Comment on gmd-2022-32

Anonymous Referee #2

Referee comment on "Towards variance-conserving reconstructions of climate indices with Gaussian Process Regression in an embedding space" by Marlene Klockmann et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-32-RC2, 2022

I would like to first apologize to the authors for the very long delay in the submission of my review (which was partly releted to an accident).

The authors presented a new method which can eventually be used for reconstructing past climate indices from proxy data. The method is applied to pseudo-proxy records based on long model simulations and the results are presented and analyzed methodically. While the paper underlines some limitations and challenges of the method, it is well written and undoubtedly provides an important step in the development and application of Gaussian Process Regression methods to the area of paleoclimatology, and I look forward to future work utilizing real proxy data.

The quality of the paper is generally high, the analysis appears accurate from my knowledge, and there are no major points which I think should prevent its publication. There are however several minor points and suggestions which I identified and reported below, which I think could improve the paper further.

Lines 33-34: "however" is awkward in the middle of the negation.

Line 134: Might be useful to describe the term 'hyperparameter' for those outside the machine learning field, how is it different from a normal 'parameter'?

Lines 148: So the batch size is the total number of observations across records (i.e. 5 records with 100 observations plus 2 records with 200 observations would mean a batch size of 900)?

Lines 163-166: What is the batch size of the full GP here in comparison to the sparse GP?

Lines 180-183: How are the irregular resolution of the proxies handled (for the realistic P2k case)? Or are the pseudo-proxies created at annual resolution? In which case it would be helpful to briefly hint in the discussion how realistic irregular proxies could be used in the future and how it might affect the results.

Equation 3: Should RA be SR?

Lines 248-249: So the embedding coordinates are calculated from the distance matrix via multi-dimensional scaling? Might not be obvious for the layman (including me) how the x matrix is obtained, could it be made more explicit in an Appendix?

Line 259: "obatain"

Equation 6,7,8,9: Shouldn't the matrix be equal to something for an equation? Make explicit which one is the input and which the output.

Lines 299-300: Could it be because CCSM4 is more homogeneous (less spatial degrees of freedom) since the mean embedding distance between the records is smaller? Just a thought.

Figure 2: Unclear to me what the 95% confidence intervals represent in the temporal domain? Are they the spread of the unsmoothed data? Also for the spectra it is not explained what the confidence intervals are, simply chi-square CI with 2 degrees of freedom? It is a question of style and not necessary, but I personally like smoothing the spectra to make for a clearer comparison.

(e.g. Using a Gaussian smoothing kernel as in JW Kirchner, Aliasing in $1/f(\alpha)$ noise spectra: Origins, consequences, and remedies. Phys Rev E Stat Nonlin Soft Matter Phys 71, 066110; 2005).

Caption Figure 2 and 3: "powerspectra" -> "power spectra"

Figure 3: I don't understand the indicated 95% CI number. Do those correspond to the same CI shown on Figure 2?

Section 3.2: I would generally favour calculating the statistics for individual ensemble members and reporting the mean+/- standard deviation rather than calculating them with respect to the ensemble mean. Similarly, I would show the mean of the spectra rather than the spectrum of the mean for Figure 3 b,d,f as it is more representative of the real result one would obtain.

Figure 3 b,d,f: I wonder whether the PCR has the right high-frequency amplitude for the right reason? Are the high-frequencies just noise and thus PCR doesn't perform better than the other methods or are they actually correlated with the real series?

Lines 329-331: Do the authors have an idea why the sparse can outperform the full GP? Could it be a case of overfitting when noise is present? Such that the sparse one is less sensitive to overfitting?

Lines 335-339: How are data resolution handled? If there is 5 years resolution, then are there gaps between the years or are the values interpolated? Or are the annual data used and only clipped at the end of the record?

Line 387: I would remind the reader for the discussion what AMV-relevant timescales are. Maybe write in parenthesis something like (decadal to multi-decadal).

Figure 7b,e: Is the higher variability of the sparse emGPR on sub-decadal timescales really

indicative of an improved reconstruction or is it just added noise? A suggestion to investigate this would be to calculate the coherency from the cross-spectra of the reconstruction and original signal. This could be an interesting (although not necessary) addition for all comparisons.

Lines 405-410: Would it be possible to efficiently run the hyperparameter estimation several times with different one tenth (or maybe even less) of points as inducing points to obtain an ensemble of hyperparameter estimates? I am just wondering if this could be a more computing efficient way to estimate the parameters using the whole data and if the mean of the ensemble would converge to the full emGPR hyperparameters.

Line 461: Remind the reader on which timescales the sparse emGPR has strong variance loss.

One issue I would like better discussed is the loss of variance on longer than centennial timescales in the sparse emGPR for the full 2k run. To me this is quite an important limitation since I don't think it makes sense to restrict AMV-relevant timescales to decadal to multi-decadal; there is a continuum of processes and I don't think there are reasons to believe that it would flatten out on longer than centennial timescales or be related to a separate non-AMV relevant process right?

Line 465-466: Looking forward to seeing how it compares to the traditional PCR method!