Nat. Hazards Earth Syst. Sci. Discuss., 2, 799–824, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/799/2014/
doi:10.5194/nhessd-2-799-2014

Natural Hazards and Earth System Sciences Discussions
Open Access

Bayesian trend analysis of extreme wind using observed and hindcast series off Catalan coast, NW Mediterranean Sea

M. I. Ortego[1], J. J. Egozcue[1], and R. Tolosana-Delgado[2,*]

[1]Universitat Politècnica de Catalunya, Departament de Matemàtica Aplicada III. Campus Nord, Barcelona, Spain
[2]Maritime Engineering Laboratory, Universitat Politècnica de Catalunya (LIM-UPC), Campus Nord, Barcelona, Spain
[*]now at: Helmholtz Institute Freiberg for Resource Technology, Halsbrueckerstr. 34, 09559, Freiberg, Germany

Received: 4 December 2013 – Accepted: 7 January 2014 – Published: 29 January 2014

Correspondence to: M. I. Ortego (ma.isabel.ortego@upc.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

It has been suggested that climate change might modify the occurrence rate and magnitude of large ocean-wave and wind storms. The hypothesised reason is the increase of available energy in the atmosphere-ocean system. Forecasting models are commonly used to assess these effects, given that good quality data series are often too short. However, forecasting systems are often tuned to reproduce the average behavior, and there are concerns on their relevance for extremal regimes. We present a methodology of simultaneous analysis of observed and hindcasted data with the aim of extracting potential time drifts as well as systematic regime discrepancies between the two data sources. The method is based on the Peak-Over-Threshold (POT) approach and the Generalized Pareto Distribution (GPD) within a Bayesian estimation framework. In this context, storm events are considered points in time, and modelled as a Poisson process. Storm magnitude over a reference threshold is modelled with a GPD, a flexible model that captures the tail behaviour of the magnitude distribution.

All model parameters, i.e. shape and location of the magnitude GPD and the Poisson occurrence rate, are affected by a trend in time. Moreover, a systematic difference between parameters of hindcasted and observed series is considered. Finally, the posterior joint distribution of all these trend parameters is studied using a conventional Gibbs sampler. This method is applied to compare hindcast and observed series of 10 min average wind speed at a deep buoy location off the Catalan coast (NE Spain, Western Mediterranean; buoy data from 2001; REMO wind hindcasting from 1958 on). Appropriate scale and domain of attraction are discussed, and the reliability of trends in time are addressed.

1 Introduction

Interest on natural hazard prevention, prediction and mitigation has increased along the last decades: strong wind storms, extreme wind gusts, hurricanes and tornados are not an exception. The dangers, that a climate change might induce, add an additional challenge to the statistical analysis of extreme wind data as possible trends in extremal winds might occur, increasing the inherent difficulties of extremal analysis. Performing extremal analysis requires long records spanning decades, even centuries for characterizing the rarest events. This need is exacerbated when the data series are potentially affected by a trend in time. Indeed, whichever model is fitted, the uncertainty of the estimates is large when only data series with few events are available.

Bayesian methods have been introduced in the field of natural hazards two decades ago and they have succeeded as a flexible and consistent way of controlling uncertainty (Sánchez-Arcilla et al., 2008; Coles and Tawn, 1996; Gelman et al., 1995). Simultaneously, the models for analysing extremal events have evolved. The models known as *excesses over threshold* or peak over threshold (POT) are nowadays of common use in natural hazards and, particularly, in wind hazard modelling (Walshaw, 1994; Palutikof et al., 1999; Della-Marta et al., 2009; Coles, 2001). However, trend analyses require non-stationary models while the standard POT models are based on the assumption of stationary of the time process. Non-stationary POT models are available, with applications in different frameworks (Beguería et al., 2011; Hundecha et al., 2008; Tramblay et al., 2011).

An obvious way of reducing the uncertainty of the estimates of both extremal parameters and their trends consists in analysing data spanning a longer time period. As direct observations cannot be extended beyond the availability of measuring devices at the site of interest, hindcast or forecast data series may offer an alternative source of information. However, it is known that hincast wind data seldom fit exactly the directly observed wind (Bolaños et al., 2004). There are several reasons for the misfit: possible mis-calibration of models; the fact that models actually aim at giving a kind of energy average, not quasi-instantaneous (e.g., 10 min average) winds; etc. However, when interest is focused in climatic features as extreme event statistics, hindcast data information is very valuable. Nevertheless, the joint use of directly observed and hindcast data introduces further complication in the models to be used, as a kind of conciliation between the two types of data must be considered.

This contribution aims at the joint analysis of non-overlapping hindcast wind data and data coming from direct measurements. This analysis evaluates the differences between both types of observations and check for (linear) trends in the extremal parameters. The procedure is based on a non-stationary POT model which is analysed using Bayesian techniques.

Section 2 describes the data and the site briefly. Section 3 presents the model and the estimation methods. Finally, results are presented in Sect. 4.

2 Wind data

The statistical properties of a hindcast model and a series of observations are compared. A series of 10 min average wind speed series has been constructed, joining actual data measured at a deep buoy location and hindcast data provided by a REMO model within the HIPOCAS project (Sotillo et al., 2005; Guedes Soares et al., 2002). Wind speeds at the Tarragona deep buoy (XIOM network, longitude 40.68° N, latitude 1.47° E) are intermitently available between the years 2004 and 2012, totalizing 7 yr and 158 days with observations. The series of wind speed at the nearest HIPOCAS grid point (longitude 40.50 N, latitude 1.50 E) is available for the period 3 January 1958–31 December 2001. Both locations are shown in Fig. 1. Admitting that the statistical parameters of the model may evolve over time, it is possible to assess evidences of differences between series (h for hindcast REMO and b for buoy) as well as evidences of linear trends for the parameters.

