Bayesian Approach to Extreme Value Theory:
How to study unusual weather events?

Davide Fabbrucci, Matteo Pierdomenico, Giacomo Randazzo

Bayesian statistics project
Politecnico di Milano, Italy

Abstract

In this work is developed an extreme value analysis for maximum daily wind gusts in a Bayesian framework. The first part of the article focuses on simplistic inferential procedures, investigating both the two main approaches for extreme data selection and analysis, i.e. the Block Maxima Approach and the Peaks Over threshold approach, only for specific sites. On the other hand, in the second part, a hierarchical random effects model is built in order to better incorporate in the analysis the structural complexity of the data. In particular, it attempts to identify site and seasonal effects for the marginal densities of daily maximum wind gusts, as well as the serial dependence at each location through a first order Markov chain model.

Keywords: *Extreme Value theory*, *Bayesian statistics*, *Generalized Extreme Value distribution*, *Generalized Pareto distribution*, *Hierarchical random effects model*, *Markov chains*, *wind gusts*.

1. Introduction

1.1. Extreme Value Theory

Due to their rarity, events like epidemies, high atmospheric pollutions or natural disasters have long captured the public's attention and often they are shrouded in mystery.
A particular branch of statistics, Extreme Value Theory (EVT), offers insights to their inherent scarcity and stark magnitude.
Moreover, in the few last decades, the EVT literature has grown considerably, with applied interest in engineering, oceanography, environment and

economics, among others, and that is because it allows us extrapolate information beyond the range of available data.
Unlike other statistical approaches, which are more concerned with the 'center' of probability distributions, Extreme Value Theory focuses on the tails of the distributions, i.e. either on maxima or minimal values, the extreme deviations from the median.

1.2. Approaches for practical extreme value analysis

Given an ordered set of observations of a given random variable, there exist two approaches to select the extremes for the analysis.
The first approach is called *Block Maxima* approach and it relies in dividing the dataset into blocks sufficiently large and then, picking the maximum over each sample, deriving the block maxima series.
The second approach is called *Peak Over Threshold (POT)* and it consists in picking a threshold and extracting, from a continuous record, all the observation exceeding that threshold.
We'll explore the *Block Maxima* approach in section 2 and the *Peak Over Threshold* approach in section 3.

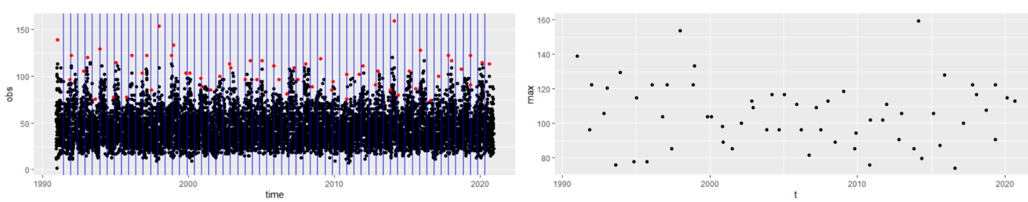

Figure 1: An example of picking Extreme values by means of the Block Maxima Approach, namely, for a site s, we have $x_{s,n}^{block-maxima} = \max\limits_{N\cdot(n-1)<t\leq N\cdot n} x_{s,t}$, where N is the dimension of the block.

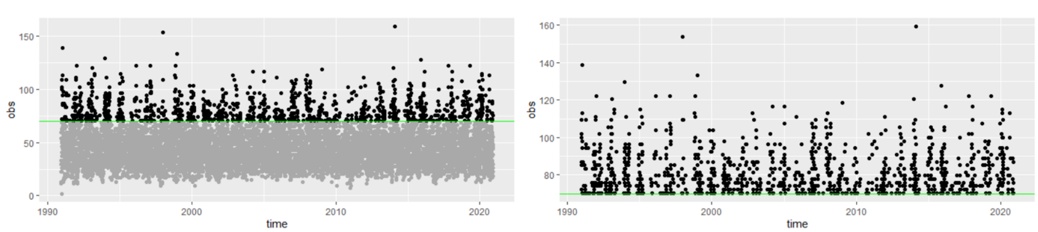

Figure 2: An example of picking Extreme values by means of the Peaks Over Threshold Approach, namely, for a site s, $x_{s,t}^{POT} = x_{s,t}\ \ s.t\ \ x_{s,t} \geq threshold.$

1.3. Extreme daily high gust in Ireland

In this paper we use a Bayesian approach to develop models for daily highest wind gusts over a region of central and southern Ireland. The data adopted for the analysis consists of daily maximum wind gusts recorded by the Irish National Meteorological Service at 5 locations (Fig 3), namely we considered data recorded by 5 different station, each of them situated in a different county (*Clare, Clarke, Dublin, Kerry* and *Westmeath*). For all the sites, data are recorded daily over a period of 30 years, from December 1st, 1990 to November 30th, 2020, constituting 54790 observations at each site.

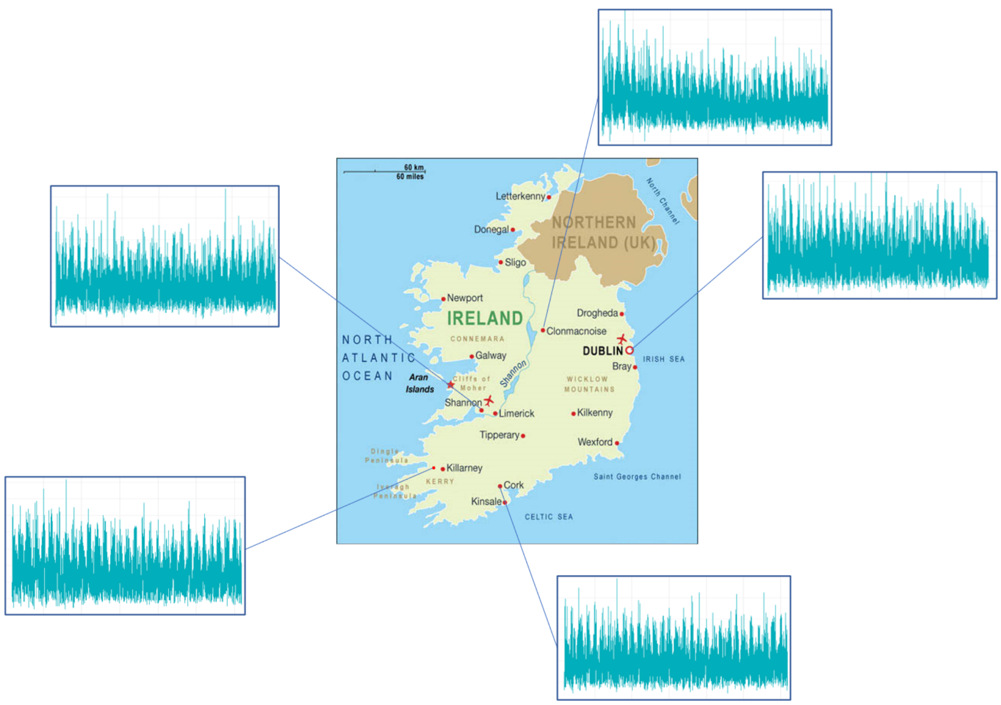

Figure 3: Map of the wind speed stations considered with the relative 30 years daily time series of maximum daily wind gust.

