Authors' Response to Reviews of

Bias Correction of Climate Models using a Bayesian Hierarchical Model

J. Carter
Geoscientific Model Development,

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

Thank you for the further review of the revised manuscript. We are very grateful for your time and have worked hard to incorporate the additional feedback into the revised manuscript. The response is detailed here and we hope that you find the further enhancements to the paper satisfactory. A revised manuscript is provided, along with a LaTex-diff document highlighting changes.

1. Referee 2 Further Comments

RC:

1. I appreciate that the proposed method would potentially help correct the bias of climate model outputs by preserving the covariance structure. Extensive experimentation on synthetic data under different distributional assumptions has also helped me better understand how the framework improves over a single-process model. However, it has not been convincingly argued that the synthetic experiments represent real-world scenarios or that the synthetic data experiments reveal how the model may perform on real climate model data. It is still unclear if the proposed methodology would be useful in real-world scenarios.

AR:

We thank the reviewer for their constructive comment and have edited the manuscript accordingly to further highlight the applicability of the methodology to real-world data sets. To start, additional writing (highlighted in blue below) is included in Sect. 3.1 of the manuscript further describing how the simulated scenarios are related to different potential real-world cases, using the study of Lima et al., 2021 as an example.

Relevant text from data generation section (added text highlighted in blue):

> Data is generated for three scenarios chosen to represent different potential real-world situations, illustrated in Fig. 3. The first scenario (Fig. 3a) represents an example case where it is expected that there is ample data provided in the form of in situ observations to capture the features of the underlying complete realisation of ϕ_Y without significant added value provided from inclusion of the climate model output during inference. The second scenario (Fig. 3b) is an adjustment where the in situ observations are relatively sparse and the underlying bias is relatively smooth. In this situation the climate model output should provide significant added value in estimating ϕ_Y across the domain since it is only afflicted by a comparatively simple bias that is easy to estimate. The final scenario (Fig. 3c) also involves sparse in situ observational data but with a reduced smoothness of the bias compared to the other scenarios. In this scenario the climate model output should provide added value in estimating ϕ_Y across the domain but this will be limited compared to scenario two due to the difficulty of disaggregating the components and estimating the comparatively more complex bias.
>
> In practice, real-world datasets are likely to be a combination of these scenarios. For example, the methodology in Lima et al., 2021 is applied to bias correcting precipitation over a domain covering South Korea and the surrounding ocean. Over the land, there is a sufficient spatial density of observational rainfall gauges to adequately capture the spatial features

of the unbiased underlying field from the observations alone (similar to scenario A). Over the ocean, rainfall gauges are very sparse and so its important to consider the spatial patterns observed from the climate model output (similar to scenario B). Not accounting for the spatial features seen in the climate model output over the ocean results in undesirable extrapolation over this region, as seen in the results presented in Lima et al., 2021. This undesirable property is something that is addressed by the methodology proposed in this paper, as illustrated by results for scenario B given in Sect. 4.1.

In addition to this, further writing has been included in the introduction detailing the flexibility and applicability of using GPs for modelling real-world datasets. Reference is made to papers from relevant text in the literature, including Zhang et al., 2021, Lima et al., 2021 and Wang and Chaib-draa, 2017.

Relevant added text to introduction:

While simple simulated scenarios are focused on in this paper, the applicability of GPs for modelling complex spatial patterns seen in real-world climatology is already illustrated in Zhang et al., 2021 and Lima et al., 2021. The non-parametric nature of GPs makes the model flexible and able to capture complex non-linear spatial relationships. Additionally, features of GPs such as uncertainty estimation, sensible extrapolation, kernel customisation and the ability to produce accurate predictions with limited data are desirable for real-world case studies. Finally, advancements in approximate inference methods have improved the scalability of GPs, improving the applicability to large climate data sets, as demonstrated in Wang and Chaib-draa, 2017. In addition to the main results presented in Sect. 4, to further demonstrate the flexibility and applicability of the methodology presented in this paper to potential real-world scenarios, some additional simulated scenarios are created with added complexity and results presented in appendix G. These additional scenarios test the robustness of the model to potential real-world situations where not all the assumptions of the model will necessarily completely hold.

Finally, a whole new section of the appendix is created to illustrate the applicability of the methodology under additional scenarios with added complexity and where some of the assumptions of the model are partially broken, as might be the case in real-world scenarios. The results from this section show the model proposed is robust and behaves desirably with potentially challenging features of real-world datasets. Despite this, importance is placed on how the purpose of this paper is not to provide a final fixed model, instead aiming to provide a flexible Bayesian framework where additional complexities present in real-world applications can be assessed on a case-by-case basis and further model adjustments made where needed to account for specific features of the real-world dataset.

