change with τ .

3.5.2. Tails of the distribution

For selecting the threshold $u_o(\mathbf{s})$, we use the second model in Table 1 with $\tau = 0.9$, providing a threshold that varies in space but not time. Figure 2 (left) shows the threshold $u_o(\mathbf{s})$ over Ireland, with cooler temperature values on the west of Ireland and coastal regions on the south and north coasts. For this $u_o(\mathbf{s})$ we estimate the threshold exceedance probability $\lambda_o(t, \mathbf{s})$ and its spatial average $\lambda_o(t) = \int_{\mathbf{s} \in \mathcal{S}_c} \lambda_o(t, \mathbf{s}) d\mathbf{s} / |\mathcal{S}_c|$. Estimates of the

Table 2. Models $M_0 - M_2$ for GPD log-scale parameter, $\log \sigma_o(t, \mathbf{s})$, along with cross-validation results and estimated shape parameter, ξ_o , with bootstrapped 95% confidence intervals. Numbers in bold font show the lowest CV values.

		ST-CV		15-CV		ξ_o
		RMSE	CRPS	RMSE	CRPS	
M_0	$\beta_0 + \beta_1 \log \sigma_c(\mathbf{s})$	0.945	0.895	0.928	0.882	-0.152(-0.237, -0.092)
M_1	$\beta_0 + \beta_1 \log \sigma_c(\mathbf{s}) + \beta_2 M^I(t)$	0.938	0.894	0.918	0.880	-0.156(-0.204, -0.110)
M_2	$\beta_0 + \beta_1 \log \sigma_c(\mathbf{s}) + \beta_2 \log C(\mathbf{s}) + \beta_3 M^I(t) + \beta_4 \log C(\mathbf{s}) M^I(t)$	0.934	0.893	0.908	0.878	-0.158(-0.194, -0.110)

latter are shown in (Healy et al. 2023, Figure 4). The $\lambda_o(t)$ estimates show an increasing exceedance rate, with the average rate over time of 0.1 reflecting the choice of the threshold. However, we find that from 1942–2020, $\lambda_o(t)$ increases by around 35% with associated 95% confidence interval 28-44%. We see the same features at individual sites, but with wider confidence intervals.

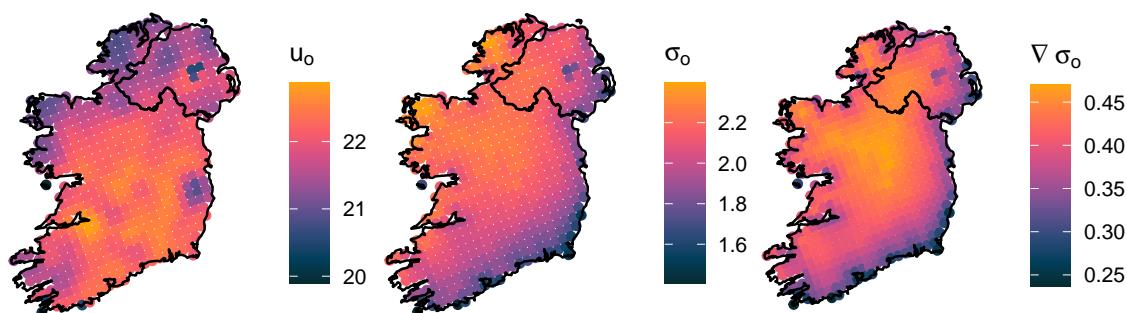


Fig. 2. Estimated values of threshold $u_o(\mathbf{s})$ (left), GPD scale parameter according to M_2 in 2020, (centre), and the estimated change in the scale parameter since 1942, $\nabla \sigma_o(\mathbf{s})$ (right).

Table 2 presents a subset of the models that we explored for the GPD scale parameter that incorporate climate model data via $\sigma_c(\mathbf{s})$, defined by expression (6), and the nearest coastal distance $C(\mathbf{s})$ from the site \mathbf{s} . The models incorporate $\log \sigma_o$ being a constant over time M_0 ; temporal non-stationarity via $M^I(t)$ with a spatially constant temporal trend in M_1 ; and with a spatial varying (with $C(\mathbf{s})$) temporal trend in M_2 . Other models were attempted with differing covariates and spline structures included but these failed to improve over models $M_0 - M_2$. However, over a range of spline models we noticed that they were consistently suggesting evidence for different temporal trends on the coast relative to inland, hence our introduction of the $C(\mathbf{s})$ covariate.

Table 2 presents our model selection diagnostics based on CV metrics (CRPS and RMSE). All four approaches favour model M_2 , with M_0 and M_1 having similar, slightly inferior performance, and we find that M_0 is too simplistic. Models $M_0 - M_2$ estimate the coefficient of $\log \sigma_c$ as close to 1 in all cases, showing that the climate model provides very helpful information as a spatial covariate. Estimates of the shape parameter ξ_o are also given in Table 2. As the GPD scale parameter model is made increasingly flexible (from model M_0 to M_2), the value of ξ_o decreases, lightening the tail decay, indicating

that each model is progressively reducing sources of variation in the tail. Since there is some uncertainty in the marginal model choice, we take M_0 , M_1 , and M_2 through the spatial dependence analysis to assess the sensitivity of the risk measures, with details for model M_2 reported here and for M_0 and M_1 in Healy et al. (2023).

Model M_2 gives that the most variable excess distribution is on the west coast (see centre plot in Figure 2), with a decay in values from west to east, so almost the opposite of the behaviour of $u_o(\mathbf{s})$. We also investigated the estimated change in the scale parameter over the observation period, denoted $\nabla\sigma_o(\mathbf{s}) = \sigma_o(2020, \mathbf{s}) - \sigma_o(1942, \mathbf{s})$, see Figure 2 (right), and found it to be largest in the centre of Ireland, with the change there being close to double that on the coast. The scale parameter is increasing over time everywhere, leading to warmer extreme temperatures.

The model selection diagnostics in Table 2 show the relative quality of the three model fits. To assess the absolute quality of the fitted model M_2 we use pooled QQ plots in Figure 3, pooling over all sites and years. Due to the spatio-temporal non-stationarity of the marginal model, we transform the data through our fitted model into a common uniform scale and to a common exponential scale (for the conditional distribution of threshold excesses). The choice of scales helps identify key departures of fit in the body and tails of the distribution respectively. We see evidence of an exceptionally strong fit in both components of the distribution, with values near the lines of equality, and in the far upper tail, all values falling within the pointwise tolerance bounds which were derived assuming independence of time and space (so are much narrower than necessary). Similar site-specific QQ plots are shown in Healy et al. (2023, Figure 13) for the five stations identified in Figure 1 and for five other randomly selected stations. These show a slightly more varied quality of fit, with the least good fits occurring on the coastlines, e.g., Malin Head, but with very good fits at most stations.

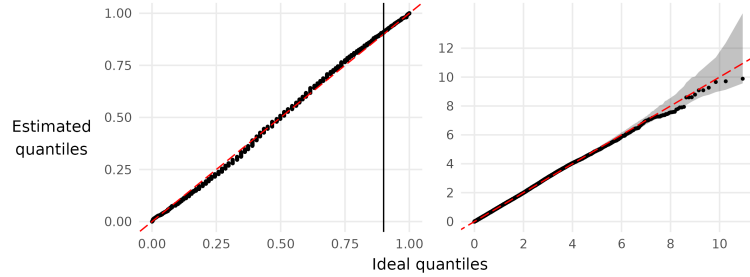


Fig. 3. Spatially and temporally pooled QQ-plots for model M_2 : (left) all data on uniform margins, threshold shown as vertical line; (right) tail model (GPD) on exponential margins. The shaded region shows pointwise 95% tolerance intervals. The lines of equality are in red.

