

Response to Reviewer 1 comments about the article “*A Bayesian Statistical Method to Estimate the Climatology of Extreme Temperature under Multiple Scenarios: the ANKIALE Package*”

ROBIN, Y., VRAC, M., RIBES, A., BARBAUX, O. and NAVEAU, P.

October 23, 2025

Note In this document, the text in regular format corresponds to the reviewers questions. The answers from authors are given in the grey blocks.

1 Reviewer 1 (Anonymous)

1.1 Global comments

I think this is an important piece of work – it extends the use of observational constraints to the estimation of the characteristics of temperature extremes for unobserved past and future periods under both “factual” and “counterfactual” conditions, using a rigorous Bayesian framework in which extreme temperature distribution is described with a suitable extreme value distribution. The ANKIALE package that implements the method should help to make this sophisticated methodology relatively accessible to a broad range of users.

We would like to thank Reviewer 1 for this summary and for recognizing our efforts to make a Bayesian-based statistical analysis method for extreme events accessible to as many people as possible.

Unfortunately, however, the paper would be VERY challenging for the target audience to understand. For this work to be impactful, I think it will be necessary for the authors to think much more carefully about presentation issues, providing more complete and more accessible explanations for the choices that they have made in implementing the package. I think they also need to provide substantially more insight into choices users will have to make when applying the package, with an emphasis on physical considerations as well as statistical and pragmatic considerations. Also, to make the paper accessible to users will require a careful redesign of the notation that is used in the paper, which is hopelessly complex.

We thank Reviewer 1 for these relevant comments, which are in line with those made by the other reviewers. We apologize that the presentation was not as accessible as we would have liked. The paper has been reworked in this regard. We rewrote the paper in order to:

- Provide the minimum theoretical tools, referring to the appendix for technical details when possible,
- Better highlight the highly extensible nature of ANKIALE: ANKIALE currently only handles a GEV model, but other statistical models (suitable for other variables) can easily be added.

1.2 Specific comments

Line 37 It is unclear why using different scenarios would result in different estimates of the counter-factual world.

It would help if a bit more were said here. Reading ahead, it turns out that this is discussed more beginning at line 113 and I can see from that discussion how this could arise. It is not made clear, however, whether the selection of a particular scenario would strongly affect inferences about observed events based on the posterior that results from using that scenario versus inferences that would be made with another scenario. Sensitivity to scenario choice, particularly if large (and especially under counterfactual conditions) would be of concern, but making that go away artificially by using information from all available scenarios doesn't really solve the problem in a satisfying way. It would remain a concern that information from models about the future can somehow affect our understanding of the past unless there is a convincing physical argument about why that makes sense. On the other hand, if the sensitivity is small then there wouldn't be a very compelling reason to bother with the added complexity of the prior and its dependence on the particular experimental design that was adopted in CMIP6. In summary, I think this is crucial point that needs clarification (and in each application, physical justification).

Indeed, our explanations here were a little brief. The estimation of the counterfactual always depends on the scenario. For example, a scenario such as SSP1-1.9 (not used here), which is very 'flat', makes the estimation of the parameters μ_1 and σ_1 that drive the trend (and which are not supposed to depend on external forcings) much more uncertain than for an SSP5-8.5 scenario, where the trend is very strong. Theoretically, since μ_1 and σ_1 do not depend on the scenario, they should be able to be estimated solely with the counterfactual covariate, restricted to the period 1850–1950 (where the anthropogenic term is very weak). However, in practice, this estimation is impossible: the design matrix during regression is ill-conditioned (no maximum rank). Our approach therefore has two advantages: it ensures a scenario-independent counterfactual, as well as estimation for scenarios where inference is difficult because the signal is weak.

Furthermore, knowledge brought by the simulations for the future period does indeed affect parameter estimation. Although this may seem strange from a physical point of view (breaking of causality), from a statistical point of view, the parameters estimation needs to rely on a continuous time series, including historical and future simulations. Therefore, the final estimation, even over the historical period, will depend on the future scenarios.

We added a sentence in the introduction stating that this applies in particular to confidence intervals, and this has been detailed in the methodology.

Line 70 What is the point of the right-hand panel in Figure 1? It adds a bit of confusion by hinting that you will make inferences about the record event during 1940 to 2024 at every land point in the domain, irrespective of when that event happened or the spatial extent of

the event that produced the record.

