Bayesian Hierarchical Model for Bias Correcting Climate Models

Jeremy Carter[1,2], Erick A. Chacón-Montalván[4], and Amber Leeson[2,3]

[1]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK
[2]Data Science Institute, University of Lancaster, Lancaster, UK
[3]Lancaster Environment Center, University of Lancaster, Lancaster, UK
[4]Escuela Profesional de Ingeniería Estadística, Universidad Nacional de Ingeniería, Lima, Peru

Correspondence: Jeremy Carter (j.carter10@lancaster.ac.uk)

Abstract. Climate models, derived from process understanding, are essential tools in the study of climate change and its wide-ranging impacts. Hindcast and future simulations provide comprehensive spatiotemporal estimates of climatology that are frequently employed within the environmental sciences community, although the output can be afflicted with bias that impedes direct interpretation. Post-processing, bias correction approaches utilise observational data to address this challenge, although are typically criticised for not being physically justified and not considering uncertainty in the correction. This paper proposes a novel Bayesian bias correction framework that propagates uncertainty robustly and models underlying spatial covariance patterns. Shared latent Gaussian processes are assumed between the in situ observations and climate model output with the aim of partially preserving the covariance structure from the climate model after bias correction, which is based on well-established physical laws. Results demonstrate added value in modelling shared generating processes under several simulated scenarios, with most value added for the case of sparse in situ observations and smooth underlying bias. Additionally, the propagation of uncertainty to a simulated final bias corrected time series is illustrated, which is of key importance to a range of stakeholders, from climate scientists engaged in impact studies, decision makers trying to understand the likelihood of particular scenarios and individuals involved in climate change adaption strategies where accurate risk assessment is required for optimal resource allocation. This paper focuses on one-dimensional simulated examples for clarity, although the code implementation is developed to also work on multi-dimensional input data, encouraging follow-on real-world application studies that will further validate performance and remaining limitations. The Bayesian framework supports uncertainty propagation under model adaptations required for specific applications, providing a flexible approach that increases the scope to data assimilation tasks more generally.

1 Introduction

20 Climate models are invaluable in the study of climate change and its impacts (Bader et al., 2008; Flato et al., 2013). Formulated from physical laws and with parameterisation and process understanding derived from past observations; climate models provide comprehensive spatiotemporal estimates of our past, current and future climate under different emission scenarios. Global climate models (GCMs) simulate important climatological features and drivers such as storm tracks and the El Niño–Southern Oscillation (ENSO) (Greeves et al., 2007; Guilyardi et al., 2009). In addition, independently developed GCMs agree on the

25 future direction of travel for many important features such as global temperature rise under continued net-positive emission scenarios (Tebaldi et al., 2021). However, GCMs are computationally expensive to run and also exhibit significant systematic errors, particularly on regional scales (Cattiaux et al., 2013; Flato et al., 2013). Regional climate models (RCMs) aim to dynamically downscale GCMs and more accurately represent climatology for specific regions of interest and have parameterisation, tuning and additional physical schemes optimised to the region (Giorgi, 2019; Doblas-Reyes et al., 2021). Despite this,

30 significant systematic errors remain, particularly for regions with complex climatology and with sparse in situ observations available to inform process understanding, such as over Antarctica (Carter et al., 2022). These systematic errors inhibit the direct interpretation of climate model output, particularly important in impact assessments (Ehret et al., 2012; Liu et al., 2014; Sippel et al., 2016).

There are many fundamental causes of systematic errors in climate models, including: the absence or imperfect represen-
35 tation of physical processes; errors in initialisation; influence of boundary conditions and finite resolution (Giorgi, 2019). The inherent complexity and computationally expensive nature of climate models makes direct reduction of systematic errors through climate model development and tuning challenging (Hourdin et al., 2017). Additionally, end users are typically interested in only a narrow aspect of the output (e.g. possibly only one or two variables), which the climate model is unlikely to be specifically tuned for. Post-processing, bias correction techniques allow improvements to the consistency, quality and value of
40 climate model output, specific to the end user's focus of interest, with manageable computational cost and without requirement of in-depth knowledge behind the climate model itself (Ehret et al., 2012). Transfer functions are derived between the climate model output and in situ observational data to correct components such as the mean (Das et al., 2022) or probability density functions (PDFs) of the data (Qian and Chang, 2021). This paper focuses on providing a novel framework for correcting systematic errors in the PDF of the climate model output at each grid point.

45 One of the fundamental issues often attached to post-processing bias correction is the lack of justification based on known physical laws and process understanding (Ehret et al., 2012; Maraun, 2016). The spatiotemporal field and associated covariance structure from the climate model, which is consistent with accepted physical laws, is typically not considered and so not preserved. Resulting corrected fields may exhibit too smooth or sharply varying behaviour over the region and discontinuities near observations. In addition, many approaches of bias correction fail to adequately handle uncertainties or estimate them
50 at all. Reliable uncertainty estimation is valuable for inclusion in impact studies to inform resulting decision making. This is especially true for regimes with tipping points, such as ice shelf collapse over Antarctica, where uncertainties in the climatology can cause a regime shift (DeConto and Pollard, 2016). In this paper these issues are partially addressed through the

development of a fully Bayesian hierarchical approach to bias correction. Parameter uncertainties are propagated through the hierarchical model and underlying spatial covariance structures are captured with latent Gaussian processes (GPs) for both in situ observations and the climate model output.

The approach presented builds on that of Lima et al. (2021), which models the in situ observational data as generated from a GP and uses quantile mapping (Qian and Chang, 2021) to apply the correction to the climate model output. In Lima et al. (2021) the spatial covariance structure of the climate model output is not considered and uncertainty is not propagated to the final bias corrected time series. The novelty of the approach proposed here is that shared latent GPs are modelled between the climate model output and the in situ observational data, which aims to incorporate information from the physically realistic spatial patterns of the climate model output in predictions of the unbiased field. Additionally, uncertainty is propagated through the quantile mapping step, which results in uncertainty bands on the bias corrected output. The approach is developed with the focus of applying bias correction to regions with sparse in situ observations, such as over Antarctica, where capturing uncertainty in the correction is of key importance and where including data from all sources during inference is particularly valuable. Performance under simulated scenarios with differing data density and underlying covariance length scales is evaluated in this paper and the potential added value assessed when compared with the approach in Lima et al. (2021).

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.

Developing the bias correction approach in a flexible Bayesian framework means further adjustments/advancements that are necessary for real-world scenarios can easily be incorporated while maintaining robust uncertainty propagation. For example, extra predictors, such as elevation and latitude, can be included either in the mean function or covariance matrix of the latent GPs. Alternatively, the model could be expanded to incorporate a temporal component of the bias accounting for variability across different seasons. This flexibility is important and increases the scope of the work, allowing the methodology to be applied to a wide range of scenarios, including for example application to many different meteorological fields and also combining observation data from different instruments rather than necessarily with respect to climate model output. Additionally, the Bayesian framework allows incorporation of domain specific, expert knowledge of different data sources and their uncertainties through the choice of prior distributions.

2 Methodology

In a probabilistic framework, the in situ observations and climate model output are treated as realisations from latent spatiotemporal stochastic processes, denoted as $\{Y(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ and $\{Z(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ respectively. Stochastic processes are sequences of random variables indexed by a set, which in this case are the spatial and temporal coordinates in the domain $(\mathcal{S}, \mathcal{T})$. The observed data is then considered a realisation of the joint distribution over a finite set of random variables across the domain. For the purpose of evaluating the time-independent component of the climate model bias, the random variables are treated as independent and identically distributed across time. That is the collection of temporal data for a given spatial location can be considered as multiple realisations from the same random variable. The random variables for each location are distributed respectively as $Y(s) \sim f_Y(\phi_Y(s))$ and $Z(s) \sim f_Y(\phi_Z(s))$, where $\phi_Y(s)$ and $\phi_Z(s)$ represent the collection of parameters that describe the PDF. For example, if the PDF is approximated as normal then $\phi(s) = [\mu(s), \sigma(s)]$. The disparity between each of the PDF parameters for the in situ observations and climate model at each site then gives a measure of bias. The goal is to estimate the parameters $\phi_Y(s)$ and $\phi_Z(s)$ at the climate model grid points to quantify the bias and to apply quantile mapping to bias correct the climate model output. Gaussian processes are used to model the underlying spatial covariance structure of the parameters, which is required to estimate $\phi_Y(s)$ away from the location of the in situ observations. Further discussion around the definition of bias in climate models is provided in appendix A.

Consider a collection of n_Y in situ observational sites, where for each site i there exists m_i measurements of some property. In addition, consider gridded output from a climate model at n_z locations, where at each location there exists m_z measurements of the same property. The data can then be represented through the following:

$$\boldsymbol{y} = [\boldsymbol{y}_{s_1}, \ldots, \boldsymbol{y}_{s_{n_y}}] \tag{1}$$

$$\boldsymbol{y}_{s_i} = [y_{s_i,1}, \ldots, y_{s_i,m_i}] \tag{2}$$

$$\boldsymbol{z} = [\boldsymbol{z}_{s'_1}, \ldots, \boldsymbol{z}_{s'_{n_z}}] \tag{3}$$

$$\boldsymbol{z}_{s'_i} = [z_{s'_i,1}, \ldots, z_{s'_i,m_z}] \tag{4}$$

Defining the collection of in situ observation sites as $\boldsymbol{s}_y = [s_1, \ldots, s_{n_y}]$ and the collection of climate model output locations as $\boldsymbol{s}_z = [s'_1, \ldots, s'_{n_z}]$, then the collection of PDF parameter values for each set of locations is written as:

$$\boldsymbol{\phi}_Y(\boldsymbol{s}_y) = [\boldsymbol{\phi}_Y(s_1), \ldots, \boldsymbol{\phi}_Y(s_{n_y})] \tag{5}$$

$$\boldsymbol{\phi}_Z(\boldsymbol{s}_z) = [\boldsymbol{\phi}_Z(s'_1), \ldots, \boldsymbol{\phi}_Z(s'_{n_z})] \tag{6}$$

The PDF parameters are each modelled as being generated from latent stochastic processes $\{\phi_Y(s)\}$ and $\{\phi_Z(s)\}$. The latent processes that generate the parameters for climate model are considered composed of two independent processes, one that also generates the equivalent parameters for the in situ observations and another that generates some bias, such that $\{\phi_Z(s)\} = \{\phi_Y(s)\} + \{\phi_B(s)\}$. The family of GPs are chosen for the latent processes. A link function is used for the case

where the parameter space is not the same as the sample space for GPs. Considering the case of no link function, the following can then be written:

$$\phi_Y(s) \sim \mathcal{GP}(\cdot, \cdot | \boldsymbol{\theta}_{\phi_Y}) \tag{7}$$

$$\phi_B(s) \sim \mathcal{GP}(\cdot, \cdot | \boldsymbol{\theta}_{\phi_B}) \tag{8}$$

$$\phi_Z(s) \sim \mathcal{GP}(\cdot, \cdot | \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \tag{9}$$

120

The collection of hyper-parameters for the generating processes are given by $\boldsymbol{\theta}_{\phi_Y}$ and $\boldsymbol{\theta}_{\phi_B}$ respectively. The hyper-parameters used in this paper consist of a mean constant, kernel variance and kernel length scale. Note the additive property of GPs allows $\phi_Z(s)$ to also be represented by a GP, where the mean and covariances are computed from the sum of the relative values from the independent processes. Further discussion around the properties of GPs is provided in appendix B. The hierarchical model is illustrated through the plate diagram shown in Fig. 1. In addition, a specific example where the PDFs are approximated as normal is presented in appendix D.
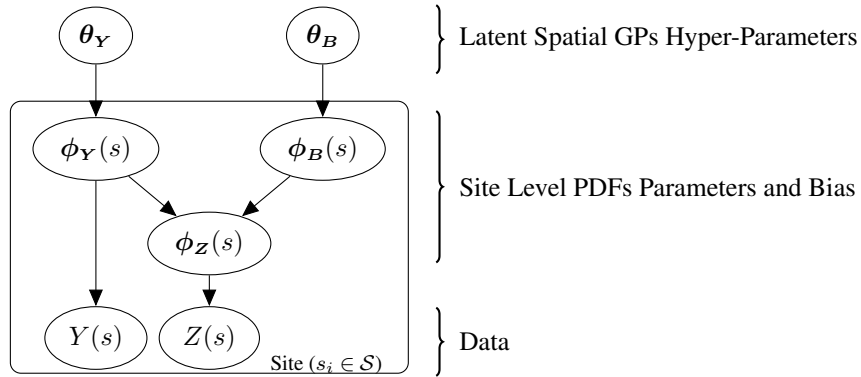
125

Figure 1. Plate diagram illustrating the full hierarchical model. The random variables for the in-situ observations $Y(s)$ and climate model output $Z(s)$ have PDFs with the collection of parameters $\phi_Y(s)$ and $\phi_Z(s)$ respectively, where $\phi_Z(s)$ is modelled as the sum of $\phi_Y(s)$ and some independent bias $\phi_B(s)$. The parameters $\phi_Y(s)$ and the corresponding bias $\phi_B(s)$ are each themselves modelled over the domain as generated from Gaussian processes with hyper-parameters $\boldsymbol{\theta}_Y$ and $\boldsymbol{\theta}_B$.

Inference on the parameters of the hierarchical model given the data is applied in a Bayesian framework, where parameters of the model are themselves treated as random variables with distributions. The distribution prior to conditioning on any data is known as the prior distribution and allows the incorporation of a domain specific expert's knowledge in the estimates of the parameters. The updated distribution after conditioning on the observed data is known as the posterior and is approximated using Markov chain Monte Carlo (MCMC) methods, which provide samples of the parameters from the distribution $P(\phi_Y(\boldsymbol{s}_y), \phi_Z(\boldsymbol{s}_z), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B} | \boldsymbol{y}, \boldsymbol{z})$. Estimates of the parameters ϕ_Y and ϕ_Z at any set of new locations $\hat{s}$ can then be made by constructing the posterior predictive distribution, in particular for the purpose of bias correction estimates of ϕ_Y at the climate model locations can be made by sampling from the posterior predictive distribution of $P(\phi_Y(\boldsymbol{s}_z) | \boldsymbol{y}, \boldsymbol{z})$.