4. Spatial models

4.1. Standardising data

When modelling dependence between variables with differing marginal distributions and covariates, it is common to first standardise the marginal variables so that they have an identical distribution over variables and covariates (Coles 2001, de Haan & Ferreira

2006). Here we transform the data to (unit) Pareto distributions, X^P , X_c^P , and X_o^P , using the same subscript notation as in Section 3. The choice of Pareto marginal scale, or the similar Fréchet scale, is ideal for studying asymptotically dependent variables, a property defined in Section 4.2, such as r -Pareto processes (de Haan & Ferreira 2006), but less ideal for asymptotically independent variables, where shorter-tailed Exponential or Laplace distributions are favoured (Wadsworth & Tawn 2022). We use of the probability integral transform, i.e.,

$$X^P(t, \mathbf{s}) = 1/[1 - F_{t,\mathbf{s}}\{X(t, \mathbf{s})\}], \text{ for all } \mathbf{s} \in \mathcal{S} \text{ and all } t, \quad (11)$$

where the marginal distribution function $F_{t,\mathbf{s}}$ takes a different estimated form below and above $u(\mathbf{s})$, see Sections 3.2 and 3.3. Thus, if the marginal model is perfectly estimated, we have $\Pr(X^P(t, \mathbf{s}) > y) = y^{-1}$ for all $y > 1$, t , and \mathbf{s} . In our uncertainty assessment in the subsequent inference, the marginal model uncertainty is accounted for through our bootstrap procedures. To transform from standard Pareto margins to the original distribution at \mathbf{s} and t , the inverse of the transformation (11) is used.

4.2. Classification of extremal dependence type

We now explore the nature of the extremal spatial dependence structure in the processes X_c^P and X_o^P . To simplify notation we omit the temporal dimension of these spatial processes but always consider the process on the same day at different locations. Following Coles et al. (1999), we estimate the pairwise coefficient of asymptotic dependence, χ . Specifically, for the process X^P at sites \mathbf{s}_i and \mathbf{s}_j , $\chi = \chi(\mathbf{s}_i, \mathbf{s}_j)$ is defined by

$$\chi(\mathbf{s}_i, \mathbf{s}_j) = \lim_{v \rightarrow \infty} \Pr(X^P(\mathbf{s}_j) > v | X^P(\mathbf{s}_i) > v). \quad (12)$$

If $\chi(\mathbf{s}_i, \mathbf{s}_j) > 0$ (or equals 0) then X^P is said to be asymptotically dependent (or asymptotically independent) respectively for these sites. The larger the value of χ ($0 < \chi \leq 1$) the stronger the asymptotic dependence.

The selection of the appropriate extremal dependence model for the data depends on whether or not the process is better approximated as being asymptotically dependent for all $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$ or not. The base quantity that is typically used to identify asymptotic dependence for a pair of sites is $\chi(\cdot, \cdot; p)$, where

$$\chi(\mathbf{s}_i, \mathbf{s}_j; p) = \Pr(X^P(\mathbf{s}_j) > v_p | X^P(\mathbf{s}_i) > v_p), \quad (13)$$

with $v_p = 1/(1 - p)$ being the p th marginal quantile of X^P . An empirical estimate of $\chi(\mathbf{s}_i, \mathbf{s}_j; p)$ exploits the replication over t by assuming spatial dependence does not change with t . We denote this estimator by $\tilde{\chi}(\mathbf{s}_i, \mathbf{s}_j; p)$. We expect approximate spatial stationarity and isotropy of the spatial extreme process. Plotting (not shown) the cloud of $\tilde{\chi}(\mathbf{s}_i, \mathbf{s}_j; p)$ against the Euclidean distance between the sites ($h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$), reveals a decay with distance that is somewhat hidden by the sampling variation of the points, with the variation depending on the overlap in time of samples at the pairs of sites. A better empirical estimate of $\chi(h; p)$, the pairwise extremal dependence at separation distance h , exploits the property that it changes smoothly over h and that we can obtain the sampling distribution of $\tilde{\chi}(\mathbf{s}_i, \mathbf{s}_j; p)$ through the bootstrap. Together these enable us

to construct a weighted estimate $\tilde{\chi}(h; p)$ from the cloud of points (using pairs with h_{ij} close to h) and obtain its sampling distribution. We used 500 bootstraps and 30 binned distances, each with an equal number of pairs of sites.

Figure 4 shows the behaviour of $\tilde{\chi}(h; p)$, for both X_c^P and X_o^P processes. It shows estimates and intervals that account for 95% of the marginal estimation uncertainty, which for X_o we used model M_2 . These estimates are shown for $p \in (0.8, 0.85, 0.9)$, the latter corresponding to only 9 days per summer. Despite the climate model having a much richer set of pairs of sites and longer simultaneous data, both processes provide very similar qualitative findings. Naturally, $\chi(h; p)$ decreases with distance but in all cases, it is far from zero, even between the most distant pairs of sites. For short distances, the estimates of $\chi_c(h; p)$ exhibit less variation than in the estimated values for $\chi_o(h; p)$. We have that $\hat{\chi}_o(h; p) \leq \hat{\chi}_c(h; p)$ for all distances, suggesting that the X_c data are overestimating the extremal dependence in X_o . This difference is important when looking at extreme events spatially, as it suggests using the climate model data alone (or when down-scaled) will lead to an overestimate of the risk of widespread heatwaves in Ireland.

Most critical for our modelling of the observed process is to assess whether, as p increases to 1, the $\hat{\chi}_o(h; p)$ values decay to zero or stabilise at a non-zero limit indicating asymptotic independence and asymptotic dependence respectively. There is a small decline, at all distances, however, even when $p = 0.9$ these estimates are far from zero for both X_c^P and X_o^P . So we conclude that it seems reasonable that maximum daily temperature data are consistent with asymptotic dependence over Ireland.

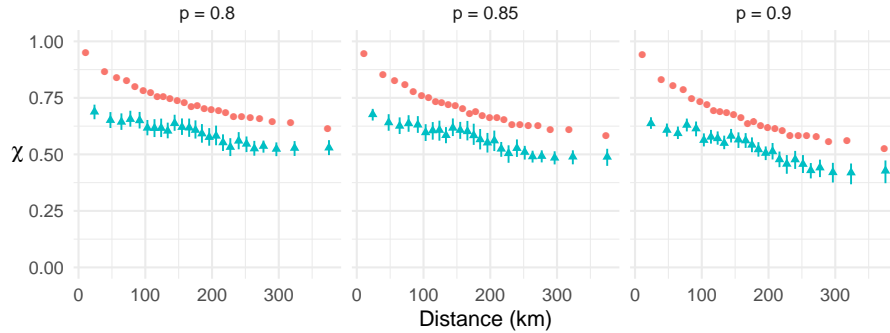


Fig. 4. Empirical estimates $\tilde{\chi}(h; p)$ plotted against inter-site distance h for the climate model (χ_c) (orange) and station data (χ_o) (blue). Plots use the marginal model M_2 for $p = 0.8, 0.85$ and 0.9 (left to right) with 95% confidence intervals shown as vertical lines.