Events have been defined in the following manner: an event starts when the recorded 10 min average wind speed in the reference time series is greater than 15 $\mathrm{m\,s^{-1}}$. It finishes when the recorded 10 min average wind speed is less than $15\,\mathrm{m\,s^{-1}}$ and remains at this level for at least 3 days. Therefore, the minimum gap between events is 3 days and we ensure that events are approximately independent. The magnitude of the event is defined as the maximum 10 min average wind speed recorded during the event, and the corresponding time of occurrence as the instant of recording of this value. The $h_0 = 15\,\mathrm{m\,s^{-1}}$ threshold ensures that the excesses have approximately a GPD distribution. As wind speed measurements may have a ratio (relative) scale, the studied magnitude is log-windspeed (Egozcue et al., 2006). The data set is displayed in Fig. 2

3 Methods

3.1 The Peak over threshold model

Models of excesses over a threshold, often named *peak over threshold* (POT Embrechts et al., 1997), have been extensively used in hazard analysis of natural phenomena (Davison and Smith, 1990; Egozcue and Ramis, 2001; Coles et al., 2003; Egozcue et al., 2006; Leadbetter, 1991). Independently of the specific estimation method, the POT framework consists of a system of modelling assumptions. The standard assumptions for stationary process (Embrechts et al., 1997) can be summarised as follows:

- events are identified as points occurring in time;
- the elapsed times between consecutive events are random, identically distributed and mutually independent;
- each event has an associated random magnitude;
- excesses of magnitudes over a given threshold are identically distributed and mutually independent;
- elapsed times and magnitude excesses are mutually independent.

In this study of wind events we adopt a slightly modified POT model in order to cope with non-stationarity. The assumption of identically distributed elapsed times and magnitude excesses is changed: both random variables have distributions in a family which parameters can change with time and other observed parameters of the event. Also the statements of independence are modified to conditional independence, i.e. for fixed parameters of the distribution of elapsed times and magnitude excesses, random observations are independent.

In the stationary case the elapsed times are usually assumed exponentially distributed, thus the number of excesses over the selected threshold of magnitude is an homogeneous Poisson process. Also a common approach is to assume that excesses over the threshold have a generalized Pareto distribution (GPD), as proposed in Davison and Smith (1990) or in Embrechts et al. (1997). In the present approach, the occurrence of events with magnitude over the threshold is assumed to be an inhomogeneous Poisson process. Excesses over the threshold are modelled by a GPD, although considering that its parameters might depend on time and data source.

3.2 Occurrence of wind events

Wind events were defined in Sect. 2 and their wind speed is assumed to be larger than the threshold $15\,\mathrm{m\,s^{-1}}$. The main assumption is that their occurrence in time corresponds to an inhomogeneous Poisson process (Grandell, 1997; Coles, 2001). In homogeneous Poisson process the parameter is the Poisson rate, interpreted as the mean number of events per year. When this mean is not constant, the Poisson rate is replaced by the Poisson intensity $\lambda(t)$. If T is the random time from an origin to the next event, and $(t, t+\mathrm{d}t)$ denotes a short enough time interval, then $\lambda(t)$ is

$$\lambda(t)\,\mathrm{d}t \simeq \Pr[t < T \leq t + \mathrm{d}t | T > t] = \frac{F_\mathrm{T}(t+\mathrm{d}t) - F_\mathrm{T}(t)}{1 - F_\mathrm{T}(t)}, \quad (1)$$

for $t > 0$, where F_T is the cumulative distribution function (cdf) of T. Therefore, $\lambda(t)dt$ is the probability of the event occurring in $(t, t+dt]$, conditional to the event not occurring before t. As a consequence of Eq. (1), the cdf and probability density function (pdf) of T, for $t > 0$, are

$$F_{\mathrm{T}}(t) = 1 - \exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right),$$

$$f_{\mathrm{T}}(t) = \lambda(t)\,\exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right),$$

respectively. The Poisson intensity is modelled as depending on time in two ways. First, we assume a linear trend on time of the Poisson rate, in order to discuss evidences in favor or against the presence of a trend in time. On the other hand, wind events come from two different sources, REMO hindcasts and observations on a buoy. The Poisson intensity corresponding to the two sources might differ. This possibility is modelled as a systematic difference, which can be viewed as a step function on time or as an indicator function of the data source.

Suppose that wind events detected by REMO took place at times $t_0, t_1, \ldots, t_n$, and those observed by the buoy at times $\tau_0, \tau_1, \ldots, \tau_m$, with $t_n < \tau_0$. The available data are the elapsed times between consecutive events, i.e. $t_i - t_{i-1}$, $i = 1, 2, \ldots, n$ (REMO events) and $\tau_i - \tau_{i-1}$, $i = 1, 2, \ldots, m$ (buoy events). Then, the whole time interval covered by observations is (t_0, τ_m). The Poisson intensity can be written as

$$\lambda(t) = \lambda_h + \frac{\alpha_\lambda}{\tau_m - t_0} t + \delta_\lambda \mathrm{I}_{\mathrm{b}}, t_0 \le t \le \tau_m,$$

where λ_h is the Poisson intensity at the first event at t_0 which corresponds to a REMO observation; δ_λ is the increment of Poisson intensity due to the fact that the observation comes from the buoy. The symbol I_{b} is an indicator with value equal 1 for event times in which events are recorded by the buoy, and 0 otherwise. The parameter α_λ is the total

NHESSD
2, 799–824, 2014

increase of the Poisson intensity from t_0 to τ_m due to the linear trend of the Poisson intensity irrespective to the type of observation.

In order to proceed to a Bayesian estimation of the parameters the likelihood function of $(\lambda_h, \alpha_\lambda, \delta_\lambda)$ given the elapsed times is required. Since the elapsed times are assumed independent, given the parameters, the required likelihood is

$$L_{\mathrm{T}}(\lambda_h, \alpha_\lambda, \delta_\lambda | \{t_i\}, \{\tau_j\}) = \prod_{i=1}^{n} \lambda(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda(s)\, \mathrm{d}s \right) \times \prod_{j=1}^{m} \lambda(\tau_j) \exp\left(-\int_{\tau_{j-1}}^{\tau_j} \lambda(s)\, \mathrm{d}s \right), \tag{2}$$

where, for simplicity, $\{t_i\}, \{\tau_j\}$ represents all available data on elapsed times between consecutive events. When interruptions of observations are present, as it is the case in the buoy data set. Some time differences $\tau_j - \tau_{j-1}$ do not correspond to elapsed times between consecutive events; these time intervals are ignored in the likelihood (see further details in Ortego et al., 2012), i.e. the likelihood (Eq. 2) is valid when there are no interruptions in the observations. After grouping integrals within the exponentials, the log-likelihood $\ell_{\mathrm{T}} = \log L_{\mathrm{T}}$, is reduced to

$$\begin{aligned} \ell_{\mathrm{T}}(\lambda_h, \alpha_\lambda, \delta_\lambda | \{t_i\}, \{\tau_j\}) = &\sum_{i=1}^{n} \log \left(\lambda_h + \frac{\alpha_\lambda (t_i - t_0)}{\tau_m - t_0} \right) \\ &+ \sum_{j=1}^{m} \log \left(\lambda_h + \delta_\lambda + \frac{\alpha_\lambda (\tau_j - t_0)}{\tau_m - t_0} \right) \\ &- \lambda_h (t_n - t_0) - (\lambda_h + \delta_\lambda)(\tau_m - \tau_0) \\ &- \frac{\alpha_\lambda}{2(\tau_m - t_0)} (t_n^2 - t_0^2 + \tau_m^2 - \tau_0^2). \end{aligned} \tag{3}$$

