Authors reply on comments of referee #1

The overall impression of the paper is that, although the topic of research is interesting the paper is not publishable in the present form.

We thank the review for these comments. The paper will be revised to take them into account.

Comment R1 #1.1. Throughout the text "burned area" should read "area burned".

Depending upon the sentence, « burned area” will be changed into “area burned, or kept, since both are possible.

Comment R1 #1.2. The authors should better explain what is the benefit of using a Bayesian framework within this context as compared with other competitor approaches. Such explanation will strengthen the paper.

A Bayesian framework has several advantages compared to standard frequentist approaches. First, explicit *a priori* assumptions can be made about the model parameters, as has been done for the shape parameter ξ (see p.6, l.20-25 of the current manuscript). Second, Bayesian methods provide a direct assessment of the uncertainty related to the parameter estimation (see, e.g. Figure 4 of the current manuscript). Generally, frequentist methods only provide confidence intervals based on theoretical results which hold true for very large samples (i.e. asymptotically). Third, in Bayesian methods, the uncertainty of a quantity of interest can be quantified by means of the predictive distribution (as explained in p.9 of the current manuscript). This predictive distribution integrates over the posterior uncertainty in the parameters and can be obtained directly from the MCMC samples. Here as well, frequentist methods provide similar theoretical results which hold true under some restrictive conditions.

We agree that these explanations would benefit to the paper, and they will be included in the revised version of the manuscript.

Comment R1 #1.3. p.2, l.30-31: the authors stress that "...our EVT framework is implemented in an explicit non-stationary context adapted to our case study". The authors should better explain, in a convincing way, to which "explicit non-stationary context" are they referring to. This is not clear neither from the introduction nor from section 3.

The non-stationary context simply refers to the fact that the frequency analysis is applied for two distinct time periods, and three different regions. Therefore, we can consider that our EVT framework is non-stationary in time and in space. However, we agree that this was not clear in the manuscript and this will be better explained in the revised version.

Comment R1 #1.4. p.2, l.33-34: replace "...to assess uncertainties in return periods and related return levels." by "...to assess uncertainties in return levels."

Ok, this will be modified.

Comment R1 #1.5. p.2, l. 34: the sentence "This allows determining if the highlighted changes are actually significant." is not understandable as it is. Please elaborate.

This sentence indicates that the comparison of the results between the different time periods and regions can provide an assessment of the differences, for example how significant are then changes in return levels between 1973-1994 and 1995-2016. This will be better explained in the revised version.

Comment R1 #1.6. p.4, l.3: replace "(or GEV)" by "(or GEV, in-short)".

Ok, this will be modified.

Comment R1 #1.7. p.4, l.3-5: the sentence "Grounding on this strong mathematical result, we assume that wildfire samples for each year are sufficiently large, so that annual maxima of BA can be considered to follow a GEV distribution." is not understandable as it is. Please elaborate.

Extreme Value Theory indicates that maxima of regular large blocks follow a GEV distribution. This means that if these maxima are taken from very large regular samples, these maxima should follow a GEV distribution. In our case, these samples are the burned areas for each year. The database contains 106,000 wildfire records, and therefore we assume that wildfire samples for each year are sufficiently large. An additional sentence in the revised manuscript should clarify this point.

Comment R1 #1.8. p.4, l.5-9: the sentence "This has already been shown to be adequate...to cite a few of them" adds anything to the paper and should be removed from the text.

In our opinion, these citations are useful and should be kept in the revised version. First, other applications of the EVT show that this theory is now widely spread in applied fields and is handled by many practitioners. Second, these references provide several examples of applications of the EVT and can help the reader to better understand how it works (for example how the block maxima are obtained, how the parameters can be estimated, etc.).

Comment R1 #1.9. p.5, l.1: the authors should include in eq (1) the range of variation of x and theta.

We thank the reviewer for this comment, this will be done.

Comment R1 #1.10. p.5, l.3: the authors should also provide an explanation to the case xi=0.

When xi=0, the GEV distribution has no bounds, this will be indicated in the revised version.

Comment R1 #1.11. p.5, l.3: remove "(the bound cannot be exceeded)" from the text.

Ok, this will be modified.

Comment R1 #1.12. p.6, l.7: the sentence "In this expression, we assume that the normalizing constant of the posterior is not known." should read "It is assumed that the normalizing constant of the posterior distribution is unknown."

Ok, this will be modified.

Comment R1 #1.13. p.6, l.8-9: the sentence "The likelihood function...parameters theta" is not understandable as it is. Please elaborate.

The likelihood function represents the conditional density of the data **D**, here the logarithm of the maximum BA, given the parameters $\boldsymbol{\theta}$, and describes the plausibility of observing the data **D** given $\boldsymbol{\theta}$. This will be clarified in the revised manuscript.