In Fig. 4 is illustrated an exploratory analysis of data from two contrasting sites, Kerry and Westmeath. The first is one of the most mountainous regions of Ireland, it is bounded on the west by the Atlantic Ocean and it is more often exposed to intense atmospheric phenomena; the second is a county in the center of Ireland, the territory is mostly constituted by hills and lakes and it has not boundaries with the sea. The effects of these differences are clearly seen in Fig 4, where a time series plot, a plot of the time series against the lagged series and an histogram of the observations is represented for both the counties.

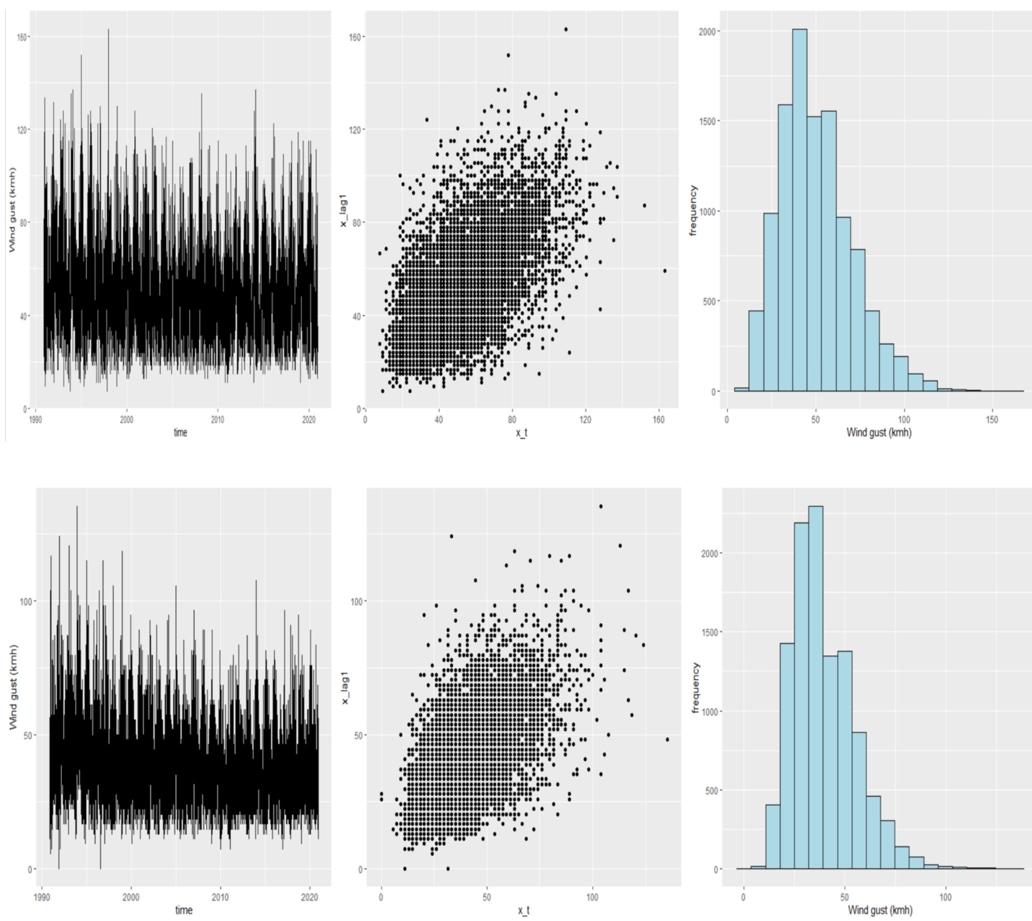

Figure 4: Time series plot, plot of the time series against the lagged series and histograms of daily maximum wind gusts in Kerry (above) and Westmeath (below).

1.4. The models

In the first part of this article we will focus on two simple models to analyze the EVT for daily highest wind gusts in **just one single site**, using respectively Block Maxima Approach and POT.
In the second part we will focus on a more complicated model which is a hierarchical random effects model that incorporates random effects for the sites, the monthly variation and the serial dependence that it is inherent in the time series of daily maximum speeds obtained at each site. As we have seen before, wind gusts can be very different in magnitude from one site to another, but it is also possible to find differences in intensity if we consider data observed in winter w.r.t data observed in summer for every single site. The aim of our hierarchical model is to use the complex structure inherent in the data to improve the "simplistic" inferential procedure adopted in the first models which are presented in section 2 and 3.
In [1],[2],[3], [4] and [8] it is highlighted in detail why using a Bayesian approach is a really good means of exploiting data on extreme events. The methods adopted to address our problems of inference are Markov Chain Monte Carlo methods, implemented by means of software such as STAN, JAGS and NIMBLE.

2. Block Maxima Approach Model

2.1. Theoretical background

As we said before, the Block Mamixa approach consists in dividing our entire dataset into blocks of size N sufficiently large and then extracting only the maximum from each block, so, per each site s, we will end up with a series of maxima $\{x_{s,1}^{block-maxima}, x_{s,2}^{block-maxima}, ..., x_{s,B}^{block-maxima}\}$, where B is the total number of blocks.
Extreme Value Theory has a central theoretical result for this approach, analogous to the Central Limit Theorem: the *Fisher-Tippet-Gnedenko theorem* (or *Extreme Value Theorem*).
This result is fundamental, indeed it states that our series of maxima follows asymptotically a *Generalized Extreme Value distribution*:

$$GEV(x) = exp\left\{ -[1+\xi(\frac{x-\mu}{\sigma})^{\frac{1}{\xi}}] \right\}, \tag{1}$$

defined on the set $\{x : 1+\xi(\frac{x-\mu}{\sigma}) > 0\}$, where μ, $\sigma > 0$ and ξ are *location, scale* and *shape* parameters respectively.

2.2. Model

This very first analysis will be focused, as we said at the beginning, on two specific counties, Kerry and Westmeath, due to their particular geographic position.
Since we know from the theory that taking blocks sufficiently large removes the short-range temporal dependence, we set the dimension equal to 14 days. Assumptions for long-range temporal dependence is granted by the Leadbetter's D condition (see [10]).
For each county, the maxima will follow the aforementioned distribution:

$$x_{s,\cdot}^{block-maxima}|\mu, \sigma, \xi \sim \mathcal{GEV}(\mu, \sigma, \xi), \tag{2}$$

where the subscript s = 1,2, used to distinguish the two counties.
In the absence of any expert prior information regarding the three parameters of the GEV distribution, we adopt a 'naive' approach using largely non–informative, independent priors for these, namely

$$\begin{aligned}
&\mu \sim \mathcal{N}(0,\ 100),\\
&\sigma \sim \mathcal{N}(0,\ 100) \quad with \ \ \sigma > 0,\\
&\xi \sim \mathcal{N}(0,\ 10).
\end{aligned}$$

We will be back on the discussion about the priors.
The algorithm is an Hamiltonian Markov Chain Monte Carlo method through STAN package in R. We run 4 chains for 6000 iterations with thinning equal to 10 and we discarded the first 1500.

