Checking model-data weather hazard occurrence fit in the context of climate change

Raimon TOLOSANA-DELGADO[1], M.Isabel ORTEGO[2], Juan José EGOZCUE[2], Agustín SÁNCHEZ-ARCILLA[1]

[1] Lab. Maritime Engineering, U. Politècnica de Catalunya, Spain
raimon.tolosana@upc.edu agustin.arcilla@upc.edu

[2] Dept. Applied Mathematics III, U. Politècnia de Catalunya, Spain

ma.isabel.ortego@upc.edu, juan.jose.egozcue@upc.edu

Abstract

In climate change impact studies it is common to run a given response model (from ecosystem changes to wavestorm or landslide occurrence) nested into one of the available long-term Global or Regional Circulation Models (GCM, RCM) reproducing the climate for the XX century or predicting it for the XXI. In this way, it is expected to capture the average behaviour of the studied system to a changing climate forcing: in other words, with such response forecasts, one does not actually expect to be able to reproduce each and every single event, but rather its statistical behaviour. Regarding weather-related hazard, the relevant statistical properties are the occurrence return period of events, and their expected magnitude. The present study focuses on wave storm occurrence, and aims at presenting a general methodology to check the adequate reproduction of the return period of hazardous weather-related events by such response forecast models. This is attained by analysing a compound data set formed by series of real data (typically of around 20-30 years in the last decades of the XX century or the beginning of the XXI one) and longer hind- or forecast series. Occurrence of a stormy event is considered to follow an inhomogeneous Poisson process, with: a linear trend to capture climate change, and a step in the junction real data-forecast data to capture systematic model biases. A Bayesian method is proposed to assess the influence of these two elements, i.e the presence/absence of a climate trend and the adequate reproduction of the statistical properties of wavestorm occurrence by forecasting models. Results suggest a non-significant trend albeit negative trend in the storm occurrence, and an inability of the used forecast model to reproduce wavestorm occurrence.

1   The data set

We focus on the region around the Ebro delta, with the aim of detecting any climate change influence in the wave storm events. The Ebro river has a relatively low sediment supply, specially after its full hydraulic regulation with the several dams and reservoirs built during the XX century. This makes the delta particularly weak nowadays against breaking storms, specially the several lateral sand bars and beaches protecting the highly productive (in ecological and agricultural terms) wet areas.

To detect climate trends, it is typical to analyse long data series obtained with local instrumentation. Unfortunately, systematic monitoring of this area started in the early 80s, thus a maximum of approx. 30years of observations may be available. This is not enough to adequately assess climate trends. For this reason, it was though to complement this "true data" with hindcasts of the second half of the XX century, obtained within the HIPOCAS project (Sotillo et al, 2005; Guedes Soares et al, 2002). HIPOCAS nested the WAM wave generation model (WAMDI group, 1988) into a REMO weather model (Jacob and Podzun, 1997), transferring the energy of daily average wind fields into wave fields. We took the series of significant waveheight at HIPOCAS node 2056046 (longitude 40.75 N, latitude 1.00 W) for the period 1958/01/03-1990/06/15, and complemented it from 1990/06/16 to 2008/12/31 with significant waveheight measurements at the Tortosa buoy (XIOM network, longitude 40.72 N, latitude 0.98 W). Significant waveheight (Hs) is a measure of the total energy contained in the wave spectrum in a given region, and it is correlated with the average height of the upper third of all measured weights. Lionello et al (2008) already used a similar strategy to infer trends on waveheight field for the XXI century.