NHESSD
2, 799–824, 2014

Bayesian trends of extreme wind

M. I. Ortego et al.

The log-likelihood in Eq. (3) is essentially that proposed in Ortego et al. (2012) for ocean waves, although a bit simpler due to the fact that in the present data set there is no overlapping in time between the hindcast and buoy observations.

3.3 Event magnitude model

The wind magnitude associated with each event can be selected in different ways. Traditionally wind speed magnitude is taken as velocity in $\mathrm{m\,s^{-1}}$ without any additional consideration. However, the natural scale and domain of wind speed should be taken into account. Large wind speeds do rather behave in a ratio scale (also known as relative scale), as shown by the thresholds chosen for most cyclone classification systems and the Beaufort scale levels of 7 or more. In fact, the absolute scale is near to be meaningless: a difference of $1\,\mathrm{m\,s^{-1}}$ on a reference wind speed of $2\,\mathrm{m\,s^{-1}}$ represents a factor 3/2, whereas the same difference on a reference wind speed of $20\,\mathrm{m\,s^{-1}}$ is considered almost irrelevant. A simple way of considering these issues on scale is to take logarithms on wind speeds, and accordingly, consider that the magnitude associated with events is the logarithm of the maximum 10 min average wind speed. Accordingly, the magnitude of a wind event is herein taken as the natural logarithm of the measured wind-speed in $\mathrm{m\,s^{-1}}$.

As we consider that events occur following a non-homogeneous Poisson process, magnitudes are assumed conditionally independent from event to event. In order to model magnitude excesses over a threshold of $h_0 = \log(15\,\mathrm{m\,s^{-1}})$, we assume that they follow a Generalized Pareto Distribution (GPD). The GPD provides a suitable asymptotic model for excesses over a high enough thresholds (Pickands III, 1975; Davison and Smith, 1990; Dupuis, 1998). Furthermore, we also consider that excesses over h_0, must be GPD in the Weibull domain of attraction, i.e. the support of the excesses has an unknown finite upper limit. This assumption is based on the fact that, on Planet Earth, wind speed is physically limited and, accordingly, the existence of such upper bound is granted, even though the limit itself is not known. This assumption was successfully taken in the extremal analysis of other weather magnitudes, as rainfall and ocean wave-heights (Pawlowsky-Glahn et al., 2005; Egozcue et al., 2005, 2006; Sánchez-Arcilla et al., 2008).

For a magnitude X, denote the excess by $Y = X - h_0$ conditional to $X > h_0$. One of the standard parameterisations of the GPD is (Embrechts et al., 1997)

$$F_Y(y) = 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi}, 0 < y < y_{\text{sup}}, \tag{4}$$

where $\beta > 0$ is a scale parameter, ξ is a shape parameter, and y_{sup} is the upper limit of the distribution support. When $\xi \geq 0$, the GPD has an unlimited tail, i.e. $y_{\text{sup}} = +\infty$, and belongs to the Fréchet domain of attraction of maxima ($\xi > 0$); for $\xi = 0$, the GPD is in the Gumbel domain of attraction. Under the assumption that the GPD has an upper bound, the shape parameter must be $\xi < 0$ (Weibull domain of attraction). In order to take into account the restriction $\xi < 0$, other parametrisation of GPD in the Weibull domain of attraction were proposed (Ortego et al., 2010, 2012); we adopt here the latter parameterisation. The new parameters, only valid for $\xi < 0$, are $\mu = \log(-\beta/\xi) = \log(y_{\text{sup}})$, $\nu = \log(-\xi)$ or, conversely, $-\xi = \exp(\nu)$, $\beta = \exp(\nu)/\exp(\mu)$. Introducing these parameters in (Eq. 4), the cumulative distribution function (cdf) and the corresponding probability density (pdf) are

$$F_Y(y|\mu,\nu) = 1 - \left(1 - \frac{1}{\exp(\mu)} y\right)^{\exp(-\nu)}, \tag{5}$$

$$f_Y(y|\mu,\nu) = \frac{1}{\exp(\nu)\exp(\mu)} [1 - \exp(-\mu) y]^{\frac{1-\exp(\nu)}{\exp(\nu)}}, \tag{6}$$

for $0 \leq y < y_{\text{sup}}$, respectively.

We have introduced two assumptions, namely, the magnitude to be treated is the log-wind speed and that the excesses over h_0 must have a limited distribution in the Weibull domain of attraction. The compatibility of both assumptions can be checked on the available data. Figure 3 shows the likelihood contours of raw and log-transformed wind speed for both data series.

In both data sources, the likelihood corresponding to the raw data cover a substantial region with $\xi > 0$ (Fréchet domain of attraction), whereas, for log-transformed data the coverage is considerably shifted towards values of $\xi < 0$ (Weibull domain of attraction). Table 1 shows the probabilities of the Weibull domain of attraction for raw and log-transformed wind speed at the location for both data series.

Buoy data are likely to correspond to a GPD with $\xi < 0$ both in the raw and log scales, but this is not the case for hindcast data. In a raw scale, the Fréchet domain of attraction is more likely for this second data series. These results suggest that compatibility between domains of attraction of REMO data and buoy data can only be achieved if both are considered in logarithms, a further argument in favor of considering a ratio scale for large wind speeds.

The upper bound parameter $\mu = \log y_{\text{sup}}$ is so large that it only depends on universal physical laws and geometric aspects of the Earth. We consider that human-scale climatic changes cannot change it, and it is therefore constant over time (Ortego et al., 2012). We also assume that μ may be different for hindcast data and data observed at the buoy. The proposed model for μ is then

$$\mu(t) = \mu_h + \delta_\mu \cdot \mathrm{I}_b, \tag{7}$$

where the difference δ_μ represents systematic regime differences between buoy observations and the hindcast data.