2.3. Results

In this section we will present the result of our model.

In Fig. 5 we have the results of the chains for the three parameters μ, σ and ξ respectively. Although we have some variation in the sampling, the parameters follow an unimodal density.

In Fig 6 we have the return level plots over 100 years and we can clearly see the difference between the two counties. Indeed, in Kerry county, which is more exposed to extreme weather event, wind gusts can reach also 140 km/h, while in the inland of Westmeath county, the maximum wind gust is considerably inferior.

In Fig 7 there are the histogram and the pointwise predictive density.

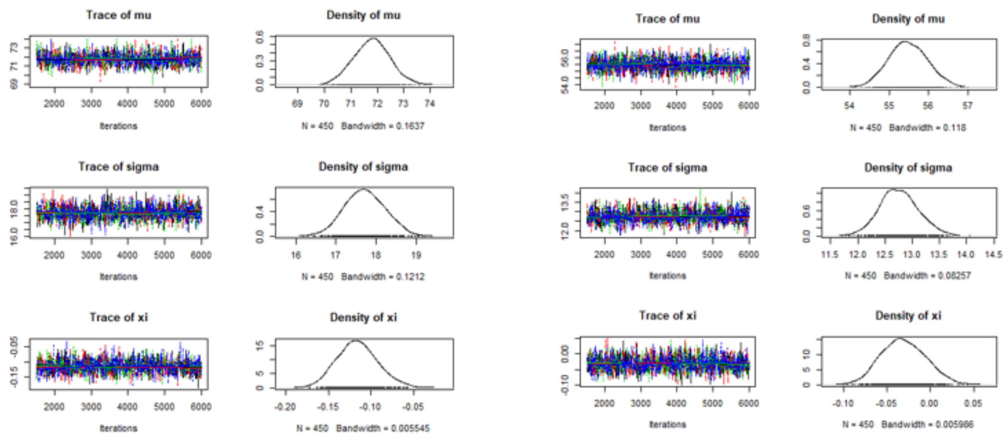

Figure 5: Trace plots: Kerry and Westmeath

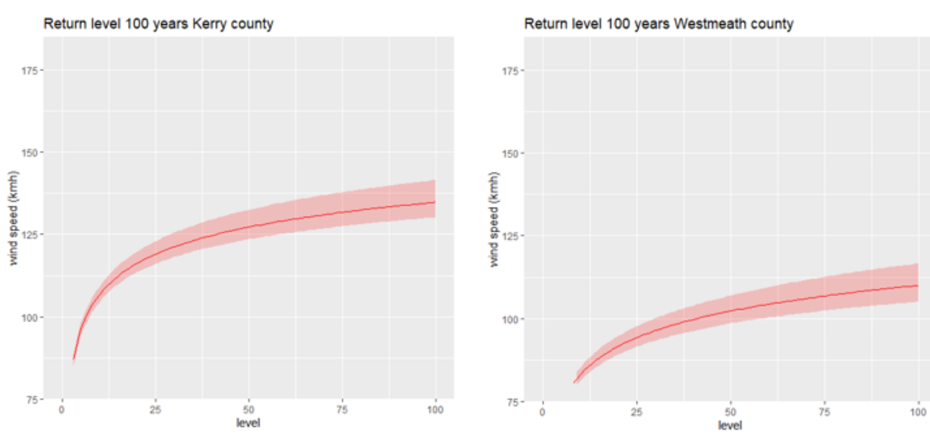

Figure 6: 100 years return level plots: Kerry and Westmeath. The dark red line represents the predictive return level curve, while the light red band is the 95% confidential interval for the values in the predictive return level curve

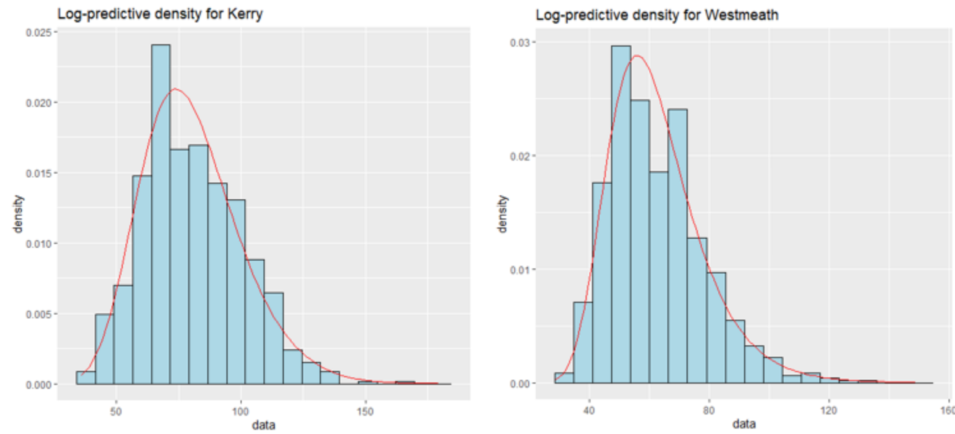

Figure 7: Pointwise predictive density with the relative histogram:
Kerry and Westmeath

2.4. Remark about priors

As we said before, we don't have any prior information about the parameters, so we ask ourselves if it would be possible, in some way, to extract information from our data.
We proceeded as follows: we divided our entire dataset in two halves and we implemented the model presented in subsection (2.2) on the first part in order to estimate mean and variance of each parameter.
We then compared two models for the second part of data via WAIC: the first one with non-informative priors again and the second one with some "informative" priors, i.e. normal distributions with mean and variance estimated in the first step. As you can see from the table below, the WAIC index evaluated through non-informative priors is always less than the other one with informative priors and this result suggests us to prefer the first one. Due to this fact we will always use non-informative priors in the following analysis.

County		*Non-informative*	*"Informative"*	*Mean of the prior*	*Variance of the prior*
Kerry	μ	68.93 (0.91)	70.90 (0.71)	73.66	1.14
	σ	16.64 (0.66)	17.93 (0.44)	18.88	0.56
	ξ	-0.13 (0.04)	-0.11 (0.001)	-0.11	0.001
Westmeath	μ	50.96 (0.63)	55.50 (0.41)	59.08	0.56
	σ	11.12 (0.46)	12.95 (0.25)	12.97	0.30
	ξ	-0.04 (0.04)	0.01 (0.001)	0.01	0.002
County		*WAIC index Non-informative*	*WAIC index "Informative"*		
Kerry		3383.7	3390.4		
Westmeath		3108.5	3167.0		

Table 1: Results of the priors analysis. In the two last columns there are the mean and the variance of the gaussians adopted as priors for the "informative" model, i.e. mean and variance of the posterior samples obtained from the first half of set of data. WAIC index evaluated as -2(lppd - p_{WAIC2}).