The resulting data series is treated with a standard Peak-Over-Threshold methodology (Embrechts et al, 1997), defining a (marked) Poisson process: an event is defined as the time while Hs is larger than a given threshold (200cm), if this time span is larger than 6 hours. The event intensity is taken as the maximum observed Hs value within that period. Two consecutive events are considered independent if the second starts more than 3 days later than the end of the first one. If two or more events are not considered independent, we just keep the largest one. These criteria define a series of "instantaneous" events with a given magnitude. Figure 1 Shows the obtained series of Hs, in m and as adimensional exceedances over the critical threshold of 200cm: $X = \log(Hs/200)$.
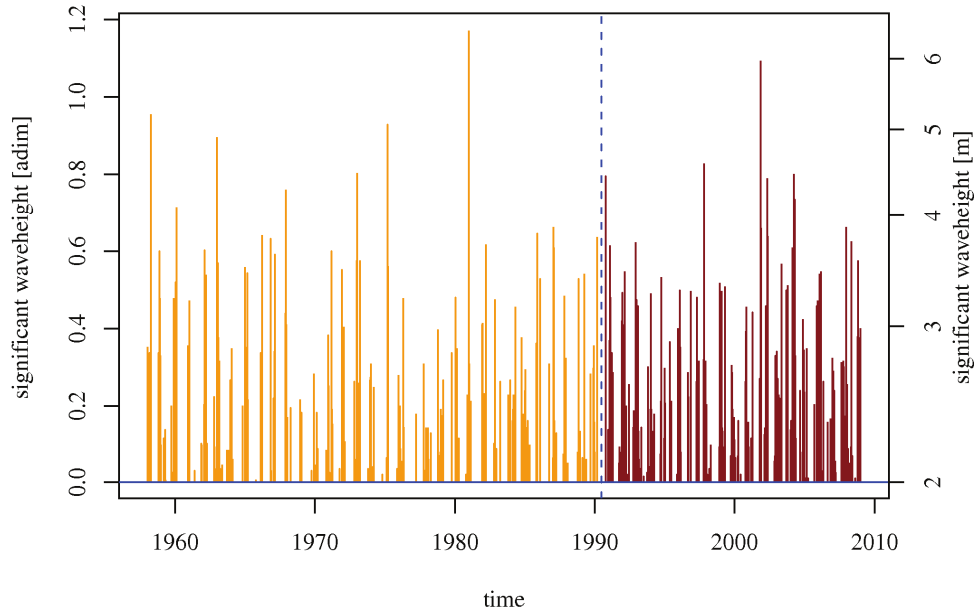
Figure 1: Data set of events and their magnitudes. The first part of the series correspond to HIPO-CAS hindcast events, the second part contains observed events at the Tortosa buoy

2   Peak-Over-Threshold occurrence model

Typically, occurrence and magnitude are modelled separately. This paper focuses on the modelling of occurrence, which is Poisson distributed. However, we assume that the intensity parameter of this Poisson process is not constant, but that it has a linear trend in time and a jump at the critical time t^c, midnight 1990/06/15 to 16 (the position of the junction between hindcast and observed data):

$$\lambda(t) = \lambda_0 + \delta_\lambda I(t > t^c) + \frac{\alpha_\lambda}{t_n} t , \qquad 0 < t < t_n .$$ (1)

With this, an event is expected to occur at a random time T, with the intensity given by:

$$\lambda(t) dt = Pr\langle t < T < t + dt \mid T > t \rangle = \frac{F_T(t + dt) - F_T(t)}{1 - F_T(t)}$$

which taking limits yields $\bullet(t) = F_T'(t)/(1 - F_T(t))$, a differential equation that may be integrated by separate variables. The resulting cumulative distribution function has as associated probability density function $f_T(t) = \lambda(t) \exp\left(-\int \lambda(\tau) d\tau\right)$. Thus, assuming that we have a series of events at occurrence times $t_1, t_2, \ldots, t_n$, we can obtain the likelihood of the model (Eq. 1) parameters $\lambda_0, \delta_\lambda, \alpha_\lambda$ as the product of the individual occurrence probabilities, or in logarithms:

$$\log L\left(\lambda_{0,}\delta_\lambda,\alpha_\lambda\right)=\sum \log\left(\lambda(t)\right)-\sum \int \lambda(\tau)d\tau =$$

$$=\sum \log\left(\lambda_0+\delta_\lambda I\left(t_i>t^c\right)+\frac{\alpha_\lambda}{t_n}t_i\right)-\left[\lambda_0 t_n-\delta_\lambda\left(t_n-t^c\right)+\alpha_\lambda\frac{t_n}{2}\right] .$$

(2)

This log-likelihood expression can be maximized to obtain the maximum likelihood estimations of the parameters $\lambda_0, \delta_\lambda, \alpha_\lambda$, or we can try a Bayesian approach.