4.3. r -Pareto Processes

We now jointly model the extreme values of the process $X^P(t, \mathbf{s})$ over $\mathbf{s} \in \mathcal{S}$, with unit Pareto distributed marginal variables. We look at spatial fields separately for each t and simplify notation by dropping the argument t . First, we define what we mean by a spatially extreme event as there is no natural ordering of multivariate or spatial processes. Here the level of extremity of the stochastic process $X^P := \{X^P(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ is determined by a risk function $r(X^P) \in [0, \infty)$, where the only constraint on r is

that it is homogeneous of order 1, i.e., $r(c\mathbf{x}) = cr(\mathbf{x})$ for any constant $c > 0$ and with $\min(\mathbf{x}) > 0$. de Fondeville & Davison (2018, 2022) suggest taking r as the magnitude at one particular site, or the spatial mean, median, maximum, or minimum over \mathcal{S} .

Under weak conditions on X^P , de Fondeville & Davison (2018) report that

$$\Pr \{ [v^{-1}X^P(\mathbf{s}) : \mathbf{s} \in \mathcal{S}] \in \cdot \mid r(X^P) > v \} \rightarrow \Pr \{ [Y_r^P(\mathbf{s}) : \mathbf{s} \in \mathcal{S}] \in \cdot \} \quad (14)$$

as $v \rightarrow \infty$, where $\{Y_r^P(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\} := Y_r^P$ is marginally non-degenerate for all $\mathbf{s} \in \mathcal{S}$, with Y_r^P an r -Pareto process. Limit (14) implies that scaled events of the X^P process with risk exceeding a threshold of v are increasingly well-approximated by an r -Pareto process, as the risk threshold increases to infinity. The limit (14) is used for statistical modelling by taking it as an equality for a suitably large value for v , denoted by v_r with $v_r > 1$, then those spatial events with a risk function exceeding v_r are treated as realisations from an r -Pareto process. Specifically, taking a set $B \subseteq \{\mathbf{x} : r(\mathbf{x}) > 1\}$ leads to the modelling assumption that

$$\Pr \{ X^P \in v_r B \mid r(X^P) > v_r \} = \Pr \{ Y_r^P \in B \}. \quad (15)$$

Hence defining the set $A = v_r B$ and un-doing the conditioning on the left-hand side of equality (15), for any $A \subseteq \mathcal{A}_r := \{\mathbf{x} : r(\mathbf{x}) > v_r\}$ we obtain that

$$\Pr \{ X^P \in A \} = \Pr \{ r(X^P) > v_r \} \Pr \{ v_r Y_r^P \in A \}. \quad (16)$$

The r -Pareto process exhibits properties which can be exploited for efficient evaluation of $\Pr \{ X^P \in A \}$. Specifically, Y_r^P decomposes into two independent components:

$$Y_r^P(\mathbf{s}) = RW_r(\mathbf{s}) \text{ for all } \mathbf{s} \in \mathcal{S}, \quad (17)$$

where R is unit Pareto distributed and is interpreted as the risk of the process, and $W_r := \{W_r(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ is a stochastic process which describes the spatial profile of the extreme event, i.e., the proportion of the risk function r contributed by each site. The limiting dependence structure of X^P is entirely determined by the stochastic properties of W_r . By construction $R = r(Y_r^P)$ and $r(W_r) = 1$. A consequence of the limiting approximation (15) holding above v_r and R having a Pareto distribution is that expression (16) simplifies as we have $\Pr \{ r(X^P) > v_r \} = c_r v_r^{-1}$ where $c_r \in (0, 1]$ depends only on the choice of risk functional and the dependence structure of Y_r^P . However, for our choice of r , given by expression (20), $c_r = 1$ always, see Coles & Tawn (1994).

The characterisation (17) is powerful for the extrapolation to larger events than those observed due to R having a known distribution and the independence property ensuring that the spatial profiles of larger events have exactly the same stochastic properties for any event with a risk greater than 1. For any r -Pareto process and a set $A \subseteq \mathcal{A}_r$, there exists a constant $b_A \in [1, \infty)$ such that for any $b \in [1, b_A]$ we have that

$$\Pr \{ v_r Y_r^P \in A \} = b^{-1} \Pr \{ b v_r Y_r^P \in A \}. \quad (18)$$

Although the two sides of this expression are equal, the two probabilities are not, with Opitz et al. (2021) noting that the latter is much more efficient to estimate using Monte Carlo methods. Taking $b > b_A$ will give bias, as some smaller outcomes in A will be

missed by simulations of $bv_r Y_r^P$ while $b < b_A$ leads to unnecessary variability in the empirical estimator. So we look to scale by b_A in Section 4.4, where we discuss how to obtain b_A and illustrate its usage in estimating the right-hand side of expression (16).

The above shows that inference for any extreme events is relatively straightforward once we have a model for the process W_r . We follow Engelke et al. (2015) and de Fondeville & Davison (2018) by modelling W_r as a spatial stationary isotropic log-Gaussian stochastic process which is determined solely by a variogram $\gamma(h)$, for inter-site distance $h \geq 0$. We use the Matérn variogram family

$$\gamma_{\text{mat}}(h; t) = \alpha \left\{ 1 - (2\sqrt{\nu}h/\phi)^\nu 2^{1-\nu} \Gamma(\nu)^{-1} K_\nu(2\sqrt{\nu}h/\phi) \right\}, \quad (19)$$

where K_ν is a modified Bessel function of the second kind and the positive parameters ($\alpha = \alpha_t, \phi = \phi_t, \nu = \nu_t$) determine the variance, range, and smoothness respectively at time t . Our choice of a bounded variogram was based on the evidence from Figure 4 which suggested that the summer temperature process is asymptotically dependent, even at the longest distances in Ireland. See Healy et al. (2023, Section 7) for additional discussion on the choice of variogram function and isotropy.

The issue of missing data in spatial extremes applications seems to be rarely discussed. A possible reason for this is that with composite likelihood fitting methods for max-stable processes (Padoan et al. 2010) the implications are restricted as only pairwise joint likelihood contributions are used so the impact of missing data is limited. This is not the case for r -Pareto processes, which model jointly across all sites, and the issue of missing data in this context does not seem to have been discussed. When encountering missing data it is tempting to remove all observations at that time point from all sites in the network. However, we observed in Section 2 that this tactic would leave us with no data! Fortunately, thanks to the properties of the log-Gaussian process, it is possible to show that the model is closed under marginalisation (Engelke & Hitz 2020).

With missing data, we need to be careful in selecting a suitable risk function r . The choice of risk function needs to be invariant to the changing dimension of partially observed events, whatever their missing patterns. Hence in our statistical inference at time t , for all $t \in \mathcal{T}_o$, we take the risk function r_t to be the average of the standardised variables over the stations which were observed at time t , i.e.,

$$r_t(X_o^P(t, \mathbf{s}) : \mathbf{s} \in \mathcal{S}) = \sum X_o^P(t, \mathbf{s}_i) I_o(t, \mathbf{s}_i) / \sum I_o(t, \mathbf{s}_i), \quad (20)$$

where $I_o(t, \mathbf{s}_i)$ is the indicator variable for whether $X_o(t, \mathbf{s}_i)$ is observed or not, and the sums are from $i = 1, \dots, |\mathcal{S}_o|$. For evaluating r_t we would have liked to use a subset of stations that are observed for all t and reasonably evenly spread across Ireland. Unfortunately, this was not possible due to sampling bias and missing data. To simplify the fitting of the r -Pareto process we ensured that one site was always present. Specifically, all spatial observations we analysed contained Aldergrove (north) which has very little missing data over the period 1942–2020.

