Nat. Hazards Earth Syst. Sci. Discuss., 2, C3605–C3618, 2015
www.nat-hazards-earth-syst-sci-discuss.net/2/C3605/2015/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "How historical information can improve extreme coastal water levels probability prediction: application to the Xynthia event at La Rochelle (France)" *by* T. Bulteau et al.

T. Bulteau et al.

t.bulteau@brgm.fr

Received and published: 31 March 2015

We thank the referee for his review that led to improving our paper.

In particular, taking into account Specific comments 1, 2 and 3, as well as Technical corrections 2 to 10, we revised Section 2 entirely. The new version of Section 2 is provided in the attached pdf to alleviate the text in the point-by-point reply below.

Specific comments:

1. P7067 L13–16: It is not clear why the annual maximum is referred to when discussing return levels here. There is perhaps a confusion between return levels of peak events and annual return levels (i.e. return levels of annual events)? The annual return level x_p for return period T is the level exceeded in one of every T years on average, i.e. the level exceeded by the annual maxima maxy(WL) once on average every T years. Since in T years there are T annual maximum events, the number exceeding xp follows Binomial(T,Pr(maxy(WL) > x_p)) hence x_p satisfies the equation:

Pr(maxy(WL) > x_p) = 1/T

However, the model fits peak events $X > u$ so it is more natural to define the return level x_p for return period T as the level exceeded once on average by a peak event in T years. With n_{py} peak events per year, the number of exceedances of x_p in T years follows Binomial(Tn_{py},Pr($X > x_p$)) hence here x_p satisfies:

Pr($X > x_p$) = 1/(Tn_{py})

The paper therefore appears to be calculating return levels of peak events but presenting them as an approximation to annual return levels. If this is intended, it should be noted that the approximation holds only when the probability of exceedance is small, i.e. for high return periods. Though if annual return levels are intended, it is not clear why the approximation is required at all. Alternatively, if the use of annual return levels is not required, Eq. (3) should be simplified by removing all reference to the annual maximum.

Authors' response: We agree with the referee that the phrasing might be confusing. What we meant originally was to define the return level x_p for return period T as the level exceeded once on average by a peak event in T years. Eq.(3) was then written to state that return levels of peak events can be approximated by annual return levels. This approximation holds only when the probability of exceedance is small but this was verified in our case study: indeed, the approximation is used only for convenience in the interpretation of results for Xynthia's WL as T is read directly on a return level plot (Figure 3). We compare annual probabilities of exceedance written as 1:T years. Strictly speaking, since we deal with peak events, p=1:T years is neither an annual probability of exceedance nor an event probability of exceedance. In the last case, we should rather compute p=1: (Tn_{py}) but this value is not directly interpretable using Figure 3. That is why we prefer to use the annual probability of exceedance $p = 1-(1-1/(Tn_{py}))^{n_{py}} \approx 1:T$ years (when Tn_{py} is large enough). The manuscript will be modified to clarify this point. In particular, an Appendix will be added (see Appendix A attached), Section 2 will be revised (see Revised Section 2 attached) and Section 3.2 will be slightly modified (see below):

Modification of Section 3.2 : P7074 L25 "(Fig. 3). *Because the calculated return periods of Xynthia's WL are large (typically greater than 100 years) and it makes more sense to speak about predictive exceedance probabilities rather than predictive return periods (see Sect. 2.2), we will compare results in terms of annual probabilities of exceedance (see Sect. 2.2 and Appendix A). Then we shall recall that the prediction for a Xynthia-like WL can be interpreted as the probability that next year's maximum WL (e.g. in 2010 if we are in 2009) will exceed Xynthia's WL.* In case (...)"

2. It would appear that every observation above the threshold u is fitted to the GPD and then the extremal index is used to correct for the temporal dependence when estimating return levels. An alternative approach, which is perhaps more common, is to first identify independent peaks above the threshold and then fit only the cluster maxima to GPD. There is then no need for the correction, so long as λ represents the mean number of independent peaks above u per year. This approach would seem to be more consistent with the historical events used in the case study as they appear to represent the peaks of separate events rather than a complete list of all known time-steps when the perception threshold was exceeded.

Authors' response: We agree with the referee and we will modify the manuscript accordingly (see the revised Section 2 attached). As a result, this will also clarify the Binomial vs Poisson issue raised by the referee (see Technical correction 11). The

results with the more common approach of fitting a POT sample selected with a temporal independence criterion do not change much compared to the old version (using an extremal index to correct a posteriori for the temporal dependence between peaks).

