Stochastic Generators for Multivariate Global Spatio-temporal Climate Data

Matthew Edwards

Thesis submitted for the degree of

Doctor of Philosophy

School of Mathematics, Statistics & Physics

Newcastle University

Newcastle upon Tyne

United Kingdom

March 2020

I dedicate this work to my mum and dad.

Acknowledgements

I would like to thank Stefano Castruccio for his excellent project, support and encouragement. Without his encouragement I would have never travelled to Boulder CO to work at NCAR and would have never met Dorit Hammerling. I would like to thank Dorit Hammerling for her support and her dedication in helping me publish my work. Furthermore, I would like to thank Dorit for introducing me to so many incredible spatio-temporal statisticians as their insights have been invaluable. I would like to thank Doug Nychka at NCAR and Will Kleiber at CU Boulder for reading and discussing my work with such enthusiasm. I would like to thank Robin Henderson for adopting me as his student and for his support and dedication in helping me complete my Thesis.

I would like to thank all the students and faculty of the CDT for Cloud Computing for Big data for making my postgraduate experience so enjoyable. I would like to thank all the students and faculty at NCAR that supported me so far from home, in particular Chintan Dalal and Johnny Hendrick.

Abstract

In order to understand and quantify the uncertainties in projections and physics of a climate model (deterministic model), a collection of climate simulations (an ensemble) is typically used. Given the high-dimensionality of the input space of a climate model, as well as the complex, non-linear relationships between the climate variables, a large ensemble is often required to accurately assess these uncertainties. If only a small number of climate variables are of interest at a specified spatial and temporal scale, the computational and storage expenses can be substantially reduced by training a statistical model on a small ensemble. The statistical model then acts as a stochastic generator able to simulate a large ensemble, given a small training ensemble. Previous work on stochastic generators has focused on modeling and simulating individual climate variables (e.g. surface temperature, wind speed) independently. Here, we introduce a stochastic generator (trivariate stochastic model) that jointly simulates three key climate variables. The parameters of this nonstationary global model are estimated with a sequence of marginal likelihood functions using large-scale parallelisation across many processors for more than 80 million data points. We demonstrate the feasibility of jointly simulating climate variables by training the stochastic generator on five ensemble members from a large ensemble project, and assess the stochastic generator simulations by comparing them to the ensemble members not used in training.

The multivariate spatio-temporal model introduced in Chapter 4 was published in the Journal of Agricultural, Biological and Environmental Statistics (Edwards et al., 2019). The theory of marginally parameterised models and stepwise maximum likelihood estimation introduced in Chapter 3 was submitted to the Journal of Computational Statistics and Data Analysis (Edwards et al., 2018).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

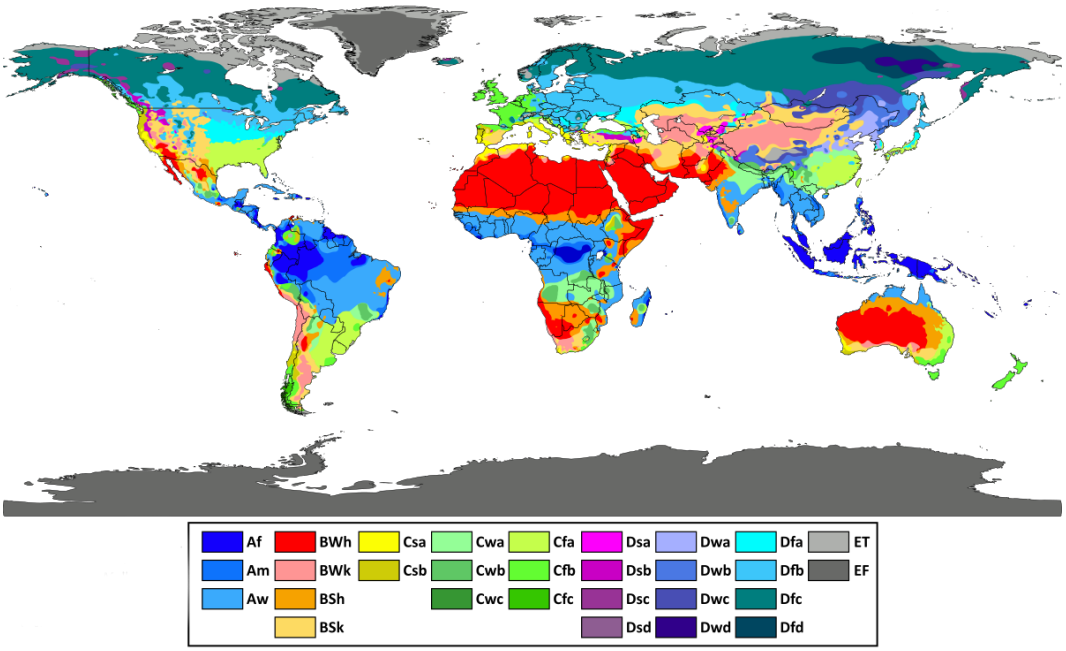
1.1 Weather and Climate

The difference between weather and climate is between actual states of the atmosphere and their short-term variations and aggregate (e.g. mean, max) states of the atmosphere and their long-term variations, respectively (McIlveen, 1991, p. 114). Weather is a local combination of meteorological components, including: temperature, atmospheric pressure, wind speed and direction, solar radiation, humidity and precipitation. In contrast, climate is a regional combination of aggregate meteorological components.

There are a number of schemes to classifying climates. The most popular is the Köppen-Geiger climate classification scheme (Geiger, 1954). Figure 1.1 is a world map of the Köppen-Geiger climate classification scheme (Peel et al., 2007). This scheme classifies climates into five major climate groups: tropical (A), arid (B), temperate (C), continental (D) and polar (E). These five major groups are then classified into minor groups. For example, the United Kingdom is classified as Cfc, i.e. temperate, without dry season (f) and cold summer (c) (Peel et al., 2007). These two minor groups depend on aggregate

meteorological components such as the precipitation of the driest and wettest month in summer and the temperature of the hottest month (Peel et al., 2007).

Figure 1.1: World map of the Köppen-Geiger climate classification scheme (Peel et al., 2007).



Variations in weather can affect individual humans from minute by minute (e.g. tornado warnings) to week by week (e.g. vacation plans). Variations in climate can affect humanity on a regional and global scale. For example, variations in climate can affect: natural and managed resources and systems, and their uses (e.g. Pednekar et al., 2005; Gonzalez et al., 2010; McGranahan et al., 2007; Perry et al., 2005; Thornton et al., 2009); human settlements, industry, and infrastructure (e.g. Willems et al., 2012; Glenn et al., 2013; Arkell and Darch, 2006) and human health, well-being, and security (e.g. Jackson et al., 2010; Webersik, 2010; Mendelsohn et al., 2006), see Field et al. (2014) for a technical summary. As a consequence, the science of climate change is extremely important to humankind.

1.2 Climate Science

Climate scientists are interested in how the climate has changed in the past, is changing at the present, and will change in the future (Farmer and Cook, 2013, p. 22). In recent decades climate scientists have focused on the present changes in the climate, commonly referred to as climate change. Climate change includes global warming; the observed increase in the average global temperature and its related effects (see Gillis, 2015). This observed global warming is unprecedented with respect to both the instrumental records, that cover the last two hundred years, and the paleoclimate proxy records that cover thousands of years (Stocker et al., 2013, p. 4). The International Panel on Climate Change (IPCC) have concluded in their Fifth Assessment Report (AR5) that it is *extremely likely* that the cause of this global warming over the last 50 years is anthropogenic (the result of human activity) (Stocker et al., 2013, p. 17). Specifically, human activities that emit greenhouse gases (GHGs), e.g carbon dioxide, methane, nitrous oxide and chlorofluorocarbons.

A GHG can absorb and reflect back thermal energy emitted from the earth that would otherwise be radiated towards space in a process named the greenhouse effect (McIlveen, 1991, pp. 592–593). It is projected that the average global temperature could increase four degrees Celsius due to the greenhouse effect in the next one hundred years (Stocker et al., 2013, pp. 89–90). Consequently, climate scientists are interested in how to decrease the rate of global warming. For this purpose, an important question for government policy makers is how different future GHG emission scenarios will affect the rate of global warming (Edenhofer et al., 2014). For the most recent assessment report the IPCC adopted the Representative Concentration Pathways (RCPs). These are four future GHG concentration (not emission) forcing scenarios (Van Vuuren et al., 2011). The GHG concentration scenarios aim to reflect a large range of possible anthropogenic GHG emission scenarios (Collins et al., 2013). For future assessment reports the IPCC will also adopt the Shared Socio-economic Pathways (SSPs). These are five future socio-economic trend scenarios

(ONeill et al., 2014). The socio-economic trend scenarios aim to integrate the future analysis of climate impact, vulnerability, adaptation and mitigation (Riahi et al., 2017). Exploring the affect of these pathways on the global climate system requires that experiments are conducted on global climate models.

1.3 Global Climate Models

Global climate models are a representation of the physical, chemical and biological processes in the atmosphere, ocean, cryosphere and lithosphere that affect the climate system. The physical atmospheric and oceanic submodels of a global climate model are general circulation models (GCMs). GCMs are represented with the Navier-Stokes equations on a rotating sphere and the thermodynamic equations (Washington and Parkinson, 2005). An oceanic GCM and atmospheric GCM can be combined to form an atmosphere-ocean coupled general circulation model (CGCM). These CGCM combined with different submodels (e.g. sea ice) form the basis for most global climate models (Sun and Hansen, 2003). Generally, most submodels are represented using a system of partial differential equations. Global climate models that have submodels that can interact during simulation are called fully coupled models. For example, The Community Earth System Model (CESM) is a fully coupled global climate model that is available to the climate science community (Hurrell et al., 2013). Administration of the CESM is maintained by the National Centre for Atmospheric Research (NCAR).

Simulating from a CGCM involves solving the systems of partial differential equations, for each GCM and submodel, numerically given some initial conditions. Initial conditions can be obtained from observations, previous CGCM simulations or both. Some experiments require multiple simulations from a CGCM. Multiple simulations constitute the members of an ensemble. Members of an ensemble can be obtained from round-off level differences in the initial conditions.

High spatial resolution simulations require a substantial computational infrastructure. For instance, the NCAR-Wyoming Supercomputing Center (NWSC) provides the computational infrastructure for NCAR to obtain simulations from the CESM. This center includes Cheyenne, a 5.34 PFLOPS[1] supercomputer with 313 TB of main memory. This is currently (June 2019) the 36th most powerful computer in the world: `https://www.top500.org`. Simulations generated from different global climate models given the same initial conditions can be qualitatively different in some aspects. Consequently, it is important to perform an experiment on a number of global climate models. This is one of the objectives of the Coupled Model Intercomparison Project (CMIP).

1.4 The Coupled Model Intercomparison Project

The objective of the CMIP is to study and compare experiments on different CGCMs. The CMIP provides a community-based framework for global climate model diagnosis, validation, intercomparison and data access. The major CMIP experiments come in phases, each addressing a series of questions. For instance, the CMIP phase 6 (Eyring et al., 2016) addresses three questions: How does the earth system respond to forcing? What are the origins and consequences of systematic model biases? How can we assess future climate changes given internal climate variability, predictability, and uncertainties in scenarios? To answer these questions the CMIP phase 6 has endorsed 21 model intercomparison projects (MIPs). These MIPs include: the Ice Sheet MIP, aimed at "improving confidence in projections of the sea level rise associated with mass loss from the ice sheets of Greenland and Antarctica" (Eyring et al., 2016, Table 3) and the Land-Use MIP aimed at "quantifying the effects of land use on climate and biogeochemical cycling (past-future), and assessing the potential for alternative land management strategies to mitigate climate change." (Eyring et al., 2016, Table 3).

[1] **p**eta (10^{15}) **fl**oating-point **o**perations **p**er **s**econd

Participating in a CMIP phase involves simulating ensembles with a large number of members (large ensembles). The computational and storage expenses associated with simulating a large ensemble are substantial. For instance, the CMIP phase 6 (Meehl et al., 2014) is expected to require approximately 1 billion core-hours of computation and 12 PB of storage (Paul et al., 2015). At the NWSC, this corresponds to approximately 290 compute days and one fifth of the file storage system.

Although a large number of climate variables are simulated during these experiments, often only a small number of them are of high interest to a climate scientist. For instance, between June 2014 and March 2018, approximately 64% of the 1,168 climate variables from the CESM Large Ensemble (Kay et al., 2015) had never been downloaded and approximately 14% of the climate variables contribute to over 90% of the 178,121 downloads (Strand and Baker, 2018). Since only a small number of climate variables are of high interest to a large proportion of the climate science community, it is valuable to focus efforts on developing methods to effectively reduce the computational and storage expenses for these high interest climate variables. Compression methods provide one approach to reducing the storage expense of large ensembles.

1.5 Compression

Compression is the reduction of the number of bits required to represent data (Sayood, 2017). Compression methods involve two algorithms: the first compresses data into a compressed representation and the second reconstructs data from that compressed representation. In lossless compression methods the reconstructed data is identical to the original (i.e. no information loss). However, in so-called lossy compression the reconstructed data is only an approximation to the original. The maximum lossless compression rate possible for data depends on the information entropy (unpredictability) of the source (see MacKay, 2003). Specifically, the higher the information entropy the lower the maximum

compression rate. For floating-point scientific data the significand is highly-entropic after several digits (Baker et al., 2014). As a consequence, the lossless compression of floating-point scientific data is limited due to a low maximum compression rate. As ensembles are floating-point scientific data, lossy compression methods are preferable.

As the application of lossy compression methods to CGCM simulations is relatively new, Baker et al. (2016) explored a number of methods and concluded that a compression rate of 5:1 could be achieved for annually averaged variables with the fpzip method (Lindstrom and Isenburg, 2006). Although lossy compression methods can reduce the storage expense of large ensembles they cannot reduce the computational expense. Stochastic generators are an alternative to compression methods that can reduce both the computational and storage expenses of large ensembles.

1.6 Stochastic Climate Generators

A stochastic generator (Tagle et al., 2017) is a statistical model that aims to represent the generating process of an ensemble. Once a stochastic generator has been trained on an ensemble (training ensemble) the stochastic generator can simulate additional ensemble members. Hence, only the stochastic generator is stored once it is trained and a large ensemble can be obtained from the stochastic generator when required. A stochastic generator reduces the storage expense, since a stochastic generator requires substantially less storage than a large ensemble, and a stochastic generator can reduce the computational expense if only a small training ensemble is required.

There have been numerous statistical models proposed as stochastic generators. Castruccio et al. (2013) proposed a model that was able to capture latitudinal non-stationarities, Castruccio and Genton (2014) proposed a non-parametric generalization to capture longitudinal non-stationarities. Castruccio and Genton (2016) generalized these models to

account for altitude, Castruccio (2016) to account for multiple climate models and Castruccio and Guinness (2017) to account for land/ocean effects. Guinness and Hammerling (2018) proposed a conditional model that was able to capture spatial non-stationarities conditional on summary statistics of the ensemble. These models have been proposed to reduce the storage expense of storing large ensembles but not to reduce the computational cost, since they were trained on all the available ensemble members. Furthermore, these are univariate models that can only capture univariate dependencies and simulate climate variables independently.

1.7 Multivariate Spatio-temporal Model

A stochastic generator can capture the multivariate dependencies between climate variables if it can simulate them jointly. Jointly simulating ensemble variables requires a multivariate global spatio-temporal model. A Gaussian model is adopted for the multivariate global spatio-temporal model, see Chapter 4. For the Gaussian model there are a large number of approaches to modelling very large spatial data sets, for example: fixed rank kriging (Cressie and Johannesson, 2008), lattice kriging (Nychka et al., 2015), predictive processes (Banerjee et al., 2008), covariance tapering (Kaufman et al., 2008), multi-resolution approximations (Katzfuss, 2017), nearest neighbour processes (Datta, Banerjee, Finley and Gelfand, 2016) and stochastic partial differential equations (Lindgren et al., 2011). There are also a number of approaches to modelling very large spatio-temporal data sets, for example: weighted composite likelihoods (Bevilacqua et al., 2012), full-scale approximations (Zhang et al., 2015), dynamic nearest neighbour processes (Datta, Banerjee, Finley, Hamm, Schaap et al., 2016) and dynamic multi-resolution spatial models (Johannesson et al., 2007). There are a small number of approaches to modelling very large multivariate spatio-temporal data sets, for example: dynamic coregionalization models (Gelfand et al., 2005).

These statistical models have been developed to reduce the cost of evaluating the Gaussian likelihood function. Hence, they are capable of reducing the computational and memory costs associated with large data sets, but not necessarily reducing the computational and memory costs associated with large parameter sets.

1.8 Computational and Memory Costs

Capturing the complex spatio-temporal and multivariate dependencies in an ensemble will required a flexible model with a large number of parameters. However, performing maximum likelihood estimation (MLE) with a large number of parameters is often infeasible due to the curse of dimensionality (Bellman, 1961) regardless of the cost of evaluating the likelihood function. Consequently, a method of estimation is required that can overcome the computational and memory costs associated with large data sets and large parameter sets.

The estimation method introduced for this purpose requires a particular model named a marginally parameterized model. The definition of a marginally parameterised model is motivated by the following observation. Given a likelihood function, if there existed a marginal likelihood function that depends on a subset of the parameters then this marginal likelihood function can be used to estimate that subset of parameters with a computational and memory cost associated with the size of the data subset and the parameter subset. A model is marginally parameterised if there exists a sequence of marginal likelihood functions that can be used to estimate all the parameters. The definition of a marginally parameterised model is motivated by the model introduced in Castruccio et al. (2013) which is marginally parameterised. The introduced estimation method, stepwise maximum likelihood estimation (SMLE), estimates the parameters of a marginally parameterised model. The proposed model for the stochastic generator is a marginally parameterised multivariate extension of the univariate global spatio-temporal model introduced

in (Castruccio et al., 2013).

1.9 Summary

It is valuable to focus efforts on developing methods to effectively reduce the computational and storage expenses for high interest climate variables. Stochastic generators are statistical models that aim to represent the generating process of an ensemble. They can reduce storage expenses since only the stochastic generator requires storage and they can reduce computational expenses if only a small training ensemble is required. For the climate variables of interest capturing multivariate and spatio-temporal dependencies with a stochastic generator requries a multivariate global spatio-temporal model. Due to the complexity of the climate system dependencies the model would required a large number of parameters. However, performing maximum likelihood estimation (MLE) with a large number of parameters is often infeasible due to the curse of dimensionality.

This work aims to provide a framework for modelling and estimation when the number of data points and parameters is large. This includes marginally parameterised models and SMLE. Since SMLE involves the estimation of parameters conditional on previously estimated parameters this work aims to provide some asymptotic guarantees on the estimators. In addition to this theoretic framework for modelling and estimation, this work also introduces the first multivariate global spatio-temporal stochastic generator that can simulate multiple climate variables jointly.

1.10 Outline

In Chapter 2 the large ensemble data set is introduced and exploratory data analysis is performed on this data set. In Chapter 3 the marginally parameterised model and the SMLE method are introduced with the associated theory and simulation study. In Chapter 4 the multivariate global spatio-temporal model is introduced. In Chapter 5 the details of the SMLE method applied to this marginally parameterised model is introduced. In Chapter 6 the univairate and multivariate diagnostics are provided to gague the the quality of the stochastic generator and in Chapter 8 the work is concluded.

Chapter 2

Data

2.1 Data Set

The CESM large ensemble (CESM-LE) (Kay et al., 2015) is a publicly available (`https://www.earthsystemgrid.org/`) ensemble intended to understand uncertainties related to climate variability and climate change. The CESM-LE consists of 33 members simulated from 1920 to 2100 on an approximately one-degree latitude-longitude grid from the fully-coupled CESM version 1 (CESM1) using the Community Atmosphere Model version 5 (CAM5) (Hurrell et al., 2013). These members where simulated on the Yellowstone supercomputer at the NWSC. The ensemble members are obtained from different initial conditions. These initial conditions are the result of round-off level perturbations of an atmospheric temperature field. Due to the chaotic nature of the global climate model–a non-linear dynamical model–and its sensitivity to initial conditions, the ensemble members are approximately independent and identically distributed (IID) (Collins and Allen, 2002; Collins, 2002; Branstator and Teng, 2010). This is an assumption adopted in Chapter 4. From 1920 to 2005 the historical GHG concentration pathway was used and from 2006 to

2100 the representative concentration pathway (RCP) 8.5 (Moss et al., 2008) was used. Each RCP is a hypothetical GHG concentration pathway used to understand the effect of particular future GHG concentrations on the climate system. Further details of the model set-up can be found in Kay et al. (2015).