130

5

After obtaining multiple realisations of $\phi_Y(s_z)$ and $\phi_Z(s_z)$ quantile mapping is then used to bias correct the climate model time series at every grid cell location. Specifically, for each value of the time series from the climate model output at a given point $(z_{s'_i,j})$, this involves finding the percentile of that value using the parameters $\phi_Z(s'_i)$ and then mapping the value onto the corresponding value of the equivalent percentile of the PDF estimated for the unbiased process, defined through the parameters $\phi_Y(s'_i)$. The cumulative density function (CDF) returns the percentile of a given value and the inverse CDF returns the value corresponding to a given percentile, which results in the following correction function $\hat{z}_{s_i,j} = F_{Y_{s_i}}^{-1}(F_{Z_{s_i}}(z_{s_i,j}))$, where F represents the CDF at a specific site. The CDF can be estimated as an integral over the parametric form assumed for the PDF. The Bayesian hierarchical model presented provides a collection of realisations for $\phi_Y(s_z)$ and $\phi_Z(s_z)$ from an underlying latent distribution. Applying quantile mapping with each set of realisations then results in a collection of bias corrected time series, with an expectation and uncertainty. The full framework for bias correction proposed in this paper is then illustrated in Fig. 2. The formulation for the posterior and posterior predictive is given in appendix C.
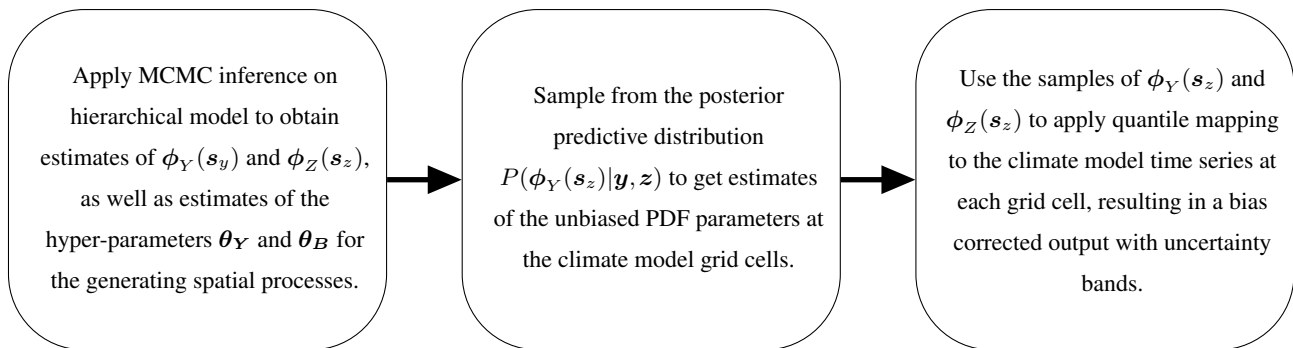

Figure 2. The full bias correction framework proposed in this paper broken down into the key steps.

3 Data Generation

The goal of the hierarchical model in Fig. 1 is primarily to estimate, with reliable uncertainties, the true unbiased values of the PDF parameters at each location of the climate model output so bias correction can be applied. Simulated examples are generated that highlight the advantage of two key features of the methodology over other approaches in the literature: modelling shared spatial covariance between the in situ data and climate model output through the inclusion of a shared generating latent process (Sect. 3.1) and the Bayesian hierarchical framework with uncertainty propagation (Sect. 3.2). One dimensional simulated examples are chosen for clarity in illustrating these features, although it is noted the implementation works for higher dimensional domains as is useful in real-world scenarios.

3.1 Non-Hierarchical Examples: Data Generation

To illustrate the potential advantage of modelling shared generating spatial processes, non-hierarchical examples are generated for simplicity. Direct measurements are assumed for one parameter of the PDFs for the in situ observations $\phi_Y(s_y)$ and for

the climate model output $\phi_Z(\boldsymbol{s}_z)$. The goal is to predict the unbiased parameter at the climate model locations $\phi_Y(\boldsymbol{s}_z)$ using information from both sets of input data $\phi_Y(\boldsymbol{s}_y)$ and $\phi_Z(\boldsymbol{s}_z)$, which are related through $\phi_Z(\boldsymbol{s}) = \phi_Y(\boldsymbol{s}) + \phi_B(\boldsymbol{s})$. Comparison is made to the approach of inferring $\phi_Y(\boldsymbol{s}_z)$ from the in situ data alone $\phi_Y(\boldsymbol{s}_y)$, as in Lima et al. (2021). Relative performance
160   is evaluated for three alternative simulated scenarios that correspond to different possible real-world situations. The data is generated assuming the model in Fig. 1, where the GPs are taken with constant mean and an radial basis function (RBF) kernel with constant kernel length scale and kernel variance. The specific values of the hyper-parameters used to generate the data and the number of observations under the different scenarios is given in Table 1.

For each scenario, a sample of the parameters $\phi_Y(\boldsymbol{s}^\star)$ and $\phi_B(\boldsymbol{s}^\star)$ is taken from the distributions $\mathcal{GP}_{\phi_Y}$ and $\mathcal{GP}_{\phi_B}$ at
165   regularly spaced, high-resolution intervals. These samples are referred to here as complete realisations and represent underlying fields for each parameter across the domain. The complete realisations of $\phi_Y(\boldsymbol{s}^\star)$ are sampled at lower-resolution, randomised locations, with the addition of some noise, to provide direct simulated 'in situ observations' of the parameter $\phi_Y(\boldsymbol{s}_y)$. In order to simulate input data for the parameter $\phi_Z(\boldsymbol{s}_z)$ of the climate model output, the complete realisations of $\phi_Y(\boldsymbol{s}^\star)$ and $\phi_B(\boldsymbol{s}^\star)$ are sampled at regularly spaced intervals to provide $\phi_Y(\boldsymbol{s}_z)$ and $\phi_B(\boldsymbol{s}_z)$, then the sum of these samples at each location is taken
170   to give $\phi_Z(\boldsymbol{s}_z)$. The input data for inference is then $\phi_Y(\boldsymbol{s}_y)$ and $\phi_Z(\boldsymbol{s}_z)$ and can be considered as the training set, while the underlying realisations generated for $\phi_Y(\boldsymbol{s}_z)$ are the test set used for validating the model performance.

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
175   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
180   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
185   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.

Dependent Variable	Model Parameters	Scenario 1	Scenario 2	Scenario 3
Unbiased PDF Parameter ϕ_Y	Kernel Variance (v_{ϕ_Y})	1.0	1.0	1.0
	Kernel Lengthscale (l_{ϕ_Y})	3.0	3.0	3.0
	Mean Constant (m_{ϕ_Y})	1.0	1.0	1.0
	Noise (σ_{ϕ_Y})	0.1	0.1	0.1
	# Observations	80.0	20.0	20.0
Bias PDF Parameter ϕ_B	Kernel Variance (v_{ϕ_B})	1.0	1.0	1.0
	Kernel Lengthscale (l_{ϕ_B})	10.0	20.0	5.0
	Mean Constant (m_{ϕ_B})	-1.0	-1.0	-1.0
Climate Model PDF Parameter ϕ_Z	# Observations	100.0	80.0	80.0

Table 1. A table showing the hyper-parameters of the two 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 three scenarios. The number of observations representing in-situ data and climate model output is also given.
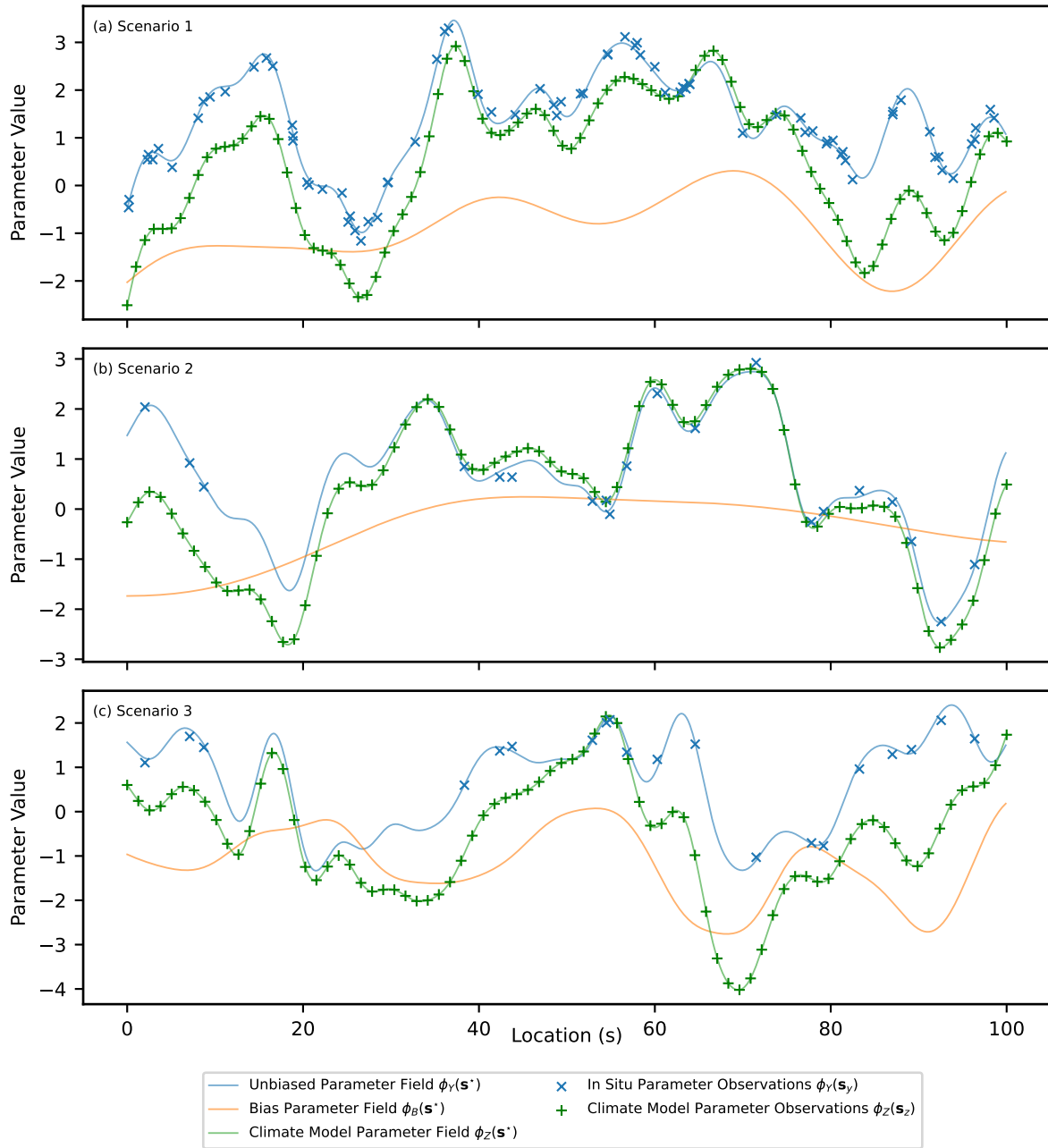
Figure 3. A figure showing simulated observed data for the PDF parameters $\phi_Y(\boldsymbol{s}_y)$ and $\phi_Z(\boldsymbol{s}_z)$, as well as the underlying complete realisations for each parameter and the underlying bias ($\phi_Y(\boldsymbol{s}^\star)$, $\phi_Z(\boldsymbol{s}^\star)$ and $\phi_B(\boldsymbol{s}^\star)$). Three scenarios are shown and correspond to data generated from parameters in Table 1.

3.2 Hierarchical Example: Data Generation

Following on from the non-hierarchical examples, to illustrate the advantage of uncertainty propagation in the Bayesian framework a hierarchical example is generated. In situ data and climate model output are simulated at each site as generated from normal distributions, such that $Y(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$ and $Z(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$ as in appendix D. The following relationship is assumed for the mean parameters $\mu_Z(s) = \mu_Y(s) + \mu_B(s)$, where $\mu_B(s)$ is the bias in the mean for the climate data. For the standard deviation, the parameters are first transformed using a logarithmic link function and then the relationship $\tilde{\sigma}_Z(s) = \tilde{\sigma}_Y(s) + \tilde{\sigma}_B(s)$ is assumed, where $\tilde{\sigma}_B(s)$ is the bias in the transformed parameter. The latent distributions that generate $\mu_Y(s)$, $\mu_B(s)$, $\tilde{\sigma}_Y(s)$ and $\tilde{\sigma}_B(s)$ across the domain are assumed as independent GPs with constant mean and an RBF kernel. The hyper-parameters for these latent generating processes are set for a single scenario, as given in Table 2 along with the number of simulated observation locations and the number of samples per location.

A sample of the parameters $\mu_Y(s^\star)$, $\mu_B(s^\star)$, $\tilde{\sigma}_Y(s^\star)$ and $\tilde{\sigma}_B(s^\star)$ is taken from the distributions $\mathcal{GP}_{\mu_Y}$, $\mathcal{GP}_{\mu_B}$, $\mathcal{GP}_{\tilde{\sigma}_Y}$ and $\mathcal{GP}_{\tilde{\sigma}_B}$ at regularly spaced, high resolution intervals. These samples are referred to as complete realisations and represent the underlying fields for each PDF parameter across the domain. The complete realisations of $\mu_Y(s^\star)$ and $\tilde{\sigma}_Y(s^\star)$ are sampled at a selection of lower-resolution, randomised locations that represent simulated in situ observation sites to give $\mu_Y(s_y)$ and $\tilde{\sigma}_Y(s_y)$. Multiple observations of $Y(s_i)$ are then generated at each in situ observation site by sampling from the corresponding normal distribution $\mathcal{N}(\mu_Y(s_i), \tilde{\sigma}_Y(s_i))$. In the case of the simulated climate model output, samples are first taken from $\mu_Y(s^\star)$, $\mu_B(s^\star)$, $\tilde{\sigma}_Y(s^\star)$ and $\tilde{\sigma}_B(s^\star)$ at regularly spaced intervals, then the sum of these samples at each location is taken to give $\mu_Z(s_z)$ and $\tilde{\sigma}_Z(s_z)$. The climate model output is then generated at each of these locations from the corresponding normal distribution $Z(s_i) \sim \mathcal{N}(\mu_Z(s_i), \tilde{\sigma}_Z(s_i))$.

In the generated example there are 40 locations corresponding to simulated in situ observation sites, where for each site 20 measurements are generated. There are 80 locations corresponding to simulated climate model output grid points and at each location 100 samples are generated. This reflects the typical scenario where the climate model output has greater spatiotemporal coverage than in situ observations but is afflicted with bias. In Fig. 4 examples of the generated samples of $Y(s_i)$ and $Z(s_i)$ are shown corresponding to the nearest sites for three locations. It is clear that, due to limited observations, there will be significant uncertainty in estimates of the mean and standard deviation parameters at each site and it's important this uncertainty is propagated to inference of the hyper-parameters for the latent GPs and also to estimates of the unbiased PDF parameters at the climate model locations ($\mu_Y(s_z)$ and $\tilde{\sigma}_Y(s_z)$). The underlying, complete realisations of the parameters $\mu_Y(s^\star)$, $\mu_Z(s^\star)$, $\sigma_Y(s^\star)$ and $\sigma_Z(s^\star)$, as well as the bias $\mu_B(s^\star)$ and $\sigma_B(s^\star)$, are shown in Fig. 5. In addition, the empirical mean value and standard deviation of the generated data is illustrated at the simulated in situ observation and climate model sites. The goal of the hierarchical model is then to predict the unbiased values for the parameters of the PDFs at the locations of the climate model output ($\mu_Y(s_z)$ and $\tilde{\sigma}_Y(s_z)$), while propagating uncertainty. An example of how the uncertainty in predictions of $\mu_Y(s_z)$ and $\tilde{\sigma}_Y(s_z)$ is propagated through quantile mapping is then provided.

