"Use of historical information in extreme surge frequency estimation: case of the marine flooding on the La Rochelle site in France" by Y. Hamdi et al.

Point by point response to Referee #1 and Referee #2 comments

By Yasser Hamdi (15 October 2014)

General comment	
Clarity and notation	See the next Table : *Specific comments* for section 3.1.1
Methodology	Section 3 was changed to talk about the GEV and GP distributions.
Application to marine flooding	This comment was also posted by referee # 2. Section 6 has been changed to take into account the comments of the reviewer and therefore the conclusion was also slightly modified.

Specific comments	
Section 1: The authors state that 'statistical characterization of extreme storm surges, using HI, has not been handled in the literature'. While not explicitly using HI, Coles and Tawn (2005) adopt a Bayesian approach to storm surge estimation that could easily accommodate HI through prior distributions. More generally, perhaps the authors might like to consider how incorporating HI compares to informative priors in a Bayesian analysis.	I added the word 'explicitly' to the sentence: "statistical characterization of extreme storm surges, using HI, has not been explicitly handled in the literature"
Section 3: If I understand correctly, type iii data comprise a time range above which no exceedances of some threshold occurred and the BMH model has be ruled out for such data. If this time range exceeds and encompasses one year, then it should be possible to treat type iii data as left-censored within the BMH model. See also general comments on censoring later.	The developed BMH model doesn't consider the 'upper bound' type of HI. In the case mentioned in the comment, this type of data can be considered as left-censored within the BMH model.
Section 3: The generalized extreme value and generalized Pareto distributions are not mentioned at all in this section, but are a crucial part of the methodology, and so should be central to the presentation of the methodology.	Ok. A sentence mentioning the GEV and the GPD is added.
Section 3.1: The 'second difficulty' highlighted by the authors is a key point, and is fundamental to the validity of the proposed methods. I can see that other users could benefit from the proposed methodology, and that it could easily be applied to other phenomena. Perhaps the authors should bring to the fore that it is vital that the mechanism behind the HI is fully understood and compatible with the systematic record, in particular that the HI should not bias the estimated distributions. Based on this, perhaps the comment on p.5651 l.29 should be checked, as presumably we want the outliers to be represented fairly.	Indeed, the second problem is a key issue. The Developed model cannot perform if the information collected is not exhaustive. As we mentioned in the modified section 6, the use of additional IH has a significant impact on the fitting and uncertainty. Therefore, completeness is a key point in the model developed.
Section 3.1.1: Quite a few entities are defined here and there seem to be quite a few combinations for each. I think there is scope to improve clarity. For example, *u* seems to have been reserved in the literature for threshold, as in Coles (2001). Furthermore, if I understand correctly, these are non-random, so lower case notation might be better. Similarly perhaps p_t might work better for the probabilities. I think if the authors can restrict themselves to *u* for thresholds, *t* for times and time periods, *n* for counts and *p* for probabilities, with suitable subscripts chosen, then this section will be much clearer, and much more consistent with typical statistical notation.	We decided to use the same notation as in the literature (Hirsh, Stedinger, Salas, … Ouarda, Gaume, Payrastre ...).
Section 3.1.2: Use of a partition, as in eq. (13), is very informative. I wonder whether it would be simpler to separate the systematic part of the model completely (as in eq. (25) for the POTH model), so that the likelihood has a term $\prod_{i=1}^{s} f_X\ x_i \mid \theta$ Then the historic part can be given separately, and partitioned according to exact, lower bound and range data. …	The section 3.1.2 is changed to consider and present the different likelihood functions for systematic, exact, lower bound and range historical data separately.
Section 3.1.2 and Section 3.2: The likelihood derivations are exhaustive, but perhaps better suited to being in an Appendix. This comment also extends to some of the calculations presented in Section 3.1.1.	We consider the likelihood developments and derivations, especially for the POTH model, a novel and key point in the present paper. That's why we prefer present them in the body of the article. In addition, the reader needs these developments for the comprehension of the models and how these models consider the different types of HI.

Technical corrections (or clarifications)	
Points	**Response**
1	Yes , the terms ' systematic data' or ' systematic period ' are often used in the literature when dealing with historical information.
2	« frequency models » refer to the frequency estimation models frequently used in hydrology.
3	Ok
4	HMax data are a type of historical data that can be used in POT model with historical information. You can see developments in section 3.2.4 for Maximum likelihood estimator for the HMax historical data.
5	Ok
6	Ok
7	Ok
8	Ok
9	Ok
10	The sentence: “However, even with important uncertainty, the use of HI is a viable mean to increase the representativity of outliers in the sample “ is replaced with: “However, even with important uncertainty, the use of HI is a viable mean to decrease the influence of outliers by increasing their representativeness in the sample “
11	Ok
12	‘HI’ was defined in the abstract
13	Ok. The paragraph regarding the use of ML was shortened.
14	Ok
15	Ok
16	There are many plotting positions formulas that can be used (Hazen, Cunnane, Weibull, etc.) to compute the empirical probabilities. The Weibull plotting positions were used in only the POTH model in which the GP distribution must be used (the exceedances converge asymptotically to a GP distribution)
17	Ok
18	It was written correctly in my word document.
19	Ok. The equation 5 is corrected. Mathematically, the integral is a summation: $\int_1^t n_{tp}\,dp = \sum_{p=1}^{t} n_{tp}$
20	The equation 6 is correct.
21	When the ‘a’ in equations 6 and 7 is equal to 0.5, the plotting positions becomes the Hazen’s one.
22	This is a very brief description for those unfamiliar with the ML method. The arrow in both directions below θ means the parameter space.
23	Ok
24	Ok. Modification done.
25	I used the index *j* for Y
26	I changed the p in page 5653 line 29 for p_{emp} (empirical probability)
27	Ok
28	Ok
29	Ok
30	Ok
31	I propose: Data and models settings
32	Ok
33	the Q-Q plot give a better visual adequacy than the return level plot. I propose to keep it.
34	Ok
35	Ok
36	Ok
37	Ok (please note that section 6 was changed in order to take into account your comments).
38	Ok
39	Ok
40	Ok
41	Ok
42	Ok
43	Ok
44	Ok
45	The independence issue when pot method is used was discussed in a previous paper (Y. HAMDI et al., 2014). The exceedances were extracted to be independent (a 72 hours between peaks, …).

Point by point response to Referee #2 comments

By Yasser Hamdi (15 October 2014)

The reviewer comment is very interesting. Section 6 has been changed to take into account this comment and therefore the conclusion was also slightly modified.