The CESM-LE is stored in Network Common Data Form (NetCDF) files in single-precision (32 bit) floating-point format. NetCDF files have been developed to store array-oriented scientific data. For the CESM-LE the arrays correspond to the values of climate variables indexed across time, longitude and latitude. For each ensemble member and climate variable there is a collection of NetCDF files for each available sampling frequency, e.g. monthly, daily and 6-hourly. Details of the climate variables and their available sampling frequencies can be found at `http://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html`. For each sampling frequency, the corresponding collection of NetCDF files are for different time periods (e.g. 2006 to 2100).

Based on consultations with climate scientists at NCAR, three high interest climate variables with important multivariate dependencies were selected: total precipitable water (TMQ), surface temperature (TS) and $10m$ wind speed (U10). TMQ is the mass of water in a column of the atmosphere per unit area (kg/m^2), TS is the radiative surface temperature (K) and U10 is the wind speed $10m$ above the surface (m/s). These climate variables are available at each spatial location of the latitude-longitude grid. To reduce the substantial computational and memory expenses required to train the stochastic generator the historical GHG concentration pathway time period (i.e. 1920 to 2005) is not considered. For the RCP 8.5 time period the three climate variables are stored in 198 NetCDF files; two NetCDF files for each combination of ensemble member and climate variable. The three climate variables are annually averaged; this is a sampling frequency not available in the CESM-LE. The climate variables are annually averaged to approximate, through the central limit theorem, the Gaussian assumption introduced in Chapter 4. Higher sampling frequencies would require a non-Gaussian model. For example, monthly wind was modelled

with a Tukey g-and-h model in Jeong et al. (2017) to capture the non-Gaussian skewness and kurtosis that appeared at this sampling frequency. In contrast to Castruccio et al. (2013), Castruccio and Genton (2014), Castruccio and Genton (2016) and Castruccio and Guinness (2017) this work includes the Arctic and Antarctic latitudinal bands (data points north of 70 degrees and south of 62 degrees). Previously, these latitudinal bands were excluded since they involve data points at the poles that are at very close proximity. Although the proximity of these data points can negatively affect estimation, this work performs estimation in such a way, see Chapter 5, that any negative effect would be isolated to the model at the poles and would not affect the model at low and middle latitudes. The corresponding ensemble array consists of 33 members, $T = 95$ years, $L = 288$ longitudes, $M = 192$ latitudes and $V = 3$ variables; a total of approximately half a billion data points. This will be referred to as the CESM-LE.

The 33 ensemble members of the CESM-LE are randomly divided into a training ensemble and a testing ensemble. The training ensemble will be used to train the multivariate global spatio-temporal model and the test ensemble will be used to test the quality of the multivariate global spatio-temporal simulations. The training ensemble is used for exploratory data analysis to eliminate bias. The test ensemble is not used before the diagnostic presented in Chapter 6. The training ensemble consists of $R = 5$ randomly selected members and the test ensemble consists of the 28 remaining members. In Jeong et al. (2018) a lack of fit index was used to concluded that $R = 5$ randomly selected members was sufficient for training with similar temporal, longitudinal and latitudinal model specifications. Consequently, the training ensemble corresponds to approximately 80 million data points.

2.2 Notation

The training ensemble consists of $R = 5$ members, $T = 95$ years, $L = 288$ longitudes, $M = 192$ latitudes and $V = 3$ variables. Let Y denote the $R \times T \times L \times M \times V$ training ensemble, in array form, where the element $\mathrm{Y}[r, t, l, m, v]$ corresponds to the value of variable v from member (realisation) r at year t, longitude l and latitude m. These indices are indexed from one with the exception of longitude l which is index from zero. Zero-based indexing for longitude is adopted for the spectral methods introduced in Section 4.4 and 4.5. Indices from the element notation $\mathrm{Y}[r, t, l, m, v]$ are omitted to denoted subarrays of Y. For example, $\mathrm{Y}[r, v]$ denotes the $T \times L \times M$ subarray (cube) indexed over year t, longitude l and latitude m for member r and variable v, see Figure 2.1. Figure 2.1 also displays a matrix $\mathrm{Y}[r, m, v]$ that corresponds to a particular latitude m and a vector (time series) $\mathrm{Y}[r, l, m, v]$ that corresponds to a particular longitude l and a latitude m. This indexing notation is used extensively in this chapter to denote data subsets. Square brackets are also used to index the rows and columns of matrices (e.g. covariance matrix).

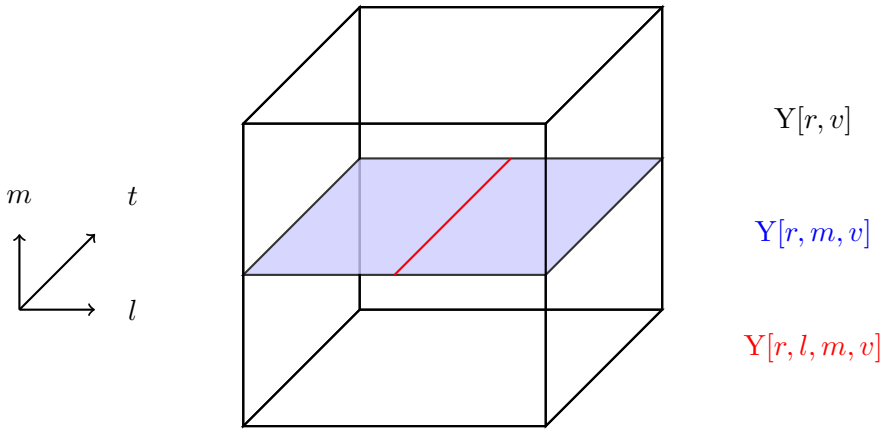


Figure 2.1: The $T \times L \times M$ subarray (cube) indexed over year t, longitude l and latitude m for member r and variable v with a matrix $\mathrm{Y}[r, m, v]$ that corresponds to a particular latitude m and a vector $\mathrm{Y}[r, l, m, v]$ that corresponds to a particular longitude l and a latitude m.

The programming language `R` is selected for computing and the training ensemble is

represented in this programming language with an array. This array is stored in memory as a vector $\mathbf{y}$ defined as vec(Y) where vec$(\cdot)$ is the vec-operator

$$\text{vec}(\text{Y}) \coloneqq \sum_{v} \cdots \sum_{r} (\mathbf{e}_{v,V} \otimes \cdots \otimes \mathbf{e}_{r,R}) \cdot \text{Y}[r, t, l, m, v]$$

and $\mathbf{e}_{k,K}$ is a vector of length K with the value one in element k and the value zero in all the other elements, see Hardy and Steeb (2010). Hence, contiguous elements of $\mathbf{y}$ are also contiguous in memory (Knuth, 1997, Section 2.2.6). As a consequence, if computation is performed in parallel over $\text{Y}[v]$ for all v then all the elements of each subarray are contiguous in memory, however, if computation is performed in parallel over $\text{Y}[r]$ for all r then none of the elements of each subarray are contiguous in memory. The more contiguous elements of an array are in memory the more efficient the computation. Computation in Chapter 5 is performed in parallel over $\text{Y}[l, m, v]$ for all l, m and v; then over $\text{Y}[m, v]$ for all m and v; and then over $\text{Y}[v]$ for all v. Hence, the order of the training ensemble dimensions, as represented in `R`, is computationally optimal.

Although, an array is convenient for representing the training ensemble in `R` and denoting data subsets for marginally parameterised models, arrays are not convenient for specifying multivariate statistical models as most multivariate statistical models (e.g. multivariate Gaussian) are defined on vectors. As a consequence, the subarray notation applied to the vector $\mathbf{y}$ is defined as the subarray notation applied to Y followed by vectorisation. As an example, $\mathbf{y}[r, l, m, v]$ defines the time series vector of length T corresponding to member r, longitude l, latitude m and variable v. This notation is used extensively in this chapter. In general, uppercase letters denote arrays, bold lowercase letters denote vectors and bold uppercase letters denote random vectors.

2.3 Exploratory Data Analysis

Exploratory data analysis is performed in order to explore the multivariate dependencies between the three selected climate variables and the spatio-temporal dependencies within them. The structure of these dependencies will be discussed in the context of the global climate system and used to inform the assumptions of the model introduced in Chapter 4.

Statistics that ignore the multivariate and spatio-temporal dependencies between data points are considered first. The statistics considered are the means $\overline{Y}_{r,v}$ and medians, upper and lower quantiles and minimums and maximums calculated over time, longitude and latitude for each ensemble member r and climate variable v. To account for the non-regular latitude-longitude grid, with more grid-points per unit area towards the poles, these statistics are area-weighted. For example, the area-weighted mean for ensemble member r and cliamte variable v is

$$\overline{Y}_{r,v} = \frac{1}{T \cdot L \cdot M} \sum_{t=1}^{T} \sum_{l=0}^{L-1} \sum_{m=1}^{M} Y[r,t,l,m] \cdot \cos(\theta_l)$$

where θ_l is the longitudinal angle of latitudinal band l. Figure 2.1 displays the means and standard errors (in parentheses) of these area-weighted statistics for each climate variable. For example, the mean and standard error of $\overline{Y}_{r,v}$ for variable v is

$$\overline{Y}_{v} = \frac{1}{R} \sum_{r=1}^{R} \overline{Y}_{r,v} \quad \text{and} \quad \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\overline{Y}_{r,v} - \overline{Y}_{v})^2}$$

respectively. The minimums are all greater than zero, since all the climate variables are non-negative valued. Consequently, adopting a Gaussian assumption for these climate variables could be problematic if the stochastic generator simulations contain negative values. The standard errors indicate that most of the area-weighted statistics are equal to one decimal place over the ensemble members. The exceptions include all maximums and the TS minimum.

Table 2.1: The mean and standard errors (in parenthesis) over the training ensemble members of the area-weighted means and medians, area-weighted lower and upper quantiles and area-weighted minimums and maximum for each climate variable.

	TMQ (kg/m^2)	TS (K)	U10 (m/s)
Minimum	0.23 (0.01)	214.60 (0.25)	0.65 (0.00)
1st Quantile	15.04 (0.03)	282.56 (0.03)	3.40 (0.00)
Median	25.31 (0.04)	294.90 (0.02)	6.53 (0.01)
Mean	28.05 (0.03)	289.64 (0.02)	6.12 (0.00)
3rd Quantile	40.66 (0.05)	300.46 (0.02)	8.00 (0.01)
Maximum	72.66 (0.70)	312.16 (0.23)	13.73 (0.09)

Figure 2.2 displays the global means over time (left panels) and the longitudinal means over latitude (right panels) for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble. The variation in global means, for each year, and longitudinal means, for each latitude, over the training ensemble members are included in grey. In the left panels of Figure 2.2 the grey is obscured since the variation in longitudinal means, for each latitude, over the training ensemble members is very small. The upper and middle left panels of Figure 2.2 indicate that global mean TMQ and TS are increasing and the rate of increase is increasing in the first half of the century and decreasing in the second half. The rates for global TMQ and TS are approximately 0.74 (kg/m^2) and 0.42 (K) per decade, respectively. The trend in TS is positively correlated to the GHG emission scenario (.i.e. RCP 8.5) due to the greenhouse effect and the trend in TMQ is positively correlated to global temperature since the atmosphere can contain more water vapour at higher temperatures (McIlveen, 1991, Section 5.4). The lower left panel of Figure 2.2 indicates that the global mean U10 is decreasing at a constant rate of approximately 0.0053 (m/s) per decade. The relationship between decreasing U10, known as global stilling, and changing climate is currently uncertain. The model introduced in Chapter 4 must include non-linear means over time to capture these

trends.

The upper and middle right panels of Figure 2.2 indicate that longitudinal mean TMQ and TS is larger towards the equator and smaller towards the poles. TS is positively correlated to absolute latitude since the annual mean solar elevation angle (from the horizon to the centre of the sun) is larger towards the equator and the larger the solar elevation angle the more solar radiation is absorbed at the surface (McIlveen, 1991, Section 8.7). The lower right panel of Figure 2.2 indicates that longitudinal mean U10 is largest between negative 60 and negative 30 degrees North. This is since the Southern Ocean is at these latitudes and annual mean wind speed is larger over ocean than land (McIlveen, 1991, Section 10.11). The model introduced in Chapter 4 must include complex non-linear means over latitude to capture these trends.

The remainder of the exploratory data analysis is derived from simple linear regression models, with year centered at 2005 as the regressor, trained at each spatial location for the three climate variables from the training ensemble. Since year is centered at 2005, the simple linear model intercepts correspond to means at 2005. Figure 2.3 displays the simple linear model intercepts (left panels) and trends (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble. The upper and middle left panels of Figure 2.3 are consistent with the upper and middle right panels of Figure 2.2, in that longitudinal mean TMQ and TS are positively correlated with absolute latitude. However, they provide a more detailed description of the means. For example, 2005 mean TMQ and TS are negatively correlated with snow cover (e.g. Greenland, Antarctica) and large mountain ranges (e.g. Himalayas, Andes). 2005 mean TMQ is lower over mountain ranges due to the higher altitudes and 2005 mean TS is lower over snow cover since snow has a high albedo, a measure of the reflection of solar radiation (McIlveen, 1991, Section 8.2). The lower left panel of Figure 2.3 indicates that 2005 mean U10 is lower over land than ocean. 2005 mean U10 is lower over the land since the land provides more friction due to mountains and trees than the

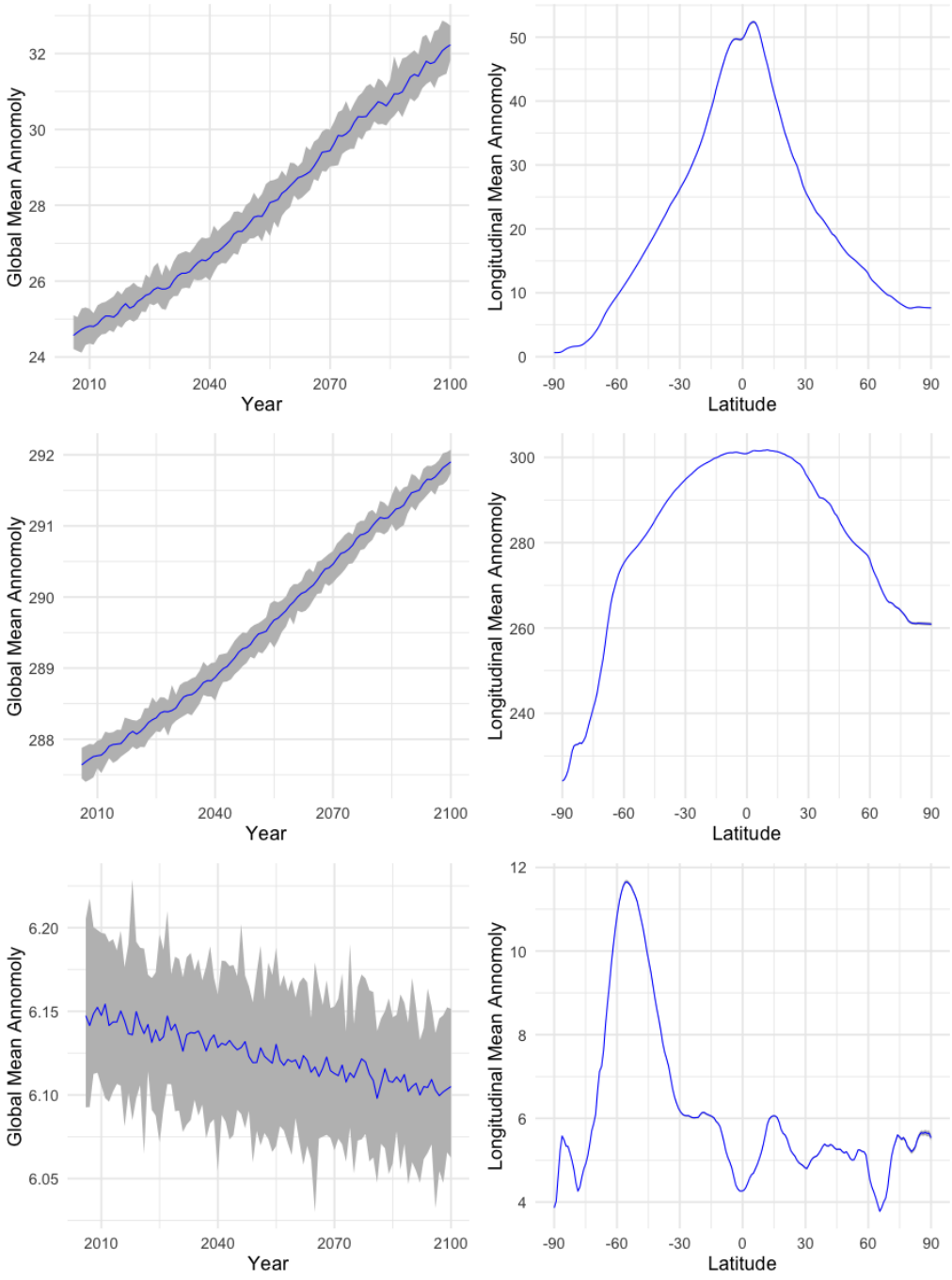


Figure 2.2: The global means over time (left panels), the longitudinal means over latitude (right panels) for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble. The maximum and minimum member global means, for each year, and longitudinal means, for each latitude, are included in grey (maximums and minimums are indistinguishable from the means in right panels).

ocean (McIlveen, 1991, Section 10.11). The lower left panel of Figure 2.3 indicates that the higher 2005 mean U10 over the Southern Ocean is due to more than just a land-ocean effect. The 2005 mean U10 over the Southern Ocean is amplified as the general circulation of the atmosphere (McIlveen, 1991, Section 4.7) is across lines of longitude (i.e. east-west) and the Southern Ocean covers entire lines of longitude. The means for the model introduced in Chapter 4 must be flexible enough to capture these 2005 means. This flexibility cannot be limited to just latitudinal and land-ocean effects or the important phenomena discussed above related to moutain ranges and the Southern Ocean, among many others, will not be well represented in the stochastic generator simulations.

As suggested in the upper and middle right panels of Figure 2.2 TMQ and TS are increasing across the the majority of the globe in Figure 2.3. The trends in TMQ are higher over the tropics (between negative and positive 23 degrees north). Paradoxically, the increase in TMQ over the tropics is considered to be due to a reduction in high-altitude clouds that absorb solar radiation (Su et al., 2017). The trends in TS are higher over land than ocean, known as the land-ocean warming contrast, and highest over the North Pole due to Arctic amplification. It has been proposed that the land-ocean warming contrast is due to there being limited moisture availability over land, which affects surface cooling (Byrne and Ogorman, 2013). Arctic amplification is considered to be primarily due to the surface albedo feedback (Serreze and Francis, 2006) where melting snow and ice causes more solar radiation to be absorbed and more snow and ice to be melted. The lower right panel of Figure 2.3 indicates that U10 is increasing and decreasing across the globe. The trends for the model introduced in Chapter 4 must be flexible enough to capture, not only non-linear trends in time, but variation in these non-linear trends over the globe. Otherwise, important phenomena such as Arctic amplification, the land-ocean warming contrast, among others, will not be well represented in the stochastic generator simulations.