The application example in Sect. 6 concerns the assignment of the maximum observed at each grid point, which is therefore given in Fig. 1b. We clarified the reference to the figure by indicating that Fig. 1b is used in Sect. 6.

Line 75 Comparison with a single long record that is almost surely inhomogeneous (e.g., due to instrument changes, observing procedure changes, development of the urban environment around the station, etc) is not going to do a great deal to increase confidence.

After verification with Météo-France, it appears that this time series has indeed not been homogenised. Then, we redid the example with ERA5 in order to have a single data set.

Line 77 Usually, “external” forcing would mean external to the climate system (solar, volcanic, ghg’s, aerosols ...) rather than external for France

That is right, we have rephrased it.

Lines 77-79 If read literally, the sentence could be interpreted as saying that you extract European mean temperatures from HadCRUT5 and global mean temperatures (more correctly, temperature anomalies) from GISTEMP. Why do you use these two datasets rather than just using one?

In addition, Fig. S1 notes the use of the BEST dataset (Berkeley Earth) – why use yet another global dataset when consistent use of one, well regarded dataset, would probably suffice?

When we began this work, we used BEST to estimate the GMST. As BEST is known to have a warm bias, we then switched to GISTEMP. Finally, for calculations on a sub-region such as Europe, we used HadCRUT, which is the data set historically used for Europe in RR20. This explains the presence of GISTEMP and HadCRUT, and the fact that BEST still appears was an error. GISTEMP was retained (i) to provide a test example where the observation series do not cover the same period during the calculation, which made it possible to verify the algorithms, and (ii) because HadCRUT does not cover all the planet in the distant past. We fixed the error for BEST, but kept two different datasets. A sentence has been added to Sect. 2.1 to explain why we have two data sets.

Lines 85-86 My understanding is that the historical forcing prescription used in CMIP6 is NOT part of the SSPs, which only cover the period from 2015 onwards.

That is right, the SSPs only cover the period after 2015, with the period 1850–2014 being given by the so-called historical scenarios. It is corrected.

Lines 90-91 I’m not aware that the IPCC assessed, in its synthesis report, that the current emissions trajectory is leading us towards SSP2-4.5. This is discussed by others, however, so I think you should provide a more suitable reference or delete this statement. Indeed, if you have high confidence in this statement, then it would seem that there would be no need to use the other scenarios.

Sentence deleted.

Line 101 The certainty expressed here that the variable of interest will be GEV distributed

seems a bit of an overstatement. The GEV distribution is a limiting distribution for block maxima that is (sometimes) achieved as the block length grows without bound. Convergence to the limiting distribution (if it happens at all) can only be demonstrated theoretically under very idealized conditions. Nature, and climate models, do not comply with those conditions (we have awkward things like an annual cycle and the presence of multiple extremes processes that complicate life considerably, with the result that the upper tail does not always behave like that expected under idealized mathematical conditions. While we can't really look into the deep upper tail with observations, and can do so, albeit with some difficulty, with climate models (Alaya et al., 2020). Experience shows that the GEV is nevertheless often useful for approximating the distribution of block maxima for blocks of even modest size (e.g., a year, which effectively only samples part of the year due to the annual cycle). The authors know all of this, and it would be good if some of this could be reflected in the paper, particularly as it is intended to introduce the methods and the ANKIALE package to a wide audience who are not as knowledgeable about the application of the GEV distribution and its limitations.

Indeed, this statement was exaggerated. The methodology section has been completely rewritten, and we are adding a discussion about this choice in the appendix.

Line 103 What is the time range considered? Also, I find the notation here somewhat confusing. Readers in a hurry will confound the index F (for factual) with “future”, and might confound the index “0” (zero) with “O” (for “observed”). A further question is whether readers should think of the three components of X as being random or fixed.

The time axis varies here from 1850 to 2100 for climate models and from 1950 to today for observations. The notations has been revised at the same time as the theoretical section.

Line 109 The use of the $*$ to indicate the reader should make a substitution for R or G is awkward and mostly just makes comprehension a bit more difficult for the reader.

The notations has been revised at the same time as the theoretical section.

Lines 114-115 Replace “supposed” with “assumed”. Also, this assumption merits some discussion.

A discussion has been added.

Lines 119-120 See the comment concerning line 37 above. This needs discussion – particularly why including different futures would affect our understanding of the past.