Comment R1 #1.14. p.6, l.13-14: the sentence "Indeed, raw BA are two-dimensional data and have very skewed distributions, which often lead to extreme value distributions with an infinite variance" is not understandable as it is. Please elaborate.

Raw BA are measures of surface, i.e. two-dimensional data. It creates a scale issue, very large BA values (e.g. 1,000 ha) being a lot larger than large BA values (e.g. 100 ha). This scale issue becomes obvious when the distribution of the raw BA is represented, this distribution being extremely skewed. As a consequence of these very skewed distributions, extreme value distributions with an infinite variance (xi>0.5) are often obtained. We will clarify this point in the revised manuscript.

Comment R1 #1.15. p.6, l.24: the authors should better explain the reasoning for considering the joint prior distribution as being the product of the marginals; in other words, what is the justification for assuming that the GEV parameters are independent? Furthermore, in the joint prior expression, the term \Phi(\sigma) should read \Phi(\log\sigma).

In the absence of *a priori* information about some type of dependence between the GEV parameters, it is usual to use independent priors (see, e.g., p.174 in Coles, 2001). This does not preclude from quantifying inter-parameter correlation within the joint posterior distribution if there is evidence in the data that such correlation actually exists. We will add this point to the revised manuscript.

Comment R1 #1.16. p.7, l.18: remove "(for example the test of Kolmogorov-Smirnov)"

We do not understand why this example of a statistical test for the equality of two distributions needs to be removed. This example of a standard statistical test can help a non-statistician reader to understand this paragraph.

Comment R1 #1.17. p.8, l.3: the authors should better justify the choice of the Bhattacharyya coefficient for measuring the distance between distribution functions. Furthermore, the authors should also highlight the advantages and limitations of such coefficient when compared with other competitors such that Mahalanobis distance or the weighted L2-Wasserstein distance, among others. Section 3: the authors should carry out the same analysis but considering several distance measures and to compare the results. This will strengthen the paper.

It is true that many options exist in terms of ‘generalized distances”, or divergence measures between two distributions. Mahalanobis (1930) first acknowledges that tests of significance are sometimes limited and for this reason he introduced the idea of divergence of two populations. In our case, Mahalanobis distance is not well suited since the sample version assumes a common standard deviation (see McLachlan 1999 p. 24). In this study, the distributions to be compared have different standard deviations (see mu for Pcr-2 in Fig. 4). However, many other alternatives could be applied. For example, the Kullback-Leibler divergence is widely spread, and the weighted L2-Wasserstein distance is another alternative. Different measures have been tested and they lead to identical conclusions. As an example, Table 1 presents the Kullback-Leibler divergence applied on posterior densities of the GEV parameters, before/after 1994, for each region, to be compared with Table 1 of the current manuscript.

In this paper, the Bhattacharyya coefficient is applied mainly because it can be easily normalized between 0 and 1, which facilitates its interpretation. However, we agree that this choice should be further motivated and discussed in the manuscript. This will be done in the revised version.

Table 1: Kullback-Leibler divergence applied on posterior densities of the GEV parameters, before/after 1994, for each region. Bold values indicate coefficients below the reference value of 2.16 (for two normal distribution N(0,1) and N(2,1)).

Zone	μ	σ	ξ
PCr-1	**9.17**	0.46	0.08
PCr-2	1.73	**3.26**	0.53
PCr-3	0.15	0.40	0.09

Comment R1 #1.18. p. 6-7: the authors should clearly explain the steps for implementing their algorithm. In my opinion the authors fail in explaining their procedure in a way understandable to the reader. Furthermore, details on how their method is implemented in practice are largely omitted which difficult the procedure understanding. For example, nothing is said about the convergence of the MCMC method. Did the authors check the convergence of the algorithm? Did the authors checked for the presence of metastability? The authors also should include a table displaying the acceptance rates for the GEV parameters within the MCMC algorithm.

At p.6 l.25-30 of the current manuscript, we indicate that we use the Metropolis-Hastings algorithm in order to sample the posterior distribution. In details, we use the function MCMCmetrop1R from the MCMCpack package in R software (R 2017). We also indicate that the multivariate normal proposal distribution is scaled using the covariance matrix of the parameter estimates when the maximum-likelihood method is applied. This is achieved using the fgev from the evd package in R.

As indicated in the manuscript, for each estimation, we produce a burn-in sample of size 100,000, and the retained sample used to represent the posterior distribution is of size M=10,000, for which a thinning interval of 10 is applied in order to reduce the autocorrelation inherent in MCMC chains produced with the Metropolis-Hastings algorithm. Such a long burn-in period usually avoids convergence issues for these dimensions (i.e. 3-dimensional parameter vectors). Table 2 presents the acceptance rates for the GEV parameters, and lie between 0.32 and 0.42. These acceptance rates are reasonable and correspond to common requirements. Indeed, acceptance rates between 0.23 and 0.5 are usually advised in order to obtain a fast convergence of the algorithm (Robert and Casella 2004).

