# Statistical approaches for understanding and correcting systematic errors in climate model estimates of Antarctic surface climatology

**Jeremy Daniel Carter, MSci**

Department of Mathematics and Statistics

Lancaster University

A thesis submitted for the degree of

*Doctor of Philosophy*

June, 2024

**Statistical approaches for understanding and correcting systematic errors in climate model estimates of Antarctic surface climatology**

Jeremy Daniel Carter, MSci.

Department of Mathematics and Statistics, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. June, 2024.

# Abstract

The current and future stability of the Antarctic ice sheet under rising global temperatures is critical to understand with wide-ranging implications, such as influencing ocean currents and having a significant contribution to global sea level rise. Sea level rise results in submergence of land as well as more regular and intense flooding, leading to wide-spread displacement of communities and collapse of coastal ecosystems. Climate models provide invaluable, spatiotemporally comprehensive estimates of past, current and future climatology that are integral for predictions of the stability of the Antarctic ice sheet - impact studies utilise the climate model product to predict events such as ice shelf collapse. Confidence in the findings of impact studies are partially limited though due to systematic errors in the climate model output that are difficult to quantify adequately across the ice sheet.

The first aim of this thesis is to fill a gap in the literature by providing a thorough examination of systematic errors in state-of-the-art regional climate model simulations over Antarctica with a focus on: how the errors vary spatially across the ice sheet; the different sources of errors and their relative contributions; errors across different temporal scales; and errors in variables important for prediction of ice shelf collapse events, including snowfall, snow melt and near-surface air temperature. Following on from this, the second aim of the thesis is to develope a novel approach for bias correction that is specifically designed for the requirements of correcting climate model output over Antarctica. The bias correction methodology is developed with and tested against several simulated data examples and then subsequently with

the real-world case study of correcting near-surface air temperature climate model output over the ice sheet using automatic weather station records. Throughout the thesis several statistical techniques are applied for the first time in this specific area of application, including techniques such as seasonal and trend decomposition using LOESS, Gaussian process regression and hierarchical Bayesian modelling. Utilising these techniques provides useful advantages over previous studies in the literature, including: presenting systematic errors at different temporal scales; explicitly modelling underlying spatial covariance patterns in the data and in systematic errors; and robustly estimating uncertainty in bias corrected climate model output.

# Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography.

Jeremy Daniel Carter

# Publications

Chapter 2:

**Jeremy Carter**, Amber Leeson, Andrew Orr, Christoph Kittel, and J. Melchior van Wessem (2022a). "Variability in Antarctic surface climatology across regional climate models and reanalysis datasets". In: *The Cryosphere* 16.9. Publisher: Copernicus GmbH. DOI: 10.5194/tc-16-3815-2022

Chapter 3:

**Jeremy Carter**, Erick Chacón-Montalván, and Amber Leeson (2024). "Bias Correction of Climate Models using a Bayesian Hierarchical Model". In: *EGUsphere*. Publisher: Copernicus GmbH. DOI: 10.5194/egusphere-2023-2536

# Contents

# Chapter 1

# Introduction

## 1.1 Thesis Structure

The focus of this thesis is to study differences in climate model outputs for snowfall, near-surface air temperature and melt over Antarctica, as well as developing a technique for correcting systematic errors using sparse in situ observations. The thesis is written in the style of a collection of papers, with chapters 2, 3 and 4 each comprising a full paper's worth of work. These primary chapters are all clearly linked, with a logical flow throughout the body of work and with each subsequent chapter being informed from the previous chapter's results. While each primary chapter contains a detailed and specific introduction to the corresponding piece of work, chapter 1 provides an overarching introduction, which motivates the body of work as a whole and highlights links between the papers. An overarching conclusion is provided in chapter 5 with future directions discussed in chapter 6.

## 1.2 Background

Climate models are invaluable tools for climate change impact studies (CCIS), such as those contained within the international panel for climate change (IPCC) assessment reports, which are of key importance to society for understanding,

mitigating and adapting to the effects of climate change (Pörtner et al., 2022). The models are underpinned by fundamental laws of atmospheric, ocean and land surface physics - providing skilful, comprehensive and physically realistic estimates of past, present and future climatology (Flato et al., 2013). Continual improvements in the models are made through for example incorporating advancements in process understanding and increasing the spatial resolution as computational capacity grows (Gutowski et al., 2020). Despite this, significant systematic model errors remain that impede direct interpretation of the results (Ehret et al., 2012) and that require thorough analysis to understand the impact on studies utilising the product (Giorgi, 2019). Climate model intercomparison projects, such as the coordinated regional downscaling experiment (CORDEX) (Gutowski Jr. et al., 2016), provide large ensembles of model outputs with consistent initialisation specifications that enable characterisation of systematic errors (Vautard et al., 2021). Within this context, this thesis aims to explore systematic errors across climate models for the specific case of estimating past and current Antarctic surface climatology. Particular attention is given to evaluating the regional dependency of systematic errors over the ice sheet and on further developing current post-processing bias correction techniques that explicitly incorporate spatial covariance.

Accurate estimates of surface climatology over Antarctica are essential for improving process understanding of events such as ice shelf collapse (Kuipers Munneke et al., 2014) and for assessing the current and future stability of the Antarctic ice sheet (Martin et al., 2019; E. Gilbert and C. Kittel, 2021). The stability of the ice sheet, often expressed in terms of the mass balance, has important implications for sea level rise (SLR) estimates (Kopp et al., 2017), which will threaten coastal regions through factors such as: submergence of land; more regular and intense flooding; increased rates of erosion; collapse of ecosystems; salinisation of land and water sources as well as reduced drainage (Oppenheimer et al., 2022). Non-linear responses of the ice sheet, potentially sensitive to small changes in the estimated surface climatology, could result in as much as 1 m of SLR by 2100 (R. M.

DeConto and Pollard, 2016), which in turn could result in hundreds of millions of people being displaced (Nicholls et al., 2011). Systematic errors in estimates of surface climatology propagate into these ice-sheet process studies and SLR estimates, being a major source of uncertainty (Verjans et al., 2021).

Weather station data on Antarctica is sparse (Yetang Wang et al., 2023) and so reliance is typically placed on climate models to accurately recreate high-resolution, widespread and comprehensive climatology estimates. A collection of state-of-the-art regional climate model (RCM) output for Antarctica, both in terms of model physics and resolution, has recently become available through the international Antarctic-CORDEX collaboration (Gutowski Jr. et al., 2016). Mottram et al., 2021 provides the first inter-comparison of these outputs, showing significant global systematic differences between outputs and focusing on the variable of surface mass balance (SMB). While SMB is useful, understanding processes such as ice shelf collapse additionally requires estimates of the components of mass change, including snowfall and snowmelt (Kuipers Munneke et al., 2014). Also of key importance is understanding how the systematic differences vary spatially across the ice sheet. Chapter 2 of this thesis aims to explore these aspects, quantifying systematic differences spatially across the ice sheet for the variables of snowfall, snow melt and near-surface air temperature (an important driver of melt). The importance of simulation features including atmospheric model dynamics, surface schemes, parametrisations, driving data, boundary conditions, domain specification, resolution and orography on systematic differences between model outputs is discussed.

Modelling all the interdependent physical processes in the climate system along with the requirement for high-resolution spatiotemporal grids make climate models computationally demanding (Washington et al., 2008) and difficult to tune to improve specific systematic errors. This points at the usefulness of statistical approaches for correcting systematic errors. Post-processing, bias correction techniques aim to improve the correspondence of the model output with observations

and with our understanding of the natural world (Ehret et al., 2012). Approaches allow end users to apply corrections to the fields most relevant to their specific area of research, providing significant added value to existing climate model simulation outputs. Typical steps in bias correction involve: selecting in situ observational data and applying quality checks, filtering outliers; interpolating the observational data onto the climate model grid or averaging observations over grid cells; applying some form of transfer function onto the climate model output timeseries at each grid cell independently to align some metric of the marginal distribution with observations (e.g. aligning the means through adding a constant to the climate model timeseries). While convenient, bias correction techniques typically lack physical justification, don't explicitly consider spatial covariance between points or uncertainty in the correction applied (Ehret et al., 2012). This is particularly important over regions such as Antarctica where observations are extremely sparse and uncertainty in any form of interpolation of the observations across the ice sheet is significant. In Lima et al., 2021 a Bayesian approach to bias correction is formulated that captures uncertainty when interpolating observations through modelling the underlying latent field as generated from a Gaussian process (GP), which explicitly considers spatial covariance between points. This approach is developed further in Chapter 3 and extended to consider shared latent generating processes between the observations and climate model output. This novel approach captures the physically realistic covariance structure produced by the climate model and essentially utilises the climate data and observations in predictions of the underlying unbiased fields. Examples are synthetically generated to demonstrate the advantage of this novel approach and to explore under what scenarios it provides the most added value.

This novel bias correction approach, while applicable to many different impact studies utilising climate data, is developed specifically for the use case of correcting climate data over Antarctica where observations are sparse. To this end, automatic weather station (AWS) output over the ice sheet, collated in the recent Yetang Wang et al., 2023 study, is utilised along with RCM output from the Antarctic-CORDEX

initiative to apply the bias correction methodology to correcting real-world near-surface air temperature estimates. Chapter 4 explores the results, providing an initial understanding of the effectiveness in real-world instances and the remaining limitations of the methodology as well as future development opportunities.

# Chapter 2

# Variability in Antarctic Surface Climatology Across Regional Climate Models and Reanalysis Datasets

Regional climate models (RCMs) and reanalysis datasets provide valuable information for assessing the vulnerability of ice shelves to collapse over Antarctica, which is important for future global sea level rise estimates. Within this context, this paper examines variability in snowfall, near-surface air temperature and melt across products from the MetUM, RACMO and MAR RCMs, as well as the ERA-Interim and ERA5 reanalysis datasets. *Seasonal and Trend decomposition using Loess* (STL) is applied to split the monthly time series at each model grid-cell into trend, seasonal and residual components. Significant, systematic differences between outputs are shown for all variables in the mean and seasonal/monthly standard deviations, occurring at both large and fine spatial scales across Antarctica. Results imply that differences in the atmospheric dynamics, parametrisation, tuning and surface schemes between models together contribute more significantly to large-scale variability than differences in the driving data, resolution, domain specification, ice

sheet mask, digital elevation model and boundary conditions. Despite significant systematic differences, high temporal correlations are found for snowfall and near-surface air temperature across all products at fine spatial scales. For melt, only moderate correlation exists at fine spatial scales between different RCMs and low correlation between RCM and reanalysis outputs. Root mean square deviations (RMSDs) between all outputs in the monthly time series for each variable are shown to be significant at fine spatial scales relative to the magnitude of annual deviations. Correcting for systematic differences results in significant reductions of RMSDs, suggesting the importance of observations and further development of bias-correction techniques.

## 2.1 Introduction

The largest source of uncertainty in 2100 Sea Level Rise (SLR) projections, for a given Representative Concentration Pathway (RCP), is from the contribution of ice sheets (Kopp et al., 2017). Non-linear instabilities in the Greenland and Antarctic ice sheets give long tails to their SLR probability projections. For example, under RCP 8.5 the median SLR from Antarctica is projected to be of the order of 20 cm, while the 95th percentile is six times higher, at 130 cm (Jonathan L. Bamber et al., 2019). The Antarctic continent is fringed by ice shelves, which act like 'ice dams', slowing down the flow of inland ice towards the sea (Rignot et al., 2004; T. A. Scambos et al., 2004). The stability of the ice shelves under a warming climate strongly determines the rate of SLR from Antarctica and it is, in part, the difficulty of modelling their complex physical dynamics, leading to retreat/collapse, that results in the large uncertainty in estimates of future SLR (Bulthuis et al., 2019).

The primary method of ice shelf retreat, when considered across the entire ice sheet, is currently through oceanic basal melting (Pritchard et al., 2012; Paolo et al., 2015), although notable exceptions are recent and dramatic collapse events, such as

the disintegration of the Larsen B ice shelf in 2002, which are linked to anomalous atmospheric conditions through the process of melt-induced hydrofracture (Ted A. Scambos et al., 2000; M. v. d. Broeke, 2005; Bell et al., 2018). Anomalously high near-surface air temperatures (leading to enhanced melt events), as well as low accumulation (leading to reduced pore space of surface snow), result in greater lateral propagation of melt water into crevasses across the ice shelf, which then deepen due to increased hydrostatic pressure (Kuipers Munneke et al., 2014). This process reduces the structural integrity of the ice shelf and, in addition to fractures created through supraglacial lake filling and drainage, can eventually lead to collapse (Banwell et al., 2013; Kuipers Munneke et al., 2014). Recent ice sheet modelling studies indicate the critical importance of atmosphere-driven hydrofracture events in distant-past SLR variation (Pollard et al., 2015) and near-future 2100-2300 SLR estimates, particularly under high-emission scenarios (R. DeConto et al., 2021). Comprehensive spatiotemporal estimates of near-surface air temperature over Antarctica, as well as the accumulation of snowfall and quantity of melt water, are thus important for SLR predictions and are typically provided by RCMs (Jan Melchior van Wessem et al., 2018; Agosta et al., 2019; Mottram et al., 2021).

RCMs are limited-area, physically-based, nested models driven at the boundaries by lower-resolution Global Climate Models (GCMs) or reanalysis datasets. The high-resolution available from RCMs is important for capturing fine-scale climatic processes in regions of complex topography, such as föhn winds that occur over ice shelves on the Antarctic Peninsula (Luckman et al., 2014). The region-specific domain enables the set-up and physical schemes of the RCM to be polar optimised (Orr et al., 2021). In addition, further added-value of RCMs is provided through inclusion of region-specific, sophisticated, surface and sub-surface schemes that capture processes such as melt water percolation (Ettema et al., 2010; Datta et al., 2019; Walters et al., 2019). Despite these features, RCMs still exhibit significant systematic errors precluding their direct interpretation in Climate Change Impact Studies (CCIS) (Christensen et al., 2008; Ehret et al., 2012).

The atmospheric model dynamics, surface scheme, parametrisation, driving data, boundary conditions, domain, resolution and orography are all examples of components that contribute to systematic error (Ehret et al., 2012; Giorgi, 2019; Mottram et al., 2021). This paper examines the magnitude and spatial distribution of systematic differences in an ensemble of RCM simulations for Antarctic-wide, 1980-2018 estimates of snowfall, near-surface air temperature and melt water. The relative contribution from different components of the simulations, such as the atmospheric model physics, are discussed. Comparisons of Antarctic-wide RCM simulations of recent-historic surface climatology are present in the literature (Mottram et al., 2021; Jan Melchior van Wessem et al., 2018; Agosta et al., 2019), although the focus is predominantly on Surface Mass Balance (SMB). Surface melt flux, when integrated over the Antarctic ice sheet, only represents a small fraction of the total SMB, which is determined predominantly by the flux of snowfall (Lenaerts et al., 2012; Agosta et al., 2019). This paper provides the first inter-comparison of recent-historic Antarctic-wide RCM simulations framed within the context of ice shelf instability and collapse events, giving specific focus to variability in near-surface air temperature, snowfall and melt water.

Six Antarctic-wide RCM simulations are compared, two from each of the Met Office Unified Model version 11.1 (MetUMv11.1), the Modèle Atmosphérique Régional version 3.10 (MARv3.10) and the Regional Atmospheric Climate Model version 2.3p2 (RACMOv2.3p2). Comparisons are also made to the reanalysis driving data of ERA-Interim and ERA5. The resulting eight Antarctic-wide datasets analysed in this paper are given in Table 2.1. MARv3.10 and RACMOv2.3p2 are both hydrostatic models specifically developed for use over polar regions and their output from Antarctic-wide simulations have been rigorously compared to one-another and against observations (Lenaerts et al., 2012; Jan Melchior van Wessem et al., 2018; Agosta et al., 2019). The MetUMv11.1 is not specifically developed with a focus on the polar regions, although it is a non-hydrostatic model meaning it can be run at and simulate atmospheric circulation features at sub-kilometer resolutions

(Orr et al., 2021), whereas MAR and RACMO are limited to maximum resolutions of 5-10 km horizontal grid spacing (J. M. van Wessem et al., 2016; Datta et al., 2019). Another feature of particular note in the MetUM simulations is that a 'zero-layer' surface scheme is used, which has been identified as a major deficiency in simulations compared with the multi-layer schemes included in MAR and RACMO due to impacts such as that on heat transfer and not representing the insulating properties of the column of snow (Slater et al., 2017; Walters et al., 2019). It is therefore expected that the MetUM, as well as the reanalysis datasets ERA-Interim and ERA5 that both use a single tile to represent snow, will produce much less physically realistic evaluations of melt than MAR and RACMO. Further details on key differences in the model specifications for the simulations analysed in this paper are presented in section 2.

Historic, evaluation simulations are chosen to remove dependency on emission scenarios, which have been shown to introduce divergent trajectories of variables such as melt (Trusel et al., 2015; E. Gilbert and C. Kittel, 2021; Christoph Kittel et al., 2021). Comparisons to observations are not included due to the sparse nature of observations available over Antarctica. Papers including observations typically require comparisons to be made across elevation bins (Mottram et al., 2021; Jan Melchior van Wessem et al., 2018; Agosta et al., 2019). In this paper comparisons are made at a 12 km grid-cell level and it is shown that variability between the simulations has greater dependency on the (latitude, longitude) location than elevation. To study the temporal dependence of variability time series decomposition is applied, separating the signal at each location into an annual, seasonal and residual component. These components are driven by different physical processes and the previous inter-comparison papers cited have not focused on examining variability at different temporal scales. Finally, despite the primary motivation for this paper focusing on surface climatology over ice shelves, the analysis is extended to the whole Antarctic ice sheet and surrounding Southern Ocean. This is done to aid discussion, as surface climatology over the ice shelves is influenced by the behaviour of the

| RCM/Reanalysis Dataset | Domain | Driving Data | Time Period of Forcing [hours] | H.Resolution [km] | Label |
|---|---|---|---|---|---|
| ERA-Interim | Global | - | - | 79 | ERAI |
| ERA5 | Global | - | - | 31 | ERA5 |
| MetUMv11.1 | Antarctica | ERA-Interim | 12 | 12 | MetUM(011) |
| MetUMv11.1 | Antarctica | ERA-Interim | 12 | 49 | MetUM(044) |
| MARv3.10 | Antarctica | ERA-Interim | 6 | 35 | MAR(ERAI) |
| MARv3.10 | Antarctica | ERA5 | 6 | 35 | MAR(ERA5) |
| RACMOv2.3p2 | Antarctica | ERA-Interim | 6 | 27 | RACMO(ERAI) |
| RACMOv2.3p2 | Antarctica | ERA5 | 3 | 27 | RACMO(ERA5) |

Table 2.1: The two reanalysis datasets and six RCM simulation outputs compared in the paper. The label with which each simulation is referred to in the paper is given.

models over the rest of the domain, and extending the analysis provides insights useful for studies not only focused on ice shelves, thus increasing the scope of the work.

## 2.2 Reanalysis Datasets and RCMs Specifications

The ensemble of Antarctic-wide RCM simulations examined in this paper are part of the Coordinated Regional Climate Downscaling Experiment (CORDEX: https://cordex.org/), which is a global project that provides coordinated sets of RCM simulations worldwide. The model specifications for each of the RCM simulations in the chosen ensemble, as well as for the ERA-Interim and ERA5 reanalysis products, are detailed here. There are significant differences, with some of the key aspects being the following: different atmospheric dynamics components; different surface schemes; differences in the vertical and horizontal resolutions, with particular interest on the performance of the high-resolution 12 km MetUM simulation against the low-resolution 49 km MetUM simulation; differences in the driving data, with particular interest on the two RACMO and two MAR simulations that are otherwise identical; and differences in the Digital Elevation Models (DEMs)

and masks used by each model, with MAR and RACMO using comparatively similar DEMs, while the MetUM uses a DEM similar to that of ERA5 (Fig. A.5).

### 2.2.1 ERA-Interim and ERA5

ERA-Interim, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), is a global reanalysis dataset spanning 1979-2019 with 6-hourly temporal resolution and approximately uniform horizontal resolution of 79 km spacing and 60 vertical levels up to 10 Pa (Dee et al., 2011). Era-Interim was world leading and is included as the specified driving data in the base criteria for the CORDEX simulations but has since been superseded by ERA5, also produced by ECMWF (Hersbach et al., 2020), with a number of ERA5 driven simulations also included in the Antarctic-CORDEX ensemble of RCM outputs. The ERA5 reanalysis dataset uses the updated Cycle 41r2 version of the Integrated Forecast System (IFS) numerical weather prediction (NWP) model, with significant developments to model physics and assimilation methods (Hersbach et al., 2020). It spans 1950-Present with an enhanced single hourly temporal resolution, horizontal resolution of 31 km and 139 vertical levels up to 1 Pa. In addition, ERA5 has uncertainty estimates derived from an ensemble of 10 data assimilations performed at a 3 hourly temporal resolution and horizontal resolution of 63 km. The elevation used by ERA-Interim comes from interpolating the GTOPO30 elevation product (ECMWF, 2009), whereas for ERA5 surface elevation is derived from interpolation of a combination of the SRTM30 elevation product along with other surface elevation datasets (ECMWF, 2016). The coupled surface schemes used for ERA-Interim and ERA5 are the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) and updated HTESSEL schemes respectively, both use a single tile to represent snow, while one of the major differences is that HTESSEL allows surface runoff (Balsamo et al., 2009).

### 2.2.2 MAR

MAR is a hydrostatic RCM, specifically developed for the polar areas (X. Fettweis et al., 2013). The Antarctic-wide simulations analysed in this paper have a spatial horizontal resolution of 35 km with a vertical resolution of 24 atmospheric levels. Specific details of the atmospheric component of MAR can be found in Gallée and Schayes (1994) and Gallée (1995). The atmospheric model is fully coupled to the 1-D SISVAT (Soil Ice Snow Vegetation Atmosphere Transfer) surface scheme (X. Fettweis et al., 2013; Xavier Fettweis et al., 2017), which uses the Crocus multi-layer surface snow model (Brun et al., 1992) that contains subroutines for processes such as snow metamorphism as well as meltwater runoff, retention, refreezing and percolation. SISVAT does not include a full radiative transfer scheme in snow/ice and surface albedo is parameterised as a function of snow grain properties (Tedesco et al., 2016). The relaxation technique is used to apply LBCs (Lateral Boundary Conditions) from the driving data every 6 hours and spectral nudging is used to constrain the large-scale behaviour in the upper atmosphere. The two Antarctic-wide MAR simulations studied in this paper are identical apart from differing driving data from ERA-Interim and ERA5 respectively. The orography used in the simulations is from BEDMAP2 (Fretwell et al., 2013). For further detail on MAR and the specific version used to generate the output examined in this paper (MARv3.10) the reader is referred to Agosta et al., 2019 and Mottram et al., 2021.

### 2.2.3 RACMO

RACMO is a hydrostatic RCM with a polar version developed to represent the climate specifically over ice sheets (Van Meijgaard et al., 2008). The RCM uses the dynamical core from HIRHAM (High Resolution Limited Area Model) (Undén et al., 2002) and the physics package CY33r1 version of the Integrated Forecast System (IFS) NWP model from ECMWF. The Antarctic-wide simulations analysed in this paper have a spatial horizontal resolution of 27 km with a vertical resolution of 40 atmospheric levels. The simulations include a multi-layer snow scheme that

simulates hydrological processes such as melt, percolation, refreezing and runoff as well as firn densification (Ettema et al., 2010). In addition, a drifting snow scheme simulates movement of snow from surface winds across the ice sheet (Lenaerts et al., 2010; Lenaerts and M. R. v. d. Broeke, 2012). A snow albedo scheme is implemented, which uses snow grain size as a prognostic variable as well as cloud optical thickness and solar zenith angle to estimate albedo (Munneke et al., 2011). The relaxation technique is used to apply LBCs from the driving data every 6 hours for the RACMO simulation driven by ERA-Interim and every 3 hours for the simulation driven by ERA5 and spectral nudging is used to constrain the large-scale behaviour in the upper atmosphere. The two simulations studied are identical apart from differing driving data from ERA-Interim and ERA5 respectively. The orography used in the simulations is the same as from J. L. Bamber et al., 2009. For further detail on RACMO and the specific version used to generate the output examined in this paper (RACMOv2.3p2) the reader is referred to Jan Melchior van Wessem et al., 2018 and Mottram et al., 2021.

### 2.2.4 MetUM

The MetUM is a non-hydrostatic climate model, not specifically developed or optimised for use over the polar regions but adapted in these simulations for use over Antarctica (Orr et al., 2021). The Regional Atmosphere physics configuration for mid-latitudes (RA1M) is used (Bush et al., 2020), which is identified as the most suitable configuration available for simulating near-surface climatology over Antarctica (E. Gilbert et al., 2020; E. M. K. Gilbert et al., 2021). The Joint UK Land Environment Simulator (JULES) (Walters et al., 2019) is used with the option of a comparatively simple zero-layer snow/soil composite scheme that does not capture processes such as refreezing of melt water (Best et al., 2011). The two Antarctic-wide MetUM simulations analysed in this paper are identical apart from their spatial horizontal resolutions of 12 km and 49 km respectively, both have a common vertical resolution of 70 atmospheric levels. These limited-area, regional

simulations are nested inside the global model configuration of the MetUM, which is itself forced using ERA-Interim reanalysis data and follows a 12 hour re-initialisation procedure that constrains the large-scale circulation in the interior of the domain and prevents it from drifting too far from the driving data (E. M. K. Gilbert et al., 2021). The global MetUM model runs for 24 hour periods, with a re-initialisation happening throughout the domain every 12 hours and boundary conditions for the nested run saved each hour. The first 12 hours of each 24 hour run is discarded as spin-up, while the second 12 hours of each run is kept as output and stitched together with following runs. The orography used in the simulations is the MetUM standard GLOBE 1 km dataset (Elvidge et al., 2019).

## 2.3 Comparison Method

The RCM simulations examined in this paper all use an equatorial rotated coordinate system, where a quasi-uniform horizontal-resolution grid is defined over the region by first specifying the grid over the equator with constant latitude and longitude spacing between each grid-cell and then applying a rotation that takes the domain over the region of interest, for example Antarctica. Direct comparisons between the model output are made by regridding onto a common grid, with a common domain and spatiotemporal coordinates. Cubic precision Clough-Tocher interpolation (Mann, 1999) is performed on the unrotated 'grid latitude' and 'grid longitude' coordinates, which are assumed approximately euclidean, to regrid all model outputs onto the MetUM(011) resolution grid. This grid is chosen as it is the highest resolution grid of the simulations examined, meaning no information is lost as part of the regridding. The domain is filtered to only include the regions common across the model outputs, see Fig. 2.1. The time series examined is filtered to the common 1981-2018 period and 3/6 hourly outputs are aggregated to monthly averages, which captures the dominant annual and seasonal dependency in the variability. For surface air temperature, filtering to only the common timestamps

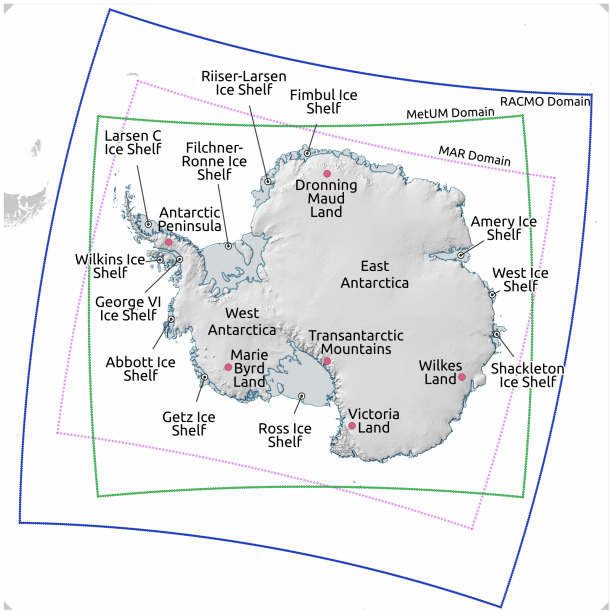



Figure 2.1: Map of Antarctica with some of the main regions and ice shelves labelled, made using the Quantarctica mapping environment (Matsuoka et al., 2021). The RCM simulation domains for the MetUM (green), RACMO (blue) and MAR (purple) are shown. A 1 km resolution hill-shade has been applied from BEDMAP2 (Fretwell et al., 2013).

across the models is first applied and then the average temperature over each month computed. The common timestamps are limited by ERA-Interim to 00 h, 06 h, 12 h and 18 h. This is not required for snowfall or melt, which are defined as fluxes in the model output.

To study annual, seasonal and monthly variability separately, *Seasonal and Trend decomposition using Loess* (STL) (Cleveland et al., 1990) is applied to the time series of each variable at each grid-cell. This results in individual trend (T), seasonal (S) and residual (R) components. The decomposition is additive, meaning for each data point $\nu$=1 to N, the components are summed to give the original time series (Y) (eq.[ 2.1]). The trend component represents the low-frequency/long-timescale pattern of the time series, after filtering out medium and high-frequency signals including the seasonal component, which captures periodic patterns, and the residual component

that explains fluctuations not caused by the long-scale trend or periodicity in the time series.

$$Y_\nu = T_\nu + S_\nu + R_\nu \tag{2.1}$$

Basic time series decomposition involves first approximating the trend component by applying a polynomial fit through the data. Subtracting this component gives the de-trended data that is then split into seasonal sub-series (e.g. January, February, ...) and an average of each sub-series gives the seasonal component of the data. Subtracting both the trend and seasonal components then gives the residual component of the series. STL is a more sophisticated procedure that allows options such as robust fitting (where the influence of outliers is limited) and also a time-varying seasonal component. The algorithm is iterative and involves two loops: the outer loop reduces the influence of outliers by assigning weights based on the magnitude of the remainder term; the inner loop involves estimation of the trend and seasonal components through iterative feedback (Cleveland et al., 1990).

The seasonal component is allowed to vary smoothly over the time series, which is done by applying a LOESS (LOcal RegrESSion) smoothing to the monthly sub-series with window length $n_s$. As $n_s \to \infty$ the LOESS smoothing becomes equivalent to simply taking the average over the sub-series. The value of $n_s$ is recommended to be greater than 7 (Cleveland et al., 1990). As the value increases, the seasonal component approaches a constant periodic state. In this work 13 is used as this allows potential decadal oscillations in the climate to be captured in the seasonal component, such as the Pacific Decadal Oscillation (PDO).

The trend component is estimated using LOESS with a window of default size ($n_t$) given by the smallest odd integer greater than the value in eq.[2.2], which for a period ($n_p$) of 12 months and seasonal smoother ($n_s$) of 13 gives $n_t = 21$. This means the seasonal component can be thought of as a 12 month periodic signal that is allowed to change gradually over a 13 year period, while the trend component can be thought of as similar to the result of taking a weighted moving average of the

deseasonalized time series over a 21 month period. The residual component is then the remaining signal not described by either the smoothly varying seasonal cycle or the long-timescale trend. An example of applying STL decomposition to the time series of snowfall, surface temperature and melt for a grid-cell on the Larsen C ice shelf is available in section A.1 of the appendix. It is noted that for the melt time series additional complications are present that impact the STL decomposition, including no melt auto-correlated austral winter periods and non-homogeneity of the variance. These are discussed in section A.1 of the appendix.'

$$n_t \geq \frac{1.5 n_p}{1 - 1.5 n_s^{-1}} \tag{2.2}$$

In this paper temporal variability between the ensemble of Antarctic-wide datasets is assessed in several ways, including: calculating the Pearson linear correlation coefficient between the outputs for each component of the time series and each variable of interest; quantifying differences in the mean of the time series as well as in the standard deviation of the seasonal and residual components; and calculating the root mean square deviation (RMSD) between the outputs for each variable of interest. Each metric is calculated for every grid-cell in the domain, with Antarctic-wide plots showing spatial patterns. Differences in the monthly mean and standard deviation of the components are calculated over the 37 year 1981-2018 period. For snowfall and melt, differences at each grid-cell are expressed as a proportion of the respective inter-annual deviations, providing some measure of the relative significance of differences at each location. The impact of systematic differences in snowfall and melt on estimates of ice shelf stability depend not only on absolute magnitudes but also on the relative magnitude against a baseline variance. The inter-annual, baseline deviation at each grid-cell is approximated as the ensemble average standard deviation in the trend component of the time series. Results presented in spatial maps then show the relative significance of systematic differences and are not simply dominated by the sites with the highest magnitude snowfall/melt.

## 2.4 Results

Variability in the ensemble of Antarctic-wide outputs (Table 2.1) for the monthly time series of snowfall, near-surface air temperature and melt are quantified across the domain through the evaluation of metrics including the correlation between the outputs, systematic differences in the mean and seasonal/residual standard deviations as well as the RMSDs between outputs. These metrics, for variability in the time series, are evaluated at each grid-cell and the main results shown in sections 2.4.1, 2.4.2 and 2.4.3. Spatial maps are used to show large and small scale patterns in the metrics across the domain. Discussion around the results, including features of variability and the relative importance of contributing factors, is given in section 2.5.

### 2.4.1 Correlation

Results are presented for the correlation in the deseasonalized and detrended, residual component of the time series between each of the 28 unique model output pairs. The correlation is computed at every grid-cell and for melt, grid-cells where the ensemble 40-year average monthly melt is less than 1 millimeter water equivalent per month ($mm\ w.e.\ m^{-1}$) are masked as these regions only experience sporadic and insignificant magnitude melt events, essentially equating to numerical noise in the simulations. The average grid-cell correlation across the entire ice sheet is then taken and the results given in Fig. 2.2. High correlation is shown for snowfall ($> 0.80$) and near-surface air temperature ($> 0.90$) across all model pairs, while results for melt show a significant divide between the reanalysis datasets and the RCMs. The correlation for melt between just the RCMs is moderate to high ($> 0.55$) across all pairs, while for the reanalysis datasets the correlation is low ($< 0.35$) for comparisons to all other models, including between ERA-Interim and ERA5. Another key feature includes the comparatively high correlation shown in every variable between simulations of the same RCM but differing resolution/driving data (MetUM(044)-

MetUM(011), MAR(ERAI)-MAR(ERA5) and RACMO(ERAI)-RACMO(ERA5)). It is noted that while not discussed further here, direct comparisons to correlations shown for the trend component of the time series are presented in section A.4 of the appendix.

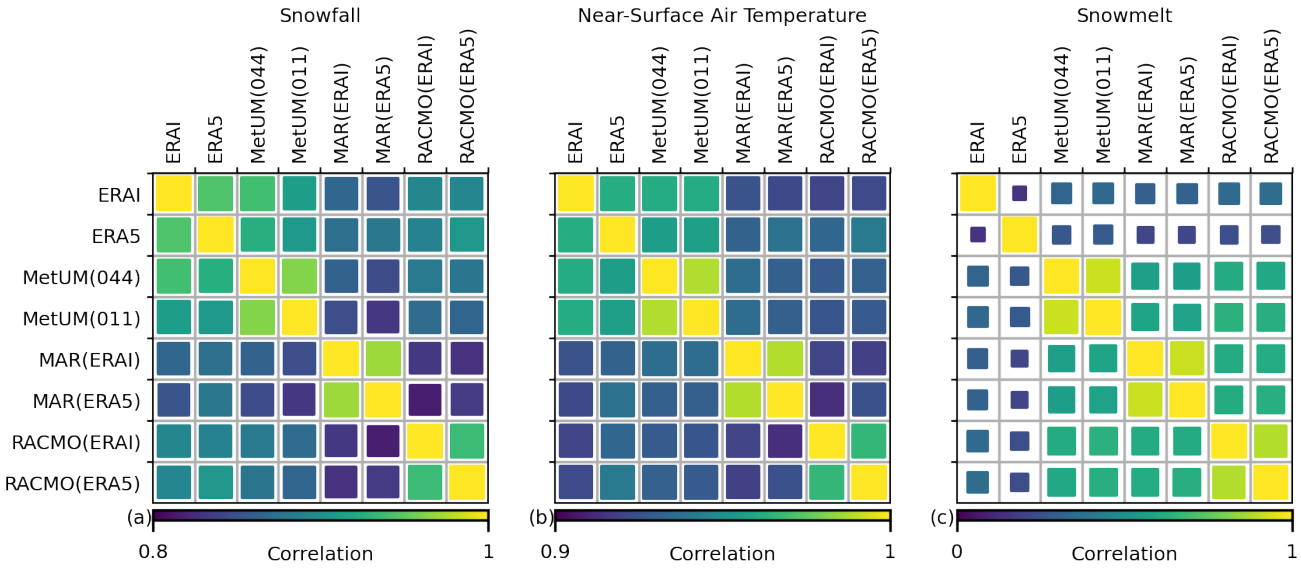



Figure 2.2: The correlation for snowfall (a), near-surface air temperature (b) and melt (c) between models averaged over the ice-sheet. The colour scale relates to the value of correlation and the scale is adjusted for each plot. The size of each square also relates to the value of correlation, although is kept constant across the figures, going from 0-1, to make comparisons clear between the different variables.

A spatial map of the median correlation in the residual component across the 28 unique model output pairs is plot in Fig. 2.3. An ice sheet-only mask is applied for melt using the high resolution shapefile from Depoorter et al., 2013, which is found to remove the most prominent edge effects caused by comparing high- and low-resolution models for a variable that is dependent on the sea/ice categorisation of the grid-cell. In addition, grid-cells where the ensemble 40-year average melt is less than 1 millimeter water equivalent per month ($mm\ w.e.\ m^{-1}$) are again masked. In Fig. 2.3 the median correlation for near-surface air temperature is shown to be

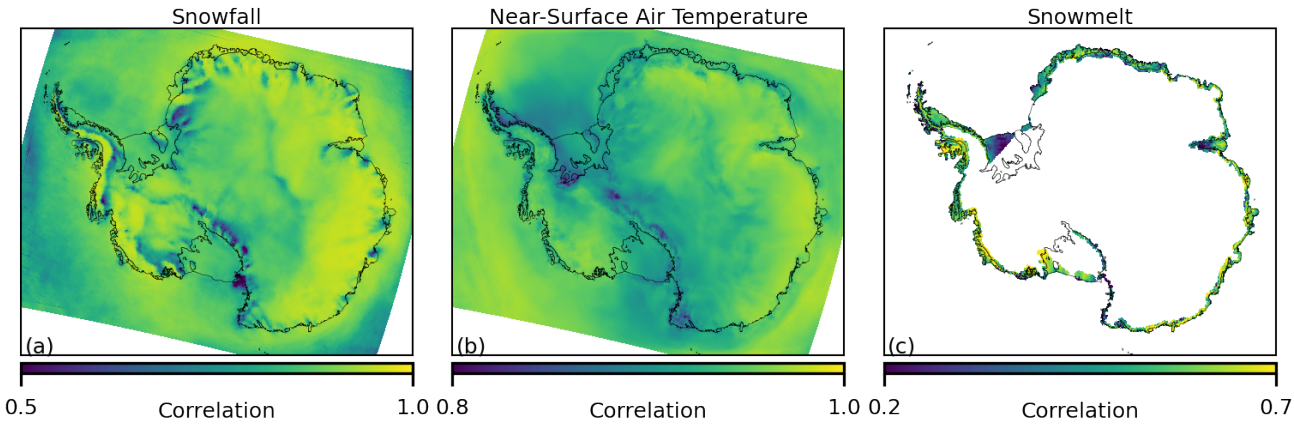



Figure 2.3: The median correlation by grid-cell in the residual component of the monthly time series between the 28 unique model pairs for snowfall (a), near-surface air temperature (b) and melt (c). The colour scale relates to the value of correlation and the scale/limits are adjusted for each plot.

high ($> 0.8$) across the ice sheet, while for snowfall the correlation remains high again across the majority of the ice sheet but is moderate to low over regions such as the Transantarctic mountains, where the topography varies sharply. For melt, the correlation is moderate over the majority of ice shelves, although is noticeably low over the Ronne ice shelf, the ice shelves bounding Victoria Land, and the interior of the Amery ice shelf.

### 2.4.2 Mean and Standard Deviation: Magnitude and Spatial Pattern of Differences

The 1981-2018 mean and standard deviation for each component of the monthly time series of the ice sheet total snowfall, average near-surface air temperature and total melt are displayed in Table 2.2. Results show that even aggregated across the entire ice sheet significant systematic differences exist between the outputs for each variable. For example, the magnitude of differences in the mean across the ensemble are comparable in magnitude to the average trend standard

deviation, which represents inter-annual variations. One particularly striking feature is the contrast between the low monthly melt of ERA5 (1.1 GT/month) compared to the high monthly melt of ERA-Interim (15.5 GT/month) and all RCMs (9.1-14.2 GT/month). It is noted that the relative magnitudes of standard deviations in each component of the time series depends on the variable and that for temperature and melt the seasonal deviation is dominant, while for snowfall both the seasonal and residual deviations have similar magnitudes. Another feature, is that systematic differences are comparatively low between simulations of the same RCM but differing resolution/driving data (MetUM(044)-MetUM(011), MAR(ERAI)-MAR(ERA5) and RACMO(ERAI)-RACMO(ERA5)) when compared with differences present between the different RCMs.

To understand how systematic differences vary spatially the 1981-2018 mean and seasonal/residual standard deviations for the monthly time series of each variable are also computed at a 12 km grid-cell level. Since it is found that systematic differences in the mean and standard deviations are most pronounced between different models in the ensemble, results presented in Fig. 2.4, 2.5 and 2.6 are filtered to only include: ERA5; MetUM(011); MAR(ERA5); and RACMO(ERA5). Differences for each model are then plotted relative to this reduced ensemble average (model-ensemble avg.). Results showing direct comparisons between same/similar model pairs are given in Fig. A.2, A.3 and A.4 in the appendix and include differences in the mean and standard deviations between: ERA-Interim and ERA5; MetUM(044) and MetUM(011); MAR(ERAI) and MAR(ERA5); RACMO(ERAI) and RACMO(ERA5). Differences in the standard deviation of the trend component are excluded from grid-cell level results as it is shown in Table 2.2 that the relative magnitude against standard deviations in the seasonal and residual components is low. For snowfall and melt, differences at each grid-cell are expressed as a proportion of the respective inter-annual deviations, approximated by the ensemble average standard deviation in the trend component.

In Fig. 2.4, it can be seen that for snowfall there exists high-magnitude, spatially-

| **Snowfall / GT** | ERAI | ERA5 | MetUM(044) | MetUM(011) | MAR(ERAI) | MAR(ERA5) | RACMO(ERAI) | RACMO(ERA5) | Average |
|---|---|---|---|---|---|---|---|---|---|
| Monthly Mean | 179.3 | 225.8 | 212.4 | 222.8 | 234.9 | 235.6 | 229.6 | 231.3 | 221.5 |
| Trend St.D. | 7.1 | 9.8 | 9.6 | 9.8 | 8.7 | 8.8 | 8.6 | 9.1 | 8.9 |
| Seasonal St.D. | 25.9 | 34.1 | 26.8 | 25.8 | 38.7 | 38.8 | 30.1 | 31.0 | 31.4 |
| Residual St.D. | 21.8 | 28.3 | 28.3 | 28.9 | 26.3 | 26.3 | 28.0 | 28.2 | 27.0 |
| **Temperature / K** | ERAI | ERA5 | MetUM(044) | MetUM(011) | MAR(ERAI) | MAR(ERA5) | RACMO(ERAI) | RACMO(ERA5) | Average |
| Monthly Mean | -32.6 | -33.3 | -34.2 | -33.9 | -32.2 | -32.2 | -34.0 | -33.8 | -33.3 |
| Trend St.D. | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.5 | 0.4 |
| Seasonal St.D. | 9.0 | 7.7 | 9.3 | 9.2 | 8.7 | 8.6 | 8.8 | 8.7 | 8.7 |
| Residual St.D. | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.2 | 1.2 | 1.1 |
| **Melt / GT** | ERAI | ERA5 | MetUM(044) | MetUM(011) | MAR(ERAI) | MAR(ERA5) | RACMO(ERAI) | RACMO(ERA5) | Average |
| Monthly Mean | 15.5 | 1.1 | 13.2 | 14.2 | 11.9 | 12.1 | 9.3 | 9.1 | 10.8 |
| Trend St.D. | 2.4 | 0.4 | 3.0 | 3.1 | 3.3 | 3.1 | 2.9 | 2.7 | 2.6 |
| Seasonal St.D. | 29.3 | 2.0 | 25.8 | 27.0 | 23.1 | 23.4 | 18.5 | 18.2 | 20.9 |
| Residual St.D. | 5.3 | 0.8 | 6.8 | 6.9 | 7.0 | 6.8 | 7.1 | 6.7 | 5.9 |

Table 2.2: After aggregating across the ice sheet, the mean and standard deviation for each component of the monthly time series for total snowfall, average near-surface air temperature and total melt are given. Values for snowfall and melt are expressed in units of gigatonnes while values for temperature are expressed in Kelvin.

coherent systematic differences over both the ocean and ice sheet, particularly in the mean of the time series (Fig. 2.4a,d,g,j), for each model relative to the ensemble average. A specific example is the strong negative difference in the mean snowfall over the ocean and strong positive difference over the majority of the ice sheet shown by MAR (Fig. 2.4g). In general, the +ve/-ve sign of the differences in the mean and standard deviations for snowfall over the interior of the ice sheet, over the Transantarctic mountains and over the oceans show a relatively large spatial correlation length scale. In contrast, near the periphery of the ice sheet, the sign of the differences exhibit a smaller correlation length scale. Regions such as the Antarctic Peninsula exhibit direction dependent length scales, with a comparatively large length scale in the latitude direction and a comparatively short length scale in the longitude direction. The magnitude of the differences shown over the ice sheet appear greater over sharply varying topography, such as the Transantarctic

mountain range and the steep coastal slopes of the ice sheet. An exception to this being high magnitude differences also shown in the mean component over the comparatively flat region of the interior of East Antarctica for the MetUM(011) and MAR(ERA5) (Fig. 2.4d,g).

It can be seen that for snowfall the difference present in the mean of the time series has a similar spatial signature and sign as the difference in the standard deviation of the seasonal and residual components (e.g. Fig. 2.4g,h,i). Exceptions to this include for example the difference in snowfall from the MetUM(011) relative to the ensemble in the interior of East Antarctica, where despite having a lower mean snowfall the standard deviation in the seasonal component is greater than the average of the ensemble (Fig. 2.4d and 2.4e).