We set the risk threshold v_r , used inequality (15) to define extreme spatial events, at the 80% sample quantile of the risk values calculated from all observed events. We explored different threshold choices and selected the lowest level we could whilst making the usual bias/variance trade-off for tail selection. Healy et al. (2023, Figures 16 and

17) show that the parametric estimate of χ_o^P , derived from the variogram, agrees well with empirical estimates for each marginal model M_0 , M_1 , and M_2 .

We explored the effect of a time-changing dependence model. We allowed for the variance and range parameters of the Matérn variogram (19) to vary over time t , while keeping ν constant. We considered a range of constant and log-linear models using each of the marginal models M_0 to M_2 . For M_0 we find some evidence, at the 5% level, for α_t increasing with $M^I(t)$ (weakening dependence over time), but not for the improved marginal models M_1 or M_2 . Evidence for a change in extremal dependence was not statistically significant and so we keep a temporally stationary r -Pareto process.

4.4. Simulation and efficient inference for spatial extreme events

We simulate spatial extreme events on the observational scale in year t by first simulating an event from the r -Pareto process and then map this pointwise to the data scale using the inverse of transform (11) for the required t . For each step of this process we use the selected statistical model and the simulated values are generated using the fitted parameters of that model, or for assessing uncertainty in the point estimates, using the bootstrapped realisations of these parameters. As the estimated r -Pareto process in our application is well-approximated by a stationary process over time, we can generate identically distributed events of the r -Pareto process to transform for each location and time t using the time-varying marginal model. The r -Pareto process simulations are generated using the R package `mvPot` (de Fondeville et al. 2021). We denote these simulations by $\mathbf{y}_1^P, \mathbf{y}_2^P, \dots, \mathbf{y}_m^P$, for m simulations, with the i th simulation consisting of the spatial realisation $\mathbf{y}_i^P = \{y_i^P(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$. For the i th realisation of the r -Pareto process, \mathbf{y}_i^P , we define, $r_i = r(\mathbf{y}_i^P) > 1$ as the risk, $\mathbf{w}_i = \mathbf{y}_i^P / r_i$ as the spatial profile.

Healy et al. (2023, Figure 21) shows five simulated extreme events, transformed to observational scale under 2020 conditions and the exact same events in 1942 conditions (presented as a difference in temperatures at each site, for each event). A positive difference shows the equivalent event in the two years to be hotter in 2020 than 1942, with that difference found to be largest for the hotter events. As the r -Pareto process realisations can have marginal values in the range $(0, 1)$ at some sites, i.e., outside the domain of a marginal Pareto variable, so we follow de Fondeville & Davison (2018) and for transformation to the observational space we use Fréchet marginals, not Pareto.

Although expression (16) provides a basis for inference for the probability of occurrence in any extreme event $A \subseteq \mathcal{A}_r$ by the process X^P , we are most interested in making inference about spatial events of the observational process which exceed a critical temperature of $T^\circ\text{C}$ somewhere over Ireland at time t . We denote these events by

$$A_{t,\mathcal{S}}(T) = \{X_o(t, \mathbf{s}), \mathbf{s} \in \mathcal{S} : \exists \mathbf{s}_0 \in \mathcal{S} \text{ with } X_o(t, \mathbf{s}_0) > T\}. \quad (21)$$

After the marginal transformation to Pareto margins, this event is equal to

$$A_{t,\mathcal{S}}(T) = \{X_o^P(t, \mathbf{s}), \mathbf{s} \in \mathcal{S} : \exists \mathbf{s}_0 \in \mathcal{S} \text{ with } X_o^P(t, \mathbf{s}_0) > T^P(t, \mathbf{s}_0)\}, \quad (22)$$

where $T^P(t, \mathbf{s})$ is the mapping of T through the transformation (11) at time t and for site \mathbf{s} . To use the r -Pareto approximation, all elements of $A_{t,\mathcal{S}}(T)$ must have a risk exceeding v_r ; this imposes a lower bound $T > 20.6^\circ\text{C}$ across all t . We also focus on

marginal extreme events, which restricts $T \geq \max_{i \in \mathcal{S}} u_i = 22.9^\circ\text{C}$. As all the results we present in Section 5 are for $T \geq 26^\circ\text{C}$ this lower bound is not restrictive for our purposes.

To estimate $\Pr \{A_{t,\mathcal{S}}(T)\}$, there have been a set of possibilities proposed, see Healy et al. (2023, Section 6.1). We focus on the most efficient of these estimators, which exploits the independence property (17), and the scaling property (18). Specifically, \mathbf{w}_i and r_j are independent realisations for all i, j , and there is no reason to restrict ourselves to the observed r_j as we know they are unit Pareto realisations. So we supplement the information to have $\{r_j^P; j = 1, \dots, L\}$, which are i.i.d. realisations of a unit Pareto variable, where L is taken as large as possible to improve computational efficiency. To find the optimal scaling factor $b_{T(t)}$ we first define component-wise maxima of the simulated Pareto processes scaled to have unit cost, i.e., $\omega_{(m)}(\mathbf{s}) = \max_{i=1, \dots, m} \{y_i^P(\mathbf{s})/r_i\}$, for each $\mathbf{s} \in \mathcal{S}$. At time t , we want to scale these component-wise maxima by as much as possible without producing a scaled event with an exceedance of $T^P(t, \mathbf{s})$ for some $\mathbf{s} \in \mathcal{S}$. The appropriate scaling is then $b_{T(t)} = \min_{\mathbf{s} \in \mathcal{S}} \{T^P(t, \mathbf{s})/\omega_{(m)}(\mathbf{s})\}$. Here $b_{T(t)} = v_r b_A$ in expression (18). Combined together, these give a form of importance sampling estimator

$$\hat{\Pr}_{imp} \{A_{t,\mathcal{S}}(T)\} = \frac{1}{b_{T(t)} m L} \sum_{i=1}^m \sum_{j=1}^L I \left\{ \exists \mathbf{s}_0 \in \mathcal{S} : r_j^P b_{T(t)} \frac{y_i^P(\mathbf{s}_0)}{r_i} > T^P(t, \mathbf{s}_0) \right\}, \quad (23)$$

see Healy et al. (2023, Algorithm 2). With this scaling choice, and the extrapolation from the $r_j^P > \max(r_1, \dots, r_m)$, we are guaranteed to have at least L out of the mL simulated fields which achieve at least temperature $T^\circ\text{C}$ somewhere in \mathcal{S} in year t .

5. Temporal changes in spatial extreme events

We present a range of summaries detailing how spatial daily maximum temperature extreme events in Ireland are changing over the period 1942–2020. First, we look at the changes in the marginal quantiles. Figure 5 shows estimates of the level exceeded by daily values with probability $1/9200$, i.e., that of a 100-year return level if the process was stationary in time. For simplicity, we refer to these as the 100-year levels changing over time. For model M_2 , we show these estimates for 2020 and also present the estimated difference between them for 2020 and 1942. In the latter, a positive value represents a warming of temperatures. The 100-year return level has increased between $1.2 - 2.2^\circ\text{C}$ across Ireland, with the larger increases away from the south and east coasts. These changes in extreme temperatures over this time period are substantially larger than the 1°C change of M^I and M^G , illustrating that climate change is more radically affecting summer temperature extreme events than mean levels.