Figure 2.4 displays the simple linear model residual standard deviations (left panels)

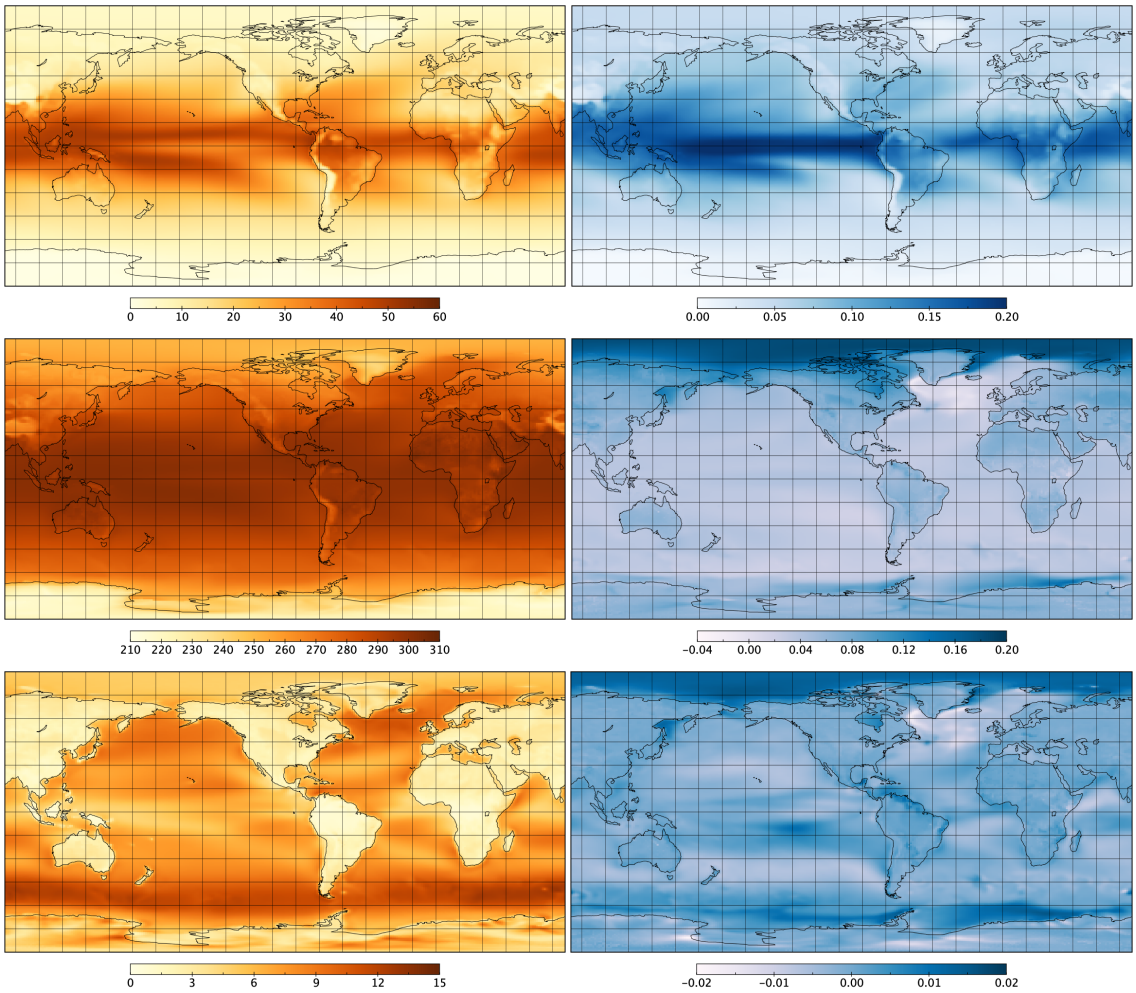


Figure 2.3: The simple linear model, with year as the regressor, 2005 means (left panels) and trends (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble.

and lag-one residual temporal auto-correlations (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble. The left panels of Figure 2.4 indicates that the TMQ, TS and U10 residual standard deviations are larger over the tropical Pacific Ocean. The tropical Pacific Ocean is related to the El Niño-Southern Oscillation (ENSO), a complex non-periodic variation in sea surface temperatures and wind speeds that affect precipita-

tion (McIlveen, 1991, Section 14.3). The ENSO phenomenon is captured by the larger residual standard deviations in this region. The middle and lower left panels of Figure 2.4 indicates that the TS and U10 residual standard deviations are larger over the poles and oceans, respectively. The standard deviations in the model introduced in Chapter 4 must be flexible enought to capture these variation in residual standard deviation over the globe. Otherwise important phenomena such as ENSO will not be well represented in the stochastic generator simulations.

The right panels of Figure 2.4 suggest that the TMQ lag-one residual temporal auto-correlations are unstructured, since the highs and lows do not correspond to any specific climate regions, but the TS and U10 lag-one residual temporal auto-correlation highs and lows do. For example the TS lag-one residual temporal auto-correlations are higher over the Greenland current. The Greenland current is a cold and low salinity (i.e. low salt level) ocean current that travels clock-wise round the southern coast of Greenland. The lower right panel of Figure 2.4 indicates that U10 lag-one residual temporal auto-correlations are higher over mountain ranges. The temporal component of the model introduced in Chapter 4 must be flexible enough to capture these variation in lag-one residual temporal auto-correlation over the globe.

Figure 2.5 displays the simple linear model lag-one residual longitudinal auto-correlations (left panels) and lag-one residual latitudinal auto-correlations (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble. Lag-one residual longitudinal and latitudinal auto-correlations are the auto-correlations between the residuals at one grid-point and the residuals at one grid-point north and east respectively. The panels in Figure 2.5 indicate that the lag-one residual longitudinal and latitudinal auto-correlations are positive with only a few exceptions and approximately constant over the oceans. This suggested that a stationary isotropic assumption would be reasonable over the ocean for the spatial component of the model introduced in Chapter 4. The structure of the spatial dependen-

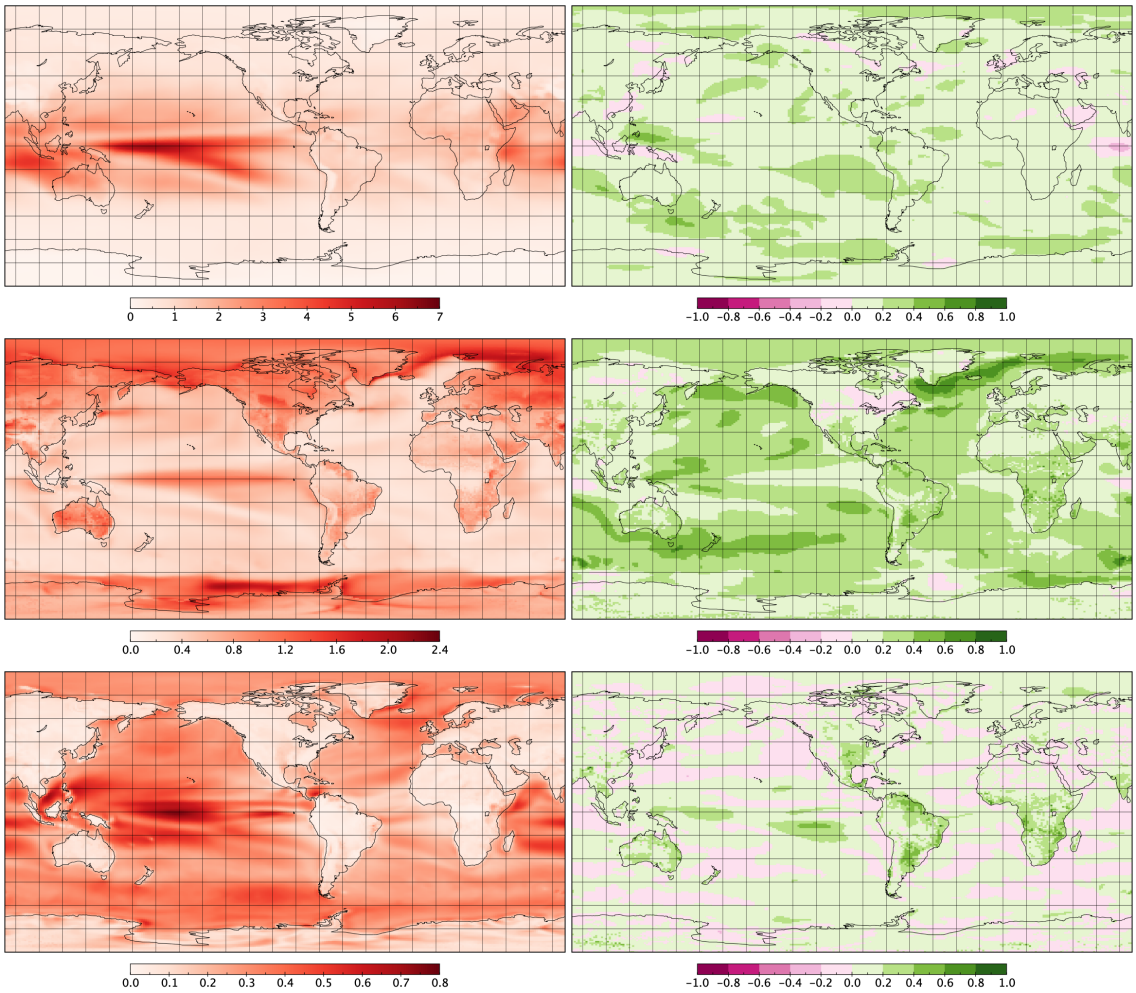


Figure 2.4: The simple linear model, with year as the regressor, residual standard deviations (left panels) and lag-one residual temporal auto-correlations (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble.

cies over land is more complex. Both lag-one residual longitudinal and latitudinal auto-correlations are reduced at some coastal and mountainous regions (e.g. Australian East coast, Andes). These are even reduced to negative lag-one residual auto-correlations on the Australian West coast. Capturing these variations in lag-one residual auto-correlation over the land would require a non-stationary assumption in the spatial component of the model introduced in Chapter 4.

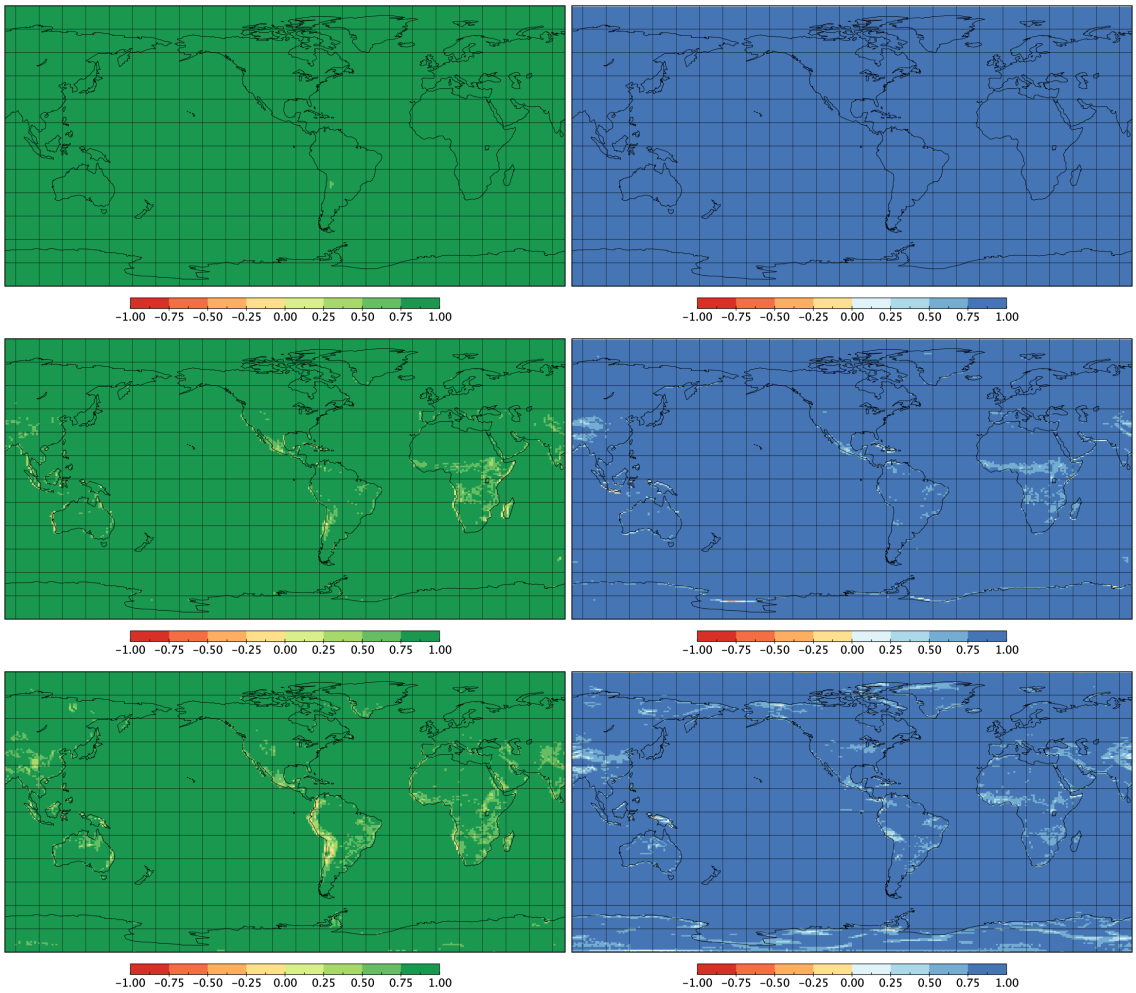


Figure 2.5: The simple linear model, with year as the regressor, lag-one residual longitudinal auto-correlations (left panels) and lag-one residual latitudinal auto-correlations (right panels) at each spatial location for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables from the training ensemble.

Figure 2.6 displays the simple linear model residual cross-correlations at each spatial location between the TMQ and TS (upper panel), TMQ and U10 (middle panel) and TS and U10 (lower panel) climate variables from the training ensemble. The upper panel of Figure 2.6 indicates that the residual cross-correlation between TMQ and TS is mostly positive but with negative residual cross-correlation in regions such as Australia, Western United States and Somalia. Saturation vapour pressure, a measure of the maximum

amount water vapour per unit of the atmosphere can contain, is positively correlated with temperature (McIlveen, 1991, Section 2.3). Hence, if TS is high and the atmosphere is rising, the temperature of the corresponding column of atmosphere is increasing and therefore TS is positively correlated with TMQ. In hot desert regions (e.g. Australia) TS is high but the corresponding column of atmosphere is falling and not necessarily increasing in temperature. This can result in a negative cross-correlation between TS and TMQ. The lower panel of Figure 2.6 indicates that the residual cross-correlation between TS and U10 is mostly low over oceans in the Frigid zone (approximately below and above negative and positive 66 degrees north, respectively) and mostly high over oceans in the Tropic, Subtropic and Temperate zones (not the Frigid zone). The variation in residual cross-correlations over land is more complex. To capture these multivariate dependencies over the globe the model introduced in Chapter 4 must have a complex multivariate specification.

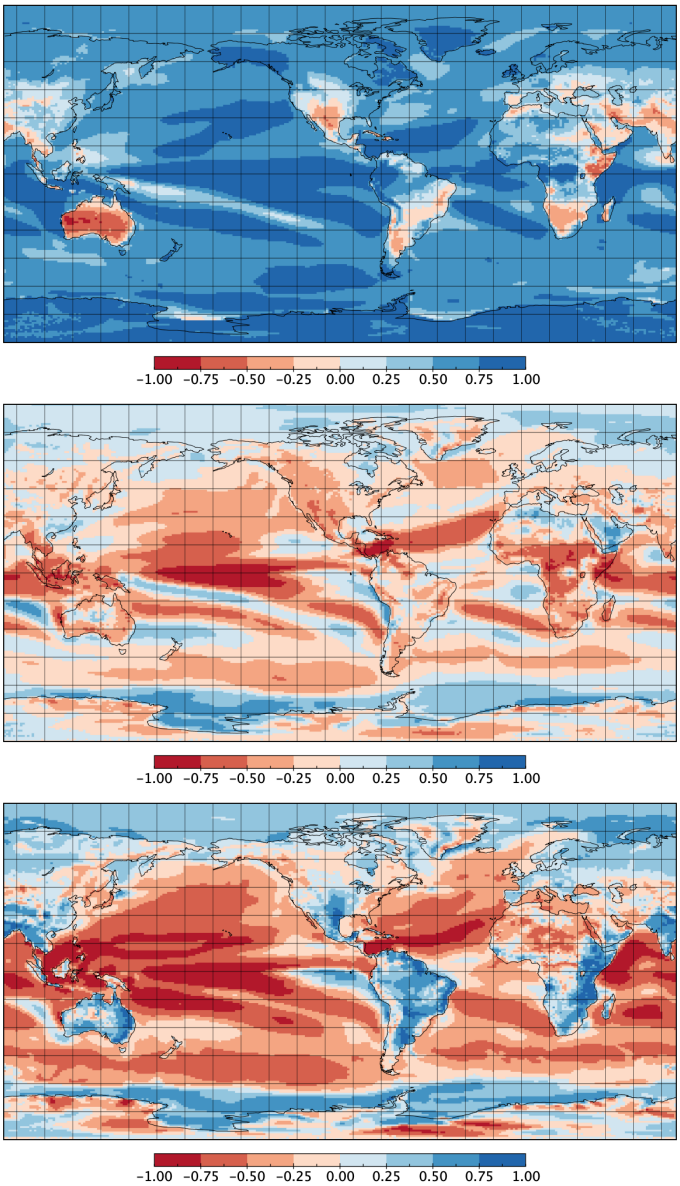


Figure 2.6: The simple linear model, with year as the regressor, residual cross-correlations at each spatial location between the TMQ and TS (upper panel), TMQ and U10 (middle panel) and TS and U10 (lower panel) climate variables from the CESM large ensemble.

Chapter 3

Theory

The training ensemble consists of approximately 80 million data points. These data points have complex multivariate and spatio-temporal dependencies that represent important climate phenomena (e.g. land-ocean warming contrast, Arctic amplification). Accurately simulating these phenomena will require a flexible multivariate global spatio-temporal model with a large number of parameters. However, estimating a large number of parameters for a large number of data points requires substantial computational and memory costs. To overcome these costs a class of models is introduced with an associated estimation method that allows the estimation of a large number of parameters with a substantial reduction in costs.

3.1 Numerical Optimisation Cost

To understand why estimating a large number of parameters for a large data set requires substantial computational and memory costs, the computational and memory costs of maximum likelihood estimation (MLE) must be understood in detail. Maximum likelihood

estimates are the values of the parameters that maximize the likelihood or log-likelihood function. Hence, obtaining maximum likelihood estimates is an optimisaiton problem. Unless estimates can be obtained in closed-form, estimates must be obtained approximately through numerical optimisation (Nocedal and Wright, 1999). In the multivariate spatio-temporal context, estimates are typically not available in closed-form. As a consequence it is assumed that closed-form estimates are not available and the costs of MLE correspond with those of numerical optimisation. A numerical optimisation algorithm includes a number of iterations, each iteration includes a number of evaluations and each evaluation includes some computational and memory costs. As a consequence, the costs associated with a numerical optimisation algorithm depend on: the costs associated with evaluating the likelihood function, the number of evaluations included in each iteration and the number of iterations.

3.1.1 Evaluation Cost

Denote $\mathbf{y}$ a data set consisting of N data points and denote $\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta})$ a corresponding likelihood function that depends on a parameter vector $\boldsymbol{\theta}$ consisting of P parameters. Evaluating the likelihood function at $\boldsymbol{\theta}_0$ requires computing the value of $\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}_0)$. The computational and memory costs of computing $\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}_0)$ depend on the number of floating-point operations performed (flops) and the number of floating-point numbers stored (memory). In this context a floating-point number is a computer representation of a real-valued number and a floating-point operation is a CPU operation (e.g. addition, multiplication) on these floating-point numbers that can be performed in a single clock cycle. The exact number of flops and memory depend on the algorithm used to compute $\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}_0)$. It is assumed that addition and multiplication are available and that higher order operations are not. This is important, since for modern processing units that can perform matrix multiplications in a single clock cycle (e.g. Tensor Core GPU), flops are not a good measure of computational cost.

For a given algorithm the number of flops and memory required can be represented as a function of the number of data points. For example, $2N-1$ flops are required to compute $\|\mathbf{y}\|^2$ (N multiplications and $N-1$ additions). A function $f(N)$, that represents how the number of flops and memory scale with the number of data points for a particular algorithm, is often replaced with a big $\mathcal{O}$ set of functions that contains it. Table 3.1 displays five big $\mathcal{O}$ sets: the constant, logarithmic, linear, polynomial and exponential sets. Formally, a function $f(N)$ is contained in the set $\mathcal{O}(g(N))$ if there exists a N_0 such that for all $N \geq N_0$ then $f(N) \leq K \cdot g(N)$ for some constant K. Heuristically, $f(N)$ is contained in $\mathcal{O}(g(N))$ if $f(N)$ is bounded above by $g(N)$ asymptotically. These big $\mathcal{O}$ sets provide a convenient way to compare how the computational and memory costs of evaluation algorithms scale with the number of data points. This is particularly important if N is very large.