As with snowfall, there exists significant differences over both the ocean and land for near-surface air temperature between the models, again particularly in the mean of the time series (Fig. 2.5). For example, MAR shows a significant positive difference in the mean of the time series over the majority of the ice sheet (Fig. 2.5g) and a significant negative difference over the majority of the surrounding ocean. The magnitude of differences shown over the ice sheet again appear greater over regions of steep topography, particularly for the MetUM(011) and MAR(ERA5) outputs (Fig. 2.5d,g). The spatial patterns of differences in near-surface air temperature differ in shape compared to those present for snowfall. In particular, near the edge of the ice sheet there are less positive-to-negative fluctuations with changing longitude and instead the patterns are more parallel to the coastline (Fig. 2.5d,g). While there are similar spatial patterns between the mean temperature difference and the seasonal standard deviation difference (as for Fig. 2.4), the sign of the differences in Fig. 2.5 is in general shown to change, for example over the majority of the ice sheet in Fig. 2.5d compared with Fig. 2.5e. In Fig. A.1e in the appendix, which gives the temperature profiles from each simulation over an example grid-cell on the Larsen C ice shelf, a colder mean temperature is shown to be the result of similar summer temperatures with more severe winter temperatures.

A land-only mask has been applied for melt in Fig. 2.6 as well as a filter masking any grid-cells where the ensemble mean average monthly melt is less than 1 $mm\ w.e.\ m^{-1}$. This limits discussion of the patterns in differences of the mean and standard deviations to the peripheral areas, which are predominantly ice shelves. The magnitude of the differences present is, as for snowfall, significant relative to the inter-annual variability of melt at each grid-cell. Unlike for snowfall and near-surface temperature, the relative magnitude of differences across the ensemble in the standard deviation of both the seasonal and residual components of the time series are greater than differences present in the mean of the time series. It is noted that for melt, which occurs primarily over just the summer months, greater values of standard deviation in the seasonal component are expected to represent both/either higher magnitudes over peak months and/or a prolonged melt season. As with near-surface temperature and snowfall there are both short and long spatial length scale patterns. An example of a relatively localised spatial pattern is that of the strong positive difference shown by MAR over the interior of the Amery ice shelf in the mean of the time series, as well as the standard deviation of the seasonal and residual components. An example of a large-scale pattern is that ERA5 shows a considerable negative difference in the mean and standard deviations of melt over the majority of ice shelves.

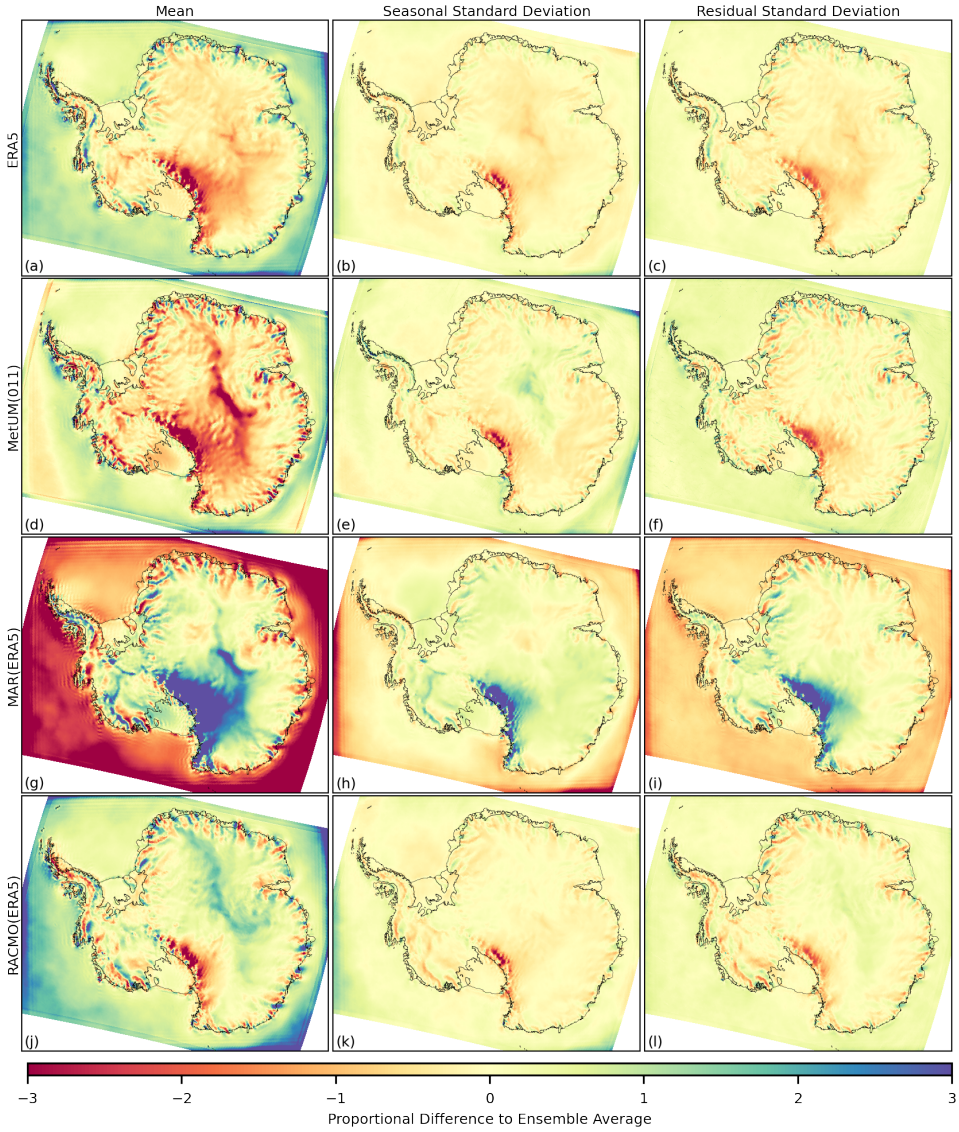



Figure 2.4: The difference to the ensemble average (model-ensemble avg.) for the 1981-2018 time series of snowfall, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l). The ensemble includes: ERA5 (a,b,c); MetUM(011) (d,e,f); MAR(ERA5) (g,h,i); and RACMO(ERA5) (j,k,l). Differences at each grid cell are expressed as a proportion of average inter-annual variation and so do not have units.

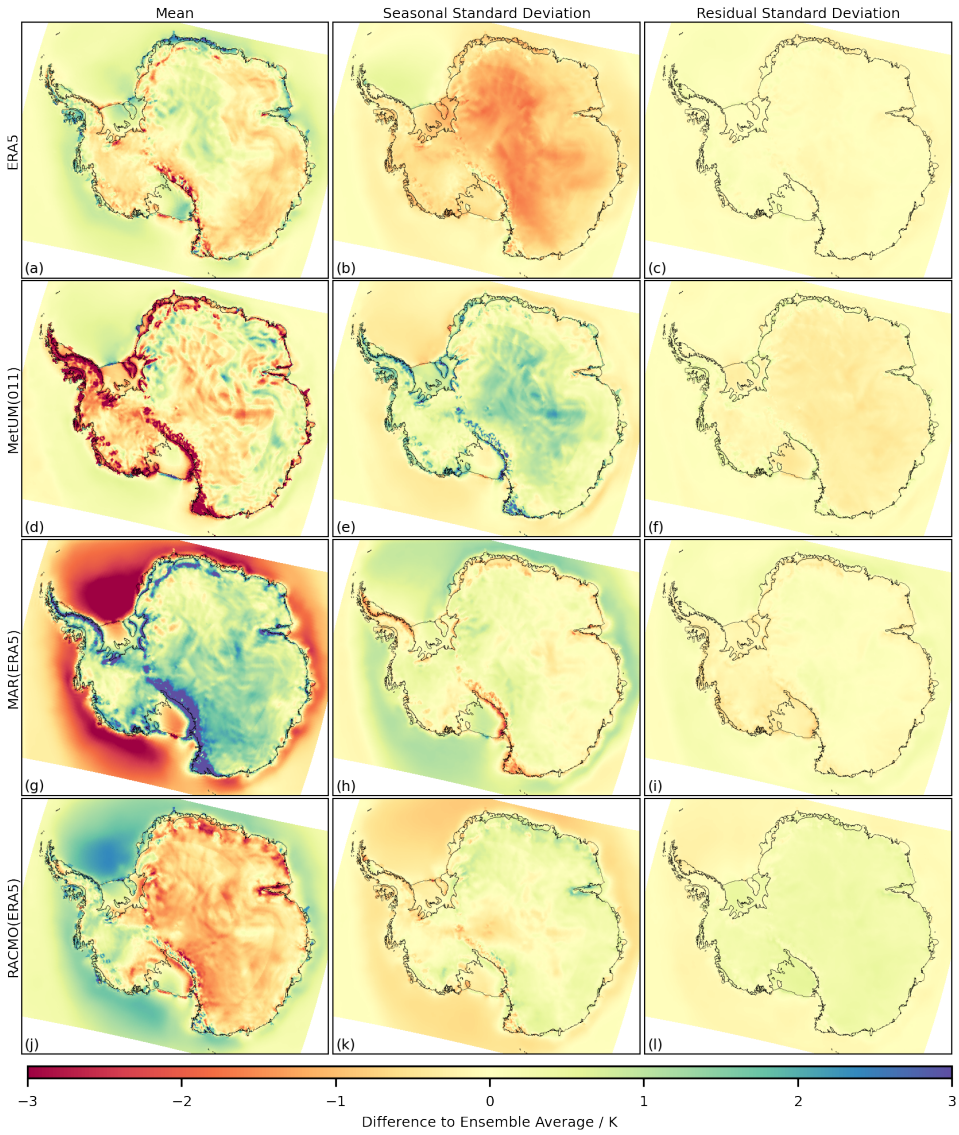



Figure 2.5: The difference to the ensemble average (model-ensemble avg.) for the 1981-2018 time series of near-surface air temperature, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l). The ensemble includes: ERA5 (a,b,c); MetUM(011) (d,e,f); MAR(ERA5) (g,h,i); and RACMO(ERA5) (j,k,l).

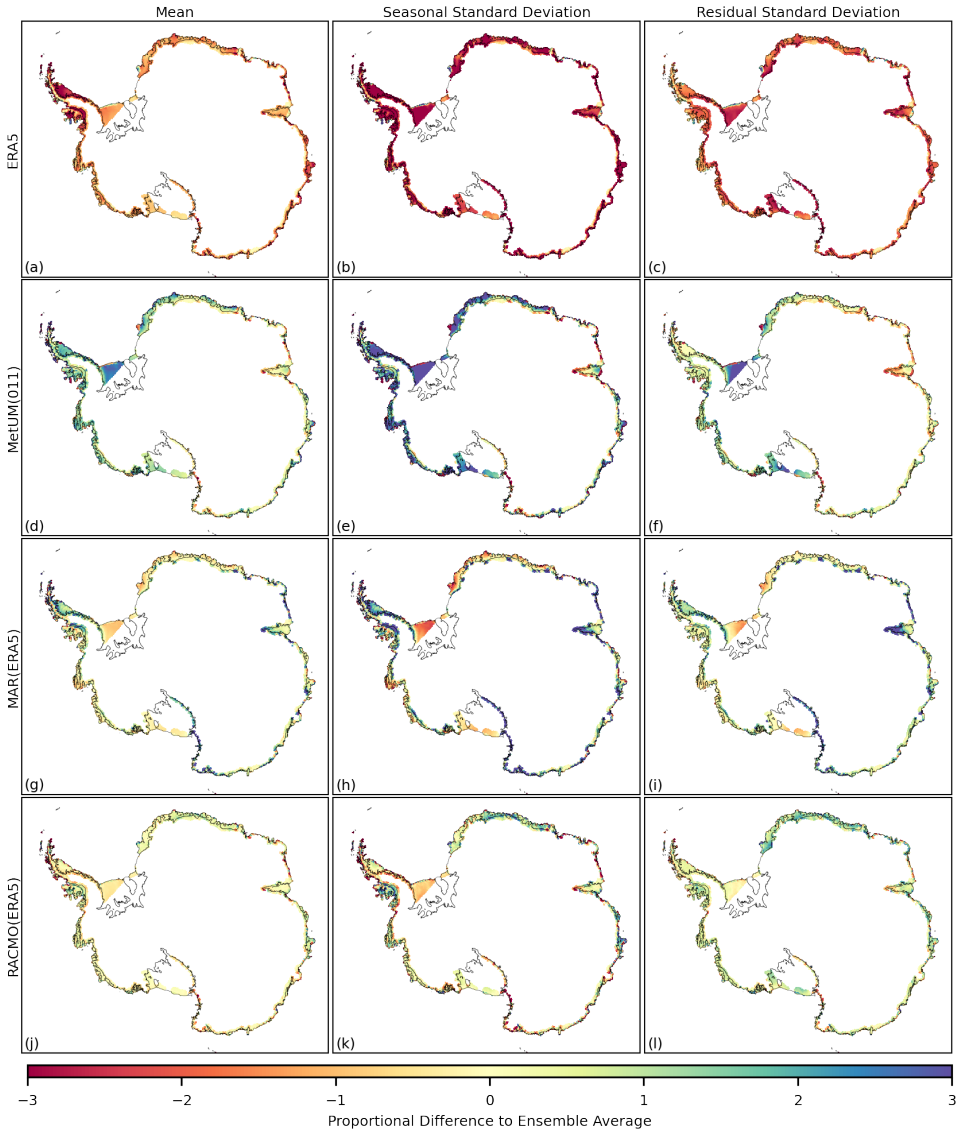



Figure 2.6: The difference to the ensemble average (model-ensemble avg.) for the 1981-2018 time series of melt, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l). The ensemble includes: ERA5 (a,b,c); MetUM(011) (d,e,f); MAR(ERA5) (g,h,i); and RACMO(ERA5) (j,k,l). Differences at each grid cell are expressed as a proportion of average inter-annual variation and so do not have units. Grid-cells where the ensemble mean average monthly melt is less than 1 $mm\ w.e.\ m^{-1}$ are masked.

### 2.4.3 RMSD

The RMSD of the monthly time series is evaluated at each grid-cell for each of the 28 unique output pairs of the ensemble. For snowfall and melt, the metric is scaled at each grid-cell by the ensemble average inter-annual standard deviation, described here as the proportional RMSD value. The average is then taken across the ice sheet for each variable and results given in Fig. 2.7a-c. The average RMSD across the ice sheet provides a measure of the average deviation between the time series of 2 model outputs at each grid cell, while the average proportional RMSD gives a measure of the average relative magnitude of deviations with respect to inter-annual variability.

To evaluate what proportion of the RMSD is the result of systematic differences, the RMSD is recalculated after adjusting the time series at each grid cell to have the same mean, seasonal standard deviation and residual standard deviation ($RMSD_{Adj}$). This is taken as the ensemble average value at each grid cell. The adjusted RMSD is then averaged across the ice sheet and the percentage change computed as $((\overline{RMSD} - \overline{RMSD_{Adj}})/\overline{RMSD})$, with results given in Fig. 2.7d-f.

From Fig. 2.7 it can be seen the average values for the RMSD/proportional RMSD are significant for all variables, with upper thresholds of 3 K for near-surface air temperature and proportional values of 4 for snowfall and 10 for melt. Values are comparatively low between simulations of the same RCM but differing resolution/driving data (MetUM(044)-MetUM(011), MAR(ERAI)-MAR(ERA5) and RACMO(ERAI)-RACMO(ERA5)) when compared with differences present between the different RCMs. For melt, ERA-Interim has noticeably higher values of proportional RMSD compared to the other models, while for snowfall and temperature differences are less pronounced but the two simulations from MAR show higher average proportional RMSD for snowfall compared with the other models.

The percentage change in RMSD/proportional RMSD after adjusting for equal means and seasonal/residual standard deviations is significant for all variables, as shown in Fig. 2.7. Upper thresholds on the percentage reduction are 50%, 70%

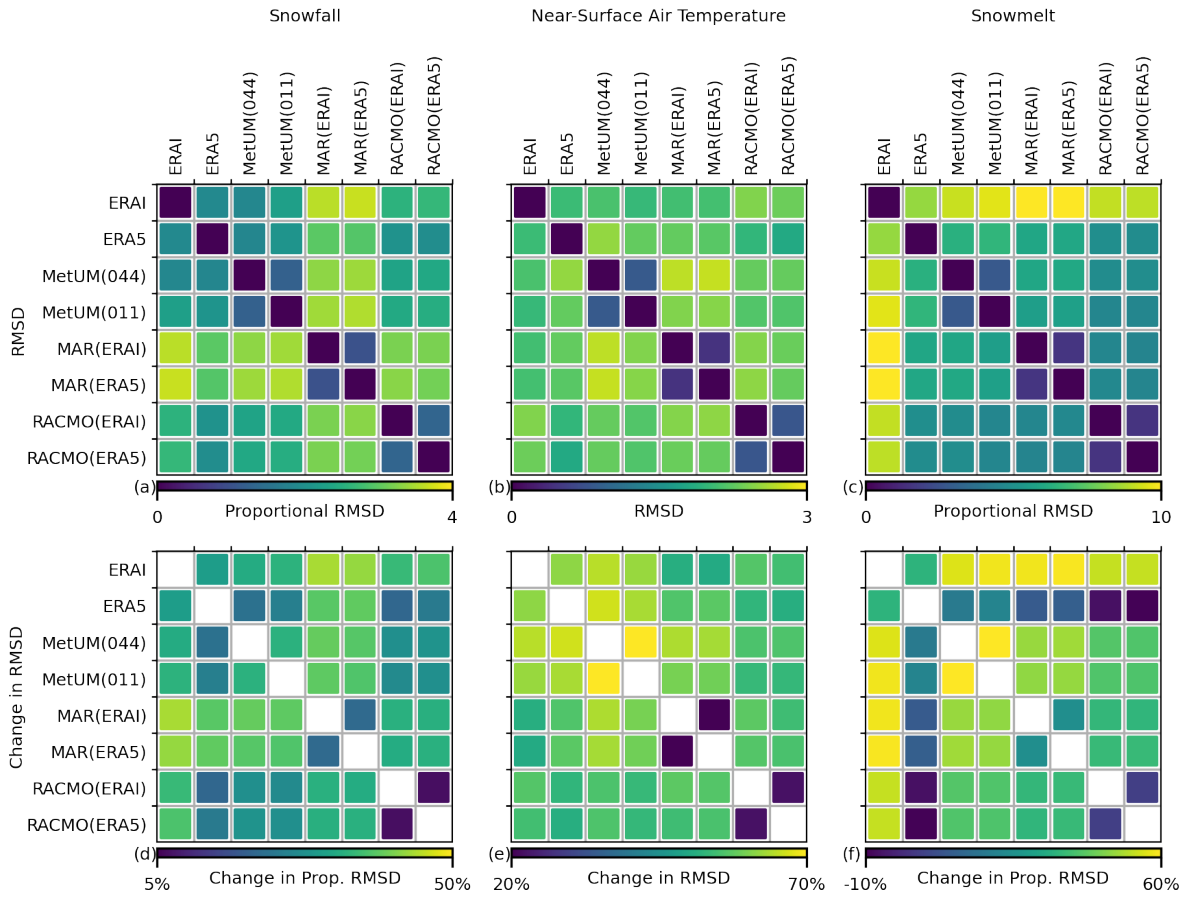



Figure 2.7: The RMSD/proportional RMSD for snowfall (a), near-surface air temperature (b) and melt (c) between models averaged over the ice-sheet. It is noted for melt that grid-cells where the ensemble average is less than than 1 $mm\ w.e.\ m^{-1}$ are excluded. After adjusting the mean and seasonal/residual standard deviations of all outputs to the ensemble average the percentage reduction in RMSD/proportional RMSD is plot for snowfall (d), near-surface air temperature (e) and melt (f).

and 60% for snowfall, near-surface air temperature and melt respectively. For melt, the most significant reductions are for ERA-Interim, while ERA5 shows the least significant reductions with proportional RMSD actually increasing between ERA5 and the RACMO products. Across the variables it can be seen that the percentage reduction in RMSD between the high/low resolution MetUM simulation pairs is of greater magnitude than reductions between the two ERA-Interim/ERA5 driven

RACMO pairs and two ERA-Interim/ERA5 driven MAR pairs. An interesting feature shown for NST is that the percentage change in RMSD shown between the MetUM simulations is significantly higher in magnitude than shown between the MAR and RACMO simulations (Fig. 2.7e). This indicates that the impact of the higher-resolution MetUM simulation is primarily on the systematic qualities of the time series rather than on the occurrence and movement of climatic events.

## 2.5 Discussion

The results presented in this paper show that for all variables studied, when considered across the entire ice sheet, the outputs that came from the same model (MetUM(011/044), MAR(ERAI/ERA5), RACMO(ERAI/ERA5)) exhibit the highest correlations in the time series as well as the smallest systematic differences and RMSDs. This is despite significant differences in resolution between the MetUM runs, which span the highest and lowest resolution RCM simulations made available from the Antarctic-CORDEX project, as well as significant differences in the driving data for the two MAR and RACMO runs. Note that, although ERA5 is an update to ERA-Interim, the results in Table 2 and in section A.2 of the appendix show that the magnitude of systematic differences in the mean and standard deviations between the reanalysis datasets are of similar or greater magnitude to that of differences between different RCM outputs (Fig. 2.4, 2.5 and 2.6). Updates in the model physics and assimilation techniques used by ERA5 (Hersbach et al., 2020) compared to ERA-Interim are hypothesised to be the primary reason for large-scale differences in snowfall and near-surface air temperature identified between the reanalysis outputs. While EAR-Interim and ERA5 exhibit large differences in their DEMs (Fig. A.5), it is argued in section 2.5.1 that differences in DEM are not primary contributors to systematic differences in the models output. The particularly significant difference (over an order of magnitude) for ice-sheet-wide melt between ERA-Interim (15.5 GT) and ERA5 (1.1 GT) is

hypothesised to be primarily due to an updated surface scheme (HTESSEL) used in ERA5 that allows run-off (Balsamo et al., 2009).

Results therefore suggest, that differing resolution and driving data are not primary contributors to large-scale spatial variability across the ensemble. Similarity in the spatial and temporal patterns between Antarctic-wide outputs of the same RCM with different driving data agrees with findings from (Agosta et al., 2019), where outputs from MAR are compared with differing reanalysis driving data from ERA-Interim, JRA-55 and MERRA-2. Similarity in results aggregated over the ice sheet (Table 2.2) between Antarctic-wide outputs of the same RCM with different driving data agrees with findings from Mottram et al., 2021 where SMB for two simulations of differing resolution (12/50 km) for the RCM HIRHAM5 are compared against other RCMs. At finer, more localised scales differing resolution is shown to create significant differences in the mean and seasonal/residual standard deviations for the monthly time series of each variable, see Fig. A.2, A.3 and A.4 (d-f) in the appendix that show direct comparisons between the high- and low-resolution MetUM simulations.

The magnitude of differences in snowfall and near-surface air temperature due to resolution are greatest over regions of sharply varying topography, such as: the Transantarctic mountains; the coastal slopes of the ice sheet; and the Antarctic Peninsula. The representation of atmospheric processes occurring over mountainous regions including foehn winds that occur over the Antarctic Peninsula and katabatic winds occurring over the coastal slopes of East Antarctica are known to be resolution dependent (Orr et al., 2014; Heinemann and Zentek, 2021; Orr et al., 2021). Foehn and katabatic winds have been shown to impact climate over ice shelves, which are often in close vicinity of steep terrain, and are an important driver of surface melt (Bromwich, 1989; Cape et al., 2015; Lenaerts et al., 2017; Datta et al., 2019; Elvidge et al., 2020). In Fig. A.3d the difference in the mean near-surface air temperature, due to resolution, extends over ice shelves such as the interior of the Amery ice shelf, which is a well-known katabatic wind confluence zone (Parish and

Bromwich, 2007). Despite this influence of resolution on the climatology over ice shelves, greater systematic differences in melt shown in Fig. 2.6 compared with Fig. A.4(d-f) indicate the potentially more significant importance of differences in surface schemes across the particular ensemble of RCMs studied. It is expected that even at 12 km resolution climatically important terrain-induced atmospheric processes, such as foehn/katabatic winds, are not being realistically resolved as is shown in Orr et al., 2021 where output from the MetUM RCM at 4 km, 1.5 km and 0.5 km during a foehn wind event on the Larsen C ice shelf show no obvious convergence towards observations during the event.

The same-model RCM simulations in the ensemble, as well as having identical model physics, parametrisation and tuning, also share factors such as the domain specification, ice mask applied, digital elevation model and boundary conditions. The relative contribution of these additional factors is explored in section 2.5.1 and from this it is argued that the joint influence of choices in model physics, parametrisation and tuning is the primary factor influencing large-scale variability across the ensemble. The impact of parameter tuning is discussed in Gallée and Gorodetskaya, 2008 where it is shown that adjusting the relative contribution of snow particles compared to ice particles in MAR's radiative scheme has a significant impact on near-surface air temperature. A higher relative contribution of snow particles leads to greater flux in long-wavelength downwards radiation. In addition to exploring the relative contribution of different factors to the large spatial scale variability in the ensemble, in section 2.5.2 specific features of the variability that are mentioned in results section 2.4 are discussed and the nature of variability for different variables, regions and time-scales is examined.

### 2.5.1 Contribution to Variability from the choice of Domain, Ice Mask, Boundary Conditions and DEM

The exact spatial domains differ between the RCM simulations as shown in Fig. 2.1. However, the spatial domain for all RCM simulations examined is Antarctic-wide

and domain boundaries all exist over the ocean, implying there should be no strong local forcing at any of the boundaries. The effect of increasing domain size over the ocean on the output of simulations from MAR over the Greenland ice sheet has previously been studied and found to not significantly impact results over the ice sheet (Franco et al., 2012). In general, the domain size should be great enough such that the buffer zone, in which boundary conditions are applied, does not intersect the region of interest, which in this case is the Antarctic ice sheet. It is found that for the MetUM(044) run, the buffer zone intersects some areas of the periphery of the ice sheet, shown clearly in Fig. A.2d. Despite this, it can be seen that effects are localised to the buffer zone boundary, and that even for the regions of the ice sheet that intersect this the relative impact on systematic differences appears less significant than other factors explored. Overall it is assumed that, for the ensemble of RCM simulations studied, differences in the domains does not have a significant effect on the model output for surface climatology over the ice sheet.

As well as having differences in the outer domain boundaries, the different models also have slight differences in the specified boundaries of the ice sheet due to different coordinates and ice masks used. This creates edge effects at the periphery of the ice sheet, particularly noticeable for melt in for example Fig. 2.5d at the edges of the Filchner-Ronne and Ross ice shelves. It is shown in Mottram et al. (2021) and Hansen et al. (2022) that these edge effects, due to inconsistent ice masks, can have a significant impact on the total estimated SMB over the ice sheet. In this paper, although ice sheet wide totals are computed (Table 2.2), the focus is primarily on evaluating variability in the time series at each 12 km grid cell after regridding products to a common high-resolution grid. Results for spatial maps of differences for melt are masked using an ice-sheet only mask from Depoorter et al. (2013), which is found to exclude the most significant edge effects from areas where low-resolution models overestimate the extent of the ice sheet after regridding. The same mask is applied before calculating average correlations and RMSDs also reducing the impact of edge effects. Results presented and discussed here, particularly regarding large-

scale spatial patterns, are therefore assumed to not be significantly impacted by the different ice masks used in the ensemble of simulations.

Another important consideration when comparing RCM simulations is how the method of applying boundary conditions varies across the ensemble. In particular, although all RCMs examined are nudged at the boundaries within buffer zones, MAR also uses spectral nudging and RACMO indiscriminate nudging that constrains the large scale circulation in the interior of domain, while the MetUM instead uses a re-initialisation procedure. Spectral nudging involves applying the relaxation technique throughout the interior of the domain to the long wavelength components of the climate model fields (Storch et al., 2000). This constrains the large-scale climatology of the RCM output to that of the driving data, while allowing value-added by the RCM in the small-scale features. The indiscriminate nudging, upper-air relaxation (UAR) technique used in the RACMO simulations also improves consistency with large-scale forcing fields, although results in partially suppressing small-scale patterns in the RCM absent from the driving data. The same goals of spectral and indiscriminate nudging are aimed to be achieved with the re-initialisation of the MetUM throughout the domain every 12 h, as discussed in Sect. 2.4.

As mentioned previously, the fact that the driving data appears to not be a primary contributor to large-scale spatial variability across the ensemble of outputs also suggests the specific approaches of spectral nudging, indiscriminate nudging or re-initialisation are not one of the main features contributing to spatial variability in the mean and the seasonal and residual standard deviations of snowfall, near-surface air temperature and melt. It is noted, however, that in general the MetUM simulations, rather than the MAR and RACMO simulations, show slightly higher correlation to the reanalysis driving data across the ice sheet for the monthly time series of snowfall and surface temperature (Fig. 2.2), indicating that the re-initialisation procedure potentially constrains the output across the ice sheet more than spectral or indiscriminate nudging.

The differences between DEMs used across the ensemble are plot in Fig. A.5 in the appendix. The elevation profiles can be split into three main groups: the coarse elevation profile of ERA-Interim (Fig. A.5a); the elevation profiles of ERA5 and the MetUM high- and low-resolution runs (Fig. A.5b,c,d); and the elevation profiles of MAR and RACMO (Fig. A.5e,f). Differences in the DEMs do not mirror the systematic differences shown in section 2.4.2. For example, while MAR and RACMO share comparatively similar DEMs, the models do not share similar patterns in systematic differences (Fig. A.2, A.3 and A.4). This indicates differences in the DEMs are not primary contributors to systematic differences in the models output, which is further supported by results displayed in Fig. A.6 where weak linear correlation is found between differences in elevation and differences in mean near-surface air temperature.

In this section, features including the domain specification, ice mask applied, digital elevation model and boundary conditions applied are argued to not be the primary contributors responsible for the large-scale systematic differences between the ensemble of model outputs. This result, in addition to the previously discussed secondary contributions of resolution and driving data towards large-scale differences, by way of elimination gives that the joint influence of choices in model physics, parametrisation and tuning is the primary factor influencing large-scale systematic differences across the ensemble.

### 2.5.2 Specific Features of the Variability

Specific features in the variability, identified and mentioned in section 2.4, are discussed here. In section 2.4.1 it is mentioned that for melt there is a clear divide in the average correlation in the residual component of the time series between reanalysis datasets compared with RCMs (Fig. 2.2). That is the correlations between different RCMs are greater than between reanalysis datasets and RCM outputs. This is not the case for snowfall and near-surface air temperature, suggesting the divide in correlation for melt is primarily due to differences in the

sophistication and polar specific tuning of the surface schemes used for the RCM simulations and the global reanalysis products. It is shown in Hansen et al. (2021) that the subsurface scheme and the handling of layers within the scheme can have a significant impact on melt.

In section 2.4.1 it is also shown that, particularly for snowfall and melt, the median correlation between the outputs is strongly dependent on the specific region and topography. For melt, three regions are highlighted that show low correlation including the Ronne ice shelf, the ice shelves bounding Victoria Land, and the interior of the Amery ice shelf. In the case of the Ronne ice shelf, the low correlation in melt is due to relatively low average melt occurring over the region, so fluctuations away from no melt are small and erratic. Low correlation over ice shelves bounding Victoria Land is expected to be caused by a combination of their fine scale and the sharply varying topography in the region, making the climatology around them difficult to resolve with the resolution available in the climate models. Finally, the pattern of low correlation around regions such as the interior of the Amery ice shelf is likely the result of atmospheric processes difficult to represent fully in the models, for example: katabatic winds, driven by gravity, flowing from the interior of the ice sheet to the exterior down elevation channels have a significant impact on the climate on the Amery ice shelf, particularly near the grounding line (Lenaerts et al., 2017).

As with for correlation, the systematic differences shown between the outputs in the ensemble vary depending on the region and topography, see section 2.4.2. This is true at large and small spatial scales and for all variables. An example of a dependency at large scale is in Fig. 2.4g where MAR shows a significant positive difference in the 40-year mean monthly snowfall relative to the other outputs over the majority of the ice sheet and a significant negative difference over the majority of the surrounding ocean. In the case of MAR this is hypothesized to be due to a couple of reasons: MAR is forced at the boundaries by humidity and needs time to transform this into precipitation; MAR allows precipitation to be advected through

the atmospheric layers until reaching the surface. The advection of precipitation in MAR through each atmospheric layer, in comparison to the instantaneous depositing of precipitation by RACMO, leads to increased snowfall towards the interior of the ice sheet, previously identified in Agosta et al., 2019.

In section 2.4.3 the RMSD between each model pair, calculated at each grid cell and then averaged across the ice sheet, is presented. This metric of average deviation is dependent on the temporal correlation and presence of systematic differences between the outputs. High values of proportional RMSD for melt, shown in Fig. 2.7 are the result of relatively low temporal correlations between models as well as relatively high systematic differences. It is noted that for melt, despite there being a clear divide in temporal correlations between reanalysis datasets and RCMs (Fig. 2.2), the RMSD between ERA5 and the RCMs is of comparable magnitude to values between the RCMs. This is due to particularly low values of total melt exhibited from ERA5 (Table 2.2) and resulting low magnitude fluctuations of melt. The percentage change in RMSD, after adjusting the mean and seasonal/residual standard deviations of all outputs to the ensemble average, further supports this as it can be seen for melt that ERA5 exhibits the smallest reduction in RMSD after adjustments (Fig. 2.7). The converse of the above argument is true for ERA-Interim, which shows particularly high values of total melt and so particularly significant values of RMSD and of percentage reductions after adjusting the mean and seasonal/residual standard deviations. Differences in the model outputs for melt across Antarctica remain significant with respect to inter-annual deviations even after adjusting for systematic differences in the mean and standard deviation of the seasonal/residual components of the time series, indicating the importance of improving surface schemes across the models.

## 2.6 Conclusion

The spatial nature and magnitude of variability present in an ensemble of current, state-of-the-art Antarctic-wide RCM outputs and global reanalysis datasets is examined for snowfall, near-surface air temperature and melt. This is done at a 12 km grid level, rather than across elevation bins, which reveals significant spatial patterns in correlation and systematic differences in the mean and seasonal/residual standard deviation. Time series decomposition is used to split comparisons across an approximately inter-annual trend component, a periodic seasonal component and a monthly residual component, which is useful for impact assessments where knowledge of variability in the climate data across different time-scales and climate drivers is needed.

It is found that the RCM outputs and reanalysis datasets show high correlation for the monthly time series of snowfall and surface temperature across the majority of Antarctica and the bounding Southern Ocean. Despite this, there exists significant differences, with respect to both magnitude and spatial scale, in the mean and seasonal/residual standard deviations of the time series. In addition, high RMSD between the outputs is found for all variables and is particularly significant for melt with respect to the proportional values relative to annual fluctuations. The primary sources of large-scale, systematic differences between the simulations, for all variables and components, are identified as deriving from differences in the model dynamical core, the surface scheme and the parametrisation and tuning. Differences in driving data, resolution, domains, ice masks, DEMs and boundary conditions are identified as having a secondary contribution. On local, fine spatial scales the relative contribution from different factors is more complex and differences in for example resolution are shown to have a more significant impact.

The variability in snowfall, near-surface air temperature and melt shown is expected to introduce significant uncertainty in estimates of the ice shelf stability with regard to collapse events, which as discussed may have an important contribution to future SLR estimates. It is suggested that the magnitude and

scale of systematic differences across the ensemble precludes the direct use and interpretation of individual outputs in impact assessments regarding ice shelf collapse. Results show that removing systematic differences between the ensemble of outputs, significantly reduces the average RMSD. Therefore, as concluded in Mottram et al. (2021), there is an importance on observational campaigns to correct for systematic differences. Improved coverage and quality of observations will provide greater constraints with which to both tune and update the model physics and parametrisations, as well as to use and reduce uncertainties in post-processing bias correction. In addition, further development of RCMs, with particular focus on improvements to the performance of surface schemes over regions of high-melt, is needed to reduce uncertainties around collapse events and future SLR. Finally, it is suggested that further development of sophisticated techniques for bias correction are needed, that are compatible with sparse observations and make use of factors such as the spatial distribution of variability identified in this paper.

# Chapter 3

# Bias Correction of Climate Models using a Bayesian Hierarchical Model

Climate models, derived from process understanding, are essential tools in the study of climate change and its wide-ranging impacts on the biosphere. 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. Bias correction approaches using observational data aim to address this challenge. However, approaches are typically criticised for not being physically justified and not considering uncertainty in the correction. These aspects are particularly important in cases where observations are sparse, such as for weather station data over Antarctica. This paper attempts to address both of these issues through the development of a novel Bayesian hierarchical model for bias prediction. The model propagates uncertainty robustly and uses latent Gaussian process distributions to capture underlying spatial covariance patterns, partially preserving the covariance structure from the climate model which is based on well-established physical laws. The Bayesian framework can handle complex modelling structures and provides

an approach that is flexible and adaptable to specific areas of application, even increasing the scope of the work to data assimilation tasks more generally. Results in this paper are presented for one-dimensional simulated examples for clarity, although the method implementation has been developed to also work on multidimensional data as found in most real applications. Performance under different simulated scenarios is examined, with the method providing most value added over alternative approaches in the case of sparse observations and smooth underlying bias. A major benefit of the model is the robust propagation of uncertainty, 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.

## 3.1 Introduction

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 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, 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., 2022b). 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 representation 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 model development and tuning challenging (Hourdin et al., 2017). End users are typically interested in only a narrow aspect of the output (e.g. possibly only one or two variables), which the model is unlikely to be specifically tuned for. Post-processing, bias correction techniques allow improvements to the consistency, quality and value of 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). Different end users are focused on different types of systematic errors, whether that's errors in the mean climatology, the multi-year trends or in other features of the output such as the covariance structure. This paper follows a common approach to focus on systematic errors in the parameters that describe the probability density function (PDF) at each site. Further, detailed discussion of this is given in Sect. 3.2 as are approaches to bias correction within this context.

One of the fundamental issues often attached to bias correction is the lack of justification based on known physical laws and process understanding (Ehret et al., 2012; Maraun, 2016). Transfer functions are derived that are applied to the climate data to improve some aspect of consistency with observations, such as the mean in for example the delta method (Das et al., 2022) or the overall PDF in the case of

quantile mapping (Qian and Chang, 2021). 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 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 (R. M. DeConto and Pollard, 2016).

In this paper a fully Bayesian approach using a hierarchical structure and latent Gaussian processes (GP) is proposed for bias correction, discussed in detail in Sect. 3.3. Parameter uncertainties are propagated through the model and the underlying covariance structure is derived both from observations and the climate model 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 is of key importance and where including data from all sources during inference is particularly valuable. In the method, climate model output is assumed to be generated from two underlying and independent stochastic processes, one relating to the true underlying field of interest (that also generates the in situ observations) and one that generates the bias present in the climate data. The aim is to separate these two processes and to infer their covariance structures. Posterior predictive estimates of the true underlying field across the region can then be made, which in turn can be used for bias correction. The ability of the model in doing this depends on factors such as the density of observations and the relative smoothness of the truth and bias components. Simulated data is used to test the performance under scenarios with differing data density and latent covariance length scales, with results and discussion presented in Sect. 3.4.

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 Yali Wang and Chaib-draa, 2017. In addition to the main results presented in Sect. 3.4, 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 B.4. 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.

The model is developed in a flexible Bayesian framework, where adjustments 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 model 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.

## 3.2 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. 3.2.1. 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. 3.2.2. 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. 3.2.3.

### 3.2.1 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 $(\mu_Z, \mu_Y)$ and the standard deviations $(\sigma_Z, \sigma_Y)$. This function can be defined in different ways, such as the difference $b(\mu_Z, \mu_Y) = \mu_Z - \mu_Y$ or the ratio $b(\sigma_Z, \sigma_Y) = \sigma_Z/\sigma_Y$.

### 3.2.2 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. 3.3.3. Bias is then defined by some discrepancy function of these spatially varying parameters, such as $b(\mu_Z(s), \mu_Y(s)) = \mu_Z(s) - \mu_Y(s)$.

### 3.2.3 Bias Correction

Climate model simulations are useful for impact assessments due to the their comprehensive spatiotemporal coverage and ability to predict climate change signals, although bias in the output precludes direct interpretation (Ehret et al., 2012). Bias correction involves using observational data to predict and then reduce bias in the climate model output. Techniques vary in focus and complexity, from using observations to apply an adjustment to the mean in the case of the delta change method (Das et al., 2022), to adjusting the whole PDF of the climate model output in the case of quantile mapping (Qian and Chang, 2021), and to approaches that use Generalised Additive Models (GAMs) to approximate transfer functions between the climate data and the observed values (Beyer et al., 2020). Various studies compare relative performance between methods (Teutschbein and Seibert, 2012; Räty et al., 2014; Beyer et al., 2020; Mendez et al., 2020). Typically all approaches fail to capture uncertainty and explicitly model the spatial dependency between points and of the processes under study, not considering correlation between nearby measurements.

The approach proposed in this paper combines the use of a Bayesian hierarchical model for predicting bias across the region, while explicitly modelling underlying spatial structures and capturing uncertainty, with the established approach of quantile mapping for applying the final correction to the climate model output. The details of the Bayesian hierarchical model are given in Sect. 3.3. The data is treated as generated from latent stochastic processes, as in Sect. 3.2.2, and estimates are made for parameters of the site-level PDFs of the observations and climate model output across the domain. This allows quantile mapping at each grid point of the climate model output to then be applied. 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 and then mapping the value onto the corresponding value of the equivalent percentile of the PDF estimated from observational data. 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 $\hat{\boldsymbol{z}}_{\boldsymbol{s_i,j}} = F^{-1}_{Y_{s_i}}(F_{Z_{s_i}}(\boldsymbol{z}_{s_i,j}))$, where $F_{,}s_i$ represents the CDF at the site $s_i$. 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 the PDF parameters at each site 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 approach presented builds upon that of Lima et al., 2021 where the spatial dependency between in situ observations is modelled when estimating the PDF parameters and then quantile mapping applied. In Lima et al., 2021 the spatial structure in the climate model output is not explicitly modelled though, whereas in this paper the shared spatial structure between the observations and the climate model output is accounted for when conditioning, discussed further in Sect. 3.3.3. This is particularly important for regions of sparse in situ observations and results in conserving some of the information available from the climate model on the true spatial variation of parameters. Incorporating and partially conserving the spatial covariance structure of the climate model is desirable as it is derived from well established physical laws and reflects the assumption that the climate model itself provides skill in assessing the site level parameters across the domain (Ehret et al., 2012). Additionally, in this paper the hierarchical model is embedded in a fully Bayesian framework and uncertainty in the PDF parameter estimates is propagated through the quantile mapping procedure to the bias corrected climate time series, which is missing from Lima et al., 2021.

## 3.3 Bias Prediction Methodology

The goal of the methodology developed and presented in this paper is to evaluate the bias in the climate model output across the domain in a framework that captures uncertainty robustly and that preserves information available from both the in

situ observations and climate model output on underlying spatial structures. The resulting predictive bias can be coupled with known bias correction methods, such as quantile mapping, with the benefits of uncertainty quantification and inherited spatial structure. The overarching approach is summarised in Sect. 3.3.1 with a specific example given in Sect. 3.3.2. The properties of GPs are discussed in Sect. 3.3.3.

### 3.3.1 Model Overview

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 ($\mathcal{S}$,$\mathcal{T}$). A random variable is attributed to each spatiotemporal coordinate $(Y(s,t), Z(s,t))$. The data observed is then considered a realisation of the joint distribution over a finite set of random variables across the domain.

For the purpose of bias prediction, the random variables are treated as independent and identically distributed across time, such that for a given location $s$, $Y(s,t) \mid \boldsymbol{\phi}_Y(s) \overset{i.i.d.}{\sim} \mathcal{F}_Y(\boldsymbol{\phi}_Y(s))$ and $Z(s,t) \mid \boldsymbol{\phi}_Z(s) \overset{i.i.d.}{\sim} \mathcal{F}_Z(\boldsymbol{\phi}_Z(s))$, where $\mathcal{F}_Y(\cdot)$ and $\mathcal{F}_Z(\cdot)$ represent some generic site-level distributions with spatially varying vector parameters $\boldsymbol{\phi}_Y(s)$ and $\boldsymbol{\phi}_Z(s)$. This follows from evaluating the time-independent component of the climate model bias. Consider evaluating bias in the values of January midday near-surface temperature over a region. While the values of nearby days are clearly dependent on each other, since focus is on evaluating time-independent bias, the time component of the data is dropped and only the marginal distribution considered. The marginal distribution in this case gives the probability of a certain value of January midday temperature just based on location and could for example be approximated as normal with mean and variance parameters, as mentioned in Sect. 3.2.1. In the case of other climatological fields such as rainfall

a more appropriate distribution might be that of a Bernoulli-Gamma with its own collection of parameters, as used in Lima et al., 2021. Caution in this treatment should be applied in cases where, for example, the observational site only has a limited number of days of data and these are bunched around the same relatively short time period, since this period is unlikely to be representative of the time series as a whole.