See our response to the comment concerning line 37 above.

Lines 126-127 It might just be a French/English problem, but what this first step in the procedure entails could be better explained. This would include saying what the assumptions are that are implicit in calculating the uncertainty covariance matrix. It is not clear from the notation if there is one such covariance matrix that describes the spread amongst the different θ_m (which I am guessing is the case) or whether each θ_m has its own uncertainty matrix.

We have completely rewritten the entire methodology section.

Lines 131-132 It would be much better if this paper could be self-contained rather than sending readers off to another reference for the parts of the methodology that have not changed.

The paper has now been rearranged to contain all the necessary information.

Lines 137, 139 While the paper is generally readable, there are many minor grammatical errors. Two examples are mentioned here. This is less excusable these days given the wide availability of tools for polishing text (assuming that GMD authors are permitted to use them).

- At line 137, replace “3-days moving average” with “3-day moving averages”.
- At line 139, where the sentence seems unclear. In that sentence, rather than “are”, do you mean “are estimated with”?

We apologise for these errors and have corrected the text.

Line 147 The white noise assumption needs some justification. This might be roughly suitable for European regional mean annual surface air temperature anomalies, but the white noise assumption seems a bit less obvious for annual global mean surface temperature anomalies.

This issue was addressed by Ribes et al. (2021) and Qasmi and Ribes (2022). Their conclusion is to use white noise in climate models, but to use a mixture of two AR(1) processes (a slow one for processes of the order of decades, a fast one for inter-annual processes) for the internal variability of observations during constraint. We have modified the text accordingly.

Lines 148-154 I find myself struggling to understand what is really done here, both because of the notation, which is increasingly complex, and because it is not obvious what model output is being used. If a model has 50 ensemble members, do you use all 50? And if so, do you treat that model differently from a model with only 1 ensemble member? What period is considered, how was the choice to use a smoothing spline with only 6 degrees of freedom made, how are the knots placed, do you worry about the fact that the knot placement is arbitrary and that this imposes wave-like fluctuations that are probably not part of the forcing response?

We have rewritten the whole Section 3. In a nutshell:

- All members of a model are used, and therefore models with one member are treated in the same way as those with 50 members. However, the latter have a lower degree of uncertainty, which affects the construction of the prior.
- The use of 6 degrees of freedom for splines was established in the original article by Ribes et al. (2020). It is true that this estimate was valid for the RCP 8.5 (CMIP5) scenario, with natural splines; whereas we use *B*-splines with different CMIP6 scenarios. ANKIALE has now been modified to be able to play on all these parameters simultaneously: base size and degree of freedom can be imposed for each scenario/covariate.

This information has been added to the article.

Line 156 Figure S2 is referenced well before Fig S1...

It is now corrected.

Line 177 This statement is made with a lot of certainty and conviction, but whether an event would be judged to be impossible, even under anthropogenic forcing, is highly uncertain. It seems clear from Fig. S3 that the value of the shape parameter is driven by the extreme temperature that is farthest from the location parameter and hence must be very uncertain. This relation between the shape parameter and the most extreme temperature presumably occurs because the parameter estimation process enforces the feasibility of the fitted GEV distribution to a variable, temperature, that tends to have light-tailed extreme value distributions.

Indeed, it is well known that the GEV law tends to underestimate the true upper bound of a variable (if it exists). The sentence has been removed.

Lines 184-185 Why is this assumption needed to construct the prior? It's a prior distribution (i.e., a proposal) that will be updated using the observations when the posterior is derived. It seems to me that this assumption is not needed to construct the prior. Given the way the prior is constructed, it would certainly be helpful if we can regard the models as being indistinguishable from each other (i.e., something hopefully like a simple random sample from model space), but even without that, couldn't we construct a prior from the models, understanding that it may not do a good job of representing model uncertainty? The updating does require us to have a joint distribution for model simulated and observed quantities, and developing that joint model may require some additional assumptions – perhaps that's where the “indistinguishable from the truth” assumption comes into play?

Indeed, this hypothesis is not used in itself for the construction of the prior; it serves to justify that this multi-model synthesis method proposed by Ribes et al. (2017) constructs a prior in which the observations are hidden, and therefore that it is reasonable to take it as the prior of reality. The section has been rewritten.

Lines 189-190 I think this needs discussion – in particular, why internal variability plays a role at all (haven't you filtered it out with the splines?) and what is being partitioned into two components. What is being referred to when you talk about the “common part” of internal variability and each model's additional internal variability??