Additional Sect. G of the appendix:

Real-world scenarios are expected to have more complex spatial features than the simulated examples presented in Sect. 4.1, with some of the assumptions of the model expected to be partially broken, such as stationarity and independence of the latent processes. To explore the performance of the methodology under scenarios with more complex spatial features, as in real-world problems, results for several additional simulated examples (A-D) are presented in Fig. G1. The hyper-parameter values used to generate the data are presented in Table G1, while the summary statistics for the posterior distributions after fitting the model proposed in this paper are presented in Table G2.

Scenario A represents a potential real-world scenario where the covariance length scale changes across the domain. This could be due to topographic features and a shift from relatively smooth topography to sharp mountainous terrain. For this scenario, the length scale of the latent unbiased process changes abruptly at $x = 50$ with a length scale of 5 for $x < 50$ and a length scale of 1 for $x > 50$. The length scale of the biased process is left constant across the domain in scenario A, although in scenario B it is made to also change abruptly at $x = 50$ to show a case where the latent spatial structure of the bias is also dependent on an extra factor such as topography.

Scenario C represents the potential real-world scenario where there are multiple sources of variation in the climate with

different covariance length scales. An example of this could be the combined influence of large-scale upper-atmosphere circulation patterns and small-scale topographic changes over the domain. The data is generated from the unbiased process after defining the kernel as the sum of two independent components, one with a variance of 1 and length scale of 3 and the other with a variance of 0.2 and length scale of 0.6. Finally, scenario D represents a potential real-world case where the bias in the parameter of study is dependent on the parameter value itself, as might be the case if for example the output from temperature sensors were skewed by over-heating. This correlation is induced between the bias and the unbiased process by generating the data for the bias as the sum of $\phi_B(s) = 0.2 * \phi_Y(s) + \phi_{B_{ind.}}(s)$, where $\phi_{B_{ind.}}(s)$ is an independent bias as generated in the other examples.

The result of fitting the model presented in this paper to each scenario is displayed in Table G2 and Fig. G1. From Table G2 it is clear that in cases where multiple length scales are used in generating the data, the expected value of the assumed single length scale is in-between the true values tending more to the smallest length scale. The reason the expectation of the single length scale tends towards the shorter values present in generating the data is hypothesised to be the result of more spatial features (peaks and troughs) being present for the shorter length scale component. The model is better able to explain the data observed with a length scale closer to the shortest value present and the 95% credible interval for the single length scale does not necessarily cover the multiple values used in generating the data.

In Fig. G1 it can be seen that, despite the additional complexities, the predictions on the unbiased parameter and on the bias are reasonable and capture the main spatial patterns. This demonstrates the flexibility of GPs and the robustness of the methodology proposed to fit different types of real-world data where some of the assumptions made in the model partially don't hold. Some features of the results due to not fully capturing the dependencies involved in generating the data are described here. In scenario A the length scale of the unbiased process is estimated close to the value used in generating the data for $x > 50$, which results in greater uncertainty than expected between nearby observations in the region $x < 50$ where the length scale is greater. For example, in the extrapolation range of $x < 0$ the prediction in the unbiased parameter values returns sharply to the mean and with uncertainty independent of observed points, whereas if the length scale was correctly estimated in this region the predictions would remain dependent on the data observed at $x = 0 - 10$ for longer. The same is true in scenario B with the addition of the estimates of the bias being effected, making disaggregating the climate model output into an unbiased and biased component more challenging, as seen at $x = 30$. In scenario C again by only modelling a single length scale for the unbiased process, disaggregating the climate model output into its two components is effected and the longer length scale peak present at $x = 20$ is attributed to the bias incorrectly. Finally in scenario D, not accounting for the correlation between the unbiased values and the bias results in a slightly greater uncertainty in predictions than could be achieved by incorporating this relationship.

Overall the model is shown to perform adequately and not be over-sensitive to some of the assumptions being partially broken, which supports the usefulness of the methodology to real-world applications. In addition, other methodologies currently applied to bias correction are likely more affected in these complex scenarios. It is noted that the purpose of this paper is not to provide a final fixed model however, instead aiming to provide a framework where additional complexities present in real-world applications can be assessed on a case-by-case basis and further model adjustments made where needed to account for specific features of the real-world dataset. The model could be modified for each scenario to take into account the extra complexity, something that a fixed-type model for bias correction would not be able to handle.