The disparity between the spatially varying parameters $\boldsymbol{\phi}_Y(s)$ and $\boldsymbol{\phi}_Z(s)$ in the site-level marginal distributions serves as a measure of bias. Specifically, as in Sect. 3.2.2, the bias for each parameter $\phi_i$ can be defined by some discrepancy function $\phi_{B,i}(s) = b_i(\phi_{Y,i}(s), \phi_{Z,i}(s))$. Alternatively, the parameters associated with the climate model output $\phi_{Z,i}(s)$ can be defined as a function of the unbiased parameters $\phi_{Y,i}(s)$ and a latent bias function $\phi_{B,i}(s)$. In this paper an additive relationship is used, such that $\phi_{Z,i}(s) = \phi_{Y,i}(s) + \phi_{B,i}(s)$. Additionally, the bias $\phi_{B,i}(s)$ is considered independent of the value of $\phi_{Y,i}(s)$. To estimate the parameters across the domain and quantify the bias, these spatially varying parameters are modelled as spatial stochastic processes with hyper-parameters $\boldsymbol{\theta}$. It's important to note that since the collection of parameters may not necessarily all belong to the same parameter space, their representation can be standardized by applying a link function transformation to some of the parameters $\tilde{\phi}_i = h_i(\phi_i)$ so that all parameters can be represented by the same family of stochastic processes. In the methodology presented in this paper the family of Gaussian processes is used to model spatial dependencies. The full hierarchical model is then the following, with dependencies illustrated through the plate diagram shown in Fig. 3.1.

$$Y(s,t) \mid \boldsymbol{\phi}_Y(s) \overset{i.i.d.}{\sim} \mathcal{F}_Y(\boldsymbol{\phi}_Y(s)) \quad (3.1)$$

$$Z(s,t) \mid \boldsymbol{\phi}_Z(s) \overset{i.i.d.}{\sim} \mathcal{F}_Z(\boldsymbol{\phi}_Z(s)) \quad (3.2)$$

$$\begin{cases} \phi_{Z,i}(s) = \phi_{Y,i}(s) + \phi_{B,i}(s) \quad \text{if correct support,} \\ \phi_{Y,i}(s) \perp\!\!\!\perp \phi_{B,i}(s) \\ \phi_{Y,i}(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\phi_{Y,i}}) \\ \phi_{B,i}(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\phi_{B,i}}) \\ \\ \tilde{\phi}_{Z,i}(s) = \tilde{\phi}_{Y,i}(s) + \tilde{\phi}_{B,i}(s) \quad \text{if link function required for correct support.} \\ \tilde{\phi}_{Y,i}(s) \perp\!\!\!\perp \tilde{\phi}_{B,i}(s) \\ \tilde{\phi}_{Y,i}(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\tilde{\phi}_{Y,i}}) \\ \tilde{\phi}_{B,i}(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\tilde{\phi}_{B,i}}) \end{cases} \quad (3.3)$$

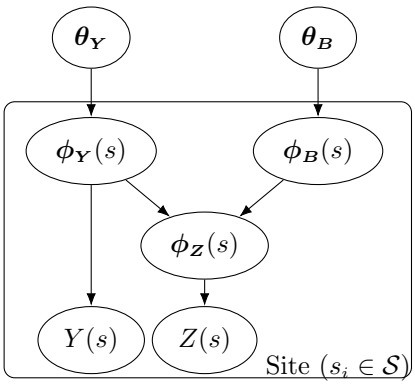



Figure 3.1: Plate diagram showing a generic version of the full hierarchical model. The in-situ observations $Y$ and climate model output $Z$ are generated from distributions with the collection of parameters $\boldsymbol{\phi}_Y$ and $\boldsymbol{\phi}_Z$ respectively. The parameters $\boldsymbol{\phi}_Z$ are modelled as some function of the parameters $\boldsymbol{\phi}_Y$ and some independent bias $\boldsymbol{\phi}_B$. The parameters $\boldsymbol{\phi}_Y$ and the corresponding bias $\boldsymbol{\phi}_B$ are each themselves modelled over the domain as generated from Gaussian processes with hyper-parameters $\boldsymbol{\theta}_Y$ and $\boldsymbol{\theta}_Z$.

Gaussian processes naturally introduce spatial structure into the parameters and enable inference with misaligned data. Predictive estimates of the PDF parameters for each data source can be made for any set of locations across the domain. Estimates at the climate model output locations are needed for bias correction, while there's also the possibility to compute estimates at higher resolution and combine with a downscaling approach, as in Lima et al., 2021. Additional added benefits of GPs include properties that facilitate inference, for example the additive property where the sum of two independent GPs is itself also represented as a GP. More details following on from this and the application of GPs in the methodology is provided in Sect. 3.3.3.

Inference on the parameters of site-level and spatial distributions of the model given the data is applied in a Bayesian hierarchical framework, where parameters of the model are 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 from the posterior. An important advantage of this framework is it allows flexible extensions of the model while automatically maintaining robust uncertainty estimation. This results in the model being applicable to a wide range of problems and domains, especially important for correcting climate model output since there's a broad range of users interested in different variables and domains with varying levels of complexity.

### 3.3.2 Specific Model Example

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

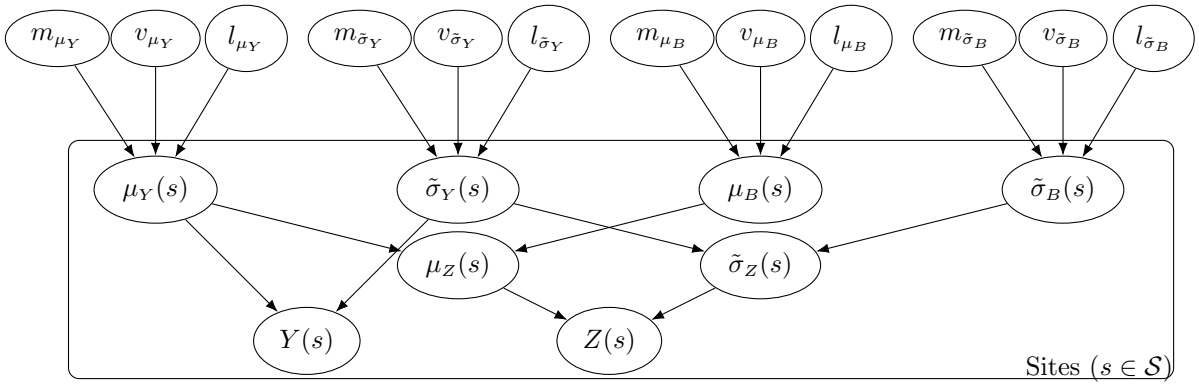
$\{Z(s,t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ respectively. To evaluate bias the time-independent marginal distributions are taken and the data treated as realisations from the spatial stochastic processes $\{Y(s) : s \in \mathcal{S}\}$ and $\{Z(s) : s \in \mathcal{S}\}$. Temperature is known to 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 output, 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. 3.3.3.

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.3.2 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 $\mu_Y$ and $\tilde{\sigma}_Y$ at all in situ observation locations and all climate model output locations, as well as realisations from the posterior of $\mu_B$ and $\tilde{\sigma}_B$ 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 $[\mu_Y, \tilde{\sigma}_Y, \mu_B, \tilde{\sigma}_B]$ 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 climate model output. The parameters $\mu_Z$ and $\tilde{\sigma}_Z$ can then be computed and quantile mapping 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 3.2: Plate diagram with directed acyclic graph showing the full hierarchical model for the case where the site-level distributions are assumed normal with parameters $\mu$ and $\sigma$. The distribution of these parameters across the domain is modelled with Gaussian processes.

### 3.3.3 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. (3.4) assumes a second order polynomial in two predictors, where the predictors $x_1(s)$ and $x_2(s)$ could be elevation and latitude.

$$m(s) = b_0 + b_1 \cdot x_1(s) + b_2 \cdot x_2(s) + b_3 \cdot x_1(s) \cdot x_2(s) + b_4 \cdot x_1(s)^2 + b_5 \cdot x_2(s)^2 = \boldsymbol{x}(s)^T \boldsymbol{\beta} \quad (3.4)$$

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. (3.5) 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. (3.6), 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. 3.3, an example of how the covariance decays with distance is given for the RBF kernel and realisations of a conditioned GP with the equivalent kernel are illustrated.

$$k_{RBF}(\boldsymbol{s}, \boldsymbol{s}') = v \cdot exp\left(-\frac{d(\boldsymbol{s}, \boldsymbol{s}')^2}{2l^2}\right) \quad (3.5)$$

$$d(\boldsymbol{s}, \boldsymbol{s}') = \sqrt{(x_3(\boldsymbol{s}') - x_3(\boldsymbol{s}))^2 + (x_4(\boldsymbol{s}') - x_4(\boldsymbol{s}))^2} \quad (3.6)$$

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 methodology presented in this paper and in the simulated examples given in Sect. 3.4. The validity of the stationarity assumption should be assessed on an application basis, with factors such as complex topography contributing to non-stationarity.

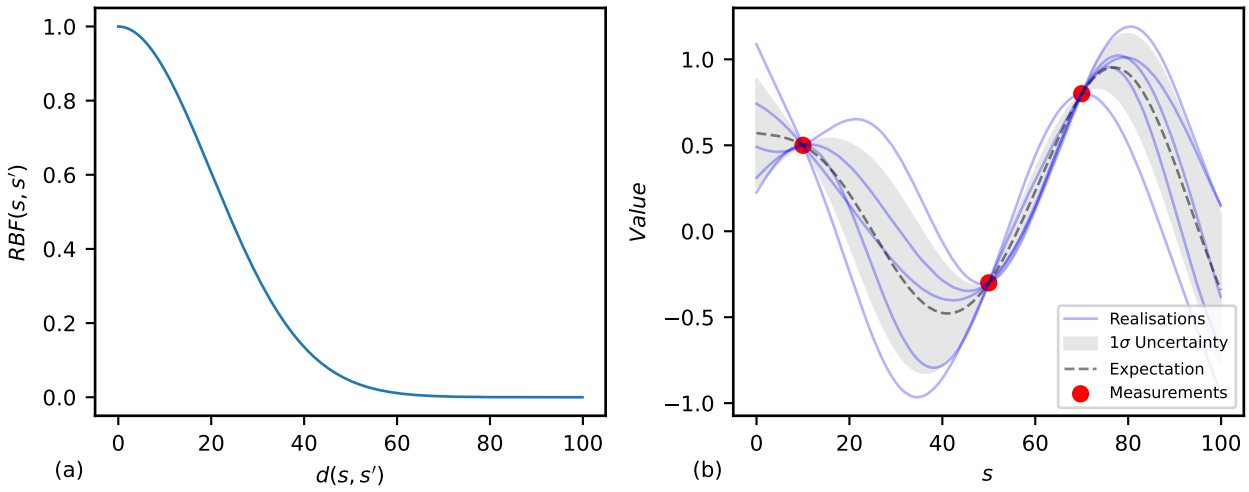



Figure 3.3: A) Values of the RBF function with a kernel variance equal to 1 and length scale equal to 20. B) Realisations of the GP with the equivalent kernel as in A and conditioned on 3 data points. The expectation and uncertainty of the distribution are shown.

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 $\phi_Y$ and the bias $\phi_B$ 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 $\tilde{\phi}_Y$ and $\tilde{\phi}_B$ are modelled as independent and generated from GPs. Assuming an additive relationship results in an easy to define distribution for $\phi_Z$ (or $\tilde{\phi}_Z$), which is a GP where the mean and covariances are simply the sum of the values from the independent GPs:

$$m_{\phi_Z} = m_{\phi_Y} + m_{\phi_B} \tag{3.7}$$

$$k_{\phi_Z}(s, s') = k_{\phi_Y}(s, s') + k_{\phi_B}(s, s') \tag{3.8}$$

This 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. This relationship is shown graphically in Fig. 3.4. Additionally, an illustration for simulated realisations of $\phi_Y(s)$ and $\phi_B(s)$ from two underlying, independent latent GPs is provided in Fig. 3.5, where $\phi_Z(s)$ is also shown as the sum of the two realisations.

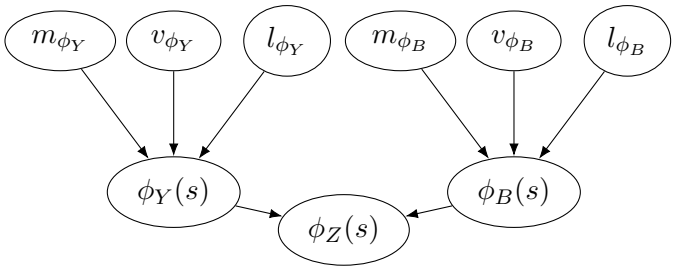



Figure 3.4: Directed acyclic graphs showing the joint dependency of the population parameters from the observations and the climate model.

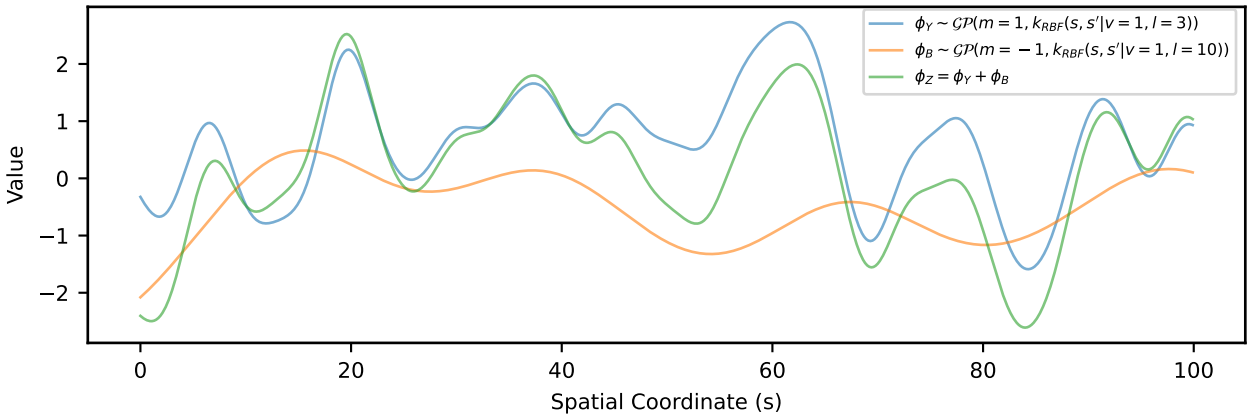



Figure 3.5: An illustration in 1 dimension of the parameters $\phi_Y$ and $\phi_B$ across some domain $0 \leq s \leq 100$. The parameters are realisations generated from GPs with different means and length scales. The parameter $\phi_Z$ is given as the sum of $\phi_Y$ and $\phi_B$.

In order to estimate the parameters relating to for example the in situ observations across the domain at some new locations $\phi_Y(\hat{s})$, conditioning is then performed on both the values of the parameter observed at the observation locations

$\phi_Y(s^*)$ and the values of the parameter for the climate model output at the locations predicted by the climate model $\phi_Z(s')$.

## 3.4 Simulated Examples

The goal of the model 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. The model additionally infers the spatial structure of these parameters and their bias. Results are presented to 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. 3.4.1) and the Bayesian hierarchical nature and uncertainty propagation (Sect. 3.4.2). 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. The steps for generating the data and the results are presented separately for each example, while the discussion of results is done together in Sect. 3.5.

Inference of the parameters of the models 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). The number of MCMC steps for warm up is taken as 1000, while the number of MCMC samples to take of parameters after warmup is 2000 and the number of chains is taken as 2. The parameters are treated as random variables with associated probability distributions. A prior distribution is set for each 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 is the updated distribution after observing

and conditioning on the data. Estimates of the parameters $\boldsymbol{\phi}_Y(s)$, $\boldsymbol{\phi}_Z(s)$ and the corresponding bias $\boldsymbol{\phi}_B(s)$ across the domain given the posterior and the observed data is then referred to as the posterior predictive.

### 3.4.1 Shared Latent Generating Processes: Non-Hierarchical Example

A non-hierarchical example is presented where direct measurements are assumed for one parameter of the PDFs for the in situ observations $\phi_Y(s)$ and for the climate model output $\phi_Z(s)$. The goal is to predict the parameter $\phi_Y(s)$ across the spatial domain using information from both the simulated in situ observations and climate model output, which are related through $\phi_Z = \phi_Y + \phi_B$. The parameters $\phi_Y(s)$ and $\phi_B(s)$ are considered independent and generated from Gaussian processes. Comparison is made to the approach of inferring $\phi_Y(s)$ just from the in situ data, as in Lima et al., 2021. The purpose of this example is to illustrate the advantage of modelling shared latent generating processes between the observational data and the climate model output, as in Fig. 3.4. Relative performance is evaluated for three alternative simulated scenarios that correspond to different possible real-world situations.

#### 3.4.1.1 Simulated Data

The simulated data in this example is generated assuming the dependency in Fig. 3.4 and the relationship $\phi_Z = \phi_B + \phi_Y$, where $\phi_Y$ and $\phi_B$ are assumed independent. The latent Gaussian process distributions that generate $\phi_Y$ and $\phi_B$ across the domain are taken with constant mean and an RBF kernel (Eq. (3.5)). The hyper-parameters of these latent distributions and the number of simulated observations are set for three scenarios, as given in Table 3.1. The prior distributions of the parameters are taken as the same for each scenario. Specifics of the data generation are given in Sect. B.2 of the appendix.

The three scenarios represent different potential real-world situations and the

data generated for each is shown in Fig. 3.6. The first scenario (Fig. 3.6a) 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 $\phi_Y$ without significant added value provided from inclusion of the climate model output during inference. The second scenario (Fig. 3.6b) 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 $\phi_Y$ across the domain since it is only afflicted by a comparatively simple bias that is easy to estimate. The final scenario (Fig. 3.6c) also involves sparse in situ observational data but with a reduced smoothness of the bias compared to other scenarios. In this scenario the climate model output should provide added value in estimating $\phi_Y$ across the domain but this will be limited compared to scenario two due to the difficulty of disaggregating and estimating the comparatively more complex bias.

| | Scenario 1 | Scenario 2 | Scenario 3 |
|---|---|---|---|
| In-Situ Kernel Variance ($v_{\phi_Y}$) | 1.0 | 1.0 | 1.0 |
| In-Situ Kernel Lengthscale ($l_{\phi_Y}$) | 3.0 | 3.0 | 3.0 |
| In-Situ Mean Constant ($m_{\phi_Y}$) | 1.0 | 1.0 | 1.0 |
| In-Situ Observation Noise ($\sigma_{\phi_Y}$) | 0.1 | 0.1 | 0.1 |
| Bias Kernel Variance ($v_{\phi_B}$) | 1.0 | 1.0 | 1.0 |
| Bias Kernel Lengthscale ($l_{\phi_B}$) | 10.0 | 20.0 | 5.0 |
| Bias Mean Constant ($m_{\phi_B}$) | -1.0 | -1.0 | -1.0 |
| # In-Situ Observations | 80.0 | 20.0 | 20.0 |
| # Climate Model Predictions | 100.0 | 80.0 | 80.0 |

Table 3.1: A table showing the hyper-parameters of the two latent Gaussian processes used to generate the complete underlying realisations of $\phi_Y$, $\phi_B$ and hence $\phi_Z$, as well as observations of $\phi_Y$ and $\phi_Z$, on which inference is done for three scenarios. The number of observations representing in-situ data and climate model output is also given.

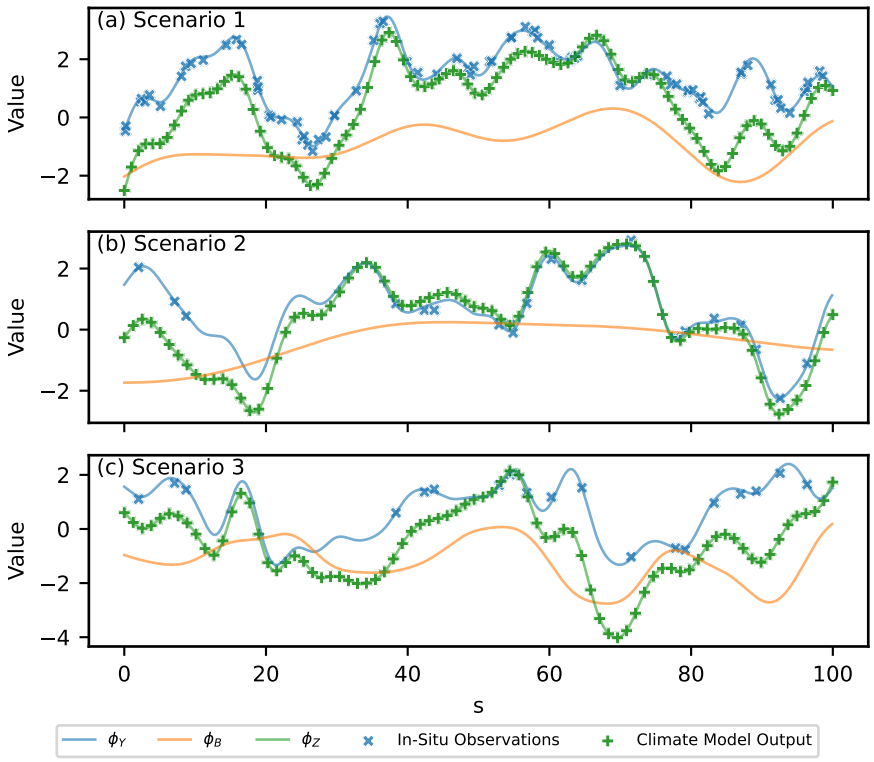



Figure 3.6: A figure showing simulated observed data for the parameters $\phi_Y$, $\phi_Z$ as well as the underlying latent functions for each parameter and the underlying bias $\phi_B$, defined through $\phi_Z = \phi_B + \phi_Y$. Three scenarios are shown and correspond to data generated from parameters in Table 3.1. The Std.Dev of the noise term when generating the in situ observations is 0.1.

#### 3.4.1.2 Results

The expectation, standard deviation and 95% credible intervals for the prior distribution and posterior distribution after inference of each parameter under the three different scenarios is given in Table 3.2. Comparisons are shown in the statistics between the posterior distributions of the full model presented in this paper, referred to as the shared process model, and the case where only the parameters for the in situ data are modelled as generated from a latent Gaussian processes, referred to as the single process model. In Sect. B.3 of the appendix, an illustration is given of the prior and posterior distributions of each parameter after inference with the

shared process model for scenario one.

Under all scenarios and for both the shared process and single process models 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 shared process model is in general closer to the specified value than the single process model and the range of the credible interval is smaller. In scenario one the differences between the models posteriors are relatively insignificant, although the shared process model does show a reduction in the uncertainty of the length scale for the latent process generating $\phi_Y$. In scenario two the difference is more significant and clear improvement is shown in both the expectation and uncertainty of latent parameter estimates for the shared process model. Improvement is also clear in estimates from the 3rd scenario, although the relative difference in performance between models is less significant.

Predictions for the underlying fields of the parameters $\phi_Y$(s), $\phi_Z(s)$ and the corresponding bias $\phi_B(s)$ across the domain given the data, referred to as the posterior predictive, are shown in Fig. 3.7 for each scenario and for both the shared and single process models. The true underlying fields that the simulated observations were sampled from is also shown. The single process model is only concerned with estimating the underlying field of $\phi_Y(s)$ across the domain given observations of the parameter for the in situ data, so in Figs. 3.7a, 3.7c and 3.7e the climate model output and bias fields are excluded. To perform bias correction of the climate model output through quantile mapping, posterior predictive estimates of $\boldsymbol{\phi}_Y(s)$ at the climate model output locations are required. The relative ability of the shared and single process models to estimate this is further assessed through $R^2$ scores, presented in Table 3.3.

In Fig. 3.7 it can be seen that the predictions of $\phi_Y(s)$ across the domain in the shared process case (Fig. 3.7b, 3.7d and 3.7f) are closer to the true underlying field and with smaller but still realistic uncertainty compared to the single process model. In addition to visual examination, further validation of the credible

intervals could be performed by re-running the results for multiple realisations of the underlying data for the scenarios, keeping the parameter values used for generating the data the same. For a specific location the credible interval could be evaluated by computing the proportion of the time the actual value is within the credible interval of the prediction. In scenario one, the difference between the posterior predictive distributions for $\phi_Y(s)$ across the domain 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 $\phi_Y(s)$ 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 there is again a significant difference in the estimates of $\phi_Y(s)$ between the models, with $R^2$ scores of 0.74 and 0.52 respectively, although the difference is reduced compared with scenario two.

| **Scenario 1** | Specified | Prior Distribution | | | | Posterior Dist. (Shared Process) | | | | Posterior Dist. (Single Process) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Value | 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. |
| In Situ Kernel Variance $v_{\phi_Y}$ | 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 |
| In Situ Kernel Lengthscale $l_{\phi_Y}$ | 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 |
| In Situ Mean Constant $m_{\phi_Y}$ | 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 |
| In Situ Observation Noise $\sigma_{\phi_Y}$ | 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 Kernel Variance $v_{\phi_B}$ | 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 |
| Bias Kernel Lengthscale $l_{\phi_B}$ | 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 |
| Bias Mean Constant $m_{\phi_B}$ | -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 |
| **Scenario 2** | Specified | Prior Distribution | | | | Posterior Dist. (Shared Process) | | | | Posterior Dist. (Single Process) | | | |
| | Value | 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. |
| In Situ Kernel Variance $v_{\phi_Y}$ | 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 |
| In Situ Kernel Lengthscale $l_{\phi_Y}$ | 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 |
| In Situ Mean Constant $m_{\phi_Y}$ | 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 |
| In Situ Observation Noise $\sigma_{\phi_Y}$ | 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 Kernel Variance $v_{\phi_B}$ | 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 |
| Bias Kernel Lengthscale $l_{\phi_B}$ | 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 |
| Bias Mean Constant $m_{\phi_B}$ | -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 |
| **Scenario 3** | Specified | Prior Distribution | | | | Posterior Dist. (Shared Process) | | | | Posterior Dist. (Single Process) | | | |
| | Value | 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. |
| In Situ Kernel Variance $v_{\phi_Y}$ | 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 |
| In Situ Kernel Lengthscale $l_{\phi_Y}$ | 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 |
| In Situ Mean Constant $m_{\phi_Y}$ | 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 |
| In Situ Observation Noise $\sigma_{\phi_Y}$ | 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 Kernel Variance $v_{\phi_B}$ | 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 |
| Bias Kernel Lengthscale $l_{\phi_B}$ | 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 |
| Bias Mean Constant $m_{\phi_B}$ | -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 3.2: 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.

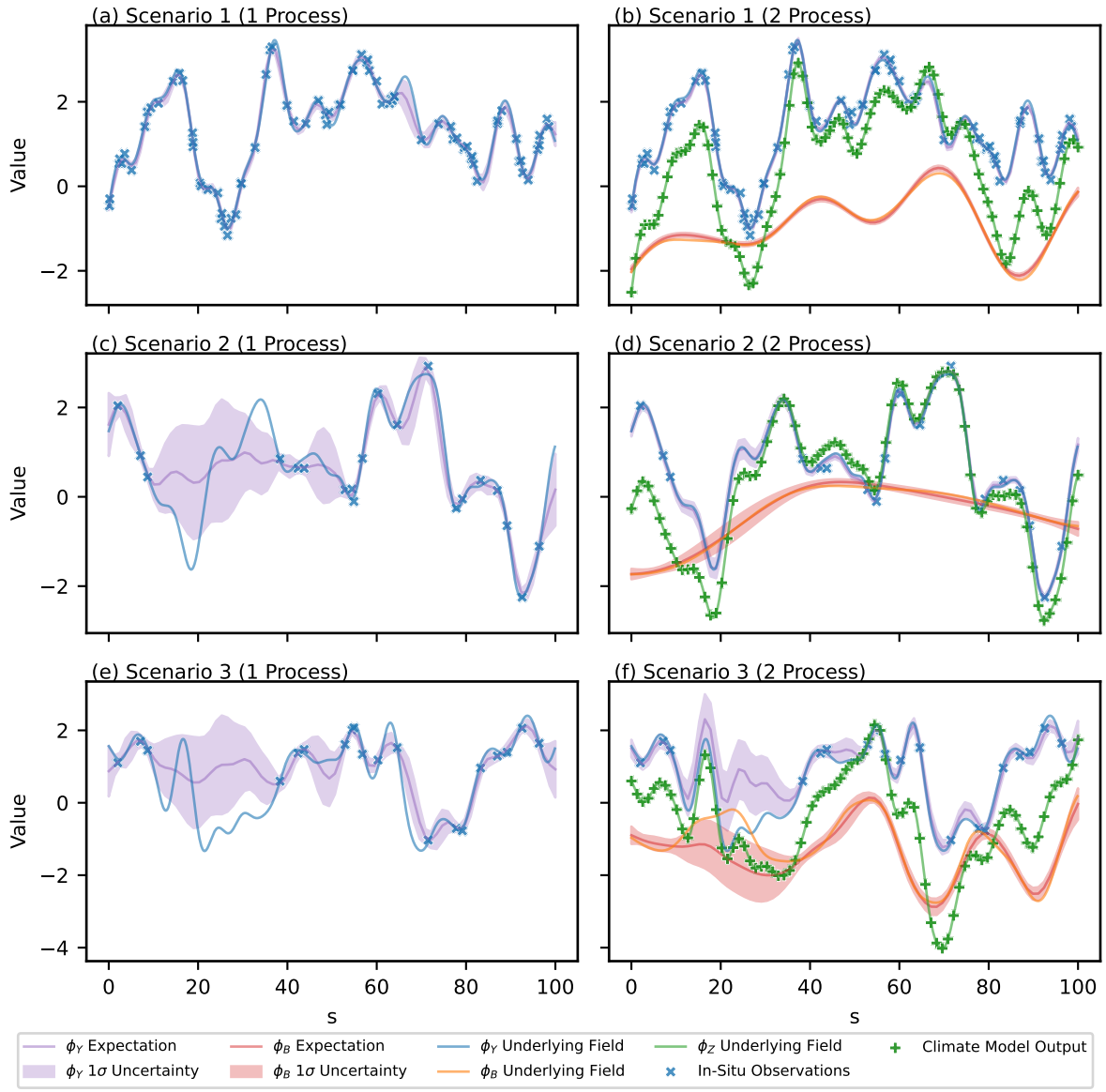



Figure 3.7: Expectation and 1$\sigma$ uncertainty of the posterior predictive distributions across the domain for the parameter $\phi_Y(s)$ and the corresponding bias $\phi_B(s)$ for three scenarios. The underlying latent functions that the data are measurements of is included.

| | Shared Process Model | | Single Process Model | |
|---|---|---|---|---|
| | Exp. | Std.Dev. | Exp. | Std.Dev. |
| Scenario 1 | 0.99 | 0.00 | 0.97 | 0.01 |
| Scenario 2 | 0.99 | 0.01 | 0.68 | 0.07 |
| Scenario 3 | 0.74 | 0.12 | 0.52 | 0.10 |

Table 3.3: A table showing the expectation and standard deviation of $R^2$ scores for the posterior predictive estimates of $\phi_Y$ at the climate model output locations for the shared and single process models for each scenario.

### 3.4.2 Bayesian Framework: Hierarchical Example

A hierarchical example is presented in this section where the in situ data and climate model output are simulated at each site as generated from normal distributions, as in the specific example given in Sect. 3.3.2. The goal of the model is the same as in Sect. 3.4.1, that is to predict the parameters of the PDFs for the climate model output and in situ observations at the locations of the climate model output. An example of how uncertainty in these predictions can be propagated through bias correction techniques such as quantile mapping is then presented. The purpose of this section is to demonstrate the model working in the intended hierarchical structure and to illustrate the benefit of having a fully Bayesian hierarchical model for uncertainty estimation.

#### 3.4.2.1 Simulated Data

The simulated data in this example is generated assuming the dependencies in Sect. 3.3.2 and Fig. 3.2. Defining $Y(s,t)$ and $Z(s,t)$ as the in-situ data and climate model output respectively, then the time-independent PDF at each site is taken as normal such that $Y(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$ and $Z(s) \sim \mathcal{N}(\mu_Y(s), \sigma_Y(s))$. 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 3.4. Further specifics of the data generation is provided in Sect. B.2 of the appendix.

There are 40 locations corresponding to simulated in situ observation sites, where for each site 20 measurements are generated. Likewise, there are 80 locations corresponding to simulated climate model output 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 also afflicted with greater bias. In Fig. 3.8 examples of the generated samples 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 when estimating the parameters across the domain. The underlying, complete realisations of the parameters $\mu_Y(s)$, $\mu_Z(s)$, $\sigma_Y(s)$ and $\sigma_Z(s)$, as well as the bias $\mu_B(s)$ and $\sigma_B(s)$, are shown in Fig. 3.9. In addition, the mean value and standard deviation of the generated data is given at the simulated in situ observation and climate model sites.

| | Hierarchical Scenario |
|---|---|
| In-Situ Mean, Kernel Variance ($v_{\mu_Y}$) | 1.0 |
| In-Situ Mean, Kernel Lengthscale ($l_{\mu_Y}$) | 3.0 |
| In-Situ Mean, Mean Constant ($m_{\mu_Y}$) | 1.0 |
| In-Situ Transformed Variance, Kernel Variance ($v_{\tilde{\sigma}^2_Y}$) | 1.0 |
| In-Situ Transformed Variance, Kernel Lengthscale ($l_{\tilde{\sigma}^2_Y}$) | 3.0 |
| In-Situ Transformed Variance, Mean Constant ($m_{\tilde{\sigma}^2_Y}$) | 1.0 |
| Bias Mean, Kernel Variance ($v_{\mu_B}$) | 1.0 |
| Bias Mean, Kernel Lengthscale ($l_{\mu_B}$) | 10.0 |
| Bias Mean, Mean Constant ($m_{\mu_B}$) | -1.0 |
| Bias Transformed Variance, Kernel Variance ($v_{\tilde{\sigma}^2_B}$) | 1.0 |
| Bias Transformed Variance, Kernel Lengthscale ($l_{\tilde{\sigma}^2_B}$) | 10.0 |
| Bias Transformed Variance, Mean Constant ($m_{\tilde{\sigma}^2_B}$) | -1.0 |
| # Spatial Locations of In-Situ Observations | 40.0 |
| # Spatial Locations of Climate Model Predictions | 80.0 |
| # Samples per Location of In-Situ Observations | 20.0 |
| # Samples per Location of Climate Model Predictions | 100.0 |

Table 3.4: 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.

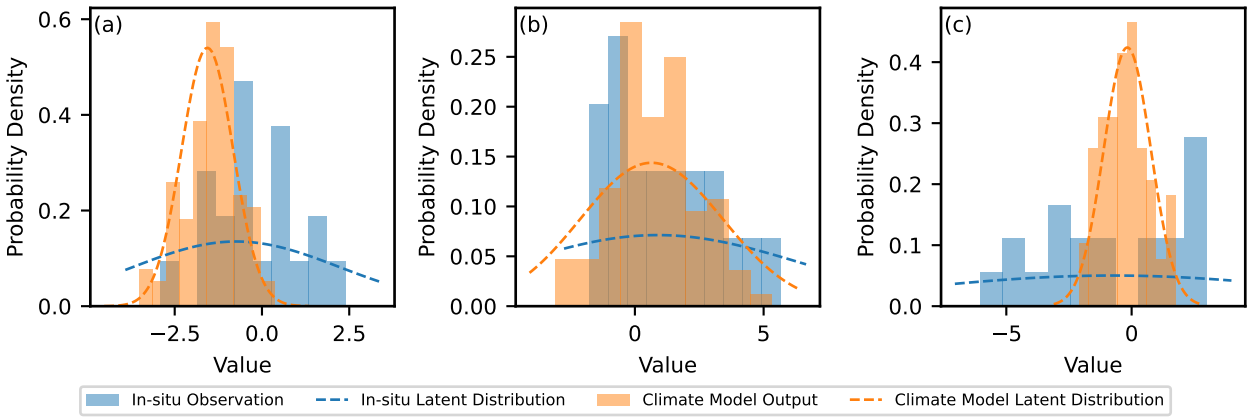



Figure 3.8: Histograms for the climate model output at three locations and the corresponding data from the nearest in situ observation site. The locations are a) s=11.4, b) s=46.8 and c) s=79.7. The latent normal distribution the data was generated from is illustrated as a dotted line.

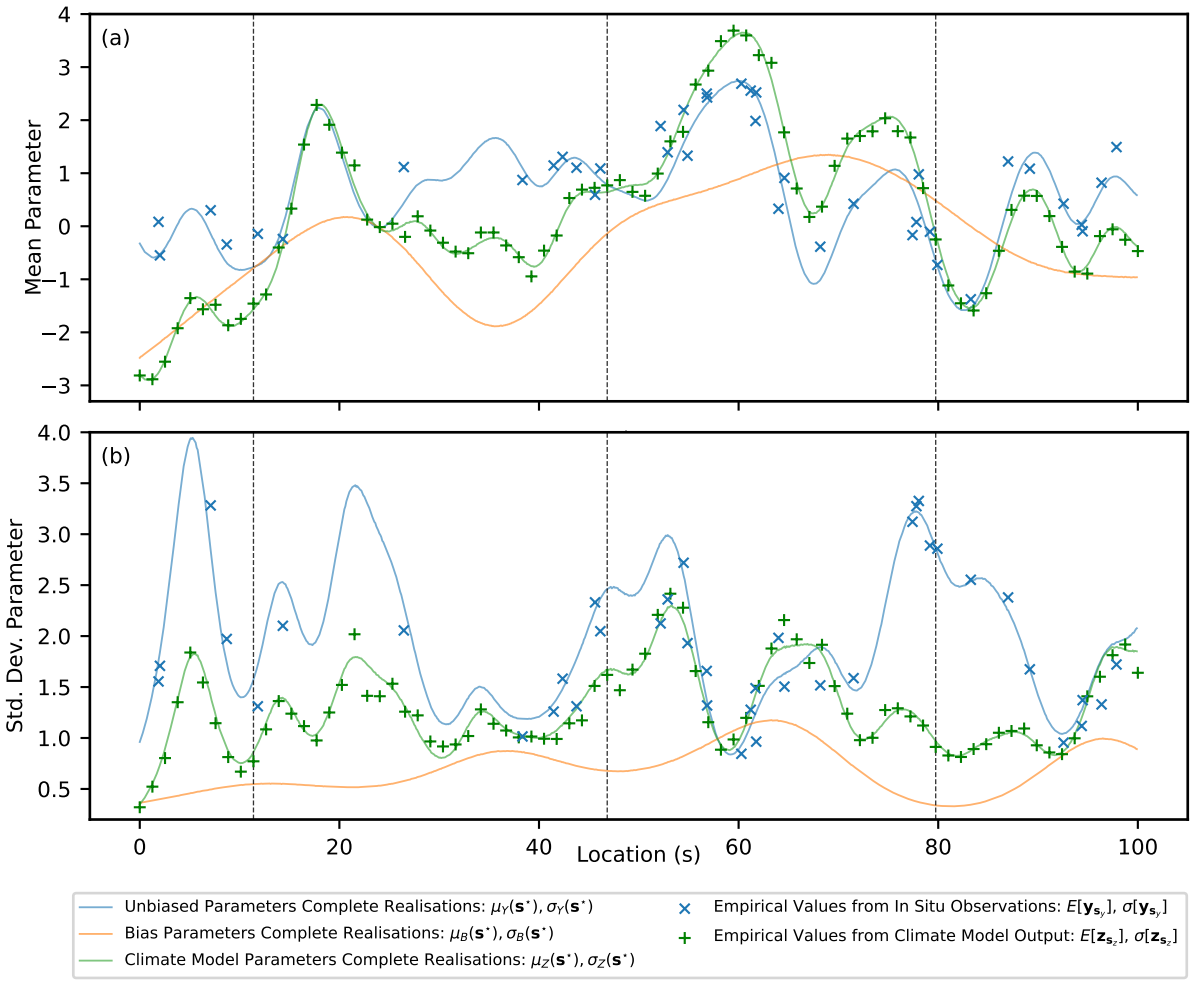



Figure 3.9: Simulated underlying functions for the parameters $\mu_Y(s)$, $\mu_B(s)$, $\mu_Z(s)$, $\tilde{\sigma}_Y(s)$, $\tilde{\sigma}_B(s)$ and $\tilde{\sigma}_Z(s)$ as well as the values at the observation locations for the in situ and climate model data. Vertical lines indicate values referred to in Fig. 3.8.

#### 3.4.2.2 Results

The expectation, standard deviation and 95% credible intervals for the prior and posterior distributions of each parameter are given in Table 3.5. 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 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 parameter are plot in Fig. B.2 of the appendix.

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

Table 3.5: 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 parameter used to generate the data is also shown.

The posterior predictive estimate for the underlying fields of $\mu_Y(s)$, $\mu_B(s)$, $\sigma_Y(s)$ and $\sigma_B(s)$ across the domain given the data is shown in Fig. 3.10. The true underlying fields of the parameters are also shown, as are the mean and standard deviation values of the samples of simulated in situ observations and climate model outputs at the locations where they are sampled. The posterior predictive appears to perform well at capturing the spatial features of the underlying fields while also exhibiting a reasonable one sigma uncertainty range that bounds the majority of the underlying function. For example, in the range of $s \in [15, 25]$, where the main data source is the biased climate model output, the prediction accurately captures features of the true, unobserved latent mean $\mu_Y(s)$ and standard deviation $\sigma_Y(s)$. Uncertainty in the parameters of $\mu_Y(s)$ and $\sigma_Y(s)$ at the observation sites, due to limited samples, is propagated through the model. This is reflected in the

uncertainty shown in estimates of the posterior predictive at the observation sites.

Bias correction of samples from the climate model output for a single site is shown in Fig. 3.11. The site chosen is at $s = 11.4$ and is the same as in Fig. 3.8a. A generic time series for the climate model output and in situ observations is generated from the correct mean and standard deviations of the samples. Quantile mapping of the climate model time series is performed for each posterior predictive realisation of $\mu_Y(s)$, $\mu_Z(s)$, $\sigma_Y(s)$ and $\sigma_Z(s)$. This results in multiple realisations of bias corrected time series with an expectation and uncertainty.

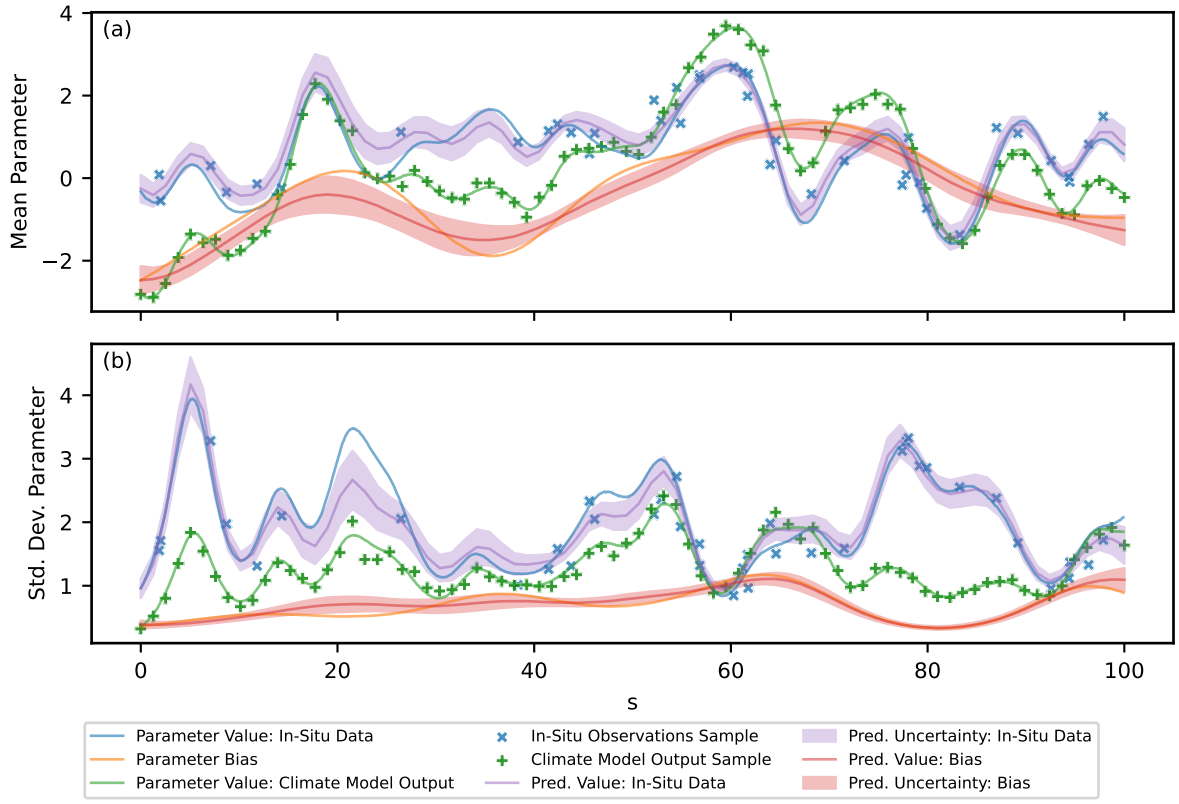



Figure 3.10: A figure showing the expectation and one sigma uncertainty of the posterior predictive distribution across the domain for the parameters $\mu_Y(s)$, $\mu_B(s)$, $\sigma_Y(s)$ and $\sigma_B(s)$ as well as their true underlying latent functions.

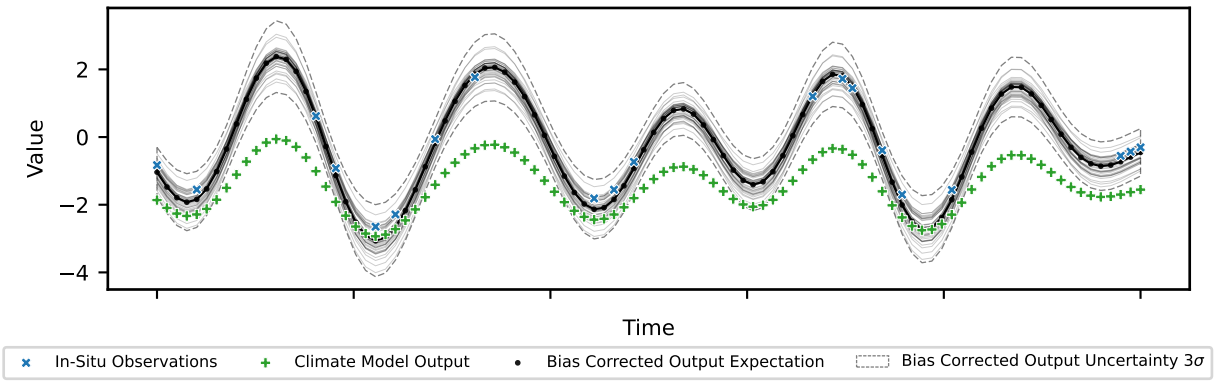



Figure 3.11: Simulated time series for the climate model output at location $s = 11.4$ and for the nearest in situ observation site. Realisations of the climate model bias corrected time series are shown along with the expectation and three sigma uncertainty range.

## 3.5 Discussion

The methodology presented in this paper assumes that each spatially varying parameter of the PDF for the climate model output is generated from two independent, latent GPs. One of these latent processes is also modelled to generate the equivalent parameter for the PDF of the in situ data, while the other latent process generates the bias. This reflects the belief that the climate model provides skilful estimates of these parameters across the domain and that the spatial covariance structure, generated from equations based on established physical laws, has similar features to the true underlying structure. The climate model output, while afflicted with bias, has comprehensive spatiotemporal coverage and provides useful information in the inference of the true values of the parameters across the domain, assuming the bias signal can be adequately deconstructed from the climate model output with the use of in situ observations. Incorporating this into the model presented in this paper provides added value over approaches where the parameters $\phi_Y(s)$ are inferred from the in situ observations alone, as in Lima et al., 2021. This is demonstrated in Sect. 3.4.1, where added value is assessed for three scenarios with differing density of observations and complexity of the bias signal.