On the other hand, the GPD parameter ν might be affected by a (linear) trend in time, apart of the possible systematic regime differences between the buoy and the REMO series. Accordingly, the model proposed is

$$\nu(t) = \nu_h + \delta_\nu \cdot \mathrm{I}_b + \alpha_\nu \cdot t, \quad 0 < t \leq t_N - t_0, \tag{8}$$

where δ_ν is the difference in ν between the series h (hindcast REMO) and b (buoy). The parameter α_ν is the total drift of ν from t_0 to t_N, from the start to the hindcast series to the end of buoy observations.

NHESSD
2, 799–824, 2014

The excess magnitude Y has been recorded for each event, for both series h and b. The data set of pairs of occurrence times and excesses is $\{(t_i, y_i), i = 1, \ldots, N\}$. It is not necessary to consider the distinction of occurrence times of series h and b. Times are generically denoted as t_i and the notation in Sect. 3.2 is no longer necessary: $N = n + m$ and τ_i's are a subset of the t_i's in the present section. The likelihood of the model parameters, given the sample is

$$L(\mu_0, \delta_\mu, \nu_0, \delta_\nu, \alpha_\nu | \{(t_i, y_i)\}) = \prod_{i=1}^{N} f_Y(y_i | \mu(t_i), \nu(t_i))$$
$$= \prod_{i=1}^{N} \frac{1}{\exp(\nu(t_i))\exp(\mu(t_i))} [1 - \exp(-\mu(t_i)) y_i]^{\frac{1-\exp(\nu(t_i))}{\exp(\nu(t_i))}} . \tag{9}$$

The corresponding log-likelihood, for $0 < y_i < y_{\text{sup}}$, is

$$\ell(\mu_0, \delta_\mu, \nu_0, \delta_\nu, \alpha_\nu | \{(t_i, y_i)\}) = -\sum_{i=1}^{N} (\nu(t_i) + \mu(t_i))$$
$$+ \left(\frac{1 - \exp[\nu(t_i)]}{\exp[\nu(t_i)]} \right) \sum_{i=1}^{N} \log[1 - \exp(-\mu(t_i)) y_i], \tag{10}$$

which will be used in the Bayesian estimation of the parameters.

3.4 Bayesian estimation

Bayesian methods (e.g. Gelman et al., 1995) are very useful in contexts where data are scarce, such as hazard assessment problems. Bayesian methods allow to assess the uncertainty of the estimates of parameters, usually large in these situations. Parameters are assumed to be random variables, and uncertainty is described through their distribution.

A prior distribution is established for all parameters. This distribution encodes our knowledge about their likely values. As a priori assumption, parameters of the event magnitude distribution and parameters of the occurrence model are taken as independent in this analysis, i.e. parameters $\boldsymbol{\theta} = (\mu_0, \nu_0, \delta_\mu, \delta_\nu, \alpha_\nu)$ and $\boldsymbol{\phi} = (\lambda_h, \lambda_b, \alpha_\lambda)$ are independent. Therefore, the prior joint density of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ is $\pi^0(\boldsymbol{\theta}) \times \pi^0(\boldsymbol{\phi})$.

The likelihood of the parameters given the data, $\boldsymbol{\phi}, \boldsymbol{\theta}$, factorizes into two terms depending only on $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ respectively: $L(\boldsymbol{\theta}|D) \times L(\boldsymbol{\phi}|D)$, where D denotes the sample of occurrence times and excess magnitudes for the two series h and b.

The posterior distribution of the parameters is proportional to the product of the priori and the likelihood:

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}) \propto \pi^0(\boldsymbol{\theta}) \times \pi^0(\boldsymbol{\phi}) \times L(\boldsymbol{\theta}|D) \times L(\boldsymbol{\phi}|D). \tag{11}$$

The likelihood of the parameters is concentrated in a finite range (mainly because of the assumption of a GPD model with an upper bound). An improper uniform prior has been assumed for both sets of parameters, according to this finite-range concentration feature (Box and Tiao, 1973, p.21). The prior density for the occurrence parameters, $\pi^0(\boldsymbol{\phi})$ has been assumed to be (improper) uniformly distributed on $\lambda_h > 0, \lambda_b > 0$, $\lambda_h + \alpha_\lambda > 0$, $\lambda_b + \alpha_\lambda > 0$. The prior density for the magnitude parameters, $\pi^0(\boldsymbol{\theta})$, has been assumed to be uniformly distributed on $\mu_0, \delta_\mu, \nu_0, \delta_\nu, \alpha_\nu$.

The whole shape of the posterior density gives an assessment of the uncertainty in the estimation procedure. The posterior $\pi(\boldsymbol{\phi}, \boldsymbol{\theta})$ may also be used to derive a point estimate, e.g. the most likely value (posterior mode) or the expected value.

The posteriori in Eq. (11) has a quite complex form when considering all parameters simultaneously. A first simplification comes from the factorisation of the posterior density into terms containing the occurrence parameters $\boldsymbol{\phi}$ and the excesses parameters $\boldsymbol{\theta}$. Therefore, the estimation of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ can be carried out separately. For both terms in the posterior density, fixing all but one of the involved parameters, the conditional log-posterior density becomes a tractable univariate log-density, which can be satisfactorily sampled using a Gibbs sampler (Robert and Casella, 2000).

NHESSD
2, 799–824, 2014

4 Results and discussion

The wind dataset described in Sect. 2 has been modelled using the POT-GPD framework defined in Sects. 3.2 and (3.3). A sample of the posteriori density in Eq. (11) has been obtained using a Gibbs sampling algorithm, with three chains, 10 000 draws and a thinning ratio of 1 : 10. A burn-in of 50 % has been applied. The convergence of the joint chain has been assessed using the Gelman criterion (Gelman et al., 1995).

The log-wind speed magnitude (excesses over $\log 15\,\mathrm{m\,s^{-1}}$) has been modelled using a GPD with the proposed (ν, μ) parametrisation. The presence of time trends and differences between REMO and deep buoy data can be assessed using Figs. 5 and 4.