In Figure 5 we also report the change in the 0.025 and 0.975 quantiles of change in return level from 1942 - 2020 across all bootstrap samples. In both cases, and at all sites, these changes are positive, with the rate of change in these features being greater than that of the point estimates. Although we present the results for the 100-year return levels, similar results hold for all high quantiles and for the finite upper endpoints of the marginal distribution; the latter as the estimated GPD shape parameter is negative (see Table 2). Healy et al. (2023, Figures 14 and 15) gives equivalent figures for M_0 and M_1 .

Next, we consider summaries that also reflect the dependence structure of extreme events. There are no established analytical closed-form expressions of such changes.

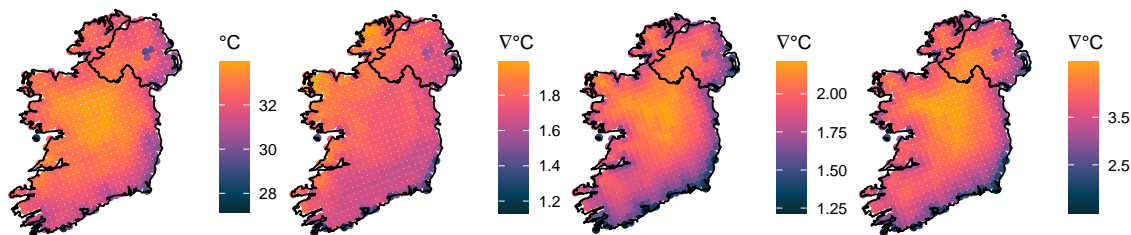


Fig. 5. (1st plot) Estimated 100-year marginal return level for the year 2020; (3rd plot) estimated change in 100-year marginal return level from 1942 to 2020; (2nd and 4th plots) lower and upper 95% CI limits respectively for the change in 100-year return level from 1942 to 2020.

Instead, we revert to using simulated fields of extreme events and presenting risk measures based on empirical summaries using large samples of these fields. The simulation strategy set out in Section 4.4 gives replicated independent spatial fields. In particular, we focus on the occurrence of events $A_{t,\mathcal{S}}(T)$, i.e., an extreme temperature of at least $T^\circ\text{C}$ somewhere in Ireland (determined by the set of locations as required), and then summaries of the properties of such events. We estimate $\Pr\{A_{t,\mathcal{S}}(T)\}$ using the estimator $\hat{\Pr}_{imp}$, with $m = 25,000$ and $L = 300$. Figure 6 shows this estimated probability (expressed as a return period) for a range of temperatures $T \in [26, 34]$ for years 1942 and 2020 separately for \mathcal{S}_o and \mathcal{S}_c . For \mathcal{S}_o , the plot reveals a marked change with estimated return periods being shorter in 2020 compared to 1942 for the same T . To illustrate this, consider the event with the hottest temperature observed anywhere at the station network, a temperature of 33°C at Phoenix Park, Dublin, July 2022. The spatial event $A_{t,\mathcal{S}}(33)$ changes from being a 1 in 182-year event in 1942 to a 1 in 8.7-year event in 2020. Furthermore, the model estimates that a temperature in excess of 34°C , i.e., a value not yet recorded in Ireland, changes from a 1 in 1,588-year event to a 1 in 27.5 year event over this time window.

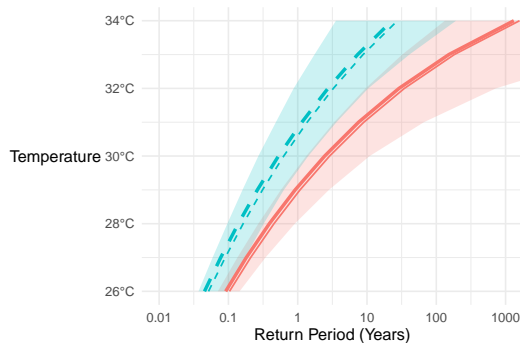


Fig. 6. Return period of the event $A_{t,\mathcal{S}}(T)$ where an extreme temperature exceeding $T^\circ\text{C}$ occurs somewhere on the Irish station network, \mathcal{S}_o . Blue dashed (solid orange) lines correspond to $t = 2020$ (1942). Shaded regions show pointwise 95% confidence intervals for the return periods. The higher bold curves show the corresponding point estimates for the climate model grid \mathcal{S}_c .

Figure 6 shows that, for a given return period, hotter temperatures are expected somewhere in \mathcal{S}_c than on \mathcal{S}_o , as the former is denser and with better coverage than the station network. The difference between the results for the two collections of sites is very small. This slight change shows that the station network, when all gauges are working, has the ability to fully capture all extreme temperature events over Ireland. Such information has not been available previously given the complexity of addressing spatial dependence, marginal non-stationary, and missing data in the station network.

We now propose risk metrics to summarise the features of events satisfying $A_{t,\mathcal{S}}(T)$. First consider a measure of pairwise dependence, which extends the idea behind χ but applied to the data scale, so it combines the effects of changes over time in the marginal distributions and the estimated extremal dependence structure. Specifically, we define

$$\chi_o^{obs}(h | A_{t,\mathcal{S}}(T)) = \Pr\{X_o(t, \mathbf{s}^h) > T \mid A_{t,\mathcal{S}}(T)\},$$

where \mathbf{s}^h is a randomly selected site in \mathcal{S} with $\|\mathbf{s}^h - \mathbf{s}_0\| = h$, i.e., the conditional probability of the observational process exceeding temperature T on day t at a site which is a distance h away from a site \mathbf{s}_0 that has a temperature exceeding T on that day. We also investigate the associated unconditional risk measure

$$\chi_o^{obs}(h; T, t) = \Pr(\exists \mathbf{s}_0 \in \mathcal{S} : \min[X_o(t, \mathbf{s}_0), X_o(t, \mathbf{s}^h)] > T).$$

Figure 7 presents estimates of each of these two risk measures for a range of h and T , between 1942 and 2020. On the observed data scale, we find that extremal spatial dependence decreases with distance as would be expected, but beyond this, there are quite different findings from the two measures. Risk measure $\chi_o^{obs}(h | A_{t,\mathcal{S}}(T))$ is broadly stable over the presented range of T and t whilst the unconditional $\chi_o^{obs}(h; T, t)$ is substantially different. The former is perhaps not too surprising given that the model is asymptotically dependent (the dependence structure is invariant to any change in the extremity of an event) and the extremal dependence structure is estimated to be stationary over time. However, as the event is on the marginal scale and the marginal distributions are changing with time, this finding was not anticipated. For $\chi_o^{obs}(h; T, t)$ we do see that the joint probability of temperature being above T at sites h apart changes notably with time, e.g., taking $h = 100$ km, we find that $\chi_o^{obs}(h; T, t)$ has increased by a factor of 2.8, 3.5, and 4.7 for $T = 28, 29$, and 30°C respectively.