Table 3.1: Five big $\mathcal{O}$ sets of functions: the constant, logarithmic, linear, polynomial and exponential sets.

Big $\mathcal{O}$ Set	Description
$\mathcal{O}(1)$	Constant
$\mathcal{O}(\log(N))$	Logarithmic
$\mathcal{O}(N)$	Linear
$\mathcal{O}(N^c)$	Polynomial
$\mathcal{O}(\exp(N))$	Exponential

Since the model introduced in Chapter 4 is multivariate Gaussian the computational and memory costs of evaluating the Gaussian likelihood function are considered. The multivariate Gaussian log-likelihood function is

$$\ln \mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta}) = -\frac{1}{2}\ln|2\pi\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\theta}))^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\theta})) \tag{3.1}$$

where $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean vector-valued function and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix-valued

function of the parameter vector $\boldsymbol{\theta}$. The multivariate Gaussian likelihood function is denoted as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y})$ where the dependence of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ on the parameter vector $\boldsymbol{\theta}$ is implicit. Since a covariance matrix is symmetric and positive-definite (SPD), the Cholesky decomposition routine that exploits SPD matrix structure (Golub and Van Loan, 2012, Section 4.2) can be used to evaluate (3.1) in $\mathcal{O}(N^3)$ flops and $\mathcal{O}(N^2)$ memory. This is the most computational and memory efficient method to evaluate (3.1) if the covariance matrix has no additional structure. Hence, this is the control to which other approaches are compared.

Since the training set contains 80 million data points, evaluating the multivariate Gaussian likelihood with the Cholesky decomposition routine would require on the order of 5.12×10^{23} flops and 6.4×10^{15} memory. This correspond to over three years of compute time on the Cheyenne supercomputer and 25.6 PB of memory (4 byte floating-point number) for one evaluation. Note that a non-Gaussian likelihood evaluation often requires considerably more flops and memory. For example, a max-stable likelihood evaluation requires $\mathcal{O}(B_N)$ flops, where B_N is the Bell number of order N (Castruccio et al., 2016). Hence, considering a non-Gaussian likelihood function to reduce the number of flops and memory required for likelihood evaluation is not common.

One approach to reducing the number of flops and memory required to evaluate (3.1) is to include additional covariance matrix structure (e.g. sparse, low-rank, circulant) (Golub and Van Loan, 2012; Davis, 2012) that a linear algebra routine can exploit. For example, the Whittle method (Whittle, 1954) includes a circular covariance matrix structure that can be exploited by the fast Fourier transform (FFT) routine to evaluate (3.1) in $\mathcal{O}(N \log(N))$ flops and $\mathcal{O}(N)$ memory. The Whittle method is introduced in detail in Section 4.4. In a spatial data context, other examples of this approach include: stochastic partial differential equations (Lindgren et al., 2011) that include sparse inverse covariance matrix (precision matrix) structure and predictive processes (Banerjee et al., 2008) that include low-rank covariance matrix structure into (3.1). Section 1.7 includes a large

number of methods used in the context of spatial, spatio-temporal and multivariate spatio-temporal data sets. The motivation for these methods is to reduce the number of flops and memory required to evaluate the likelihood function. Hence, they are ideal for large data sets.

3.1.2 Number of Iterations

Numerical optimisation algorithms include a number of iterations and, in the context of likelihood function optimisation, each iteration includes a number of likelihood evaluations. For example, in line search methods (Nocedal and Wright, 1999, Chapter 3) one or more likelihood evaluations are included for each iteration to calculate the search direction and step length. Since, each iteration of most numerical optimisation algorithms include only one or two likelihood evaluations, e.g. Nelder-Mead (Nelder and Mead, 1965) and Conjugate Gradient (Hestenes and Stiefel, 1952), the computational and memory costs of an iteration are equal to those of a likelihood evaluation in terms of big $\mathcal{O}$ sets. As a consequence, only the costs of likelihood evaluation in terms of big $\mathcal{O}$ sets and the number of iterations are considered.

The number of iterations a numerical optimisation algorithm requires depends on the number of parameters P (Vavasis, 1991). How the number of iterations scale with the number of parameters depends on the optimisation algorithm and properties of the likelihood function. The property of the likelihood function that is the most important is concavity. Given a likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$, where $\boldsymbol{\theta} \in \Theta$ and Θ is a convex set in a real vector space, the likelihood function is concave if for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ and $\lambda \in [0, 1]$

$$\mathcal{L}(\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_2 \mid \mathbf{y}) \geq \lambda\mathcal{L}(\boldsymbol{\theta}_1 \mid \mathbf{y}) + (1-\lambda)\mathcal{L}(\boldsymbol{\theta}_2 \mid \mathbf{y}).$$

Heuristically, a function is concave if the line segment between any two points on the function is below or on the function. The distinction between concave and non-concave

likelihood functions is important, since how the number of iterations scales with the number of parameters depends substantially on this property. If the likelihood function is concave then numerical optimisation algorithms exist where the number of iterations is polynomial in P. For example, the ellipsoid algorithm (Yudin and Nemirovskiĭ, 1977) requires $\mathcal{O}(P^2)$ iterations. However, if the likelihood function is non-concave, algorithms *only* exist where the number of iterations is exponential in P (Yudin and Nemirovskiĭ, 1983). This is often described as the curse of dimensionality (Bellman, 1961).

To exploit concave multivariate Gaussian likelihood functions it would be useful to know some conditions under which a covariance matrix-valued function of the parameter vector $\boldsymbol{\theta}$ results in a concave multivariate Gaussian likelihood function. Questions regarding covariance matrix-valued functions and concave Gaussian likelihood functions are interesting. However, this work does not attempt to answer these questions. Instead, it is assumed throughout that equation (3.1) is non-concave. Since this implies that the number of iterations is exponential in P, an exponential lower bound on the function of P, big $\mathcal{O}$ sets are inappropriate as they represent upper bounds. In this work big Ω sets are used to replace the function of P. Formally, a function $f(P)$ is a member of the set $\Omega(g(P))$ if there exists a P_0 such that for all $P \geq P_0$ then $f(P) \geq K \cdot g(P)$ for some constant K. Heuristically, $f(P)$ is contained in $\Omega(g(P))$ if $f(P)$ is bounded below by $g(P)$ asymptotically. Therefore, for a non-concave likelihood function any numerical optimisation algorithm will require $\Omega(c^P)$ iterations for some constant $c > 1$. Hence, even if the cost of evaluation is one flop, if $c = 2$ and $P = 100$, any numerical optimisation algorithm will require over 7.5 millions years of compute time on the Cheyenne supercomputer to obtain a global optimum. Consequently, estimating a large number of parameters from a non-concave Gaussian likelihood function is often infeasible regardless of the computational and memory costs of likelihood evaluation.

3.2 Multi-stage Approaches

Multi-stage approaches are stepwise estimation methods often used in applications with different types of dependence (e.g. temporal, spatial) such as non-linear mixed models (Giltinan and Davidian, 1995). Schabenberger and Gotway (2017, pp. 431-433) proposed a two-stage approach for spatio-temporal data where the sampling design consists of regularly spaced time points at a fixed (in time) number of spatial locations. In this approach, time series submodels are trained at each spatial location in the first stage and then a spatial submodel is trained to the residuals in the second stage. Hence, temporal parameters are estimated in the first stage and spatial parameters are estimated in the second stage. A large number of multi-stage approaches, with more than two stages, have been proposed in the context of global climate data (Castruccio et al., 2013; Castruccio and Guinness, 2017) where temporal, longitudinal and latitudinal parameters are estimated in three stages. An extension has been proposed in Castruccio and Genton (2016) where altitudinal parameters are estimated in a fourth stage. In the context of neuroscience a multi-stage approach for whole-brain data was proposed in Castruccio et al. (2018). These models have a reduced evaluation cost, as each submodel is defined over a data subset consisting of less than N data points. They also require a reduced number of iterations, since each submodel depends on less than P parameters. Therefore, they are ideal for dealing with large spatio-temporal data sets that require complex models with a large number of parameters.

These multi-stage approaches have been introduced in the aforementioned statistical literature as a collection of *ad-hoc* approaches for modelling large complex global spatio-temporal data where the problem suggests that modelling at multiple scales is useful for learning properties of interest. The absence of an underlying and unifying framework has prevented a full understanding of the generalisability of such approaches. Perhaps more importantly, a multi-stage approach does not guarantee an underlying joint model and

hence probabilistic statements about the data are not available. For example, without a joint model, minimum mean squared error prediction cannot be performed as conditional models are not available. This is particularly problematic in a spatio-temporal context where minimum mean squared error prediction (Kriging) is often the primary objective. It is also problematic for simulation since there is no joint model from which to simulate. Furthermore, the absence of an underlying and unifying framework has prevented the development of multi-stage asymptotic results.

3.3 Marginally Parameterised Model

Denote $\mathbf{y}$ a data set consisting of N data points and denote $\mathcal{L}(\mathbf{y} \mid \boldsymbol{\theta})$ a corresponding joint likelihood function that depends on a parameter vector $\boldsymbol{\theta}$ consisting of P parameters. The following definition of a marginally parameterised model provides an underlying and unifying framework for the multi-stage approaches discussed in Subsection 3.2.

Definition 1 (Marginally Parametrised Model)**.** A model for $\mathbf{y}$ is marginally parameterised if there exists a finite sequence of $K > 1$ data subsets $(\mathbf{y}_k)$ such that the marginal model of $\mathbf{y}_k$ depends on a parameter subset with a partition $\boldsymbol{\theta}_k, \boldsymbol{\eta}_k$ where $\boldsymbol{\theta}_k \neq \varnothing$ and $\boldsymbol{\eta}_k \subseteq \boldsymbol{\theta}_1 \cup \cdots \cup \boldsymbol{\theta}_{k-1}$ $(\boldsymbol{\eta}_1 = \varnothing)$ for $k = 1, \ldots, K$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ is a partition of $\boldsymbol{\theta}$.

The sequence of data subsets $(\mathbf{y}_k)$ corresponds to a sequence of marginal models. Each marginal model depends on a parameter subset that is partitioned into a set of primary and nuisance parameters ($\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ respectively). The primary parameters correspond to the parameters previous marginal models in the sequence *do not* depend on; whereas, the nuisance parameters correspond to the parameters previous models in the sequence *do* depend on. Heuristically, the primary parameters of each marginal model control the dependencies *only* within data subsets; hence, marginally parameterized. As a consequence of these conditions, the parameter set of a marginally parameterised model can be

estimated with a sequence of marginal (estimated) likelihood functions (Pawitan, 2001, Section 10.6), see Subsection 3.4. They are estimated likelihood functions if they depend on parameter estimates.

The primary parameter subset $\boldsymbol{\theta}_k$ controls the dependencies *only* within the data subset $\mathbf{y}_k$. As a consequence, marginally parameterised models are ideal for modelling non-stationarities, since dependencies can vary across data subsets (in time and space). An implication of this is that marginally parameterised models can have a large number of parameters that require multiple realisations to estimate. Subsection 3.3.1 describes how a large number of parameters can be estimated in a spatio-temporal context without multiple realisations. However, for now only the multiple realisation context will be considered. One of the simplest marginally parameterised models is the bivariate Gaussian model where

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{bmatrix}$$

and $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Denote $\mathbf{y}^{(r)} = (y_1^{(r)}, y_2^{(r)})$ as the r-th realisation from the bivariate Gaussian model. To prove that this is a marginally parameterised model it must be demonstrated that there exists a finite sequence of data subsets that satisfy the conditions of Definition 1. Consider the finite sequence of three data subsets $\{y_1^{(1)}, \ldots, y_1^{(R)}\}$, $\{y_2^{(1)}, \ldots, y_2^{(R)}\}$ and $\{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(R)}\}$. The marginal model for the first and second data subsets are a product of R identical univariate Gaussian models. The first and second marginal models depend on parameters subsets with partitions $\boldsymbol{\theta}_1, \boldsymbol{\eta}_1$ and $\boldsymbol{\theta}_2, \boldsymbol{\eta}_2$, respectively, where $\boldsymbol{\theta}_1 = \{\mu_1, \sigma_1\}$, $\boldsymbol{\eta}_1 = \varnothing$, $\boldsymbol{\theta}_2 = \{\mu_2, \sigma_2\}$ and $\boldsymbol{\eta}_2 = \varnothing$. The (marginal) model for the third data subset is a product of R identical bivariate Gaussian models. This (marginal) model depends on a parameter subset with a partition $\boldsymbol{\theta}_3, \boldsymbol{\eta}_3$ where $\boldsymbol{\theta}_3 = \{\rho\}$ and $\boldsymbol{\eta}_3 = \boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2$. Clearly, this finite sequence of three data subsets satisfies the conditions of Definition 1. How these parameters are estimated with the SMLE method is introduced in Section 3.4.

3.3.1 Diagonal VARMA Model

In a spatio-temporal context it is not often that multiple realisations are available. Hence, an marginally parameterised model is required such that a large number of parameters can be estimated without multiple realisations. This section provides a framework for the two-stage approach proposed in Schabenberger and Gotway (2017). Hence, consider $\mathbf{y}$ as a spatio-temporal data set where the sampling design consists of T regularly spaced time points $t \in \mathbb{Z}_{>0}$ at S fixed (in time) spatial locations $\mathbf{x}_s \in \mathcal{M}^d$, where $\mathcal{M}^d$ is a d-dimensional manifold (e.g. plane, sphere); for a total of $N = S \cdot T$ data points. The spatio-temporal data set $\mathbf{y}$ is modelled with a vector auto-regressive moving average (VARMA) model (Lütkepohl, 2005, Chapter 11) with diagonal auto-regressive (AR) and moving average (MA) matrices (diagonal VARMA or DVARMA). The AR and MA orders are p and q, respectively. With $\mathbf{Y}_t$ denoting the random vector corresponding to the data points at all spatial locations at time point t, the DVARMA model is defined as

$$\mathbf{Y}_t - \boldsymbol{\mu}_t = \Sigma \mathbf{U}_t + \sum_{i=1}^{p} \Phi_i (\mathbf{Y}_{t-i} - \boldsymbol{\mu}_{t-i}) + \sum_{i=1}^{q} \Pi_i \Sigma \mathbf{U}_{t-i} \quad \text{where} \quad \boldsymbol{\mu}_t \coloneqq \sum_{i=0}^{d} \boldsymbol{\beta}_i z_{t,i}, \tag{3.2}$$

$\boldsymbol{\beta}_i = (\beta_{i,\mathbf{x}_s})$ are vectors of mean parameters, $z_{t,i}$ are regressor variables from a $T \times (d+1)$ regressor matrix Z, $\Sigma = \operatorname{diag}(\sigma_{\mathbf{x}_s})$ is the diagonal matrix of standard deviation parameters, $\Phi_i = \operatorname{diag}(\phi_{i,\mathbf{x}_s})$ are the diagonal matrices of AR parameters and $\Pi_i = \operatorname{diag}(\pi_{i,\mathbf{x}_s})$ are the diagonal matrices of MA parameters. Furthermore, $\mathbf{U}_t$ are IID zero-mean and unit-variance Gaussian innovations with correlation matrix $\mathbf{R}$ and correlation parameter set $\boldsymbol{\nu}$. For this model there is a mean, standard deviation, AR and MA parameter for each spatial location.

Let us consider two points $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}^d$ and let us denote $h = |t_1 - t_2|$. The diagonal VARMA spatio-temporal covariance function for lag h is

$$C(\mathbf{x}_1, \mathbf{x}_2, h) \coloneqq \sigma^2_{\mathbf{x}_1} \sigma^2_{\mathbf{x}_2} \rho_{\mathbf{x}_1,\mathbf{x}_2} \sum_{i=0}^{\infty} \psi_{i+h,\mathbf{x}_1} \psi_{i,\mathbf{x}_2} \tag{3.3}$$

where $\rho_{\mathbf{x}_1,\mathbf{x}_2}$ is the element of $\mathbf{R}$ corresponding to the row of $\mathbf{x}_1$ and the column of $\mathbf{x}_2$, $\psi_{i,\mathbf{x}}$ are the parameters corresponding to the infinite order MA process derived from the ARMA specification at location $\mathbf{x}$, see Appendix A.1 for a proof. This new result was required to derive the marginal likelihood functions. If the AR and MA parameters at each spatial location are assumed equal, then the spatio-temporal covariance function is separable (Cameletti et al., 2011). However, since these parameters are not assumed equal, they are spatially indexed. Consequently, the infinite order MA parameters are spatially indexed in (3.3) and the spatio-temporal covariance function is non-separable. As a consequence, this model can implicitly capture cross-temporal dependence through the innovations, see Chapter 6.

To prove that (3.2) is a marginally parameterised model we must demonstrate that there exists a finite sequence of data subsets that satisfy the conditions of Definition 1. Consider the finite sequence of $K = S + 1$ data subsets $(\mathbf{y}_{\mathbf{x}_k})$ where $\mathbf{y}_k$ is the time series at spatial location $\mathbf{x}_k$ for k from 1 to S and $\mathbf{y}_K = \mathbf{y}$. From (3.3) the mean and auto-covariance function of $\mathbf{y}_k$ is $\mu_{\mathbf{x}_k}$ and $\sigma^2_{\mathbf{x}_k} \sum_{i=0}^{\infty} \psi_{i+h,\mathbf{x}_k} \psi_{i,\mathbf{x}_k}$, respectively. The marginal model of $\mathbf{y}_k$ is therefore an ARMA model that depends on a parameter subset with a partition $\boldsymbol{\theta}_k, \boldsymbol{\eta}_k$ where

$$\boldsymbol{\theta}_k = \{\mu_{\mathbf{x}_k}\} \cup \{\sigma_{\mathbf{x}_k}\} \cup \left(\bigcup_{i=1}^{\infty} \{\psi_{i,\mathbf{x}_k}\} \right) = \{\mu_{\mathbf{x}_k}\} \cup \{\sigma_{\mathbf{x}_k}\} \cup \left(\bigcup_{i=1}^{p} \{\phi_{i,\mathbf{x}_k}\} \right) \cup \left(\bigcup_{j=1}^{q} \{\pi_{j,\mathbf{x}_k}\} \right) \quad (3.4)$$

and $\boldsymbol{\eta}_k = \varnothing$ for k from 1 to S. Note that this model is marginally parameterised as a result of the AR and MA parameter matrices being diagonal. The marginal (joint) model of $\mathbf{y}_K$ is a diagonal VARMA model that depends on a parameter subset with a partition $\boldsymbol{\theta}_K, \boldsymbol{\eta}_K$ where $\boldsymbol{\theta}_K = \boldsymbol{\nu}$ and $\boldsymbol{\eta}_K = \boldsymbol{\theta}_1 \cup \cdots \cup \boldsymbol{\theta}_S$. Clearly, this finite sequence of data subsets satisfies the conditions of Definition 1. How these parameters are estimated with the SMLE method is introduced in Section 3.4. For additional flexibility a marginally parameterised non-diagonal VARMA model could be required. However, the diagonal VARMA model is currently the only marginally parameterised VARMA model known.

3.4 Stepwise Maximum Likelihood Estimation

Assume that $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ is a marginally parameterised model likelihood function for $\mathbf{y}$ and denote $(\mathbf{y}_k)$ the corresponding finite sequence of data subsets that satisfies Definition 1. Let $\mathcal{L}_k(\boldsymbol{\theta}_k, \boldsymbol{\eta}_k \mid \mathbf{y}_k)$ denote the marginal likelihood function of $\mathbf{y}_k$, that depends on a parameter subset with partition $\boldsymbol{\theta}_k, \boldsymbol{\eta}_k$ for $k = 1, \ldots, K$. Instead of estimating $\boldsymbol{\theta}$ with the marginally parameterised joint likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ in one step (MLE), the SMLE method estimates $\boldsymbol{\theta}_1$ with the marginal likelihood function $\mathcal{L}_1(\boldsymbol{\theta}_1 \mid \mathbf{y})$ in step one (since $\boldsymbol{\eta}_1 = \varnothing$) and estimates $\boldsymbol{\theta}_k$ with the marginal estimated likelihood function $\widehat{\mathcal{L}}_k(\boldsymbol{\theta}_k \mid \mathbf{y}) = \mathcal{L}_k(\boldsymbol{\theta}_k, \widehat{\boldsymbol{\eta}}_k \mid \mathbf{y})$ in step $k = 2, \ldots, K$. Here $\widehat{\boldsymbol{\eta}}_k$ is obtained from primary parameter estimates obtained in previous steps. The SMLE method is defined in Algorithm 1.