3   Bayesian estimation

To estimate a given parameter vector ❑, Bayesian estimation methods combine the information coming from its likelihood (Eq. 2) given the data L(❑|data) with a prior assessment of the probability distribution ❑^{0}(❑) of ❑. This prior distribution is the way the analyst has to introduce all his/her knowledge about the most likely values of the parameters, prior to having a look at the data. This is obtained by perturbing the prior by the likelihood, or

$$\pi(\theta)\propto\pi^0(\theta)\dot{L}(\theta|data) .$$

In our case, we will simply assume some reasonable intervals for the parameters, and a uniform prior distribution ❑^{0}(●$_0$, ♒$_\bullet$, ♋$_\bullet$) within the hypercube (●$_0^{\mathrm{inf}}$, ●$_0^{\mathrm{sup}}$)◈(♒$_\bullet^{\mathrm{inf}}$, ♒$_\bullet^{\mathrm{sup}}$)◈(♋$_\bullet^{\mathrm{inf}}$, ♋$_\bullet^{\mathrm{sup}}$).

To characterize it, we will obtain a sample of the posterior ❑(❑) with a Gibbs sampling scheme (e.g. Robert and Casella, 2000). This is obtained following these steps

    for $k=1$, fix some feasible values for ●$_0$=●$_0^0$ and ♒$_\bullet$❒♒$_\bullet^\square$

2. compute ❑(♋$_\bullet$| ●$_0^{k-1}$, ♒$_\bullet^{k\square\square}$) for a dense set of points within (♋$_\bullet^{\mathrm{inf}}$, ♋$_\bullet^{\mathrm{sup}}$); obtain the associated cumulative distribution F(♋$_\bullet$| ●$_0^{k-1}$, ♒$_\bullet^{k\square\square}$) and use it to simulate a value ♋$_\bullet^k$

    compute ❑(♒$_\bullet$| ●$_0^{k-1}$,♋$_\bullet^k$) for a dense set of points within (♒$_\bullet^{\mathrm{inf}}$, ♒$_\bullet^{\mathrm{sup}}$); obtain the associated distribution F(♒$_\bullet$| ●$_0^{k-1}$,♋$_\bullet^k$) and use it to simulate a value of ♒$_\bullet$

4. compute ❑(●$_0$| ♒$_\bullet^k$,♋$_\bullet^k$) for a dense set of points within (●$_\square^{\mathrm{inf}}$, ●$_\square^{\mathrm{sup}}$); obtain the associated cumulative distribution F(●$_\square$ ♒$_\bullet^k$♋$_\bullet^k$) and use it to simulate a value of ●$_\square^k$