Added value is assessed with respect to: summary statistics for the posterior distributions of the parameters of the latent GPs, visual examination of the expectation and standard deviation for posterior predictive estimates of the site-level PDF parameters across the domain, and comparison of $R^2$ scores for the PDF parameter $\phi_Y(s)$ at the locations of the climate model output. In Sect. 3.4.1.2 it is shown that most added value is provided, across all these measures, in the case of scenario two, where in situ observations are sparse compared to the climate model output and the underlying bias is relatively smooth compared to the unbiased signal. The bias can be estimated with high accuracy and precision, despite sparse in situ observations, since it varies smoothly across the domain, which also means the climate model output can be disaggregated and the unbiased component estimated across the domain with high accuracy and precision.

It is also shown in Sect. 3.4.1.2 that if 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. This is demonstrated in scenario one, where the number of in situ observations is sufficient to adequately capture the spatial features of the underlying process (Fig. 3.7) as well as the latent spatial covariance structure, encoded through the hyper-parameter estimates of the latent GP (Table 3.2). Additionally, if the complexity of the bias signal is increased, through for example reducing the length scale of the latent generating process, then again added value is reduced. This is shown in scenario three, where a relatively more complex bias compared with scenario two makes it more difficult to disaggregate the climate model output into its biased and unbiased components, thus reducing the benefit provided in estimates of $\phi_Y(s)$. It is noted that, while added value is reduced relative to scenario two, benefits are still shown for scenarios one and three and incorporating the climate model output in inference improves overall performance. It is noted, in real-world applications there are various methods to evaluate the model performance. One method is Leave-One-Out Cross-Validation (LOOCV), where a single observation is left out and used as a validation point. The validation is repeated for each observation in turn and an average performance metric computed.

In addition to shared latent processes, another important feature of the methodology presented in this paper is the Bayesian framework, where the parameters of the model are treated as random variables with associated distributions. This framework is flexible and allows for robust uncertainty propagation, which is important for making the model applicable to a wide range of real-world applications where bias prediction is required. Additionally, expert knowledge can be incorporated in the inference through the choice of prior distributions, which is especially important where the data is sparse. In Sect. 3.4.2 results for a simulated one dimensional hierarchical example illustrate uncertainty propagation between parameter values of the PDF at each sample site and the values of 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 PDF parameters across the domain. These posterior predictive estimates can be used in bias correction techniques, such as quantile mapping, which is illustrated in Fig. 3.11. This results in multiple realisations of the final bias corrected time series, with an expectation and uncertainty range. Robust uncertainty computation that incorporates the spatial relationships between points is important for impact assessments and resulting decision making. Having multiple realisations for the final bias corrected time series is also useful for further propagation of uncertainty in process models driven by climate model output, such as land surface models Liu et al., 2014.

## 3.6 Conclusions

Current approaches for bias prediction and correction do not aim to preserve the spatial covariance structure of the climate model output (Ehret et al., 2012). Climate models are fundamentally based on established physical laws and so the covariance structures are desirable since it is reasonable to assume that they are physically realistic. In addition, current approaches typically either neglect uncertainty or inadequately model uncertainty propagation through the model. In this paper a fully Bayesian hierarchical model for bias correction is presented where latent GP distributions are used to capture and preserve underlying covariance structures. The Bayesian nature allows robust uncertainty propagation under a flexible modelling framework where the model is easily expanded for specific real-world scenarios, increasing the scope of the work.

Simple simulated examples are chosen to illustrate the key features of the model. In Sect. 3.4.1, results are displayed for a non-hierarchical example where the focus is on illustrating the nature of GPs and how assuming a shared latent GP between the in situ data and climate model output allows inference on the unbiased field

from both sources of data. This is shown to be particularly important in the case of sparse data and a simple bias, where the climate model output provides significant value added in predictions. In Sect. 3.4.2, 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 parameter predictions that are used in bias correction. Uncertainty in the parameter estimates is easily propagated in bias correction of the time series from the climate model at every location through the existing approach of quantile mapping. This results in a bias corrected time series with uncertainty bands, which is desirable for use in impact studies that compute predictions on responses to climate change and for informing decisions based on these. This is especially true in areas where the climatology is hard to model and in situ observations are sparse, such as Antarctica, meaning the uncertainty is expected to be significant (Carter et al., 2022b).

The model presented is a step towards adequately capturing uncertainty and incorporating underlying spatial covariance structures from the climate model in bias correction. The primary limitation is the assumption that the spatial structure of the site-level parameters can be adequately modelled through a stationary GP. Over large and complex topographic regions it is likely that the covariance length scale will vary across the domain and this is something that will need assessing for each specific application. Additionally, many real-world applications will necessitate specific model adjustments, such as incorporating a mean function dependent on factors like elevation and latitude, handling non-Gaussian data, and accounting for other bias structures. Since the model is developed in a Bayesian framework and inference on the parameters conducted with MCMC, model adjustments are simple to incorporate with adequate uncertainty propagation. The next step then is to apply the methodology to a real-world dataset, incorporating additional modelling components and further exploring advantages as well as limitations that arise.

# Chapter 4

# Applying Bias Correction to Surface Climatology over Antarctica

Accurate and comprehensive estimates of climatology over Antarctica are essential for modelling the current and future stability of the ice sheet. Climate models provide invaluable estimates of climatology useful for driving physical process models, although the impact of biases in the output is not always considered or adequately modelled. This paper provides the first real-world application case study of a novel Bayesian hierarchical model for bias prediction and correction. This novel approach considers underlying spatial covariance patterns in the data and uncertainty in the correction applied. Near-surface air temperature output from the state-of-the-art Modèle Atmosphérique Régional climate model is utilised and along with in situ automatic weather station records, compiled across the ice sheet in a recent study. Results are promising, with the methodology resistant to overfitting, producing spatially smooth predictions over the ice sheet that improve agreement with in situ observations and that demonstrate realistic estimates of uncertainty.

## 4.1 Introduction

Antarctic surface climatology is an important driver of changes in mass balance over the continent, with wide-ranging implications for the biosphere (Oppenheimer et al., 2022). Non-linear, physical processes, such as ice shelf collapse (Banwell et al., 2013; Pollard et al., 2015) and albedo feedbacks (Buzzard et al., 2018) among others, can result in regime shifts where dramatic mass change can occur (Shuman et al., 2011). In R. M. DeConto and Pollard, 2016 it is hypothesised that under certain regimes by 2100 Antarctica could contribute over 1 m of sea level rise, displacing hundreds of millions of people (Nicholls et al., 2011). Accurate estimates of surface climatology across Antarctica are therefore imperative to inform impact assessments and our understanding of the likelihood of regime shifts, for which the community is primarily dependent on physically based climate models due to sparse in situ observations.

To understand the current state and predict the future state of the ice sheet requires high-resolution, skilful estimates of local climatology, typically provided by dynamical downscaling from regional climate models (RCMs) (Luckman et al., 2014). These are comparatively high-resolution climate models nested inside lower-resolution Global Climate Models (GCMs) or reanalysis datasets. The RCMs are typically tuned for the specific continent/region of interest and include features such as sophisticated land-surface schemes with region-specific adaptations (Lenaerts and M. R. v. d. Broeke, 2012; Giorgi, 2019). While RCMs are shown to provide skilful estimates of surface climatology over regions such as Antarctica (Jan Melchior van Wessem et al., 2015; Agosta et al., 2019), significant systematic errors/differences between the models are also shown, both globally and regionally (Mottram et al., 2021; Carter et al., 2022a). It's important to consider the influence of these biases on impact assessments, propagating uncertainty in the driving climatology. Biases in climate models originate from factors such as required parameterisations of physical processes and incomplete process understanding. The computationally demanding nature of climate models coupled with complex topography-induced

climatology over Antarctica, such as föhn winds (Munneke et al., 2018), and complex feedback mechanisms make direct reduction in these errors through climate model development challenging. Statistical post-processing bias correction techniques have been developed to try to solve this issue, being adaptable for the specific focus of the end user and adding value to RCM simulation outputs.

These statistical bias correction techniques range from relatively simple methods, such as the delta change method (Das et al., 2022), to more involved and complex approaches using for example generalised additive models (Beyer et al., 2020). Complex modelling approaches generally provide improved consistency with observations although interpretation of the methodology is more difficult and non-physical sharp variations in the predictions can occur. Some approaches focus on correcting errors further upstream in the global driving dataset (Bruyère et al., 2014), while others focus directly on the RCM output. Across the literature, methodologies typically fail to consider uncertainty in the correction and neglect to preserve the physically-realistic spatial structures produced from the physically-based climate models, which among other aspects preserve energy and mass across the domain (Ehret et al., 2012). This is of particular importance over regions such as Antarctica, where sparse in-situ observations limit confidence in observation-based corrections. To this end a novel technique is developed in (Carter et al., 2024), which considers the in situ observations and climate model output to be generated from shared underlying Gaussian processes (GPs). In this paper this novel approach is explored for the first time with respect to real-world data and the challenging case study of bias correcting near-surface air temperature climate model output from the Modèle Atmosphérique Régional version 3.10 (MARv3.10) RCM over Antarctica.

In-situ observations of near-surface air temperature over Antarctica are predominantly provided from automatic weather stations (AWSs), which are sparsely distributed over the continent. This paper utilises AWS output from a recent compilation (Yetang Wang et al., 2023) with 267 AWS sites available. This equates to approximately 1 station per $50,000$ km$^2$ (an equivalent density to 5 stations

over the whole of the United Kingdom). Additionally, the data is distributed highly heterogeneously in space, with much lower densities occurring over much of East Antarctica. Despite limited observational data, AWS records still potentially provide useful information for bias correcting climate model output, especially with respect to any large-scale, relatively smooth spatial patterns of the bias. Significant uncertainty is expected due to the spatial and temporal data sparsity, which is important to capture. The novel approach for bias correction applied in this paper is specifically designed for this type of scenario, with robust uncertainty propagation through the model and inference on the unbiased parameters that utilises the skilful climate model output as well as the in situ observations (Carter et al., 2024). Remaining challenges and future development of the methodology, such as modelling non-stationary lengthscales are discussed.

## 4.2 Data Exploration

This section provides a brief exploration of the in situ observational data (section 4.2.1) and climate model output (section 4.2.2) used in this paper. Important features of the data are highlighted and any modelling choices as well as data filtering are discussed. Figure 4.1 shows an overview of Antarctica for reference against, with key place names and elevation contour lines shown. Ice shelves are displayed in light blue and rocky outcrops in black.


Figure 4.1: Map of Antarctica, generated using the Quantarctica mapping environment (Matsuoka et al., 2021).

### 4.2.1 Automatic Weather Station Data

The recently published AntAWS dataset is a quality controlled compilation of records from 267 AWSs covering a time period of 1980-2021 (Yetang Wang et al., 2023). Five meteorological variables are provided: air temperature; air pressure; relative humidity; wind speed and direction. This paper focuses on the air temperature product, which has the highest data continuity and integrity of the variables (Yetang Wang et al., 2023). Temporal resolutions offered for the processed data are 3 hourly, daily and monthly. Quality control measures implemented are applied to the 3 hourly records and include: removing records outside the viable

measurable range of the sensors; removing records with unrealistic, rapid changes in temperature (threshold of $\Delta 5$°C over 6 hours); removing records outside 3 standard deviations from the monthly mean; removing records for air temperature in the Antarctic summer months (Dec-Feb) with wind speeds less than 2 m $s^{-1}$ that can exhibit warm biases due to lack of ventilation; removing records that visually appear as outliers during comparison with corresponding outputs from the reanalysis product ERA5 (Hersbach et al., 2020) and output from the MARv3.10 RCM (Agosta et al., 2019). Two reliability threshold bands (25% and 75%) are provided for the daily and monthly products, which are computed from aggregations of the quality controlled 3 hourly data. The bands only give aggregated values where 25% and 75% respectively of the 3 hourly data exists for the time period. Air temperature records are measured at approximately 2-3 m height and have ventilated radiation shields, obstructing direct sunlight. Thermistor or resistive platinum probes are used for measurements and uncertainty is within ± 0.5°C (Yetang Wang et al., 2023). It is noted that the relative height change in the sensor, typically from snow accumulation, partially influences the temporal component of records, particularly at sites in the Antarctic Plateau during winter where significant temperature inversions exist (Genthon et al., 2021).

In this paper the 75% reliability threshold monthly product is used. The higher reliability threshold is chosen since bias induced in the AWS records themselves are not considered, while the monthly timescale is chosen for computational reasons and since significant systematic errors are known to exist at this temporal scale (Carter et al., 2022a). An example monthly timeseries output from the Manuela AWS is illustrated in Fig. 4.2 with the corresponding marginal PDF also shown. This site has one of the longest and most complete records with measurements from 1984-2021, only containing significant gaps in the years 2006 and 2008. The majority of sites have much fewer records (as shown in Fig. 4.5a.), either running for a much shorter period of time or with longer gaps in the records. The PDF over all months, as given in Fig. 4.2b, contains multiple peaks corresponding to seasonal temperature

variations. The PDFs for the individual months of January (austral summer) and June (austral winter) are shown in Fig. 4.2c and can be approximated as normal with mean and variance parameters. For this reason, bias correction techniques are typically applied to each month separately (Lima et al., 2021) and this paper follows this approach, focusing on correcting bias in just June temperatures for simplicity and clarity of results presented.

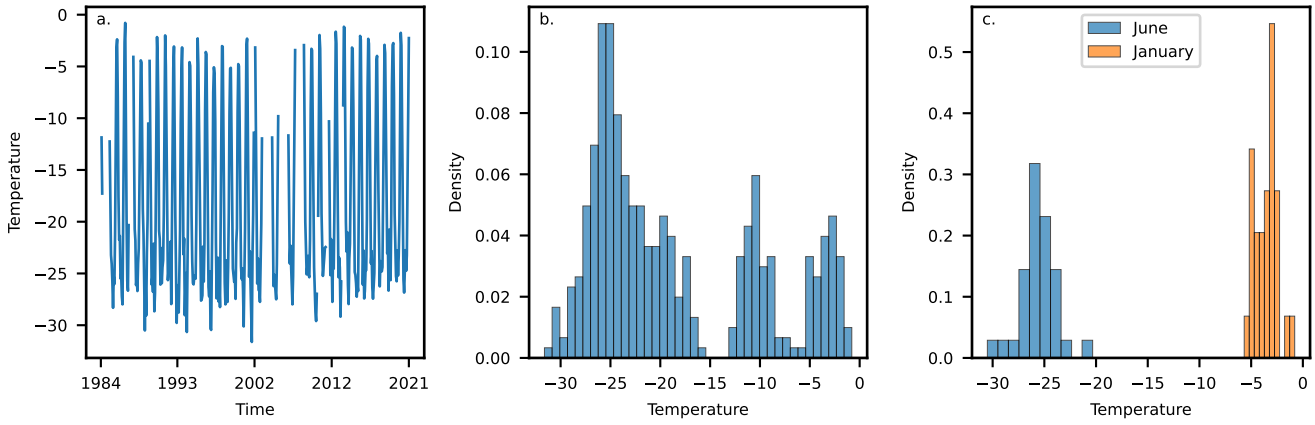



Figure 4.2: Monthly near-surface air temperature records from the Manuela AWS (lat:-74.95,lon:163.69) located near the coast over Victoria Land. The timeseries is shown in (a) with the corresponding empirical marginal PDF given in (b) and the marginal distributions filtered to just January and June illustrated in (c).

The number of AWS monthly temperature records, aggregated across all stations, is shown against the year and month in Fig. 4.3. There is an overall increasing trend with records against year (the exception being over the last 5 years of records), indicative of improvements in technology and Antarctic climate modelling efforts. In theory this should translate to a reduction in uncertainty for later years in any bias corrected product. In reality, bias is typically considered time-independent, which is also the approach taken in this paper. The distribution of records against the month of the year (Fig 4.3b) shows the number of records is lower in the austral summer months (December, January and February) than the austral winter months (June, July and August). This disparity would induce an observational bias in

the marginal PDF over all months (Fig. 4.2b) and again provides motivation for handling bias correction in the climate model output for each month separately.

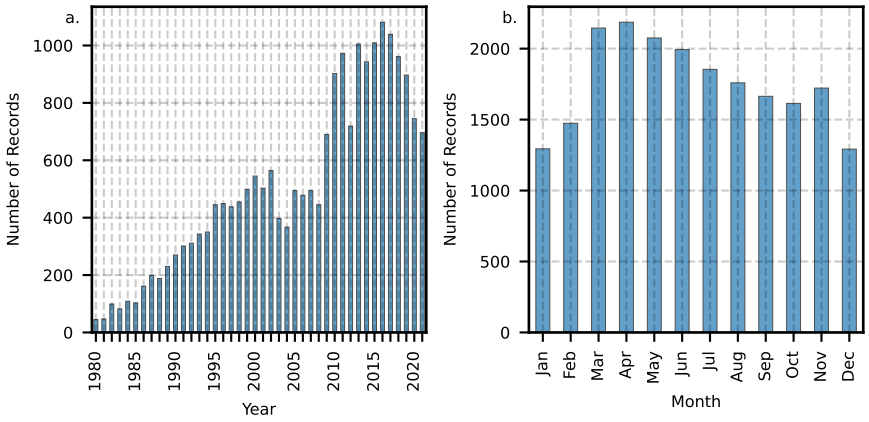



Figure 4.3: The number of AWS records across the entire ice sheet given by the year (a) and the month of the year (b).

The locations of AWSs over Antarctica and the number of monthly records for each site is displayed in Fig. 4.4. The total number across all months is given in Fig. 4.4a, while in Fig. 4.4b data is filtered to just June records and sites that are outside the shape file for the main ice sheet and or have less than 5 years of June monthly records are excluded. This filtering is done to again simplify the bias correction task and improve interpretability of results, with sites very close to the periphery and on islands expected to exhibit more challenging climatologies to model. In Fig. 4.4 stations are shown to be distributed highly heterogeneously in space, in particular with respect to the concentration of records around the south-east side of the Ross ice shelf. There is also a disproportionately high number of sites over West Antarctica compared to East.

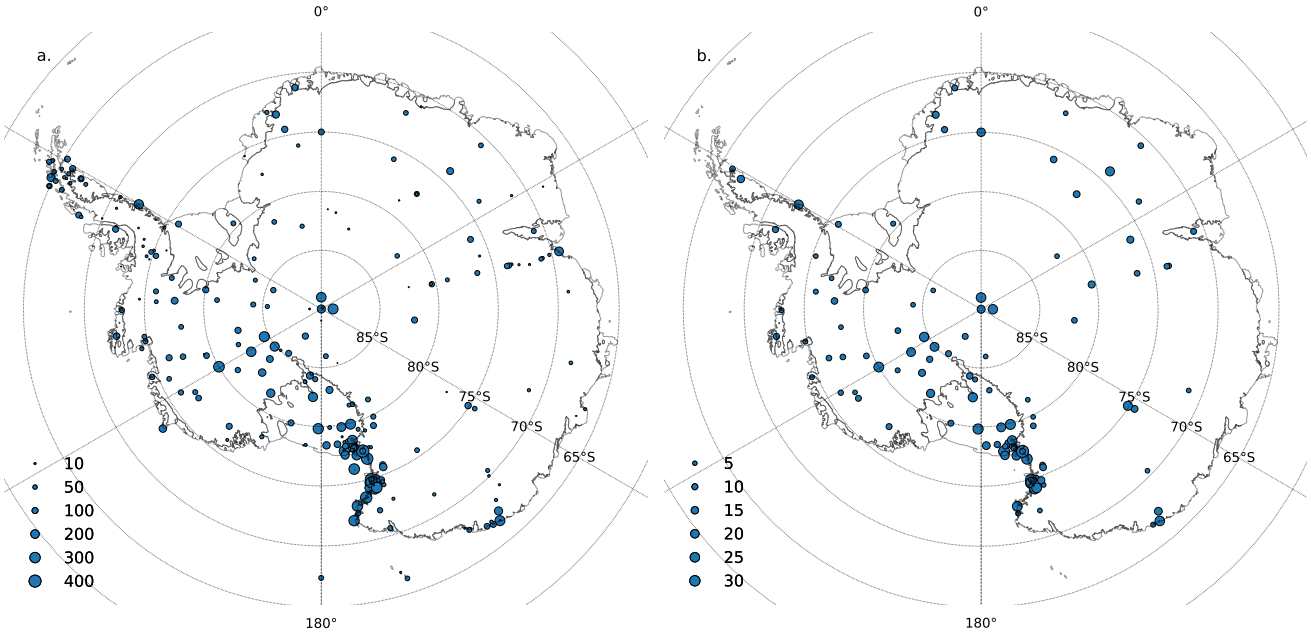



Figure 4.4: Locations of AWSs over Antarctica with the number of monthly records illustrated for each site. The sum over all months is given in (a), while (b) is filtered to just June aw well as only sites within the main ice sheet shapefile and with more than 5 years worth of data.

Histograms of the number of AWS sites against the total number of monthly records (a) and against elevation (b) and latitude (c) are illustrated in Fig. 4.5. Greater uncertainty near sites with fewer records is important to consider as part of the bias correction methodology and is done in this paper. Additionally, an uneven distribution across elevation and latitude bands is handled through modelling their dependency with near-surface air temperature. Various factors impact the location of weather stations, including accessibility and research interest, with areas such as ice shelves being high-priority due to events such as ice shelf collapse.

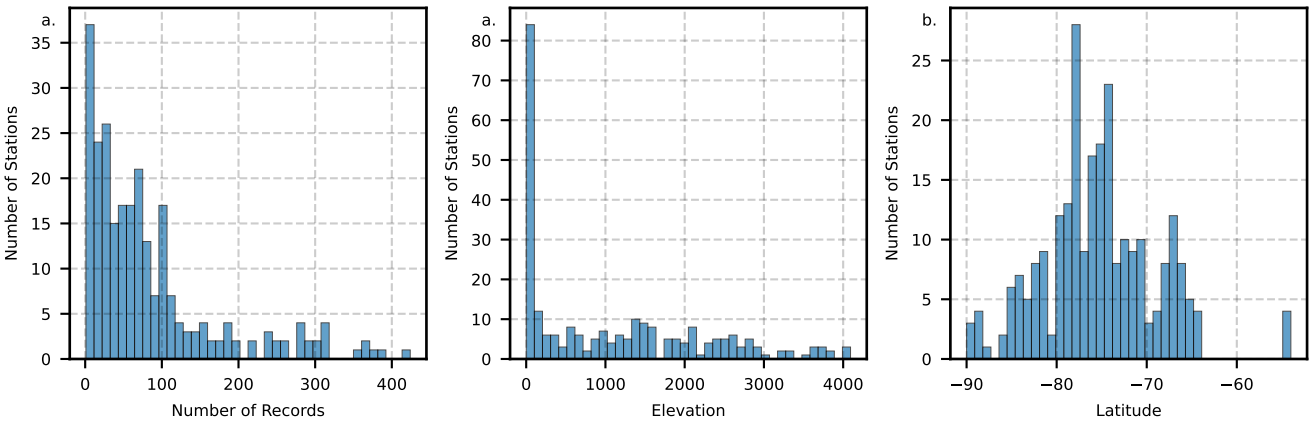



Figure 4.5: Histograms showing the frequency of AWS stations against the number of monthly records at each station (a), as well as against elevation (b) and latitude bands (c).

### 4.2.2 Climate Model Output

The climate model output used in this paper is an Antarctic-wide hindcast simulation from the MARv3.10 RCM (Agosta et al., 2019), driven at the domain boundaries by global ERA5 reanalysis output. This model is specifically designed for polar regions and is coupled with sophisticated surface schemes that account for processes such as melt water run-off and retention (X. Fettweis et al., 2013; Xavier Fettweis et al., 2017). The native spatial resolution is 35 km and temporal resolution 3 hourly. In this paper the monthly near-surface air temperature product from the RCM intercomparison of Carter et al., 2022a is used, which as well as being aggregated in time is has been interpolated onto a common 12.25 km grid for intercomparison. This product is used for convenience and so that the methodology can be easily extended to incorporate more RCM outputs.

The empirical mean June near-surface air temperature and annual standard deviation are plot in Fig. 4.6. The mean June temperature is lowest over central East Antarctica and highest near the periphery of the ice sheet, over the ice shelf grounding lines and over the Antarctic Peninsula. The standard deviation

in June temperatures varies more sporadically over the ice sheet, with various peaks including over the interior of Wilkes Land and over regions of the Filchner-Ronne and Ross ice shelves, as well as various troughs such as over the Transantarctic Mountains, Oates Land and Dronning Maud Land.

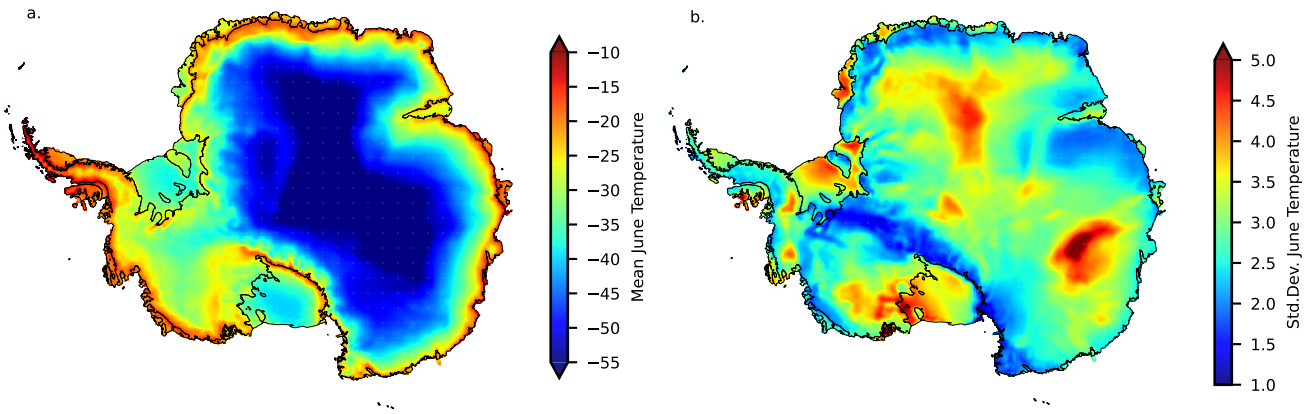



Figure 4.6: The empirical mean (a) and standard deviation (b) for June near-surface air temperatures from an Antarctic-wide MAR RCM simulation.

The pairwise distributions between the mean June near-surface air temperature, standard deviation, elevation and latitude are given in Fig. 4.7. Strong correlation (-0.87) is shown between the mean temperature and elevation, as well as a moderate correlation (0.41) given between the mean temperature with latitude. The equivalent correlations between elevation and latitude with standard deviation are weak (0.21 and -0.06 respectively). This is incorporated in the bias correction methodology, where domain-wide relationships between the mean temperature with elevation and latitude are modelled, while the standard deviation is considered independent of these predictors. It's interesting to note of a moderate correlation between the mean temperature and standard deviation themselves. In the bias correction methodology presented this is not considered as the parameters are treated as independent, but this could be an area for future improvements.

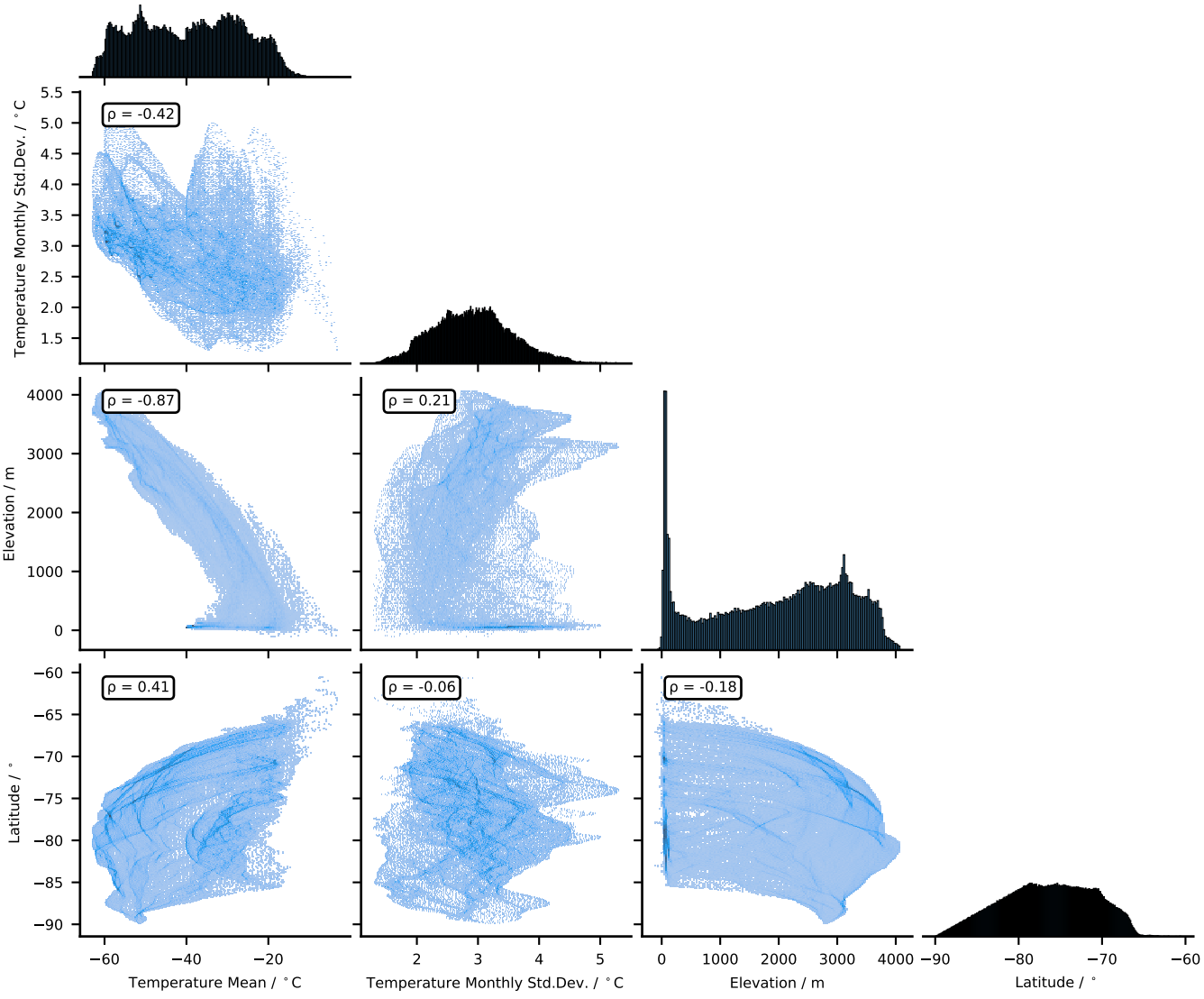



Figure 4.7: A corner plot showing pairwise relationships on the off diagonals between the mean June temperature, standard deviation in June temperatures, elevation and latitude. Marginal distributions along the diagonal are also given for each variable.

### 4.2.3 Comparison and Bias Correction

The climate model output for the nearest grid cell to the Manuela AWS is plotted against the station records in Fig. 4.8. The marginal PDF is also given for all months and for just June (Fig. 4.8b,c). The climate model provides a comprehensive time series, useful for process studies and impact assessments, although has a clear bias captured here as differences in the PDFs. In the case of June temperatures the PDF can be approximated as normal and differences in the PDF summarised by differences in the mean and standard deviation parameters. Quantile mapping is a bias correction technique that maps the PDF for the climate model output onto the unbiased observational PDF. The task is then to acquire estimates of the unbiased mean and standard deviation at each climate model grid cell (the focus of the methodology presented) so that quantile mapping can be applied across the ice sheet. Uncertainty in the unbiased parameter estimates can be propagated through the quantile mapping procedure to give multiple realisations of the final bias corrected timeseries with uncertainty bands.

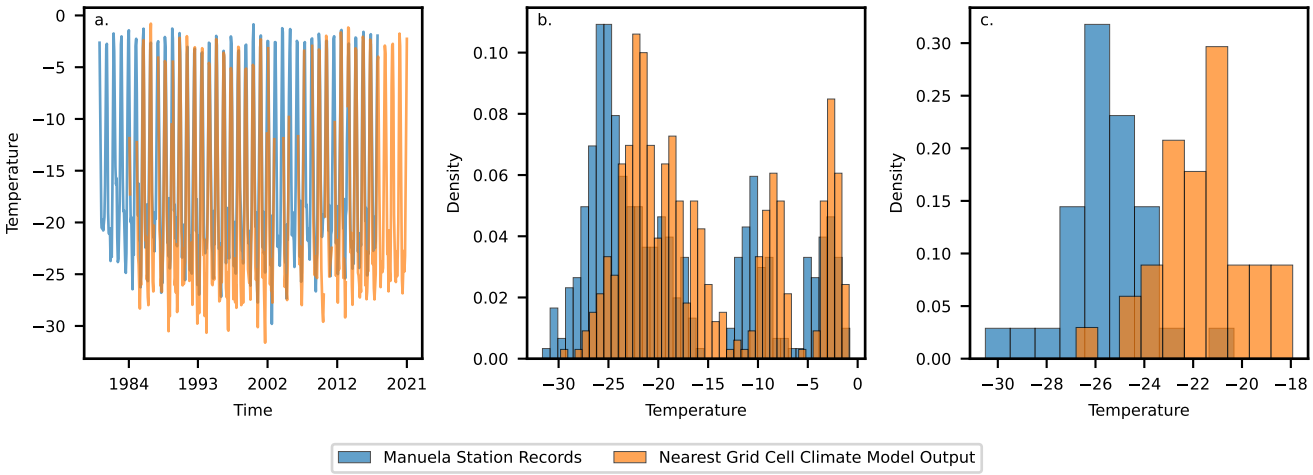



Figure 4.8: Comparison of records from the Manuela AWS with output from the nearest grid cell of the MAR RCM simulation (a). Marginal distributions across all months (b) and across just June (c) are also given.

An important assumption in the methodology utilised in this paper for predicting

the underlying unbiased parameters across the domain is that the climate model output itself provides skilful estimates of these parameters and reflects what is seen in the in situ observations with the addition of a bias component. Figure 4.9 shows the correlation between the AWS records and the output of the nearest grid cell from the climate model for the mean and standard deviation in June temperatures. Higher correlation is shown in the mean (0.98) than the standard deviation (0.65) indicating the the climate model is better at accounting for the underlying physical processes that impact the mean than the standard deviation. For both metrics the climate model provides useful information for estimating the true values across the domain. Global bias is exhibited in these parameters, as shown for example in Fig. 4.9a where the climate model output has a tendency to overpredict the mean June temperature. Additionally, significant regional patterns in bias of the parameters are expected, making it important to model the spatial covariance in bias between points.

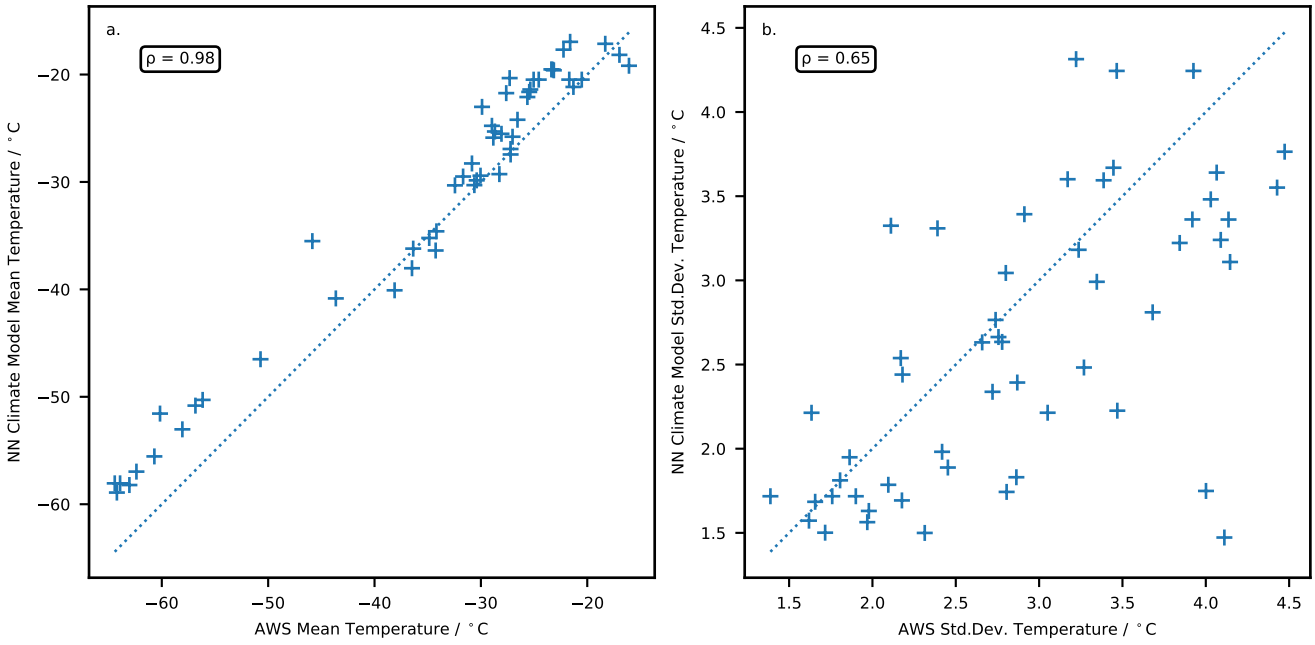



Figure 4.9: A scatter plot showing correlation between empirical values of the mean (a) and standard deviation (b) from the climate model output and AWS records.

## 4.3 Methodology

The methodology for bias prediction and correction in this paper follows the approach of (Carter et al., 2024). In this section the key components of the model are described in brief, with adaptations incorporated due to the specific real-world application of correcting near-surface air temperature over Antarctica highlighted and discussed.

### 4.3.1 Model Overview

Data from the AWS in situ observations and output from the MAR climate model simulation are considered 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. Meaning a random variable is attributed to each spatiotemporal coordinate $(Y(s,t), Z(s,t))$ and the data observed at a set of locations and times is then considered a realisation of the joint distribution of that particular finite set of random variables. For the purpose of bias prediction, the time component is marginalised and the data instead treated as realisations from the spatial stochastic processes $\{Y(s) : s \in \mathcal{S}\}$ and $\{Z(s) : s \in \mathcal{S}\}$. In the case of monthly temperature averages filtered to just June the site-level marginal distributions can be approximated as normal, such that $Y(s_i) \sim \mathcal{N}(\mu_Y(s_i), \sigma_Y(s_i))$ and $Z(s_i) \sim \mathcal{N}(\mu_Y(s_i), \sigma_Y(s_i))$.

Bias at each site is then defined with respect to deviations in the mean and standard deviations parameters, such that: $\mu_B(s_i) = \mu_Y(s_i) - \mu_Z(s_i)$ and $\sigma_B(s_i) = \sigma_Y(s_i)/\sigma_Z(s_i)$. The bias in the parameters is assumed independent of the true unbiased values that generate the observations: $\mu_Y(s) \perp\!\!\!\perp \mu_B(s)$ and $\sigma_Y(s) \perp\!\!\!\perp \sigma_B(s)$. The spatial distribution of the bias and unbiased values of the parameters are modelled throughout the domain assuming the parameters are generated from independent underlying GPs. For example, considering the mean this gives $\mu_Y(s) \sim \mathcal{GP}(m_{\mu_Y}(s), k(s,s'))$ and $\mu_B(s) \sim \mathcal{GP}(m_{\mu_B}(s), k(s,s'))$, where $m(s)$ represents the

mean function and $k(s, s')$ the covariance function, otherwise known as the kernel. In the case of the standard deviation, a logarithmic transformation is utilised to map the parameter space onto that of a GP, specifically $log(\sigma(s)) = \tilde{\sigma}(s) \in \mathbb{R}$. The transformed parameter and its bias are then modelled as generated from GPs, such that: $\tilde{\sigma}_Y(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_Y}(s), k(s, s'))$ and $\tilde{\sigma}_B(s) \sim \mathcal{GP}(m_{\tilde{\sigma}_B}(s), k(s, s'))$. The parameters for the climate model output $\mu_Z(s)$ and $\tilde{\sigma}_Y(s)$ are then considered as generated from the sum of these independent latent GPs, one that generates unbiased estimates of the parameters and one that generates the bias in the climate model.

In Carter et al., 2024 the mean function for each underlying GP is taken as a constant for simplicity. In this paper, since a clear global relationship is shown between the mean June temperature with elevation and latitude in Fig. 4.7, a multiple linear relationship is included such that $m_\mu(s) = \beta_{0_\mu} + \beta_{1_\mu} \cdot x_1 + \beta_{2_\mu} \cdot x_2$. The predictors $x_1$ and $x_2$ are standardised versions of the elevation and latitude respectively. For the standard deviation in June temperatures, only a comparatively weak relationship is observed between with elevation and latitude, so the corresponding generating GP is assumed to have constant mean $m_{\tilde{\sigma}}(s) = \beta_{0_{\tilde{\sigma}}}$. Similarly, the mean function of the GPs that generate the bias in both parameters is assumed constant.

Another adaptation for the real-world scenario implemented in this paper is that the kernel is chosen to be a Matern 3/2 instead of a radial basis function (RBF). This is typically preferred for real data where the covariance is not as smooth across the domain. The distance metric between points for the kernel function is computed from approximately Euclidean rotated geodesic grid latitude and grid longitude coordinates. A single lengthscale is assumed along with kernel variance and noise hyper-parameters. Finally, in Carter et al., 2024 the model is run in a fully Bayesian hierarchical set-up, whereas here it is chosen to split the model inference into two separate stages. Firstly the mean function is fit (section 4.3.2) and then the underlying spatial covariance structures considered (section 4.3.3).

This is done for aiding understanding of model performance at each stage as well as for computational reasons. The final bias correction of the climate model output is done through quantile mapping, where uncertainty in the unbiased parameter predictions is propagated to the final bias corrected climate model timeseries, which is explained in section 4.3.4.

### 4.3.2 Stage one: Fitting the mean function

The first stage involves estimating the parameter values for the observations and climate model output at each location where there is data $(\boldsymbol{\mu}_Y, \tilde{\boldsymbol{\sigma}}_Y, \boldsymbol{\mu}_Z, \tilde{\boldsymbol{\sigma}}_Z)$. The following global mean functions are fit for each parameter, where $x_1$ and $x_2$ are standardised versions of the elevation and latitude respectively:

$$m_{\mu_Y}(s) = \beta_{0_{\mu_Y}} + \beta_{1_{\mu_Y}} \cdot x_1 + \beta_{2_{\mu_Y}} \cdot x_2 \tag{4.1}$$

$$m_{\tilde{\sigma}_Y}(s) = \beta_{0_{\tilde{\sigma}_Y}} \tag{4.2}$$

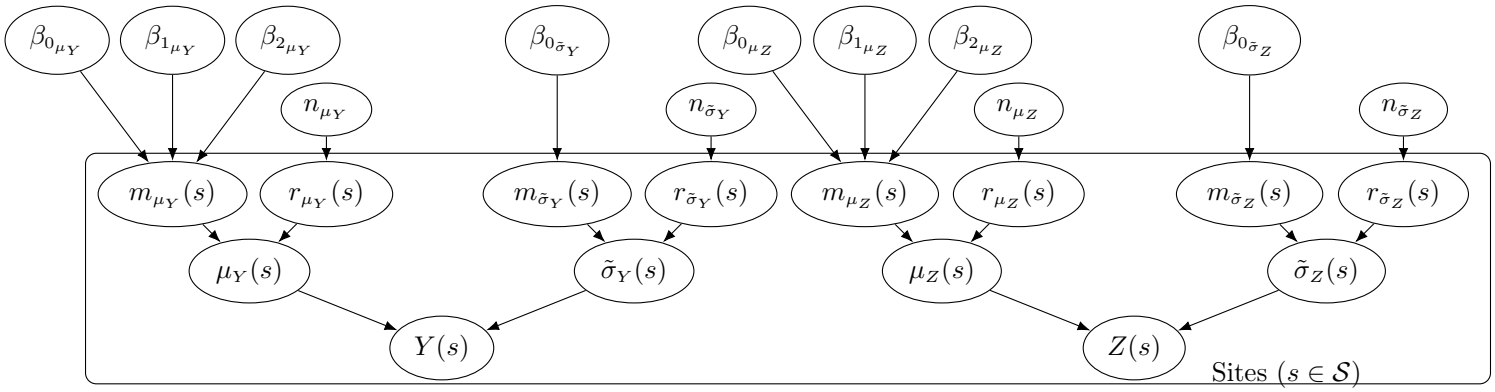
$$m_{\mu_Z}(s) = \beta_{0_{\mu_Z}} + \beta_{1_{\mu_Z}} \cdot x_1 + \beta_{2_{\mu_Z}} \cdot x_2 \tag{4.3}$$

$$m_{\tilde{\sigma}_Z}(s) = \beta_{0_{\tilde{\sigma}_Z}} \tag{4.4}$$

$$\tag{4.5}$$

The bias in the mean function components across the ice sheet are then given as $m_{\mu_B}(s) = m_{\mu_Z}(s) - m_{\mu_Y}(s)$ and $m_{\tilde{\sigma}_B}(s) = m_{\tilde{\sigma}_Z}(s) - m_{\tilde{\sigma}_Y}(s)$. The parameters $\mu(s)$ and $\tilde{\sigma}(s)$ at each location are modelled as the sum of the mean function component and some residual. The residual in this stage is modelled as generated from a normal distribution with mean zero and some noise parameter. The model is illustrated in Fig. 4.10, with underlying parameters obtained from MCMC inference with the no-u-turn sampler (NUTS) implemented from Numpyro. The residual component is the part of the data not explained by the Antarctic-wide relationships, so the component not explained by a constant or linear relationships with elevation and latitude. To estimate the residual components away from observations the underlying spatial

covariance structures are considered, which is done in stage two of the methodology (section 4.3.3).