Finally, we look at a spatial risk measure based on the proportion, C , of a spatial field over the network that exceeds $T^\circ\text{C}$ at time t . Specifically, we consider the expected value of C , denoted $E_o(C; t, T)$. We also consider the conditional expected value of C , given by $E_o\{C \mid A_{t,\mathcal{S}}(T)\}$, i.e., given that we have observed a temperature somewhere on the station network. This conditional expectation is closely related to a functional used in characterising heatwave events (Cebrián et al. 2022). These functionals and their relationships are given as follows:

$$\begin{aligned} E_o(C; t, T) &= \mathbb{E} \left(\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} I\{X_o(t, \mathbf{s}) > T\} d\mathbf{s} \right) \\ &= \mathbb{E} \left(\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} I\{X_o(t, \mathbf{s}) > T\} d\mathbf{s} \mid A_{t,\mathcal{S}}(T) \right) \times \Pr\{A_{t,\mathcal{S}}(T)\} \\ &= E_o\{C \mid A_{t,\mathcal{S}}(T)\} \times \Pr\{A_{t,\mathcal{S}}(T)\}, \end{aligned}$$

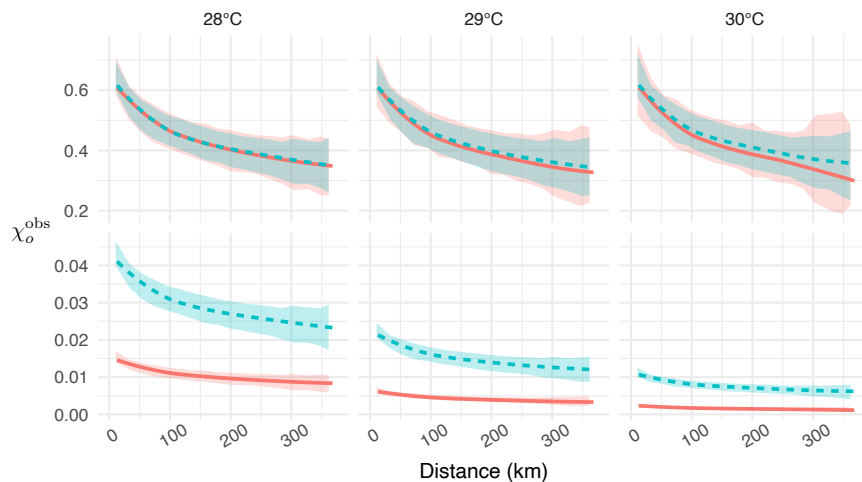


Fig. 7. Estimates of $\chi_o^{\text{obs}}(h | A_{t,S}(T))$ (top row) and $\chi_o^{\text{obs}}(h; T, t)$ (bottom row) against h (in km) for $T = 28, 29,$ and 30°C for 1942 (solid, orange line) and for 2020 (dashed, blue line) for model M_2 . Confidence intervals are based on 10,000 simulations for each 500 bootstrap sample datasets.

where $I(B)$ is the indicator function of event B .

Figure 8 shows that estimates of both of these measures for the station network over Ireland have increased from 1942 to 2020. The changes are highly significant, a factor of 90 larger for $T = 34^\circ\text{C}$ when considering the unconditional expectation $E_o(C; t, T)$. However, when conditioning on the event $A_{t,S}(T)$ this expected coverage proportion exhibits more limited changes with the largest difference being a doubling of the expected area affected when $T = 34^\circ\text{C}$. In this latter case, the estimated change is small by comparison with its associated uncertainties.

We finish by reporting on our investigation of the sensitivity of our risk measure analyses to our marginal modelling choices for temporal non-stationarity. The results for the models M_0 and M_1 , i.e., the less well-fitting models, are given in Healy et al. (2023, Section 6.2), with these being given for the same features shown for model M_2 in Section 5. Unsurprisingly, the inclusion of temporal non-stationarity in the tail model gives markedly different conclusions for all risk measures compared to those derived from the stationary model M_0 . The inclusion of a coastal proximity covariate in model M_2 leads to larger GPD scale parameter estimates inland and lower estimates in coastal regions than M_1 , see Healy et al. (2023, Figure 7). This is reflected in the estimates of $\Pr\{A_{t,S}(T)\}$ and $E_o(C; t, T)$, with estimates from M_0 being lower than both M_1 and M_2 , and M_1 giving slightly higher estimates than M_2 on the station network. This change results in the probability of observing 33°C somewhere in \mathcal{S}_o increasing by a factor of 1.5, 25, and 21 between 1942 and 2020 for models $M_0, M_1,$ and M_2 respectively, showing that the difference in key conclusions is not too large between models M_1 and M_2 .

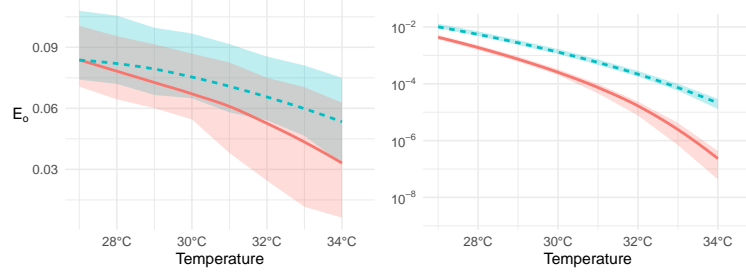


Fig. 8. Left: expected proportion, $E_o\{C \mid A_{t,S}(T)\}$, of Ireland that exceeds a temperature of $T^\circ\text{C}$ in an extreme event given that at least one site in Ireland (at the station network) exceeds $T^\circ\text{C}$ according to M_2 . Right: the equivalent unconditioned estimates, i.e., estimates of $E_o(C; t, T)$. Estimates are plotted against T for 1942 (solid, orange line) and for 2020 (dashed, blue line). The shaded regions give associated pointwise 95% confidence intervals, based on 10,000 simulated fields for each 500 bootstrap sample datasets.

6. Conclusions

We have presented some novel candidate approaches to merge information from spatially and temporally complete climate models into the spatial extreme value analysis of sparse and temporally incomplete observed temperatures from available meteorological stations. New methodological features include using outputs from an extreme value analysis of the climate model data to provide a covariate for the equivalent analysis of observational data, and dealing with r -Pareto processes in a missing data framework. We also presented novel metrics, combining both marginal and dependence features, to describe changes in spatial risk over time.

Our analysis was for daily maximum summer temperatures over Ireland. We found that the climate model data were more helpful for marginal modelling of the observational data than for dependence modelling, as they overestimate extremal spatial dependence relative to the observational data. We pooled data from across stations to fit our model and found evidence that the Irish summer temperature anomalies were the best-fitting covariate, appearing to mostly affect marginal behaviour with minimal effect on spatial extremal dependence, see Healy et al. (2023, Section 2 and 7.2). We found that from 1942 to 2020 the occurrence rates of high threshold exceedances have increased by 35%, with 95% confidence interval 28 – 44%, and extreme quantiles have increased by 1.2 – 2.2°C, the latter $\approx 1^\circ\text{C}$ greater than the change of mean summer temperature anomalies for Ireland and globally. Finally, we found that spatial heatwave events over thresholds that are critical for society have become much larger, having at least doubled in spatial extent for 28°C, with this change increasing at more extreme temperatures.

7. Discussion

7.1. Use of climate models

In the analysis of climate extremes, practitioners tend to make use of some combination of observations and climate models, but without necessarily recognising the inherent

trade-offs between them and the respective limitations in the data sources. Perhaps more critically we believe further thought could be given to the synergies which might enable a better set of tools for practitioners. Downscaling from climate models to match observational properties is a major area in statistical climate science, but it risks introducing biases in spatial dependence from climate models given its focus on the marginal agreement. Here we have illustrated what appears to be an effective new approach aiming instead to enhance the observational analysis by exploiting the strong information about the physical properties of the climate system and the greater spatial coverage of information embedded in the climate model data.

