Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2020-454-RC2, 2021


# Comment on amt-2020-454

Anonymous Referee #3

---

Referee comment on "Global ensemble of temperatures over 1850–2018: quantification of uncertainties in observations, coverage, and spatial modeling (GETQUOCS)" by Maryam Ilyas et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2020-454-RC2, 2021

---

Summary of paper: In recent years, there have been a number of new reconstructions of the global temperature dataset, including at least three major government agencies (NOAA and NASA in the US, the Meteorological Office in the UK, maybe I should count Japan as a fourth) and some private efforts (the Berkeley Earth group and the Cowtan-Way-Robert papers). In an earlier paper published in 2017, three of the present authors sought to improve previously established methods by using the "LatticeKrig" approach to do spatial interpolation. LatticeKrig is a very sophisticated form of spatial statistics modeling that uses multi-resolution basis function expansions combined with Markov random fields for the coefficients. It's a much more advanced method than simple kriging, but the latter is still used by many authors (including some of the competing global temperature groups) for this kind of work. The present paper further extends this by using the approximate Bayesian computing (ABC) approach to construct a Bayesian version of the same algorithm. The advantage of the Bayesian approach is that it incorporates uncertainty in some of the spatial model parameters and thereby, in principle, should result in more precise assessments of uncertainty in the final product. The project required an enormous amount of computing (estimated at 62 months computer time, reduced to 8 months by parallelization) but the end result is a new data product that can realistically be compared with the other constructions that have been made.

Assessment: In the end, it seems to me that the authors come out with an incremental improvement on their own previous construction and it doesn't seem to be the "game changer" that the authors were possibly hoping for when they set out along this path. For example, the uncertainty bounds in Figure 5 are only very slightly different from those in their own previous construction, and the extended example on pages 9-11 seems to have been specially constructed as a worst-case scenario (as is explained in the paper, May 1861 was the month with the least spatial coverage of the entire dataset, therefore, presumably, the one for which the advantages of this kind of approach should be most clearly seen). Nevertheless, the work seems to be been competently carried out and it is always useful to have another dataset for comparison - I fully support publication.

My main concern about the paper is that the manuscript itself seems to have been put together in some haste. Prior to reading this, I was familiar with the LatticeKrig approach but not with the intimate details. Some of those details are important, e.g. the definitions of the lambda and aw parameters (page 4). I believe, throughout the paper, there is a need to give more explicit detail about the method. Given the 62 months of computer time, it seems unlikely that anyone else would want to exactly reconstruct this dataset, but nevertheless, I still feel that this should be a requirement of publication, that the method should be described in sufficient detail that anyone who wants to reconstruct the result has all the information required to do so.

In my initial correspondence with the editor, I queried why the authors had not produced a "supplemental materials" document - as I understand, such a document would be supported by the journal and I might suggest that the authors take advantage of this when revising the paper. Specifically, the kind of detail that is probably not relevant to the casual reader of the paper, but would be needed by anyone actually intending to try to reproduce the results, could very well go into an online appendix.

One further "general" comment - there is a passing reference to HadCRUT5 (Morice et al. 2020), which is the latest version of the U.K. Met. Office model. I wonder if maybe the authors should say a little more about this. I presume the production of HadCRUT5 overlapped the present effort but would the authors like to comment on how HadCRUT5 improves on HadCRUT4 and specifically how it compares with the present work?

Specific comments:

p. 1 line 16 - "phenomena" (plural of phenomenon)

p. 2 l. 15 Here the author Rohde is misspelled Rhode. This error also occurs in p. 2 l. 20, p. 3 l. 1 and p. 18 l. 24

p. 2 l. 34 Reference to JMA (?) where the question mark is a standard latex warning for a missing reference. Were the authors referring to the paper Ishii et al, mentioned on p. 2 l. 14?

p. 4, l. 19. Here the authors mention two parameters from Nychka et al 2015, called lambda and aw, but they never define these two parameters. The implication is that one can look up Nychka 2015 to find these definitions but I tried doing that and I think we need more assistance.

Nychka (2015) defines a parameter lambda=sigma^2/rho but they don't call it a smoothing parameter - that was my first confusion. I do note that the LatticeKrig R manual also defines lambda and does call it a smoothing parameter - the most recent version of this that I downloaded for preparing this review was version 8.4 dated November 2019 (the authors of this paper refer to version 6.4 as the one they used for the bulk of their computations). If the present authors want to call it "the smoothing parameter" without further explanation, they need to be precise about where this is defined, and the answer appears to be the LatticeKrig manual (which I'll subsequently refer to as LK), not Nychka (2015) (henceforth N15).

Both N15 and LK say that lambda and rho are computed my maximum likelihood and I understand that one of the objectives of the present paper is to extend that by using the ABC approach to approximate lambda, but what happened to rho? This isn't explained here, but later (p. 6 l. 4) they say, "both d and rho are still estimated using the maximum likelihood approach".