5. return to step 2, with $k=k+1$

We run the chain a certain burn-in period (100 samples in our case), to be sure that the algorithm has converged to sample the true posterior distribution. After that moment, we store the vectors of simulations ❑^{k}=(●$_0$, ♒$_\bullet$, ♋$_\bullet$)^k. We can then later study the statistics of this sample to characterize the posterior distribution of the parameters.
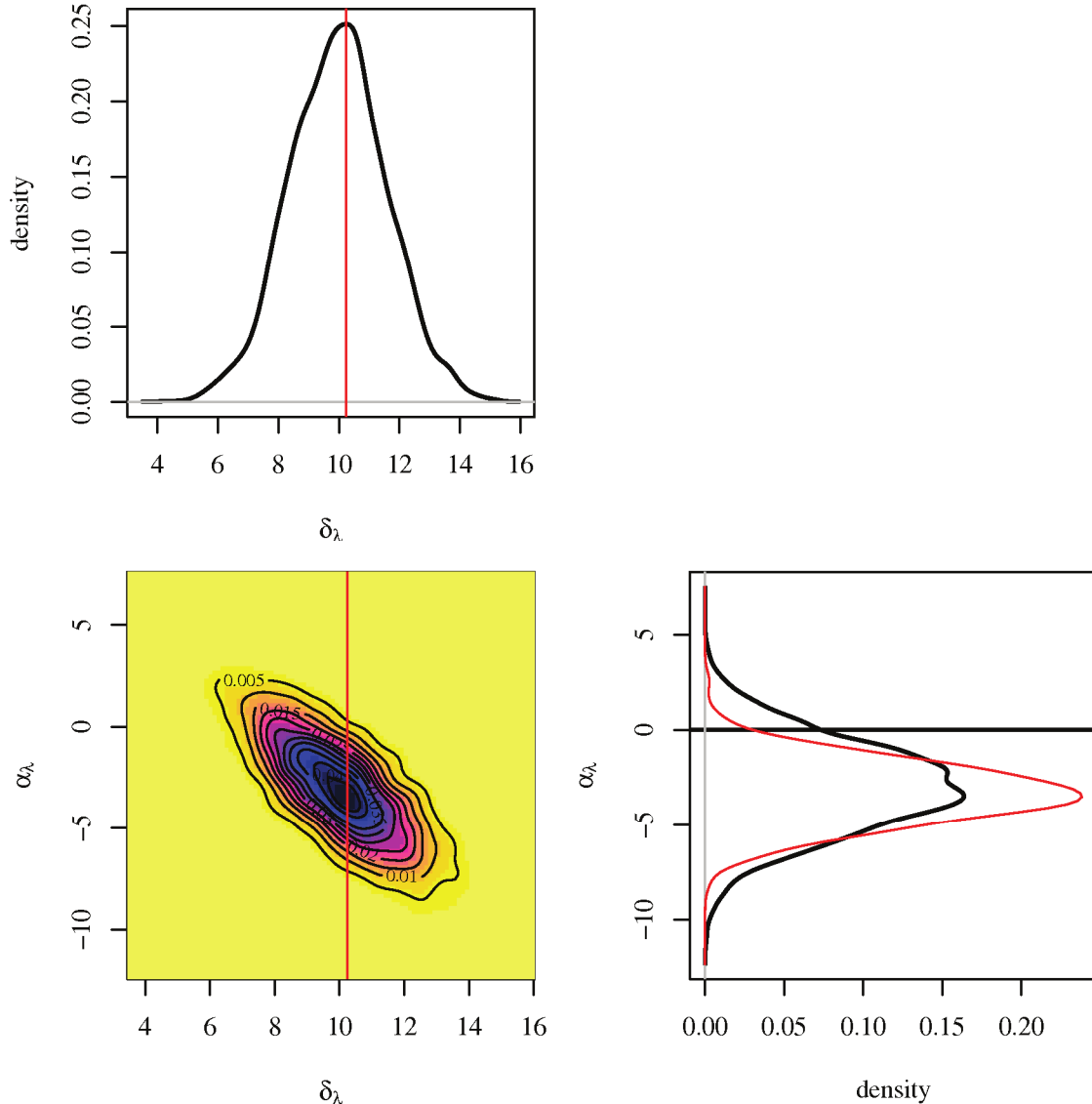
Figure 2: Posterior density estimates of the slope parameter (♎●) and jump parameter (♋●). The red vertical line marks the position of the posterior mode: the red curve in the lower right diagram shows the density profile for ♋● along this mode.

4 Results and discussion

Using the algorithm outlined before, we obtained a sample for the three parameters describing the change on event intensity along time. Figure 2 shows kernel estimates of the posterior density for the pair (♎●, ♋●) and for each marginal. Parameter ♋● represents the linear trend with time (a decreasing but not significant one), while ♎● is the jump at the transition HIPOCAS-buoy (a clearly significant increment) thus indicating the systematic differences between the model and the actual measurements. To illustrate these results, Figure 3 shows the expected number of events for a representative subsample of size ten from the sample of posterior parameters. These are compared with the number of events actually observed each year. From these diagrams it is evident that HIPOCAS does not adequately reproduce the occurrence statistics of wavestorms as shown by actual buoy data. The reasons for that undesirable behaviour might be:

- either the REMO model underpredicts winds systematically, or the WAM does not efficiently transfer wind energy to wave energy; this would produce a systematic reduction of the number and magnitude of events in the HIPOCAS set with respect to the buoy set;

- or else that predicted wave fields have "too much inertia", growing too slowly, but also decreasing too slowly, i.e. tending to smooth the series to the average; this would yield a reduction of the number of events for the modelled part, but could produce an increment of the predicted magnitude of events.