As a proof of concept, we have restricted the information we extract from the climate models to that arising from one climate model output. There exists a broad ensemble of global and regional models that could be used. Each combination of such models encapsulates different modelling assumptions and therefore would provide a distinct estimate of the behaviour of maximum temperatures over Ireland. To fully quantify the uncertainty in our estimates would require an adequate sampling strategy to select global-regional model combinations from the available ensemble.

Our analysis looks only at the change in extreme temperature events over the historical record. We do not have observations of the future. Climate models are our only reliable tool for predicting future temporally non-stationary extremal behaviour under different scenarios, so being able to link the temporal non-stationarity of observational and climate data is essential to understanding model strengths and weaknesses, one such recent example being Zhong, Brunner, Opitz & Huser (2022) for precipitation extremes. Our approach has been to only extract marginal spatial information from the climate model data, via $\sigma_c(\mathbf{s})$, and instead rely on additional measures of global and local climate change to capture temporal non-stationarity. However, for inference on a future time period, it may be beneficial to model the change in this parameter through time, i.e., $\sigma_c(t, \mathbf{s})$, but this would require strong evidence that the climate model was really capturing identifiable trends in the observed data. Incorporating future climate modelling information is particularly important for the consideration of highly nonlinear change linked to instabilities such as the possible effects of any change in the strength of the Atlantic meridional overturning circulation (AMOC) which has a profound modulating effect on Irish climate. Most global models suggest some weakening of the AMOC through to 2100, and a complete shutdown cannot be ruled out.

Finally, although we focus on temperature, many other variables (e.g., wind, rainfall) are important for assessing climate change for extreme meteorological events marginally, jointly, and integrated over different time windows. Climate model data are likely to provide improvements in extremal inference of such joint distributions under the assumption that they better capture the physical interactions between processes. This can be used to enhance the equivalent empirical information from the observational data.

7.2. *Choice of threshold in non-stationary analysis*

Threshold choice for the GPD and other tail models for identically distributed univariate extremes has been a major area of research for much of the last 40 years. Therefore, it is not surprising that there are a number of different perspectives for picking a systematic threshold selection criteria in our temporally non-stationary spatial context.

For univariate temporally non-stationary problems Eastoe & Tawn (2009) propose pre-processing the data using models fitted to the body of the distribution before modelling the extremes of the residuals with a constant threshold. Another approach is to use a conditional quantile (Northrop & Jonathan 2011). As noted in Section 3.3 it is difficult to account for the uncertainty in the threshold selection, so incorporating a temporal trend into the threshold undermines our ability to account for the uncertainty in estimating the temporal change in extreme data, which is our primary focus. We are pleased to see that even with our constant threshold at each station we found a simple model for how the GPD scale parameter changes over time and that the GPD is a good fit globally. More generally, the signal-to-noise ratio is critical in determining whether non-stationarity is accounted for in selecting the set of “extreme data” to analyse.

We had an additional threshold to select for the extreme spatial dependence modelling via the risk function r . We had to face issues of missing data, with our approach presenting the first methods, we are aware of, for this. Climate models may help here either through exploring the sensitivity of different missing data patterns or through the use of reanalyses (weather forecast models conditioned on the observed values) to replace missing data, as these will help identify the largest events over space correctly.

7.3. Choice of spatial extremes dependence model

The scale of Ireland relative to the physical systems that drive temperature extremes has also played a key role in our choice of extreme value approach for modelling spatial dependence. This enabled us to take a simple model which is asymptotically dependent at the largest required spatial separation, which we achieved via an r -Pareto process coupled to a log-Gaussian latent process with a bounded variogram. We do not believe our approach would be applicable at much broader scales.

Even over the scale of Ireland, the asymptotic dependence property will not necessarily hold for other climatic variables, e.g., precipitation, which are manifest on smaller scales and with higher variability. In such cases, the modelling approaches need to incorporate asymptotic independence and to address issues about the scale over which asymptotic dependence holds (Wadsworth & Tawn 2022, Zhong, Huser & Opitz 2022) or even whether the spatial process is stationary over different mixture type events, e.g., convective or frontal precipitation (Richards et al. 2023). Any application of this method to different regions or processes should certainly involve an assessment of the evidence for asymptotic dependence, as our impression is that this assumption is made too readily. We find that, as a good approximation, we can assume spatial stationarity and isotropy. Ireland has, relatively speaking, simple topography with few ranges of hills of significant altitude. It does not follow that the method would be readily applicable to more complex alpine regions without, at a minimum, considerable additional validation.

7.4. Choice of metrics for assessing change

Section 5 illustrates the challenge of finding effective metrics to illustrate temporal change when considering extremes of spatial fields. Metrics for marginal variables, e.g., high quantiles, are well-established and parsimonious. We see the development of spatial risk measures which enable the simple assessment of changing risk over time as an

important avenue for further research. Spatial extreme value model inference also lacks well-established diagnostic methods for assessing the fit with observed events. Pairwise, and potentially higher order versions of the measure χ , used in Sections 4.2 and 5, can be helpful but these may not be sufficient in practice. Winter et al. (2016) used severity-area frequency curves as a basis of comparison, but these focused only on assessing the performance of the model for the dependence structure. Picking metrics that directly link to risk assessment from heatwaves, such as health factors (Winter & Tawn 2016) or crop failures or forest fires (Zhang et al. 2022), is likely to be valuable for planners.

We focused on the spatial properties of the extreme events. Models for spatio-temporal extremal dependence of the process are needed to capture the evolution over time of spatial extreme events. This is an area where greater focus is required. For processes that are asymptotically dependent in space and time, some methods have been developed (Davis et al. 2013, Huser & Davison 2014), and it is pleasing to see recent extensions to incorporate asymptotic independence (Simpson & Wadsworth 2021).

Acknowledgements

Healy's work was supported by SFI grant 18/CRT/6049. Parnell's work was supported by the SFI awards 17/CDA/4695; 16/IA/4520; 12/RC/2289_P2 and we thank Simon Noone (Maynooth University) for help with data. For access to climate data, we acknowledge the World Climate Research Programme's Working Groups on Regional Climate and on Coupled Modelling, and the European Network for Earth System Modelling. We would like to thank all three referees who provided helpful and extensive constructive criticisms of the work and which has enormously improved its quality and presentation.

References

- Brown, B. M. & Resnick, S. I. (1977), 'Extreme values of independent stochastic processes', *Journal of Applied Probability* **14**(4), 732–739.
- Brown, S. J., Caesar, J. & Ferro, C. A. (2008), 'Global changes in extreme daily temperature since 1950', *Journal of Geophysical Research Atmospheres* **113**(5), D05115.
- Cebrián, A. C., Asín, J., Gelfand, A. E., Schliep, E. M., Castillo-Mateo, J., Beamonte, M. A. & Abaurrea, J. (2022), 'Spatio-temporal analysis of the extent of an extreme heat event', *Stochastic Environmental Research and Risk Assessment* **36**(9), 2737–2751.
- Chavez-Demoulin, V. & Davison, A. C. (2005), 'Generalized additive modelling of sample extremes', *Journal of the Royal Statistical Society. Series C* **54**(1), 207–222.
- Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.
- Coles, S. G., Heffernan, J. E. & Tawn, J. A. (1999), 'Dependence measures for extreme value analyses', *Extremes* **2**(4), 339–365.
- Coles, S. G. & Tawn, J. A. (1994), 'Statistical methods for multivariate extremes: an application to structural design (with discussion)', *Journal of the Royal Statistical Society, Series C* **43**(1), 1–48.