3. P7069 L2–P7070 L8: It is much simpler to derive the historical likelihood of Eq. (12) by observing that the peaks exceeding the perception threshold X_0 in n_y years occur as a Poisson process. Since the number of exceedances of u in n_y years is assumed to be Poisson(λn_y), it follows that the number of events H exceeding X_0 in n_y years is also Poisson with mean given by $\lambda n_y Pr(X > X_0|X > u) = \lambda n_y$(1-$G_\theta(X_0)$). Eq. (12) then follows directly after writing the historical likelihood as: $f(Dhist|\theta) = Pr(H|\theta)f(historicaldata|X > X_0, \theta)$

Authors' response: We agree with the referee who suggests a more elegant way of deriving Eq.(12). The manuscript will be modified accordingly (see the revised Section 2 attached).

Technical corrections:

1. P7064 L22–24: Bayesian methods are not required to incorporate historical data as censored observations; any likelihood-based method would do (e.g. maximum likelihood). So the use of a Bayesian approach should instead be justified in terms of the better representation of uncertainty, for example.

Authors' response: We agree with the referee and we will modify the manuscript as follows:

"(...) (see e.g. Baart et al., 2011). *The added value of using historical information in EVA has been widely recognised for the last 30 years in the domain of hydrology (see e.g. Benito et al. (2004) for a review). Among the statistical techniques developed to combine both sources of data (recent observations and historical information), Bayesian methods provide the most flexible and adequate framework because of their natural ability for handling uncertainties in extreme value models (Reis and Stedinger,* *2005; Coles and Tawn, 2005).* Surprisingly, we found only one reference (. . .)"

2. P7065 L5–6 and elsewhere: The approach of integrating partial historical information into an extreme value analysis is referred to as Bayesian Markov Chain Monte Carlo (or BMC2) following Reis and Stedinger (2005). However, the essence of the method is to incorporate the historical data into the model likelihood as censored observations yet this is not reflected in the title. Moreover, in general the approach need not depend upon a Bayesian model nor on Markov Chain Monte Carlo (MCMC); other numerical integration methods could be used to fit the Bayesian model and obtain the same estimates, while the modified likelihood could equally be applied in a classical maximum likelihood analysis for example to obtain similar results. I therefore suggest that the title BMC2 is replaced by something more appropriate.

Authors' response: We agree with the referee. Therefore, the method will no longer be called BMC2 but HIBEVA (Historical Information in Bayesian Extreme Value Analysis). We still keep the word Bayesian in the title as this is central in our paper (e.g. predictive distribution could not have been obtained using classical maximum likelihood analysis). In addition, we will modify the manuscript as follows:

"In the hydrology field, Reis and Stedinger (2005) developed a Bayesian Markov Chain Monte Carlo (MCMC) approach to tackle the issue of integrating partial historical information within EVA. *The essence of the approach is to incorporate partial historical data into the model likelihood as censored observations. In the present study, we build on this approach to develop a Bayesian MCMC method adapted for EVA of coastal water levels (called HIBEVA – for Historical Information in Bayesian Extreme Value Analysis, hereafter).*"

3. P7065 L22–23: There are strong arguments in favour of extrapolating to extreme values via fitting Peaks-Over-Threshold to GPD so these should be referred to (see e.g. Coles, 2001).

Authors' response: OK, the manuscript will be modified as follows (see also the re-

vised Section 2 attached):

"The model chosen to represent and extrapolate extreme values of WL is the Generalised Pareto Distribution (GPD), applied to a Peaks-Over-Threshold (POT) sample. *This extreme value model has been widely used and is most commonly recommended as it makes use of all the high values for the period under study to adjust the parametric distribution (Coles, 2001; Hawkes et al., 2008).*"

4. P7066 L4–5: The support of the GPD is stated here as $x\leq$ u$-(\sigma/\xi)$ if $\xi < 0$ and $x \in \Re$ otherwise but this does not account for the additional constraint $x > u$.

Authors' response: OK, the manuscript will be modified accordingly (see also the revised Section 2 attached):

"*The support of the distribution is* $u < x\leq$u-(σ/ξ) if $\xi < 0$ and $x > u$ if $\xi\geq$0."