It is assumed that the covariance matrix Σ_{θ_m} of the random variable θ^m of model m can be decomposed into the sum $\Sigma_{\theta_m} = \Sigma_u + \Sigma_m$ of two terms of internal variability; This internal variability is not that of global or European average temperatures, but that of the θ parameter vector. These two terms are:

- Σ_u , which does not depend on the climate model;
- Σ^m , which is a term specific to model m .

All of this has been rewritten and clarified in the supplementary material.

Line 200 I have no idea what is being referred to here (95% of the covariance matrix)...

Apologies for the expression, we are of course referring to the ellipses defined by the covariance matrices, which correspond to the 95% quantile levels of the multivariate Gaussian distribution. We have edited the article for clarity.

Line 202 Which model is excluded? Note that the UK models maybe be problematic due to a known problem in the coupling between the land surface and atmosphere that leads to

extreme high localized daily maximum temperatures. The problem is documented here¹. Note that the Australian ACCESS models, which also use a version of the UK MetOffice atmospheric model, are not affected but the Korean KACE model, which uses both the atmospheric model and the Jules land surface model, is affected (an erratum has not been published for the KACE model).

Thanks for your relevant informations. In our case, this is the Norwegian Earth System Model NCC / NorESM2-LM. We have added this information to the article. (Seland et al., 2020).

Lines 238-239 Why not use internal variability estimated from climate models rather than relying on the one, very limited realization we have been able to observe?

The Gaussian conditioning theorem requires the knowledge of an estimate for the internal variability of the observations used for the constraint. The internal variability of the models is not necessarily that of the observations.

¹<https://errata.ipsl.fr/static/view.html?uid=76b3f818-d65f-c76b-bfd8-cae5bc27825c>

Bibliography

- Alaya, M. A. B., F. Zwiers, and X. Zhang (Aug. 2020). “An Evaluation of Block-Maximum-Based Estimation of Very Long Return Period Precipitation Extremes with a Large Ensemble Climate Simulation”. In: *J. Clim.* 33.16, pp. 6957–6970. ISSN: 0894-8755, 1520-0442. DOI: 10 . 1175/JCLI-D-19-0011 . 1.
- Qasmi, S. and A. Ribes (Oct. 2022). “Reducing Uncertainty in Local Temperature Projections”. In: *Sci. Adv.* 8.41, eabo6872. DOI: 10 . 1126 / sciadv.abo6872.
- Ribes, A., S. Qasmi, and N. P. Gillett (Jan. 2021). “Making Climate Projections Conditional on Historical Observations”. In: *Sci. Adv.* 7.4, eabc0671. DOI: 10 . 1126 / sciadv.abc0671.
- Ribes, A., S. Thao, and J. Cattiaux (2020). “Describing the Relationship between a Weather Event and Climate Change: A New Statistical Approach”. In: *J. Clim.* 33.15, pp. 6297–6314. ISSN: 0894-8755, 1520-0442. DOI: 10 . 1175 / JCLI-D-19-0217 . 1.
- Ribes, A., F. W. Zwiers, J.-M. Azaïs, and P. Naveau (2017). “A New Statistical Approach to Climate Change Detection and Attribution”. In: *Clim Dyn* 48.1, pp. 367–386. ISSN: 1432-0894. DOI: 10 . 1007/s00382-016-3079-6.
- Seland, Ø., M. Bentsen, D. Olivié, T. Toniazzo, A. Gjermundsen, L. S. Graff, J. B. Debernard, A. K. Gupta, Y.-C. He, A. Kirkevåg, J. Schwinger, J. Tjiputra, K. S. Aas, I. Bethke, Y. Fan, J. Griesfeller, A. Grini, C. Guo, M. Ilicak, I. H. H. Karset, O. Landgren, J. Liakka, K. O. Moseid, A. Nummelin, C. Spensberger, H. Tang, Z. Zhang, C. Heinze, T. Iversen, and M. Schulz (Dec. 2020). “Overview of the Norwegian Earth System Model (NorESM2) and Key Climate Response of CMIP6 DECK, Historical, and Scenario Simulations”. In: *Geosci. Model Dev.* 13.12, pp. 6165–6200. ISSN: 1991-959X. DOI: 10 . 5194/gmd-13-6165-2020.