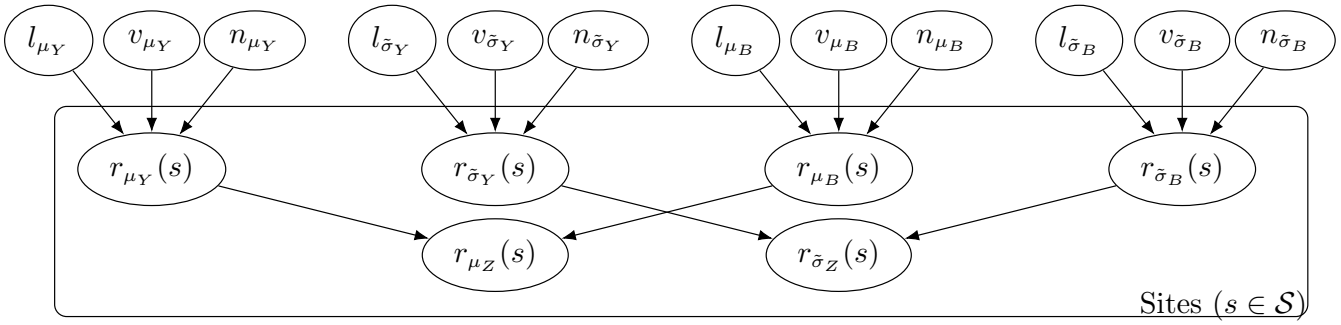

$$Y(s) \sim (\mu_Y(s), \sigma_Y(s)) \qquad \tilde{\sigma}_Y(s) = log(\sigma_Y(s)) \qquad \mu_Y(s) = m_{\mu_Y}(s) + r_{\mu_Y}(s) \qquad \mu_Z(s) = m_{\mu_Z}(s) + r_{\mu_Z}(s)$$
$$Z(s) \sim (\mu_Z(s), \sigma_Z(s)) \qquad \tilde{\sigma}_Z(s) = log(\sigma_Z(s)) \qquad \tilde{\sigma}_Y(s) = m_{\tilde{\sigma}_Y}(s) + r_{\tilde{\sigma}_Y}(s) \qquad \tilde{\sigma}_Z(s) = m_{\tilde{\sigma}_Z}(s) + r_{\tilde{\sigma}_Z}(s)$$
$$m_{\mu_Y}(s) = \beta_{0_{\mu_Y}} + \beta_{1_{\mu_Y}} \cdot x_{ele} + \beta_{2_{\mu_Y}} \cdot x_{lat} \qquad m_{\tilde{\sigma}_Y}(s) = \beta_{0_{\tilde{\sigma}_Y}} \qquad r_{\mu_Y}(s) \sim (0, n_{\mu_Y}) \qquad r_{\tilde{\sigma}_Y}(s) \sim (0, n_{\tilde{\sigma}_Y})$$
$$m_{\mu_Z}(s) = \beta_{0_{\mu_Z}} + \beta_{1_{\mu_Z}} \cdot x_{ele} + \beta_{2_{\mu_Z}} \cdot x_{lat} \qquad m_{\tilde{\sigma}_Z}(s) = \beta_{0_{\tilde{\sigma}_Z}} \qquad r_{\mu_Z}(s) \sim (0, n_{\mu_Z}) \qquad r_{\tilde{\sigma}_Z}(s) \sim (0, n_{\tilde{\sigma}_Z})$$

Figure 4.10: Plate diagram with directed acyclic graph showing the model for stage one of the bias correction methodology for the case where the site-level distributions are assumed normal with parameters $\mu$ and $\sigma$.

### 4.3.3 Stage two: Considering underlying spatial covariance structures

In stage one of the methodology posterior realisations are generated for $\boldsymbol{\mu}_Y$ and $\tilde{\boldsymbol{\sigma}}_Y$ at each in situ observation site, and for $\boldsymbol{\mu}_Z$ and $\tilde{\boldsymbol{\sigma}}_Z$ at each climate model grid cell. Additionally these parameters are modelled as being composed of a mean component and residual component, for which posterior realisations are also provided. In this second stage the residual components are estimated across the ice sheet by modelling the underlying spatial covariance in the values from stage one. In particular, the aim is to make estimates of the unbiased residuals $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ at the climate model grid cell locations for the purpose of bias correction. To do this residuals are modelled as generated from latent GPs with zero mean and hyper-parameters for the kernel including a kernel lengthscale ($l$), kernel variance ($v$) and kernel noise

$(n)$. The residuals for the climate model output are modelled as composed of the sum of the residuals from the observations and some independent bias component, such that $r_{\mu_Z}(s) = r_{\mu_Y}(s) + r_{\mu_B}(s)$ and $r_{\tilde{\sigma}_Z}(s) = r_{\tilde{\sigma}_Y}(s) + r_{\tilde{\sigma}_B}(s)$. It is then these independent components that are modelled as generated from GP distributions with Matern3/2 kernel and the model is illustrated in Fig. 4.11.




$$r_{\mu_Z}(s) = r_{\mu_Y}(s) + r_{\mu_B}(s) \qquad r_{\tilde{\sigma}_Z}(s) = r_{\tilde{\sigma}_Y}(s) + r_{\tilde{\sigma}_B}(s)$$
$$r_{\mu_Y}(s) \sim \mathcal{GP}(0, k(s, s'|v_{\mu_Y}, l_{\mu_Y})) \qquad r_{\tilde{\sigma}_Y}(s) \sim \mathcal{GP}(0, k(s, s'|v_{\tilde{\sigma}_Y}, l_{\tilde{\sigma}_Y}))$$
$$r_{\mu_B}(s) \sim \mathcal{GP}(0, k(s, s'|v_{\mu_B}, l_{\mu_B})) \qquad r_{\tilde{\sigma}_B}(s) \sim \mathcal{GP}(0, k(s, s'|v_{\tilde{\sigma}_B}, l_{\tilde{\sigma}_B}))$$

Figure 4.11: Plate diagram with directed acyclic graph showing the full hierarchical model for the case where the site-level distributions are assumed normal with parameters $\mu$ and $\sigma$. The distribution of these parameters across the domain is modelled with Gaussian processes.

Posterior realisations of the hyper-parameters of the model are again obtained from MCMC inference with the NUTS sampler. Posterior predictive estimates of $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ throughout the domain can then be made after conditioning on the values observed at the AWS sites as well as the values for $r_{\mu_Z}(s)$ and $r_{\tilde{\sigma}_Z}(s)$ at the climate model locations.

### 4.3.4 Quantile Mapping

The final bias correction of the climate model output is done through quantile mapping, which involves applying the following transformation to the climate model output ($\boldsymbol{z}_{s_{i,j}}$) for every record $j$ at each location $i$: $\hat{\boldsymbol{z}}_{\boldsymbol{s_{i,j}}} = F^{-1}_{Y_{s_i}}(F_{Z_{s_i}}(\boldsymbol{z}_{s_{i,j}}))$. The

term $F^{-1}_{Y_{s_i}}$ represents the inverse cumulative density function (CDF) for the in situ observational data at the site $s_i$, while the term $F_{Z_{s_i}}$ is the CDF for the climate model output at the equivalent site. The CDF is given as the integral over the PDF at each site and can be computed using posterior predictive realisations of $[\mu_{Y_{s_i}}, \tilde{\sigma}_{Y_{s_i}}]$ and $[\mu_{Z_{s_i}}, \tilde{\sigma}_{Z_{s_i}}]$ respectively. Making estimates of $\mu_{Y(s)}$ and $\tilde{\sigma}_{Y(s)}$ at the locations of the climate model grid cells is done through estimates of the components $m_{\mu_Y}(s)$, $r_{\mu_Y}(s)$, $m_{\tilde{\sigma}_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ at the climate model locations. The components $m_{\mu_Y}(s)$ and $m_{\tilde{\sigma}_Y}(s)$ are estimated from Eq. 4.1 and 4.2 respectively using standardised values of elevation and latitude available from the climate model. The residual components $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ are estimated at the climate model locations taking posterior predictive estimates from the corresponding generating GP with hyper-parameters given from the posterior realisations after fitting stage two of the model. Applying the quantile mapping procedure multiple times for different realisations of the parameter estimates gives an expectation and uncertainty band for the bias corrected output.

## 4.4 Results

Results are presented for bias correcting near-surface air temperature climate model output from an Antarctic-wide MAR RCM simulation using in situ AWS observations. The data is aggregated to monthly time scales and data is filtered to just the month of June, with stations that have less than five annual June records excluded, as are any sites that fall outside the main ice sheet shape file, as in Fig. 4.4b. The climate model output is scaled up for computational reasons before fitting the mean function or spatial model, averaging groups of 14 x 14 cells, reducing the resolution from $12.25 \cdot 12.25$ km$^2$ to $171.5 \cdot 171.5$ km$^2$. The difference between the elevation for the mesh of the climate model and the point like in-situ observations is accounted for by incorporating a mean function that is dependent on elevation in the model. Once the model is fit, the predictions on the climatology will depend

on the elevation product used, for example high-resolution digital elevation model output could be used to made predictions at high-resolution and downscale the climate model output. Final predictions of the unbiased parameters are made at the original, high-resolution grid. The elevation and latitude values for each data-set were standardised using the mean and standard deviation of values from the AWSs (elevation mean = 868.09, elevation std.dev. = 916.78, latitude mean = 76.65, latitude std.dev. = 5.73). Rotated geodesic grid latitude and grid longitude coordinates are chosen to compute L2 distances between points. Prior distributions for each parameter were chosen to be relatively non-informative. The full model is split into two stages as described in section 4.3. Results from the inference of each stage are presented in sections 4.4.1 and 4.4.2 respectively. The final joint prediction of the unbiased mean and log-variance are presented in section 4.4.3 along with the bias. Bias correction to the climate model timeseries of June monthly temperatures is demonstrated in section 4.4.4.

### 4.4.1 Mean Function Model Results

The mean and log-variance parameters of the marginal distribution for both the monthly June AWS output at each station site and the coarsened climate model output at each grid-cell are estimated. A multiple linear regression model is assumed for the mean parameter dependency with elevation and latitude, see section 4.3.2. The log-variance parameter is only weakly correlated with elevation and latitude, as shown in Fig. 4.7, so no equivalent dependency is assumed. The prior distributions for the parameters of the model are displayed in Fig. C.1 of the appendix, along with the posterior distributions after inference. The summary statistics for the distributions are presented here in Table 4.1. The same model is fit separately to the AWS data and the climate model output.

The standard deviations and 95% confidence intervals in the model parameters posterior distributions, shown in Table 4.1, are lower in magnitude for the climate model output than AWS data. The expectation for posterior realisations of the

intercept terms ($\beta_{0,\mu}$ and $\beta_{0,log(\sigma^2)}$) and the regression coefficient with elevation ($\beta_{1,\mu}$) are quite similar between the two datasets. The expectation for posterior realisations of the regression coefficient with latitude ($\beta_{2,\mu}$) shows greater dependency for the climate model output (2.8) than the AWS observations (1.92). Additionally, lower values for the expectations of the noise parameters ($n_\mu, n_{log(\sigma^2)}$) are shown for the climate model output compared to the AWS output. The R-squared scores for the mean $\mu$ after fitting the model are ${R^2}_{AWS} = 0.71 \pm 0.05$ and ${R^2}_{C.M.} = 0.82 \pm 0.01$ respectively.

| Dependent Variable | Model Parameters | Distribution | Exp. | Std. Dev. | 95% C.I. L.B. | 95% C.I. U.B. |
|---|---|---|---|---|---|---|
| Mean | $\beta_{0,\mu}$ | Prior | -33.70 | 10.00 | -53.30 | -14.10 |
| | | Posterior AWS | -33.61 | 0.68 | -34.90 | -32.26 |
| | | Posterior Climate Model | -33.07 | 0.30 | -33.68 | -32.47 |
| | $\beta_{1,\mu}$ | Prior | -10.00 | 5.00 | -19.80 | -0.20 |
| | | Posterior AWS | -10.13 | 0.66 | -11.42 | -8.84 |
| | | Posterior Climate Model | -10.32 | 0.25 | -10.82 | -9.85 |
| | $\beta_{2,\mu}$ | Prior | 0.00 | 5.00 | -9.80 | 9.80 |
| | | Posterior AWS | 1.92 | 0.70 | 0.58 | 3.30 |
| | | Posterior Climate Model | 2.80 | 0.22 | 2.38 | 3.22 |
| | $n_\mu$ | Prior | 5.00 | 2.88 | 0.26 | 9.75 |
| | | Posterior AWS | 6.85 | 0.51 | 5.89 | 7.84 |
| | | Posterior Climate Model | 5.03 | 0.16 | 4.70 | 5.34 |
| Log Variance | $\beta_{0,log(\sigma^2)}$ | Prior | 2.00 | 1.00 | 0.04 | 3.96 |
| | | Posterior AWS | 2.02 | 0.07 | 1.88 | 2.16 |
| | | Posterior Climate Model | 2.13 | 0.02 | 2.09 | 2.17 |
| | $n_{log(\sigma^2)}$ | Prior | 1.00 | 0.58 | 0.05 | 1.95 |
| | | Posterior AWS | 0.56 | 0.07 | 0.43 | 0.69 |
| | | Posterior Climate Model | 0.36 | 0.02 | 0.32 | 0.39 |

Table 4.1: A table showing summary statistics for the prior and posterior distributions of the parameters for the global mean function model. Statistics include 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 expectation as well as the uncertainty in estimates of the mean and log-

variance parameters for the AWS and coarsened climate model output are shown for each site and grid-cell respectively in Fig. 4.12. Strong correlation (0.90) is shown for the mean parameter expectations between the AWS sites and nearest-neighbour climate model output (Fig. 4.12a). Weaker correlation (0.62) is shown for expectations of the log-variance parameter (Fig. 4.12b). Uncertainties in the parameter estimates are given in Fig. 4.12c and d, which show greater uncertainty in the parameter estimates for the AWS sites than for the climate model output, with the uncertainty related to the number of records at each site.

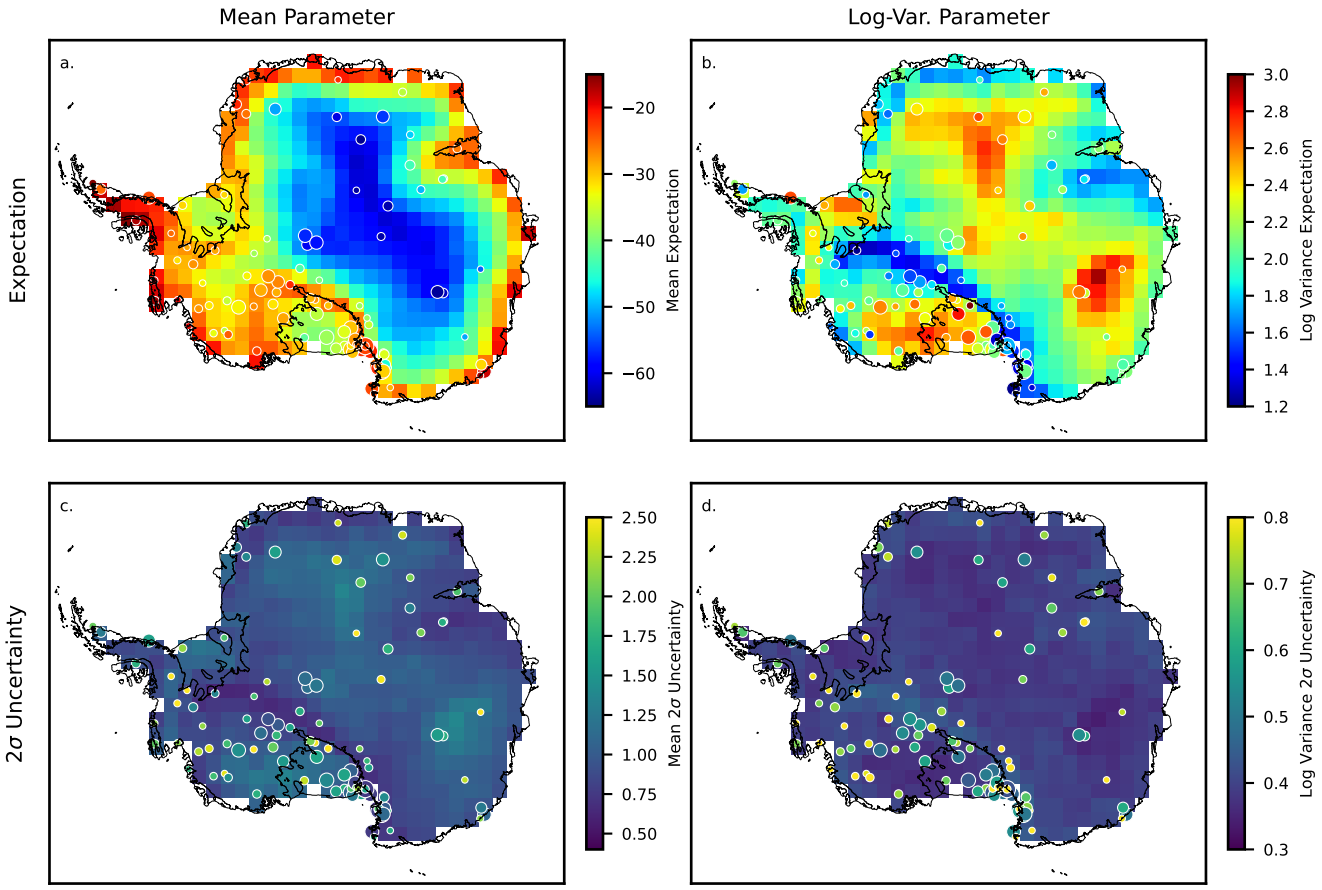



Figure 4.12: Expectations (a,c) and one sigma uncertainty (b,d) for posterior realisations of the mean and log-variance respectively in marginal distributions of the AWS at the station sites and climate model output at the grid cell locations. The size of the circles for AWS locations represents the number of monthly June records available at each site.

The parameter estimates shown in Fig. 4.12 are made considering a global mean function and some noise. The noise estimates at each location are here referred to as the residual, as mentioned in section 4.3.2. Residuals are computed for each AWS site and each climate model grid cell and shown in Fig. 4.13. There are obvious spatial correlation patterns in both sets of residuals and additionally correlation shown between the residuals of the two datasets. A zoomed-in region is illustrated over Victoria land where there's a relatively high density of AWS sites (Fig. 4.13c,d),

with sharp variations shown between values in some nearby sites. This indicates the importance of considering a noise term in stage two, where residual estimates are made across the ice sheet considering underlying GPs that explicitly capture spatial covariance. The residual estimates from Fig. 4.13 are used as input data for stage two.

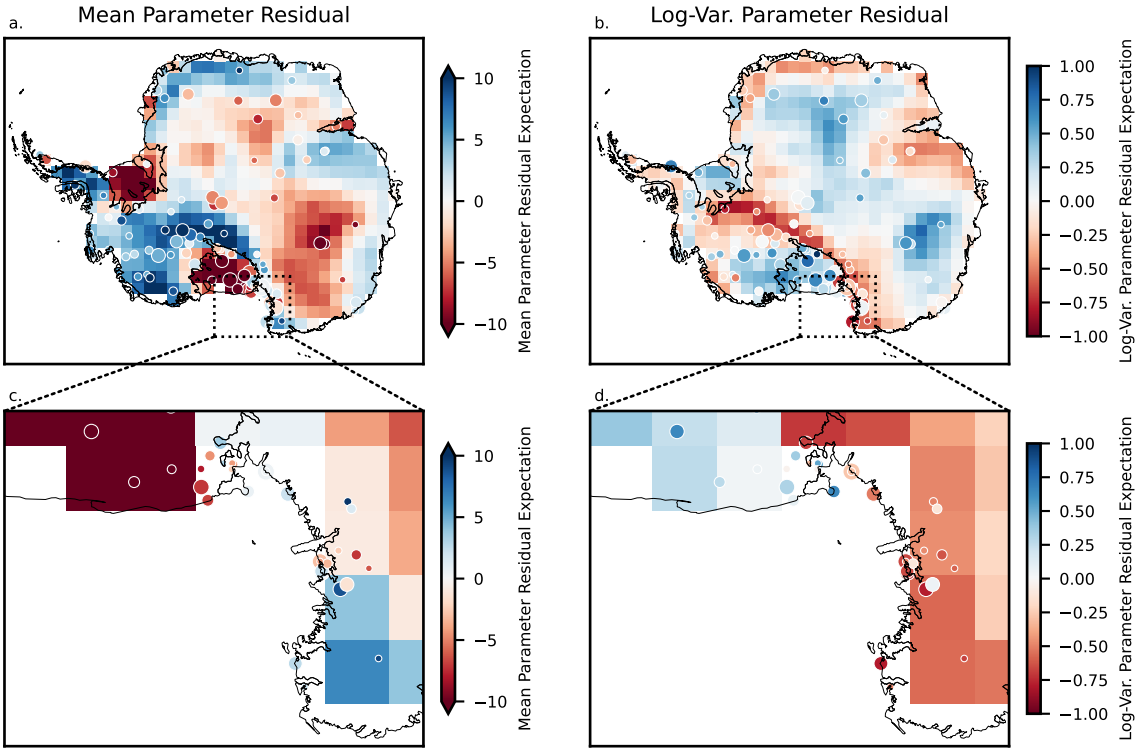



Figure 4.13: Expectations for posterior realisations of the residual in the mean and log-variance (a,b) at the station sites and grid cell locations. A zoomed in region is shown in c and d. The size of the circles for AWS locations represents the number of monthly June records available at each site.

### 4.4.2 Residuals Spatial Model Results

The components of the mean and log-variance not explained by global functions, referred to here as the residuals, are modelled across the domain assuming latent generating GPs. The residuals for the climate model output are modelled as

composed of the sum of an unbiased and biased component, as explained in section 4.3.3. The aim is to get estimates of the residuals $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ throughout the domain by considering their spatial covariance as well as spatial covariance in the bias. The input data to the model consists of the residuals from stage one, with uncertainty incorporated during inference of the GP hyper-parameters. The GP hyper-parameters consist of a lengthscale, kernel variance and additional noise term that helps account for any sharp variations. Comparisons are made between the approach outlined in this paper, referred to as the shared process approach (section 4.3.3), and the approach of just modelling covariance in the observations alone, referred to as the single process approach. The prior distributions for the hyper-parameters of the model are displayed in Fig. C.2 and C.3 of the appendix, along with the posterior distributions after inference. The summary statistics for the distributions are presented here in Table 4.2, with Fig. C.5 of the appendix providing a reference for the relative magnitude of the lengthscale values against the ice sheet extent.

The expectation of the lengthscale hyper-parameter for the unbiased residuals in the mean and log-variance ($l_{\mu_Y}$ and $l_{log(\sigma^2_Y)}$) is greater for the shared process than the single process model, while the kernel variance ($v_{\mu_Y}$ and $v_{log(\sigma^2_Y)}$) and noise terms ($n_{\mu_Y}$ and $n_{log(\sigma^2_Y)}$) are greater in the single process model. The spread of the posterior distributions for all hyper-parameter estimates, given by the standard deviation and credible intervals, is less for the shared process model than the single process model. An important benefit of the shared process model is direct modelling of the bias. The expectations for the lengthscale hyper-parameters of the latent bias processes ($l_{\mu_B}$ and $l_{log(\sigma^2_B)}$) are over twice the values for the unbiased process. Expectations for the kernel variance of the bias processes ($v_{\mu_B}$ and $v_{log(\sigma^2_B)}$) are significantly lower than for the unbiased process, as are expectations for the noise hyper-parameter for the mean ($n_{\mu_B}$). The expectation for the noise hyper-parameter of the log-variance ($n_{log(\sigma^2_B)}$) is similar to the value for the unbiased process.

| Dependent Variables | Model Parameters | Distribution | Exp. | Std. Dev. | 95% C.I. L.B. | 95% C.I. U.B. |
|---|---|---|---|---|---|---|
| Mean Residuals ($r_{\mu_Y}$,$r_{\mu_B}$) | $v_{\mu_Y}$ | Prior | 50.05 | 28.84 | 2.60 | 97.50 |
| | | Posterior Shared Process | 43.95 | 14.61 | 22.29 | 75.98 |
| | | Posterior Single Process | 46.33 | 15.14 | 22.30 | 79.11 |
| | $l_{\mu_Y}$ | Prior | 5.77 | 2.75 | 1.24 | 10.30 |
| | | Posterior Shared Process | 4.67 | 0.94 | 3.04 | 6.61 |
| | | Posterior Single Process | 3.83 | 1.13 | 1.96 | 6.05 |
| | $n_{\mu_Y}$ | Prior | 10.05 | 5.74 | 0.60 | 19.50 |
| | | Posterior Shared Process | 6.11 | 1.26 | 3.65 | 8.58 |
| | | Posterior Single Process | 9.45 | 2.88 | 4.36 | 15.33 |
| | $v_{\mu_B}$ | Prior | 15.00 | 8.66 | 0.75 | 29.25 |
| | | Posterior Shared Process | 10.09 | 6.80 | 0.09 | 23.75 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |
| | $l_{\mu_B}$ | Prior | 11.04 | 5.80 | 1.50 | 20.58 |
| | | Posterior Shared Process | 13.34 | 4.57 | 5.95 | 21.07 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |
| | $n_{\mu_B}$ | Prior | 2.00 | 1.15 | 0.10 | 3.90 |
| | | Posterior Shared Process | 0.57 | 0.56 | 0.00 | 1.74 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |
| Log Variance Residuals ($r_{log(\sigma^2_Y)}$,$r_{log(\sigma^2_B)}$) | $v_{log(\sigma^2_Y)}$ | Prior | 0.376 | 0.216 | 0.020 | 0.731 |
| | | Posterior Shared Process | 0.139 | 0.084 | 0.039 | 0.271 |
| | | Posterior Single Process | 0.153 | 0.094 | 0.039 | 0.342 |
| | $l_{log(\sigma^2_Y)}$ | Prior | 11.042 | 5.798 | 1.502 | 20.583 |
| | | Posterior Shared Process | 5.024 | 1.683 | 2.625 | 8.315 |
| | | Posterior Single Process | 5.762 | 2.349 | 2.106 | 10.697 |
| | $n_{log(\sigma^2_Y)}$ | Prior | 0.013 | 0.007 | 0.001 | 0.024 |
| | | Posterior Shared Process | 0.007 | 0.005 | 0.000 | 0.017 |
| | | Posterior Single Process | 0.010 | 0.007 | 0.000 | 0.022 |
| | $v_{log(\sigma^2_B)}$ | Prior | 0.200 | 0.115 | 0.011 | 0.390 |
| | | Posterior Shared Process | 0.068 | 0.060 | 0.002 | 0.194 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |
| | $l_{log(\sigma^2_B)}$ | Prior | 11.042 | 5.798 | 1.502 | 20.583 |
| | | Posterior Shared Process | 13.720 | 4.737 | 5.573 | 21.056 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |
| | $n_{log(\sigma^2_B)}$ | Prior | 0.013 | 0.007 | 0.001 | 0.024 |
| | | Posterior Shared Process | 0.010 | 0.006 | 0.000 | 0.021 |
| | | Posterior Single Process | N/A | N/A | N/A | N/A |

Table 4.2: A table showing summary statistics for the prior and posterior distributions of the parameters for the global mean function model. Statistics include the expectation (Exp.), standard deviation (Std. Dev.) and lower and upper bounds for the 95% credible interval (C.I. L. and C.I. U.).

Posterior realisations of the hyper-parameters allow multiple realisations of the latent generating GPs to be defined. After conditioning these GP realisations on the input data, posterior predictive estimates of the unbiased residuals ($r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$) can be made across the domain, as well as estimates of the bias in the residuals ($r_{\mu_B}(s)$ and $r_{\tilde{\sigma}_B}(s)$). Posterior predictive estimates of the unbiased residuals are generated at the locations of the high-resolution climate model grid and the expectations shown in Fig. 4.14 for both the shared and single process models, with uncertainties shown in Fig. 4.15. The posterior predictive estimates for the bias residuals are shown in Fig. C.4c,d of the appendix.

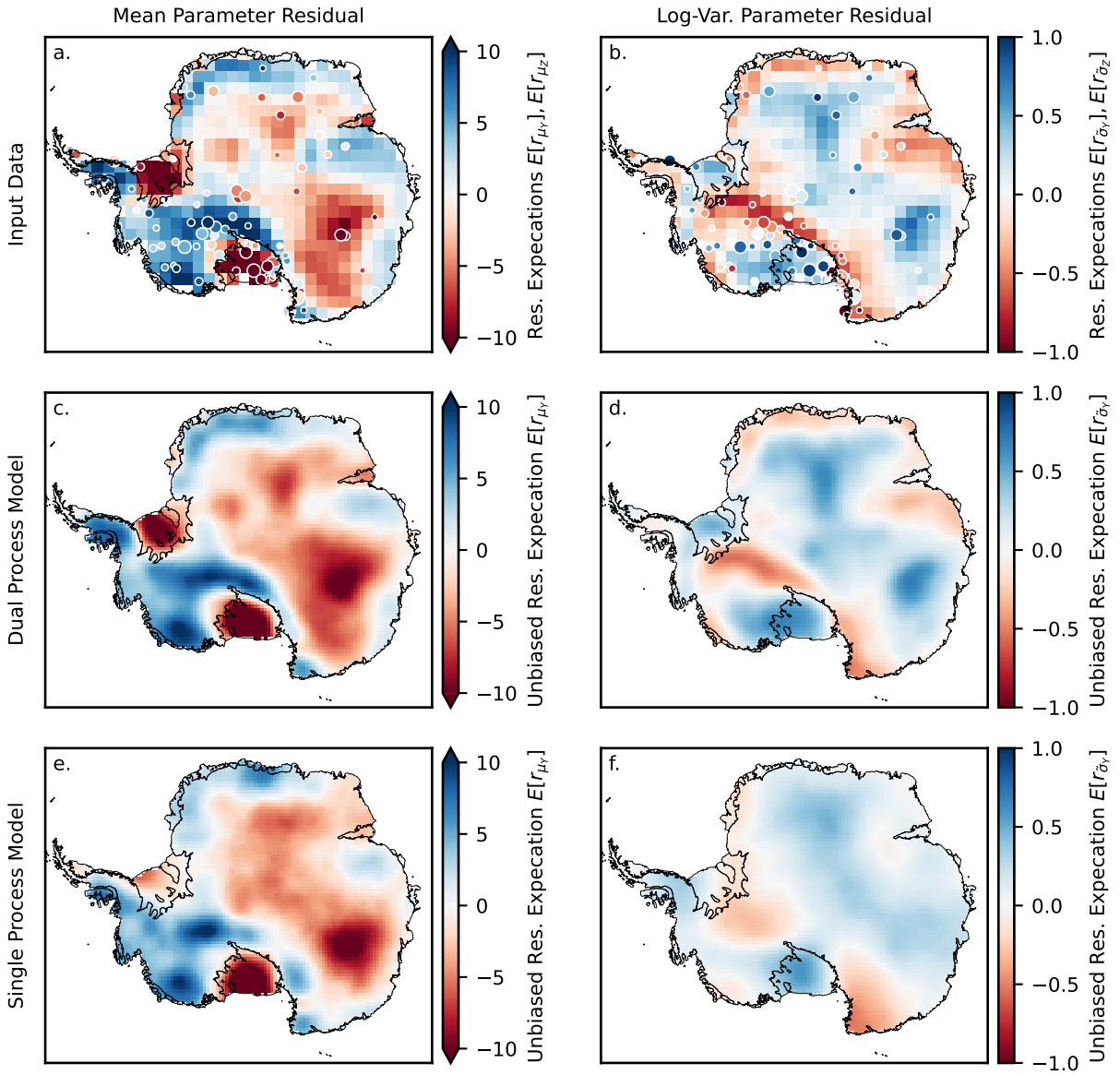



Figure 4.14: The expectation in posterior predictive estimates of the residuals $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ at the locations of the high-resolution climate model grid for the dual (c,d) and single process models (e,f). The input data of $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ at the AWS sites and $r_{\mu_Z}(s)$ and $r_{\tilde{\sigma}_Z}(s)$ at the coarse climate model grid locations are also shown (a,b).

In both the shared and single process models the posterior predictive estimates of the unbiased process across the domain are smoothly varying and appear visually

to align with the observational input data (Fig. 4.14). Posterior predictive estimates for the shared process (Fig. 4.14c,d) are additionally seen to capture and incorporate features of the climate model residuals in predictions away from observational locations. An example of this is over the Ronne ice shelf, where only one in situ AWS is present and while predictions from the single process model (Fig. 4.14e,f) tend towards zero away from this observation, predictions for the shared process model tend towards the values given by the climate model. The average uncertainty in estimates over the ice sheet is shown to be significantly greater in the single process model (Fig. 4.15c,d) than the shared process model (Fig. 4.15a,b). In both models reduced uncertainty is shown near the AWS observation locations. Patterns in the uncertainty away from observations are more pronounced in the single process model predictions.

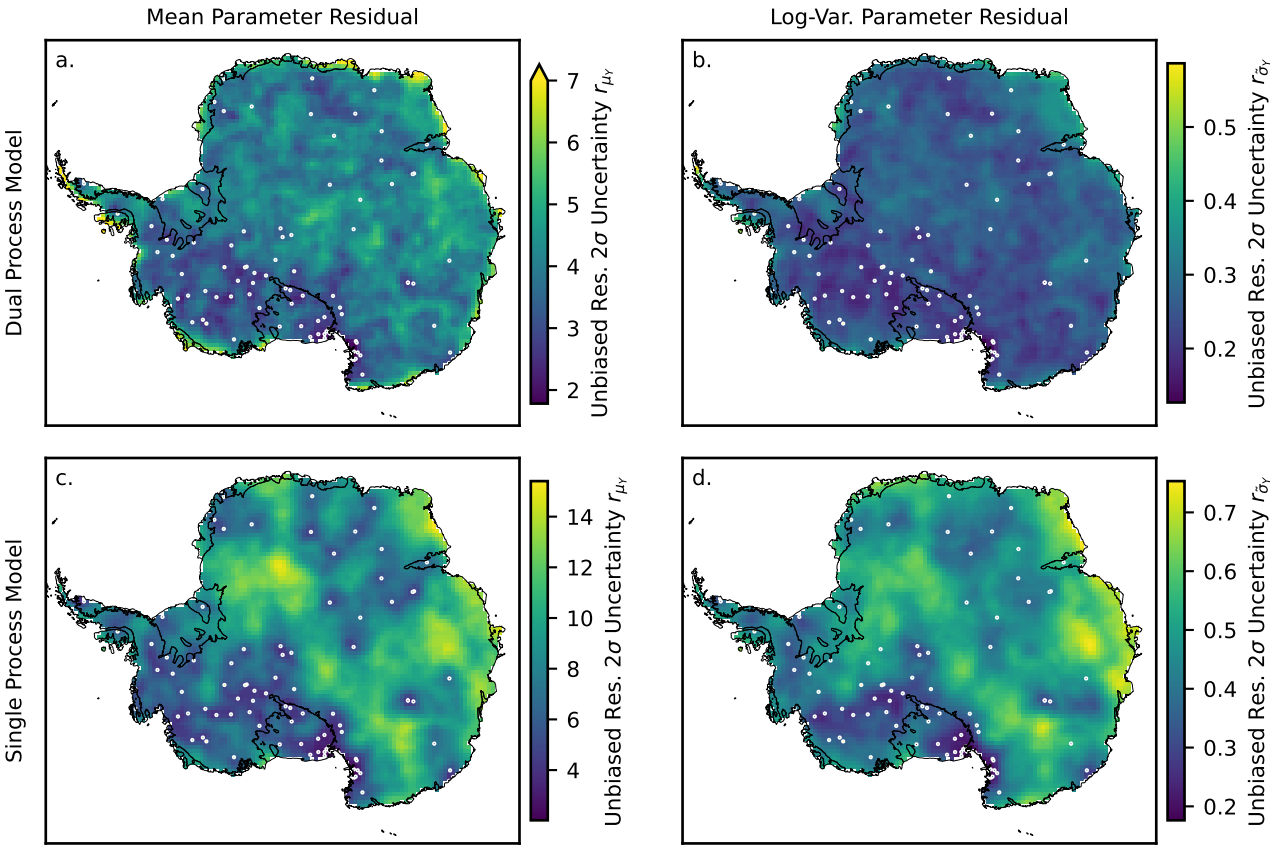



Figure 4.15: The two sigma uncertainty in posterior predictive estimates of the residuals $r_{\mu_Y}(s)$ and $r_{\tilde{\sigma}_Y}(s)$ at the locations of the climate model grid for the single process (c,d) and shared process models (a,b). The location of AWS sites are shown for reference as white markers.

### 4.4.3 Final Joint Prediction

The final predictions for the unbiased mean and log-variance across the domain are made from the sum of expectations for the mean function and residual, such that: $E[\mu_Y(s)] = E[m_{\mu_Y}(s)] + E[r_{\mu_Y}(s)]$ and $E[\tilde{\sigma}_Y(s)] = E[m_{\tilde{\sigma}_Y}(s)] + E[r_{\tilde{\sigma}_Y}(s)]$. The resulting estimates are shown in Fig. 4.16a,b. Uncertainties are computed from summing up the respective variances and taking two times the square root to get the two sigma range. The noise term of the latent GPs in section 4.4.2 that accounts for sharp variations in the residuals is also incorporated, such that: $V[\mu_Y(s)] =$

$V[m_{\mu_Y}(s)]+V[r_{\mu_Y}(s)]+E[n_{\mu_Y}(s)]$ and $V[\tilde{\sigma}_Y(s)] = V[m_{\tilde{\sigma}_Y}(s)]+V[r_{\tilde{\sigma}_Y}(s)]+E[n_{\tilde{\sigma}_Y}(s)]$. Resulting uncertainties are shown in Fig. 4.16c,d. Estimates are only shown considering the shared process model for the residuals, not the single process model.

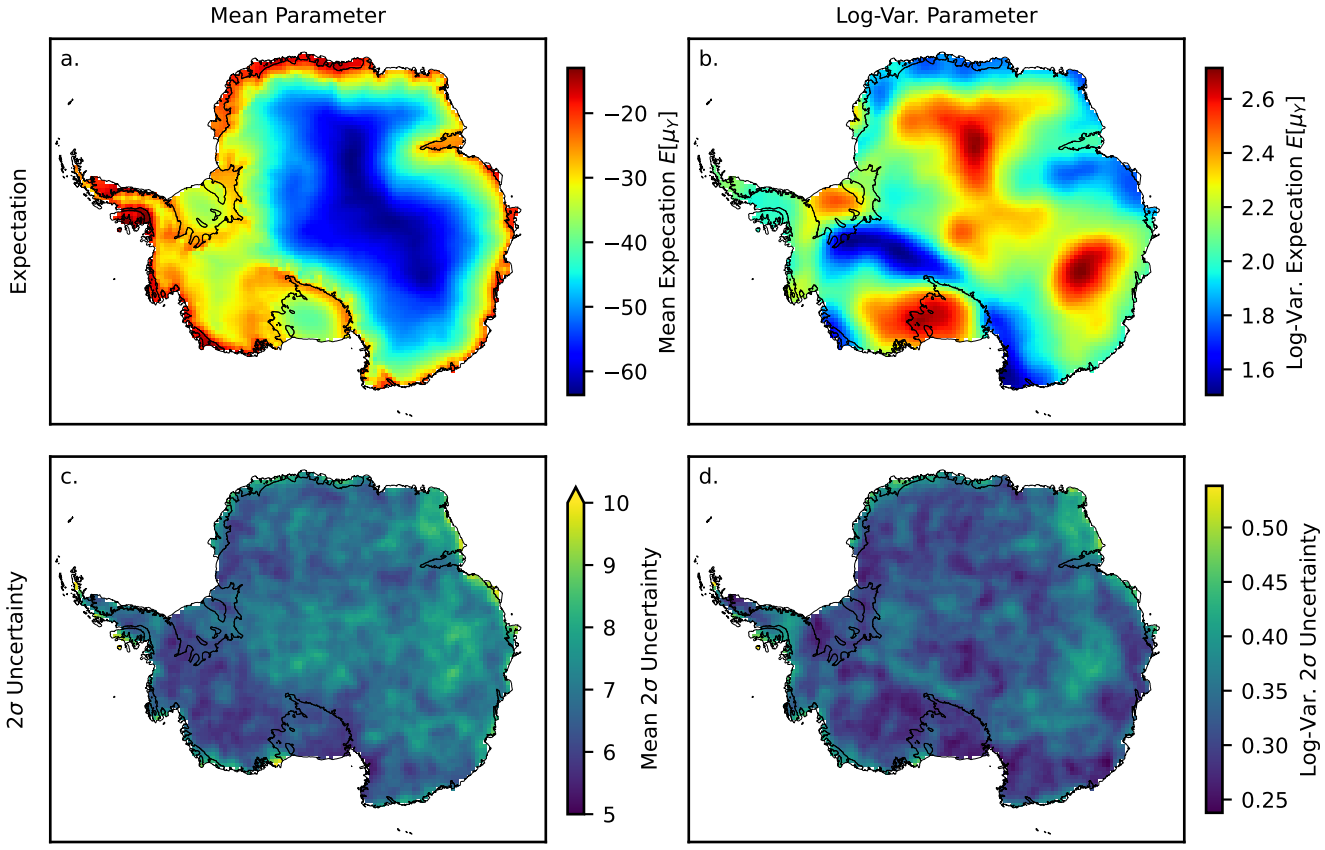



Figure 4.16: The expectations (a,b) and uncertainties (c,d) of unbiased estimates of the mean ($\mu_Y$) and log-variance ($\tilde{\sigma}_Y$) at the locations of the high-resolution climate model grid.

The expectation and uncertainty in predictions of the final joint bias across the domain are shown in Fig. 4.17. Predictions of the bias in the mean June temperatures range from approximately -5 °C. to +5 °C. Predictions of the bias in the log-variance of monthly June temperatures range from approximately -0.2 to +0.1. Taking the exponential of the bias in the log-variance and then the square root gives an estimate of the ratio between the standard deviations of the climate

model output and in situ observations ($\sigma_Z/\sigma_Y$). Estimates of this ratio then vary between approximately $\sqrt{(e^{-0.2})} = 0.90$ and $\sqrt{(e^{0.1})} = 1.05$. The uncertainty of the bias estimates incorporates a noise term in the latent GPs that generate the residual component of the bias. This noise term accounts for sharp variations not captured by modelling the residuals with a stationary, single lengthscale GP.

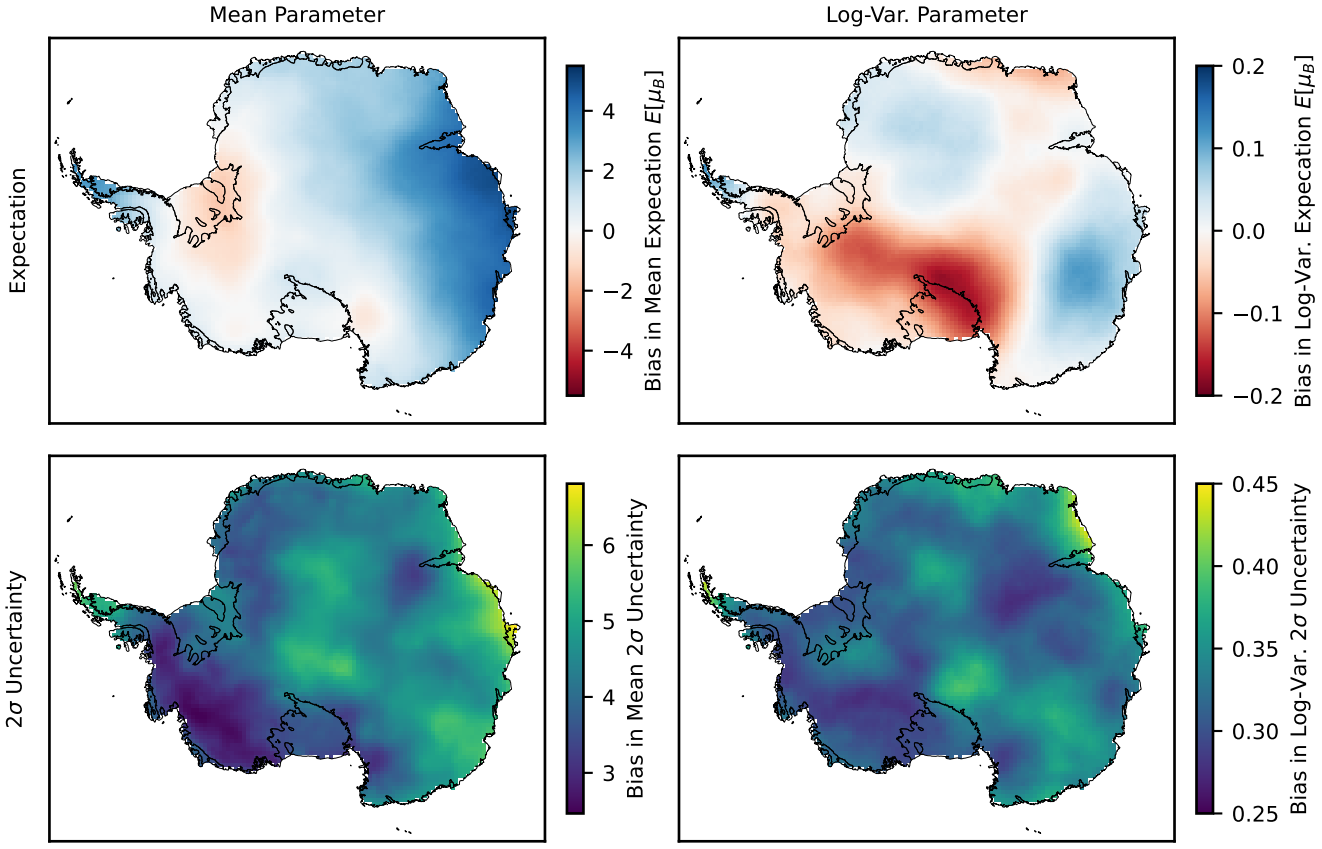



Figure 4.17: The expectations (a,b) and uncertainties (c,d) of the bias in estimates of the mean ($\mu_B$) and log-variance ($\tilde{\sigma}_B$) at the locations of the high-resolution climate model grid.

### 4.4.4 Bias Corrected Timeseries

To apply bias correction to the monthly June temperatures output from the high-resolution climate model the empirical mean and log-variance is first computed for each grid cell. Estimates of the unbiased mean and log-variance at these locations

(Fig. 4.16) are then used to perform quantile mapping to the timeseries from the climate model output. Instead of using the expectation of posterior predictive parameter estimates, realisations of the posterior predictives of the unbiased parameters are utilised and quantile mapping performed with each realisation separately. This allows propagation of uncertainty, producing multiple realisations for the corrected timeseries and thus an uncertainty band on the correction. To illustrate this six AWS sites are chosen and the timeseries from the AWS records plot alongside the timeseries for the original and corrected climate model output from the nearest grid cell. Additionally, the original and bias corrected climate model timeseries are shown for a site isolated from any nearby AWSs. These seven locations are illustrated in Fig. 4.18, with results presented in Fig. 4.19.