Figure 4 shows the joint posterior pdf of δ_ν and α_ν (lower-left panel). This joint pdf is characterized by its large dispersion thus pointing out the need of larger records to reliably estimate possible time trends in the shape of the GPD and the differences between the two data series. The marginal and conditional mode of α_ν (lower-right panel) differ substantially from 0. But the value $\alpha_\nu = 0$ is placed on the central part of the posterior marginal, thus making the possible positive trend in ν non-significative. We can conclude that there is some evidence of positive trend in the shape parameter but it should considered doubtful as there is no strong evidence against no-trend during the observed time interval. Similarly, the change in ν from hindcast data to buoy data has the mode placed at a positive value near to 1. (upper-left panel), pointing out differences between the GPD shape for the two series. However, the value $\delta_\nu = 0$ is fairly centered in the posterior marginal distribution of δ_ν, thus meaning that the change in ν is not significative and should be considered carefully.

Figure 5 (lower-left panel) shows the posterior joint pdf of total drift in the parameter μ, α_μ and the difference in μ corresponding to the two series. The value $\delta_\mu = 0$, although not centered with respect to the marginal pdf (upper-left panel), is in a 90 % credible interval. The hypothesis of no difference between REMO and buoy data is plausible. The mode of the posterior marginal of δ_μ is negative. Therefore, there is only a weak evidence against that the upper limit of the wind speed is the same for hindcast and

NHESSD
2, 799–824, 2014

buoy observations. With regard to μ_0, the posterior median estimate for μ_0 is about 0.53 (lower-right panel) which corresponds to about $y_{sup} = 82\,\mathrm{m\,s^{-1}}$. The μ_0 density conditional to the mode of (δ_μ, μ_0) (red line, lower right panel) approximately corresponds to $\mu_0 = 0.82$, i.e. an upper limit of wind speed of $145\,\mathrm{m\,s^{-1}}$.

Figure 6 shows the estimated posterior density for the intensity difference $\lambda_b - \lambda_h$ and the linear trend parameter α_λ which is the total drift of λ along the observation time. The marginal histograms for these parameters are shown in the secondary panels. The equality of initial Poisson intensities for REMO and buoy series is assessed graphically by means of the line $\lambda_b - \lambda_h = 0$. The line lies in the lower tail of the posterior pdf, leading to a small Bayesian p value when testing for the equality of both values, i.e. the difference of the reference Poisson intensities is significant. The deep buoy series predicts about 2 events per year more than those hindcasted by REMO model. Regarding to the trend α_λ, its histogram (lower-left panel) is approximately centered at 0. This provides a Bayesian p value near 0.5 when testing no linear trend in the Poisson intensity. Therefore, there is a non-significant trend in the intensity of the Poisson process.

It is still a matter of debate to what extent the frequency and intensity of windstorms may change as a consequence of the hypothetical climate change in the future. The results obtained for $\lambda(t)$ are non contradictory with other author's works, mainly devoted to investigate the changes in extreme winds, with methods based on global or regional climate models (e.g. Bolaños et al., 2004; Rockel and Woth, 2007). The slight and non-significant positive time trend observed for $\lambda(t)$, corresponding to an increase of events, is in agreement to the hypothesis of climate change considered in IPCC reports (IPCC, 2007). However, this result cannot be considered confirmatory. We consider that the hindcast model may have a stronger inertia than the buoy measurements; REMO winds are daily averaged wind fields, which have less variability and more inertia than true winds. Under steady, non extremal stormy conditions, hindcast winds would have more energy than true winds, leading to an overestimation of winds. However, after the analysis of this data set, no significant change of neither upper limit and shape of wind speed excesses distribution has been detected.

5 Conclusions

A dataset of hincast wind speed (REMO) has been analysed, together with a wind speed dataset registered in a deep buoy in the Tarragona coast. A non-stationary Poisson/GPD model accounting for linear time trends and differences between the hindcast and buoy series has been assessed. The wind speed was log-transformed to deal with its ratio scale. The parameterisation of GPD of excesses over 15 $\mathrm{m\,s^{-1}}$ has been adopted to restrict distributions to be have finite tail i.e. within the Weibull domain of attraction of maxima. The model was fitted using a Bayesian procedure.

The results confirm that the joint analysis of hindcasted and directly observed wind speeds is useful to enlarge existing records used in extremal analysis. No significative time trends have been detected in occurrence rates of events and shape parameter of GPD. Importantly, there were no evidences in favour of the existence of differences in shape and upper limit of the GPD for excesses between the two sources of information, thus supporting the idea of using hindcast data for extremal analysis. Nonetheless, there were significant differences in the rate of occurrence of wind events recorded by hindcast and directly observable events, being the latter substantially higher, about 2 events per year.

Although the total time of observation has been substantially increased by incorporating hindcast data, the uncertainty of the estimates is too large to attain conclusive results. This is the case of the time trend on the shape of GPD, represented by the parameter α_v, with a marginal distribution suggesting a positive trend, but without a clear statistical significance.

Acknowledgements. This research has been funded by the Spanish Ministry of Economy and Competitiveness under the projects “METRICS” (Ref. MTM2012-33236), “CODA-RSS” (Ref. MTM2009-13272). The last author acknowledges also funding within the “Juan de la Cierva” program (ref. JCI-2008-1835) and the COVARIANCE research project (ref: CTM2010-19709) of the Spanish Ministry of Education and Science, as well as the 7th FP EU Project FIELD-AC.

References

Beguería, S., Angulo-Martínez, M., Vicente-Serrano, S. M., López-Moreno, J. I., and El-Kenawy, A.: Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006, Int. J. Climatol., 31, 2102–2114, doi:10.1002/joc.2218, 2011. 801

Bolaños, R., Sánchez-Arcilla, A., Gómez, J., Cateura, J., and Sairouni, A.: Limits of operational wave prediction in the Northwestern Mediterranean., in: Proceedings of the 29th International Conference on Coastal Engineering, Lisbon, Portugal, 2004. 801, 813

Box, G. E. P. and Tiao, G. C.: Bayesian inference in statistical analysis, Addison Wesley, Reading, Mas, USA, 588 pp., 1973. 811

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, GB, 208 pp., 2001. 801, 804

Coles, S., Pericchi, L. R., and Sisson, S.: A fully probabilistic approach to extreme rainfall modelling, J. Hydrol., 273, 35–50, 2003. 803

Coles, S. G. and Tawn, J. A.: A Bayesian analysis of extreme rainfall data, Appl. Statist., 45, 463–478, 1996. 801

Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds (with discussion), J. Roy. Stat. Soc. B, 52, 393–442, 1990. 803, 804, 807

Della-Marta, P. M., Mathis, H., Frei, C., Liniger, M. A., Kleinn, J., and Appenzeller, C.: The return period of wind storms over Europe, Int. J. Climatol., 29, 437–459, doi:10.1002/joc.1794, 2009. 801

Dupuis, D.: Exceedances over high thresholds: a guide to threshold selection, Extremes, 1, 251–261, 1998. 807

Egozcue, J. J. and Ramis, C.: Bayesian hazard analysis of heavy precipitation in Eastern Spain, Int. J. Climatol., 21, 1263–1279, 2001. 803

Egozcue, J. J., Pawlowsky-Glahn, V., and Ortego, M. I.: Wave-height hazard analysis in Eastern Coast of Spain - Bayesian approach using generalized Pareto distribution, Adv. Geosci., 2, 25–30, doi:10.5194/adgeo-2-25-2005, 2005. 808

Egozcue, J. J., Pawlowsky-Glahn, V., Ortego, M. I., and Tolosana-Delgado, R.: The effect of scale in daily precipitation hazard assessment, Nat. Hazards Earth Syst. Sci., 6, 459–470, doi:10.5194/nhess-6-459-2006, 2006. 803, 808

Embrechts, P., Klüppelberg, C., and Mikosch, T.: Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin, Germany, 663 p, 1997. 803, 804, 808

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, Wiley, New York, NY, USA, 1995. 801, 810, 812

Grandell, J.: Mixed Poisson Processes, Chapman & Hall, London, GB, 268 p, 1997. 804

Guedes Soares, C., Carretero Albiach, J. C., Weisse, R., and Alvarez-Fanjul, E.: A 40 yr hindcast of wind, sea level and waves in European waters, in: Proceedings of the 21st International Conference on Offshore Mechanics and Arctic Engineering, 669–675, Oslo, Norway, 2002. 802

Hundecha, Y., St-Hilaire, A., Ouarda, T. B. M. J., El Adlouni, S., and Gachon, P.: A non-stationary extreme value analysis for the assessment of changes in extreme annual wind speed over the gulf of St. Lawrence, Canada, J. Appl. Meteor. Climatol., 47, 2745–2759, doi:10.1175/2008JAMC1665.1, 2008. 801

IPCC: IPCC Fourth Assessment Report: Working Group II Report "Impacts, Adaptation and Vulnerability", Cambridge University Press, 2007. 813

Leadbetter, M. R.: On a basis for peaks over threshold modeling, Stat. Probabil. Lett., 12, 357–362, doi:10.1016/0167-7152(91)90107-3, 1991. 803

Ortego, M. I., Gibergans-Báguena, J., Tolosana-Delgado, R., Egozcue, J. J., and Llasat, M. C.: Bayesian trend analysis for daily rainfall series of Barcelona, Adv. Geosci., 26, 71–76, doi:10.5194/adgeo-26-71-2010, 2010. 808

Ortego, M. I., Tolosana-Delgado, R., Gibergans-Báguena, J., Egozcue, J. J., and Sánchez-Arcilla, A.: Assessing wavestorm hazard evolution in the NW Mediterranean with hindcast and buoy data, Climatic Change, 113, 713–731, 2012. 806, 807, 808, 809

Palutikof, J. P., Brabson, B. B., Lister, D. H., and Adcock, S. T.: A review of methods to calculate extreme wind speeds, Meteorol. Appl., 6, 119–132, doi:10.1017/S1350482799001103, 1999. 801

Pawlowsky-Glahn, V., Tolosana-Delgado, R., and Egozcue, J. J.: Scale effect in hazard assessment - application to daily rainfall, Adv. Geosci., 2, 117–121, doi:10.5194/adgeo-2-117-2005, 2005. 808

Pickands III, J.: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, 1975. 807
Robert, C. P. and Casella, G.: Monte Carlo Statistical Methods, Springer, New York, NY, USA, 507 p, 2000. 811
Rockel, B. and Woth, K.: Extremes of near-surface wind speed over Europe an their future changes as estimated from an ensemble of RCM simulations, Climatic Change, 81, 267–280, 2007. 813
Sánchez-Arcilla, A., Gómez-Aguar, J., Egozcue, J. J., Ortego, M. I., Galiatsatou, P., and Prinos, P.: Extremes from scarce data, the role of Bayesian and scaling techniques in reducing uncertainty, J. Hydraul. Res., 46, 224–234, 2008. 801, 808
Sotillo, M. G., Ratsimandresy, A. W., Carretero, J. C., Bentamy, A., Valero, F., and Gonzalez-Rouco, F.: A high-resolution 44 yr atmospheric hindcast for the Mediterranean Basin: contribution to the regional improvement of global reanalysis, Clim. Dynam., 25, 219–236, 2005. 802
Tramblay, Y., Neppel, L., and Carreau, J.: Brief communication "Climatic covariates for the frequency analysis of heavy rainfall in the Mediterranean region", Nat. Hazards Earth Syst. Sci., 11, 2463–2468, doi:10.5194/nhess-11-2463-2011, 2011. 801
Walshaw, D.: Getting the most from your extreme wind data: a step by step guide, J. Res. Natl. Inst. Stand., and Technol., 99, 399–411, 1994. 801

NHESSD
2, 799–824, 2014

Bayesian trends of extreme wind

M. I. Ortego et al.

Table 1. Probability of GPD belonging to the Weibull attraction domain, $\xi < 0$, for the buoy and REMO series, for raw (in $\mathrm{m\,s^{-1}}$) and log-transformed data.