3. Peaks Over Threshold Model

3.1. Theoretical Background

The idea of the *Peaks Over Threshold* approach is to pick a high-enough threshold and fit a *Generalized Pareto Distribution (GPD)* to observations above the threshold. If x is an observation and u is the threshold, we'll call x *exceedance* if $x > u$ and we'll call $x - u$ *excess* over u. The fundamental theoretical result driving this approach is the *Pickans-Balkema-de Haan theorem*, which states that, under very broad conditions the limiting distribution of $(X - u|X > u)$, the excesses over the threshold u, as $u \to \infty$ is a *GPD*. The *CDF* for the *GPD* is:

$$H(x) = 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)_{+}^{-1/\xi}$$

defined on the set $\{x : u \leq x \wedge \xi \geq 0, u \leq x \leq u - \sigma/\xi \wedge \xi < 0\}$, where $a_{+} = \max 0, a$, $\sigma > 0$ is the *scale* parameter and ξ is the *shape* parameter.

3.2. Model

Also for the *Peaks Over Threshold* approach we will focus on data from the Kerry and Westmeath counties. Conventionally, the threshold u is fixed before fitting the model. It is picked by means of graphical exploration, for instance by looking at the *Mean residual life plot*. This has the effect of ignoring the uncertainty around the threshold selection for the rest of the analysis, therefore we decided instead to build a new model that assigns a prior to the threshold u. We will elicit this prior by looking at the results of the usual graphical exploration methodology and we'll include a degree of uncertainty that we find fitting. For each county s, we'll employ the following model for wind speed values:

$$x_{s,\cdot}^{POT,u_1}|\sigma, u, \xi \sim \begin{cases} \frac{1}{2}\frac{1}{(u-u_1)} & \text{if } u_1 \leq x < u \\ \frac{1}{2}h(x|\sigma, u, \xi) & \text{if } x \geq u \end{cases},$$

where u_1 and u_2 are the lower and upper bound for the threshold, $h(x|\sigma, u, \xi)$ is the *PDF* of the *GPD*. We don't fit the model with all of our data, instead we first follow a procedure called *declustering*, in order to overcome serial dependence of our extremes data: we consider two consecutive exceedances over u_1 to be in different clusters if we registered wind speeds below u_1 for

three days in a row between them. We consider as observation the maximum wind speed in each 3-days observations cluster.
We fit a uniform as part of the mixture for the observations below the threshold in order to let the model update the beliefs on the threshold. We tried several solutions including: fitting a gamma distribution below the threshold, having as mixture a convex combination between a gamma distribution and a *GPD* with threshold 0, ignoring the center of the distribution. The solution of which the results were the most sensible is the one of fitting a uniform for the center of the distribution. This solution could be interpreted as putting less weight on likelihoods for which the threshold takes a large value. As priors we use:

$$\log \sigma \sim \mathcal{N}(0,\ 10),$$
$$u \sim \mathcal{N}(u_s^{guess},\ \sigma^2_{u_s^{guess}}), \quad u \in (u_1, u_2),$$
$$\xi \sim \mathcal{N}(0,\ 10), \quad \xi \in \left(-\frac{\sigma}{x_{max} - u}, +\infty\right),$$

where u_s^{guess} is the threshold estimated by graphical means by us, and $\sigma^2_{u_s^{guess}}$ is the associated variance.

3.3. Results

We sample from the posterior thanks to JAGS. We ran 4 chains for 3000 iterations with thinning equal to 10 and we discarded the first 1500 iterations. In Fig. 8 we show the chains for σ, u and ξ, in Fig. 9 we show the predictive return level plots up to 100 years.

In order to compute the N-year return level z_N we consider the predictive distribution for the p_N-quantile of the *GPD* with parameters u, σ and ξ, given n_{py} the average number of observations per year after declustering, p_N is computed as:

$$p_N = 1 - \frac{1}{n_{py} * N}$$

We plot the median, 0.25-quantile and 0.75 quantile of the predictive distribution for the p_N-quantile of the *GPD* with parameters u, σ and ξ, as N goes from 1 to 100.
Finally in Fig. 10 we show the predictive density of the observations, notice that in both plots the fit in the center of the distribution is clearly not good. We are aware of this fact but decided to ignore it since the only reason we fit the center of the distribution, instead of only the tail, is for having a prior on the threshold.