5. P7066 L5–6: The text states "σ represents the width of the distribution" but I suggest 'scale' is used rather than 'width' here to avoid confusion with the width of the support of the distribution, which is referred to in the preceding sentence.

Authors' response: OK, the manuscript will be modified as follows (see also the revised Section 2 attached):

"Whereas σ represents *the scale of the distribution (in units of x)*, (. . .)"

6. P7067 L1: I assume the non-informative prior distribution applied was the improper flat prior ($f(\theta) \propto 1$) since the uniform distribution cannot be used for variables with infinite support.

Authors' response: Yes, the referee is right. The manuscript will be modified as follows to clarify this point (see also the revised Section 2 attached):

"Consequently, we use a *non-informative flat prior ($f(\theta) \propto 1$)* (Payrastre et al., 2011)."

7. P7067 L5: MCMC algorithms are very flexible but some would argue that they are not a very efficient sampling method since they can take many iterations to converge etc.

Authors' response: OK, we will delete the word "efficiently" in the manuscript (see the revised Section 2 attached).

8. P7067 L10: It is not obvious how the mode can be retrieved from a set of samples of continuous variables since every sample value is likely to be unique. Nor is it clear why it is particularly useful to extract maximum likelihood estimates when a Bayesian approach is being used.

Authors' response: In the particular case of using a non-informative flat prior in the Bayesian model, the posterior distribution of θ is proportional to the likelihood of data. Thus, among the 40,000 vectors θ generated, the one giving the maximum value of the likelihood is also the one maximising the posterior distribution of θ, i.e. the mode of the distribution. The manuscript will be modified to clarify this point (see also the revised Section 2 attached):

"In particular, the mode of the set of vectors θ can be retrieved *as the vector maximizing the likelihood function (because of the proportionality between $f(\theta|D)$ and $f(D|\theta)$). The associated quantiles x_T correspond therefore to the maximum likelihood estimates for WL*."

Moreover, the interest of extracting maximum likelihood estimates relies on the possibility to calculate both standard estimative return levels (equal to what would have been obtained using a classical maximum likelihood estimator) and predictive return levels within the Bayesian framework. In addition, it is also of interest to visualize credibility intervals as it indicates the width of the posterior distribution of quantiles (and therefore the statistical uncertainty on the results). A passage will be added in the manuscript to stress this on (see also the revised Section 2 attached):

P7068 just before L12 "The formulation of the likelihood (...)": "*Within the Bayesian*

framework, it is therefore possible to calculate and compare both standard estimative return levels x_T (equal to what would have been obtained using a classical maximum likelihood estimator) and predictive return levels $\tilde{x}_T$. While the predictive return levels incorporate all the uncertainty information, standard estimative return levels can be associated with credibility intervals which provide an overview of the uncertainty related to the quantiles x_T when visualised on a return level plot."

9. P7068 L14: The non-historical data are first referred to as 'systematic' here in passing but this has not been defined. Nor is it clear why 'systematic' is an appropriate name for the non-historical data.

Authors' response: The word 'systematic' refers to the systematic gauging era. It will be defined in the manuscript P7064 L14 and P7068 L14 (see below). We used this term as it is commonly used in the hydrology field.

P7064 L14: "In the past, *before the era of systematic gauging*, extreme events also happened." P7068 L14: "(...), thus separating the systematic period *(with systematic tide gauge records) and the historical period*:"

10. P7068 L14: It is not clear what is gained by partitioning the data into 'systematic' and 'historical'. Mathematically, it would seem that the systematic data is treated the same as the historical data for the special case that the perception threshold is equal to u and there are no censored observations (i.e. $h_2 = h_3 = 0$). Since ultimately a collection of perception thresholds are applied, each with separate sets of known and censored observations, the first of these could be taken as the systematic data to simplify the presentation.

Authors' response: We agree with the referee that mathematically, the systematic data is treated the same way as the historical data for the special case that the perception threshold is equal to u and there are no censored observations (i.e. $h_2 = h_3 = 0$). However, we believe it is better for the presentation to separate the two datasets as it eases the understanding of what is new in the likelihood function compared to a classical case with only known values. Indeed, the historical information is often partial (i.e. $h_1 = 0$) whereas the systematic data are known. Moreover, whereas the systematic data are originally corrected from the mean sea-level rise, this correction needs to be done in the historical likelihood for the historical perception threshold (cf Eq. 11 in the revised Section 2).