Dependent Variable	Model Parameters	Scenario A	Scenario B	Scenario C	Scenario D
Unbiased PDF Parameter ϕ_Y	Kernel Variance (v_{ϕ_Y})	1.0	1.0	1.0 & 0.2	1.0
	Kernel Lengthscale (l_{ϕ_Y})	5.0 (x<50), 1.0 (x>50)	5.0 (x<50), 1.0 (x>50)	3.0 & 0.6	3.0
	Mean Constant (m_{ϕ_Y})	1.0	1.0	1.0	1.0
	Noise (σ_{ϕ_Y})	0.1	0.1	0.1	0.1
	# Observations	40.0	40.0	40.0	40.0
Bias PDF Parameter ϕ_B	Kernel Variance (v_{ϕ_B})	1.0	1.0	1.0	1.0
	Kernel Lengthscale (l_{ϕ_B})	10.0	10.0 (x<50), 2.0 (x>50)	10.0	10.0
	Mean Constant (m_{ϕ_B})	-1.0	-1.0	-1.0	-1.0
Climate Model PDF Parameter ϕ_Z	# Observations	80.0	80.0	80.0	80.0

Table G1. A table showing the hyper-parameters of the latent Gaussian processes used to generate the complete underlying realisations of $\phi_Y(\boldsymbol{s}^\star)$, $\phi_B(\boldsymbol{s}^\star)$ and hence $\phi_Z(\boldsymbol{s}^\star)$, as well as observations of $\phi_Y(\boldsymbol{s}_y)$ and $\phi_Z(\boldsymbol{s}_z)$, on which inference is done for the additional scenarios. The number of observations representing in-situ data and climate model output is also given.

(a) Scenario A and B

		Scenario A Posterior Dist.				Scenario B Posterior Dist.			
Dependent Variable	Model Parameter	Exp.	Std. Dev.	C.I. L.	C.I. U.	Exp.	Std. Dev.	C.I. L.	C.I. U.
Unbiased PDF Parameter ϕ_Y	Kernel Variance v_{ϕ_Y}	0.87	0.17	0.55	1.18	0.85	0.18	0.53	1.20
	Kernel Lengthscale l_{ϕ_Y}	1.06	0.06	0.94	1.19	1.12	0.07	0.99	1.24
	Mean Constant m_{ϕ_Y}	0.78	0.15	0.48	1.07	1.42	0.17	1.09	1.77
	Noise σ_{ϕ_Y}	0.12	0.03	0.06	0.17	0.15	0.05	0.05	0.24
Bias PDF Parameter ϕ_B	Kernel Variance v_{ϕ_B}	1.15	0.86	0.19	2.87	0.85	0.39	0.32	1.60
	Kernel Lengthscale l_{ϕ_B}	10.34	1.93	6.86	14.32	3.50	0.82	2.21	5.05
	Mean Constant m_{ϕ_B}	-0.68	0.49	-1.59	0.34	-0.79	0.27	-1.33	-0.25

(b) Scenario C and D

		Scenario C Posterior Dist.				Scenario D Posterior Dist.			
Dependent Variable	Model Parameter	Exp.	Std. Dev.	C.I. L.	C.I. U.	Exp.	Std. Dev.	C.I. L.	C.I. U.
Unbiased PDF Parameter ϕ_Y	Kernel Variance v_{ϕ_Y}	0.52	0.10	0.34	0.72	0.88	0.23	0.49	1.33
	Kernel Lengthscale l_{ϕ_Y}	0.75	0.08	0.61	0.92	2.95	0.06	2.82	3.07
	Mean Constant m_{ϕ_Y}	1.03	0.11	0.81	1.25	0.93	0.26	0.42	1.44
	Noise σ_{ϕ_Y}	0.18	0.06	0.09	0.30	0.10	0.02	0.07	0.13
Bias PDF Parameter ϕ_B	Kernel Variance v_{ϕ_B}	0.57	0.44	0.11	1.38	0.42	0.22	0.14	0.84
	Kernel Lengthscale l_{ϕ_B}	6.71	2.11	3.19	10.84	4.31	0.53	3.26	5.33
	Mean Constant m_{ϕ_B}	-0.12	0.30	-0.73	0.44	-0.02	0.21	-0.46	0.38

Table G2. Tables showing summary statistics for the posterior distributions including the expectation (Exp.), standard deviation (Std. Dev.) and lower and upper bounds for the 95% credible interval (C.I. L. and C.I. U.). The prior distributions are the same non-informative distributions given in Table 3.