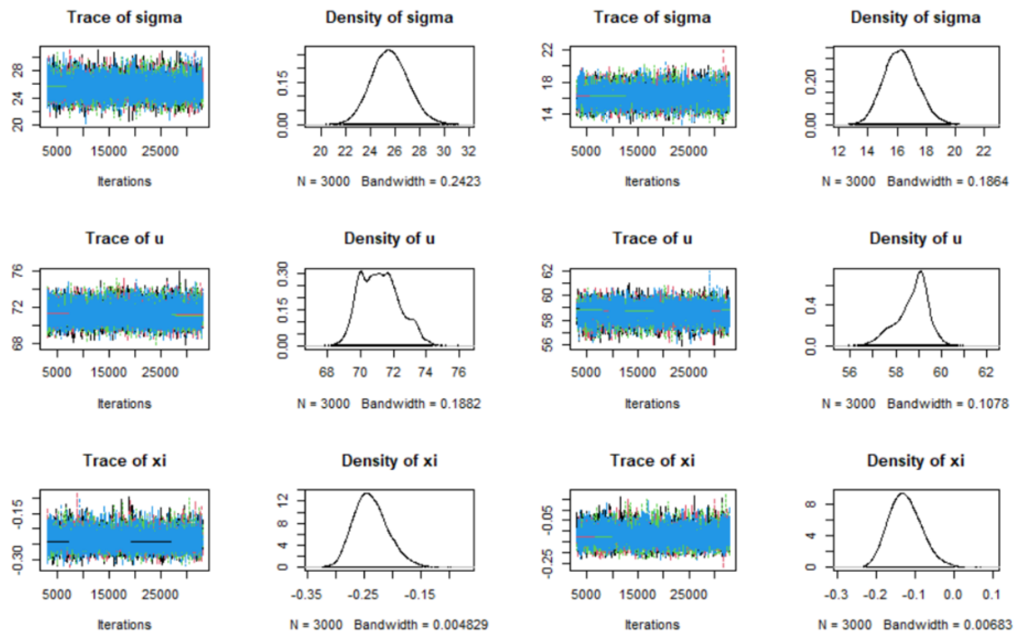

Figure 8: Trace plots: on the left Kerry, on the right Westmeath

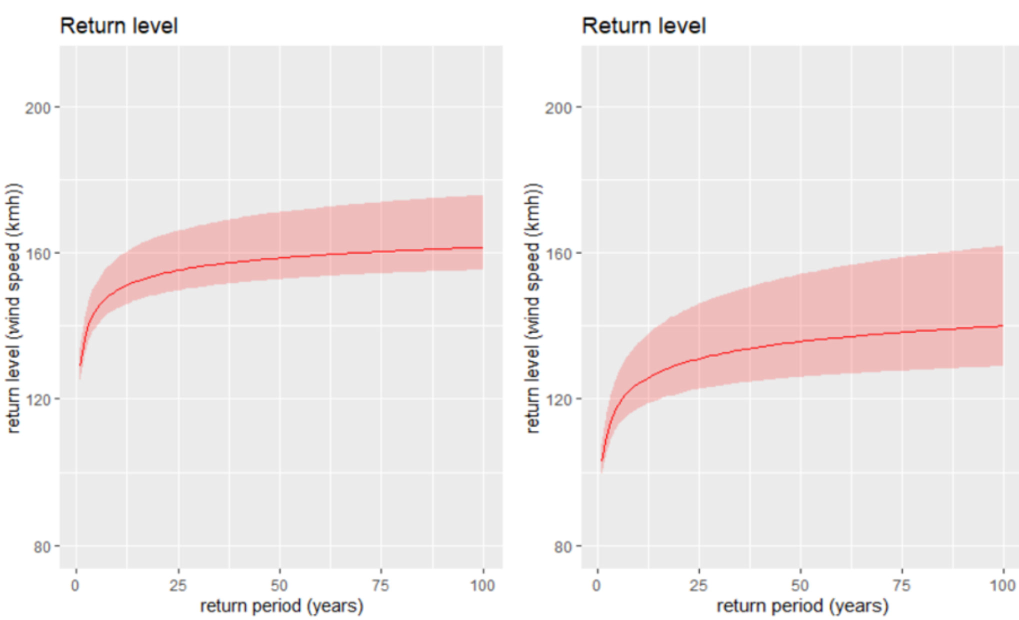

Figure 9: Predictive return level plots: on the left Kerry, on the right Westmeath

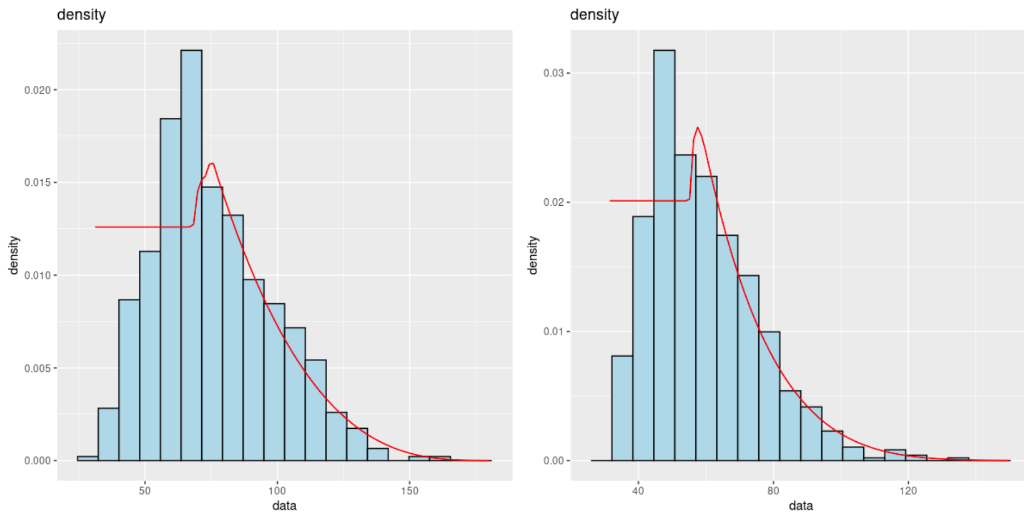

Figure 10: Predictive density: on the left Kerry, on the right Westmeath

	Kerry	*Westmeath*
u	71.20 (1.16)	58.74 (0.71)
σ	25.61 (1.50)	16.25 (1.15)
ξ	-0.24 (0.03)	-0.12 (0.04)
u_1	40	40
u_2	140	140

Table 2: Posterior mean and standard deviation for the parameter u, σ, ξ; u_1 and u_2 for the two different counties.

4. Hierarchical random effects model

4.1. Overview

As seen before, although the Block Maxima Approach is a valid model for EVT analysis, it can lead to a loss of significally consistent observations. For this reason, for the hierarchical model, we adopted a POT approach using the generalized Pareto distribution, which is a more natural way of modelling extremes of time series, as already considered in section (3).
For our purposes, we need the GPD parameters to vary across sites and seasonally.
In order to analyze seasonality, we simply partitioned the annual cycle into 12 months. Thus, our hierarchical model will be characterized by parameters pairs $\sigma_{s,m}, \xi_{s,m}$ for $s = 1, ..., 5$ sites and $m = 1, ..., 12$ months.
It is also necessary to allow the threshold u that is used for the selection of the extremes to vary, since there are different criteria about what constitutes an extreme value for each combination of season and site. So this yield to different thresholds $u_{s,m}$ for each site s and month m.
Even though in section (3) we considered the threshold as a parameter of the model, so that it was possible to include the variability of the threshold into the analysis, in the following hierarchical model, due to an already high complexity of the model, the values of the thresholds $u_{s,m}$ have been chosen empirically, by means of the empirical Mean Residual Life plot and the Threshold Choice plot.

4.2. Temporal dependence

In (3), we adopted a common and simple approach to handle the serial correlation concerning the extremes of our data, consisting in to identify "clusters" of extremes and to model only the cluster peaks, which are taken to be a set of independent exceedances.
However, it can be demonstrated (see [5]) that this procedure can be wasteful, since it discards the several extreme observations, and it can introduce bias into return level estimation, which would not otherwise be present. For those reasons, we decided to explicitily model the temporal dependence.
An investigative analysis of our data have shown that the partial autocorrelation functions for all sites indicate a very large value at lag 1, with all lags greater than 1 showing negligible partial autocorrelation. So we decided to use a simple first-order Markov chain model for the serial dependence of our data. Although this not necessarily imply that the temporal dependence is

entirely explained by the first-order dependence for the extreme values, it is a reasonable assumption for modelling.
The stochastic properties of such chain are completely determined by the joint distribution of successive pairs, so, given, for example, a set of observations $\{x^i_{s,m}\}_{i=1}^{N_{s,m}}$ for a particular site s in a particular month m, with $N_{s,m}$ denoting the number of observations in that site and month, we have that the likelihood for our parameters $(\sigma_{s,m}, \xi_{s,m})$ is:

$$
\begin{aligned}
L(\sigma_{s,m}, \xi_{s,m}, &\alpha_s | u_{s,m}, x^1_{s,m}, ..., x^{N_{s,m})}_{s,m} = \\
&= \prod_{i=2}^{N_{s,m}} f(x^i_{s,m}|x^{i-1}_{s,m}, \sigma_{s,m}, \xi_{s,m}, u_{s,m}, \alpha_s) \cdot f(x^1_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m}) \\
&= \frac{f(x^1_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m}) \cdot \prod_{i=2}^{N_{s,m}} f(x^i_{s,m}, x^{i-1}_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m}, \alpha_s)}{\prod_{i=2}^{N_{s,m}} f(x^i_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m})}
\end{aligned}
\tag{3}
$$