11. P7070 L1–2: The number of exceedances of u is assumed to be Poisson when deriving the historical likelihood. However, earlier in the paper when defining the return level it is stated that there are n events per year which would imply that the number of exceedances should be Binomial. A reason should be given for this discrepancy.

Authors' response: In the paper, we first extract from the original time series of WL every peak separated by at least 9 hours (so as to select every high water). Then, every peak above the threshold u is fitted to the GPD and then the extremal index is used to correct for the temporal dependence when estimating return levels (cf Specific comment 2). The "Binomial aspect" comes from the fact that the tidal signal is deterministic and cyclic. For example, on the French Atlantic coast, which is a macrotidal environment, we know there are about 706 high tides per year. Since the duration of storms is typically larger than 12 hours, the odds are the maximum WL (tide + storm surge) during a storm will occur at high tide (or close to high tide). As a result, the maximum possible number of peaks above u is equal to the number of high tides per year, n (defined P7067). Then, strictly speaking, the number of exceedances of u in n_y years is Binomial(nn_y, p), with $p = Pr(X > u)$.

However, the Binomial distribution converges towards the Poisson distribution when $nn_y \rightarrow \infty$ and $nn_y p \rightarrow \nu$ constant. In our case study, it can be verified that Poisson(λn_y) $\sim$ Binomial(nn_y, p) and consequently that Poisson$(\lambda n_y(1 - G_\theta(X_0))) \sim$ Binomial$(nn_y, p(1-G_\theta(X_0)))$. Moreover, considering that the threshold u is adequately chosen and assuming we are in the domain of asymptotic validity of the GPD, the number of events above u occurring in a fixed interval of time follows implicitly a Poisson distribution (Coles, 2001). This is consistent when considering that extreme water lev-

els are a manifestation of storms whose occurrences are by nature random.

To answer the referee's comment, there is no real discrepancy as both distributions (Poisson and Binomial) may be equally applied in our case to describe the number of exceedances of a threshold. We prefer to use Poisson as it seems more intuitive and consistent with the occurrence of storms. With the new version of the manuscript (see Revised section 2 attached), following referee's specific comment 2, we first select a POT sample of WL using a physical threshold u_p (Bernardara et al., 2014) with an independence criterion: every peak must be separated from the others by at least 72h (typical storm duration on the French Atlantic coast). That way, the number n of peak events per year is no longer fixed, only its empirical mean can be calculated. As a result, it is more straightforward to see the number of events above the statistical threshold u_s (Bernardara et al., 2014) occurring in a fixed interval of time as a Poisson distribution of annual rate $\lambda = np$.

12. P7070 L2: The threshold exceedance rate λ is first introduced when defining the return level and applied again to derive the historical likelihood. However, the paper does not state how it is estimated. While the rate could be treated as uncertain under the Bayesian framework, the case study implies that it is instead fixed at the proportion observed in the synthetic data. This should be clarified.

Authors' response: We agree with the referee on the fact that the rate could be treated as uncertain under the Bayesian framework. However, we believe it would have raised the complexity of the case study. The aim of the paper is indeed to show the potential of the HIBEVA method rather than to present a holistic treatment of La Rochelle case study as it is highlighted in the discussion and conclusion. The manuscript will be modified as follows to indicate the possibility of treating λ as uncertain and to state how it is estimated:

P7073 L6 "*The first step of the double-threshold approach detailed in Sect. 2.1 is the physical declustering of systematic data. With a minimal duration of 72 h (typical* *storm duration on the French Atlantic coast) between two peaks to ensure their independence, the physical threshold u_p is chosen so that n, the mean number of peak events per year, is about 10. Then, the statistical threshold u_s is selected using the classical tools described in Sect. 2. This provides a threshold u_s= 6.68 m for the case with the smallest dataset, i.e. case 1. For this case, the mean number of peak WL that exceed that threshold per year is λ= 2.9. It is estimated as the number of peak WL exceeding u_s divided by the effective duration of the systematic period (about 26 years for case 1). For sake of intercomparison, the threshold u_s is kept constant for every case (1 to 4). It should be noted that the rate λ could be treated as uncertain under the Bayesian framework, thus making the problem tridimensional. In that case, the likelihood of systematic data (cf Eq. (7)) should be modified to account for the probability of observing s peak WL during the systematic period. However, to simplify the presentation, we chose to fix λ at the proportion observed in the systematic dataset.* Results are presented (...)"