10

Dependent Variable	Model Parameters	Hierarchical Scenario
Unbiased PDF Mean μ_Y	Kernel Variance (v_{μ_Y})	1.0
	Kernel Lengthscale (l_{μ_Y})	3.0
	Mean Constant (m_{μ_Y})	1.0
Unbiased PDF Transformed Variance $\tilde{\sigma}_Y^2$	Kernel Variance ($v_{\tilde{\sigma}_Y^2}$)	1.0
	Kernel Lengthscale ($l_{\tilde{\sigma}_Y^2}$)	3.0
	Mean Constant ($m_{\tilde{\sigma}_Y^2}$)	1.0
Bias PDF Mean μ_B	Kernel Variance (v_{μ_B})	1.0
	Kernel Lengthscale (l_{μ_B})	10.0
	Mean Constant (m_{μ_B})	-1.0
Bias PDF Transformed Variance $\tilde{\sigma}_B^2$	Kernel Variance ($v_{\tilde{\sigma}_B^2}$)	1.0
	Kernel Lengthscale ($l_{\tilde{\sigma}_B^2}$)	10.0
	Mean Constant ($m_{\tilde{\sigma}_B^2}$)	-1.0
Unbiased Output Y	# Observation Sites	40.0
	# Observations per Site	20.0
Climate Model Output Z	# Observation Sites	80.0
	# Observations per Site	100.0

Table 2. A table showing the hyper-parameters used to generate the complete underlying realisations and the measurement data on which inference is done for the hierarchical scenario. The number of sites where data is generated along with the number of samples for each site is also given.
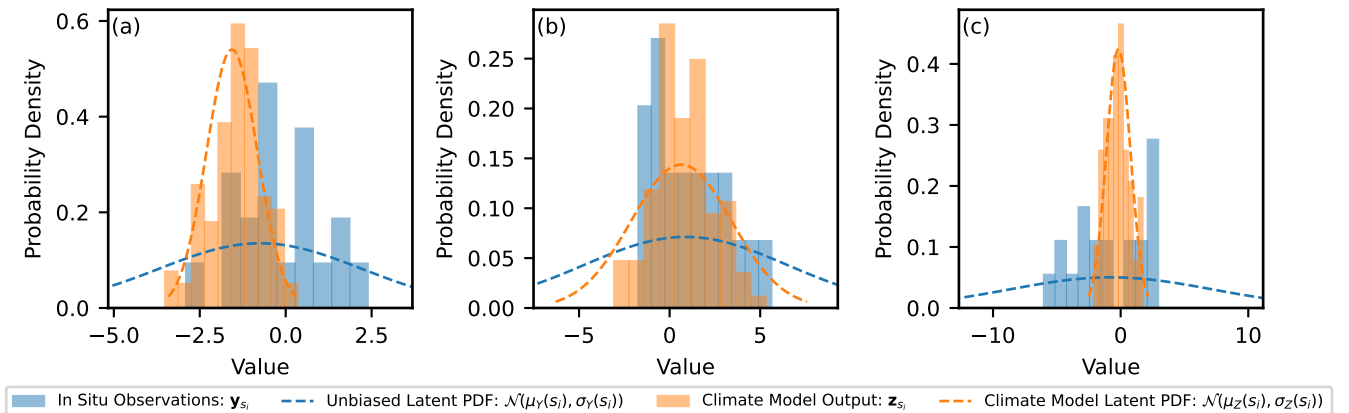

Figure 4. Histograms for the climate model output at three locations and the corresponding data from the nearest in situ observation site. The locations are a) s=11.4, b) s=46.8 and c) s=79.7. The latent normal distribution the data was generated from is illustrated as a dotted line.
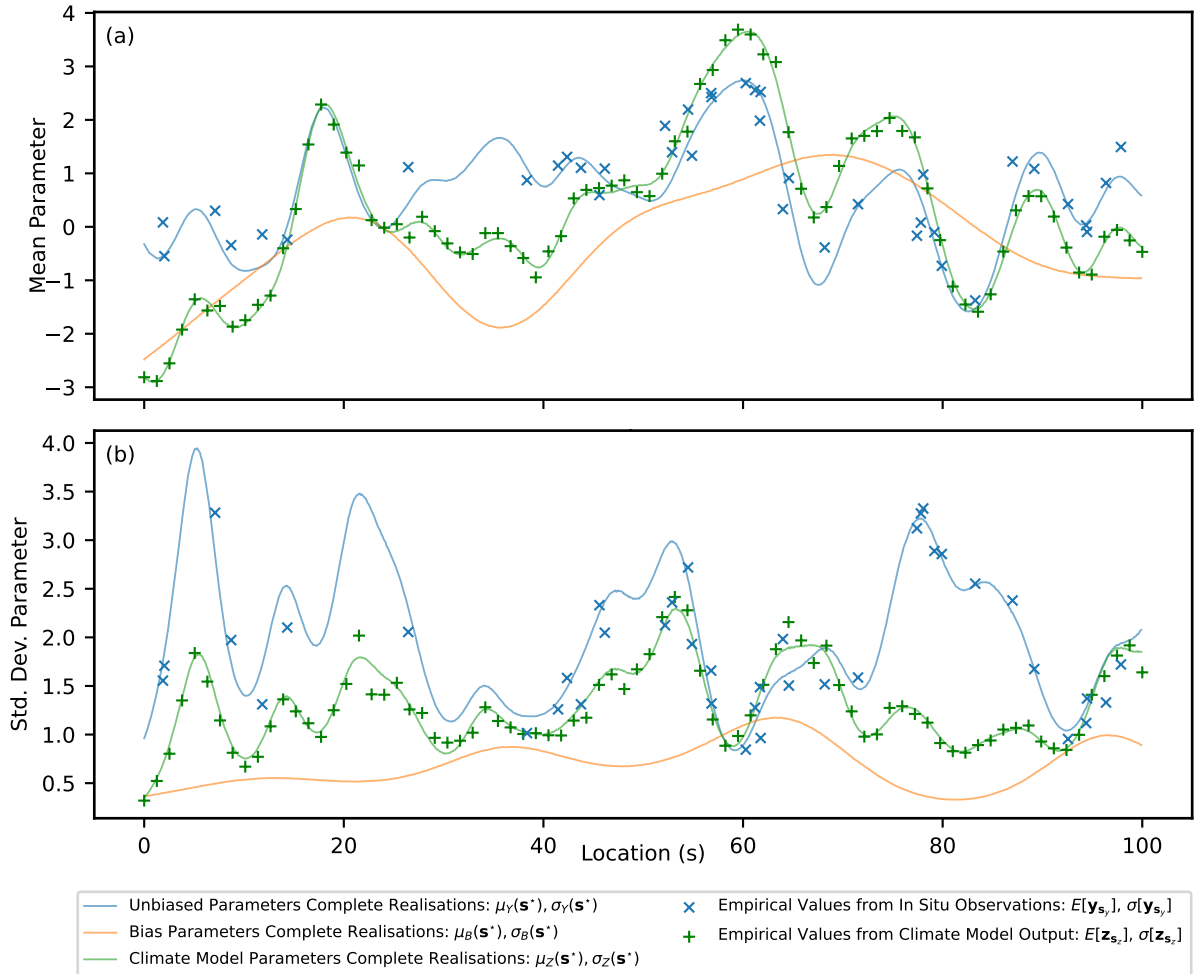
Figure 5. Simulated complete realisations for the parameters $\mu_Y(\boldsymbol{s}^\star)$, $\mu_B(\boldsymbol{s}^\star)$, $\mu_Z(\boldsymbol{s}^\star)$, $\tilde{\sigma}_Y(\boldsymbol{s}^\star)$, $\tilde{\sigma}_B(\boldsymbol{s}^\star)$ and $\tilde{\sigma}_Z(\boldsymbol{s}^\star)$ as well as the empirical values at the observation locations for the in situ and climate model data.

4 Results

Inference is done in a Bayesian framework using MCMC and the No-U-Turn Sampler (NUTS) algorithm (Hoffman and Gelman, 2014) implemented in Numpyro (Phan et al., 2019). For the MCMC sampling 1000 iterations were used for warm-up and then 2000 samples taken, which was found to be adequate for convergence. The parameters/hyper-parameters are treated as random variables with associated probability distributions. A prior distribution is set for each hyper-parameter and represents the belief on the distribution before observing any data, which typically incorporates knowledge from application specific experts. In the examples presented, relatively non-informative priors are chosen since the data is simulated and represents generic examples. The posterior distribution of each parameter/hyper-parameter is the updated distribution after observing and condi-

230 tioning on the data. Estimates of the PDF parameters $\phi_Y(\hat{s})$, $\phi_Z(\hat{s})$ and the corresponding bias $\phi_B(\hat{s})$ at new locations away from the observation sites, are then referred to as samples from the posterior predictive.

4.1 Non-Hierarchical Examples: Results

The shared latent process model presented in this paper is fit to the three non-hierarchical example scenarios, as discussed in Sect. 3.1. Input data for $\phi_Y(s_y)$ and $\phi_Z(s_z)$ are provided and the hyper-parameters for the latent GPs that generate the

235 unbiased and biased components inferred. Comparisons in estimates of the hyper-parameters for the unbiased process (m_{ϕ_Y}, v_{ϕ_Y} and l_{ϕ_Y}) are made to the approach of only fitting to the in situ data, referred to here as the single process approach since the latent process generating the bias is not modelled. The difference between the shared and single process approaches is detailed further in appendix E. The expectation, standard deviation and 95% credible intervals for the prior and posterior distributions of the hyper-parameters under the three different scenarios is given in Table 3. Illustrations of the prior and posterior distributions

240 for each hyper-parameter are plot in Fig. F1 of the appendix.

Under all scenarios the 95% credible interval of the posterior for every hyper-parameter bounds the value specified in generating the data. The expectation for the posterior distribution of the unbiased hyper-parameters is in general closer to the specified value in the shared process model compared with the single process model and the range of the credible interval is smaller. In scenario one and three the differences between the shared and single process models posteriors are relatively

245 insignificant for the mean constant (m_{ϕ_Y}) and kernel variance (v_{ϕ_Y}), although the shared process model shows a noticeable reduction in the uncertainty of the kernel length scale (l_{ϕ_Y}). In scenario two the difference is more significant and clear improvement is shown for the shared process model, both in the expectation and uncertainty of hyper-parameter estimates.

After applying MCMC inference on the parameters/hyper-parameters that generate the data, posterior predictive estimates are made for the unbiased PDF parameter values at the simulated locations of the climate model ($\phi_Y(s_z)$). These estimates

250 are presented in Fig. 6 for each scenario and for both the shared and single process models. Additionally, estimates of the underlying bias $\phi_B(s)$ are shown for the shared process model, since the bias is explicitly modelled. The relative performance of the shared and single process models is quantified by computing R^2 scores between the predictions of $\phi_Y(s_z)$ and the actual values used in generating the data (although not used in training), with results presented in Table 4. In Fig. 6 it can be seen that the predictions of $\phi_Y(s_z)$ in the shared process case (Fig. 6d, 6e and 6f) are closer to the true underlying field and with

255 smaller but still realistic uncertainty compared to the single process model. In scenario one, the difference between the posterior predictive distributions for $\phi_Y(s_z)$ between the two approaches is not substantial, with both models performing adequately, having R^2 scores of 0.99 and 0.97 respectively. In scenario two, the difference between estimates of $\phi_Y(s_z)$ between the models is significant with R^2 scores of 0.99 and 0.68 for the shared and single process models respectively. Finally, in scenario three the difference is again significant with R^2 scores of 0.74 and 0.52 respectively, although less significant compared with

260 scenario two.

(a) Scenario 1

Dependent Variable	Model Parameter	Specified Value	Prior Distribution				Posterior Dist. (Shared Process)				Posterior Dist. (Single Process)			
			Exp.	Std. Dev.	C.I. L.	C.I. U.	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}	1.0	0.67	0.67	0.02	2.46	1.25	0.30	0.73	1.86	1.04	0.31	0.57	1.69
	Kernel Lengthscale l_{ϕ_Y}	3.0	15.00	8.66	3.09	36.12	2.96	0.06	2.85	3.08	2.73	0.20	2.32	3.10
	Mean Constant m_{ϕ_Y}	1.0	0.00	2.00	-3.92	3.92	1.14	0.28	0.61	1.68	1.23	0.26	0.74	1.76
	Noise σ_{ϕ_Y}	0.1	2.00	2.00	0.05	7.38	0.11	0.01	0.09	0.12	N/A	N/A	N/A	N/A
Bias PDF Parameter ϕ_B	Kernel Variance v_{ϕ_B}	1.0	15.00	8.66	3.09	36.12	2.10	1.30	0.48	4.72	N/A	N/A	N/A	N/A
	Kernel Lengthscale l_{ϕ_B}	10.0	0.00	2.00	-3.92	3.92	11.45	1.28	9.07	14.00	N/A	N/A	N/A	N/A
	Mean Constant m_{ϕ_B}	-1.0	0.25	0.14	0.01	0.49	-1.00	0.64	-2.31	0.24	N/A	N/A	N/A	N/A

(b) Scenario 2

Dependent Variable	Model Parameter	Specified Value	Prior Distribution				Posterior Dist. (Shared Process)				Posterior Dist. (Single Process)			
			Exp.	Std. Dev.	C.I. L.	C.I. U.	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}	1.0	0.67	0.67	0.02	2.46	1.13	0.28	0.66	1.66	1.49	0.53	0.65	2.55
	Kernel Lengthscale l_{ϕ_Y}	3.0	15.00	8.66	3.09	36.12	2.97	0.06	2.86	3.09	3.70	0.44	2.83	4.56
	Mean Constant m_{ϕ_Y}	1.0	0.00	2.00	-3.92	3.92	0.70	0.27	0.15	1.22	0.69	0.40	-0.14	1.44
	Noise σ_{ϕ_Y}	0.1	2.00	2.00	0.05	7.38	0.12	0.03	0.08	0.18	N/A	N/A	N/A	N/A
Bias PDF Parameter ϕ_B	Kernel Variance v_{ϕ_B}	1.0	15.00	8.66	3.09	36.12	1.24	0.99	0.16	3.23	N/A	N/A	N/A	N/A
	Kernel Lengthscale l_{ϕ_B}	20.0	0.00	2.00	-3.92	3.92	23.69	5.79	12.29	34.90	N/A	N/A	N/A	N/A
	Mean Constant m_{ϕ_B}	-1.0	0.25	0.14	0.01	0.49	-0.66	0.64	-1.87	0.62	N/A	N/A	N/A	N/A

(c) Scenario 3

Dependent Variable	Model Parameter	Specified Value	Prior Distribution				Posterior Dist. (Shared Process)				Posterior Dist. (Single Process)			
			Exp.	Std. Dev.	C.I. L.	C.I. U.	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}	1.0	0.67	0.67	0.02	2.46	1.18	0.33	0.62	1.83	0.85	0.33	0.30	1.50
	Kernel Lengthscale l_{ϕ_Y}	3.0	15.00	8.66	3.09	36.12	3.00	0.07	2.87	3.14	3.08	0.49	2.03	3.96
	Mean Constant m_{ϕ_Y}	1.0	0.00	2.00	-3.92	3.92	0.95	0.30	0.35	1.53	0.90	0.29	0.33	1.48
	Noise σ_{ϕ_Y}	0.1	2.00	2.00	0.05	7.38	0.16	0.06	0.03	0.27	N/A	N/A	N/A	N/A
Bias PDF Parameter ϕ_B	Kernel Variance v_{ϕ_B}	1.0	15.00	8.66	3.09	36.12	1.50	1.02	0.28	3.56	N/A	N/A	N/A	N/A
	Kernel Lengthscale l_{ϕ_B}	5.0	0.00	2.00	-3.92	3.92	6.34	1.71	3.23	9.20	N/A	N/A	N/A	N/A
	Mean Constant m_{ϕ_B}	-1.0	0.25	0.14	0.01	0.49	-1.17	0.50	-2.11	-0.10	N/A	N/A	N/A	N/A