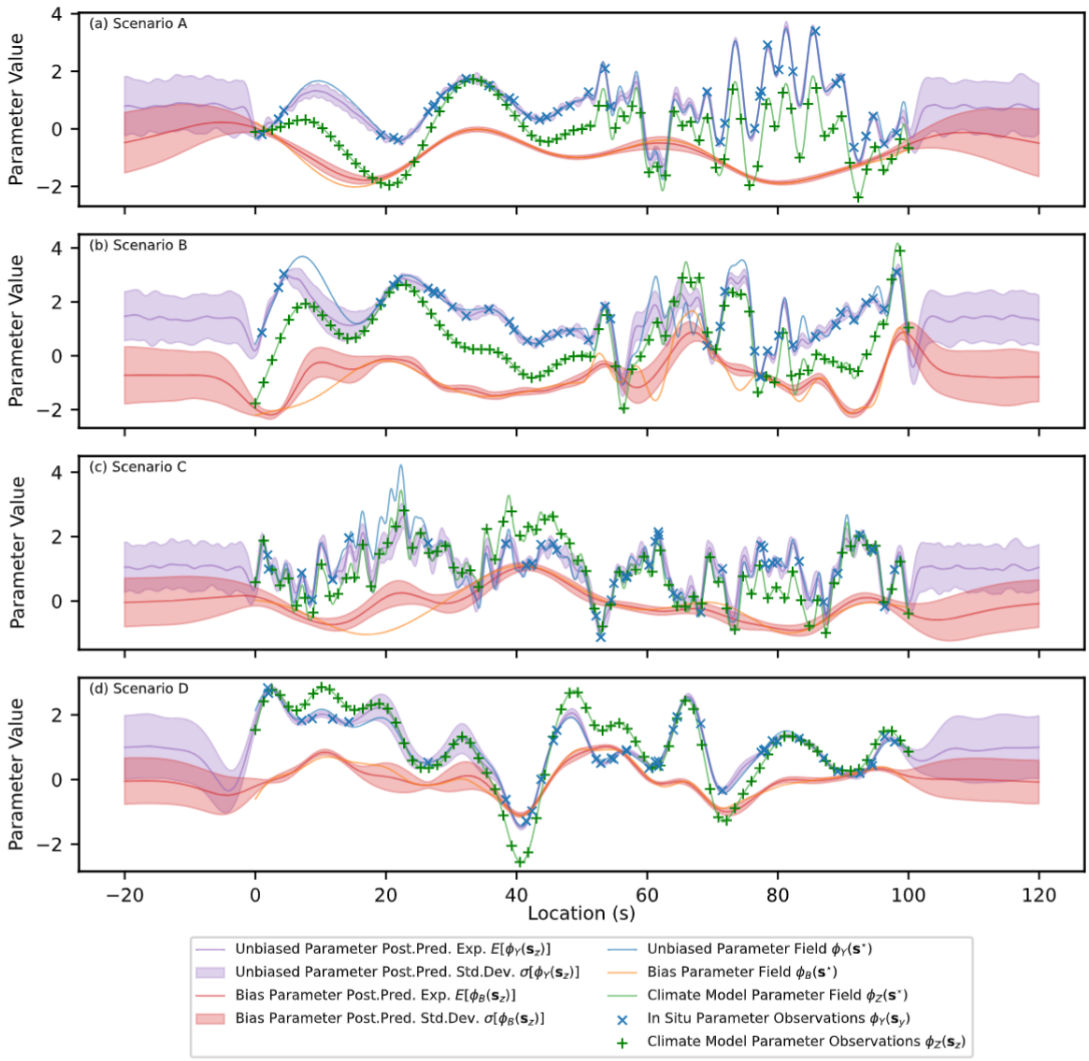

Figure G1. Expectation and 1σ uncertainty of the posterior predictive distributions of the parameter $\phi_Y(\boldsymbol{s}_z)$ and the corresponding bias $\phi_B(\boldsymbol{s}_z)$ for three scenarios. The underlying functions (complete realisations) as well as the simulated input data are also shown.

References

Lima, Carlos H. R., Hyun-Han Kwon, and Yong-Tak Kim (June 1, 2021). "A Bayesian Kriging model applied for spatial downscaling of daily rainfall from GCMs". In: *Journal of Hydrology* 597, p. 126095. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2021.126095. URL: https://www.sciencedirect.com/science/article/pii/S0022169421001426 (visited on 08/16/2021).

Wang, Yali and Brahim Chaib-draa (Jan. 1, 2017). "An online Bayesian filtering framework for Gaussian process regression: Application to global surface temperature analysis". In: *Expert Systems with Applications* 67, pp. 285–295. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2016.09.018. URL: https://www.sciencedirect.com/science/article/pii/S095741741630495X (visited on 05/14/2024).

Zhang, Yongshun, Miao Feng, Weimin Zhang, Huizan Wang, and Pinqiang Wang (July 1, 2021). "A Gaussian process regression-based sea surface temperature interpolation algorithm". In: *Journal of Oceanology and Limnology* 39.4, pp. 1211–1221. ISSN: 2523-3521. DOI: 10.1007/s00343-020-0062-1. URL: https://doi.org/10.1007/s00343-020-0062-1 (visited on 05/14/2024).