class/scale	raw	log
hindcast	0.40	**0.83**
buoy	0.89	**0.94**

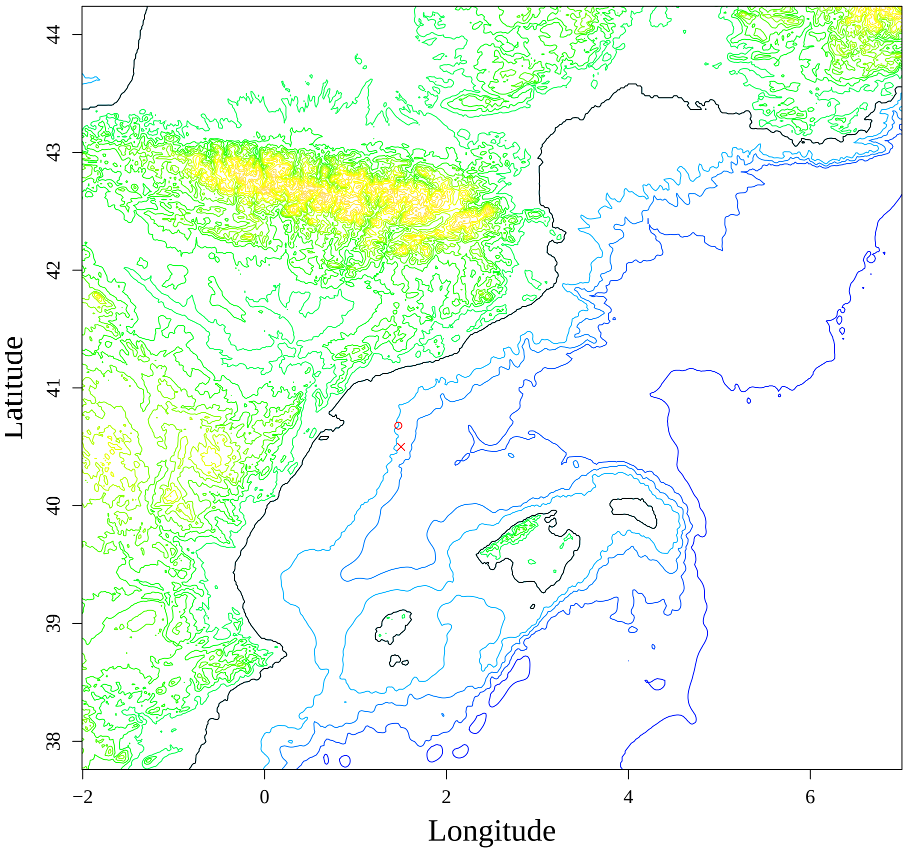

Fig. 1. Site of the data series: Buoy, circle; REMO grid-point, cross.

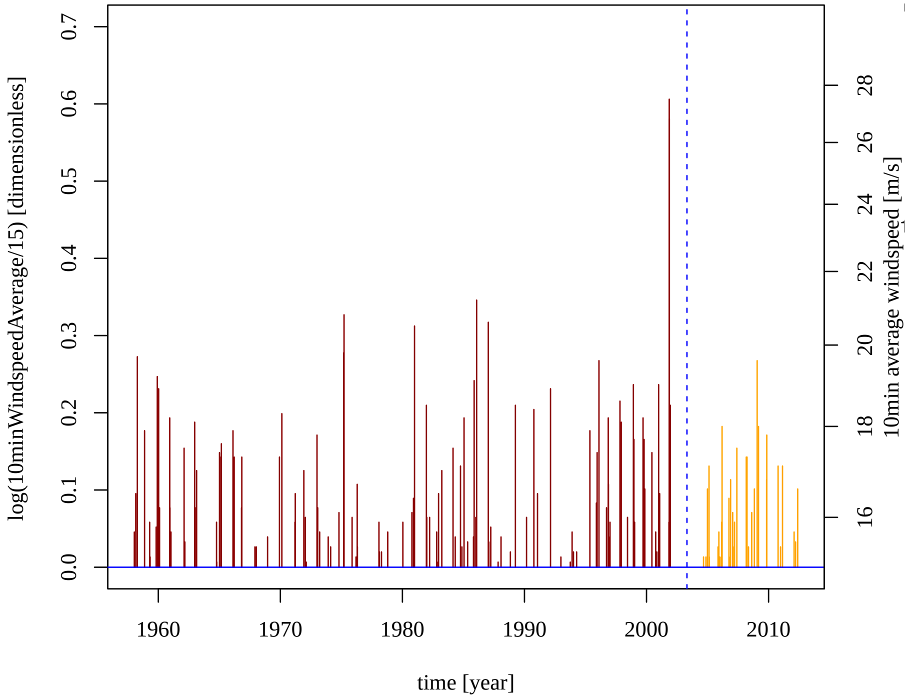

Fig. 2. Data series: REMO (brown) and buoy (orange) events.

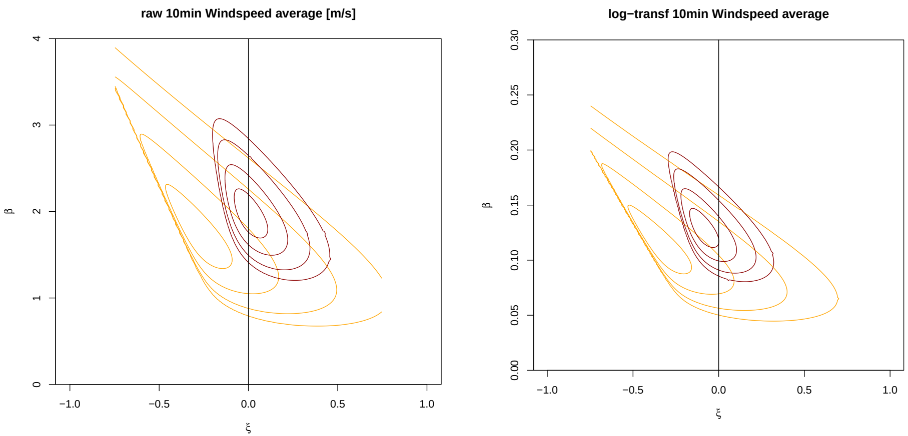

Fig. 3. Joint likelihood contours for the (ξ, β) parameters of the classical parametrisation of the GPD distribution for the raw data (left) and for the log-transformed data (right). Orange contours for buoy; brown contours for REMO.