Table 3. A table showing summary statistics for the prior and 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 posterior distributions for the shared and single process models are given. The specified value for each parameter used to generate the data is also shown.
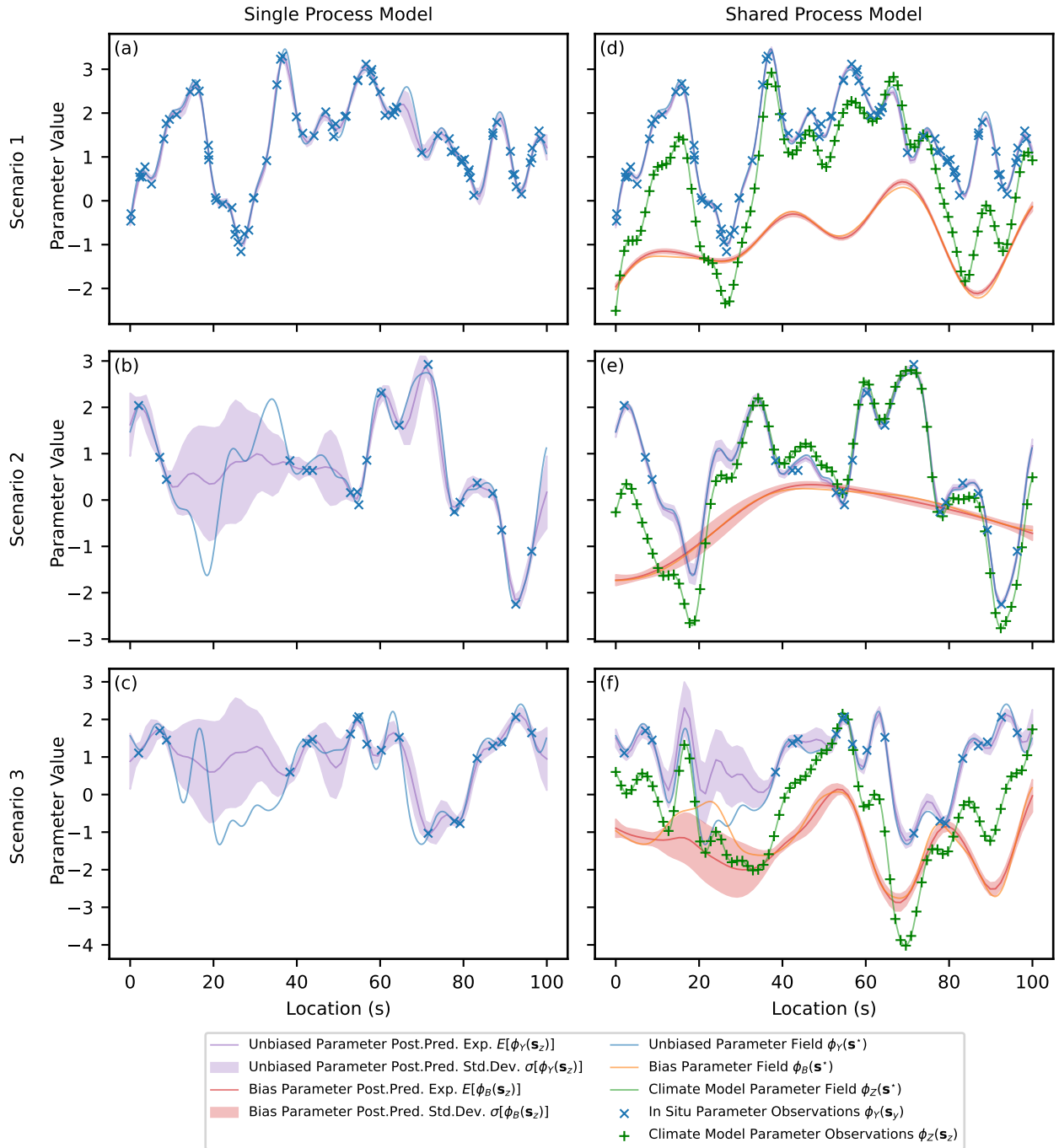
Figure 6. Expectation and 1σ uncertainty of the posterior predictive distributions of the parameter $\phi_Y(\boldsymbol{s}_z)$ and the corresponding bias $\phi_B(\boldsymbol{s}_z)$ for three scenarios. The underlying functions (complete realisations) as well as the simulated input data are also shown.

| | R^2 Scores: Posterior Predictive Estimates of $\phi_Y(\boldsymbol{s}_z)$ | | | |
| | Shared Process Model | | Single Process Model | |
Scenario	Exp.	Std.Dev.	Exp.	Std.Dev.
1	0.99	0.00	0.97	0.01
2	0.99	0.01	0.68	0.07
3	0.74	0.12	0.52	0.10

Table 4. A table showing the expectation and standard deviation of R^2 scores for the posterior predictive estimates of the unbiased PDF parameter at the climate model output locations $\phi_Y(\boldsymbol{s}_z)$ for the shared and single process models for each scenario.

4.2 Hierarchical Example

The hierarchical model presented in this paper is fit to the hierarchical example from Sect. 3.2. The expectation, standard deviation and 95% credible intervals for the prior and posterior distributions of each hyper-parameter of the latent generating processes are given in Table 5. The 95% credible interval of the posterior for every hyper-parameter bounds the value specified in generating the data. As expected the posterior distribution for each hyper-parameter is concentrated closer to the value specified when generating the data than the relatively non-informative prior distributions. The prior and posterior distributions for each hyper-parameter are plot in Fig. F2 of the appendix.

PDF Parameter	Model Hyper-Parameter	Specified Value	Prior Distribution Exp.	Std. Dev.	C.I. L.	C.I. U.	Posterior Dist. Exp.	Std. Dev.	C.I. L.	C.I. U.
Unbiased Mean μ_Y	Kernel Variance v_{μ_Y}	1.0	0.67	0.67	0.02	2.46	1.00	0.32	0.49	1.63
	Kernel Lengthscale l_{μ_Y}	3.0	15.00	8.66	3.09	36.12	3.00	0.22	2.56	3.43
	Mean Constant m_{μ_Y}	1.0	0.00	2.00	-3.92	3.92	0.73	0.28	0.17	1.26
Unbiased Transformed Variance $\tilde{\sigma}_Y$	Kernel Variance $v_{\tilde{\sigma}_Y^2}$	1.0	0.67	0.67	0.02	2.46	0.70	0.25	0.30	1.17
	Kernel Lengthscale $l_{\tilde{\sigma}_Y^2}$	3.0	15.00	8.66	3.09	36.12	2.94	0.24	2.47	3.40
	Mean Constant $m_{\tilde{\sigma}_Y^2}$	1.0	0.00	2.00	-3.92	3.92	1.12	0.24	0.66	1.61
Bias Mean μ_B	Kernel Variance v_{μ_B}	1.0	0.67	0.67	0.02	2.46	1.38	0.63	0.42	2.58
	Kernel Lengthscale l_{μ_B}	10.0	15.00	8.66	3.09	36.12	12.02	3.59	5.08	18.50
	Mean Constant m_{μ_B}	-1.0	0.00	2.00	-3.92	3.92	-0.78	0.56	-1.89	0.29
Bias Transformed Variance $\tilde{\sigma}_B$	Kernel Variance $v_{\tilde{\sigma}_B^2}$	1.0	0.67	0.67	0.02	2.46	0.92	0.48	0.24	1.86
	Kernel Lengthscale $l_{\tilde{\sigma}_B^2}$	10.0	15.00	8.66	3.09	36.12	8.97	1.96	5.07	12.58
	Mean Constant $m_{\tilde{\sigma}_B^2}$	-1.0	0.00	2.00	-3.92	3.92	-0.86	0.42	-1.73	-0.06

Table 5. A table showing summary statistics for the prior and 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 specified value for each hyper-parameter used to generate the data is also shown.

After fitting the model, posterior predictive estimates are made of the unbiased mean and standard deviation parameters at the simulated locations of the climate model ($\mu_Y(\boldsymbol{s}_z)$ and $\sigma_Y(\boldsymbol{s}_z)$). Additionally, estimates of the bias in the parameters are
270 made ($\mu_B(\boldsymbol{s}_z)$ and $\sigma_B(\boldsymbol{s}_z)$). These estimates along with the true underlying values are shown in Fig. 7. The empirical mean and standard deviation of the input data is also given at the locations where they are sampled. The estimates visually appear to perform well at capturing the spatial features of the underlying fields and at estimating a one sigma uncertainty range. For example, in the range of $s \in [28, 38]$, where the main data source is the biased climate model output, the prediction accurately captures the spatial features of the unbiased parameters ($\mu_Y(s)$ and $\sigma_Y(s)$) with an uncertainty that bounds the true underlying
275 value over most of the region. Additionally, uncertainty in the unbiased parameters at the in situ observation sites ($\mu_Y(\boldsymbol{s}_y)$ and $\sigma_Y(\boldsymbol{s}_y)$) due to limited samples is clearly reflected in the estimates.
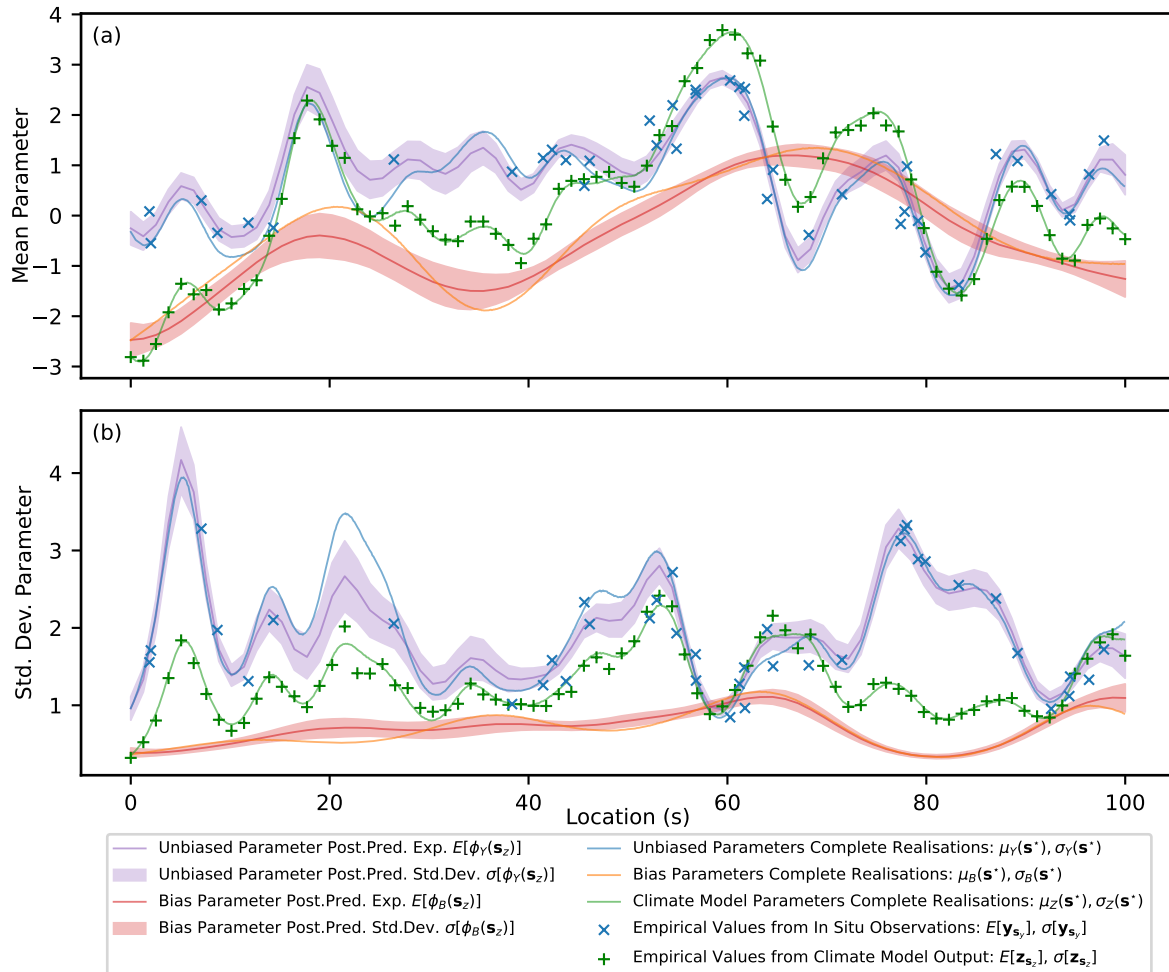
Figure 7. A figure showing the expectation and one sigma uncertainty of the posterior predictive distribution across the domain for the parameters $\mu_Y(\boldsymbol{s}_z)$, $\mu_B(\boldsymbol{s}_z)$, $\sigma_Y(\boldsymbol{s}_z)$ and $\sigma_B(\boldsymbol{s}_z)$ as well as the true underlying functions.

17

Quantile mapping is applied to the climate model output for a single site (z_{s_i}) and the bias corrected time series ($\hat{z}_{s_i}$) is shown in Fig. 8. The site chosen is at $s = 11.4$ and is the same as in Fig. 4a. A generic time series for the climate model output and nearest in situ observations is generated from the correct mean and standard deviations of the samples. Quantile mapping

280   is performed for each posterior predictive realisation of $\mu_Y(s_i)$, $\mu_Z(s_i)$, $\sigma_Y(s_i)$ and $\sigma_Z(s_i)$. This results in multiple realisations of bias corrected time series with an expectation and uncertainty.
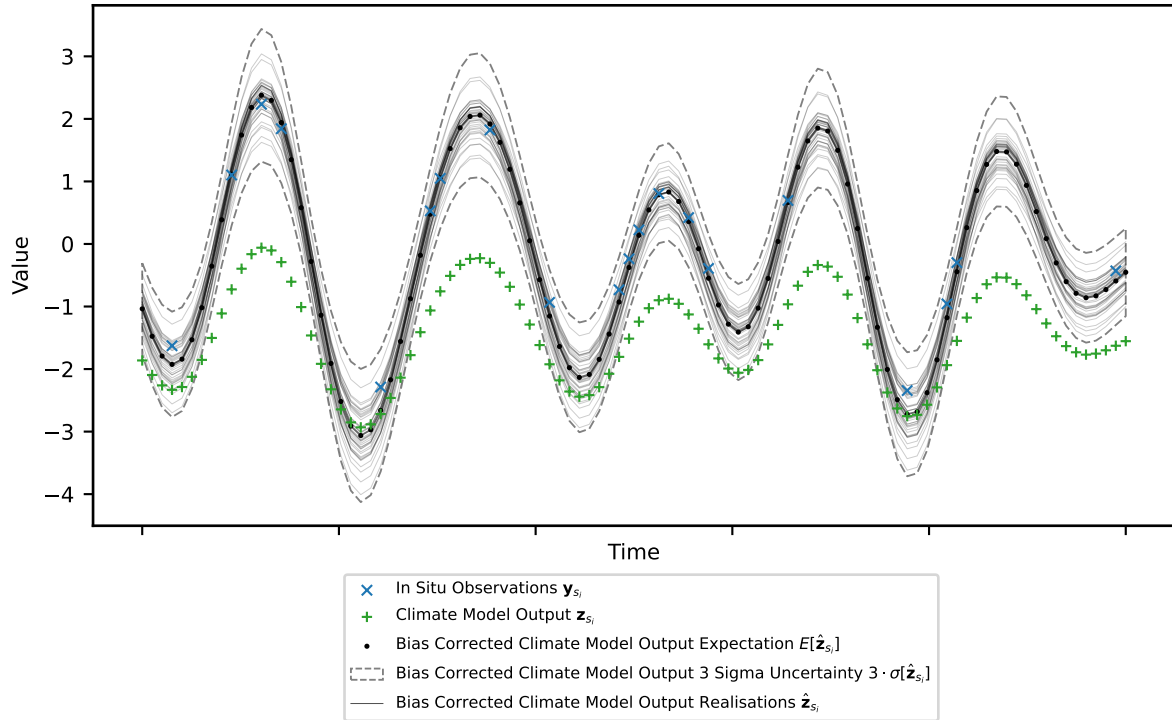
Figure 8. Simulated time series for the climate model output at location $s = 11.4$ and for the nearest in situ observation site. Realisations of the climate model bias corrected time series are shown along with the expectation and three sigma uncertainty range.