Contributions to the numerator can be modelled by using an appropriate bivariate extreme value model. One of the most flexible and accessible is the logistic model (see [9]).
For consecutive threshold exceedances, the appropriate form of this model is

$$F(x^{i-1}_{s,m}, x^i_{s,m}) = 1 - \{Z(x^{i-1}_{s,m})^{\frac{1}{\alpha_s}} + Z(x^i_{s,m})^{\frac{1}{\alpha_s}}\}^{\alpha_s}, \tag{4}$$

where the transformation Z, for a generic threshold $u_{s,m}$, GPD parameters $\sigma_{s,m}$ and $\xi_{s,m}$ and an exceedance rate per observation $\lambda_{s,m}$ is given by

$$Z(x) = \lambda^{-1}_{s,m} \cdot max(0, (1 + \frac{\xi_{s,m}(x - u_{s,m})}{\sigma_{s,m}})^{\frac{1}{\xi_{s,m}}})$$

and ensures that the margins are also of GPD form.
The parameter $\alpha_s \in (0, 1]$ measures the strength of dependence between consecutive extremes, with smaller values indicating stronger dependence. On the other hand, independence is obtained when $\alpha_s = 1$.
The parameter $\lambda_{s,m}$ is the empirical proportion of threshold exceedances w.r.t. all the observations for each location s and each month m.
Since we are considering the whole set of observations in our analysis, inference for this model is complicated by the fact that a bivariate pair may

exceed a specific threshold in just one of its components, namely let

$$\begin{aligned}
R_{0,0} &= (0, u_{s,m}) \times (0, u_{s,m}), \\
R_{1,0} &= [u_{s,m}, +\infty) \times (0, u_{s,m}), \\
R_{0,1} &= (0, u_{s,m}) \times [u_{s,m}, +\infty), \\
R_{1,1} &= [u_{s,m}, +\infty) \times [u_{s,m}, +\infty),
\end{aligned}$$

so that, for example, a point $(x^{i-1}_{s,m}, x^{i}_{s,m}) \in R_{1,0}$ if $x^{i-1}_{s,m}$ exceeds the threshold but $x^{i}_{s,m}$ does not. For points in $R_{1,1}$, model (4) applies and the relative density represents $f(x^{i}_{s,m}, x^{i-1}_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m})$ in 3. However for a point in $R_{1,0}$ and $R_{0,1}$ the corresponding likelihood component is given respectively by $\frac{\partial F}{\partial x^{i-1}_{s,m}}\Big|_{(x^{i-1}_{s,m}, u_{s,m})}$ and $\frac{\partial F}{\partial x^{i}_{s,m}}\Big|_{(u_{s,m}, x^{i}_{s,m})}$.
If $(x^{i-1}_{s,m}, x^{i}_{s,m}) \in R_{0,0}$ the contribution is give by the distribution function (4) evaluated at the threshold $u_{s,m}$.

For what concern the function $f(x^{i}_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m}) \quad \forall i = 1, ..., N_{s,m}$ in (3), it is defined as

$$f(x^{i}_{s,m}|\sigma_{s,m}, \xi_{s,m}, u_{s,m}) = \begin{cases} \frac{\lambda_{s,m}}{\sigma_{s,m}}(1 + \frac{\xi_{s,m}(x^{i}_{s,m} - u_{s,m})}{\sigma_{s,m}})^{(\frac{-1}{\xi_{s,m}} - 1)}, & if \;\; x^{i}_{s,m} \geq u_{s,m}, \; \xi_{s,m} \neq 0 \\ \\ \frac{\lambda_{s,m}}{\sigma_{s,m}}(exp(\frac{-(x^{i}_{s,m} - u_{s,m})}{\sigma_{s,m}})), & if \; x^{i}_{s,m} \geq u_{s,m}, \; \xi_{s,m} \to 0 \\ \\ \frac{1-\lambda_{s,m}}{u_{s,m}} & if \;\; x^{i}_{s,m} < u_{s,m} \end{cases}$$

which correspond to a GPD density function for the daily wind gusts greater or equal to the threshold and to a uniform density function over the interval $(0, u_{s,m})$ for the observations below the threshold.

4.3. Construction of the model

Summing up, we are considering a GPD model for exceedances over an high threshold, allowing this threshold and the GPD parameters to vary across sites and seasons. Since there is not a strong correlation in extremes across sites with the distance involved and the temporal dependence in extremes of wind speed is typically short lived, we make the assumption that extremes between months and between sites are independent, but successive extremes within a month have a first-order Markov dependence structure and

so can be modelled by means of the equation (3).
We also make the assumption that there is no interaction between seasonal and site effects. Moreover, due to the already high complexity of the model, we assume that both spatial effects and monthly effects are exchangeable a priori, losing little benefits in the precision of the analysis, but optimizing the overall model structure.
We denote by $(sigma_{s,m}, xi_{s,m})$ the parameters of the model (3) valid for threshold exceedances in month m and site s. The logistic dependence parameter α_s varies only between sites, considering the serial dependence in extremes fairly constant across all the months (see [6]).
The assumptions above lead us to specify the following random-effects model:

$$log(\sigma_{s,m}) = \epsilon_\sigma^{(s)} + \gamma_\sigma^{(m)},$$
$$\xi_{s,m} = \epsilon_\xi^{(s)} + \gamma_\xi^{(m)}$$

where we take $log(\sigma_{s,m})$ to retain the positivity of the scale parameter $\sigma_{s,m}$. Moreover, we add a constraint on xi in order to handle the threshold dependence of the model, namely

$$\xi_{s,m} > \frac{\sigma_{s,m}}{\max_i(x^i_{s,m}) - u_{s,m}} \quad \forall s = 1, ..., 5 \;\; and \;\; m = 1, ...12.$$

All random effects are normally and independently distributed:

$$\begin{aligned} &\epsilon_\sigma^{(s)} \sim \mathcal{N}(a_\sigma, 1/\zeta_\sigma), \\ &\epsilon_\xi^{(s)} \sim \mathcal{N}(a_\xi, 1/\zeta_\xi), \quad for \quad s = 1, ..., 5 \end{aligned}$$

for the site effect and

$$\begin{aligned} &\gamma_\sigma^{(m)} \sim \mathcal{N}(0, 1/\tau_\sigma), \\ &\gamma_\sigma^{(m)} \sim \mathcal{N}(0, 1/\tau_\xi), \quad for \quad m = 1, ..., 12 \end{aligned}$$

for the monthly effect, and

$$\alpha_s \sim uniform(0,\ 1) \quad for \quad s = 1, ..., 5$$

for the logistic dependence parameter variation between sites.

The final layer is then:

$$\begin{aligned} &a_\sigma \sim \mathcal{N}(0,\ 1000), \quad && a_\xi \sim \mathcal{N}(0,\ 1000), \\ &\zeta_\sigma \sim uniform(0,\ 1), && \zeta_\xi \sim uniform(0,\ 1), \\ &\tau_\sigma \sim uniform(0,\ 1), && \tau_\xi \sim uniform(0,\ 1), \end{aligned}$$

where all the hyperparameters of the distributions are chosen to give a highly non-informative specification, since the absence of expert prior knowledge.