Algorithm 1 Sequential Maximum Likelihood Estimation

1: $\widehat{\boldsymbol{\theta}}_1 \leftarrow \arg\max_{\boldsymbol{\theta}_1} \mathcal{L}_1(\boldsymbol{\theta}_1 \mid \mathbf{y})$
2: **for** $k \leftarrow 2$ *to* K **do**
3: $\widehat{\boldsymbol{\eta}}_k$ obtained from $\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{k-1}$
4: $\quad \widehat{\boldsymbol{\theta}}_k \leftarrow \arg\max_{\boldsymbol{\theta}_k} \mathcal{L}_k(\boldsymbol{\theta}_k, \widehat{\boldsymbol{\eta}}_k \mid \mathbf{y})$
5: **end for**
6: **return** $\widehat{\boldsymbol{\theta}}_k \leftarrow \arg\max_{\boldsymbol{\theta}_k} \widehat{\mathcal{L}}_k(\boldsymbol{\theta}_k \mid \mathbf{y})$

In general this is a sequential algorithm where each step is performed in a sequence. However, under certain conditions, some sequences of steps can be parallelised. In this case it is said that these steps can be performed in one stage. For example, the n steps from $k+1$ to $k+n$ can be performed in one stage if $\widehat{\boldsymbol{\eta}}_{k+j}$ for $j = 1, \ldots, n$ can be obtained after step k, since all the marginal estimated likelihood functions can be obtained after step k. Formally, the n steps from $k+1$ to $k+n$ can be performed in one stage if $\boldsymbol{\eta}_{k+j} \subseteq \boldsymbol{\theta}_1 \cup \cdots \cup \boldsymbol{\theta}_k$ for $j = 1, \ldots, n$. Note that the SMLE method can be performed in one stage when estimating the parameters of the ARMA models of the diagonal VARMA model introduced in Subsection 3.4.2. Hence, the parameters of the diagonal VARMA

model can be obtained in two stages.

Consider the bivariate Gaussian model introduced in Section 3.3 and the three data subsets used to demonstrate that this model is marginally parameterised. The marginal models that correspond to the first two data subsets can be used to estimate $\boldsymbol{\theta}_1 = \{\mu_1, \sigma_1\}$ and $\boldsymbol{\theta}_2 = \{\mu_2, \sigma_2\}$ in one stage. Furthermore, these estimates can be obtained in closed-form as

$$\widehat{\boldsymbol{\theta}}_i = \left\{ \frac{1}{R}\sum_{r=1}^{R} y_i^{(r)}, \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(y_i^{(r)} - \frac{1}{R}\sum_{r=1}^{R} y_i^{(r)}\right)^2} \right\}$$

for $i = 1, 2$. The (marginal) model that corresponds to the third data subset can be used to estimate $\boldsymbol{\theta}_3 = \{\rho\}$ in the second stage. The second stage is different from the first stage since $\boldsymbol{\eta}_3 = \{\mu_1, \mu_2, \sigma_1, \sigma_2\} \neq \varnothing$. Hence, the correlation parameter is estimated conditional on the estimates of the mean and standard deviation parameters. Often, the conditioning can be represented through a transformation. For example, the likelihood function for $\mathbf{y}^{(r)}$ can be represented as

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}^{(r)}) = \mathcal{N}(\mathbf{0}, \mathbf{R} \mid \mathbf{u}^{(r)}) \quad \text{where} \quad \mathbf{R} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}$$

and $u_i^{(r)} := (y_i^{(r)} - \mu_i)/\sigma_i$ for $i = 1, 2$. In this case the conditioning transformation is the standardisation transformation that results from subtracting the mean and dividing by the standard deviation. With this transformation estimating $\boldsymbol{\theta}_3$ conditional on $\widehat{\boldsymbol{\eta}}_3$ is equivalent to estimating $\boldsymbol{\theta}_3$ with the estimated likelihood function

$$\widehat{\mathcal{L}}_3(\boldsymbol{\theta}_3 \mid \mathbf{y}) = \prod_{r=1}^{R} \mathcal{N}(\mathbf{0}, \mathbf{R} \mid \widehat{\mathbf{u}}^{(r)})$$

where $\widehat{\mathbf{u}}^{(r)}$ is the standardisation transformation with respect to the estimated parameters. A conditioning transformation can be very important for reducing the computational and memory costs of successive steps and stages. The conditioning transformation after stage one for the DVARMA model, see Subsection 3.4.2, results in multiple realisations to

estimate the correlation parameters.

3.4.1 Consistency

Assume that the data set $\mathbf{y}$ is from a marginally parameterised model with true parameter set $\boldsymbol{\theta}^*$ and denote $\mathbf{Y}$ as the corresponding random vector. Define $\widehat{\boldsymbol{\theta}}_1(\mathbf{Y})$ as the estimator of $\boldsymbol{\theta}_1$, define $\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \boldsymbol{\eta}_k)$ as the estimator of $\boldsymbol{\theta}_k$ given $\boldsymbol{\eta}_k$ for $k = 2, \ldots, K$, define $\widehat{\boldsymbol{\theta}}_k{}'(\mathbf{Y}, \boldsymbol{\eta}_k)$—if it exists—as its Jacobian matrix (derivative with respect to $\boldsymbol{\eta}_k$) and define $\widehat{\boldsymbol{\eta}}_k(\mathbf{Y})$ as the estimator of $\boldsymbol{\eta}_k$. Furthermore, let n_k quantify the information contained in $\mathbf{Y}$ relevant to the estimation of $\boldsymbol{\theta}_k$. For the diagonal VARMA model $n_k = T$ for k from 1 to S and $n_K = S$. The SMLE consistency theorem provides the conditions under which $\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_k(\mathbf{Y}))$ is consistent for $k = 2, \ldots, K$.

Theorem 3.1 (SMLE Consistency)**.** *Suppose that*

$$\widehat{\boldsymbol{\theta}}_1(\mathbf{Y}) \xrightarrow{P} \boldsymbol{\theta}_1^* \quad as \quad n_1 \to \infty, \tag{3.5}$$

$$\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \boldsymbol{\eta}_k^*) \xrightarrow{P} \boldsymbol{\theta}_k^* \quad as \quad n_k \to \infty, \tag{3.6}$$

for $k = 2, \ldots, K$ where $\boldsymbol{\eta}_k^*$ and $\boldsymbol{\theta}_k^*$ are the true parameter sets for all k and P implies convergence in probability. Furthermore, assume that there exists a $n_{k0} < \infty$ such that for all $n_k > n_{k0}$, $\widehat{\boldsymbol{\theta}}_k{}'(\mathbf{Y}, \boldsymbol{\eta}_k)$ exists and is uniformly bounded in an open neighborhood of $\boldsymbol{\eta}_k^*$ almost surely for $k = 2, \ldots, K$. Then

$$\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_k(\mathbf{Y})) \xrightarrow{P} \boldsymbol{\theta}_k^* \quad as \quad n_1, \ldots, n_k \to \infty,$$

for $k = 2, \ldots, K$.

The assumptions of the SMLE consistency theorem are required for the Spall consistency theorem (Spall, 1989, Theorem 1) used in the inductive hypothesis of the proof provided

in Appendix A.2. The contribution to Theorem 3.5 is the inductive structure allowing the Spall consistency theorem to be applied to multiple steps. Heuristically, for step $k = 2, \ldots, K$ the theorem states that if $\widehat{\boldsymbol{\eta}}_k$ is a consistent estimator, $\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \boldsymbol{\eta}_k^*)$ is a consistent estimator and $\widehat{\boldsymbol{\theta}}_k{}'(\mathbf{Y}, \boldsymbol{\eta}_k)$ exists and is well-behaved near $\boldsymbol{\eta}_k^*$, then $\widehat{\boldsymbol{\theta}}_k(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_k)$ is a consistent estimator. Section 3.5 provides a simulation study to explain the key results of Theorem (3.5).

3.4.2 Diagonal VARMA Estimation

As $\boldsymbol{\eta}_k = \varnothing$ for k from 1 to S, the parameters $\boldsymbol{\theta}_k$ for k from 1 to S can be estimated in parallel with the ARMA likelihood functions in one stage. In the second stage, the correlation parameters are estimated conditional on the estimates of the mean, standard deviation, AR and MA parameters (3.4). The conditioning can be represented through the transformation

$$\mathbf{u}_t \coloneqq \Sigma^{-1} \left(I + \sum_{i=1}^{q} \Pi_i \mathrm{B}^i \right)^{-1} \left(I - \sum_{i=1}^{p} \Phi_i \mathrm{B}^i \right) (\mathbf{y}_t - \boldsymbol{\mu}_t) \quad \text{where} \quad \boldsymbol{\mu}_t \coloneqq \sum_{i=0}^{d} \boldsymbol{\beta}_i z_{t,i} \tag{3.7}$$

and B is the vector back-shift operator such that $B\mathbf{y}_t = \mathbf{y}_{t-1}$ and $B\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1}$, see Appendix A.3 for a proof. This residual transformation is very important for reducing the computation and memory costs of the second SMLE stage. This is since the step S estimated likelihood function has an innovation form

$$\widehat{\mathcal{L}}_{S+1}(\boldsymbol{\theta}_{S+1} \mid \mathbf{y}) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{0}, \mathbf{R} \mid \widehat{\mathbf{u}}_t)$$

where $\widehat{\mathbf{u}}_t$ is the residual transformation (3.7) with respect to the estimated parameters. As a consequence of this innovation form, $\widehat{\mathbf{u}}_t$ is only calculated once for all the evaluations. Furthermore, it is now as if there are T realisations of the innovations. This allows for nonstationary spatial specifications. The temporal parameter estimates obtained in the first stage are consistent in T (Hamilton, 1994, Section 5.8) and satisfy assumptions (3.5)

and (3.6) of the SMLE consistency theorem (3.1) where $n_k = T$ for k from 1 to S. The simulation study in Section 3.5 suggests that the spatial parameters of a stationary Matérn specification converge in probability as T and S approach infinity.

This two-stage approach has some advantages. First, ARMA model selections can be performed in the first stage in parallel, i.e. the AR and MA orders of the ARMA models can vary across space. Second, innovation model selection and specification can be performed in the second stage. Hence, a spatial specification can be informed by an exploratory analysis of the spatial residuals. Note that if the innovation model is a marginally parameterised model with a parameter set that can be estimated in s stages, then the parameter set of the diagonal VARMA model can be estimated in $s + 1$ stages.

3.5 Simulation Study

The first simulation study compares the estimated biases (difference between true and estimated parameter values) and standard errors of the maximum likelihood and stepwise maximum likelihood estimators for the diagonal VARMA model (3.2) introduced in Subsection 3.3.1 with isotropic innovations. The second simulation study is used to corroborate the SMLE consistency Theorem 3.1 introduced in Section 3.4.1 for the same model.

3.5.1 Simulation Model

The diagonal VARMA model (3.2) is used with zero mean, AR order two and MA order zero, i.e. centered diagonal VAR(2), with stationary innovations. The number of parameters is restricted so that the SMLE method can be directly compared with the MLE method where the number of evaluations grows exponentially with the number of param-

eters. For all T and S considered in this section define the standard deviation and AR parameters as $\sigma_{\mathbf{x}_s} = 1.2$, $\phi_{1,\mathbf{x}_s} = 0.50$ and $\phi_{2,\mathbf{x}_s} = 0.25$ for all s. The stationary innovations are modelled with the Matérn correlation function, which for distance h has the following form:

$$\frac{\pi^{1/2}}{2^{\kappa-1}\Gamma(\kappa+1/2)\alpha^{2\kappa}}\,(\alpha|h|)^{\kappa}\,K_{\kappa}(\alpha|h|),$$

where the inverse scale parameter $\alpha > 0$ controls the range of correlation, the smoothness parameter $\kappa > 0$ controls the mean-square differentiability of the process and $K_{\kappa}(\cdot)$ is a modified Bessel function (Stein, 2012, p. 31). In this section we define the inverse scale and smoothness parameters as $\alpha = 0.8$ and $\kappa = 1.5$ respectively. This model corresponds to $N = T \cdot S$ data points, where spatial locations are distributed regularly on a line, and $P = 3 \cdot S + 2$ parameters, i.e. three temporal parameters for each time series and two spatial parameters.

3.5.2 Set-up

Since, for MLE, the number of evaluations grows exponentially with the number of parameters and $P = 3 \cdot S + 2$, the number of spatial locations S is restricted so that the SMLE method can be directly compared to the MLE method. Furthermore, since the computational cost of evaluating S AR(2) likelihood functions and T Gaussian likelihood functions in parallel for SMLE is $\mathcal{O}(T + S^3)$, the diagonal VAR(2) likelihood function is evaluated in parallel with identical computational cost for direct comparison. As a consequence, the difference in the computational cost between MLE and SMLE is from the number of iterations. Therefore, for the purpose of comparison, we let $T = 50$ and $S = 20$ (so that $N = 50 \cdot 20 = 1,000$ and $P = 3 \cdot 20 + 2 = 62$). For both methods the algorithm is initialized at the true parameter values to eliminate the effects of initial value selection and standard errors are obtained by performing both the MLE and SMLE methods for 30 independent simulations. For MLE and SMLE numerical optimization is performed in `R` with the Nelder-Mead simplex algorithm (Nelder and Mead, 1965). Computation

is performed on a 2.6 GHz Intel Core i7 (8 virtual cores) CPU and 8 GB of RAM and parallel computation is performed via multi-threading.

3.5.3 Results

Table 3.2 displays the mean estimates (averaged over s) and standard errors, in parenthesis, of $\sigma_{\mathbf{x}_s}$, $\phi_{1,\mathbf{x}_s}$, $\phi_{2,\mathbf{x}_s}$, α and κ obtained from SMLE, Full MLE and Fixed MLE ($\sigma_{\mathbf{x}_s}$ fixed for all s) using 30 independent simulations. The estimates from Full MLE suggest that $\sigma_{\mathbf{x}_s}$ for all s are difficult to identify in the diagonal VARMA model. Variations on this simulation study suggest that the standard deviation parameters are difficult to identify when α is small. Table 3.3 demonstrates that for $\alpha = 0.3$ the estimates of the standard deviation parameters are more biased. As a consequence, Fixed MLE was included to provide a comparison with SMLE. The mean estimates and standard errors obtained from Fixed MLE aim to approximate those from Full MLE. The approximated estimates are expected to be more statistically efficient as fixing the standard deviation parameters has increased the data to parameter ratio. There is a relatively small difference in the sample biases of $\widehat{\phi}_{1,\mathbf{x}_s}$ and $\widehat{\phi}_{2,\mathbf{x}_s}$ between SMLE and Fixed MLE, however, the 57% relative efficiencies (ratio of standard errors) demonstrate that the estimates from SMLE are less efficient, as expected. The sample biases of $\widehat{\alpha}$ and $\widehat{\kappa}$ are relatively small with relative efficiencies 67% and 74% respectively. Considering that these are conservative estimates and the number of data points is small ($N = 1,000$) there appear to be relatively estimated biases. The fact that SMLE was capable of estimating $\sigma_{\mathbf{x}_s}$ for all s demonstrates one advantage of estimating with marginal likelihood functions.

Table 3.4 displays the mean time (seconds) and mean number of iterations required by both estimation methods and the two steps of the SMLE method. On average MLE took approximately 9,200 times longer (47 minutes 42 seconds) and required approximately 740 times more iterations than SMLE.

Table 3.2: Mean estimates (averaged over s) and standard errors, in parenthesis, of $\sigma_{\mathbf{x}_s}$, $\phi_{1,\mathbf{x}_s}$, $\phi_{2,\mathbf{x}_s}$, α and κ obtained from SMLE, Full MLE and Fixed MLE using 30 independent simulations.

	$\widehat{\sigma}$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\alpha}$	$\widehat{\kappa}$
SMLE	1.17 (0.12)	0.49 (0.14)	0.22 (0.14)	0.86 (0.12)	1.51 (0.19)
Full MLE	6.54 (9.34)	0.13 (0.71)	-0.12 (0.54)	0.53 (0.43)	1.66 (1.20)
Fixed MLE	NA	0.49 (0.08)	0.24 (0.08)	0.86 (0.08)	1.63 (0.14)
True Values	1.20	0.50	0.25	0.80	1.50

Table 3.3: Mean estimates (averaged over s) and standard errors, in parenthesis, of $\sigma_{\mathbf{x}_s}$, $\phi_{1,\mathbf{x}_s}$, $\phi_{2,\mathbf{x}_s}$, α and κ obtained from SMLE, Full MLE and Fixed MLE using 30 independent simulations.

	$\widehat{\sigma}$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\alpha}$	$\widehat{\kappa}$
SMLE	1.16 (0.12)	0.51 (0.13)	0.21 (0.14)	0.33 (0.04)	1.54 (0.07)
Full MLE	95.13 (1746.27)	0.30 (0.59)	-0.03 (0.52)	0.15 (0.11)	1.27 (0.40)
Fixed MLE	NA	0.50 (0.02)	0.24 (0.02)	0.30 (0.02)	1.51 (0.04)
True Values	1.20	0.50	0.25	0.30	1.50

Estimators are expected to have zero bias and variance as T and S increase to infinity according to the consistency theorem. However, for clarity we provide plots of the bias, variance and MSE of estimators as T increases with S fixed then as S increases with T fixed. Figure 3.1 displays the bias (squared), variance and MSE profile plots of $\widehat{\alpha}$ and $\widehat{\kappa}$ from SMLE as the number of time points T increases to 100 with $S = 20$ fixed and then as the number of spatial locations S increases to 45 with $T = 100$ fixed. These plots provide numerical evidence of the SMLE consistency theorem in Subsection 3.4.1 as they

Table 3.4: Mean time (seconds) and mean number of iterations required by both estimating methods and the two steps of the SMLE method. Means and standard deviations, in parenthesis, are calculated from the 30 simulations.

	SMLE			**MLE**
	Step 1	Step 2	Total	Total
Time (sec)	0.10 (0.01)	0.21 (0.06)	0.31 (0.06)	2862.43 (5344.8)
# Iterations	135.87 (20.84)	66.6 (20.02)	202.47 (29.79)	149667.8 (36135.3)

demonstrate how the MSE decreases for $\widehat{\alpha}$ and $\widehat{\kappa}$ conditional on the temporal parameter estimates as T increases for a fixed number of spatial locations ($S = 20$) and as S increases for a fixed number of time points ($T = 100$). For both $\widehat{\alpha}$ and $\widehat{\kappa}$ the variance dominates the bias for all values of T. Furthermore, the plots demonstrate that the reductions in MSE are mostly attributable to reductions in variance with small bias even for a few time points and spatial locations. Similar results where obtained for higher AR orders and different AR parameter values.