5   Discussion

The bias correction framework proposed in this paper models the parameters of the PDFs for the in situ observations and climate model output across the domain using a Bayesian hierarchical model. This allows estimates to be made of the unbiased

285   PDF parameters at the climate model locations and quantile mapping can then be applied to bias correct the climate model time series. The hierarchical model uses GPs to model the spatial covariance structure of the PDF parameters and assumes that each parameter of the climate model output is generated from two independent GPs: One that generates an unbiased component and another that generates a bias. The GP that generates the unbiased component is also modelled as generating the equivalent PDF parameters for the in situ data. This approach reflects the belief that the climate model provides skillful estimates of the PDF

parameters across the domain and that the spatial covariance structure, generated from equations based on established physical laws, has spatial features similar to the true unbiased PDF parameter values. The climate model output, while afflicted with bias, provides comprehensive spatiotemporal coverage and useful information in the inference of the unbiased PDF parameters across the domain. This is assuming the bias signal can be adequately deconstructed from the climate model output with the use of in situ observations. In Sect. 4.1 of the results, the added value of modelling shared latent GPs between the in situ observations and climate model output is demonstrated. This is compared with the approach of modelling a latent GP for the in situ observations alone and inferring the unbiased PDF parameters without incorporating information from the climate model output, as in Lima et al. (2021) and here referred to as the single process approach.

The added value is assessed for three scenarios with differing density of observations and spatial complexity of the bias signal. The methods used to assess the added value include: comparisons of summary statistics for the posterior distributions of the GP hyper-parameters; visual comparisons of the expectation and standard deviation for posterior predictive estimates of the unbiased PDF parameters across the domain; and comparison of R^2 scores for the unbiased PDF parameter at the locations of the climate model output. It is shown that across all these measures, the most added value is provided where the in situ observations are relatively sparse compared with the climate model output and the underlying bias is relatively spatially smooth compared with the the unbiased signal, as in scenario two. In this scenario, despite sparse in situ observations, since the bias signal varies smoothly across the domain the climate model output can be accurately and precisely disaggregated into its unbiased and biased components. This leads to improved estimates of the unbiased PDF parameters at the climate model locations when considering shared GPs compared with the single process approach that uses in situ observations alone, see Fig. 6.

As the density of in situ observations is increased to similar levels as the climate model output itself, then the value added from the climate model output in inference of the unbiased parameters is reduced, illustrated through results for scenario one. The number of in situ observations is sufficient to adequately capture the spatial features of the underlying process (Fig. 6a) as well as the latent spatial covariance structure, encoded through the hyper-parameter estimates of the latent GP (Table 3). Additionally, as the complexity of the bias signal is increased, through for example reducing the length scale of the latent generating process, as in scenario three, then again added value is reduced. The relatively more complex bias structure compared with scenario two makes it more difficult to disaggregate the climate model output into its biased and unbiased components. Despite this, while added value is reduced for scenarios one and three relative to scenario two, incorporating the climate model output in inference is still shown to improve overall performance. Modelling the generating process for the bias explicitly also provides informative information that is potentially useful for future climate model development.

In addition to modelling shared latent processes, another important feature of the methodology presented in this paper is the Bayesian framework. In this framework the parameters/hyper-parameters of the hierarchical model are treated as random variables with associated probability distributions. Uncertainty is inherently propagated through the framework, making the code implementation flexible to further development and so applicable to a wide range of real-world scenarios. Additionally, a Bayesian framework allows expert knowledge to be incorporated in the inference through the choice of prior distributions, which can be especially important where the data is sparse. In Sect. 4.2, results for a simulated hierarchical example illustrate

uncertainty propagation between the PDF parameter values and the hyper-parameters of the latent generating processes. Uncertainty present in the different levels of the hierarchical model are incorporated in the final posterior predictive estimates of the unbiased PDF parameters at the climate model locations, see Fig. 7. Multiple realisations from the posterior predictive can then be used in quantile mapping to produce multiple realisations of the final bias corrected time series, with an expectation and uncertainty range, illustrated in Fig. 8. Reliable uncertainty bands on the final bias corrected time series is important for impact assessments and resulting decision making. Additionally, having multiple realisations for the final bias corrected time series allows further propagation of uncertainty in process models driven by climate model output, such as land surface models (Liu et al., 2014).

The simulated examples presented provide an initial proof of concept, although future studies validating the methodology against real-world applications are important for understanding the remaining limitations and areas for further development. The current primary limitation is expected to be that the underlying spatial covariance structures are assumed stationary. That is that the covariance length scale is assumed constant across the domain, whereas for real-world applications over large and complex topographic domains the length scale will be expected to change depending on the specific topography of the region. Further development of the methodology to incorporate non-stationary kernels would therefore be valuable, although is beyond the scope of this paper. Another important limitation to consider is the assumption that the bias is time independent. In situations where the bias varies gradually through time and uniformly across the domain, the methodology can be further developed such that the mean function of the GPs is modelled with a time dependency. If the bias varies in time non-uniformly across the domain, spatiotemporal GPs will need to be considered, which is again beyond the scope of this paper. Secondary limitations, include the assumption that the unbiased and biased components of the PDF parameter values are independent. In situations where there is a dependence between these components, the methodology presented is still expected to perform adequately, although information is lost by not modelling the dependency explicitly. Additionally, many real-world applications will necessitate specific model adjustments, such as incorporating a mean function dependent on factors like elevation and latitude. Finally, the computational complexity of the model is an important remaining consideration, with inference time of GPs scaling as the cube of the number of data points. Incorporating techniques from the literature such as using sparse variational GPs (SVGP) (Hensman et al., 2015), nearest-neighbor GPs (NNGP) (Datta et al., 2016) or upscaling the climate model output, while outside the scope of this paper, will aid computational performance under demanding real-world scenarios and will facilitate further model development.

6 Conclusion

Current approaches for bias prediction and correction do not aim to preserve the spatial covariance structure of the climate model output and typically either neglect uncertainty or inadequately model uncertainty propagation (Ehret et al., 2012). This paper presents a novel fully Bayesian hierarchical framework for bias correction with uncertainty propagation and latent GP distributions used to capture and preserve underlying covariance structures. In this framework bias is considered in the parameters of the time-independent PDF at each site. Estimates of the unbiased PDF parameters are made at the climate model

locations and then quantile mapping is applied to produce the final bias corrected time series. The novelty of the approach lies in the fully Bayesian implementation, assuming shared latent GPs between the in situ data and climate model output and in propagating uncertainty through the quantile mapping step.

Simple simulated examples are chosen to illustrate key features of the framework. In Sect. 4.1, results are displayed for non-hierarchical examples where the focus is on illustrating the advantage of modelling spatial covariance in both the in situ data and climate model output, assuming shared latent GPs. This is shown to be particularly important in the case of sparse in situ observations and bias that varies smoothly across the domain, where the climate model output itself provides significant added value in predictions of the unbiased PDF parameters. In Sect. 4.2, results are presented for a hierarchical case and focus is on illustrating how the model propagates uncertainty between the different levels and to the final unbiased PDF parameter predictions at the climate model locations. In addition, a simulated example of propagating this uncertainty through quantile mapping is then provided to demonstrate how this results in a bias corrected time series with uncertainty bands, which is desirable for use in impact studies and for informing decision making. Adequately modelling uncertainty in the bias corrected time series is expected to be especially important over areas where the climatology is hard to model and in situ observations are sparse, such as over Antarctica (Carter et al., 2022). The framework presented provides a step towards adequately capturing uncertainty and incorporating underlying spatial covariance structures from the climate model in bias correction. While initial results are promising, further studies applied to real-world datasets are important to further validate the approach and explore remaining limitations. The Bayesian implementation provides a flexible modelling framework, where adjustments to the methodology needed for specific applications can be made while inherently propagating uncertainty.

Appendix A: Bias in Climate Models

Bias in climate models is defined in a number of similar ways across different papers. In Maraun (2016) it is defined as the systematic difference between any statistic derived from the climate model and the real-world true value of that statistic. While in Haerter et al. (2011), bias is defined as the time-independent part of the error between the climate model simulated values and the observed values. In general, across the community involved with climate change impact studies, bias is used to refer to any deviation of interest between the model output and that of the true value (Ehret et al., 2012). Deviations of interest are typically with respect to the statistical properties of the data, for example the mean and variance as well as spatial properties such as the covariance length scale. The methodology developed in this paper treats bias with respect to deviations in the PDFs of the climate model output and observations at each site. Assuming a parametric form for the PDF, this translates to evaluating bias in the parameters of the site-level PDFs, as discussed in Sect. A1. In order to model bias in real-world phenomena while considering the intrinsic spatial structure, the parameters are allowed to vary spatially using stochastic processes, see Sect. A2. After evaluating bias across the domain, the methodology in this paper can be combined with current approaches of correcting bias in climate models, such as quantile mapping, discussed further in Sect. A3.

A1   Bias in Random Variables

Consider a specific in situ observation site (e.g. an automatic weather station) with measurements of some variable $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$, such as midday temperature, and also comprehensive predictions from a climate model at the same location $\boldsymbol{z} = [z_1, z_2, \ldots, z_k]$. In this scenario, bias can be defined in terms of discrepancy between the PDFs of the in situ observations and the climate model predictions. In particular, assuming a parametric density function for both random variables, bias is translated to the discrepancy between the parameters of the PDFs. For example, assuming the observation measurements are independent and identically distributed (i.i.d.) with normal distribution $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$ and the equivalent for the climate model outcomes $Z \sim \mathcal{N}(\mu_Z, \sigma_Z)$, then bias can be quantified by some discrepancy function of the mean parameters (μ_Z, μ_Y) and the standard deviations (σ_Z, σ_Y). This function can be defined in different ways, such as the difference $b(\mu_Z, \mu_Y) = \mu_Z - \mu_Y$ or the ratio $b(\sigma_Z, \sigma_Y) = \sigma_Z / \sigma_Y$.

A2   Bias with Spatially Varying Parameters

Real-world applications, such as impact studies, typically require bias to be evaluated over a spatial region rather than just at a point. Consider a collection of n observational sites $[\boldsymbol{y}_{s_1}, \ldots, \boldsymbol{y}_{s_n}]$, where for each site i there exists m daily measurements of some property such as midday temperature $\boldsymbol{y}_{s_i} = [y_{s_i,1}, \ldots, y_{s_i,m}]$. In addition, consider gridded output from a climate model of the same property at different locations s^*. In this scenario, instead of defining bias in terms of the discrepancy in the PDFs at a single point, bias can be defined with respect to the two latent spatial processes underlying the observed data $\{Y(s)\}$ and the climate model output $\{Z(s)\}$. This allows bias to be estimated across the domain.

As an example, assume that observations and the climate model output come from the spatial stochastic processes $\{Y(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))\}$ and $\{Z(s) \sim \mathcal{N}(\mu_Z(s), \sigma_Z(s))\}$ respectively, where the distribution of data at each location s is normal with

spatially varying parameters $[\mu(s), \sigma(s)]$. The spatial structures of the latent processes are inherited from the spatial structures in the parameters, which are themselves modelled throughout the domain as spatial stochastic processes $\{\mu_Y(s)\}$, $\{\sigma_Y(s)\}$, $\{\mu_Z(s)\}$ and $\{\sigma_Z(s)\}$. In this paper, GPs are used to model the spatial structures, which explicitly capture relationships for the expectation and covariance between points across the domain, see Sect. B of the appendix. Bias is then defined by some discrepancy function of these spatially varying parameters, such as $b(\mu_Z(s), \mu_Y(s)) = \mu_Z(s) - \mu_Y(s)$.

A3 Bias Correction

Bias correction involves using observational data to predict and then reduce bias in the climate model output, with techniques varying in focus and complexity. For example, the delta change method aims to simply apply an adjustment to the mean of the variable under study at each location (Das et al., 2022), while quantile mapping aims to correct the whole PDF of the climate model output (Qian and Chang, 2021). In (Beyer et al., 2020) generalised additive models (GAMs) are used to approximate transfer functions between the climate data and the observed values. The relative performance between methods is assessed in various studies (Teutschbein and Seibert, 2012; Räty et al., 2014; Beyer et al., 2020; Mendez et al., 2020). Typically approaches fail to adequately capture uncertainty and to explicitly model spatial dependencies between locations across the domain. The approach proposed in this paper combines the use of a Bayesian hierarchical model for predicting bias across the region with the established approach of quantile mapping for applying the final correction to the climate model output. Uncertainty is propagated through the framework and underlying spatial structures are explicitly modelled through latent spatial stochastic processes.

Appendix B: Capturing Spatial Structure with Gaussian Processes

A collection of random variables $\boldsymbol{\phi} = [\phi_{s_1}, \phi_{s_2}, ..., \phi_{s_k}]$ indexed according to location in a domain can be modelled through a spatial stochastic process, such as $\{\phi(s) : s \in \mathcal{S}\}$ (shorthand $\{\phi(s)\}$), where $\mathcal{S}$ represents the region under study. The family of Gaussian processes (Rasmussen, 2004) have the property that any finite subset of random variables across the domain are modelled as multivariate normal (MVN) distributed. Consider a collection of k random variables, then the joint distribution between these variables is MVN with $\boldsymbol{\phi} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\phi}$ is some k dimensional random vector, $\boldsymbol{\mu}$ is some k dimensional mean vector and $\boldsymbol{\Sigma}$ is some $k \cdot k$ dimensional covariance matrix. Parameterising the mean and covariance of the MVN distribution then gives the GP, which provides a distribution over continuous functions $\phi(s) \sim \mathcal{GP}(m(s), k(s, s'))$. The collection of parameters for the mean and covariance functions are often referred to as hyper-parameters.

The mean function $m(s)$ of a GP gives the expectation of the parameter at the location index, allowing global relationships for the variable given predictors. In this paper the mean function is considered as a constant across the domain for simplicity, such that $m(s) = m$. In real-world applications a more complex relationship is likely to be useful, for example Eq. (B1) assumes a second order polynomial in two predictors, where the predictors $x_1(s)$ and $x_2(s)$ could be elevation and latitude.

$$m(s) = b_0 + b_1 \cdot x_1(s) + b_2 \cdot x_2(s) + b_3 \cdot x_1(s) \cdot x_2(s) + b_4 \cdot x_1(s)^2 + b_5 \cdot x_2(s)^2 = \boldsymbol{x}(s)^T \boldsymbol{\beta} \tag{B1}$$

The kernel (covariance) function is typically some function of distance between points $d(\boldsymbol{s}, \boldsymbol{s}')$, parameterised by a length scale l and kernel variance v, for example Eq. (B2) gives the well known radial basis function (RBF) for the kernel. The function of distance could be Euclidean or geodesic and arbitrarily complex, including factors such as wind paths, etc. The 2D Euclidean case is given in Eq. (B3), where predictors $x_3(s)$ and $x_4(s)$ could for example be latitude and longitude, which for relatively small distances near the equator are approximately Euclidean. In Fig. B1, an example of how the covariance decays with distance is given for the RBF kernel (a) and realisations of a conditioned GP with the equivalent kernel are illustrated (b).

$$k_{RBF}(\boldsymbol{s}, \boldsymbol{s}') = v \cdot exp\left(-\frac{d(\boldsymbol{s}, \boldsymbol{s}')^2}{2l^2}\right) \tag{B2}$$

$$d(\boldsymbol{s}, \boldsymbol{s}') = \sqrt{(x_3(\boldsymbol{s}') - x_3(\boldsymbol{s}))^2 + (x_4(\boldsymbol{s}') - x_4(\boldsymbol{s}))^2} \tag{B3}$$

The kernel is often assumed stationary for simplicity, as in Lima et al. (2021), meaning that the relationship between covariance and distance is consistent across the domain of study. This assumption is used in the simulated results presented in this paper. The validity of the stationarity assumption should be assessed on an application basis, with factors such as complex topography contributing to non-stationarity.