NHESSD
2, 799–824, 2014

Bayesian trends of extreme wind

M. I. Ortego et al.

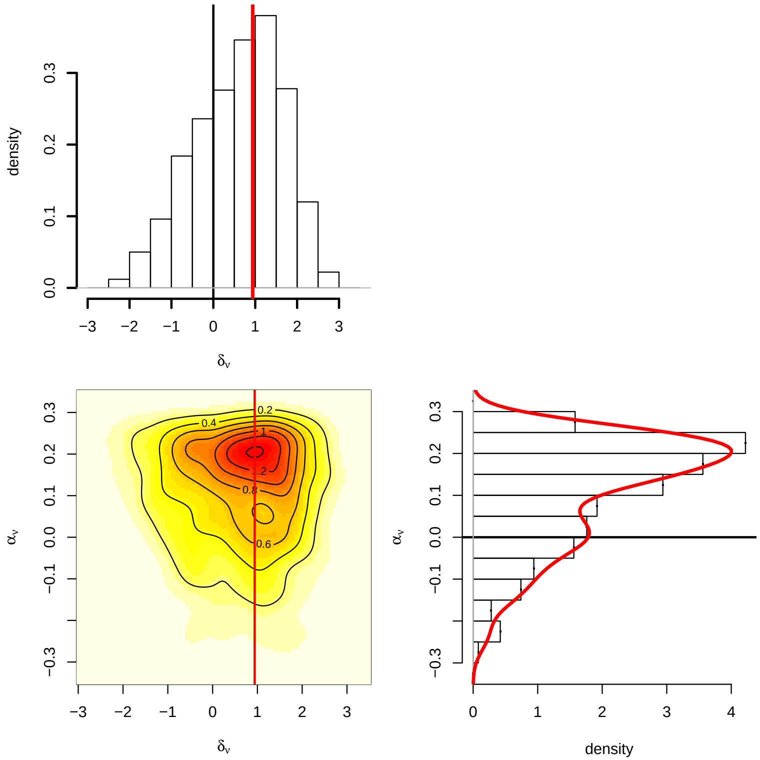

Fig. 4. Joint posterior density for (δ_v, α_v), the difference between the two data series and the total drift along the observed time interval for the shape parameter v (lower-left panel). The upper and lower-right panels show the marginal histograms for the two parameters and their pdf conditional to the posterior joint mode.

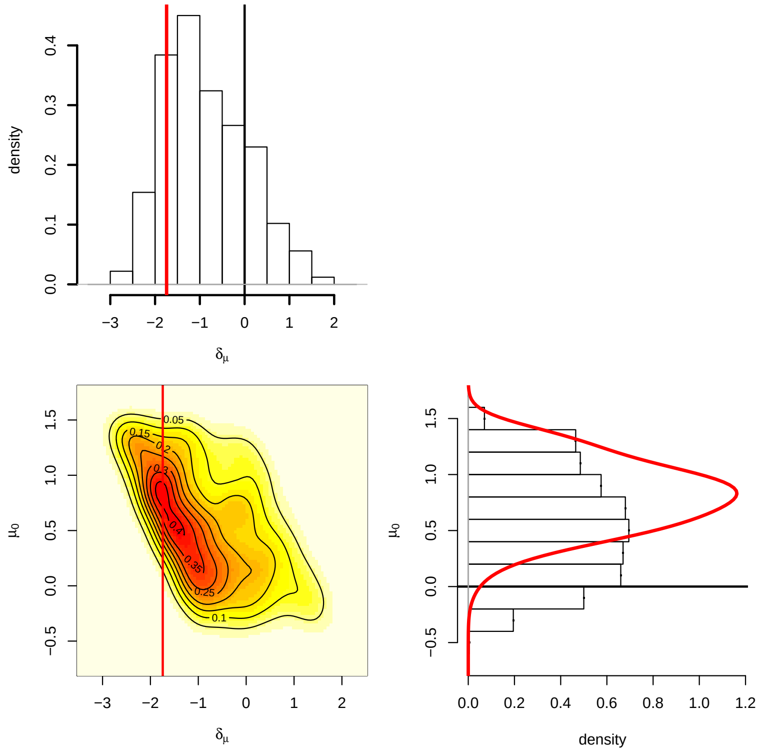

Fig. 5. Joint posterior density for (δ_μ, μ_0), the difference between the two data series and the parameter determining the upper limit of the GPD (lower left panel). The upper and lower-right panels show the marginal histograms for the two parameters and their pdf conditional to the posterior joint mode.

NHESSD
2, 799–824, 2014

Bayesian trends of extreme wind

M. I. Ortego et al.

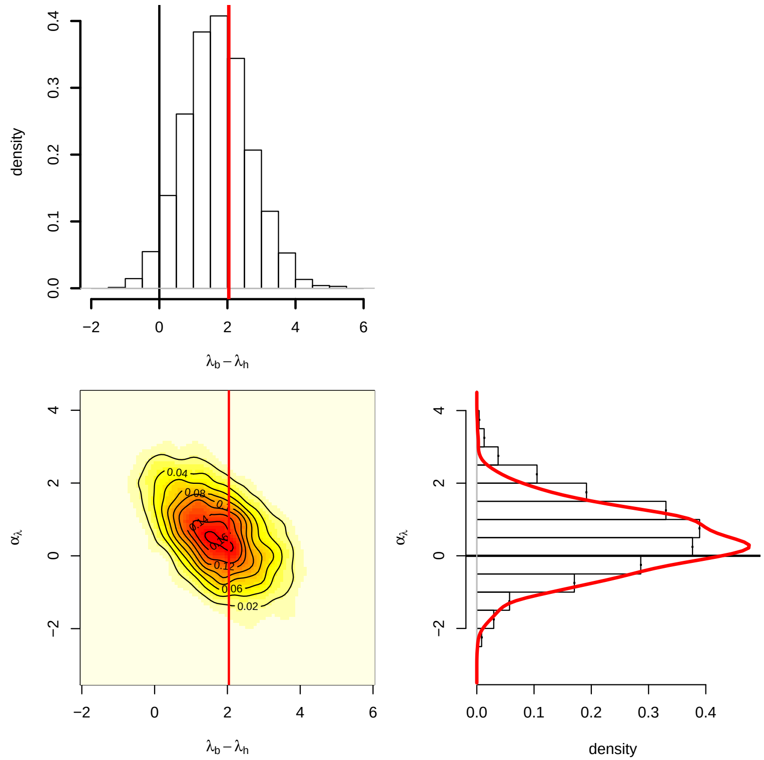

Fig. 6. Joint posterior density for $(\lambda_b - \lambda_h)$, the difference on Poisson intensity for both signals, and α_λ), the total drift of the Poisson intensity (lower-left panel). The upper and lower-right panels show the marginal histograms for the two parameters and their pdf conditional to the posterior joint mode.

NHESSD
2, 799–824, 2014

Bayesian trends of extreme wind

M. I. Ortego et al.