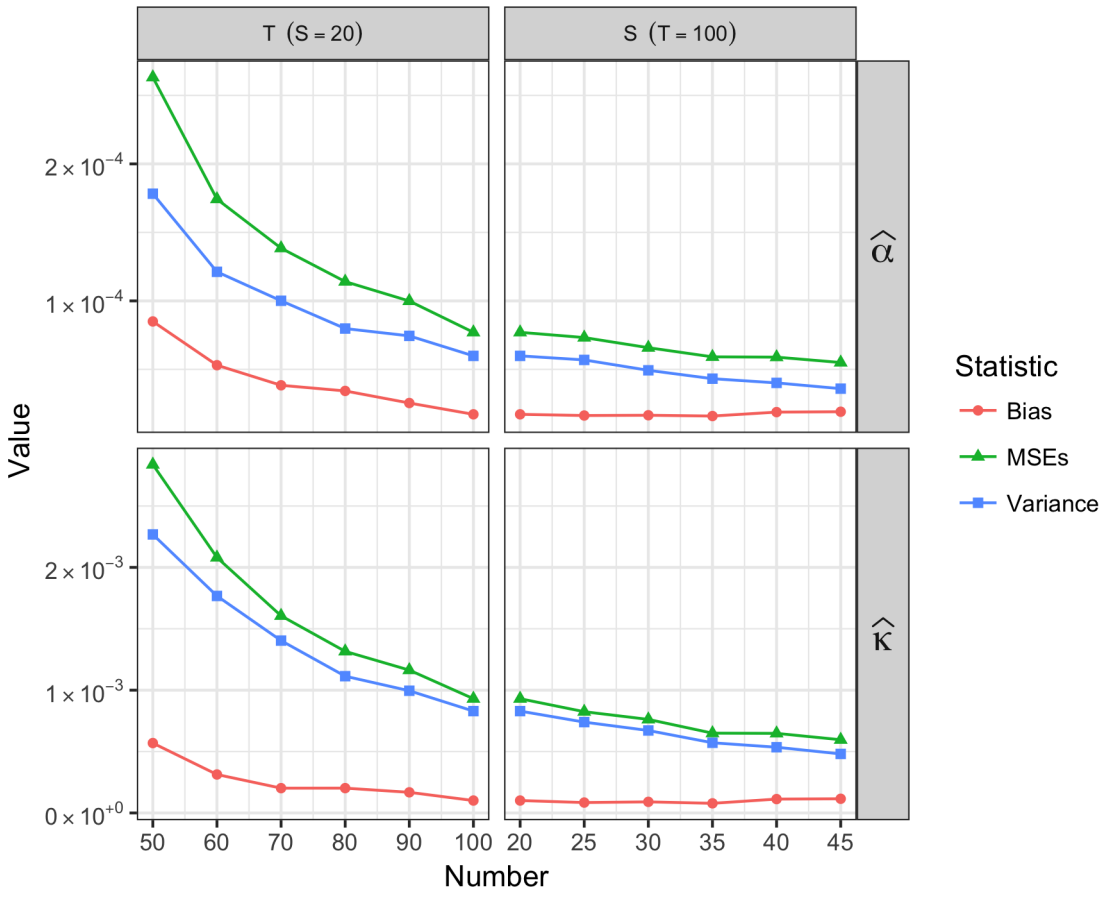


Figure 3.1: Bias (squared), variance and mean square error (MSE) of $\widehat{\alpha}$ and $\widehat{\kappa}$ from SMLE as the number of time points T increases to 100 with $S = 20$ fixed and then as the number of spatial locations S increases to 45 with $T = 100$ fixed.

Chapter 4

Modelling

In this chapter and Chapter 4 the parameters of a particular model are considered as elements of a parameter set rather than elements of a parameter vector. This is so that set operators and relations (e.g. union operator and subset relation) can be used to define the parameter sets of marginal models and their relations to the joint model. This is particularly useful in the context of marginally parameterised models that are defined with marginal models that correspond to data subsets, see Definition 1. As a consequence, a parameter set $\boldsymbol{\theta}$ consisting of P parameters is defined as $\{\theta_1, \ldots, \theta_P\}$.

4.1 Ensemble Model

Define $\mathbf{y}_r$ as the vectorisation of $\mathrm{Y}[r]$ for all r. These vectors corresponds to the members of the training ensemble. It is assumed that the training ensemble members $\mathbf{y}_r$ for all r are independent and identically distributed (IID) realisations of a stochastic model, hence the index r. This assumption depends on the chaotic nature of the non-linear dynamical system that generated the ensemble (i.e. global climate system) and its sensitivity to initial

conditions (see Berliner, 1992). The rate of divergence from initial conditions for the TS climate variable is discussed in Collins and Allen (2002), Collins (2002) and Branstator and Teng (2010). These discussions suggest that this assumption is appropriate for the TS climate variable. This assumption will be checked on the training ensemble in Chapter 5. It is also assumed that $\mathbf{y}_r$ for all r are realisations from a multivariate Gaussian stochastic model with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ for all r. Note that the climate variables were annually averaged to approximate this assumption through the central limit theorem. Again, this assumption will be checked on the training ensemble in Chapter 5.

As a consequence of the small training ensemble ($R = 5$), in contrast to Castruccio et al. (2013), Castruccio and Genton (2016) and Castruccio and Guinness (2017), we do not estimate $\boldsymbol{\mu}$ non-parametrically using a restricted likelihood (Patterson and Thompson, 1971). Instead, $\boldsymbol{\mu}$ is incorporated parametrically into the temporal model component specified in Subsection 4.2.

The remaining sections of this chapter will introduce the multivariate global spatio-temporal model, component by component. These model components correspond to SMLE stages. Hence, after each model component is specified, the corresponding sequence of data subsets, marginal models and primary parameter subsets are provided. Furthermore, the corresponding conditioning transformation is provided and the next model component is specified with respect to this transformation.

4.2 Temporal Model Component

Define $\mathbf{Y}_r$ as the random vector corresponding to $\mathbf{y}_r$. The random vectors $\mathbf{Y}_r$ for all r are modelled with a vector auto-regressive moving average model (Lütkepohl, 2005, Chapter 11) with diagonal auto-regressive and moving average matrices (DVARMA). The auto-regressive (AR) and moving average (MA) orders are p and q, respectively. The

DVARMA model is specified as

$$\mathbf{Y}_r[t] - \boldsymbol{\mu}_t = \Sigma \mathbf{U}_r[t] + \sum_{i=1}^{p} \Phi_i(\mathbf{Y}_r[t-i] - \boldsymbol{\mu}_{t-i}) + \sum_{i=1}^{q} \Pi_i \Sigma \mathbf{U}_r[t-i] \tag{4.1}$$

where

$$\boldsymbol{\mu}_t \coloneqq \sum_{i=0}^{d} \boldsymbol{\beta}_i z_{t,i},$$

$\boldsymbol{\beta}_i = (\beta_{i,l,m,v})$ are vectors of mean parameters, $z_{t,i}$ are regressor variables from a $T \times (d+1)$ regressor matrix Z, $\Sigma = \text{diag}(\sigma_{l,m,v})$ is the diagonal matrix of standard deviation parameters, $\Phi_i \coloneqq \text{diag}(\phi_{i,l,m,v})$ are the diagonal matrices of AR parameters, $\Pi_i \coloneqq \text{diag}(\pi_{i,l,m,v})$ are the diagonal matrices of MA parameters. Furthermore, $\mathbf{U}_r[t]$ are IID zero-mean and unit-variance multivariate-spatial Gaussian innovations with correlation matrix $\mathbf{R}$. The regressor matrix Z is the orthonormal matrix "Q" from the QR decomposition (Golub and Van Loan, 2012, Section 5.2) of the $T \times (d+1)$ Vandermonde matrix (Golub and Van Loan, 2012, Subsection 4.6)

$$V = \begin{bmatrix} 1 & 1-\bar{t} & (1-\bar{t})^2 & \dots & (1-\bar{t})^d \\ 1 & 2-\bar{t} & (2-\bar{t})^2 & \dots & (2-\bar{t})^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & T-\bar{t} & (T-\bar{t})^2 & \dots & (T-\bar{t})^d \end{bmatrix}. \tag{4.2}$$

where $\bar{t}$ is the mean year $(T+1)/2$. The columns of V are polynomials from order zero to d evaluated at $t-\bar{t}$ for all t. The columns of Z are the orthonormalised columns of V. The columns of V are orthonormalised so that the mean parameter estimators are uncorrelated and more efficient. Heuristically, the mean specification is designed to capture polynomial trends in time.

This DVARMA model can represent variation in mean, variance and temporal autocorrelation over space and climate variables. As described in Section 3.3.1, since the AR and MA parameters are spatially indexed, the spatio-temporal covariance function (3.3)

is non-separable. As a consequence, this model can implicitly capture cross-temporal dependence through the innovations.

The parameters of the temporal model component are estimated in the first SMLE stage. The corresponding sequence of data subsets are $\mathbf{y}[l, m, v]$ for all l, m and v. Each data subset is an $R \times T$ matrix consisting of R realisations of a time series of length T. Since, there is a data subset for each spatial location and climate variable there are a total of $288 \cdot 192 \cdot 3 = 165,888$ of them. The marginal likelihood function that corresponds to each of these data subsets is a product of R identical ARMA likelihood functions. This follows directly from the proof in Appendix A.1. The computational and memory cost of evaluating each ARMA likelihood function is $\mathcal{O}(T)$ and $\mathcal{O}(T)$, respectively. Furthermore, the primary and nuisance parameter subsets corresponding to the marginal likelihood function of $\mathbf{y}[l, m, v]$ are

$$\boldsymbol{\theta}_{l,m,v} \coloneqq \left(\bigcup_{i=1}^{d}\{\beta_{i,l,m,v}\}\right) \cup \left(\bigcup_{i=1}^{p}\{\phi_{i,l,m,v}\}\right) \cup \left(\bigcup_{i=1}^{q}\{\pi_{i,l,m,v}\}\right) \cup \{\sigma_{l,m,v}\} \quad \text{and} \quad \boldsymbol{\eta}_{l,m,v} \coloneqq \varnothing, \tag{4.3}$$

respectively. The primary parameter subsets include the mean, AR, MA and standard deviation parameters; and the nuisance parameter subsets are empty. Since, these primary parameter subsets are mutually disjoint they can be estimated in parallel. Let $\boldsymbol{\theta}^1$ and $\boldsymbol{\eta}^1$ denote all the SMLE stage one primary and nuisance parameter sets, respectively. After stage one the conditioning transformation is the residual transformation (3.7) where $\mathbf{u}_t$ is replaced with $\mathbf{u}_r[t]$.

4.3 Innovation Assumption

Let $\mathbf{U}_{r,t} \coloneqq \mathbf{U}_r[t]$ denote the IID zero-mean and unit-variance multivariate-spatial innovation for member r and year t. It is assumed that

$$\text{corr}(\mathbf{U}_{r,t}[l+h, m_1, v_1], \mathbf{U}_{r,t}[l, m_2, v_2]) = C(h, m_1, m_2, v_1, v_2),$$

where h is the longitudinal lag and $C(\cdot)$ is a positive definite multivariate-spatial cross-correlation function. This assumption is a multivariate extension of axial symmetry (Jones, 1963). This extension is the multivariate analogue to the altitudinal extension introduced in Castruccio and Genton (2016). Given the multivariate axial symmetry assumption, all the longitudinal bands are jointly stationary. Let $\mathbf{R}_{m_1,m_2,v_1,v_2}$ denote the $L \times L$ cross-correlation matrix between the longitudinal bands $\mathbf{U}_{r,t}[m_1, v_1]$ and $\mathbf{U}_{r,t}[m_2, v_2]$. Since, the longitudinal bands are circular and jointly stationary the cross-correlation matrices are circulant (Davis, 2012). A $k \times k$ matrix is circulant if it has the form

$$\mathbf{C} = \begin{bmatrix} c_0 & c_{k-1} & c_{k-2} & \dots & c_1 \\ c_1 & c_0 & c_{k-1} & \dots & c_2 \\ c_2 & c_1 & c_0 & \dots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{k-1} & c_{k-2} & c_{k-3} & \dots & c_0 \end{bmatrix}.$$

and is fully specified by its first column $\mathbf{c} = (c_0, c_1, c_3, \dots, c_{k-1})$. Circulant matrices have an eigenvalue decomposition in terms of a discrete Fourier transform (DFT) matrix. This is one of the reasons a circulant matrix is often defined with respect to a vector $\mathbf{c}$ indexed from zero. In terms of the cross-correlation matrix $\mathbf{R}_{m_1,m_2,v_1,v_2}$ the eigenvalue decomposition is

$$\mathbf{R}_{m_1,m_2,v_1,v_2} = \mathrm{W}^{-1}\text{diag}(\mathrm{W}\mathbf{r}_{m_1,m_2,v_1,v_2})\mathrm{W}, \tag{4.4}$$

where W is the $L \times L$ DFT matrix, with element $\exp(-2\pi ijk/L)$ in row j and column k where $i \coloneqq \sqrt{-1}$, and $\mathbf{r}_{m_1,m_2,v_1,v_2}$ is the first column of $\mathbf{R}_{m_1,m_2,v_1,v_2}$ (Davis, 2012, Section 3.2). Since the cross-correlation function corresponding to $\mathbf{R}_{m_1,m_2,v_1,v_2}$ is defined as $R_{m_1,m_2,v_1,v_2}(h) \coloneqq C(h, m_1, m_2, v_1, v_2)$, if follows that $\mathbf{r}_{m_1,m_2,v_1,v_2}$ defines the cross-correlation function. Furthermore, since the cross-spectral mass function corresponding to $\mathbf{R}_{m_1,m_2,v_1,v_2}$ is defined as

$$f_{m_1,m_2,v_1,v_2}(c) \coloneqq \sum_{h=0}^{L-1} R_{m_1,m_2,v_1,v_2}(h) \exp\left(-2\pi ich/L\right),$$

if follows that $\mathrm{W}\mathbf{r}_{m_1,m_2,v_1,v_2}$ defines the cross-spectral mass function. Analogously to how a covariance function can be decomposed into two variance functions and a correlation function, the cross-spectral mass function can be decomposed as

$$f_{m_1,m_2,v_1,v_2} = f_{m_1,v_1}^{1/2} \cdot f_{m_2,v_2}^{1/2} \cdot \rho_{m_1,m_2,v_1,v_2} \cdot \exp(i\phi_{m_1,m_2,v_1,v_2}), \tag{4.5}$$

where $f_{m_1,v_1} \coloneqq f_{m_1,m_1,v_1,v_1}$ is a spectral mass function, ρ_{m_1,m_2,v_1,v_2} is a coherence mass function and ϕ_{m_1,m_2,v_1,v_2} is a phase mass function (Koopmans, 1995, pp. 137). The spectral mass function models the dependencies within longitudinal bands and the coherence mass function and phase mass function functions model the dependencies between longitudinal bands. Furthermore, analogously to correlation functions, $\rho_{m_1,m_2,v_1,v_2} \equiv 1$ and $\phi_{m_1,m_2,v_1,v_2} \equiv 0$ when $m_1 = m_2$ and $v_1 = v_2$. These functions, see Subsections 4.4 and 4.5, fully specify a model for the multivariate-spatial innovations given the multivariate axially-symmetric assumption. Note that the model is specified in the spectral domain as a result of the multivariate axial symmetry assumption that resulted in circulant cross-correlation matrices that resulted in cross-spectral mass functions. The coherence mass function can be used to specify latitudinal dependence for one variable (e.g. (Castruccio et al., 2013)) or to specify latitudinal and multivariate dependence for two or more variables. Furthermore, since the number of Fourier coefficients in $W\mathbf{r}_{m_1,m_2,v_1,v_2}$ is finite and the process is a linear combination of these Fourier coefficients the process is well defined

as long as the variance of all these Fourier coefficients is finite. Note that the requirements are considerably more complicated if the number of Fourier coefficients is countable or uncountable infinite.

4.4 Longitudinal Model

The longitudinal model is the model for dependencies within longitudinal bands. Hence the longitudinal model requires the specification of the correlation matrices, e.g. $\mathbf{R}_{m,m,v,v}$. Since $\rho_{m,m,v,v} \equiv 1$ and $\phi_{m,m,v,v} \equiv 0$ for $\mathbf{R}_{m,v} \coloneqq \mathbf{R}_{m,m,v,v}$ it is fully specified by the spectral mass function $f_{v,m}$. Furthermore, since the multivariate-spatial innovations are unit-variance

$$\sum_{c \in \mathbb{Z}_L} f_{m,v}(c) = L, \tag{4.6}$$

see Appendix A.4 for a proof, hence the spectral mass functions do not require scale parameters and only require specification up to a constant of proportionality. Since scale parameters are included in Castruccio et al. (2013), Castruccio and Genton (2016) and Castruccio and Guinness (2017), the relationship (4.6) resulted in a reduction in parameters. The Modified Matérn spectral mass function introduced in Castruccio et al. (2013) is a Matérn spectral mass function modified for circular processes (e.g. longitudinal bands). It assumes that there is a smooth spectrum transition at high frequencies. This assumption is appropriate for TS, however, it is not appropriate for TMQ or U10 since the corresponding mean periodograms do not always suggest a smooth spectral transition at high frequencies, see Figure 5.10 for examples. In order to weaken this assumption a γ-Modified Matérn spectral mass function is proposed, with a parameter that controls the spectrum transition at high frequencies. The γ-Modified Matérn is defined as

$$f_{m,v}(c) \propto \left\{\alpha_{m,v}^2 + \gamma_{m,v} A^2(c) + (1 - \gamma_{m,v}) B^2(c)\right\}^{-\kappa_{m,v} - 1/2} \tag{4.7}$$

where

$$A(c) \coloneqq 2\sin(\pi c/L), \quad B(c) \coloneqq 2 - |4c/L - 2|,$$

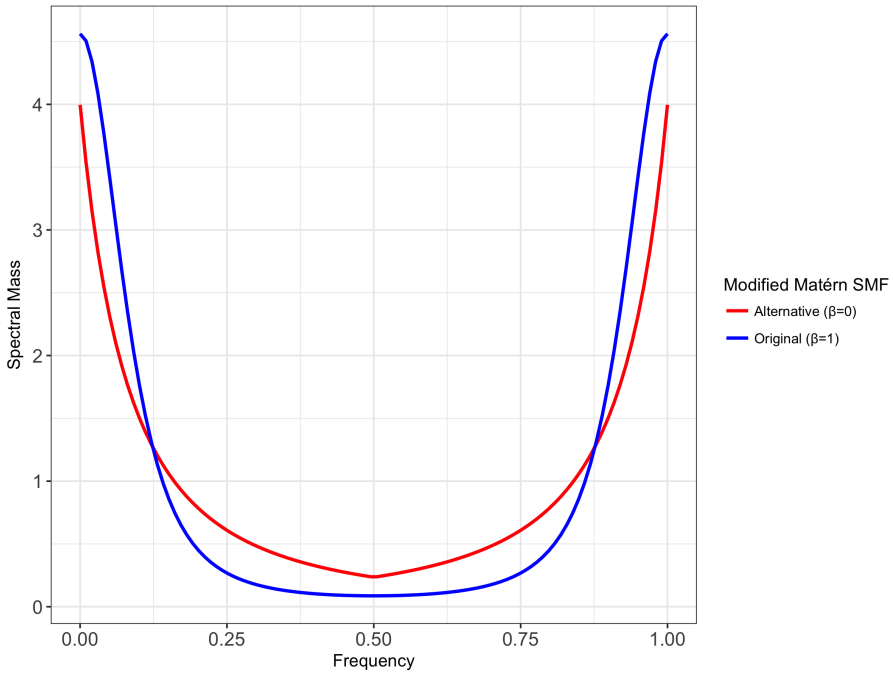
$\alpha_{m,v}$ is the inverse range parameter, $\gamma_{m,v}$ is the transition parameter and $\kappa_{m,v}$ is the smoothness parameter that controls the increased decay rate in spectral mass for larger wavenumbers, see Edwards et al. (2019). These parameters are allowed to vary over variables and latitudes. When $\gamma_{m,v} = 1$ the Modified Matérn is recovered and when $\gamma_{m,v} = 0$ an alternative Modified Matérn is recovered that does not assume that there is a smooth spectrum transition at high frequencies. Figure 4.1 displays the γ-Modified Matérn spectral mass function for $\alpha = 0.8$, $\gamma = 0, 1$ (i.e. alternative and original Modified Matérn spectral mass function) and $\kappa = 1.5$ where both are scaled to unit variance. At high frequencies (i.e. $1/2$), the original Modified Matérn spectral mass function is smooth and the alternative Modified Matérn spectral mass function is non differentiable. Note that these functions are symmetric around $1/2$ since they are the spectral mass functions for a circular process.