Gaussian processes have the property that the sum of independent GPs is also a GP. This property is utilised in this paper as the additive relationship $\phi_Z = \phi_Y + \phi_B$ is assumed, where ϕ_Y and the bias ϕ_B are taken as independent and generated from latent Gaussian processes. Note that in the case of different supports between the parameter space and that of the sample space of a Gaussian process, then a link function is included and the relationship $\tilde{\phi}_Z = \tilde{\phi}_Y + \tilde{\phi}_B$ assumed, where the parameters
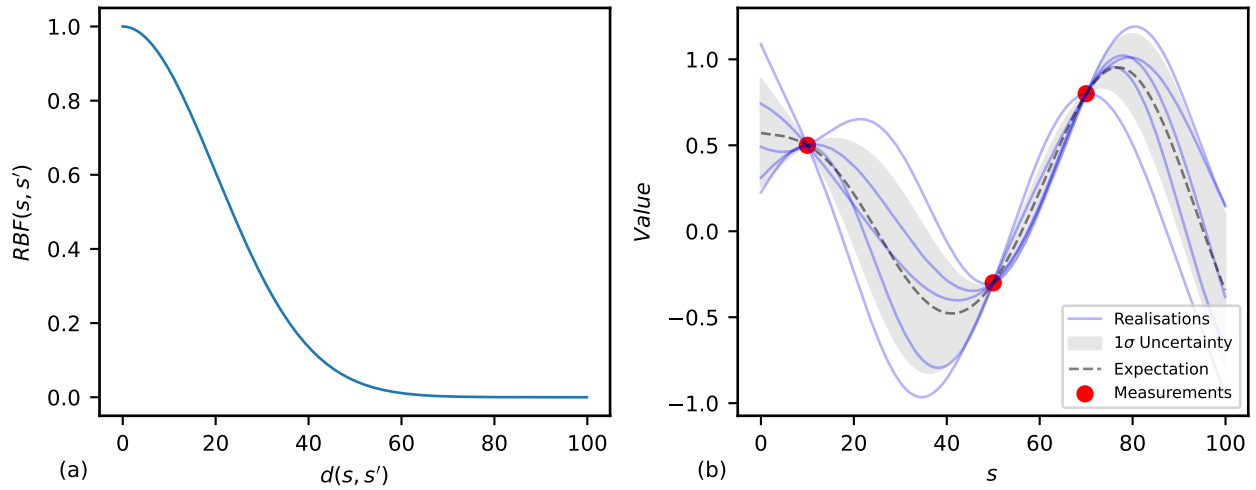
Figure B1. A) Values of the RBF function with a kernel variance equal to 1 and length scale equal to 20. B) Realisations of the GP with the equivalent kernel as in A and conditioned on 3 data points. The expectation and uncertainty of the distribution are shown.

$\tilde{\phi}_Y$ and $\tilde{\phi}_B$ are modelled as independent and generated from GPs. Assuming an additive relationship results in an easy to define distribution for ϕ_Z (or $\tilde{\phi}_Z$), which is a GP where the mean and covariances are simply the sum of the values from the independent GPs (Eq. B4 and B5). The additive relationship captures the belief that the climate model output has some shared latent spatial covariance structure with the in situ observations but is inflicted by an independent bias. In order to make predictions from the unbiased process across the domain $\phi_Y(\hat{s})$ conditioning is then performed on both the observed in situ data $\phi_Y(s_y)$ and the observed climate model output $\phi_Z(s_z)$.

$$m_{\phi_Z} = m_{\phi_Y} + m_{\phi_B} \tag{B4}$$

$$k_{\phi_Z}(s, s') = k_{\phi_Y}(s, s') + k_{\phi_B}(s, s') \tag{B5}$$

Appendix C: Posterior and Posterior Predictives Formulation

C1 Full Hierarchical Model

The in situ observations and climate model output are treated as realisations from the stochastic processes $\{Y(s)\}$ and $\{Z(s)\}$ respectively, where the random variables for a given site are distributed as $Y(s) \sim f(\phi_Y(s))$ and $Z(s) \sim f(\phi_Z(s))$. The symbols $\phi_Y(s)$ and $\phi_Z(s)$ represent the collection of parameters that describe the PDF at the site. The collection of in situ observation sites is given as $s_y = [s_1, \ldots, s_{n_y}]$ and the collection of climate model output locations as $s_z = [s'_1, \ldots, s'_{n_z}]$, then the collection of PDF parameter values for each set of locations is written as $\phi_Y(s_y) = [\phi_Y(s_1), \ldots, \phi_Y(s_{n_y})]$ and $\phi_Z(s_z) = [\phi_Z(s'_1), \ldots, \phi_Z(s'_{n_z})]$. Each parameter of the PDFs are modelled as generated from latent Gaussian processes, one that generates the unbiased component and one that generates the bias, such that $\phi_Y(s) \sim \mathcal{GP}(\cdot, \cdot | \theta_{\phi_Y})$, $\phi_B(s) \sim \mathcal{GP}(\cdot, \cdot | \theta_{\phi_B})$ and $\phi_Z(s) \sim \mathcal{GP}(\cdot, \cdot | \theta_{\phi_Y}, \theta_{\phi_B})$. The symbols θ_{ϕ_Y} and θ_{ϕ_B} represent the collection of hyper-parameters for the Gaussian processes. The posterior distribution for the model can then be written as:

$$P(\phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B} | y, z) = \frac{P(y, z | \phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B}) \cdot P(\phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B})}{P(y, z)} \tag{C1}$$

The first part of the numerator for the fraction can be broken down into:

$$P(y, z | \phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B}) = P(y | \phi_Y(s_y)) \cdot P(z | \phi_Z(s_z)) \tag{C2}$$

The second part of the numerator for the fraction can be broken down into:

$$P(\phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B}) = P(\phi_Y(s_y) | \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B}) \cdot P(\phi_Z(s_z) | \theta_{\phi_Y}, \theta_{\phi_B}) \cdot P(\theta_{\phi_Y}) \cdot P(\theta_{\phi_B}) \tag{C3}$$

The above equations are inherently incorporated into the code implementation through the model definition using the Numpyro python package (Phan et al., 2019). The posterior distribution is approximated using MCMC, which returns realisations of $\phi_Y(s_y)$, $\phi_Z(s_z)$, θ_{ϕ_Y} and θ_{ϕ_B} from the posterior. The posterior predictive estimates of for example $\phi_Y(\hat{s})$ at any set of new locations $\hat{s}$ across the domain is then given by the following:

$$P(\phi_Y(\hat{s}) | y, z) = \int P(\phi_Y(\hat{s}), \phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B} | y, z) d\phi_Y(s_y) d\phi_Y(s_z) d\phi_B(s_z) d\theta_{\phi_Y} d\theta_{\phi_B} \tag{C4}$$

Where the integrand can be broken down into:

$$P(\phi_Y(\hat{s}), \phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B} | y, z) = P(\phi_Y(\hat{s}) | \phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B}) \cdot P(\phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B} | y, z)$$
$$\tag{C5}$$

The second part of this expression is equivalent to the posterior distribution defined earlier. The realisations from the posterior provided through the MCMC inference can be used as parameter values in the first part of the expression above to give a distribution that when sampled from provides posterior predictive realisations for $\phi_Y(\hat{s})$. In the case of Gaussian processes the distribution of $P(\phi_Y(\hat{s})|\phi_Y(s_y), \phi_Z(s_z), \theta_{\phi_Y}, \theta_{\phi_B})$ can be formulated in the following way, where to start take the joint distribution:

$$
\begin{bmatrix} \phi_Y(\hat{s}) \\ \phi_Y(s_y) \\ \phi_Z(s_z) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m_{\phi_Y}(\hat{s}) \\ m_{\phi_Y}(s_y) \\ m_{\phi_Z}(s_z) \end{bmatrix}, \begin{bmatrix} K_{\phi_Y}(\hat{s}, \hat{s}) & K_{\phi_Y}(\hat{s}, s_y) & K_{\phi_Y}(\hat{s}, s_z) \\ K_{\phi_Y}(s_y, \hat{s}) & K_{\phi_Y}(s_y, s_y) & K_{\phi_Y}(s_y, s_z) \\ K_{\phi_Y}(s_z, \hat{s}) & K_{\phi_Y}(s_z, s_y) & K_{\phi_Z}(s_z, s_z) \end{bmatrix} \right) \tag{C6}
$$

Note, that since $\phi_Y(s)$ and $\phi_B(s)$ are independent and $\phi_Z(s) = \phi_Y(s) + \phi_B(s)$, the covariance between the parameters $\phi_Y(s)$ and $\phi_Z(s)$ is simply $COV\left(\phi_Y(s), \phi_Z(s')\right) = COV\left(\phi_Y(s), \phi_Y(s')\right) = K_{\phi_Y}(s, s')$. Additionally, the mean and covariance terms for the process that generates $\phi_Z(s)$ are computed as $m_{\phi_Z}(s) = m_{\phi_Y}(s) + m_{\phi_B}(s)$ and $K_{\phi_Z}(s, s') = K_{\phi_Y}(s, s') + K_{\phi_B}(s, s')$.

Then, defining the following:

$$
U_1 = \begin{bmatrix} \phi_Y(\hat{s}) \end{bmatrix}, U_2 = \begin{bmatrix} \phi_Y(s_y) \\ \phi_Z(s_z) \end{bmatrix}, \boldsymbol{U} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, M_1 = \begin{bmatrix} m_{\phi_Y}(\hat{s}) \end{bmatrix}, M_2 = \begin{bmatrix} m_{\phi_Y}(s_y) \\ m_{\phi_Z}(s_z) \end{bmatrix}, \boldsymbol{M} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \tag{C7}
$$

$$
K_{11} = \begin{bmatrix} K_{\phi_Y}(\hat{s}, \hat{s}) \end{bmatrix}, K_{12} = \begin{bmatrix} K_{\phi_Y}(\hat{s}, s_y) & K_{\phi_Y}(\hat{s}, s_z) \end{bmatrix} \tag{C8}
$$

$$
K_{21} = \begin{bmatrix} K_{\phi_Y}(s_y, \hat{s}) \\ K_{\phi_Y}(s_z, \hat{s}) \end{bmatrix}, K_{22} = \begin{bmatrix} K_{\phi_Y}(s_y, s_y) & K_{\phi_Y}(s_y, s_z) \\ K_{\phi_Y}(s_z, s_y) & K_{\phi_Z}(s_z, s_z) \end{bmatrix} \tag{C9}
$$

The distribution can be written as:

$$
\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \right) \tag{C10}
$$

Where the conditional distribution $P(U_1|U_2)$ is well known for Gaussian distributions and is given as:

$$
P(U_1|U_2) = \mathcal{N}(M_{1|2}, K_{1|2}) \tag{C11}
$$

With parameter values:

$$
M_{1|2} = M_1 + K_{12} K_{22}^{-1}(U_2 - M_2) \tag{C12}
$$

$$
K_{1|2} = K_{11} - K_{12} K_{22}^{-1} K_{21} \tag{C13}
$$

510   This provides the distribution $P(U_1|U_2)$, which is equivalent to the distribution $P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Y(\boldsymbol{s}_z), \boldsymbol{\theta}_{\phi_Y})$ that is needed to compute the posterior predictive.

C2   Non-hierarchical Case

In the non-hierarchical case used in Sect. 4.1, direct observations are assumed for $\boldsymbol{\phi}_Y(\boldsymbol{s}_y)$ and $\boldsymbol{\phi}_Z(\boldsymbol{s}_z)$. In this case the posterior for the model can be written out as:

515
$$P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = \frac{P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)|\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \cdot P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B})}{P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z))} \tag{C14}$$

Where the first expression of the numerator can be broken down into:

$$P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)|\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y)|\boldsymbol{\phi}_Z(\boldsymbol{s}_z), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \cdot P(\boldsymbol{\phi}_Z(\boldsymbol{s}_z)|\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \tag{C15}$$

While the second part of the numerator can be split due to independence between the generating processes, such that:

$$P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{\theta}_{\phi_Y}) \cdot P(\boldsymbol{\theta}_{\phi_B}) \tag{C16}$$

520   As with the full hierarchical model, the above equations are inherently incorporated into the non-hierarchical code implementation, with the posterior distribution approximated using MCMC, which returns realisations of $\boldsymbol{\theta}_{\phi_Y}$ and $\boldsymbol{\theta}_{\phi_B}$ from the posterior. The posterior predictive estimates of for example $\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})$ at any set of new locations $\hat{\boldsymbol{s}}$ across the domain is then given by the following:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = \int P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z))d\boldsymbol{\theta}_{\phi_Y}d\boldsymbol{\theta}_{\phi_B} \tag{C17}$$

525   Where the integrand can be broken down into:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}, \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) \cdot P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) \tag{C18}$$

The second part of this expression is equivalent to the posterior distribution defined earlier. The realisations from the posterior provided through the MCMC inference can be used as parameter values in the first part of the expression above to give a distribution that when sampled from provides posterior predictive realisations for $\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})$. The distribution in the first part of

530 the expression can be formulated in the same way as presented in Sect. C1.

Appendix D: Specific Example with Temperature

Take the case of evaluating bias in the output of near-surface temperature from a climate model relative to some in situ observations. The output from the in situ observations and the climate model are each considered as realisations from latent spatiotemporal stochastic processes, $\{Y(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ and $\{Z(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ respectively. To evaluate bias the time-independent marginal distributions are taken and the data treated as realisations from the spatial stochastic processes $\{Y(s) : s \in \mathcal{S}\}$ and $\{Z(s) : s \in \mathcal{S}\}$. Temperature is known to have diurnal and seasonal dependency and so for the in situ observation measurements to be representative of the time-independent marginal distribution there must be an equal spread of the data over the time of day and season. To reduce this requirement the data can be filtered to just midday January values. Filtering the data has the added benefit of simplifying the marginal distribution and so also the interpretation of bias, allowing the bias to be evaluated for different seasons individually. In the case of January midday temperature, the site-level marginal distributions can be approximated as normal, such that $Y(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$ and $Z(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$.