To distinguish one from the other, it would be good to model event magnitude as well. This is a matter of further research.
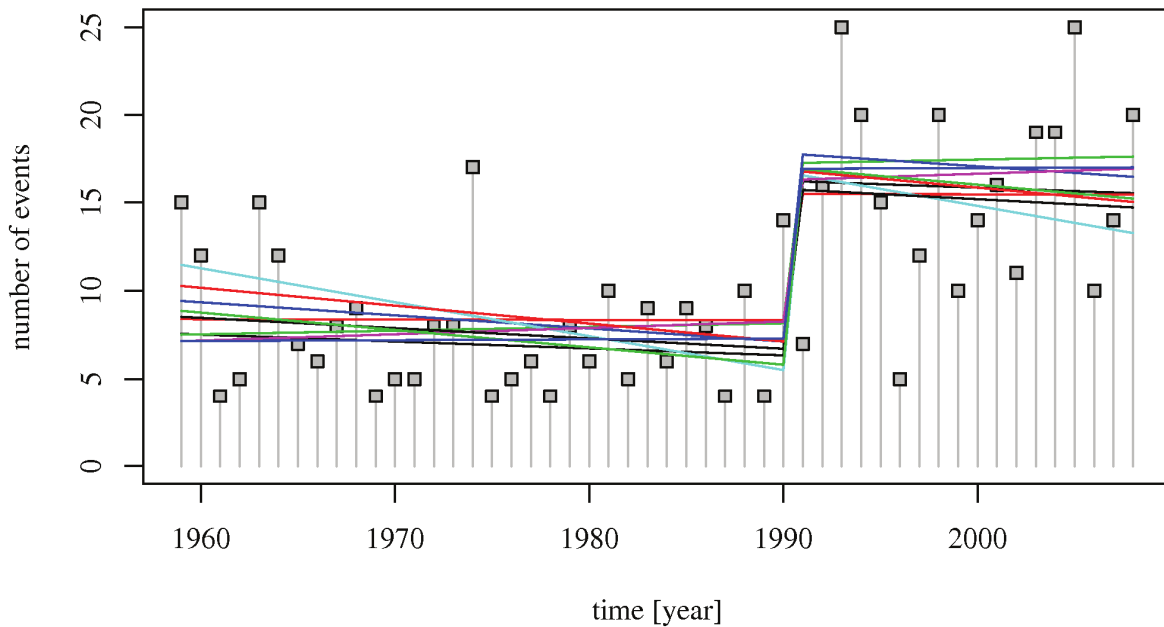

Figure 3: Number of events per year, comparing observed values with expected values for a selected set of posterior estimates

5 Conclusions

When modelling the effects of climate change with a numerical model nested on a GCM/RCM, it is important to be sure that the chosen model adequately reproduces the "climate" of the present or past situations (i.e. the statistical properties of the phenomenon under study). It is actually more important to ensure this statistical reproduction than the ability to predict one by one each possible individual event. In the case of hazardous phenomena (droughts, fires, landslides, surges, wavestorms) it is reasonable to use the Peak-Over-Threshold methodology to model their extremes, those actually producing the damages. In a POT framework, we model events as marked Poisson processes, where occurrence is captured by a conventional Poisson process and event magnitude corresponds to the marking. This paper presented a methodology that allows to check the adequate statistical reproduction of occurrence of hazardous events by models with respect to data. This is obtained analysing with conventional Bayesian methods an inhomogeneous Poisson process, where event occurrence intensity is considered a linear function of time with a different intercept for the observed and the modelled series: significant differences of

~~these intercepts will indicate a systematic under/overestimation of the event occurrence by mod-~~ els with respect to data.

This framework was applied to the study of wave storm occurrences as predicted by the HIPO-CAS project of wave field hindcasting for the XX century, in the region of the Ebro delta. Results indicate a strong underestimation of event occurrence with respect to buoy data, tentatively related to a high inertia of models, that reduce their ability to follow sharp space-time variations of wind fields observed in reality in the studied region.

Acknowledgements

References

EMBRECHTS P, KLÜPPELBERG C, MIKOSCH T (1997): Modelling extremal events for insurance and finance. Springer-Verlag, Berlin, Germany, 663 p

JACOB D, PODZUN R (1997): Sensitivity studies with the regional climate model REMO. Meteorol Atmos Phys Vol. 63, 119-129

GUEDES-SOARES C, CARRETERO-ALBIACH J, WEISSE R, , ALVAREZ-FANJUL E (2002): A 40 years hindcast of wind, sea level and waves in european waters. Oslo, Norway, pp. 669-675

LIONELLO P, COGO S, GALATI M, SANNA A (2008): The Mediterranean surface wave climate inferred from future scenario simulations. Global and Planetary Change Vol. 63, 152-162

ROBERT C, CASELLA G (2000) Monte Carlo Statistical Methods. Springer, New York, NY, USA, 507 p

SOTILLO M, RATSIMANDRESY A, CARRETERO J, BENTAMY A, VALERO F, GON-ZALEZ-ROUCO F (2005) A high-resolution 44-year atmospheric hindcast for the mediterra-nean basin: contribution to the regional improvement of global reanalysis. Climate Dynamics Vol. 25: 219-236

WAMDI-Group (1988) The wam model - a third generation ocean wave prediction model. J Phys Oceanogr. Vol. 18: 1775-1810