The parameters of the longitudinal model component are estimated in the second SMLE stage. The corresponding sequence of data subsets are $\mathbf{y}[m, v]$ for all m and v. Each data subset is an $R \times T \times L$ array consisting of $R \cdot T$ realisations of longitudinal bands of length L. Since, there is a data subset for each latitude and climate variable there are a total of $192 \cdot 3 = 576$ of them. The estimated marginal likelihood function for $\mathbf{y}[m, v]$ can be represented as

$$\widehat{\mathcal{L}}_{m,v}(\boldsymbol{\theta}_{m,v} \mid \mathbf{y}[m, v]) = \prod_{r=1}^{R} \prod_{t=1}^{T} \mathcal{N}(\mathbf{0}, \mathbf{R}_{m,v} \mid \widehat{\mathbf{u}}_{r,t}[m, v]) \tag{4.8}$$

where $\widehat{\mathbf{u}}_{r,t}$ represents the residual transformation (3.7) with respect to the estimated parameters. The estimated likelihood function $\mathcal{N}(\mathbf{0}, \mathbf{R}_{m,v} \mid \widehat{\mathbf{u}}_{r,t}[m, v])$ is an estimated Whittle likelihood function (Whittle, 1954) with spectral mass function (4.7), see Appendix A.5 for a proof. The computational and memory cost of evaluating each estimated Whittle likelihood function is $\mathcal{O}(L \ln L)$ and $\mathcal{O}(L)$, respectively. The primary and nuisance

Figure 4.1: The γ-Modified Matérn spectral mass function for $\alpha = 0.8$, $\gamma = 0, 1$ (i.e. alternative and original Modified Matérn spectral mass function) and $\kappa = 1.5$ where both are scaled to unit variance.



parameter subsets corresponding to the estimated Whittle likelihood function of $\mathbf{y}[m, v]$ are

$$\boldsymbol{\theta}_{m,v} \coloneqq \{\alpha_{m,v}\} \cup \{\gamma_{m,v}\} \cup \{\kappa_{m,v}\} \quad \text{and} \quad \boldsymbol{\eta}_{m,v} \coloneqq \left(\bigcup_{l=0}^{L-1} \boldsymbol{\theta}_{l,m,v} \right) \cup \left(\bigcup_{l=0}^{L-1} \boldsymbol{\eta}_{l,m,v} \right),$$

respectively. The primary parameter subsets include the inverse range, transition, and smoothness parameters; and the nuisance parameters correspond to those parameters used for the conditional transformation. Since these primary parameter subsets are mutually disjoint they can be estimated in parallel. Let $\boldsymbol{\theta}^2$ and $\boldsymbol{\eta}^2$ denote all the SMLE stage two primary and nuisance parameter sets, respectively.

4.5 Latitudinal and Multivariate Model

The latitudinal and multivariate model is the model for dependencies between longitudinal bands. Hence the latitudinal and multivariate model requires the specification of the cross-correlation matrices, e.g. $\mathbf{R}_{m_1,m_2,v_1,v_2}$. Since f_{m_1,v_1} and f_{m_2,v_2} are already specified in (4.7) only the coherence mass functions and phase mass functions require specification. The coherence mass functions and phase mass functions are specified implicitly through a dynamical model. The dynamical model allows for efficient simulation, see Chapter 5 for details. Let $\widetilde{\mathbf{U}}_{r,t}[m,v]$ denote the discrete Fourier transform of $\mathbf{U}_{r,t}[m,v]$ and let $\widetilde{\mathbf{V}}_{r,t}[m,v] \coloneqq \widetilde{\mathbf{U}}_{r,t}[m,v]\mathbf{F}_{m,v}^{-1/2}$ denote the standardised Fourier coefficients where $\mathbf{F}_{m,v} \coloneqq \text{diag}(f_{m,v}(c))$. The standardised Fourier coefficients are modelled with a diagonal vector AR (DVAR) model of order one (Edwards et al., 2019). The diagonal AR model of order one has a closed form solution for the coherence and phase in terms of the AR parameters, see below. However, closed form solutions for higher orders are unknown. The DVAR model is specified as

$$\widetilde{\mathbf{V}}_{r,t}[c,m] = \Psi_{c,m}\widetilde{\mathbf{V}}_{r,t}[c,m-1] + \mathbf{W}_{r,t}[c,m],$$

where $\widetilde{\mathbf{V}}_{r,t}[c,m]$ are vectors of standardised Fourier coefficients, $\Psi_{c,m} \coloneqq \text{diag}(\psi_{c,m,v})$ are the diagonal matrices of AR parameters and $\mathbf{W}_{r,t}[c,m]$ are zero-mean and unit-variance multivariate innovations with correlation matrices $\boldsymbol{\Xi}_{c,m}$. Under the constraint that

$$\boldsymbol{\Xi}_{c,m}[v_1,v_2] \coloneqq \boldsymbol{\Xi}_c[v_1,v_2] \cdot (1 - \psi_{c,m,v_1} \cdot \psi_{c,m,v_2}) \quad \text{for} \quad m > 1,$$

and $\boldsymbol{\Xi}_{c,1} \coloneqq \boldsymbol{\Xi}_c$, the cross-spectral mass functions are

$$f_{m_1,m_2,v_1,v_2}(c) = f_{m_1,v_1}^{1/2}(c) \cdot f_{m_2,v_2}^{1/2}(c) \cdot \boldsymbol{\Xi}_c[v_1,v_2] \cdot \left(\prod_{j=m_1+1}^{m_2} \psi_{c,j,v_2} \right), \tag{4.9}$$

for $m_1 < m_2$, m_1 and m_2 are exchanged for $m_2 < m_1$, see Appendix A.6 for a proof. For $m_1 = m_2$ the purely multivariate dependence is controlled by $\boldsymbol{\Xi}_c$ only and for $v_1 = v_2$ the purely latitudinal dependence is controlled by the AR parameters only. Since, the AR parameters are variable indexed, the cross-spectral mass function (4.9) is non-separable. Furthermore, the purely latitudinal dependence is the same as that in Castruccio and Guinness (2017). Following Castruccio and Guinness (2017) the AR parameters are parametrized as

$$\psi_{c,m,v} = \delta_{m,v} \left(1 + 4 \sin^2 \frac{\pi c}{L}\right)^{-\tau_{m,v}}, \tag{4.10}$$

where the scale parameter $\delta_{v,m}$ controls the rate of decay in coherence, over all wavenumbers, as the distance between latitudes increases and the smoothness parameter $\tau_{v,m}$ controls the increased decay rate in coherence for larger wavenumbers. These parameters are allowed to vary over variables and latitudes.

The multivariate model is fully specified by $\boldsymbol{\Xi}_c$. Since the mean cross-periodograms between the climate variables are not easily captured with a parametric model (see Figure 4.2), for each climate variable the moduli and arguments of $\boldsymbol{\Xi}_c[v_1, v_2]$ are modelled with a natural cubic spline (Friedman et al., 2001, Subsection 5.2.1) over the wavenumbers. Note that $\boldsymbol{\Xi}_c[v_1, v_2] = \boldsymbol{\Xi}_c[v_2, v_1] = 0$ implies that the climate variables v_1 and v_2 are independent. The parameters of the latitudinal model component are estimated in the third SMLE stage. The corresponding sequence of data subsets are $\mathbf{y}[v]$ for all v. Each data subset is an $R \times T \times L \times M$ array consisting of $R \cdot T$ realisations of $L \times M$ spatial fields. Since there is a data subset for each climate variable there are a total of three of them. The estimated marginal likelihood function for $\mathbf{y}[v]$ can be represented as

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}_v \mid \mathbf{y}[v]) = \prod_{r=1}^{R} \prod_{t=1}^{T} \mathcal{N}(\mathbf{0}, \mathbf{R}_v \mid \widehat{\mathbf{u}}_{r,t}[v])$$

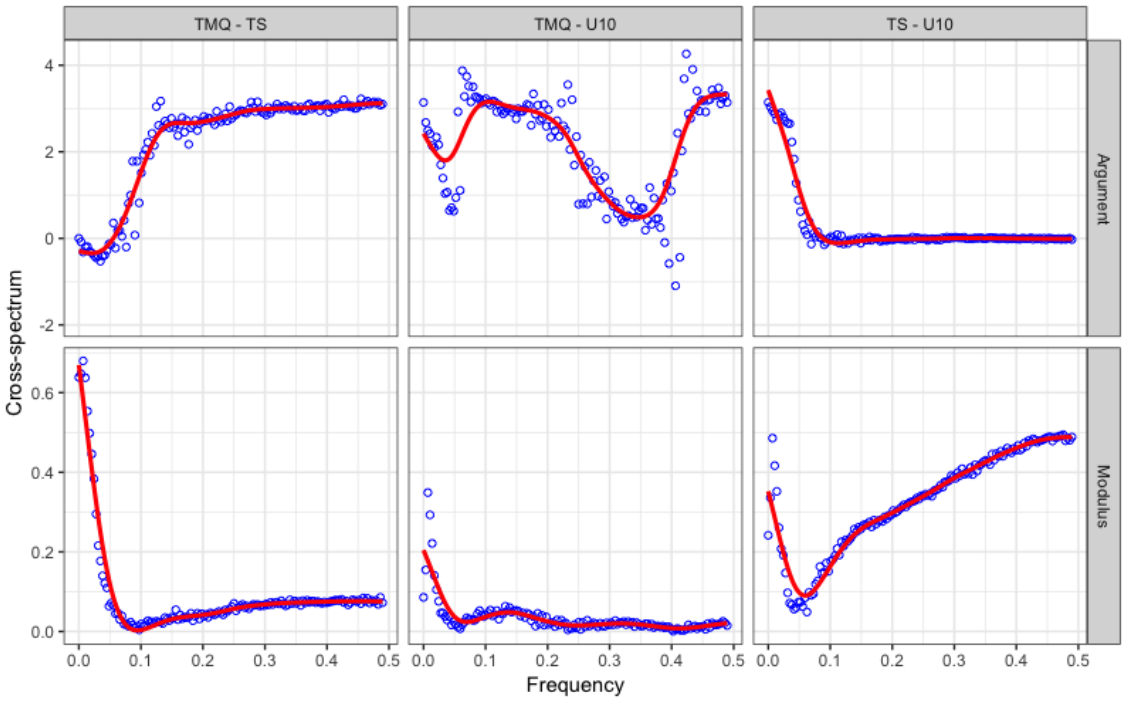


Figure 4.2: The moduli and arguments of mean cross-periodograms (blue) and the natural cubic splines (red) between the TMQ and TS climate variables, the TMQ and U10 climate variables and the TS and U10 climate variables over frequency.

where

$$\mathbf{R}_v = \begin{bmatrix} \mathbf{R}_{1,1,v,v} & \dots & \mathbf{R}_{1,M,v,v} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{M,1,v,v} & \dots & \mathbf{R}_{M,M,v,v} \end{bmatrix}$$

is a $M \times M$ block circulant matrix and $\widehat{\mathbf{u}}_{r,t}$ represents the residual transformation (3.7) with respect to the estimated parameters. Since the estimated likelihood function $\mathcal{N}(\mathbf{0}, \mathbf{R}_v \mid \widehat{\mathbf{u}}_{r,t}[v])$ has a block circulant matrix it can be expressed as a product of L estimated multivariate Gaussian likelihood functions of dimension M, see Appendix A.7. Consequently, the computational and memory cost of evaluating the multivariate Gaussian likelihood functions is $\mathcal{O}(M^3)$ and $\mathcal{O}(M^2)$, respectively. The primary and nuisance parameter sub-

sets corresponding to the estimated likelihood function of $\mathbf{y}[v]$ are

$$\boldsymbol{\theta}_v \coloneqq \left(\bigcup_{m=1}^{M} \{\delta_{m,v}\} \right) \cup \left(\bigcup_{m=1}^{M} \{\tau_{m,v}\} \right) \quad \text{and} \quad \boldsymbol{\eta}_v \coloneqq \left(\bigcup_{m=1}^{M} \boldsymbol{\theta}_{m,v} \right) \cup \left(\bigcup_{m=1}^{M} \boldsymbol{\eta}_{m,v} \right)$$

respectively. The primary parameter subsets include the scale and smoothness parameters; and the nuisance parameters correspond to those parameters used for the conditional transformation. Since, these primary parameter subsets are mutually disjoint they can be estimated in parallel. Let $\boldsymbol{\theta}^3$ and $\boldsymbol{\eta}^3$ denote all the SMLE stage three primary and nuisance parameter sets, respectively.

4.6 Model Summary

The multivariate global spatio-temporal model is a VARMA model with multivariate-spatial innovations where the mean, standard deviation, AR and MA parameters vary over variables and space. Hence, the multivariate-spatial fields are non-stationary in mean and variance over variables and space. The multivariate-spatial innovations are axially symmetric (the longitudinal bands are jointly stationary). This implies that the spatial fields are stationary in correlation over longitudes but non-stationary in correlation over latitudes. One consequence of this is that the model can not distinguish between land and ocean. One approach to overcoming this is to use evolutionary spectra (Castruccio and Guinness, 2017). Land-ocean dependencies are very important, however, the diagnostics in Figure 6 suggest that these dependencies are primarily captured by the mean and standard deviation parameters, since the dependencies that remain in the residuals are primarily coastal and mountainous. Modelling these sparse coastal and mountainous regions with an evolutionary spectral approach could improve the model, however, an alternative model is suggested in Chapter 7. This suggested model is fully non-stationary, rather than piece-wise stationary, and would be more computationally efficient. However, the details of SMLE for this model require significantly more work. Due to the axially

symmetric assumption, the multivariate-spatial innovations are fully specified by spectral mass functions and coherence mass functions. The γ-Modified Matérn spectral mass function is introduced to provide additional flexibility where the assumption of a smooth transition at high frequencies is not appropriate. The coherence mass functions are defined implicitly through a complex diagonal VAR model over latitude. These coherence mass functions have two very important properties: the parameters that control the dependencies within variables can be estimated with each variable's data subset independently and the parameters that control the dependencies between pairs of variables can be estimated with each pair of variable's data subsets independently. These properties are the main reason that a multivariate global spatio-temporal model for three global spatio-temporal variables can be trained.

Chapter 5

Training

As demonstrated in Chapter 4 the multivariate global spatio-temporal model can be trained in four stages with SMLE. Before each SMLE stage, diagnostics are used to check model assumptions; during each SMLE stage, model selection and parameter estimation are performed; and after each SMLE stage, diagnostics are used to check the trained model. This chapter discusses each SMLE stage, regarding model selection, parameter estimation and diagnostics, in addition to the general details of performing SMLE.

5.1 Numerical Optimisation and Computation

Before discussing the specifics of each SMLE stage, first a general approach to performing SMLE is discussed. For each SMLE stage, a large number of marginal likelihood functions are numerically optimised. In order to ensure that near globally optimal solutions are obtained the optimisations require tuning. However, effectively tuning a large number of optimisations manually is very difficult. As a consequence, an approach is required that can produce reliable estimates for each SMLE stage without manual tuning. The first

component of the approach is the Nelder-Mead optimisation algorithm (Nelder and Mead, 1965). Although this optimisation algorithm is not the most efficient it is robust and can provide reliable estimates. The second component of the approach involves the initial values. The marginal likelihood functions are optimised twice: the first optimisations are performed with the initial values set to zero, the second optimisations are performed with the initial values set to the estimates of the first optimsations smoothed over spatial locations. This second component assumes that the parameters of the marginal models vary smoothly over space. The plots in Section 5.3, 5.4 and 5.5 suggest that this assumption is reasonable.

5.2 Ensemble Assumptions

Section 4.1 introduced two assumptions: first the realisations $\mathbf{y}_r$ for all r are multivariate Gaussian distributed, and second the realisations are independent, i.e. Gaussian, independent and identically distributed (IDD). To check the first assumption Gaussian quantile-quantile (QQ) plots are used. These Gaussian QQ plots check if the realisations are marginally Gaussian. Since a random vector can be marginally Gaussian but not jointly Gaussian, these Gaussian QQ plots cannot check the jointly Gaussian assumption. Figure 5.1 displays Gaussian QQ plots for the training ensemble TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at four random spatio-temporal locations. Only a random selection of Gaussian QQ plots are displayed since there are over $95 \times 288 \times 192 \times 3 \approx 15$ million of them (one for each time point, spatial location and climate variable). The Gaussian QQ plots suggest that for all the climate variables a marginally Gaussian assumption is reasonable.

Since, uncorrelated implies independent for Gaussian random variables and Figure 5.1 suggest that the realisations are marginally Gaussian, the marginally independent assumption requires a figure that can suggest that the realisations are marginally uncorrelated.

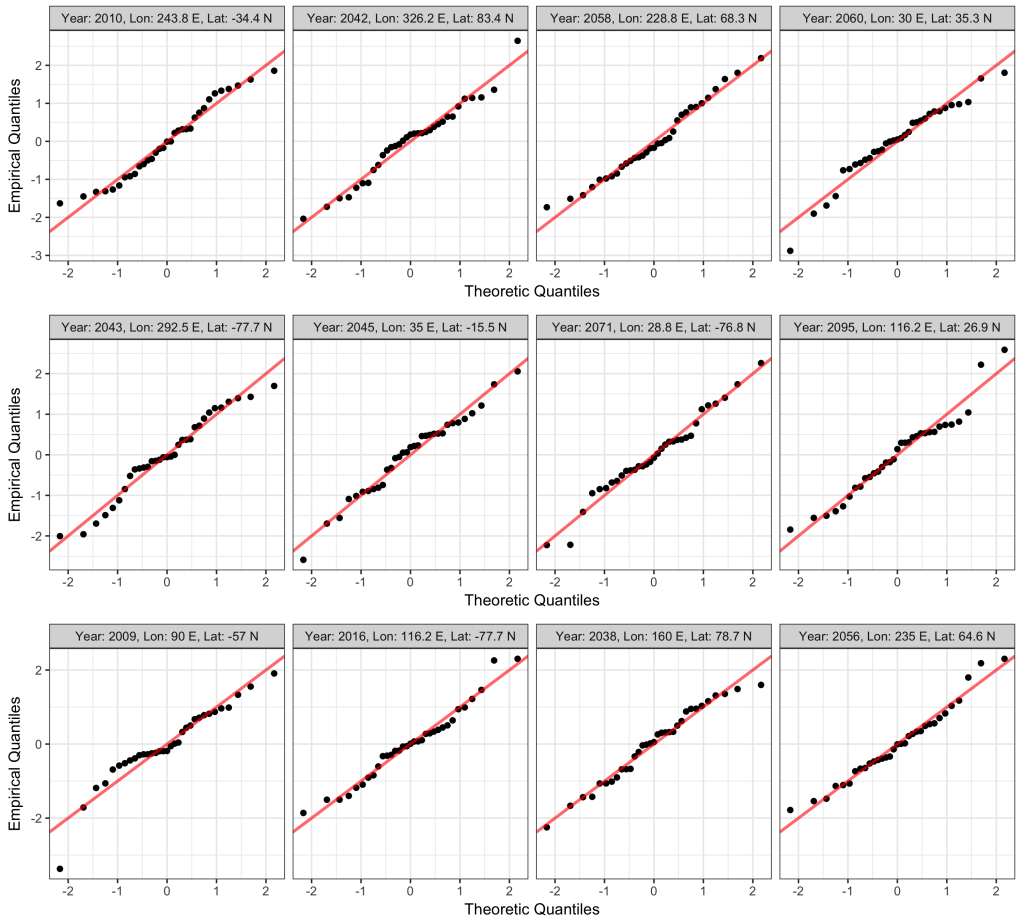


Figure 5.1: Gaussian quantile-quantile plots between the training ensemble members for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at four random spatio-temporal locations.

As the realisations are generated in sequence to form the ensemble, auto-correlation plots can be used to check if the realisations are marginally uncorrelated. Figure 5.2 displays auto-correlation plots for the training ensemble TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at the four random spatio-temporal locations. These plots suggest that all the climate variables are marginally uncorrelated and hence, independent under the marginally Gaussian assumption. Again only a random selection of auto-correlation plots are displayed since there are over 15 million of them.