4.4. MCMC simulations and results

The hierarchical random effect model presented has been implemented both in STAN and in NIMBLE. In STAN, the software uses an Hamiltonian MCMC, on the other hand, in NIMBLE, a random walk Metropolis-Hasting within Gibbs is adopted.
Even though the goodness of the posterior samples in STAN was already relevant running 3 chains for 1000 iterations, with 800 of burn-in, and thinning equal to 1, we decided to use both STAN and NIMBLE because the first one was particularly slow, due to the high complexity of the model (to run the 3 chain for 1000 iteration with 800 of burn-in it takes more than 14 hours on a common laptop). On the other hand, NIMBLE was considerably faster and very intuitive, but often it is less reliable, especially in the convergence of the chains, so, in order to better evaluate the reliability of the results with NIMBLE, we compared them with the ones in STAN, obtaining similar posterior samples.
To follow, we will present the results obtained using NIMBLE, namely we considered 3 chains and for each chain 15000 iterations, with 10000 iterations for the burn-in and thinning every 10th observation.
The results consist in samples from the approximate posterior distribution for the 5 site effect parameters and the 12 monthly effect parameters for each of $log(\sigma_{s,m})$ and $\xi_{s,m}$, and for the 5 site effect parameters α_s. This gives a total of 39 samples. In Fig 11 we show the resulting trace plot for Kerry and Westmeath, two "contrasting" sites, in December and July, which are the months in which it has been observed respectively the highest and the lowest maximum daily wind gusts. The trace plots demonstrate good mixing and apparent good convergence of the chains. In Fig 12 we show the trace plots for the logistic dependence parameter α_s, for Kerry and Westmeath. Table 3 summarizes the results shown in terms of the posterior means and standard deviations.

In particular, in the second part of table 3, we can see how the hierarchical model can fully catch the variability inherent in the data, highlighting differences between months and site, leading to a more precise and complete analysis of daily maximum wind gusts in the counties considered and improving the simplistic inferential procedure adopted in the sect.s (2) and (3). In fig 13 is represented the pointwise predictive density and how it fits our

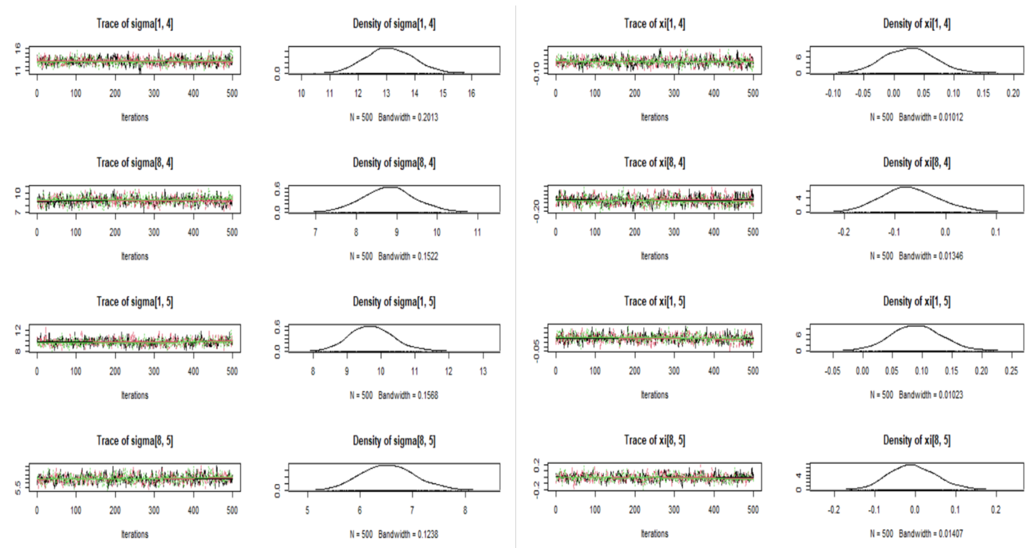

Figure 11: Trace plots of the parameters of the GPD, obtained as a sum of random effects. In the first 2 lines of plots are represented σ and ξ for Kerry, respectively in December and July. In the second 2 lines of plots are represented σ and ξ for Westmeath, respectively in December and July.

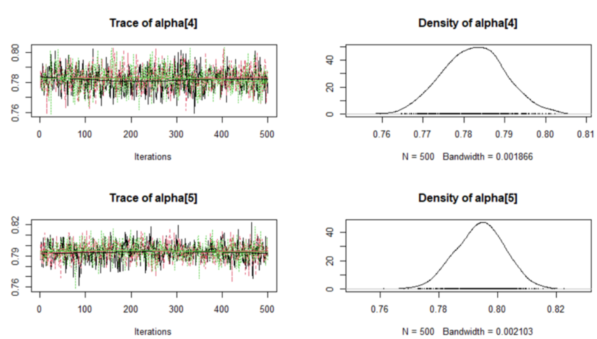

Figure 12: Trace plots of the dependence parameters in the logistic model 4. In the first line is represented α for Kerry, in the second line is represented α for Westmeath.

Effects		*Mean for Kerry, December*	*Mean for Kerry, July*	*Mean for Westmeath, December*	*Mean for Westmeath, July*
Monthly	$\gamma_\sigma^{(m)}$	-19.35 (0.38)	-19.75 (0.39)	-19.35 (0.38)	-19.75 (0.39)
	$\gamma_\xi^{(m)}$	-6.90 (0.30)	-7.00 (0.30)	-6.90 (0.30)	-7.00 (0.30)
Site	$\epsilon_\sigma^{(s)}$	21.93 (0.38)	21.93 (0.38)	21.63 (0.38)	21.63 (0.38)
	$\epsilon_\xi^{(s)}$	6.92 (0.30)	6.92 (0.30)	6.99 (0.30)	6.99 (0.30)
Parameters					
$\sigma_{s,m}$		13.11 (0.82)	8.81 (0.63)	9.73 (0.65)	6.54 (0.5)
$\xi_{s,m}$		0.026 (0.041)	-0.07 (0.05)	0.09 (0.04)	-0.006 (0.06)
α_s		0.782 (0.008)	0.782 (0.008)	0.794 (0.009)	0.794 (0.009)

Table 3: List of all the principal results of the analysis for Kerry and Westmeath in December and July, in particular, above are reported mean and, in brackets, the SD for the random effects and below the mean and the SD for the resulting parameters of the GPD and the logistic model.

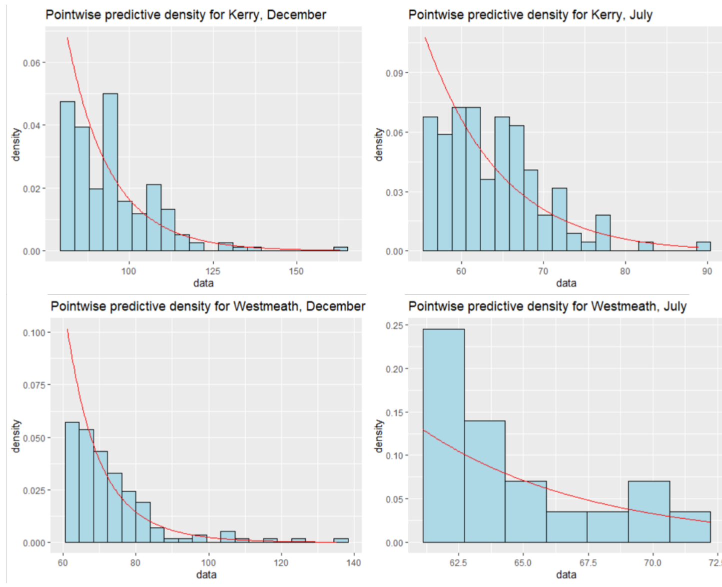

Figure 13: Pointwise predictive density (in red) with the relative histogram of the exceedances over the threshold for, starting from the figure in the higher left corner, Kerry in December, Kerry in July, Westmeath in December, Westmeath in July.

data, i.e. the exceedances over a selected threshold, which are illustrated by means of an histogram, for each site and each month.