Treating the site-level distributions as normal results in bias being defined in terms of disparities in the mean and standard deviation parameters between in situ observations and climate model ouput, such that $\mu_B(s) = b_1(\mu_Y(s), \mu_Z(s))$ and $\sigma_B(s) = b_2(\sigma_Y(s), \sigma_Z(s))$. Bias in the climate model output and the parameters of the in situ observations are considered independent and both generated from separate spatial stochastic processes. For example, the bias in the mean $\mu_B(s)$ is considered independent of the mean of the in situ observations $\mu_Y(s)$ and both are modelled as generated from separate GPs: $\mu_Y(s) \sim \mathcal{GP}(m_{\mu_Y}, k_{RBF}(s, s'|v_{\mu_Y}, l_{\mu_Y}))$ and $\mu_B(s) \sim \mathcal{GP}(m_{\mu_B}, k_{RBF}(s, s'|v_{\mu_B}, l_{\mu_B}))$. In this example the mean function of the GP is considered a constant and the kernel/covariance function is considered a radial basis function parameterised by a kernel variance and length scale. Defining the relationship $\mu_Z(s) = \mu_Y(s) + \mu_B(s)$ allows advantage of the property that the sum of 2 independent GPs is itself a GP, such that $\mu_Z(s) \sim \mathcal{GP}(m_{\mu_Y} + m_{\mu_B}, k_{RBF}(s, s'|v_{\mu_Y}, l_{\mu_Y}) + k_{RBF}(s, s'|v_{\mu_B}, l_{\mu_B}))$. see Sect. B of the appendix.

In the case of the standard deviation the parameter space ($\sigma(s) \in \mathbb{R}_{>0}$) is not the same as the sample space of a GP ($\mathbb{R}$) and so a link function is applied $log(\sigma(s)) = \tilde{\sigma}(s) \in \mathbb{R}$. The transformed parameters are then modelled as being generated from GPs: $\tilde{\sigma}_Y(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_Y}, k_{RBF}(s, s'|v_{\tilde{\sigma}_Y}, l_{\tilde{\sigma}_Y}))$ and $\tilde{\sigma}_B(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_B}, k_{RBF}(s, s'|v_{\tilde{\sigma}_B}, l_{\tilde{\sigma}_B}))$. To again take advantage of the property that the sum of 2 independent GPs is itself a GP, the relationship $\tilde{\sigma}_Z(s) = \tilde{\sigma}_Y(s) + \tilde{\sigma}_B(s)$ is defined. The parameter $\tilde{\sigma}_Z(s)$ is then distributed as: $\tilde{\sigma}_Z(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_Y} + m_{\tilde{\sigma}_B}, k_{RBF}(s, s'|v_{\tilde{\sigma}_Y}, l_{\tilde{\sigma}_Y}) + k_{RBF}(s, s'|v_{\tilde{\sigma}_B}, l_{\tilde{\sigma}_B}))$. After predictions across the domain are made of the transformed parameter the inverse link function can be applied to get estimates of the non-transformed parameter.

The diagram in Fig.D1 gives a representation of this full model in a hierarchical framework. Applying MCMC inference provides posterior realisations of the parameters of the model. This includes realisations from the posterior distribution of μ_Y and $\tilde{\sigma}_Y$ at all in situ observation locations, as well as realisations from the posterior of μ_Z and $\tilde{\sigma}_Z$ at all the climate model output locations. These realisations in addition to those of the parameters from the generating GPs can be used to compute the posterior predictive distribution of the parameters $[\mu_Y, \tilde{\sigma}_Y, \mu_B, \tilde{\sigma}_B, \mu_Z, \tilde{\sigma}_Z]$ everywhere in the domain. For the purpose of applying bias correction, the posterior predictive distribution for these parameters can be evaluated at the locations of the

565 climate model output. Quantile mapping is then applied to transform the predicted distribution of the climate model output onto that of the predicted distribution for in situ observations. Applying quantile mapping or alternative methods for multiple realisations of the parameters provides an expectation and uncertainty band for the bias corrected output.
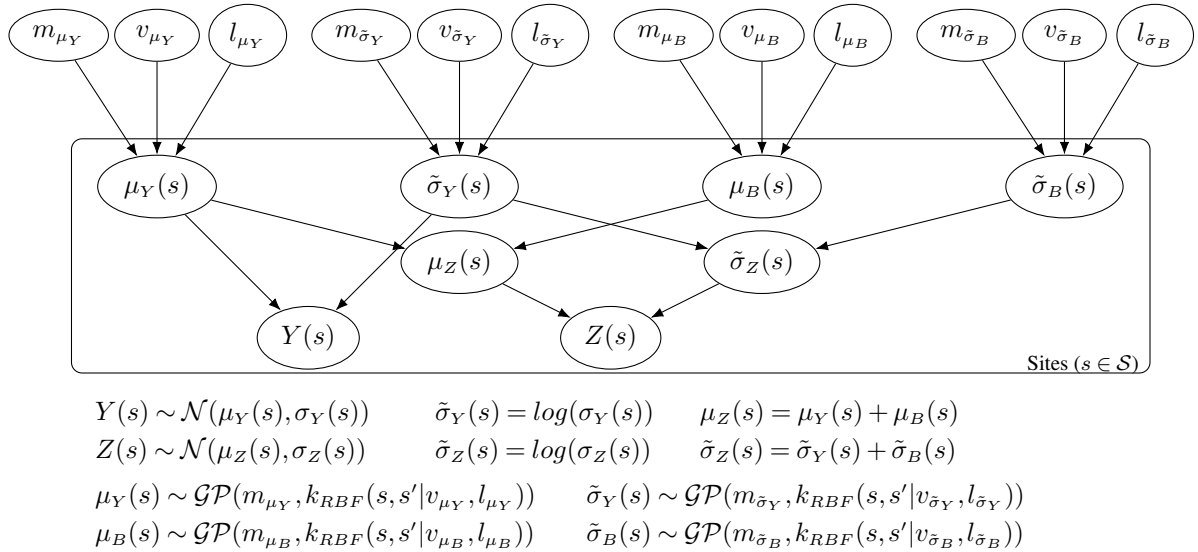

$$Y(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s)) \qquad \tilde{\sigma}_Y(s) = log(\sigma_Y(s)) \qquad \mu_Z(s) = \mu_Y(s) + \mu_B(s)$$
$$Z(s) \sim \mathcal{N}(\mu_Z(s), \sigma_Z(s)) \qquad \tilde{\sigma}_Z(s) = log(\sigma_Z(s)) \qquad \tilde{\sigma}_Z(s) = \tilde{\sigma}_Y(s) + \tilde{\sigma}_B(s)$$
$$\mu_Y(s) \sim \mathcal{GP}(m_{\mu_Y}, k_{RBF}(s, s'|v_{\mu_Y}, l_{\mu_Y})) \qquad \tilde{\sigma}_Y(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_Y}, k_{RBF}(s, s'|v_{\tilde{\sigma}_Y}, l_{\tilde{\sigma}_Y}))$$
$$\mu_B(s) \sim \mathcal{GP}(m_{\mu_B}, k_{RBF}(s, s'|v_{\mu_B}, l_{\mu_B})) \qquad \tilde{\sigma}_B(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_B}, k_{RBF}(s, s'|v_{\tilde{\sigma}_B}, l_{\tilde{\sigma}_B}))$$

Figure D1. Plate diagram with directed acyclic graph showing the full hierarchical model for the case where the site-level distributions are assumed normal with parameters μ and σ. The distribution of these parameters across the domain is modelled with Gaussian processes.

Appendix E: Shared and Single Process Model Comparison

In the main paper comparisons are made made between the shared process and single process models. The shared process
model is the hierarchical model proposed in this paper, while the single process model represents a similar approach taken
from the literature, see Lima et al. (2021). The two models are shown in Fig. E1. In both models the random variables for
the in-situ observations $Y(s)$ and climate model output $Z(s)$ have PDFs with the collection of parameters $\phi_Y(s)$ and $\phi_Z(s)$
respectively. In the case of the shared process model, $\phi_Z(s)$ is modelled as the sum of $\phi_Y(s)$ and some independent bias $\phi_B(s)$.
The parameters $\phi_Y(s)$ and the corresponding bias $\phi_B(s)$ are each themselves modelled over the domain as generated from
Gaussian processes with hyper-parameters $\boldsymbol{\theta}_Y$ and $\boldsymbol{\theta}_B$. The unbiased parameters $\phi_Y(s)$ and hyper-parameters $\boldsymbol{\theta}_Y$ are inferred
from both the in situ data and climate model output. Posterior predictive estimates of $\phi_Y(\boldsymbol{s}_z)$ are made by conditioning on both
sets of data. In the case of the single process model, the PDF parameters for the climate model output and in situ observations
are treated as independent. Only one latent GP is considered with the unbiased parameters $\phi_Y(s)$ and hyper-parameters $\boldsymbol{\theta}_Y$
inferred from in situ observations alone. Posterior predictive estimates of $\phi_Y(\boldsymbol{s}_z)$ are also made by conditioning on just in situ
observations.
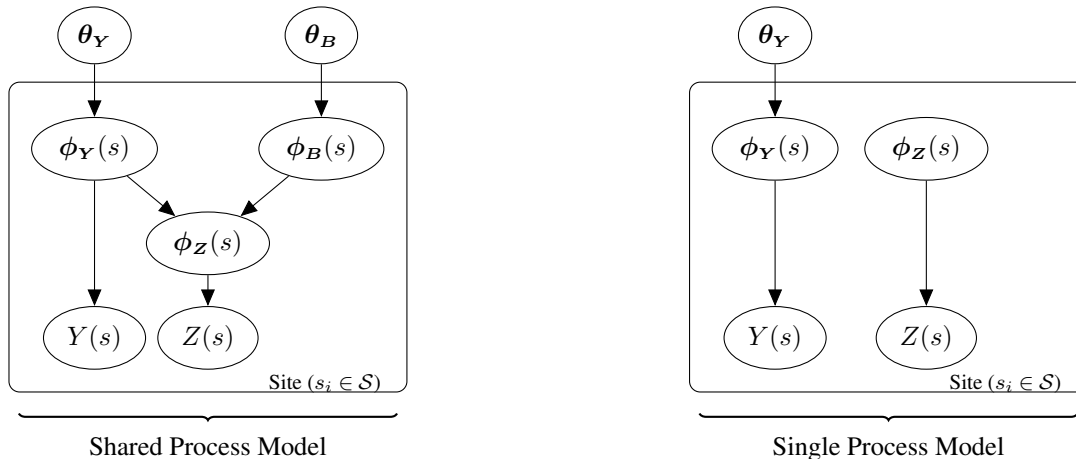
Figure E1. Plate diagram illustrating the difference between the shared process hierarchical model presented in this paper and the single
process model that comparisons are made against.

Appendix F: Prior and Posterior Hyper-Parameter Distributions

A Bayesian framework considers the parameters of the model as random variables with probability distributions. The prior is the assumed distribution before observing any data and the posterior is the updated distribution after observing data. Figures F1 and F2 illustrate the prior and posterior distributions for the non-hierarchical and hierarchical simulated examples respectively. The parameter value specified when generating the data is also shown.
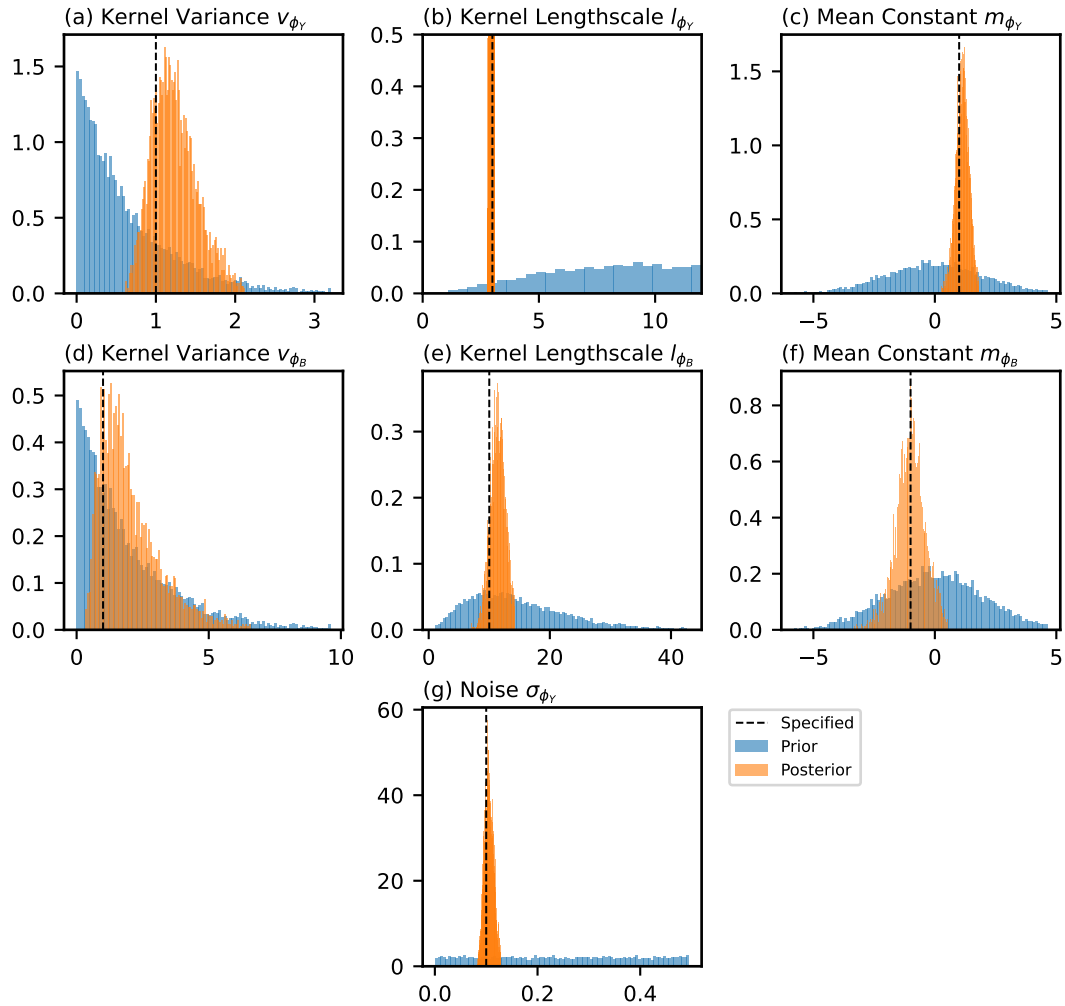
Figure F1. A figure illustrating the prior and posterior distributions for the parameters of the model in the case of scenario one. The value that was specified when generating the data is also shown.
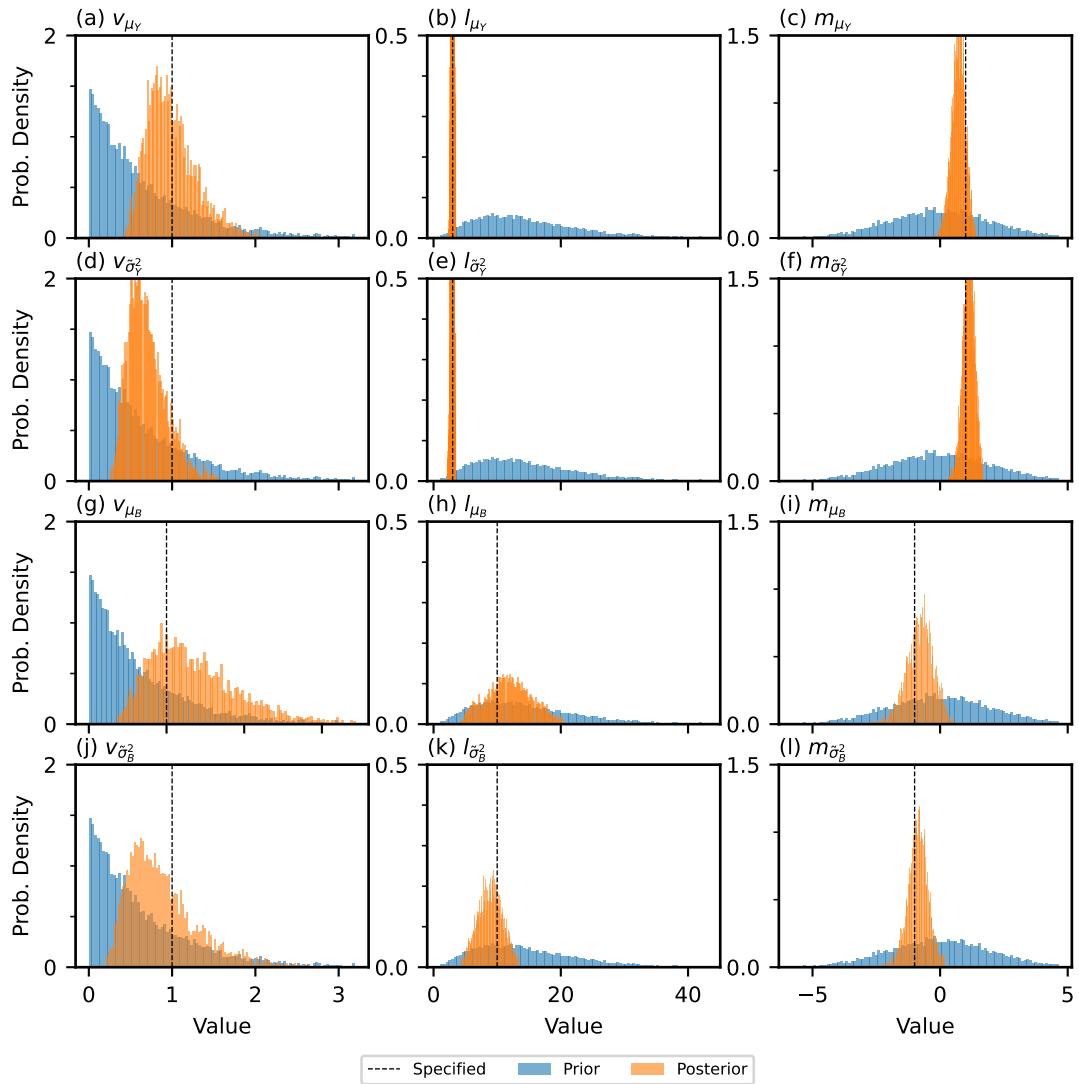
Figure F2. A figure illustrating the prior and posterior distributions for the parameters of the model in the case of the 1D hierarchical example. The value that was specified when generating the data is also shown.