In fact d and rho are respectively an overall mean and a variance (scaling) parameter of a multivariate normal distribution and it is well known and trivial to implement that these parameters can be integrated out analytically, so as to focus attention on the spatial correlation parameters - in my first reading of the paper I assumed this was what they had done. But the comment on p. 6 l. 4 makes me wonder about this point. In summary, we need clarification of what the authors actually did. (If it really was a Bayesian approach, we also need to discuss the prior distribution, since rho in particular may require a proper prior.)

Now let me turn to the other parameter, aw, referred to as "autoregressive weights". This is based on the fact that at each level of the multiresolution process that defines the spatial model, the coefficients of the radial basis functions have the structure of a lattice process that is assumed to be of conditional autoregressive (CAR) form. However, here is no fixed structure for this and no single parameter called the autoregressive weight (or weights - it's not clear whether the authors actually meant to use a plural form here). p. 584 of N15 refers to a weight matrix B where the off-diagonal entries are -1 and the diagonal entries are of form B_{j,j}=4+kappa^2. In this case the lower bound 4 arises because the sum of entries in B is required to be positive. So is kappa the autoregressive weight? LK actually use a different notation, where they define a variable a.wght (this is the nearest I can find to any variable actually called "aw") and they comment (p. 23), "in the simplest case a.wght is the central value, and should be greater than 4". So is a.wght the same as 4+kappa^2 in the N15 notation? If so, which parameterization do the present authors actually use? Later (p. 8 l. 21) the authors give aw a prior density that is uniform on the interval (1,4), but now I'm really confused about how that particular range was determined ...

Another potential wrinkle is that N15 p. 585 explicitly mentions the possibility that the autoregressive structure may be different at different levels of the multiresolution process, but I'm reading between the lines that they didn't consider that extension in this paper.

I don't actually think any of these questions are complicated. I understand very well that there are certain model choices that you just have to make. The authors simply need to be explicit about what these choices were and how exactly the various parameters are defined.

p. 5. The flow diagram illustrating the ABC approach is clear and should be easy for the reader to follow, but again, some specific details are missing. How do they determine the number H of variogram sampling points and the specific vector of distances represented by h? I'm assuming that when you write gamma(h) you don't literally mean h as the distance (h is an index going from 1 to H) but gamma(d_h) for some distance d_h, but then the same question, what values did you actually use and how were they chosen?

p. 6 equations (5) and (6). I don't think we need you to define every symbol here but please give an exact source for these equations, which I assume are somewhere in N15?

p. 6 lines 16,17 - reference to a new dataset HadCRUT5 (which I wasn't aware of myself until reading this paper). I presume HadCRUT5 came out while this paper was being developed. I think this sentence should be moved to the discussion section and the authors should discuss how the two approaches compare and contrast with one another - are there any features in which HadCRUT5 improves on the present approach?

p. 7, l. 22 - "since the last few decades" - slightly awkward English construction here, maybe "during the last few decades" would be better. I am aware that the word "since" would be used in several other languages, for example "depuis" in French.

p. 7, l. 28. Unclear why you set any negative value to 0.0. While I'm well aware that we all talk about global warming and not global cooling, I don't think the possibility of cooling is excluded by basic atmospheric physics - if a stochastic model occasionally produces a negative value, why not include it in the analysis? From a political point of view, the authors should take care to avoid any implication that their approach was predetermined to result in a warming outcome.

p. 8 l. 28. This review is already getting rather lengthy and by this point I was definitely suffering from reviewer fatigue, but if I'm not mistaken, this is the first time in the paper there is any parameter called alpha. Please, either define it, or give an explicit prior reference where it is defined.

pp. 9-11. The authors are quite explicit that May 1861 was chosen for this illustration because it was the month with the poorest spatial coverage, and therefore presumably the one that best illustrates the advantages of using a more refined spatial approach, but I think it would be helpful to have at least some comparisons with other months. Are these kinds of plots typical of what we would expect if we just chose a month at random?

p. 10 lines 3-4: so part of the reason for the difference is that the LKrig function was improved between the two versions of LatticeKrig that were used for the 2017 paper and this one? Could you expand on that a bit - was that a major factor?

p. 12 l. 24 and Figure 5b: what exactly is the "median time series"? I'm inferring that each section of the time series was centered about its median value but what time scale was used for calculating the medians?

p. 14, this is an additional feature that is only introduced later in the paper and somewhat complicated to evaluate. It seems that the authors do not intend to publish their full 100,000-member ensemble but only a subset selected by a conditional latin hypercube sampling (CLHS) approach? I'm sure there are good reasons for doing that but at least from the appearance of Fig. 6, there appear to be some nontrivial differences between the two approaches, or am I misinterpreting this figure? Once again, the fact that they have shown this figure only for May 1861 may in some sense be a worst case scenario, but it would be helpful to clarify that point.