The convergence was checked visually. As an example, Figure 1 shows the traces of the sampled parameters and density estimates for each parameter, for the region PCr-3 and period 1995-2016. The trace is regular and no jumps can be observed. This assessment did not reveal any convergence issue.

The details about the implementation of the algorithm will be added to the revised version of the manuscript, as well as Table 2 displaying the acceptance rates for the GEV parameters.

Table 2: acceptance rates from the Metropolis-Hastings algorithm.

Zone	1960-1994	1995-2016
PCr-1	0.42	0.34
PCr-2	0.35	0.32
PCr-3	0.33	0.40

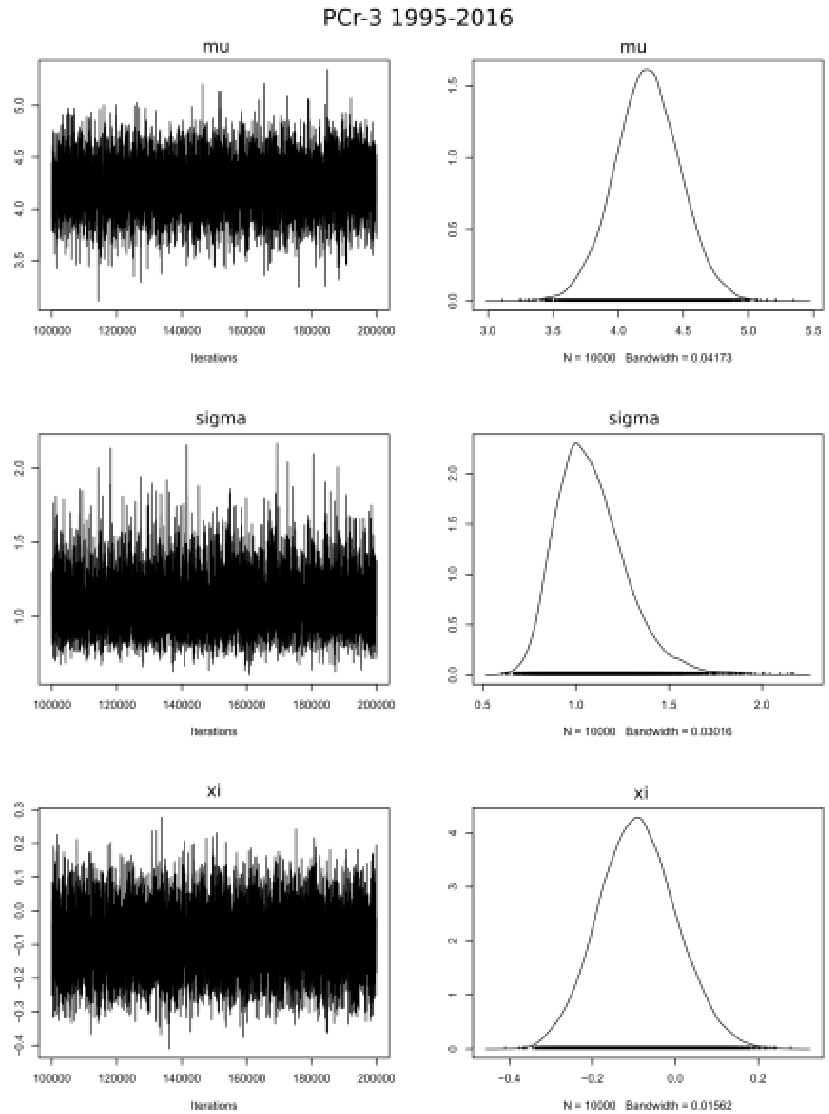

Figure 1: Trace of the sampled parameters and density estimates for each parameter, for the region PCr-3 and period 1995-2016.

Comment R1 #1.19. p.8, l.20: the sentence "Posterior distributions of parameters..." should read "Posterior C3 distributions estimates for the extreme value parameters..."

Ok, this will be modified.

Comment R1 #1.20. p.8, l.20: replace "sigma" by "log(sigma)"

Ok, this will be modified.

Comment R1 #1.21. p.9, l.1: replace "...corresponding BA is ...." by "...corresponding BA estimate is ...."

Ok, this will be modified.

Comment R1 #1.22. p.9, l.2: replace "...predictive distribution of these..." by "...predictive distribution estimate of these..."

Ok, this will be modified.

References

McLachlan, G. J. 1999. “Mahalanobis Distance.” *Resonance* 4 (6): 20–26. https://doi.org/10.1007/BF02834632.

R. 2017. *R: A Language and Environment for Statistical Computing* (version 3.4.0). R Foundation for Statistical Computing. Vienna, Austria: ISBN 3-900051-07-0.

Robert, Christian, and George Casella. 2004. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. New-York: Springer-Verlag.