Appendix G: Complex Scenarios

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

34

$x > 50$, 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_X})	1.0	1.0	1.0 & 0.2	1.0
	Kernel Lengthscale (l_{ϕ_X})	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_X})	1.0	1.0	1.0	1.0
	Noise (σ_{ϕ_X})	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 $\phi_Y(s^\star)$, $\phi_B(s^\star)$ and hence $\phi_Z(s^\star)$, as well as observations of $\phi_Y(s_y)$ and $\phi_Z(s_z)$, 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

Dependent Variable	Model Parameter	Scenario A Posterior Dist.				Scenario B Posterior Dist.			
		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_X}	0.87	0.17	0.55	1.18	0.85	0.18	0.53	1.20
	Kernel Lengthscale l_{ϕ_X}	1.06	0.06	0.94	1.19	1.12	0.07	0.99	1.24
	Mean Constant m_{ϕ_X}	0.78	0.15	0.48	1.07	1.42	0.17	1.09	1.77
	Noise σ_{ϕ_X}	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

Dependent Variable	Model Parameter	Scenario C Posterior Dist.				Scenario D Posterior Dist.			
		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_X}	0.52	0.10	0.34	0.72	0.88	0.23	0.49	1.33
	Kernel Lengthscale l_{ϕ_X}	0.75	0.08	0.61	0.92	2.95	0.06	2.82	3.07
	Mean Constant m_{ϕ_X}	1.03	0.11	0.81	1.25	0.93	0.26	0.42	1.44
	Noise σ_{ϕ_X}	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.
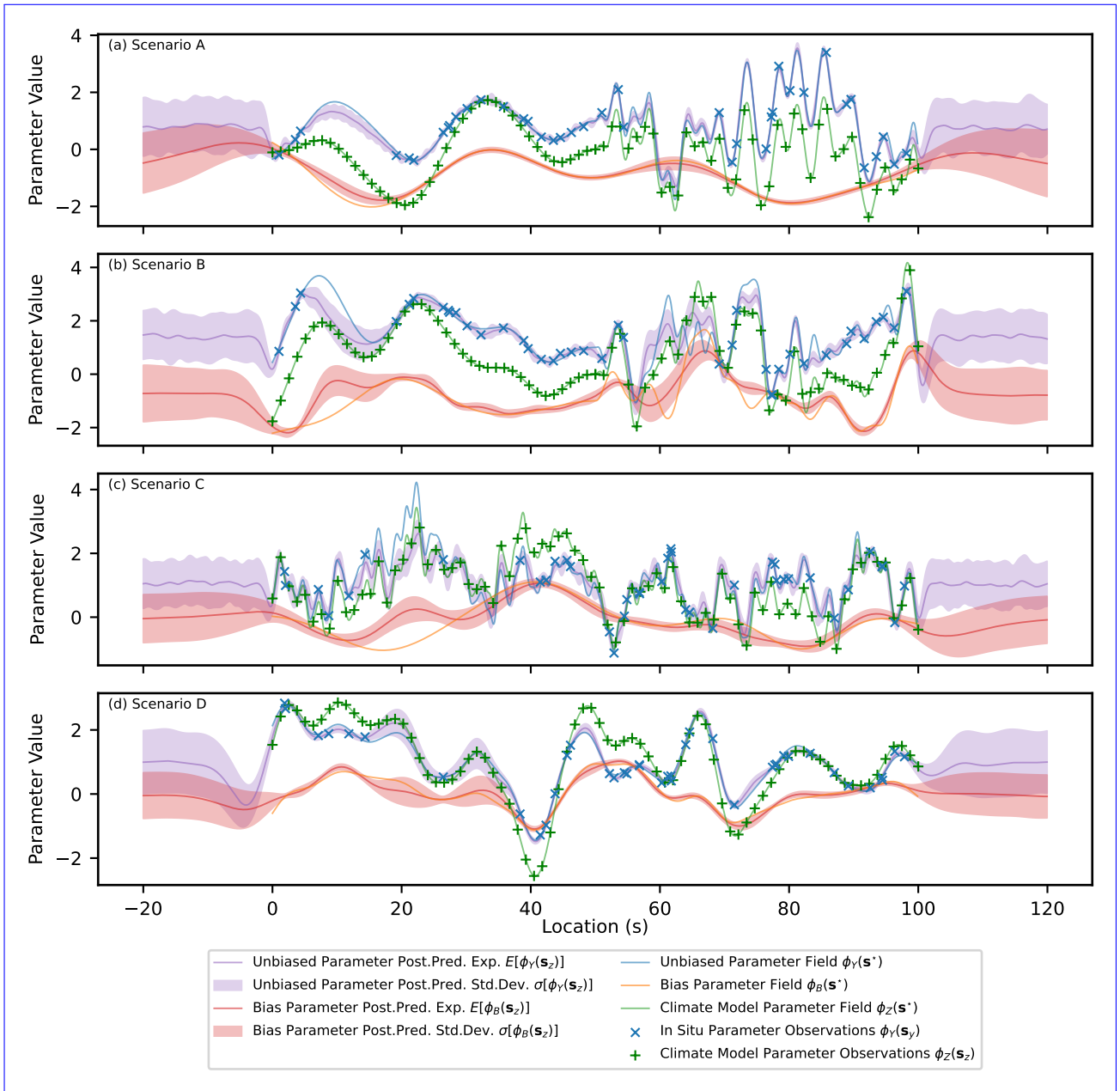
Figure G1. Expectation and 1σ uncertainty of the posterior predictive distributions of the parameter $\phi_Y(\boldsymbol{s}_{\tilde{z}})$ and the corresponding bias $\phi_B(\boldsymbol{s}_{\tilde{z}})$ for three scenarios. The underlying functions (complete realisations) as well as the simulated input data are also shown.

640

References

645 Bader, D., Covey, C., Gutowski, W., Held, I., Kunkel, K., Miller, R., Tokmakian, R., and Zhang, M.: Climate Models: An Assessment of Strengths and Limitations, Climate Models: An Assessment of Strengths and Limitations, 2008.

Beyer, R., Krapp, M., and Manica, A.: An empirical evaluation of bias correction methods for palaeoclimate simulations, Climate of the Past, 16, 1493–1508, https://doi.org/10.5194/cp-16-1493-2020, publisher: Copernicus GmbH, 2020.

Carter, J.: Bias Correction of Climate Models using a Bayesian Hierarchical Model: Code, https://doi.org/10.5281/zenodo.10053653, a.

650 Carter, J.: Data used in generation of results in 'Bias Correction of Climate Models using a Bayesian Hierarchical Model' J.Carter et. al., https://doi.org/10.5281/zenodo.10053531, b.

Carter, J., Leeson, A., Orr, A., Kittel, C., and van Wessem, M.: Variability in Antarctic Surface Climatology Across Regional Climate Models and Reanalysis Datasets, https://doi.org/10.5281/zenodo.6367850, 2022.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Climate
655 Dynamics, 41, 2889–2907, https://doi.org/10.1007/s00382-013-1731-y, 2013.

Das, A., Rokaya, P., and Lindenschmidt, K.-E.: The impact of a bias-correction approach (delta change) applied directly to hydrological model output when modelling the severity of ice jam flooding under future climate scenarios, Climatic Change, 172, 19, https://doi.org/10.1007/s10584-022-03364-5, 2022.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E.: Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatisti-
660 cal Datasets, Journal of the American Statistical Association, 111, 800–812, https://doi.org/10.1080/01621459.2015.1044091, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2015.1044091, 2016.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591–597, https://doi.org/10.1038/nature17145, number: 7596 Publisher: Nature Publishing Group, 2016.

Doblas-Reyes, F. J., Sörensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T.,
665 Lamptey, B. L., Maraun, D., Stephenson, T. S., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking global to regional climate change, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 1363–1512, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
670 https://doi.org/10.1017/9781009157896.001, 2021.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrology and Earth System Sciences, 16, 3391–3404, https://doi.org/10.5194/hess-16-3391-2012, publisher: Copernicus GmbH, 2012.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C.,
675 Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, in: Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Intergovernmental Panel on Climate Change (IPCC), pp. 741–866, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781107415324.020, 2013.

Foreman-Mackey, D.: dfm/tinygp: The tiniest of Gaussian Process libraries, https://doi.org/10.5281/zenodo.7646759.

680 Giorgi, F.: Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?, Journal of Geophysical Research: Atmospheres, 124, 5696–5723, https://doi.org/10.1029/2018JD030094, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JD030094, 2019.

Greeves, C., Pope, V., Stratton, R., and Martin, G.: Representation of Northern Hemisphere winter storm tracks in climate models, Clim. Dynam., 28, 683–702, https://doi.org/10.1007/s00382-006-0205-x, 2007.

685 Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, C., Capotondi, A., Van Oldenborgh, G. J., and Stockdale, T.: Understanding El Niño in Ocean-Atmosphere General Circulation Models: Progress and Challenges, Bulletin of the American Meteorological Society, 90, 325–340, https://doi.org/10.1175/2008BAMS2387.1, 2009.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrology and Earth System Sciences, 15, 1065–1079, https://doi.org/10.5194/hess-15-1065-2011, publisher: Copernicus GmbH, 2011.

690 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, number: 7825 Publisher: Nature Publishing Group, 2020.

Hensman, J., Matthews, A. G., Filippone, M., and Ghahramani, Z.: MCMC for Variationally Sparse Gaussian Processes, in: Advances
695 in Neural Information Processing Systems, vol. 28, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2015/hash/6b180037abbebea991d8b1232f8a8ca9-Abstract.html, 2015.

Hoffman, M. D. and Gelman, A.: The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, Journal of Machine Learning Research, 15, 1593–1623, http://jmlr.org/papers/v15/hoffman14a.html, 2014.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C.,
700 Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, Bulletin of the American Meteorological Society, 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2017.

Hunter, J. D.: Matplotlib: A 2D Graphics Environment, Computing in Science Engineering, 9, 90–95, https://doi.org/10.1109/MCSE.2007.55, conference Name: Computing in Science Engineering, 2007.

705 Lima, C. H. R., Kwon, H.-H., and Kim, Y.-T.: A Bayesian Kriging model applied for spatial downscaling of daily rainfall from GCMs, Journal of Hydrology, 597, 126 095, https://doi.org/10.1016/j.jhydrol.2021.126095, 2021.

Liu, M., Rajagopalan, K., Chung, S. H., Jiang, X., Harrison, J., Nergui, T., Guenther, A., Miller, C., Reyes, J., Tague, C., Choate, J., Salathé, E. P., Stöckle, C. O., and Adam, J. C.: What is the importance of climate model bias when projecting the impacts of climate change on land surface processes?, Biogeosciences, 11, 2601–2622, https://doi.org/10.5194/bg-11-2601-2014, publisher: Copernicus GmbH, 2014.

710 Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Mendez, M., Maathuis, B., Hein-Griggs, D., and Alvarado-Gamboa, L.-F.: Performance Evaluation of Bias Correction Methods for Climate Change Monthly Precipitation Projections over Costa Rica, Water, 12, 482, https://doi.org/10.3390/w12020482, number: 2 Publisher: Multidisciplinary Digital Publishing Institute, 2020.

715 Phan, D., Pradhan, N., and Jankowiak, M.: Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro, https://doi.org/10.48550/arXiv.1912.11554, arXiv:1912.11554 [cs, stat], 2019.

Qian, W. and Chang, H. H.: Projecting Health Impacts of Future Temperature: A Comparison of Quantile-Mapping Bias-Correction Methods, International Journal of Environmental Research and Public Health, 18, 1992, https://doi.org/10.3390/ijerph18041992, 2021.

Rasmussen, C. E.: Gaussian Processes in Machine Learning, in: Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures, edited by Bousquet, O., von Luxburg, U., and Rätsch, G., Lecture Notes in Computer Science, pp. 63–71, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-28650-9_4, 2004.

Räty, O., Räisänen, J., and Ylhäisi, J. S.: Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations, Climate Dynamics, 42, 2287–2303, https://doi.org/10.1007/s00382-014-2130-8, 2014.

Sippel, S., Otto, F. E. L., Forkel, M., Allen, M. R., Guillod, B. P., Heimann, M., Reichstein, M., Seneviratne, S. I., Thonicke, K., and Mahecha, M. D.: A novel bias correction methodology for climate impact simulations, Earth System Dynamics, 7, 71–88, https://doi.org/10.5194/esd-7-71-2016, publisher: Copernicus GmbH, 2016.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., Moss, R., Bauer, S. E., Boucher, O., Brovkin, V., Byun, Y.-H., Dix, M., Gualdi, S., Guo, H., John, J. G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, publisher: Copernicus GmbH, 2021.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Wang, Y. and Chaib-draa, B.: An online Bayesian filtering framework for Gaussian process regression: Application to global surface temperature analysis, Expert Systems with Applications, 67, 285–295, https://doi.org/10.1016/j.eswa.2016.09.018, 2017.

Zhang, Y., Feng, M., Zhang, W., Wang, H., and Wang, P.: A Gaussian process regression-based sea surface temperature interpolation algorithm, Journal of Oceanology and Limnology, 39, 1211–1221, https://doi.org/10.1007/s00343-020-0062-1, 2021.