- Davis, R. A., Kluppelberg, C. & Steinkohl, C. (2013), ‘Statistical inference for max-stable processes in space and time’, *Journal of the Royal Statistical Society. Series B* **75**(5), 791–819.
- Davison, A. C., Padoan, S. A. & Ribatet, M. (2012), ‘Statistical modeling of spatial extremes’, *Statistical Science* **27**(2), 161–186.
- Davison, A. C. & Smith, R. L. (1990), ‘Models for exceedances over high thresholds (with discussion)’, *Journal of the Royal Statistical Society: Series B* **52**(3), 393–442.
- de Fondeville, R., Belzile, L. & Thibaud, E. (2021), ‘mvPot: Multivariate peaks-over-threshold modelling for spatial extreme events’.
- de Fondeville, R. & Davison, A. C. (2018), ‘High-dimensional peaks-over-threshold inference’, *Biometrika* **105**(3), 575–592.
- de Fondeville, R. & Davison, A. C. (2022), ‘Functional peaks-over-threshold analysis’, *Journal of the Royal Statistical Society: Series B* **84**(4), 1392–1422.
- de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer, New York.
- Eastoe, E. F. & Tawn, J. A. (2009), ‘Modelling non-stationary extremes with application to surface level ozone’, *Journal of the Royal Statistical Society. Series C* **58**(1), 25–45.
- Engelke, S. & Hitz, A. S. (2020), ‘Graphical models for extremes (with discussion)’, *Journal of the Royal Statistical Society, Series B* **82**(4), 871–932.
- Engelke, S., Malinowski, A., Kabluchko, Z. & Schlather, M. (2015), ‘Estimation of Hüsler–Reiss distributions and Brown-Resnick processes’, *Journal of the Royal Statistical Society. Series B* **77**(1), 239–265.
- Fuentes, M., Henry, J. & Reich, B. (2013), ‘Nonparametric spatial models for extremes: Application to extreme temperature data’, *Extremes* **16**(1), 75–101.
- Giorgi, F. (2019), ‘Thirty years of regional climate modeling: where are we and where are we going next?’, *Journal of Geophysical Research: Atmospheres* **124**(11), 5696–5723.
- Gneiting, T. & Katzfuss, M. (2014), ‘Probabilistic forecasting’, *Annual Review of Statistics and Its Application* **1**, 125–151.
- Hastie, T., Tibshirani, R. & Friedman, J. (2001), *The Elements of Statistical Learning*, Springer, New York.
- Healy, D., Tawn, J. A., Thorne, P. & Parnell, A. (2023), Supplementary material for extreme spatial temperature events in a changing climate with application to Ireland.
- Huser, R. & Davison, A. C. (2014), ‘Space-time modelling of extreme events’, *Journal of the Royal Statistical Society. Series B* **76**(2), 439–461.
- Huser, R. & Wadsworth, J. L. (2022), ‘Advances in statistical modeling of spatial extremes’, *WIREs Computational Statistics* **14**(1), e1537.

- IPCC (2021), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, in V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu & B. Zhou, eds, ‘Cambridge University Press’, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 2061–2086.
- Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes.*, Springer, New York.
- Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A. & Mearns, L. O. (2017), ‘Towards process-informed bias correction of climate change simulations’, *Nature Climate Change* **7**(11), 764–773.
- McElwain, L. & Sweeney, J. (2007), *Key Meteorological Indicators of Climate Change in Ireland*, Technical report, Environmental Protection Agency, Wexford, Ireland.
- Met Office (2012), ‘Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current)’, *NCAS British Atmospheric Data Centre* **2**, 1–30.
- Northrop, P. J., Attalides, N. & Jonathan, P. (2017), ‘Cross-validators extreme value threshold selection and uncertainty with application to ocean storm severity’, *Journal of the Royal Statistical Society. Series C: Applied Statistics* **66**(1), 93–120.
- Northrop, P. J. & Jonathan, P. (2011), ‘Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights’, *Environmetrics* **22**(7), 799–809.
- Opitz, T., Allard, D. & Mariethoz, G. (2021), ‘Semi-parametric resampling with extremes’, *Spatial Statistics* **42**, 100445.
- O’Sullivan, J., Sweeney, C. & Parnell, A. C. (2020), ‘Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland’, *Environmetrics* **31**(5), e2621.
- Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010), ‘Likelihood-based inference for max-stable processes’, *Journal of the American Statistical Association* **105**(489), 263–277.
- Pickands, J. (1975), ‘Statistical inference using extreme order statistics’, *The Annals of Statistics* **10**(3), 119–131.
- Richards, J., Tawn, J. A. & Brown, S. J. (2023), ‘Joint estimation of extreme precipitation aggregates at different spatial scales through mixture modelling’, *Spatial Statistics* **53**, 100725.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle,

- B. A., Hartig, F. & Dormann, C. F. (2017), ‘Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure’, *Ecography* **40**(8), 913–929.
- Schratz, P., Muenchow, J., Iturrutxa, E., Richter, J. & Brenning, A. (2019), ‘Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data’, *Ecological Modelling* **406**, 109–120.
- Simpson, E. S. & Wadsworth, J. L. (2021), ‘Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures’, *Spatial Statistics* **41**, 100482.
- Thibaud, E. & Opitz, T. (2015), ‘Efficient inference and simulation for elliptical Pareto processes’, *Biometrika* **102**(4), 855–870.
- Wadsworth, J. L. & Tawn, J. A. (2022), ‘Higher-dimensional spatial extremes via single site conditioning’, *Spatial Statistics* **51**, 100677.
- Winter, H. C. & Tawn, J. A. (2016), ‘Modelling heatwaves in central France: a case-study in extremal dependence’, *Source: Journal of the Royal Statistical Society. Series C* **65**(3), 345–365.
- Winter, H. C., Tawn, J. A. & Brown, S. J. (2016), ‘Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia’, *Annals of Applied Statistics* **10**(4), 2075–2101.
- Wood, S. N. (2006), *Generalized Additive Models: an Introduction with R*, Chapman and Hall/CRC, New York.
- Youngman, B. D. (2019), ‘Generalized additive models for exceedances of high thresholds with an application to return level estimation for U.S. wind gusts’, *Journal of the American Statistical Association* **114**(528), 1865–1879.
- Youngman, B. D. (2022), ‘evgam: An R package for generalized additive extreme value models’, *Journal of Statistical Software* **103**(3), 1–26.
- Yu, K. & Moyeed, R. A. (2001), ‘Bayesian quantile regression’, *Statistics and Probability Letters* **54**(4), 437–447.
- Zamo, M. & Naveau, P. (2018), ‘Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts’, *Mathematical Geosciences* **50**(2), 209–234.
- Zhang, L., Shaby, B. A. & Wadsworth, J. L. (2022), ‘Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations’, *Journal of the American Statistical Association* **117**(539), 1357–1369.
- Zhong, P., Brunner, M., Opitz, T. & Huser, R. (2022), ‘Spatial modeling and future projection of extreme precipitation extents’, *arXiv preprint arXiv:2212.03028*.
- Zhong, P., Huser, R. & Opitz, T. (2022), ‘Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude’, *The Annals of Applied Statistics* **16**(1), 272–299.