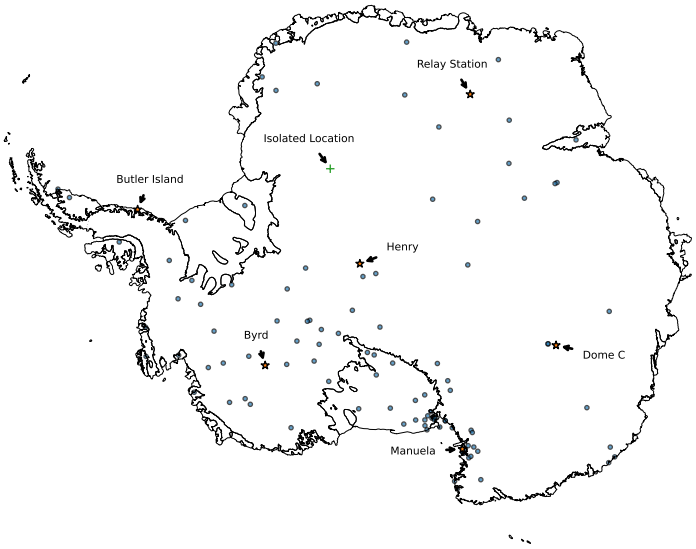



Figure 4.18: Locations of AWS sites used in predictions across the ice sheet. Six AWS sites are highlighted where results for the bias corrected timeseries will be compared against. An isolated location is also marked where the bias corrected timeseries is examined.

The expectation of realisations for the bias corrected time series in Fig 4.19 shows a visual shift in the corrected climate model output towards closer agreement with

nearby AWS records. The two sigma uncertainty band of the correction covers the AWS timeseries at each site. In the case of the isolated location the expectation for the corrected time series is very close to the original climate model output, although with large uncertainty bands. A couple of interesting features are shown for the AWS records, including an anomalously high temperature-record for the Butler Island AWS (Fig. 4.19c) occurring in the year 1988 and an apparent step change in the bias for the Relay AWS (Fig. 4.19e) when comparing the period 1995-2005 with the period 2010-2021.

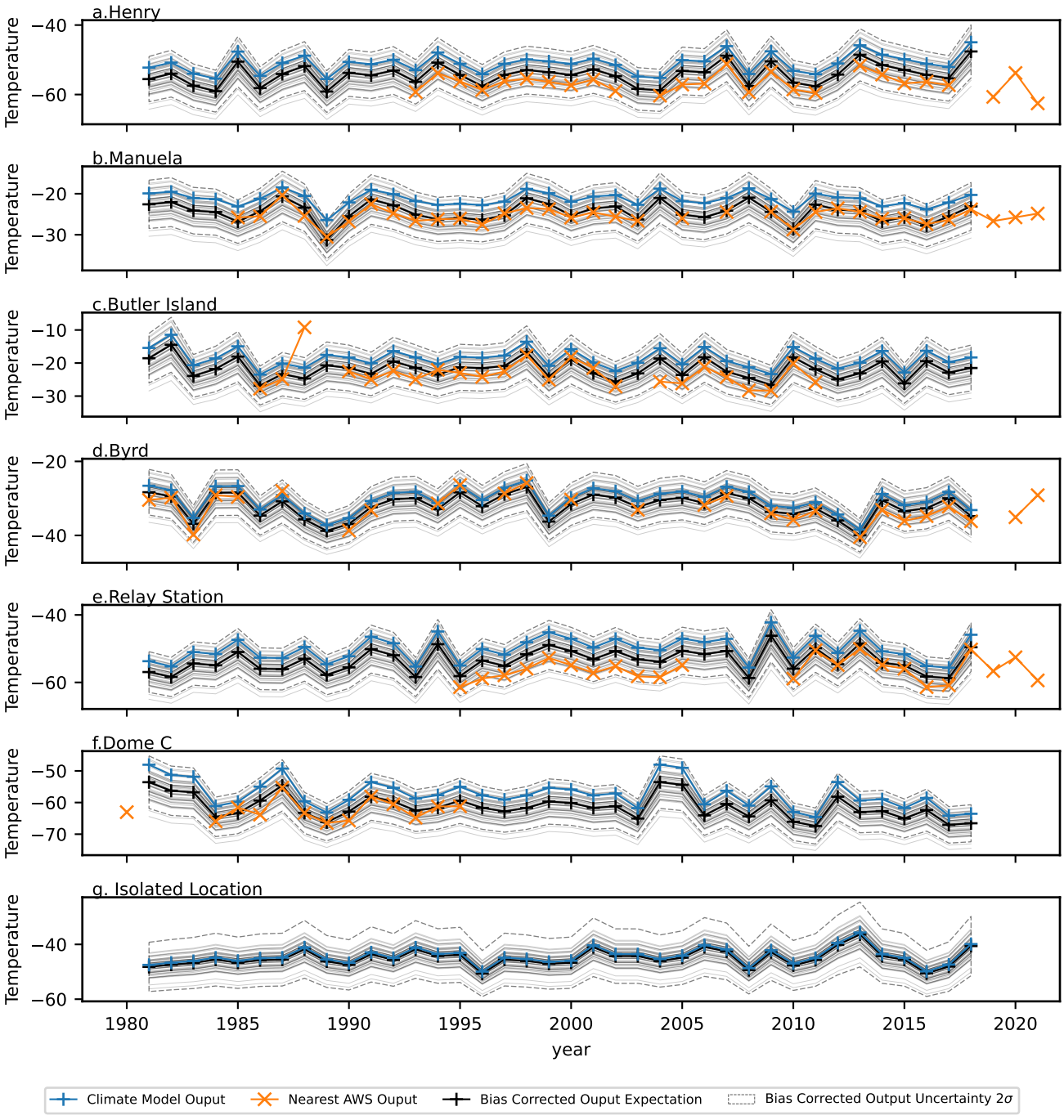



Figure 4.19: Average June temperature records for six AWS sites and the equivalent output from the nearest grid cell for the original and bias corrected climate model timeseries (a-f). The timeseries for the original and bias corrected climate model output is also demonstrated for a site isolated from any nearby AWSs (g).

## 4.5 Discussion

A novel methodology for bias prediction and correction of climate model output is applied to the case study of correcting bias in monthly near-surface air temperature output over Antarctica from the MARv3.10 RCM. Bias is defined with respect to the time-independent marginal PDFs at each site. To simplify the form of the PDF the data is filtered to just the month of June, for which the distribution can be approximated as normal with mean and variance parameters. Bias is then evaluated in these parameters and bias correction involves estimating the true unbiased values at each grid cell. These estimates are used to apply quantile mapping to the climate model output at each location to improve agreement with AWS records.

The unbiased parameter values across the domain are inferred from both AWS in situ observations and the climate model output itself. This is done in two stages: fitting a global mean function dependent on elevation and latitude (section 4.4.1) and subsequently fitting a spatial model to the residuals from the first step to make estimates away from the AWS sites and across the domain (section 4.4.2). Exploratory analysis showed a clear dependency between the mean June temperature with elevation and latitude, so multiple linear regression is used to capture this in stage one. A comparatively weak relationship was observed between the same predictors and the log-variance in June temperatures and so for the log-variance only a mean constant is fit. Results show that for both the climate model and AWS output, most of the variability in the mean June temperature can be explained by linear relationships with elevation and latitude ($R^2{}_{AWS} = 0.71 \pm 0.05$ and $R^2{}_{C.M.} = 0.82 \pm 0.01$ respectively). This highlights the potential benefit and importance of downscaling using higher resolution elevation data than was utilised in the RCM simulation, as for example is done in Noël et al., 2023. Differences in the relationships with elevation and latitude between the AWS records and the climate model output induce a bias, with the most significant difference shown for the relationship with latitude. The component of the mean and log-variance not explained by the relationship with elevation and latitude is referred to as the residual.

The expectations for the residuals show spatial covariance patterns across the ice sheet (Fig. 4.13) and strong correlations between expectations for the AWS records and climate model output. This indicates the advantage of incorporating patterns shown in the climate model output as well as in AWS records when estimating the unbiased component across the ice sheet, which is done in stage two of the methodology.

In stage two of estimating the unbiased PDF parameters across the ice sheet, the residual estimates from stage one are modelled across the ice sheet considering latent GP distributions that inherently capture spatial covariance. Residuals from the climate model output are considered the sum of an unbiased and biased component, where the unbiased component also generates the AWS residuals directly. This is referred to as the shared process or dual process model. Considering shared latent processes between the two data sets allows estimates to incorporate spatial features from both datasets, which is illustrated in Fig. 4.14. Direct comparisons are made to an approach proposed in Lima et al., 2021 where the unbiased component is estimated from in situ observations alone, here referred to as the single process model, with hyper-parameter estimates for the two models illustrated in Table 4.2. Reduced credible intervals shown in the case of the shared process model show an advantage of providing extra information in the form of residuals from the climate model output. This extra information also results in smaller expectation values for the kernel variance and noise hyper-parameters. Conditioned forms of the latent GP distributions allow estimation of the unbiased residuals and the bias across the domain. Performance of the two models near observations appears visually similar, but away from observations some clear differences in the expectations (Fig. 4.14) and uncertainties (Fig. 4.15) exist with the shared process model capturing features from the climate model, such as the high negative values for the residuals over the south side of the Ronne ice shelf, that the single process model misses. This behaviour is desirable and reflects the belief that the climate model produces skilful estimates over the ice sheet, while afflicted with bias. Having direct estimates of the

bias from the shared process is also an advantage, with predictions across the ice sheet displayed in Fig. C.4 of the appendix.

The final predictions for PDF parameters and their bias across the ice sheet are composed from the sum of the mean function estimate, dependent on elevation and latitude, and the residual estimate, dependent on spatial patterns observed in the climate model and AWS datasets. Predictions for the unbiased component of the mean and log-variance across the ice sheet reflect similar patterns to those observed in the climate model output itself (Fig. 4.16), although smoothed and with closer agreement to AWS observations. The smooth nature of predictions is desirable in that it indicates the model is not over-fitting. While in reality the true underlying fields will contain more sharp variations and exhibit a spatial covariance lengthscale that varies depending on factors such as topography, to be able to statistically model this adequately would require a greater spatial density and coverage of AWS sites or the additional use of other datasets such as satellite products. Improved spatial density of unbiased observations would motivate the use of more complex latent processes in the modelling, with for example, multiple lengthscales considered and non-stationarity.

Considering the various sources of uncertainty in the final predictions for the mean and variance across the ice sheet, as well as their comparative magnitude, reveals where most advancements in predictions can be made. Uncertainty can be thought of as split between the estimation of the mean component and residual component. Uncertainty in the mean component comes from uncertainty in the parameters in stage one of the methodology. Uncertainty in the residual is the combined result of uncertainty in the input data (residuals from stage one), uncertainty in the hyper-parameters of the generating latent GPs, uncertainty in predictions due to the noise term in the GPs that account for variation not captured by a single lengthscale and uncertainty in predictions resulting from making predictions away from observations after conditioning on the input data. While the primary source of uncertainty is in the residual component rather than the mean

function component, it's important to note that including a more sophisticated mean function that captures more of the total variance would in turn likely contribute to a reduction in sources of uncertainty for the residual component. With respect to the residual component it's interesting to compare uncertainty between the single process and shared process models. Uncertainty in predictions from the single process model seem dominated by regions away from observations with clear peaks forming in the regions with the lowest AWS density (Fig. 4.15c,d). Uncertainties from the shared process are both smaller in magnitude and don't show as clear a dependency on AWS density, which is the result of utilising the climate model data itself. It is proposed that focusing on increasing spatial coverage of AWS sites and incorporating other forms of remote observations (such as satellite measurements) is the most obvious next step to improve bias correction of climate model output over the ice sheet. This would both reduce uncertainty in current techniques and allow further development of bias correction techniques to incorporate features such as non-stationary covariance lengthscales over the ice sheet.

Uncertainties propagated to the final bias corrected timeseries for June temperatures are illustrated for seven sites (Fig. 4.19). Six of these sites are deliberately chosen near AWS sites that contain many records and additionally are distributed widely across the ice sheet. For these locations improvement in the agreement with in situ observations is shown for the bias corrected climate model product and realistic uncertainty bands are demonstrated. Additionally, an isolated site chosen far away from any AWS, illustrates the desirable property that predictions of the unbiased component tend towards that of the climate model output far away from any counter information. This is desirable since the climate model is assumed to produce skilful estimates of the parameters. Interesting features are found in the AWS records themselves, including an apparent temperature anomaly and an apparent shift in bias after a break in AWS records for one of the sites. The temperature anomaly, not explained by the climate model output, is an important consideration in impact assessments. If physically meaningful and not just an instrumentation

error, it highlights a potential deficiency in the climate model output that will not be improved considering time-independent errors. Similarly, the step change in bias observed will also not be reflected in the bias corrected output if only time-independent biases are considered. If the step change is instead the result of biases induced by measurement techniques, it's important to consider this type of bias in the observations themselves. Although not conducted here, it is noted that the methodology could be further validated using a train-test approach such as Leave-One-Out Cross-Validation (LOOCV). In LOOCV a single AWS would be left out and used as a validation point, the approach is repeated for each observation in turn and an average performance metric computed.

## 4.6 Conclusions

In this paper monthly near-surface air temperature output from the MARv3.10 RCM is bias corrected using AWS in situ records. The novel bias correction approach proposed in Carter et al., 2024 is applied for the first time on real-world data. The approach models the data using latent generating GP distributions that explicitly consider underlying spatial covariance structures and uncertainties in predictions, which is of key importance for regions with sparse in situ observations. Shared generating processes are assumed for the climate model output and AWS records, which allows incorporation of spatial patterns seen in the climate model output to predictions of the unbiased underlying process. Initial data exploration motivates additionally fitting a global mean function capturing the dependency of temperature with elevation and latitude. For interpretability and computational reasons, inference for parameters of the global mean function is done separately to inference on parameters of the covariance structure for latent GPs. Bias in the climate model output is defined in terms of deviation in the time-independent marginal PDF at each location compared to the true distribution. To simplify the form of the PDF, bias correction can be applied to each month separately, while

in this paper results for the month of June are focused on. The marginal PDF distribution for monthly June temperatures can be approximated as normal with mean and variance parameters. Bias can then be defined in terms of deviations in these parameters from their true values, with AWS records assumed to provide unbiased estimates of the true values. Using AWS records and the climate model output itself, estimates of the unbiased parameters are then made across the domain.

The methodology shows promising results, with smoothly varying estimates of the unbiased marginal PDF parameters across the domain and with the final bias corrected time series showing improved agreement with AWS records as well as realistic uncertainty. The PDF parameters are modelled across the domain as composed of a mean component and a residual component. The mean component is modelled with global relationships, such as the dependency with elevation and latitude, with the primary source of bias in this component resulting from differences in the dependency with latitude between the AWS records and climate model output. The residual component is modelled spatially using latent GPs. Bias in the residual components are found to exhibit longer lengthscale spatial patterns compared to the unbiased residual components. Improvements in predictions for the unbiased residual component are demonstrated when considering shared latent GPs between the AWS records and climate model output, compared to the approach of only considering latent processes for the AWS records. The relative impact of different sources of uncertainty on predictions are discussed and it is hypothesised that the greatest improvements in corrections would come from increasing spatial coverage of AWS records. Increasing observational coverage would in turn motivate further development of bias correction techniques that incorporate factors such as non-stationary covariance lengthscales over the ice sheet.

# Chapter 5

# Conclusions

This thesis provides an in-depth examination of systematic differences in surface climatology across state-of-the-art Antarctic-wide RCM outputs, which motivates the subsequent development of a novel approach to post-processing bias correction. Chapter 2 provides the first inter-comparison of recent-historic Antarctic-wide RCM simulations framed within the context of ice shelf instability that specifically examines spatial patterns in systematic differences and that utilises time series decomposition to consider variations in different components of the time series separately. Significant magnitude differences are shown in the inter-annual mean and amplitude of periodic seasonal variations as well as monthly residual fluctuations. It is suggested that systematic differences across the ensemble limits the validity of utilising individual outputs in impact assessments and directly interpreting results, such as for prediction of the vulnerability of ice shelves to collapse events. In addition to considering the impact of differences across the ensemble on impact assessments, emphasis is also put on the importance of future observational campaigns with improved coverage and quality. These campaigns are useful for providing tighter constraints with which to both tune and update the climate model physics and parametrisations, as well as for direct use in reducing systematic differences through post-processing bias correction techniques.

Large-scale spatial patterns in the systematic differences are shown across the

ice sheet and are attributed primarily to differences in the RCM dynamical core, the surface scheme and the parametrisation and tuning. It is therefore suggested that further development of sophisticated techniques for bias correction are needed that explicitly incorporate underlying spatial covariances. Robust uncertainty estimation is also required due to both the sparsity of in situ observations and magnitude of systematic differences. Chapter 3 develops a novel Bayesian hierarchical model for bias correction that fills these requirements. Latent GP distributions are used to capture underlying covariance structures in both the observations and climate model output. A shared generating GP is assumed for both data sources, with an additional GP considered that generates bias in the climate model output. Simulated examples that cover a range of real-world scenarios show promising results, with improvements in the $R^2$ score for predictions of the unbiased field compared with the methodology from Lima et al., 2021 that doesn't consider shared processes between the data sources. Uncertainty estimation appears robust and uncertainty propagation to the final corrected time series is demonstrated, desirable for informing impact studies and the resulting decision making. Particular advantage of the methodology is demonstrated for the case of sparse observational data and a smooth bias, where considering shared latent processes provides significant value added in predictions. The methodology is therefore desirable for correcting surface climatology on Antarctica, where there exists sparse in situ observations and systematic differences between RCM outputs that exhibit large-scale spatial patterns as shown in chapter 2.

In chapter 4 the bias correction methodology developed in chapter 3 is tested further on the real-world scenario of correcting near-surface air temperature output from the Modèle Atmosphérique Régional version 3.10 RCM. Sparse automatic weather station records are used from the recent ice-sheet wide collation performed in Yetang Wang et al., 2023. The methodology shows promising results, producing smooth estimates for the unbiased underlying fields with realistic uncertainty. These estimates align with patterns shown in station observations and also capture

some features of the climate model output over regions far away from weather stations, where the skilful climate model predictions are utilised. Expectations and uncertainty bands are demonstrated for the bias corrected timeseries over several regions spanning the ice sheet. Comparisons at some of the sites to nearby weather station records show improved consistency and realistic uncertainty estimation. The implications of this work are significant, allowing impact studies to incorporate spatially dependent and robust uncertainty of any driving climate model output afflicted with bias. Additionally it provides support and information useful for increasing efforts towards in situ observational campaigns, suggesting a focus on improving spatial coverage of measurements to be incorporated into the bias correction model. While developed specifically for Antarctic climatology, the scope of the work is wide-ranging with the methodology being adaptable to other continents, variables of interest and even detection of bias in other sources of data than climatology. Future developments of the work are suggested in chapter 6.

# Chapter 6

# Future Directions

There are many exciting directions for further methodology development and topic research that are outside of the scope for this thesis but are discussed briefly here. These include aspects such as: expanding the bias correction methodology to incorporate non-stationary lengthscales for the covariance; including more sophisticated modelling of the relationships between temperature with elevation and latitude; considering time dependent systematic errors and time dependency between months; incorporating additional datasets in the bias correction, such as satellite measurements and other RCM simulations; capturing systematic errors in the in situ observations themselves; considering differences in the spatial support of the climate model (gridded area average) and in situ observations (point-like); including higher-resolution elevation data for downscaling; applying the methodology to additional variables and considering correlation across response variables in multi-output models; and improving the computational efficiency of the model through advancements in the model code and incorporation of techniques such as sparse GP regression. Non-stationary covariance lengthscales, time-dependent bias and including additional RCM simulations are discussed further below.

In the bias correction methodology presented, underlying Gaussian processes are modelled to have a constant spatial covariance lengthscale over the ice sheet. In reality, the lengthscale varies depending on factors such as the topography

with greater lengthscales observed over regions where the topography varies more smoothly, such as over the Antarctic plateau compared with the Antarctic Peninsula. Considering a single lengthscale with a noise term for the entire ice sheet captures the structure of the covariance over large distances but not at more local scales. This results in predictions of for example the unbiased mean temperature expectation being over-smoothed for regions with sharply varying topography, such as over the Antarctic Peninsula. Incorporating a non-stationary lengthscale in the modelling, where the value is allowed to change across the domain, adds complexity but would improve estimates over these regions of sharply varying topography, which is particularly important in the case of predicting collapse events for ice shelves off the Antarctic Peninsula where significant warming is observed.

An additional component of the bias correction methodology that could be further developed to improve estimates is consideration of time-dependence in the response variable and its bias. Currently the model assumes time-independent bias and additionally models bias in each month separately without considering temporal correlation between months. It is suggested that to improve this GPs could also be used to provide a distribution over the whole timeseries for each location. Bias can then be considered in the hyper-parameters of the temporal GPs, with the current methodology developed able to consider the spatial correlation in hyper-parameters. The mean function of the temporal GPs at each site could have some simple time-dependent relationship for example a linear trend with intercept and gradient hyper-parameters. In the case of near-surface air temperature, this formulation would allow the unbiased mean and variance to be predicted for every spatial location and time point. Quantile mapping could then be applied at every climate model output time and spatial coordinate separately to get a final bias corrected timeseries.

Finally, currently the model only incorporates climate model output from a single simulation. An ensemble of Antarctic-wide RCM simulations are available through the CORDEX project, as compared in chapter 2. The methodology could be extended to consider additional RCM output by considering each additional dataset

as generated by the sum of two latent generating processes, one of which is shared with the previous datasets and the other which generates a new bias to estimate. An assumption would be that the biases in each climate model are independent and the aim is to disaggregate an unbiased signal from the ensemble. All these advancements are exciting but outside the scope of the PhD, they would be expected to improve estimates but come at the cost of increased computational demands and complexity that potentially makes interpretation of findings more challenging.

# Appendix A

# Variability in Antarctic Surface Climatology Across Regional Climate Models and Reanalysis Datasets

## A.1 STL Decomposition

Figure A.1 shows an example of applying STL decomposition to the time series of snowfall, surface temperature and melt for a grid-cell on the Larsen C ice shelf. The decomposition has been applied to each of the 8 model outputs examined in this paper. The trend, seasonal and residual components are shown next to the original time series. Decomposing the time series into these components allows some features to be extracted. For example, in the case of snowfall and surface temperature the models all show high correlation in the inter-annual trend, although there exists a significant systematic difference in the mean between the models. Similarly, for snowfall and surface temperature there is high correlation in the residual term between the models but there is a systematic difference between the models in the standard deviation of that component. In the case of melt, the correlation is more moderate for the trend and residual components, meaning systematic differences are less obvious. The seasonal and residual components in STL decomposition are defined to have approximately zero mean.

The melt time series over Antarctica generally consists of two distinct periods, being the austral summer months (December, January and February) and then all other months of the year. Across all non-austral summer months, the melt magnitude is typically zero. The periods are therefore distinct in that the magnitude of melt over the austral summer doesn't inform the magnitude over the following non-summer months. Additionally, the variances over both periods are non-homogeneous. The impact this has on the STL decomposition is that any deviations in the magnitude of melt occurring over austral summer months, compared to the mean over these months, gets primarily treated as noise rather than as a component of the trend signal. Additionally, in the estimated noise component of the time series, non-zero magnitude melt events are attributed to non-austral summer months to account for the smooth trend component. To resolve these issues in future work, the distinct periods could be separated and the time series for melt over just the austral summer months analysed.

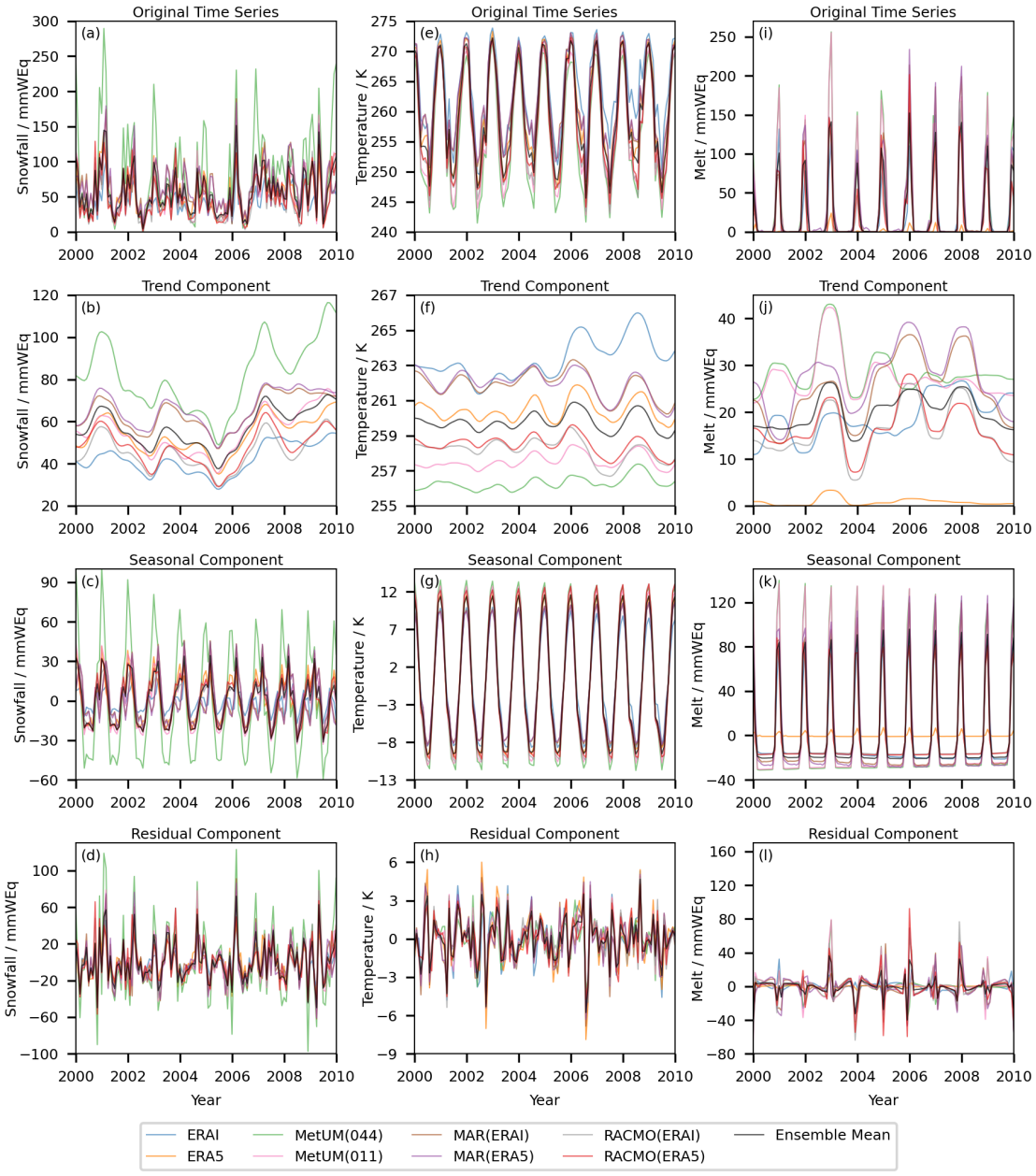



Figure A.1: An example of STL decomposition applied to the monthly time series of snowfall (a,b,c,d), surface temperature (e,f,g,h) and melt (i,j,k,l) for a grid-cell near the grounding line on the Larsen C ice shelf. The original time series for the years 2000-2010 are shown (a, e, i), as are the trend (b, f, j), seasonal (c, g, k) and residual (d, h, l) decompositions. The model is additive meaning the sum of trend, seasonal and residual components returns the original time series. Parameter values are $n_s = 13$ and $n_t = 21$.

## A.2 Same Core Model Differences

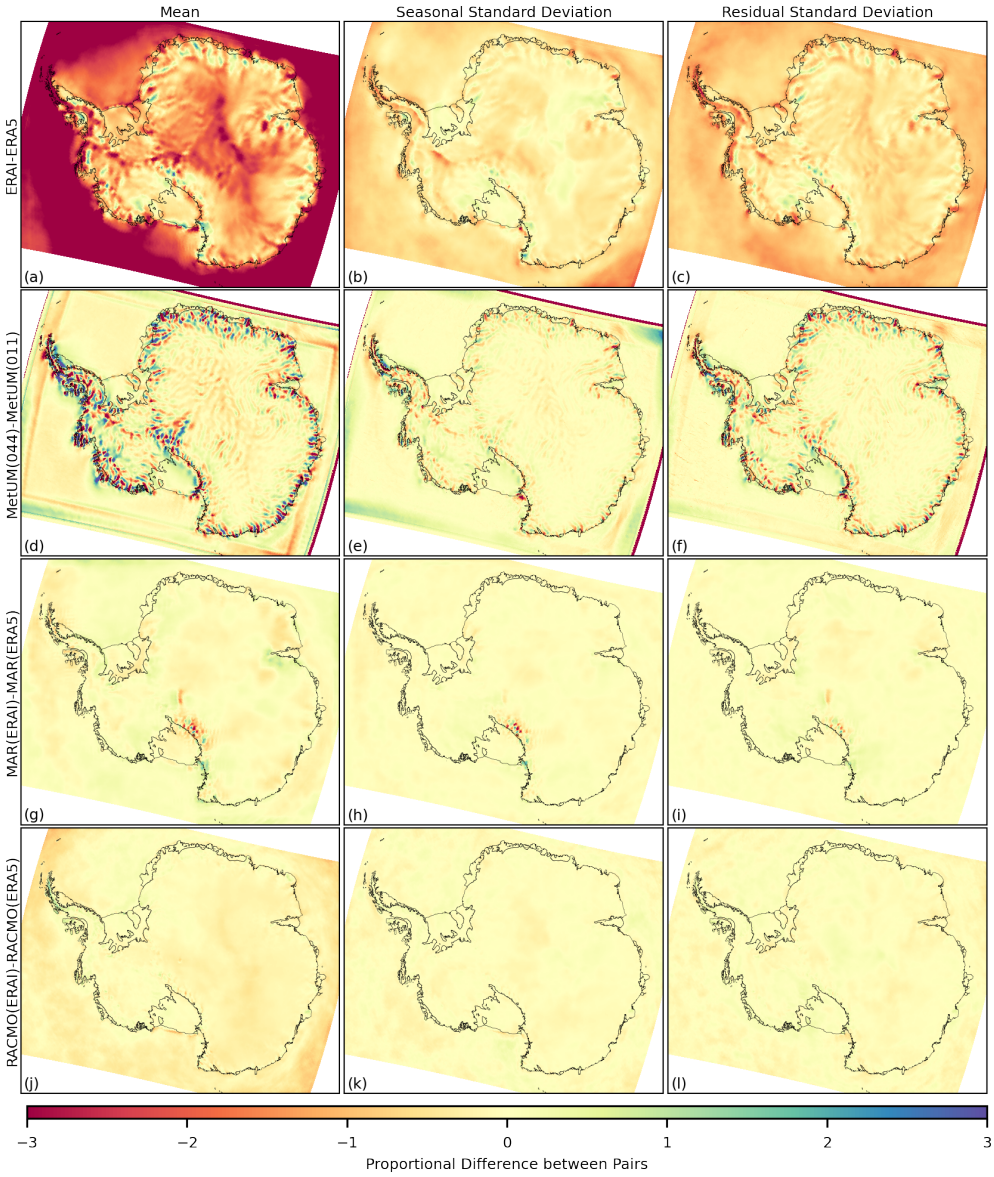



Figure A.2: The difference for the 1981-2018 time series of snowfall, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l) between the following pairs of outputs: ERA-Interim relative to ERA5 (a,b,c); MetUM(044) relative to MetUM(011) (d,e,f); MAR(ERAI) relative to MAR(ERA5) (g,h,i); RACMO(ERAI) relative to RACMO(ERA5) (j,k,l). Differences at each grid cell are expressed as a proportion of average inter-annual variation and so do not have units.

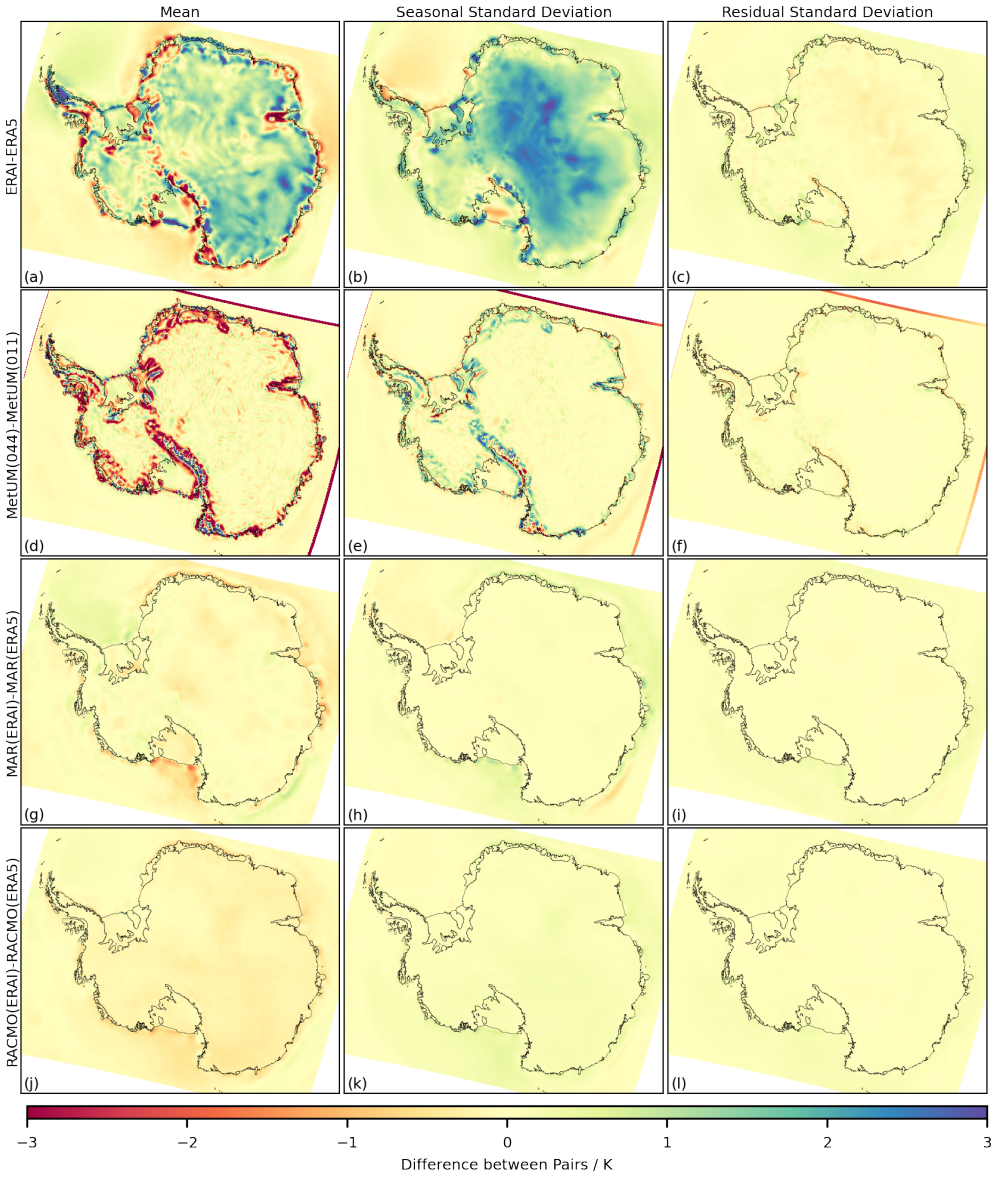



Figure A.3: The difference for the 1981-2018 time series of near-surface air temperature, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l) between the following pairs of outputs: ERA-Interim relative to ERA5 (a,b,c); MetUM(044) relative to MetUM(011) (d,e,f); MAR(ERAI) relative to MAR(ERA5) (g,h,i); RACMO(ERAI) relative to RACMO(ERA5) (j,k,l).

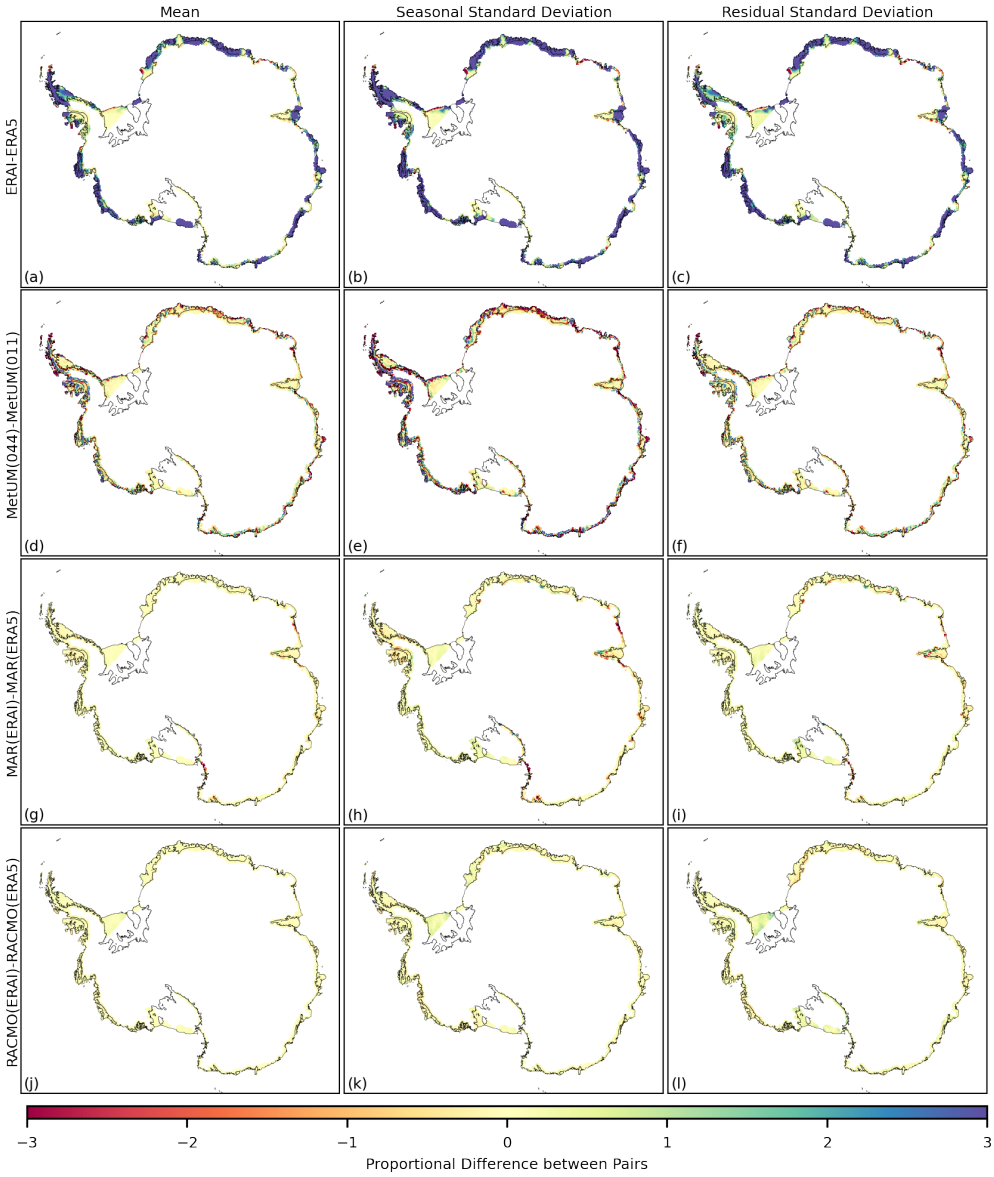



Figure A.4: The difference for the 1981-2018 time series of melt, in the mean (a,d,g,j), the standard deviation of the seasonal component (b,e,h,k) and the standard deviation of the residual component (c,f,i,l) between the following pairs of outputs: ERA-Interim relative to ERA5 (a,b,c); MetUM(044) relative to MetUM(011) (d,e,f); MAR(ERAI) relative to MAR(ERA5) (g,h,i); RACMO(ERAI) relative to RACMO(ERA5) (j,k,l). Differences at each grid cell are expressed as a proportion of average inter-annual variation and so do not have units. Grid-cells where the ensemble mean average monthly melt is less than 1 $mm\ w.e.\ m^{-1}$ are masked.

## A.3 DEM Differences

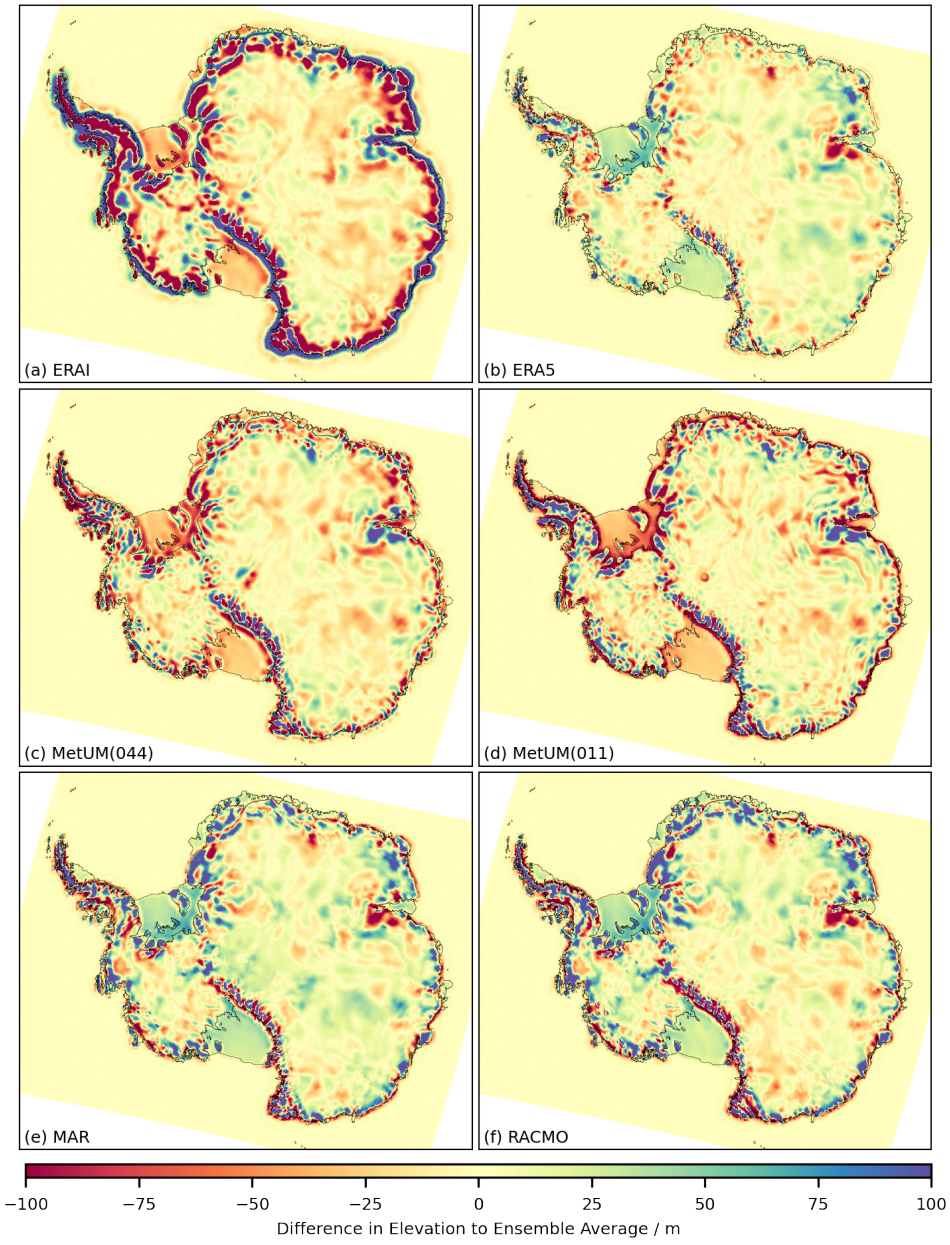



Figure A.5: The difference between the DEM used by each climate model is plot relative to the ensemble average (a. ERA-Interim, b. ERA5, c. MetUM(044), d. MetUM(011), e. MAR and f. RACMO). The DEMs are regrid onto the MetUM(011) 12.5 $km^2$ grid for comparison. Units are meters of elevation difference.

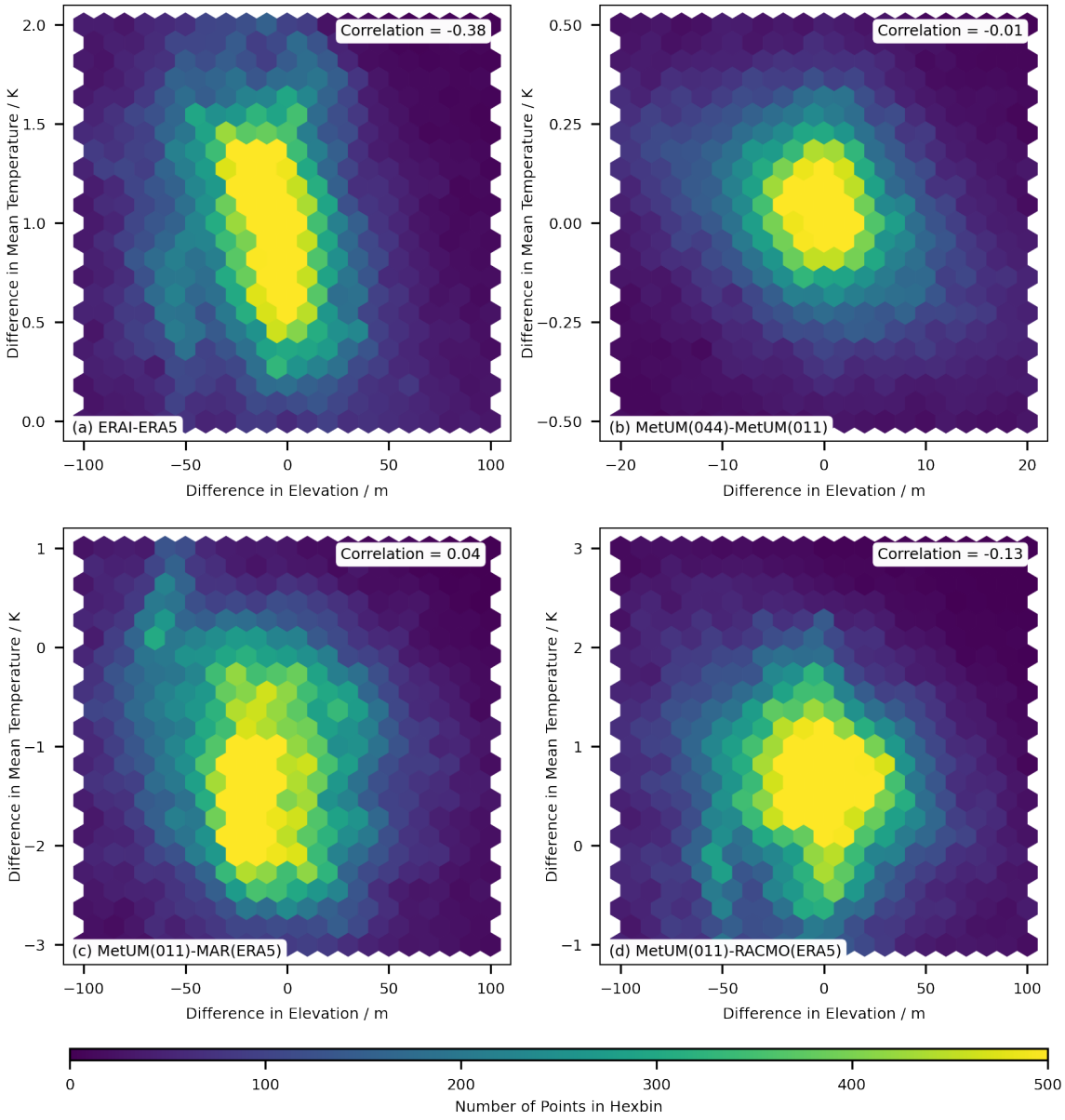



Figure A.6: A density scatter plot showing the correlation between the difference in elevation for each model relative to the ensemble and the difference for near-surface temperature in the mean of the time series (a), the standard deviation of the seasonal component (b) and the standard deviation of the residual component (c). The linear Pearson correlation coefficient is given for each plot.