In the last figure, fig 14, are shown the most important results in the analysis of extreme events: the estimate of extreme quantiles expressed as *return levels*. They gives a good approximation to the value of daily maximum wind gusts which are exceeded once every r years. They are often applied to understand the strength requirements for buildings and other large structures so that a certain level of wind speed can be withstood.
Here we represent the 100-years return levels for the two conties considered, Kerry and Westmeath, both in December and July.
A relevant more precision for the estimates of the return levels can be obtained by using the presented hierarchical model, which is a very important result to better study and evaluate the risks related to extreme wind speeds.

4.5. Considerations about the use of conjugate priors

In the construction of the model, subsection (4.3), we also considered the use of conjugate priors for ζ_σ, ζ_ξ, τ_σ and τ_ξ, i.e. non-informative Gammas

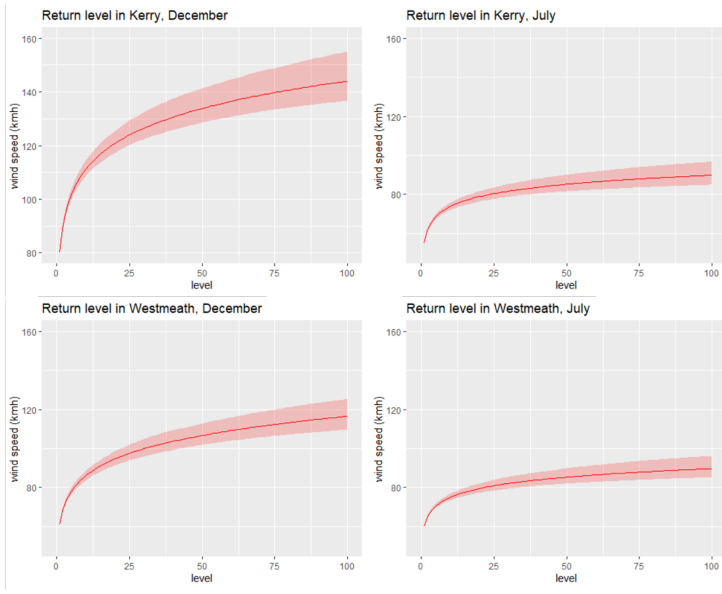

Figure 14: Return levels: In the first line of plots are represented the return levels of daily maximum wind gusts for Kerry, respectively in December and July. In the second line of plots are represented the ones for Westmeath.

instead of uniforms.
From an analytical point of view, this would have considerably simplified the sampling from the posterior distributions for those parameters, however, in STAN, since the software is based on Hamiltonian MCMC in which conjugacy isn't considered, the introduction of gamma priors only increased by far the complexity of the analysis and made the sampling quality worse.
In NIMBLE instead, since the software uses Gibbs sampling when there is conjugacy, it has been possible to compare the two models, the one with Gamma priors and the one with uniform priors presented before. In order to compare the two models we adopted the WAIC computed in NIMBLE, which is slightly different from the one presented at the end of section (2), namely here it is considered $WAIC = lppd - p_{WAIC2}$.
The values obtained are $WAIC = -300624$ for the model in which ζ_σ, ζ_ξ, τ_σ and τ_ξ have $\Gamma(0.01, 0.01)$ priors and $WAIC = -296723$ for the model presented in subsection (4.3), showing that the preferable model is the one considered in our analysis above.

5. Conclusions

In this article, we considered Extreme value theory to study Irish daily wind gusts, in order to obtain a behaviour and a good estimation of extreme daily wind speeds in 5 different counties of Ireland. The first two models, presented in sections (2) and (3), focus on the study of all the extremes in the 30 years observations just in one single site at time. The lower complexity of the structure of the two models carries a more simple inference, incorporating however less variability inherent in the data and bringing a lack in the precision of the analysis, especially in the estimation of return levels.

For what concern the Block Maxima Approach, already in the selection of the extremes, values meaningful to the inference can be discarded, that's why, especially in the EVT applied to weather data, the use of Peaks Over Threshold is often preferred. Moreover, in the Peaks Over Threshold Approach presented, an important result has been attained: we managed to include in the model the uncertainty related to the choice of the threshold, which often in literature is done empirically, losing all the variability included in the selection of the threshold, useful for a more accurate analysis.

In the end, a hierarchical model for the Irish daily wind gusts has been employed, as a natural framework for modelling the variation that is inherent in the data under consideration. The underlying structure of the data, in terms of monthly and spatial variation, has been estimated through the model, allowing the sharing of information between sites and months. This provides apparent benefits in terms of a greatly improved precision in estimation of the model parameters and, moreover, estimative inferences for the return levels have also improved in precision as a consequence.

Although we used non-informative priors in our analysis, it is of course possible and appealing to incorporate other sources of information through prior distribution if they are present, especially because extreme data usually, by their very origin, are scarce. Naturally, the model framework developed is transferable to other environmental variables for which measurements are available over a spatial network of locations.

References

[1] L.Fawcett,D. Walshaw. *Modelling Environmental Extremes* Short Course for the 19th Annual Conference of The International Environmetrics Society, The University of British Columbia kanagan, Kelowna, Canada (2008)

[2] Jacob M., Neves C., Vukadinović Greetham D. *Extreme Value Theory. In: Forecasting and Assessing Risk of Individual Electricity Peaks* Mathematics of Planet Earth. Springer (2020), Cham.

[3] Behrens, C.N., Lopes H.F., Gamerman D. *Bayesian analysis of extreme events with threshold estimation* Statist. Mod. (2004), 4, 227–244.

[4] N.A.M. Amina, M.B. Adam, A.Z. Aris *Bayesian Extreme for modeling high PM10 concentration in Johor* Procedia Environmental Sciences 30 (2015) 309 – 314

[5] Fawcett L. and Walshaw D. , *Improved estimation for temporally clustered extremes.* Environmetrics: The official journal of the International Environmetrics Society 18.2 (2007): 173-188.

[6] Fawcett L. and Walshaw D., *A hierarchical model for extreme wind speeds.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 55: 631-646 (2006).

[7] Scarrott C., and MacDonald A. *A review of extreme value threshold estimation and uncertainty quantification.* REVSTAT–Statistical Journal 10.1 (2012): 33-60.

[8] Coles, Stuart G., and Jonathan A. Tawn. *A Bayesian Analysis of Extreme Rainfall Data.* Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 45, no. 4, (1996), pp. 463–478.

[9] Coles S., Bawa J., Trenner L. and Dorazio, P. *An introduction to statistical modeling of extreme values* (Vol. 208, p. 208). London: Springer, (2001).

[10] Leadbetter, M.R. *Extremes and local dependence in stationary sequences.* Z. Wahrscheinlichkeitstheorie verw Gebiete 65, 291–306 (1983).