13. P7074 L16–18: What is meant by “The standard estimation of return period is” here? Is this calculated from the predictive exceedance probability, perhaps after averaging out the parameter uncertainty? Alternatively, is the return period derived as a function of the parameters and then summarised by the mean or median estimate?

Authors' response: The standard estimation of return period is the return period associated with the quantile x_T, derived from the maximum likelihood estimates of the posterior distribution of the parameters (see e.g. Technical correction 8). Therefore, the standard estimation of return period is the return period that would have been obtained using a classical maximum likelihood estimator for the GPD parameters.

14. P7079 L4–5: It should be made clear here that the plotting positions are not used for the model fitting (as they are in some classical estimation methods) but are used only for plotting return level estimates in Fig. 3. Appendix A would appear to only show how exceedance probabilities are assigned to observed values Xi but there is no mention of how the censored observations are plotted.

Authors' response: As suggested by the referee, we will clarify the manuscript as follows:

P7079 L5 "(...) on the formulation of Hirsch and Stedinger (1987). *The plotting positions are used only for plotting return level estimates in Fig. 3, they are not involved in the model fitting process.*"

In addition, Appendix B (in the new version of the manuscript, the plotting positions are indeed detailed in Appendix B) will be clarified to indicate how each type of historical information (i.e. known values – h_1, only lower bounds – h_2, range of values – h_3) is dealt with in the method. For h_2 type, X_i is taken as the corresponding lower bound whereas for h_3 type, X_i is taken as the middle value of the corresponding range:

P7079 L9 "(...) (systematic or historical). *In the case of historical censored observations, X_i is taken either as the corresponding lower bound for historical events that exceeded a value but whose exact water levels are not known, or as the middle value of the corresponding range for historical events whose water levels are known to be within a given range of values.* Let (...)"

15. P7080 L12: This should cite Cunnane (1978): Cunnane, C. (1978) Unbiased plotting positions - A review. J. Hydrol., 37, 205– 222.

Authors' response: The citation will be added in the reference section of the manuscript.

16. Table 2: What is meant by “Standard estimative return values” here? The text implies that return levels are estimated by solving for $\tilde{x}_p$ after averaging over the parameter uncertainty (P7068 L7). However, the same paragraph states that credibility intervals are no longer defined, in which case how are those in Table 2 constructed? As with return period above, it is possible to algebraically solve for return level conditional upon the parameters which then provides a probability distribution for each return level estimate. If this was used, was the best estimate taken as the mean, median or mode?

Similarly, the table caption states that it provides 95% credibility intervals but are these central probability intervals (corresponding to the 2.5% and 97.5% quantiles) or perhaps a highest posterior density interval? The table caption should either indicate what the four cases are in the first column or refer to where they are defined (i.e. in Fig. 3).

Authors' response: For the first part of the referee's comment, we believe it is the same remark as Technical correction 13. Our response is detailed for Technical corrections 8 and 13. Considering the second part of the referee's comment, the standard estimative return levels are calculated from the maximum likelihood estimates of the posterior distribution of the parameters, which correspond to the mode of the posterior distribution of the parameters (see again Technical corrections 8 and 13). The 95% credibility intervals are central probability intervals, the caption of Table 2 will be clarified accordingly. Finally, we will follow the referee's suggestion by indicating in the table caption what the four cases are.

17. Figure 3: Of the four cases defined in this plot, the historical data is used in cases 3 and 4 while the 2010 data is used in cases 1 and 4. The order would perhaps be more intuitive if the 2010 data was instead used in cases 2 and 4 (i.e. if cases 1 and 2 were swapped).

Authors' response: We agree with the referee. The order of the four cases will be modified accordingly.

References:

Bernardara, P., Mazas, F., Kergadallan, X. and Hamm,L.: A two-step framework for over-threshold modelling of environmental extremes, Nat. Hazards Earth Syst. Sci., 14, 635–647, 2014.

Hawkes, P. J., Gonzales-Marco, D., Sánchez-Arcilla, A. and Panayotis, P.: Best practice for the estimation of extremes: a review, J. Hydraul. Res. 46(2), 324–332, 2008.

Please also note the supplement to this comment:
http://www.nat-hazards-earth-syst-sci-discuss.net/2/C3605/2015/nhessd-2-C3605-2015-supplement.pdf

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 2, 7061, 2014.