## A.4 Correlation in the Trend Component

Figure A.7 compares the correlation in the residual (a,b,c) and trend component of the time series (d,e,f). The magnitudes and patterns shown are similar for the two components, with the greatest difference visible for NST. The pattern shown for NST where the MetUM simulations are more correlated with the reanalysis models than the MAR and RACMO simulations, is not visible for the trend component. This is hypothesised to indicate that the reinitialisation procedure that the MetUM simulations utilise provides most added value in improving correlation with the driving data over approaches in MAR and RACMO simulations for short frequency climatic events.

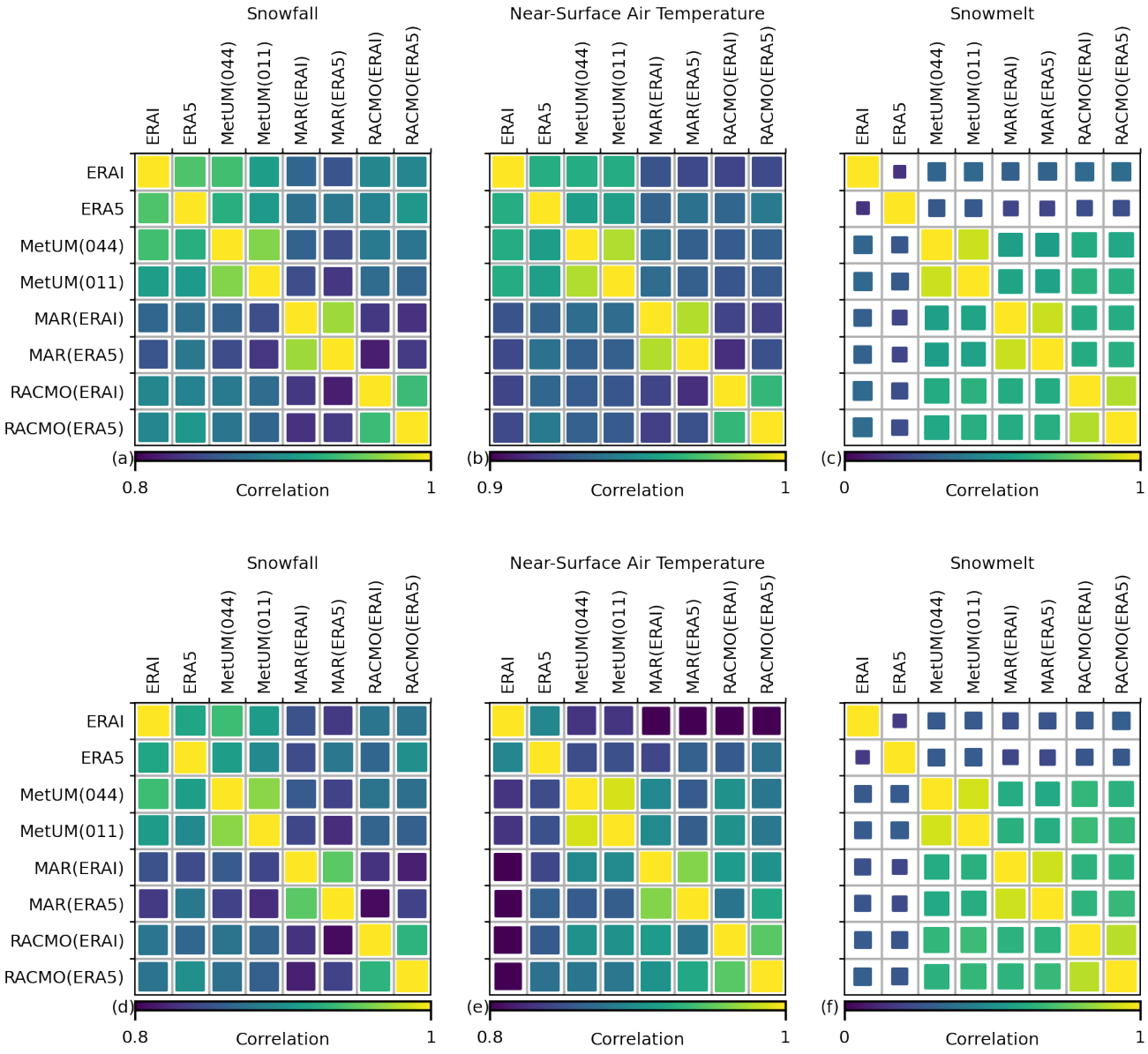



Figure A.7: The correlation in the residual (a,b,c) and trend component of the time series (d,e,f) for snowfall (a,d), near-surface air temperature (b,e) and melt (c,f) between models averaged over the ice-sheet. The colour scale relates to the value of correlation and the scale is adjusted for each plot. The size of each square also relates to the value of correlation, although is kept constant across the figures, going from 0-1, to make comparisons clear between the different variables.

# Appendix B

# Bias Correction of Climate Models using a Bayesian Hierarchical Model

# B.1 Posterior and Posterior Predictives

## B.1.1 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:

$$Y(s) \sim \mathcal{F}(\boldsymbol{\phi}_Y(s)) \tag{B.1}$$

$$Z(s) \sim \mathcal{F}(\boldsymbol{\phi}_Z(s)) \tag{B.2}$$

The symbols $\boldsymbol{\phi}_Y(s)$ and $\boldsymbol{\phi}_Z(s)$ represent the collection of parameters that describe the PDF at the site. For example if the PDF is normal, then $\boldsymbol{\phi}_Y(s) = [\mu_Y(s), \sigma_Y(s)]$ and $\boldsymbol{\phi}_Z(s) = [\mu_Z(s), \sigma_Z(s)]$.

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:

$$\boldsymbol{y} = [\boldsymbol{y}_{s_1}, \ldots, \boldsymbol{y}_{s_{n_y}}] \tag{B.3}$$

$$\boldsymbol{y}_{s_i} = [y_{s_i,1}, \ldots, y_{s_i,m_i}] \tag{B.4}$$

$$\boldsymbol{z} = [\boldsymbol{z}_{s_1}, \ldots, \boldsymbol{z}_{s_{n_z}}] \tag{B.5}$$

$$\boldsymbol{z}_{s_j} = [z_{s_j,1}, \ldots, z_{s_j,m_z}] \tag{B.6}$$

Also, 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:

$$\boldsymbol{\phi}_Y(\boldsymbol{s}_y) = [\boldsymbol{\phi}_Y(\boldsymbol{s}_1), \ldots, \boldsymbol{\phi}_Y(\boldsymbol{s}_{n_y})] \tag{B.7}$$

$$\boldsymbol{\phi}_Z(\boldsymbol{s}_z) = [\boldsymbol{\phi}_Z(\boldsymbol{s}'_1), \ldots, \boldsymbol{\phi}_Z(\boldsymbol{s}'_{n_z})] \tag{B.8}$$

The PDF parameters are themselves 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 be be written:

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

$$\phi_B(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\phi_B}) \tag{B.10}$$

$$\phi_Z(s) \sim \mathcal{GP}(\cdot,\cdot|\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B}) \tag{B.11}$$

The collection of hyper-parameters for the generating processes are given by $\boldsymbol{\theta}_{\phi_Y}$ and $\boldsymbol{\theta}_{\phi_B}$ respectively. 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. The posterior distribution for the model can then be written as:

$$P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B}|\boldsymbol{y},\boldsymbol{z}) = \frac{P(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B})\cdot P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B})}{P(\boldsymbol{y},\boldsymbol{z})} \tag{B.12}$$

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

$$P(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{y}|\boldsymbol{\phi}_Y(\boldsymbol{s}_y))\cdot P(\boldsymbol{z}|\boldsymbol{\phi}_Z(\boldsymbol{s}_z)) \tag{B.13}$$

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

$$P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y)|\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B})\cdot P(\boldsymbol{\phi}_Z(\boldsymbol{s}_z)|\boldsymbol{\theta}_{\phi_Y},\boldsymbol{\theta}_{\phi_B})\cdot P(\boldsymbol{\theta}_{\phi_Y})\cdot P(\boldsymbol{\theta}_{\phi_B}) \tag{B.14}$$

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 $\boldsymbol{\phi}_Y(\boldsymbol{s}_y)$, $\boldsymbol{\phi}_Z(\boldsymbol{s}_z)$, $\boldsymbol{\theta}_{\boldsymbol{\phi}_Y}$ and $\boldsymbol{\theta}_{\boldsymbol{\phi}_B}$ 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:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{y},\boldsymbol{z}) = \int P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}),\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\boldsymbol{\phi}_Y},\boldsymbol{\theta}_{\boldsymbol{\phi}_B}|\boldsymbol{y},\boldsymbol{z})d\boldsymbol{\phi}_Y(\boldsymbol{s}_y)d\boldsymbol{\phi}_Y(\boldsymbol{s}_z)d\boldsymbol{\phi}_B(\boldsymbol{s}_z)d\boldsymbol{\theta}_{\boldsymbol{\phi}_Y}d\boldsymbol{\theta}_{\boldsymbol{\phi}_B} \tag{B.15}$$

Where the integrand can be broken down into:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}),\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\boldsymbol{\phi}_Y},\boldsymbol{\theta}_{\boldsymbol{\phi}_B}|\boldsymbol{y},\boldsymbol{z}) = P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\boldsymbol{\phi}_Y},\boldsymbol{\theta}_{\boldsymbol{\phi}_B})\cdot P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\boldsymbol{\phi}_Y},\boldsymbol{\theta}_{\boldsymbol{\phi}_B}|\boldsymbol{y},\boldsymbol{z}) \tag{B.16}$$

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 $\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})$. In the case of Gaussian processes the distribution of $P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}})|\boldsymbol{\phi}_Y(\boldsymbol{s}_y),\boldsymbol{\phi}_Z(\boldsymbol{s}_z),\boldsymbol{\theta}_{\boldsymbol{\phi}_Y},\boldsymbol{\theta}_{\boldsymbol{\phi}_B})$ can be formulated in the following way, where to start take the joint distribution:

$$\begin{bmatrix}\phi_Y(\hat{\boldsymbol{s}})\\ \phi_Y(\boldsymbol{s}_y)\\ \phi_Z(\boldsymbol{s}_z)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}m_{\phi_Y}(\hat{\boldsymbol{s}})\\ m_{\phi_Y}(\boldsymbol{s}_y)\\ m_{\phi_Z}(\boldsymbol{s}_z)\end{bmatrix}, \begin{bmatrix}K_{\phi_Y}(\hat{\boldsymbol{s}},\hat{\boldsymbol{s}}) & K_{\phi_Y}(\hat{\boldsymbol{s}},\boldsymbol{s}_y) & K_{\phi_Y}(\hat{\boldsymbol{s}},\boldsymbol{s}_z)\\ K_{\phi_Y}(\boldsymbol{s}_y,\hat{\boldsymbol{s}}) & K_{\phi_Y}(\boldsymbol{s}_y,\boldsymbol{s}_y) & K_{\phi_Y}(\boldsymbol{s}_y,\boldsymbol{s}_z)\\ K_{\phi_Y}(\boldsymbol{s}_z,\hat{\boldsymbol{s}}) & K_{\phi_Y}(\boldsymbol{s}_z,\boldsymbol{s}_y) & K_{\phi_Z}(\boldsymbol{s}_z,\boldsymbol{s}_z)\end{bmatrix}\right) \tag{B.17}$$

Note, that since $\boldsymbol{\phi}_Y(s)$ and $\boldsymbol{\phi}_B(s)$ are independent and $\boldsymbol{\phi}_Z(s) = \boldsymbol{\phi}_Y(s) + \boldsymbol{\phi}_B(s)$, the covariance between the parameters $\boldsymbol{\phi}_Y(s)$ and $\boldsymbol{\phi}_Z(s)$ is simply $COV\Big(\phi_Y(s),\phi_Z(s')\Big) = COV\Big(\phi_Y(s),\phi_Y(s')\Big) = K_{\phi_Y}(s,s')$. Additionally, the mean and covariance terms for the process that generates $\boldsymbol{\phi}_Z(s)$ 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:

$$U_1 = \left[\phi_Y(\hat{\boldsymbol{s}})\right], U_2 = \begin{bmatrix}\phi_Y(\boldsymbol{s}_y)\\ \phi_Z(\boldsymbol{s}_z)\end{bmatrix}, \boldsymbol{U} = \begin{bmatrix}U_1\\ U_2\end{bmatrix}, M_1 = \left[m_{\phi_Y}(\hat{\boldsymbol{s}})\right], M_2 = \begin{bmatrix}m_{\phi_Y}(\boldsymbol{s}_y)\\ m_{\phi_Z}(\boldsymbol{s}_z)\end{bmatrix}, \boldsymbol{M} = \begin{bmatrix}M_1\\ M_2\end{bmatrix} \tag{B.18}$$

$$K_{11} = \left[K_{\phi_Y}(\hat{\boldsymbol{s}}, \hat{\boldsymbol{s}})\right], K_{12} = \begin{bmatrix}K_{\phi_Y}(\hat{\boldsymbol{s}}, \boldsymbol{s}_y) & K_{\phi_Y}(\hat{\boldsymbol{s}}, \boldsymbol{s}_z)\end{bmatrix} \tag{B.19}$$

$$K_{21} = \begin{bmatrix}K_{\phi_Y}(\boldsymbol{s}_y, \hat{\boldsymbol{s}})\\ K_{\phi_Y}(\boldsymbol{s}_z, \hat{\boldsymbol{s}})\end{bmatrix}, K_{22} = \begin{bmatrix}K_{\phi_Y}(\boldsymbol{s}_y, \boldsymbol{s}_y) & K_{\phi_Y}(\boldsymbol{s}_y, \boldsymbol{s}_z)\\ K_{\phi_Y}(\boldsymbol{s}_z, \boldsymbol{s}_y) & K_{\phi_Z}(\boldsymbol{s}_z, \boldsymbol{s}_z)\end{bmatrix} \tag{B.20}$$

The distribution can be written as:

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

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

$$P(U_1|U_2) = \mathcal{N}(M_{1|2}, K_{1|2}) \tag{B.22}$$

With parameters values:

$$M_{1|2} = M_1 + K_{12}K_{22}^{-1}(U_2 - M_2) \tag{B.23}$$

$$K_{1|2} = K_{11} - K_{12}K_{22}^{-1}K_{21} \tag{B.24}$$

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.

### B.1.2 Non-hierarchical Case

In the non-hierarchical case used in Sect. 3.4.1, 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:

$$P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B} | \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = \frac{P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z) | \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \cdot P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B})}{P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z))} \quad \text{(B.25)}$$

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

$$P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z) | \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{\phi}_Y(\boldsymbol{s}_y) | \boldsymbol{\phi}_Z(\boldsymbol{s}_z), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \cdot P(\boldsymbol{\phi}_Z(\boldsymbol{s}_z) | \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) \quad \text{(B.26)}$$

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

$$P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}) = P(\boldsymbol{\theta}_{\phi_Y}) \cdot P(\boldsymbol{\theta}_{\phi_B}) \quad \text{(B.27)}$$

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:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}) | \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = \int P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B} | \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) d\boldsymbol{\theta}_{\phi_Y} d\boldsymbol{\theta}_{\phi_B} \quad \text{(B.28)}$$

Where the integrand can be broken down into:

$$P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}), \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B} | \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) = P(\boldsymbol{\phi}_Y(\hat{\boldsymbol{s}}) | \boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B}, \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) \cdot P(\boldsymbol{\theta}_{\phi_Y}, \boldsymbol{\theta}_{\phi_B} | \boldsymbol{\phi}_Y(\boldsymbol{s}_y), \boldsymbol{\phi}_Z(\boldsymbol{s}_z)) \quad \text{(B.29)}$$

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 the expression can be formulated in the same way as presented in Sect. B.1.1.

## B.2 Data Generation

### B.2.1 3.4.1

Define $\phi_Y$ as one parameter of the probability density function for the in situ observations and $\phi_Z$ as the corresponding parameter for the climate model output. The following relationship is then assumed $\phi_Z = \phi_Y + \phi_B$, where $\phi_B$ is the bias in the parameter and is assumed independent of $\phi_Y$. The latent distributions that generate $\phi_Y$ and $\phi_B$ across the domain are assumed GPs with mean and covariance functions. The mean function is assumed constant for simplicity and the covariance function is taken as an RBF kernel with a kernel variance and length scale parameter (Eq. (3.5)). These hyper-parameters of the two latent generating processes are set for three scenarios, as given in Table 3.1.

For each scenario, a sample of the parameters $\phi_Y$ and $\phi_B$ is taken from the distributions $\mathcal{GP}_{\phi_Y}$ and $\mathcal{GP}_{\phi_B}$ at regularly spaced, high-resolution intervals. These samples are referred to here as complete realisations and represent underlying fields for each parameter across the domain, which the model aims to estimate. Direct 'observations' of the parameter $\phi_Y$ from the underlying field are simulated at lower-resolution, randomised locations after conditioning the distribution $\mathcal{GP}_{\phi_Y}$ on the complete realisation and introducing some noise. In order to simulate direct measurements of the parameter $\phi_Z$ of the climate model output, samples are first generated for $\phi_Y$ and $\phi_B$ at regularly spaced intervals after conditioning the distributions $\mathcal{GP}_{\phi_Y}$ and $\mathcal{GP}_{\phi_B}$ on the complete realisations, then the sum of these samples at each location is taken to give $\phi_Z$. The number of direct observations/measurements for each parameter under the different scenarios is given in Table 3.1.

### B.2.2 3.4.2

A sample of the parameters $\mu_Y$, $\mu_B$, $\tilde{\sigma}_Y$ and $\tilde{\sigma}_B$ 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 parameter across the domain. A sample of the parameters for the in situ data ($\mu_Y$ and $\tilde{\sigma}_Y$) are also generated at a selection of lower-resolution, randomised locations after conditioning the latent distributions on the complete realisations. Observations of the in situ data $Y$ are then generated at these locations from the corresponding normal distribution. In the case of the climate model data $Z$, samples are first generated for $\mu_Y$, $\mu_B$, $\tilde{\sigma}_Y$ and $\tilde{\sigma}_B$ at regularly spaced intervals after conditioning the latent distributions on the complete realisations, then the sum of these samples at each location is taken to give $\mu_Z$ and $\tilde{\sigma}_Z$. The climate model data is then generated at these locations from the corresponding normal distribution. The number of locations and the number of samples per location are given in Table 3.4.

## B.3 Prior and Posterior Distribution Examples

### B.3.1 3.4.1

In Figure B.1 the prior and posterior distributions are illustrated for each parameter given the measurement data in scenario one. As expected, it can be seen that the density of the posterior distribution for each parameter is concentrated closer to the value specified when generating the data than in the case of the prior distributions. The extent of the variation between the prior and posterior distributions depends on the specific parameter and the impact that parameter has on the likelihood of the measured data. As an example, the posterior distribution for the length scale of the latent GP that generates $\phi_Y$ across the domain (Fig. B.1b) is more concentrated around the specified value compared to the equivalent length scale for the bias (Fig. B.1e).

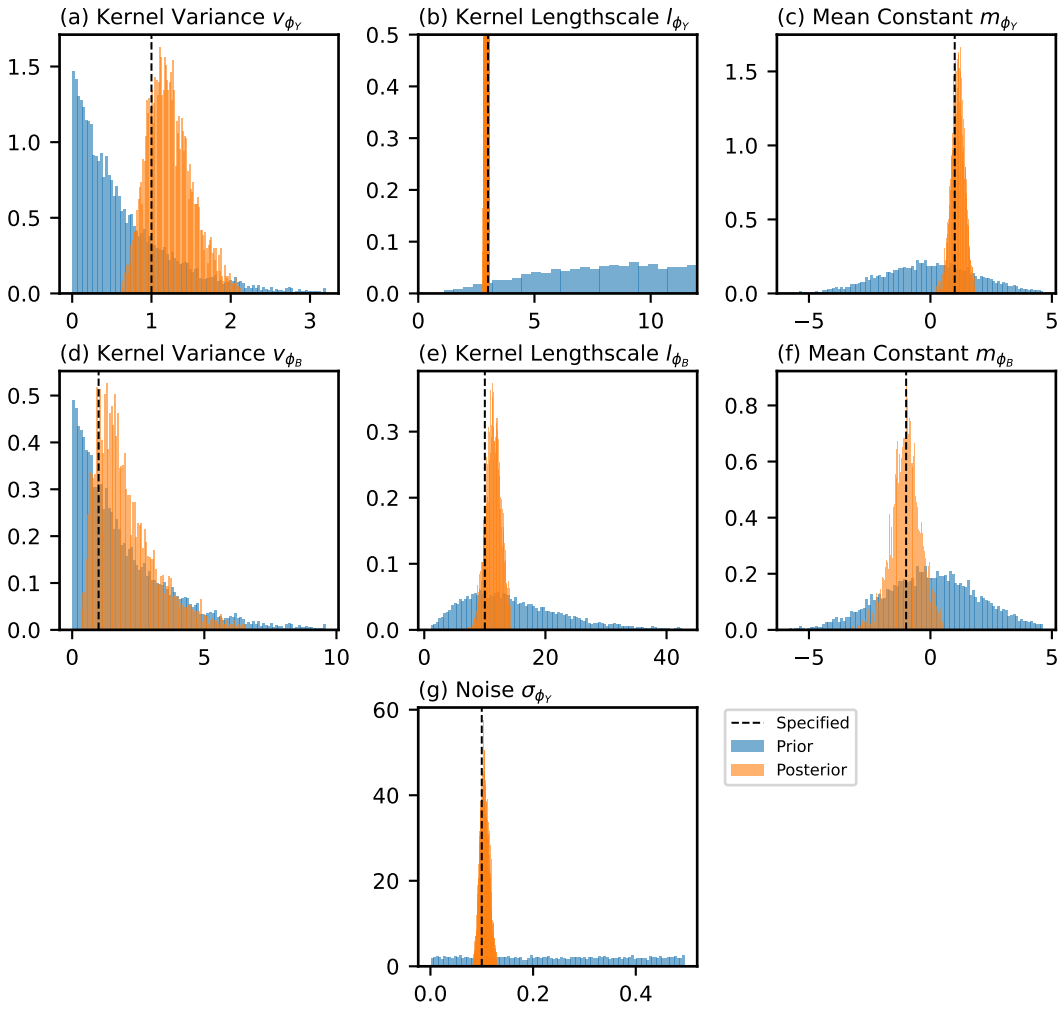



Figure B.1: A figure illustrating the prior and posterior distributions for the parameters of the model in the case of scenario one. The value that was specified when generating the data is also shown.

### B.3.2 3.4.2

The prior and posterior distributions for all parameters of the hierarchical model is presented in Figure B.2. Inference on the parameters was performed using MCMC in a hierarchical Bayesian framework. Relatively non-informative prior distributions are chosen and are equivalent to the choice of priors in Sect. 3.4.1. The model assumes 4 generating GP distributions, one for each of the parameters $\mu_Y$, $\mu_B$, $\tilde{\sigma}_Y$

and $\tilde{\sigma}_B$. Each GP distribution has 3 associated parameters, being the mean constant ($m$), the kernel variance ($v$) and the kernel length scale ($l$). The values specified in generating the data are shown as a dotted line for each parameter.

As expected, the posterior distribution for each parameter is concentrated closer to the value specified when generating the data than the relatively non-informative prior distributions. As with the non-hierarchical case, more confidence is shown on the value of some parameters than others in the posterior distributions, such as the posterior of the length scale of the latent GP that generates $\mu_Y$ across the domain (Fig. B.2b) compared to the equivalent length scale for that of the bias (Fig. B.2h).

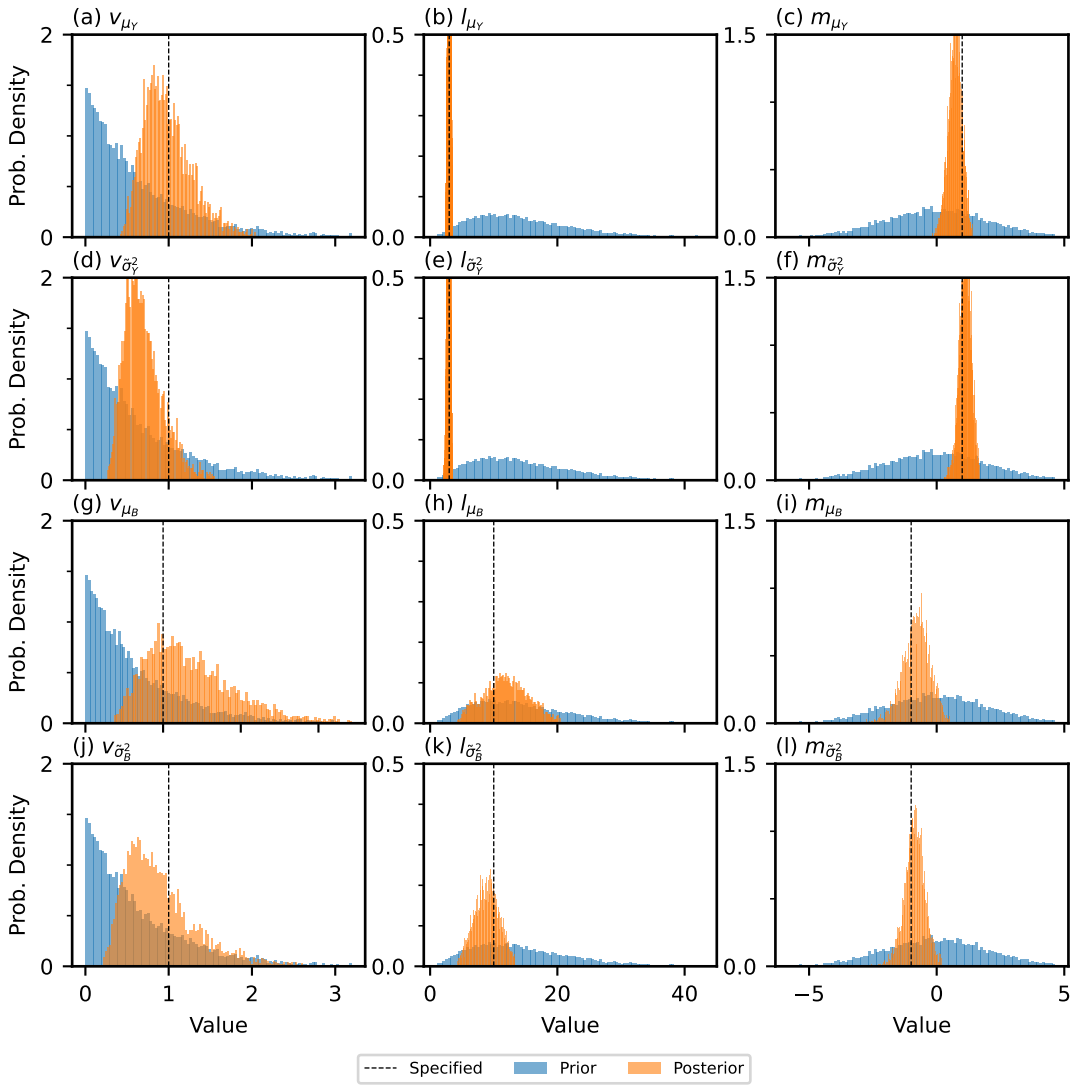



Figure B.2: A figure illustrating the prior and posterior distributions for the parameters of the model in the case of the 1D hierarchical example. The value that was specified when generating the data is also shown.

## B.4 Complex Scenarios

Real-world scenarios are expected to have more complex spatial features than the simulated examples presented in Sect. 3.4.1, 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. B.3. The hyper-parameter values used to generate the data are presented in Table B.1, while the summary statistics for the posterior distributions after fitting the model proposed in this paper are presented in Table B.2.

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 B.2 and Fig. B.3. From Table B.2 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. B.3 it can be seen that, despite the additional complexities, the predictions on the unbiased parameter and on the bias are reasonable and capture the main spatial patterns. This demonstrates the flexibility of GPs and the robustness of the methodology proposed to fit different types of real-world data where some of the assumptions made in the model partially don't hold. Some features of the results due to not fully capturing the dependencies involved in generating the data are described here. In scenario A the length scale of the unbiased process is estimated close to the value used in generating the data for $x > 50$, which results in greater uncertainty than expected between nearby observations in the region $x < 50$ where the length scale is greater. For example, in the extrapolation range of $x < 0$ the prediction in the unbiased parameter values returns sharply to the mean and with uncertainty independent of observed points, whereas if the length scale was correctly estimated in this region the predictions would remain dependent on the data observed at $x = 0 - 10$ for longer. The same is true in scenario B with the addition of the estimates of the bias being effected, making disaggregating the climate model output into an unbiased and biased component more challenging, as seen at $x = 30$. In scenario C again by only modelling a single length scale for the unbiased process, disaggregating the climate model output into its two components is effected and the longer length scale peak present at $x = 20$ is attributed to the bias

incorrectly. Finally in scenario D, not accounting for the correlation between the unbiased values and the bias results in a slightly greater uncertainty in predictions than could be achieved by incorporating this relationship.

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

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

Table B.1: A table showing the hyper-parameters of the latent Gaussian processes used to generate the complete underlying realisations of $\phi_Y(\boldsymbol{s}^\star)$, $\phi_B(\boldsymbol{s}^\star)$ and hence $\phi_Z(\boldsymbol{s}^\star)$, as well as observations of $\phi_Y(\boldsymbol{s}_y)$ and $\phi_Z(\boldsymbol{s}_z)$, on which inference is done for the additional scenarios. The number of observations representing in-situ data and climate model output is also given.

(a) Scenario A and B

| Dependent Variable | Model Parameter | Scenario A Posterior Dist. Exp. | Std. Dev. | C.I. L. | C.I. U. | Scenario B Posterior Dist. Exp. | Std. Dev. | C.I. L. | C.I. U. |
|---|---|---|---|---|---|---|---|---|---|
| Unbiased PDF Parameter $\phi_Y$ | Kernel Variance $v_{\phi_Y}$ | 0.87 | 0.17 | 0.55 | 1.18 | 0.85 | 0.18 | 0.53 | 1.20 |
| | Kernel Lengthscale $l_{\phi_Y}$ | 1.06 | 0.06 | 0.94 | 1.19 | 1.12 | 0.07 | 0.99 | 1.24 |
| | Mean Constant $m_{\phi_Y}$ | 0.78 | 0.15 | 0.48 | 1.07 | 1.42 | 0.17 | 1.09 | 1.77 |
| | Noise $\sigma_{\phi_Y}$ | 0.12 | 0.03 | 0.06 | 0.17 | 0.15 | 0.05 | 0.05 | 0.24 |
| Bias PDF Parameter $\phi_B$ | Kernel Variance $v_{\phi_B}$ | 1.15 | 0.86 | 0.19 | 2.87 | 0.85 | 0.39 | 0.32 | 1.60 |
| | Kernel Lengthscale $l_{\phi_B}$ | 10.34 | 1.93 | 6.86 | 14.32 | 3.50 | 0.82 | 2.21 | 5.05 |
| | Mean Constant $m_{\phi_B}$ | -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. Exp. | Std. Dev. | C.I. L. | C.I. U. | Scenario D Posterior Dist. Exp. | Std. Dev. | C.I. L. | C.I. U. |
|---|---|---|---|---|---|---|---|---|---|
| Unbiased PDF Parameter $\phi_Y$ | Kernel Variance $v_{\phi_Y}$ | 0.52 | 0.10 | 0.34 | 0.72 | 0.88 | 0.23 | 0.49 | 1.33 |
| | Kernel Lengthscale $l_{\phi_Y}$ | 0.75 | 0.08 | 0.61 | 0.92 | 2.95 | 0.06 | 2.82 | 3.07 |
| | Mean Constant $m_{\phi_Y}$ | 1.03 | 0.11 | 0.81 | 1.25 | 0.93 | 0.26 | 0.42 | 1.44 |
| | Noise $\sigma_{\phi_Y}$ | 0.18 | 0.06 | 0.09 | 0.30 | 0.10 | 0.02 | 0.07 | 0.13 |
| Bias PDF Parameter $\phi_B$ | Kernel Variance $v_{\phi_B}$ | 0.57 | 0.44 | 0.11 | 1.38 | 0.42 | 0.22 | 0.14 | 0.84 |
| | Kernel Lengthscale $l_{\phi_B}$ | 6.71 | 2.11 | 3.19 | 10.84 | 4.31 | 0.53 | 3.26 | 5.33 |
| | Mean Constant $m_{\phi_B}$ | -0.12 | 0.30 | -0.73 | 0.44 | -0.02 | 0.21 | -0.46 | 0.38 |

Table B.2: 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 3.2.

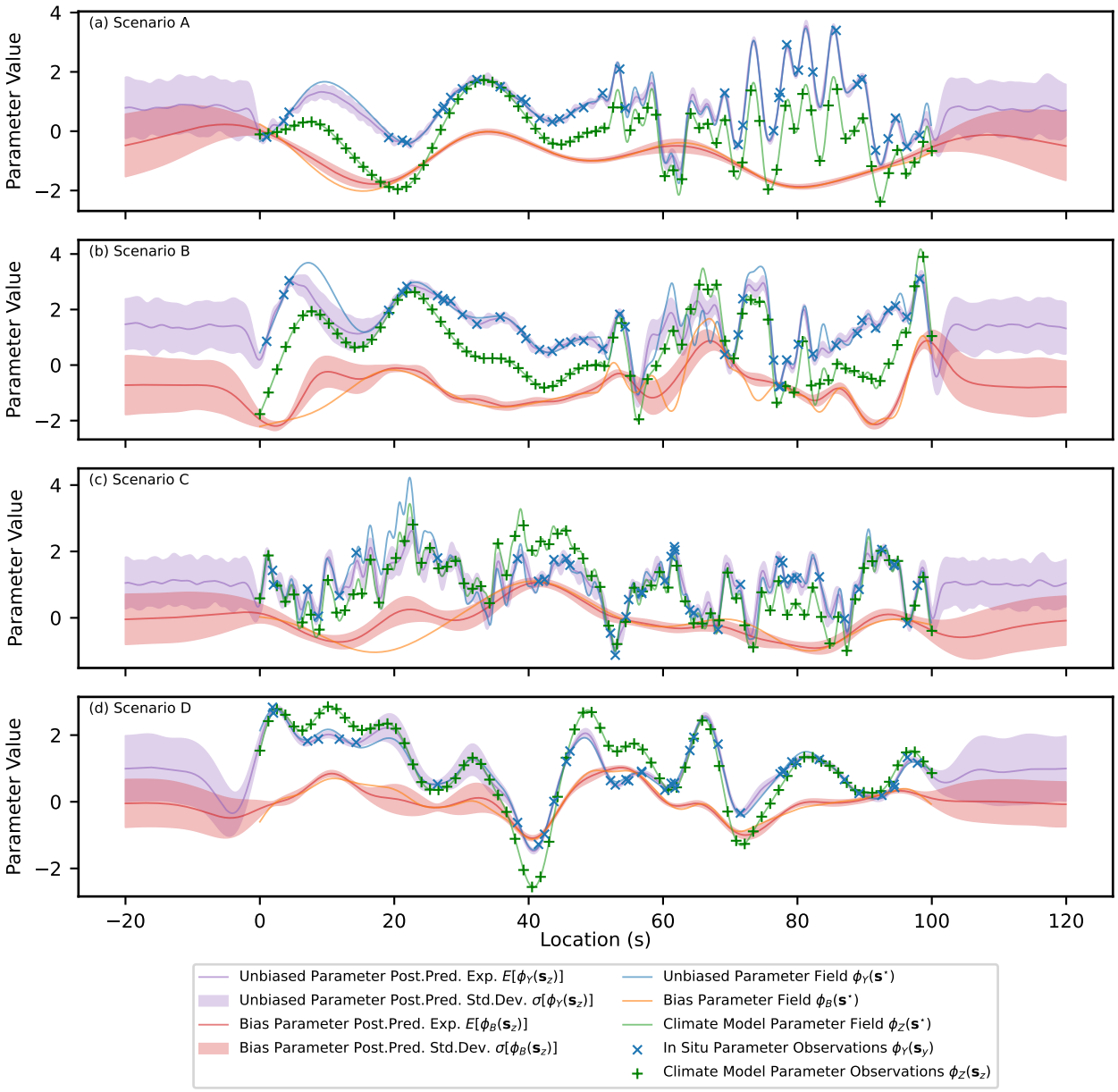



Figure B.3: Expectation and $1\sigma$ uncertainty of the posterior predictive distributions of the parameter $\phi_Y(\boldsymbol{s}_z)$ and the corresponding bias $\phi_B(\boldsymbol{s}_z)$ for three scenarios. The underlying functions (complete realisations) as well as the simulated input data are also shown.

# Appendix C

# Applying Bias Correction to Surface Climatology over Antarctica

## C.1 Mean function model parameter prior and posterior distributions

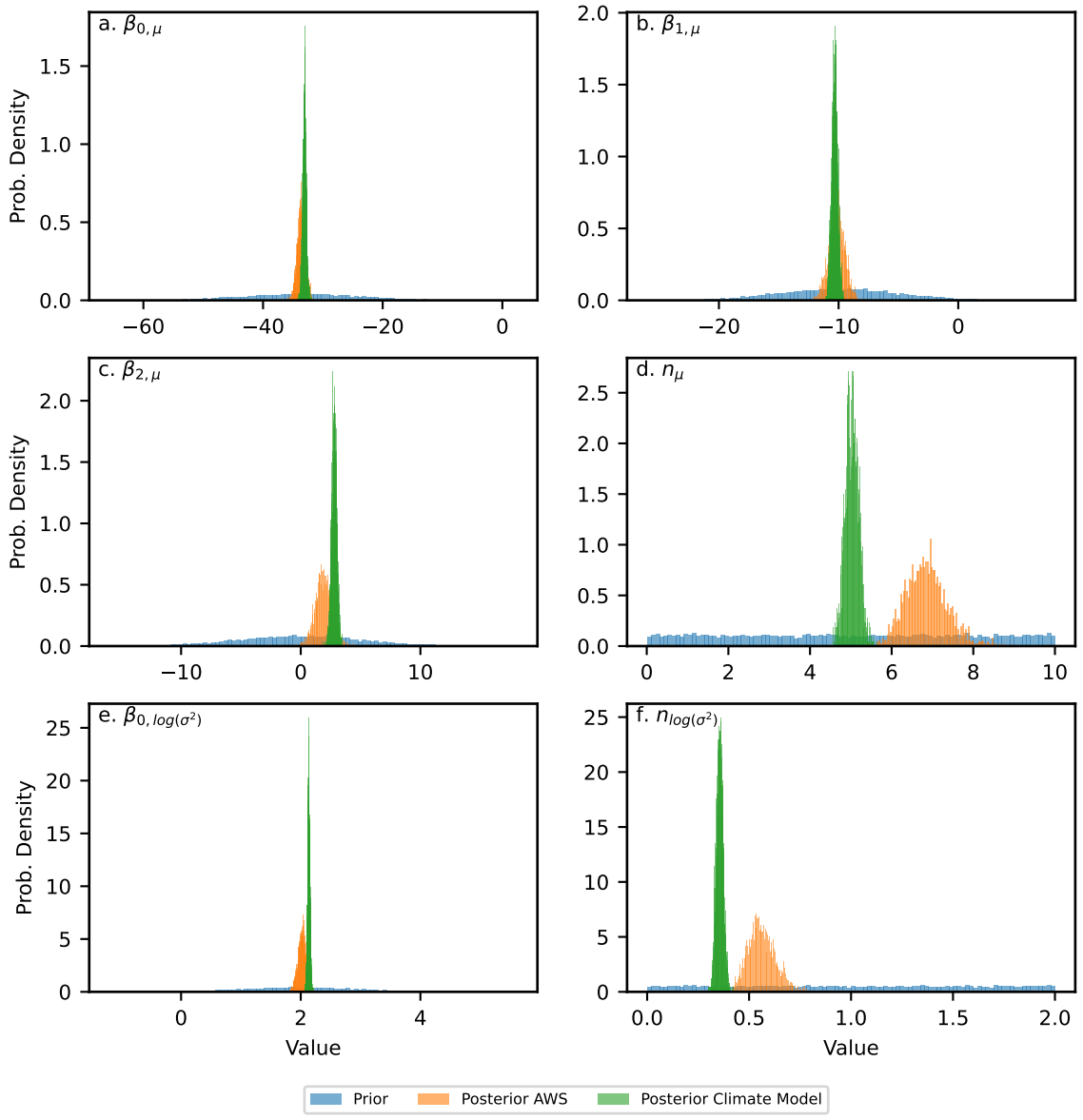



Figure C.1: Prior and posterior distributions for the parameters of the global mean function model fit to both the AWS and climate model output separately.

## C.2 Residuals spatial model parameter prior and posterior distributions

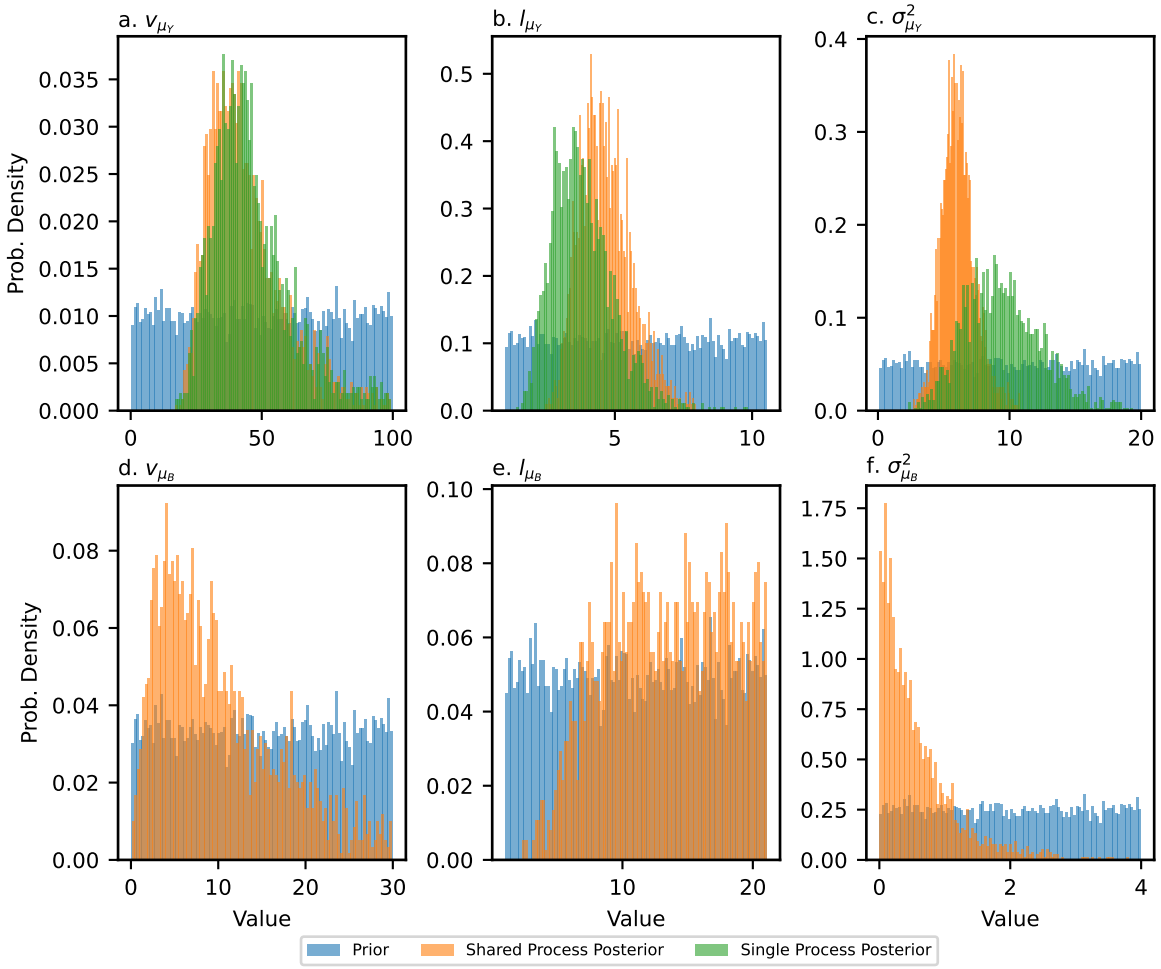



Figure C.2: Prior and posterior distributions for the parameters of the spatial model fit to the residual components of the mean parameter.

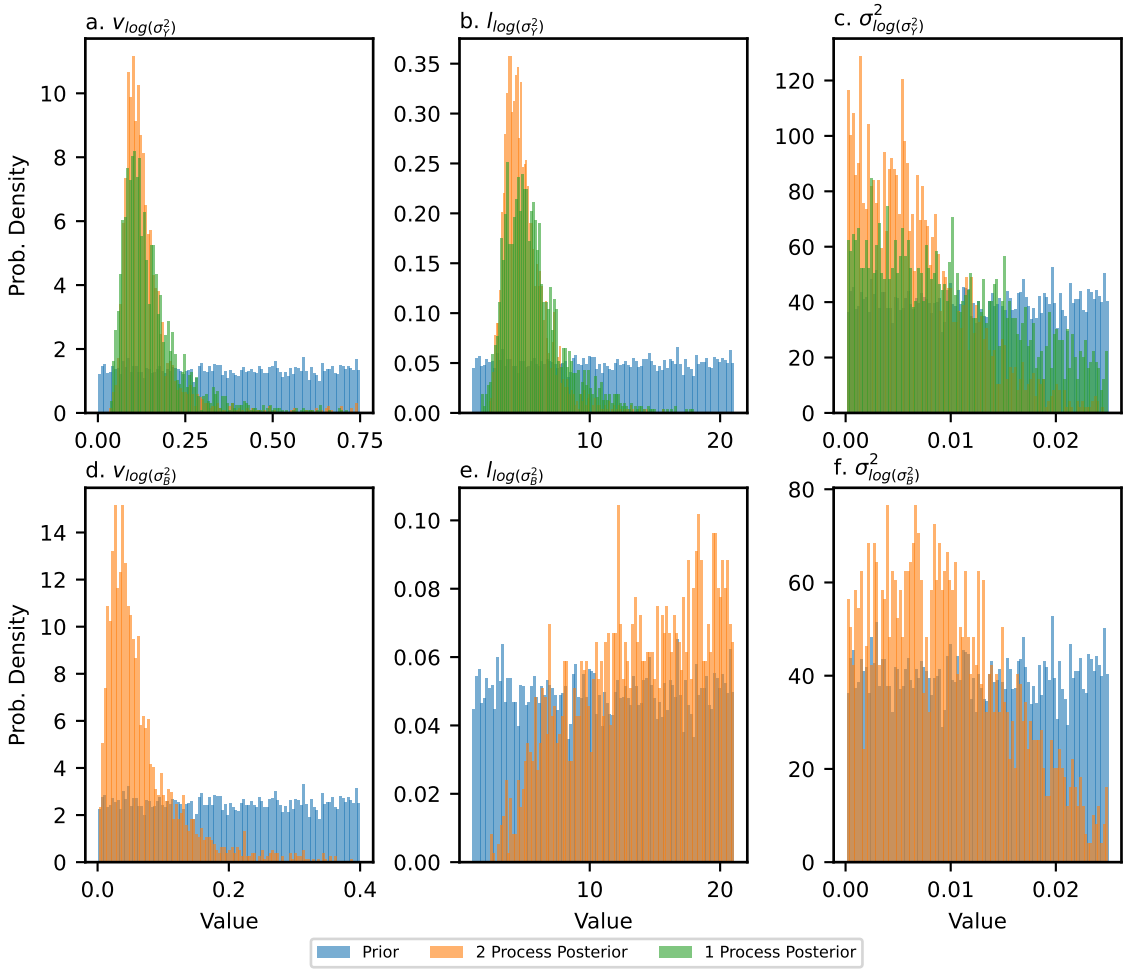



Figure C.3: Prior and posterior distributions for the parameters of the spatial model fit to the residual components of the log-variance parameter.

## C.3 Bias Components

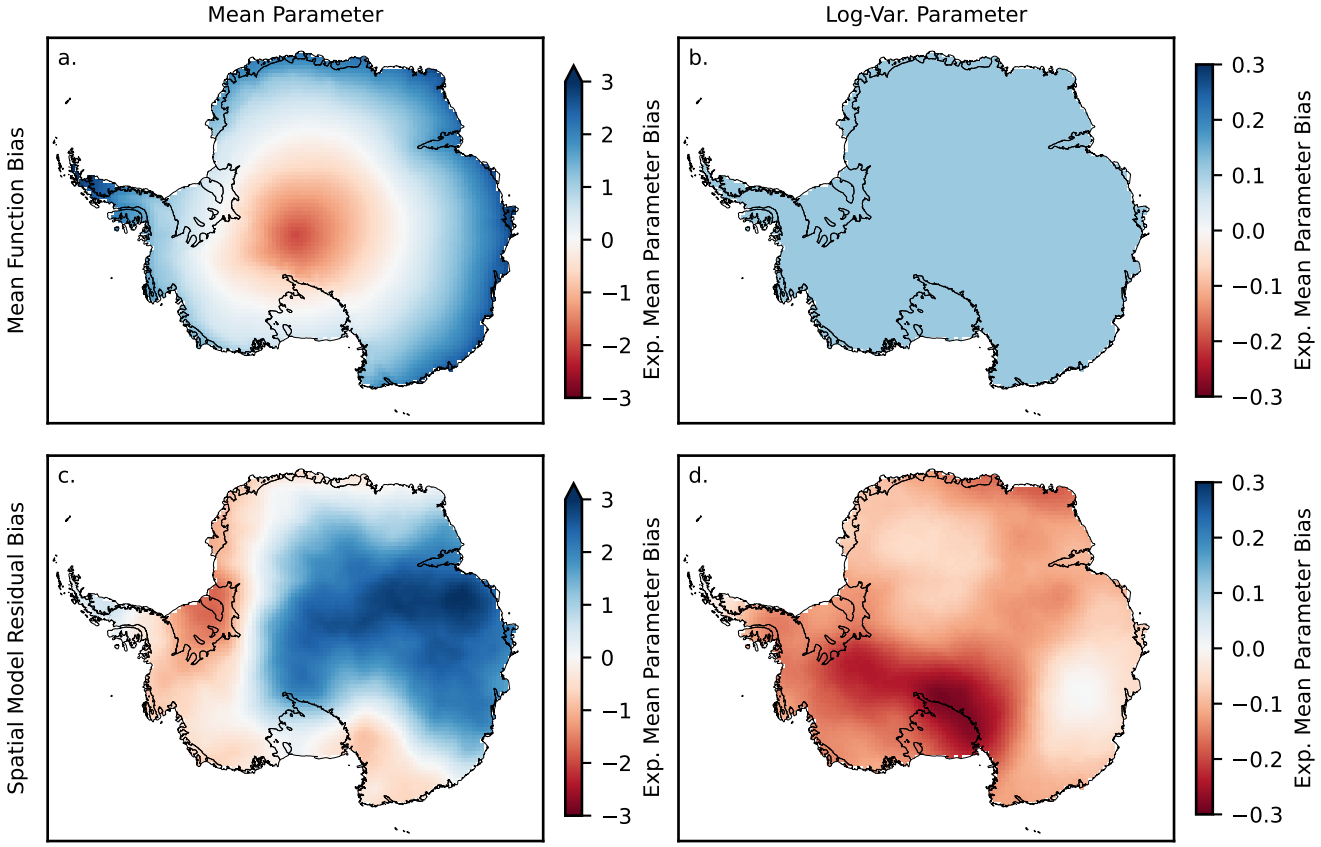



Figure C.4: Bias in the mean and log-variance parameters split into components for the mean function (a,b) and residual (c,d).

## C.4 Spatial Model Grid

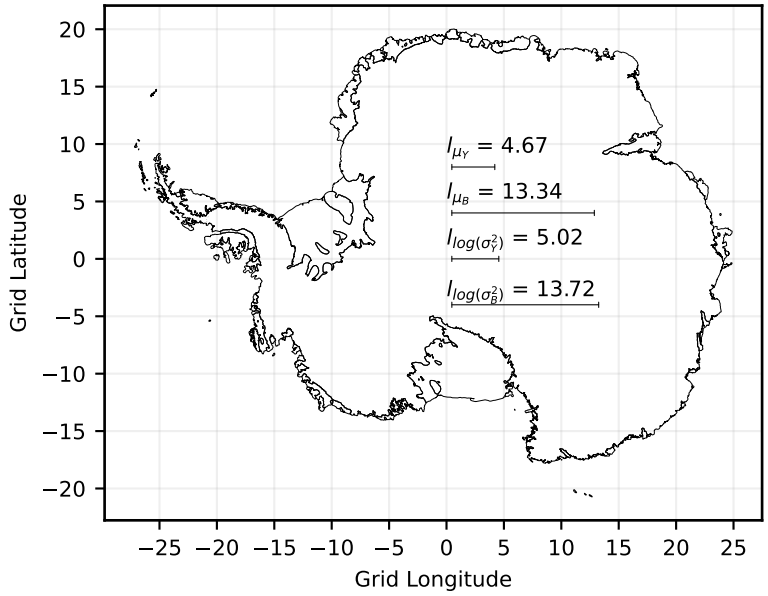



Figure C.5: A figure providing reference for the values obtained for the lengthscales of the latent GPs in stage two of the model against the extent of the ice sheet.

# References


Agosta, Cécile, Charles Amory, Christoph Kittel, Anais Orsi, Vincent Favier, Hubert Gallée, Michiel R. van den Broeke, Jan T. M. Lenaerts, Jan Melchior van Wessem, Willem Jan van de Berg, and Xavier Fettweis (2019). "Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes". In: *The Cryosphere*. DOI: https://doi.org/10.5194/tc-13-281-2019.

Bader, David, Curt Covey, William Gutowski, Isaac Held, Kenneth Kunkel, Ronald Miller, Robin Tokmakian, and Minghua Zhang (2008). "Climate Models: An Assessment of Strengths and Limitations". In: *Climate Models: An Assessment of Strengths and Limitations*.

Balsamo, Gianpaolo, Anton Beljaars, Klaus Scipal, Pedro Viterbo, Bart van den Hurk, Martin Hirschi, and Alan K. Betts (2009). "A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System". In: *Journal of Hydrometeorology*. DOI: 10.1175/2008JHM1068.1.

Bamber, J. L., J. L. Gomez-Dans, and J. A. Griggs (2009). "A new 1 km digital elevation model of the Antarctic derived from combined satellite radar and laser data – Part 1: Data and methods". In: *The Cryosphere*. DOI: 10.5194/tc-3-101-2009.

Bamber, Jonathan L., Michael Oppenheimer, Robert E. Kopp, Willy P. Aspinall, and Roger M. Cooke (2019). "Ice sheet contributions to future sea-level rise

from structured expert judgment". In: *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.1817205116.

Banwell, Alison F., Douglas R. MacAyeal, and Olga V. Sergienko (2013). "Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes". In: *Geophysical Research Letters*. DOI: 10.1002/2013GL057694.

Bell, Robin E., Alison F. Banwell, Luke D. Trusel, and Jonathan Kingslake (2018). "Antarctic surface hydrology and impacts on ice-sheet mass balance". In: *Nature Climate Change*. DOI: 10.1038/s41558-018-0326-3.

Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding (2011). "The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes". In: *Geoscientific Model Development*. DOI: 10.5194/gmd-4-677-2011.

Beyer, Robert, Mario Krapp, and Andrea Manica (2020). "An empirical evaluation of bias correction methods for palaeoclimate simulations". In: *Climate of the Past*. DOI: 10.5194/cp-16-1493-2020.

Broeke, Michiel van den (2005). "Strong surface melting preceded collapse of Antarctic Peninsula ice shelf". In: *Geophysical Research Letters*. DOI: 10.1029/2005GL023247.

Bromwich, David H. (1989). "Satellite Analyses of Antarctic Katabatic Wind Behavior". In: *Bulletin of the American Meteorological Society*. DOI: 10.1175/1520-0477(1989)070<0738:SAOAKW>2.0.CO;2.

Brun, E., P. David, M. Sudul, and G. Brunot (1992). "A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting". In: *Journal of Glaciology*. DOI: 10.1017/s0022143000009552.

Bruyère, Cindy L., James M. Done, Greg J. Holland, and Sherrie Fredrick (2014). "Bias corrections of global models for regional climate simulations of high-impact weather". In: *Climate Dynamics*. DOI: 10.1007/s00382-013-2011-6.

Bulthuis, Kevin, Maarten Arnst, Sainan Sun, and Frank Pattyn (2019). "Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change". In: *Cryosphere*. DOI: 10.5194/tc-13-1349-2019.

Bush, Mike, Tom Allen, Caroline Bain, Ian Boutle, John Edwards, Anke Finnenkoetter, Charmaine Franklin, Kirsty Hanley, Humphrey Lean, Adrian Lock, James Manners, Marion Mittermaier, Cyril Morcrette, Rachel North, Jon Petch, Chris Short, Simon Vosper, David Walters, Stuart Webster, Mark Weeks, Jonathan Wilkinson, Nigel Wood, and Mohamed Zerroukat (2020). "The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1". In: *Geoscientific Model Development*. DOI: 10.5194/gmd-13-1999-2020.

Buzzard, Sammie, Daniel Feltham, and Daniela Flocco (2018). "Modelling the fate of surface melt on the Larsen C Ice Shelf". In: *Cryosphere*. DOI: 10.5194/tc-12-3565-2018.

Cape, M. R., Maria Vernet, Pedro Skvarca, Sebastian Marinsek, Ted Scambos, and Eugene Domack (2015). "Foehn winds link climate-driven warming to ice shelf evolution in Antarctica". In: *Journal of Geophysical Research-Atmospheres*. DOI: 10.1002/2015JD023465.

Carter, Jeremy, Erick Chacón-Montalván, and Amber Leeson (2024). "Bias Correction of Climate Models using a Bayesian Hierarchical Model". In: *EGUsphere*. DOI: 10.5194/egusphere-2023-2536.

Carter, Jeremy, Amber Leeson, Andrew Orr, Christoph Kittel, and J. Melchior van Wessem (2022a). "Variability in Antarctic surface climatology across regional climate models and reanalysis datasets". In: *The Cryosphere*. DOI: 10.5194/tc-16-3815-2022.

Carter, Jeremy, Amber Leeson, Andrew Orr, Christoph Kittel, and Melchior van Wessem (2022b). *Variability in Antarctic Surface Climatology Across Regional Climate Models and Reanalysis Datasets*. DOI: 10.5281/zenodo.6367850.

Cattiaux, Julien, Hervé Douville, and Yannick Peings (2013). "European temperatures in CMIP5: origins of present-day biases and future uncertainties". In: *Climate Dynamics*. DOI: 10.1007/s00382-013-1731-y.

Christensen, Jens H., Fredrik Boberg, Ole B. Christensen, and Philippe Lucas-Picher (2008). "On the need for bias correction of regional climate change projections of temperature and precipitation". In: *Geophysical Research Letters*. DOI: 10.1029/2008GL035694.

Cleveland, Robert B., William S. Cleveland, and Irma Terpenning (1990). "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". In: *Journal of Official Statistics*. URL: https://www.proquest.com/docview/1266805989/abstract/6B2D09F43DE44231PQ/1 (visited on 06/03/2024).

Das, Apurba, Prabin Rokaya, and Karl-Erich Lindenschmidt (2022). "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". In: *Climatic Change*. DOI: 10.1007/s10584-022-03364-5.

Datta, Rajashree Tri, Marco Tedesco, Xavier Fettweis, Cecile Agosta, Stef Lhermitte, Jan T. M. Lenaerts, and Nander Wever (2019). "The Effect of Foehn-Induced Surface Melt on Firn Evolution Over the Northeast Antarctic Peninsula". In: *Geophysical Research Letters*. DOI: 10.1029/2018GL080845.

DeConto, Robert M. and David Pollard (2016). "Contribution of Antarctica to past and future sea-level rise". In: *Nature*. DOI: 10.1038/nature17145.

DeConto, Robert, David Pollard, Richard Alley, Isabella Velicogna, Edward Gasson, Natalya Gomez, Shaina Sadai, Alan Condron, Daniel Gilford, Erica Ashe, Robert Kopp, Dawei Li, and A. Dutton (2021). "The Paris Climate Agreement and future sea-level rise from Antarctica". In: *Nature*. DOI: 10.1038/s41586-021-03427-0.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M.

Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart (2011). "The ERA-Interim reanalysis: configuration and performance of the data assimilation system". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.828.

Depoorter, Mathieu A., Jonathan L. Bamber, Jennifer Griggs, Jan T. M. Lenaerts, Stefan R. M. Ligtenberg, Michiel R. van den Broeke, and Geir Moholdt (2013). *Antarctic masks (ice-shelves, ice-sheet, and islands), link to shape file*. DOI: 10.1594/PANGAEA.819147.

Doblas-Reyes, Francisco J., Anna A. Sörensson, M. Almazroui, Alessandro Dosio, William J. Gutowski, Rein Haarsma, Rafiq Hamdi, Bruce Hewitson, Won-Tae Kwon, Benjamin L. Lamptey, Douglas Maraun, Tannecia S. Stephenson, Izuru Takayabu, Laurent Terray, Andrew Turner, and Zhiyan Zuo (2021). "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*. Ed. by Valérie Masson-Delmotte, Panmao Zhai, Anna Pirani, Sarah L. Connors, Clotilde Péan, Sophie Berger, Nada Caud, Yang Chen, Leah Goldfarb, Melissa I. Gomis, Mengtian Huang, Katherine Leitzell, Elisabeth Lonnoy, J. B. Robin Matthews, Thomas K. Maycock, Tim Waterfield, Ozge Yelekçi, Rong Yu, and Botao Zhou. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. DOI: 10.1017/9781009157896.001.

ECMWF (2009). "IFS Documentation CY33R1 Part IV: Physical Processes". In: *IFS Documentation CY33R1*. ECMWF. DOI: 10.21957/8o7vwlbdr.

ECMWF (2016). "IFS Documentation CY41R2 - Part IV: Physical Processes". In: *IFS Documentation CY41R2*. ECMWF. DOI: 10.21957/tr5rv27xu.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert (2012). "HESS Opinions "Should we apply bias correction to global and regional climate model

data?"". In: *Hydrology and Earth System Sciences*. DOI: 10.5194/hess-16-3391-2012.

Elvidge, Andrew D., Peter Kuipers Munneke, John C. King, Ian A. Renfrew, and Ella Gilbert (2020). "Atmospheric drivers of melt on Larsen C Ice Shelf: Surface energy budget regimes and the impact of foehn". In: *Journal of Geophysical Research: Atmospheres*. DOI: 10.1029/2020JD032463.

Elvidge, Andrew D., Irina Sandu, Nils Wedi, Simon B. Vosper, Ayrton Zadra, Souhail Boussetta, François Bouyssel, Annelize van Niekerk, Mikhail A. Tolstykh, and Masashi Ujiie (2019). "Uncertainty in the Representation of Orography in Weather and Climate Models and Implications for Parameterized Drag". In: *Journal of Advances in Modeling Earth Systems*. DOI: 10.1029/2019MS001661.

Ettema, J., M. R. van den Broeke, E. van Meijgaard, W. J. van de Berg, J. E. Box, and K. Steffen (2010). "Climate of the Greenland ice sheet using a high-resolution climate model – Part 1: Evaluation". In: *The Cryosphere*. DOI: 10.5194/tc-4-511-2010.

Fettweis, X., B. Franco, M. Tedesco, J. H. van Angelen, J. T. M. Lenaerts, M. R. van den Broeke, and H. Gallée (2013). "Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR". In: *The Cryosphere*. DOI: 10.5194/tc-7-469-2013.

Fettweis, Xavier, Jason E. Box, Cécile Agosta, Charles Amory, Christoph Kittel, Charlotte Lang, Dirk van As, Horst Machguth, and Hubert Gallée (2017). "Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model". In: *The Cryosphere*. DOI: 10.5194/tc-11-1015-2017.

Flato, Gregory, Jochem Marotzke, Babatunde Abiodun, Pascale Braconnot, Sin Chan Chou, William Collins, Peter Cox, Fatima Driouech, Seita Emori, Veronika Eyring, Chris Forest, Peter Gleckler, Eric Guilyardi, Christian Jakob, Vladimir Kattsov, Chris Reason, and Markku Rummukainen (2013). "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*. Ed. by Intergovernmental Panel on Climate Change (IPCC). Cambridge: Cambridge University Press. ISBN: 978-1-107-05799-9. DOI: 10.1017/CBO9781107415324.020.

Franco, B., X. Fettweis, C. Lang, and M. Erpicum (2012). "Impact of spatial resolution on the modelling of the Greenland ice sheet surface mass balance between 1990–2010, using the regional climate model MAR". In: *The Cryosphere*. DOI: 10.5194/tc-6-695-2012.

Fretwell, P., H. D. Pritchard, D. G. Vaughan, J. L. Bamber, N. E. Barrand, R. Bell, C. Bianchi, R. G. Bingham, D. D. Blankenship, G. Casassa, G. Catania, D. Callens, H. Conway, A. J. Cook, H. F. J. Corr, D. Damaske, V. Damm, F. Ferraccioli, R. Forsberg, S. Fujita, Y. Gim, P. Gogineni, J. A. Griggs, R. C. A. Hindmarsh, P. Holmlund, J. W. Holt, R. W. Jacobel, A. Jenkins, W. Jokat, T. Jordan, E. C. King, J. Kohler, W. Krabill, M. Riger-Kusk, K. A. Langley, G. Leitchenkov, C. Leuschen, B. P. Luyendyk, K. Matsuoka, J. Mouginot, F. O. Nitsche, Y. Nogi, O. A. Nost, S. V. Popov, E. Rignot, D. M. Rippin, A. Rivera, J. Roberts, N. Ross, M. J. Siegert, A. M. Smith, D. Steinhage, M. Studinger, B. Sun, B. K. Tinto, B. C. Welch, D. Wilson, D. A. Young, C. Xiangbin, and A. Zirizzotti (2013). "Bedmap2: improved ice bed, surface and thickness datasets for Antarctica". In: *The Cryosphere*. DOI: 10.5194/tc-7-375-2013.

Gallée, Hubert (1995). "Simulation of the Mesocyclonic Activity in the Ross Sea, Antarctica". In: *Monthly Weather Review*. DOI: 10.1175/1520-0493(1995)123<2051:SOTMAI>2.0.CO;2.

Gallée, Hubert and Irina V. Gorodetskaya (2008). "Validation of a limited area model over Dome C, Antarctic Plateau, during winter". In: *Climate Dynamics*. DOI: 10.1007/s00382-008-0499-y.

Gallée, Hubert and Guy Schayes (1994). "Development of a Three-Dimensional Meso-$\gamma$ Primitive Equation Model: Katabatic Winds Simulation in the Area of

Terra Nova Bay, Antarctica". In: *Monthly Weather Review*. DOI: 10.1175/1520-0493(1994)122<0671:DOATDM>2.0.CO;2.

Genthon, Christophe, Dana Veron, Etienne Vignon, Delphine Six, Jean-Louis Dufresne, Jean-Baptiste Madeleine, Emmanuelle Sultan, and François Forget (2021). "10 years of temperature and wind observation on a 45 m tower at Dome C, East Antarctic plateau". In: *Earth System Science Data*. DOI: 10.5194/essd-13-5731-2021.

Gilbert, E. and C. Kittel (2021). "Surface Melt and Runoff on Antarctic Ice Shelves at 1.5°C, 2°C, and 4°C of Future Warming". In: *Geophysical Research Letters*. DOI: 10.1029/2020GL091733.

Gilbert, E., A. Orr, J. C. King, I. A. Renfrew, T. Lachlan-Cope, P. F. Field, and I. A. Boutle (2020). "Summertime cloud phase strongly influences surface melting on the Larsen C ice shelf, Antarctica". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: https://doi.org/10.1002/qj.3753.

Gilbert, Ella M. K., Andrew Orr, John C. King, Ian Renfrew, and Tom A. Lachlan-Cope (2021). *A 20-year study of melt processes over Larsen C Ice Shelf using a high-resolution regional atmospheric model: Part 1, Model configuration and validation*. preprint. DOI: 10.1002/essoar.10506250.1.

Giorgi, Filippo (2019). "Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?" In: *Journal of Geophysical Research: Atmospheres*. DOI: 10.1029/2018JD030094.

Greeves, C., Vicky Pope, R. Stratton, and G. Martin (2007). "Representation of Northern Hemisphere winter storm tracks in climate models". In: *Clim. Dynam.* DOI: 10.1007/s00382-006-0205-x.

Guilyardi, Eric, Andrew Wittenberg, Alexey Fedorov, C. Collins, Antonietta Capotondi, Geert Jan Van Oldenborgh, and Tim Stockdale (2009). "Understanding El Niño in Ocean-Atmosphere General Circulation Models: Progress and Challenges". In: *Bulletin of the American Meteorological Society*. DOI: 10.1175/2008BAMS2387.1.

Gutowski Jr., William J., Filippo Giorgi, Bertrand Timbal, Anne Frigon, Daniela Jacob, Hyun-Suk Kang, Krishnan Raghavan, Boram Lee, Christopher Lennard, Grigory Nikulin, Eleanor O'Rourke, Michel Rixen, Silvina Solman, Tannecia Stephenson, and Fredolin Tangang (2016). "WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6". In: *Geoscientific Model Development*. DOI: 10.5194/gmd-9-4087-2016.

Gutowski, W. J., P. A. Ullrich, A. Hall, L. R. Leung, T. A. O'Brien, C. M. Patricola, R. W. Arritt, M. S. Bukovsky, K. V. Calvin, Z. Feng, A. D. Jones, G. J. Kooperman, E. Monier, M. S. Pritchard, S. C. Pryor, Y. Qian, A. M. Rhoades, A. F. Roberts, K. Sakaguchi, N. Urban, and C. Zarzycki (2020). "The Ongoing Need for High-Resolution Regional Climate Models: Process Understanding and Stakeholder Information". In: *Bulletin of the American Meteorological Society*. DOI: 10.1175/BAMS-D-19-0113.1.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani (2011). "Climate model bias correction and the role of timescales". In: *Hydrology and Earth System Sciences*. DOI: 10.5194/hess-15-1065-2011.

Hansen, Nicolaj, Peter L. Langen, Fredrik Boberg, Rene Forsberg, Sebastian B. Simonsen, Peter Thejll, Baptiste Vandecrux, and Ruth Mottram (2021). "Downscaled surface mass balance in Antarctica: impacts of subsurface processes and large-scale atmospheric circulation". In: *The Cryosphere*. DOI: 10.5194/tc-15-4315-2021.

Hansen, Nicolaj, Sebastian B. Simonsen, Fredrik Boberg, Christoph Kittel, Andrew Orr, Niels Souverijns, J. Melchior van Wessem, and Ruth Mottram (2022). "Brief communication: Impact of common ice mask in surface mass balance estimates over the Antarctic ice sheet". In: *The Cryosphere*. DOI: 10.5194/tc-16-711-2022.

Heinemann, Günther and Rolf Zentek (2021). "A Model-Based Climatology of Low-Level Jets in the Weddell Sea Region of the Antarctic". In: *Atmosphere*. DOI: 10.3390/atmos12121635.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut (2020). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.3803.

Hoffman, Matthew D. and Andrew Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research*. URL: http://jmlr.org/papers/v15/hoffman14a.html (visited on 10/11/2023).

Hourdin, Frédéric, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, Qingyun Duan, Doris Folini, Duoying Ji, Daniel Klocke, Yun Qian, Florian Rauser, Catherine Rio, Lorenzo Tomassini, Masahiro Watanabe, and Daniel Williamson (2017). "The Art and Science of Climate Model Tuning". In: *Bulletin of the American Meteorological Society*. DOI: 10.1175/BAMS-D-15-00135.1.

Kittel, Christoph, Charles Amory, Cécile Agosta, Nicolas C. Jourdain, Stefan Hofer, Alison Delhasse, Sébastien Doutreloup, Pierre-Vincent Huot, Charlotte Lang, Thierry Fichefet, and Xavier Fettweis (2021). "Diverging future surface mass balance between the Antarctic ice shelves and grounded ice sheet". In: *The Cryosphere*. DOI: 10.5194/tc-15-1215-2021.

Kopp, Robert E., Robert M. DeConto, Daniel A. Bader, Carling C. Hay, Radley M. Horton, Scott Kulp, Michael Oppenheimer, David Pollard, and Benjamin H. Strauss (2017). "Evolving Understanding of Antarctic Ice-Sheet Physics and

Ambiguity in Probabilistic Sea-Level Projections". In: *Earth's Future*. DOI: 10.1002/2017EF000663.

Kuipers Munneke, Peter, Stefan R. M. Ligtenberg, Michiel R. van den Broeke, and David G. Vaughan (2014). "Firn air depletion as a precursor of Antarctic ice-shelf collapse". In: *Journal of Glaciology*. DOI: 10.3189/2014JoG13J183.

Lenaerts, J. T. M. and M. R. van den Broeke (2012). "Modeling drifting snow in Antarctica with a regional climate model: 2. Results". In: *Journal of Geophysical Research-Atmospheres*. DOI: 10.1029/2010JD015419.

Lenaerts, J. T. M., M. R. van den Broeke, W. J. van de Berg, E. van Meijgaard, and P. Kuipers Munneke (2012). "A new, high-resolution surface mass balance map of Antarctica (1979-2010) based on regional atmospheric climate modeling". In: *Geophysical Research Letters*. DOI: 10.1029/2011GL050713.

Lenaerts, J. T. M., M. R. van den Broeke, S. J. Déry, G. König-Langlo, J. Ettema, and P. K. Munneke (2010). "Modelling snowdrift sublimation on an Antarctic ice shelf". In: *The Cryosphere*. DOI: 10.5194/tc-4-179-2010.

Lenaerts, J. T. M., S. Lhermitte, R. Drews, S. R. M. Ligtenberg, S. Berger, V. Helm, C. J. P. P. Smeets, M. R. van den Broeke, W. J. van de Berg, E. van Meijgaard, M. Eijkelboom, O. Eisen, and F. Pattyn (2017). "Meltwater produced by wind-albedo interaction stored in an East Antarctic ice shelf". In: *Nature Climate Change*. DOI: 10.1038/NCLIMATE3180.

Lima, Carlos H. R., Hyun-Han Kwon, and Yong-Tak Kim (2021). "A Bayesian Kriging model applied for spatial downscaling of daily rainfall from GCMs". In: *Journal of Hydrology*. DOI: 10.1016/j.jhydrol.2021.126095.

Liu, M., K. Rajagopalan, S. H. Chung, X. Jiang, J. Harrison, T. Nergui, A. Guenther, C. Miller, J. Reyes, C. Tague, J. Choate, E. P. Salathé, C. O. Stöckle, and J. C. Adam (2014). "What is the importance of climate model bias when projecting the impacts of climate change on land surface processes?" In: *Biogeosciences*. DOI: 10.5194/bg-11-2601-2014.

Luckman, Adrian, Andrew Elvidge, Daniela Jansen, Bernd Kulessa, Peter Kuipers Munneke, John King, and Nicholas E. Barrand (2014). "Surface melt and ponding on Larsen C Ice Shelf and the impact of fohn winds". In: *Antarctic Science*. DOI: 10.1017/S0954102014000339.

Mann, Stephen (1999). "Cubic precision Clough-Tocher interpolation". In: *Computer Aided Geometric Design*. DOI: 10.1016/S0167-8396(98)00038-7.

Maraun, Douglas (2016). "Bias Correcting Climate Change Simulations - a Critical Review". In: *Current Climate Change Reports*. DOI: 10.1007/s40641-016-0050-x.

Martin, Daniel F., Stephen L. Cornford, and Antony J. Payne (2019). "Millennial-Scale Vulnerability of the Antarctic Ice Sheet to Regional Ice Shelf Collapse". In: *Geophysical Research Letters*. DOI: 10.1029/2018GL081229.

Matsuoka, Kenichi, Anders Skoglund, George Roth, Jean de Pomereu, Huw Griffiths, Robert Headland, Brad Herried, Katsuro Katsumata, Anne Le Brocq, Kathy Licht, Fraser Morgan, Peter D. Neff, Catherine Ritz, Mirko Scheinert, Takeshi Tamura, Anton Van de Putte, Michiel van den Broeke, Angela von Deschwanden, César Deschamps-Berger, Brice Van Liefferinge, Stein Tronstad, and Yngve Melvær (2021). "Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands". In: *Environmental Modelling & Software*. DOI: 10.1016/j.envsoft.2021.105015.

Mendez, Maikel, Ben Maathuis, David Hein-Griggs, and Luis-Fernando Alvarado-Gamboa (2020). "Performance Evaluation of Bias Correction Methods for Climate Change Monthly Precipitation Projections over Costa Rica". In: *Water*. DOI: 10.3390/w12020482.

Mottram, Ruth, Nicolaj Hansen, Christoph Kittel, J. Melchior van Wessem, Cécile Agosta, Charles Amory, Fredrik Boberg, Willem Jan van de Berg, Xavier Fettweis, Alexandra Gossart, Nicole P. M. van Lipzig, Erik van Meijgaard, Andrew Orr, Tony Phillips, Stuart Webster, Sebastian B. Simonsen, and Niels Souverijns (2021). "What is the surface mass balance of Antarctica? An

intercomparison of regional climate model estimates". In: *The Cryosphere*. DOI: 10.5194/tc-15-3751-2021.

Munneke, P. Kuipers, M. R. van den Broeke, J. T. M. Lenaerts, M. G. Flanner, A. S. Gardner, and W. J. van de Berg (2011). "A new albedo parameterization for use in climate models over the Antarctic ice sheet". In: *Journal of Geophysical Research: Atmospheres*. DOI: 10.1029/2010JD015113.

Munneke, P. Kuipers, A. J. Luckman, S. L. Bevan, C. J. P. P. Smeets, E. Gilbert, M. R. van den Broeke, W. Wang, C. Zender, B. Hubbard, D. Ashmore, A. Orr, J. C. King, and B. Kulessa (2018). "Intense Winter Surface Melt on an Antarctic Ice Shelf". In: *Geophysical Research Letters*. DOI: 10.1029/2018GL077899.

Nicholls, Robert J., Natasha Marinova, Jason A. Lowe, Sally Brown, Pier Vellinga, Diogo de Gusmão, Jochen Hinkel, and Richard S. J. Tol (2011). "Sea-level rise and its possible impacts given a 'beyond 4°C world' in the twenty-first century". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. DOI: 10.1098/rsta.2010.0291.

Noël, Brice, J. Melchior van Wessem, Bert Wouters, Luke Trusel, Stef Lhermitte, and Michiel R. van den Broeke (2023). "Higher Antarctic ice sheet accumulation and surface melt rates revealed at 2 km resolution". In: *Nature Communications*. DOI: 10.1038/s41467-023-43584-6.

Oppenheimer, M., B.C. Glavovic, J. Hinkel, R. van de Wal, A.K. Magnan, A. Abd-Elgawad, R. Cai, M. Cifuentes-Jara, R.M. DeConto, T. Ghosh, J. Hay, F. Isla, B. Marzeion, B. Meyssignac, and Z. Sebesvari (2022). "Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities". In: *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*. Ed. by Intergovernmental Panel on Climate Change (IPCC). Cambridge: Cambridge University Press. ISBN: 978-1-00-915796-4. DOI: 10.1017/9781009157964.006.

Orr, Andrew, Amélie Kirchgaessner, John King, Tony Phillips, Ella Gilbert, Andrew Elvidge, Mark Weeks, Alan Gadian, Peter Kuipers Munneke, Michiel van den

Broeke, Stuart Webster, and Daniel McGrath (2021). "Comparison of kilometre and sub-kilometre scale simulations of a foehn wind event over the Larsen C Ice Shelf, Antarctic Peninsula using the Met Office Unified Model (MetUM)". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.4138.

Orr, Andrew, Tony Phillips, Stuart Webster, Andy Elvidge, Mark Weeks, Scott Hosking, and John Turner (2014). "Met Office Unified Model high-resolution simulations of a strong wind event in Antarctica". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.2296.

Paolo, Fernando S., Helen A. Fricker, and Laurie Padman (2015). "Volume loss from Antarctic ice shelves is accelerating". In: *Science*. DOI: 10.1126/science.aaa0940.

Parish, Thomas R. and David H. Bromwich (2007). "Reexamination of the Near-Surface Airflow over the Antarctic Continent and Implications on Atmospheric Circulations at High Southern Latitudes". In: *Monthly Weather Review*. DOI: 10.1175/MWR3374.1.

Phan, Du, Neeraj Pradhan, and Martin Jankowiak (2019). *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro*. DOI: 10.48550/arXiv.1912.11554.

Pollard, David, Robert DeConto, and Richard Alley (2015). "Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure". In: *Earth and Planetary Science Letters*. DOI: 10.1016/j.epsl.2014.12.035.

Pörtner, Hans-Otto, Debra Cynthia Roberts, Melinda M. B. Tignor, Elvira S. Poloczanska, Katja Mintenbeck, Andrès Alegría, Marlies Craig, Stefanie Langsdorf, Sina Löschke, Vincent Möller, Andrew Okem, and Bardhyl Rama, eds. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.*

Pritchard, Hamish, Stefan Ligtenberg, Helen Fricker, David Vaughan, Michiel Van den Broeke, and L Padman (2012). "Antarctic ice-sheet loss driven by basal melting of ice shelves". In: *Nature*. DOI: 10.1038/nature10968.

Qian, Weijia and Howard H. Chang (2021). "Projecting Health Impacts of Future Temperature: A Comparison of Quantile-Mapping Bias-Correction Methods". In: *International Journal of Environmental Research and Public Health*. DOI: 10.3390/ijerph18041992.

Rasmussen, Carl Edward (2004). "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*. Ed. by Olivier Bousquet, Ulrike von Luxburg, and Gunnar Rätsch. Berlin, Heidelberg: Springer. ISBN: 978-3-540-28650-9. DOI: 10.1007/978-3-540-28650-9_4.

Räty, Olle, Jouni Räisänen, and Jussi S. Ylhäisi (2014). "Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations". In: *Climate Dynamics*. DOI: 10.1007/s00382-014-2130-8.

Rignot, E., G. Casassa, P. Gogineni, W. Krabill, A. Rivera, and R. Thomas (2004). "Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf". In: *Geophysical Research Letters*. DOI: 10.1029/2004GL020697.

Scambos, T. A., J. A. Bohlander, C. A. Shuman, and P. Skvarca (2004). "Glacier acceleration and thinning after ice shelf collapse in the Larsen B embayment, Antarctica". In: *Geophysical Research Letters*. DOI: 10.1029/2004GL020670.

Scambos, Ted A., Christina Hulbe, Mark Fahnestock, and Jennifer Bohlander (2000). "The link between climate warming and break-up of ice shelves in the Antarctic Peninsula". In: *Journal of Glaciology*. DOI: 10.3189/172756500781833043.

Shuman, Christopher A., Etienne Berthier, and Ted A. Scambos (2011). "2001-2009 elevation and mass losses in the Larsen A and B embayments, Antarctic Peninsula". In: *Journal of Glaciology*. DOI: 10.3189/002214311797409811.

Sippel, S., F. E. L. Otto, M. Forkel, M. R. Allen, B. P. Guillod, M. Heimann, M. Reichstein, S. I. Seneviratne, K. Thonicke, and M. D. Mahecha (2016). "A novel bias correction methodology for climate impact simulations". In: *Earth System Dynamics*. DOI: 10.5194/esd-7-71-2016.

Slater, Andrew G., David M. Lawrence, and Charles D. Koven (2017). "Process-level model evaluation: a snow and heat transfer metric". In: *The Cryosphere*. DOI: 10.5194/tc-11-989-2017.

Storch, Hans von, Heike Langenberg, and Frauke Feser (2000). "A Spectral Nudging Technique for Dynamical Downscaling Purposes". In: *Monthly Weather Review*. DOI: 10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2.

Tebaldi, Claudia, Kevin Debeire, Veronika Eyring, Erich Fischer, John Fyfe, Pierre Friedlingstein, Reto Knutti, Jason Lowe, Brian O'Neill, Benjamin Sanderson, Detlef van Vuuren, Keywan Riahi, Malte Meinshausen, Zebedee Nicholls, Katarzyna B. Tokarska, George Hurtt, Elmar Kriegler, Jean-Francois Lamarque, Gerald Meehl, Richard Moss, Susanne E. Bauer, Olivier Boucher, Victor Brovkin, Young-Hwa Byun, Martin Dix, Silvio Gualdi, Huan Guo, Jasmin G. John, Slava Kharin, YoungHo Kim, Tsuyoshi Koshiro, Libin Ma, Dirk Olivié, Swapna Panickal, Fangli Qiao, Xinyao Rong, Nan Rosenbloom, Martin Schupfner, Roland Séférian, Alistair Sellar, Tido Semmler, Xiaoying Shi, Zhenya Song, Christian Steger, Ronald Stouffer, Neil Swart, Kaoru Tachiiri, Qi Tang, Hiroaki Tatebe, Aurore Voldoire, Evgeny Volodin, Klaus Wyser, Xiaoge Xin, Shuting Yang, Yongqiang Yu, and Tilo Ziehn (2021). "Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6". In: *Earth System Dynamics*. DOI: 10.5194/esd-12-253-2021.

Tedesco, Marco, Sarah Doherty, Xavier Fettweis, Patrick Alexander, Jeyavinoth Jeyaratnam, and Julienne Stroeve (2016). "The darkening of the Greenland ice

sheet: trends, drivers, and projections (1981–2100)". In: *The Cryosphere*. DOI: 10.5194/tc-10-477-2016.

Teutschbein, Claudia and Jan Seibert (2012). "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods". In: *Journal of Hydrology*. DOI: 10.1016/j.jhydrol.2012.05.052.

Trusel, Luke D., Karen E. Frey, Sarah B. Das, Kristopher B. Karnauskas, Peter Kuipers Munneke, Erik van Meijgaard, and Michiel R. van den Broeke (2015). "Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios". In: *Nature Geoscience*. DOI: 10.1038/NGEO2563.

Undén, Per, Laura Rontu, Heikki Järvinen, Peter Lynch, Javier Calvo-Sanchez, Gerard Cats, Joan Cuxart, Kalle Eerola, Carl Fortelius, and J. García-Moya (2002). "HIRLAM-5 scientific documentation". In: URL: https://www.researchgate.net/publication/278962772_HIRLAM-5_scientific_documentation (visited on 03/01/2022).

Van Meijgaard, E., L. H. Van Ulft, W. J. Van de Berg, F. C. Bosvelt, BJJM Van den Hurk, G. Lenderink, and A. P. Siebesma (2008). "The KNMI regional atmospheric model RACMO version 2.1". In: *Tech. Note Tech. Rep.*

Vautard, Robert, Nikolay Kadygrov, Carley Iles, Fredrik Boberg, Erasmo Buonomo, Katharina Bülow, Erika Coppola, Lola Corre, Erik van Meijgaard, Rita Nogherotto, Marit Sandstad, Clemens Schwingshackl, Samuel Somot, Emma Aalbers, Ole B. Christensen, James M. Ciarlo, Marie-Estelle Demory, Filippo Giorgi, Daniela Jacob, Richard G. Jones, Klaus Keuler, Erik Kjellström, Geert Lenderink, Guillaume Levavasseur, Grigory Nikulin, Jana Sillmann, Cosimo Solidoro, Silje Lund Sørland, Christian Steger, Claas Teichmann, Kirsten Warrach-Sagi, and Volker Wulfmeyer (2021). "Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble". In: *Journal of Geophysical Research: Atmospheres*. DOI: 10.1029/2019JD032344.

Verjans, V., A. A. Leeson, M. McMillan, C. M. Stevens, J. M. van Wessem, W. J. van de Berg, M. R. van den Broeke, C. Kittel, C. Amory, X. Fettweis, N. Hansen, F. Boberg, and R. Mottram (2021). "Uncertainty in East Antarctic Firn Thickness Constrained Using a Model Ensemble Approach". In: *Geophysical Research Letters*. DOI: 10.1029/2020GL092060.

Walters, D., Anthony Baran, Ian Boutle, Melissa Brooks, Paul Earnshaw, John Edwards, Kalli Furtado, Peter Hill, Adrian Lock, James Manners, Cyril Morcrette, Jane Mulcahy, Claudio Sanchez, Chris Smith, Rachel Stratton, Warren Tennant, Lorenzo Tomassini, Kwinten Weverberg, Simon Vosper, and Mohamed Zerroukat (2019). "The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations". In: *Geoscientific Model Development*. DOI: 10.5194/gmd-12-1909-2019.

Wang, Yali and Brahim Chaib-draa (2017). "An online Bayesian filtering framework for Gaussian process regression: Application to global surface temperature analysis". In: *Expert Systems with Applications*. DOI: 10.1016/j.eswa.2016.09.018.

Wang, Yetang, Xueying Zhang, Wentao Ning, Matthew A. Lazzara, Minghu Ding, Carleen H. Reijmer, Paul C. J. P. Smeets, Paolo Grigioni, Petra Heil, Elizabeth R. Thomas, David Mikolajczyk, Lee J. Welhouse, Linda M. Keller, Zhaosheng Zhai, Yuqi Sun, and Shugui Hou (2023). "The AntAWS dataset: a compilation of Antarctic automatic weather station observations". In: *Earth System Science Data*. DOI: 10.5194/essd-15-411-2023.

Washington, Warren M, Lawrence Buja, and Anthony Craig (2008). "The computational future for climate and Earth system models: on the path to petaflop and beyond". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. DOI: 10.1098/rsta.2008.0219.

Wessem, J. M. van, S. R. M. Ligtenberg, C. H. Reijmer, W. J. van de Berg, M. R. van den Broeke, N. E. Barrand, E. R. Thomas, J. Turner, J. Wuite, T. A. Scambos, and E. van Meijgaard (2016). "The modelled surface mass balance

of the Antarctic Peninsula at 5.5km horizontal resolution". In: *Cryosphere*. DOI: 10.5194/tc-10-271-2016.

Wessem, Jan Melchior van, Willem Jan van de Berg, Brice P. Y. Noël, Erik van Meijgaard, Charles Amory, Gerit Birnbaum, Constantijn L. Jakobs, Konstantin Krüger, Jan T. M. Lenaerts, Stef Lhermitte, Stefan R. M. Ligtenberg, Brooke Medley, Carleen H. Reijmer, Kristof van Tricht, Luke D. Trusel, Lambertus H. van Ulft, Bert Wouters, Jan Wuite, and Michiel R. van den Broeke (2018). "Modelling the climate and surface mass balance of polar ice sheets using RACMO2 – Part 2: Antarctica (1979–2016)". In: *The Cryosphere*. DOI: https://doi.org/10.5194/tc-12-1479-2018.

Wessem, Jan Melchior van, Carleen H. Reijmer, Willem Jan van de Berg, Michiel R. van den Broeke, Alison J. Cook, Lambertus H. van Ulft, and Erik van Meijgaard (2015). "Temperature and Wind Climate of the Antarctic Peninsula as Simulated by a High-Resolution Regional Atmospheric Climate Model". In: *Journal of Climate*. DOI: 10.1175/JCLI-D-15-0060.1.

Zhang, Yongshun, Miao Feng, Weimin Zhang, Huizan Wang, and Pinqiang Wang (2021). "A Gaussian process regression-based sea surface temperature interpolation algorithm". In: *Journal of Oceanology and Limnology*. DOI: 10.1007/s00343-020-0062-1.