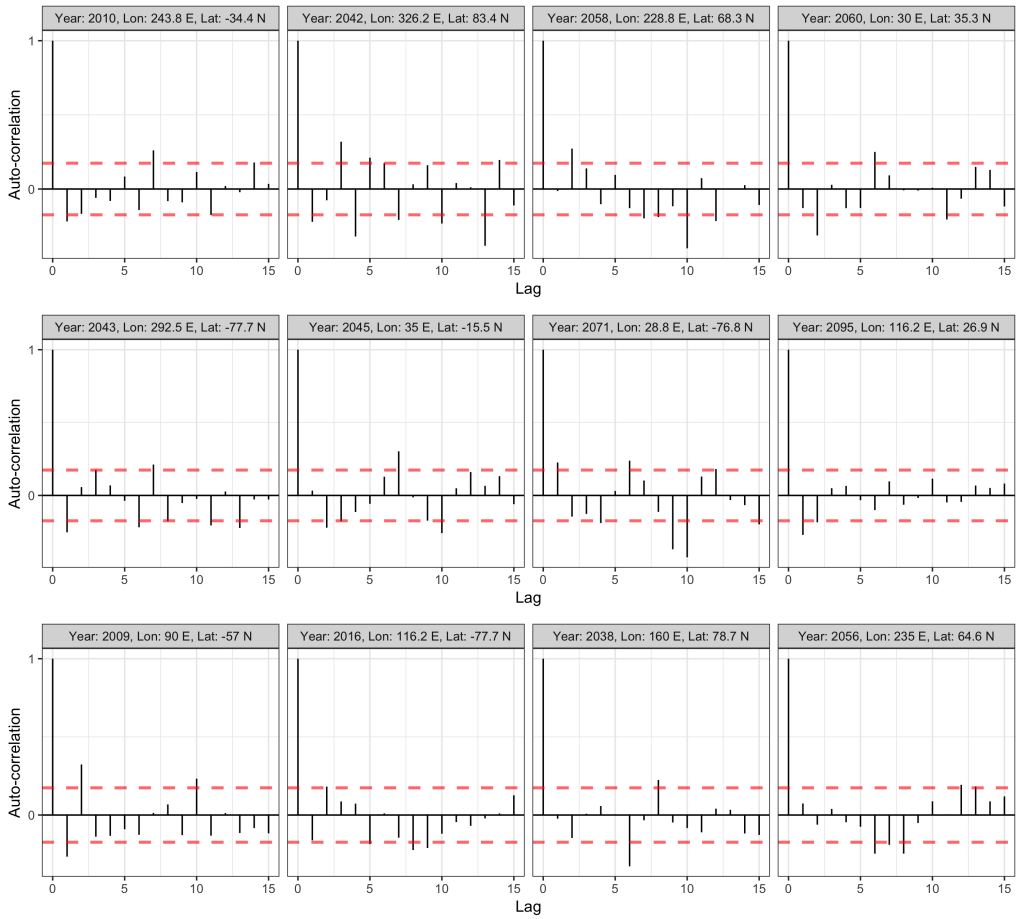


Figure 5.2: Auto-correlation plots between the training ensemble members for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at four random spatio-temporal locations.

5.3 Temporal Training

In the first SMLE stage $288 \times 192 \times 3 \approx 165$ thousand ARMA models are trained (one for each spatial location and climate variable). For each ARMA model the AR order p, the MA order q and the number of regressor variables d (excluding intercept) are selected with the Akaike information criterion (AIC). The number of regressor variables is equal to the order of the temporal trend. For example, two regressor variables corresponds to an order two quadratic temporal trend. The first time the ARMA models were trained,

almost all AR orders were between zero and three (inclusive) and almost all of the MA orders were zero. Furthermore, the spatial location of the non-zero MA orders appeared unstructured. As a consequence, all the subsequent ARMA models were trained with the restriction that $0 \leq p \leq 3$ and $q = 0$. Figure 5.3 displays the AR orders (0 - red, 1 - orange, 2 - light green, 3 - dark green) for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables. For each of the climate variables, the AR orders are generally larger towards the equator. This suggests that the range of temporal dependence is generally larger towards the equator. For the TMQ climate variable the ARMA models are almost exclusively white noise models towards the poles and for the TS climate variable the ARMA models are almost exclusively AR one models towards the north pole. There is spatial dependence between the AR orders for all three climate variables, however, this dependence is weaker for the U10 climate variable.

Figure 5.4 displays the number of regressor variables (1 - light green, 2 - dark green) for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables. Although there was no restriction on the number of regressor variables, d, only one or two regressor variables were ever selected. Note that light and dark green correspond to linear and quadratic temporal trends, respectively. These plots demonstrate that a linear or quadratic temporal trend is required for all the spatial locations and climate variables. The upper panel of Figure 5.4 suggests that the rate in change of TMQ is changing over time (quadratic trend) at most spatial locations. The middle panel of Figure 5.4 also suggests this for the TS climate variable, however, not for the north pole. The lower panel of Figure 5.4 suggests that there is generally a fixed rate of change in U10 over time (linear), however, the rate of change is changing in some regions.

The first SMLE stage involves the estimation of hundreds of thousands of parameters. For simplicity only estimates of the constant and linear temporal trend mean parameters, the standard deviation parameters and the AR order one parameters are considered. These parameter are particularly useful since they correspond to statistics introduced in

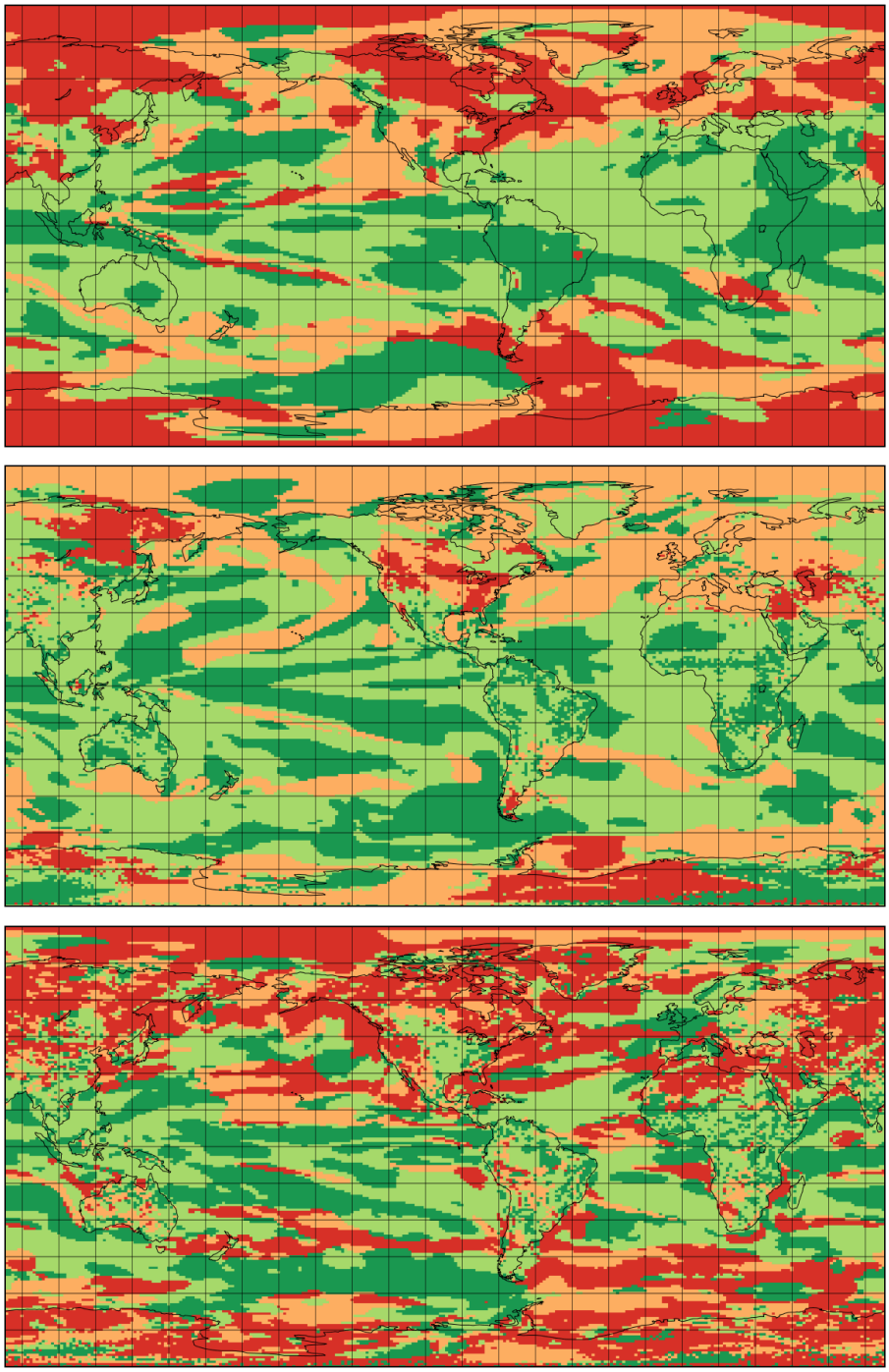


Figure 5.3: The AR orders of the ARMA models (0 - red, 1 - orange, 2 - light green, 3 - dark green) for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables.

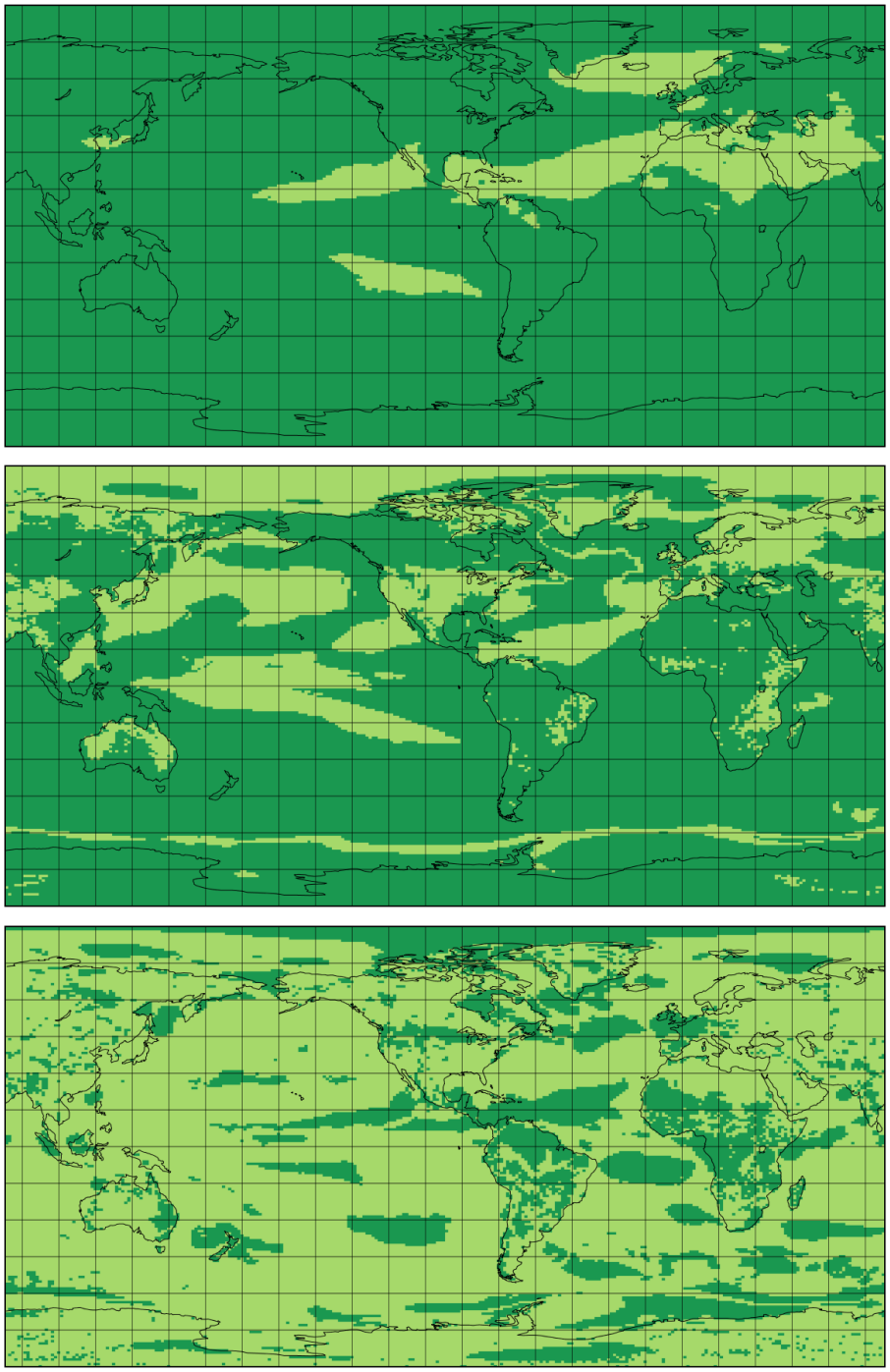


Figure 5.4: The number of regressor variables of the ARMA models (1 - light green, 2 - dark green) for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables.

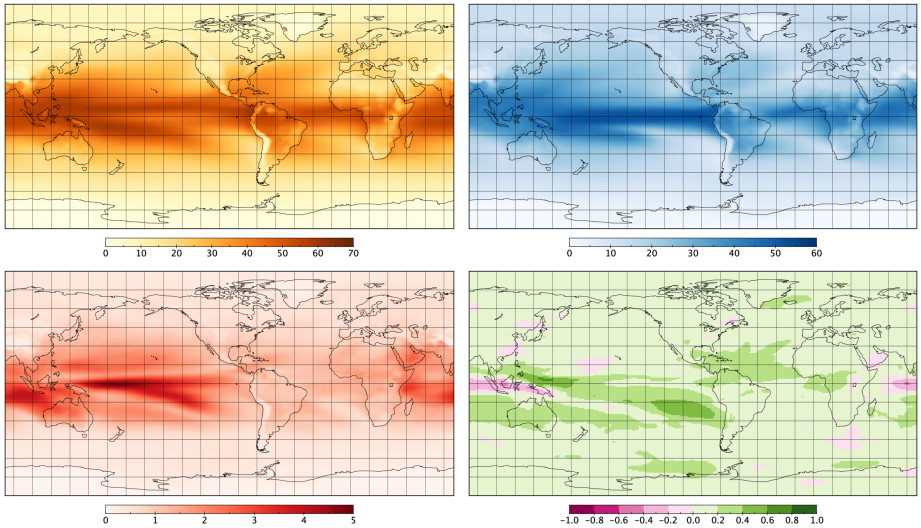


Figure 5.5: The estimated constant and linear trend mean parameters (upper left and right panels), the estimated standard deviation parameters (lower left panel) and the estimated AR order one parameters (lower right panel) for the TMQ climate variable.

Section 2.3 for exploratory data analysis. For example, the constant and linear trend mean parameters correspond to the simple linear regression model intercepts and slopes; and the standard deviation parameters and AR order one parameters correspond to the residual standard deviations and residual lag-one auto-correlations. As a consequence, the spatial structure of corresponding plots can be compared. Figures 5.5, 5.6 and 5.7 display the estimated constant and linear temporal trend mean parameters (upper left and right panels), the estimated standard deviation parameters (lower left panel) and the estimated AR order one parameters (lower right panel) for the TMQ, TS and U10 climate variables, respectively. Comparison with the statistics in Figures 2.3 and 2.4 demonstrate that the ARMA models are capable of capturing complex variations in mean, standard deviation and auto-correlation.

Figure 5.6 suggests that the diagonal VARMA model can capture some very complex phenomena important to climate scientists. The upper left panel demonstrates the model has captured the lower temperatures over Greenland and Antarctica. These lower tem-

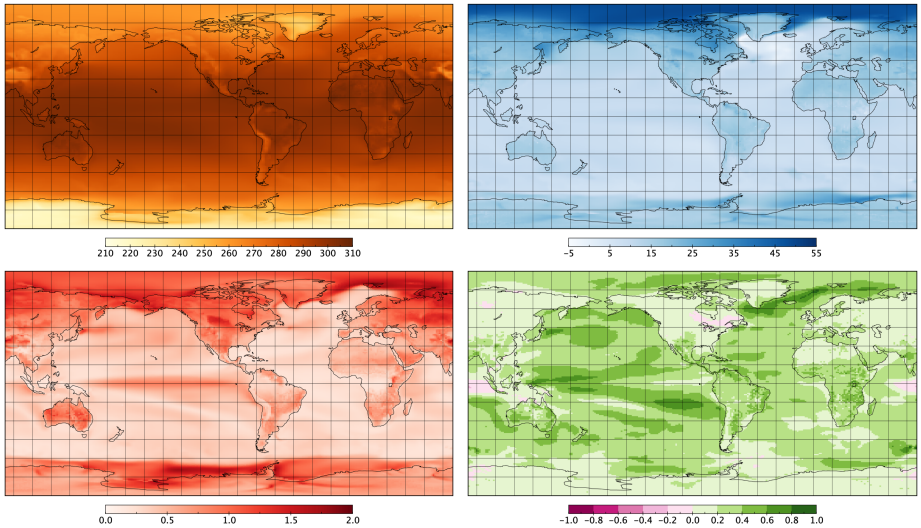


Figure 5.6: The estimated constant and linear trend mean parameters (upper left and right panels), the estimated standard deviation parameters (lower left panel) and the estimated AR order one parameters (lower right panel) for the TS climate variable.

peratures are the result of the ice and snow cover that reflect solar radiation. The upper right panel demonstrates the model has captured the extreme increase in temperature at the north pole due to global warming. The lower left panel demonstrates that the model has captured that temperature variations are larger over regions of sea-ice. The lower left panel demonstrates that the model has captured the strong temporal dependence of the east Greenland drift. Note that all of these phenomena can only be captured accurately if the constant and linear mean parameters, the standard deviation parameters and the AR order one parameters are not fixed over the globe. This is one of the substantial advantages of the specified diagonal VARMA model.

Auto-correlation plots of the ARMA model residuals are used to check if the ARMA models have captured all the temporal trends and dependence. They will also suggest whether or not the multivariate-spatial residuals are marginally independent. Checking if the multivariate-spatial residuals are jointly independent is substantially harder, since a very large number of cross-correlation plots would be required. Figure 5.8 displays the

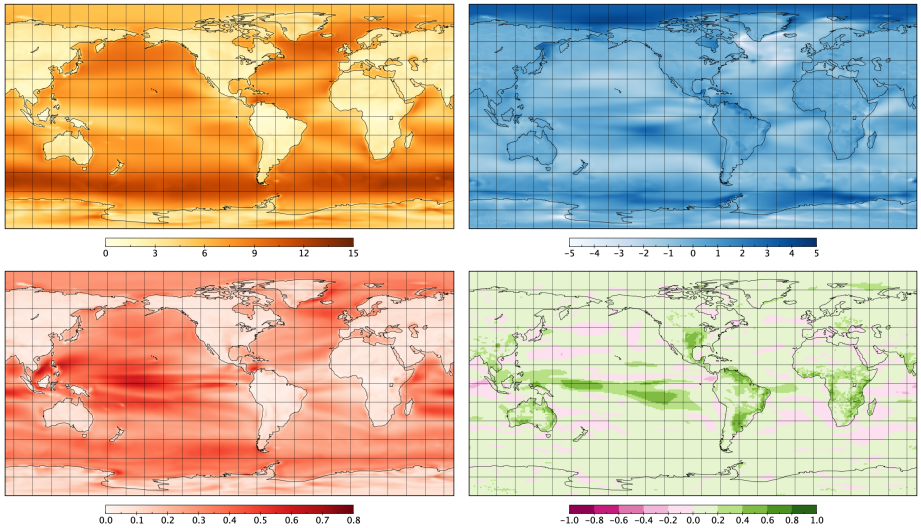


Figure 5.7: The estimated constant and linear trend mean parameters (upper left and right panels), the estimated standard deviation parameters (lower left panel) and the estimated AR order one parameters (lower right panel) for the U10 climate variable.

auto-correlation plots of the ARMA model residuals for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at the four random spatio-temporal locations. These plots suggest that the multivariate-spatial residuals are marginally independent.

5.4 Longitudinal Training

In the second SMLE stage $192 \times 3 \approx 600$ Whittle models are trained (one for each latitude and climate variable). For each Whittle model the β-Modified Matérn spectral mass function (4.7) is compared with the original Modified Matérn spectral mass function with the AIC. The percentage of Whittle models selected with the β-Modified Matérn spectral mass function (i.e. $\beta < 1$) are 67, 49 and 79 for the TMQ, TS and U10 climate variables, respectively. Figure 5.9 displays estimates of the parameters β, κ and $\log(\alpha)$ (blue circles)

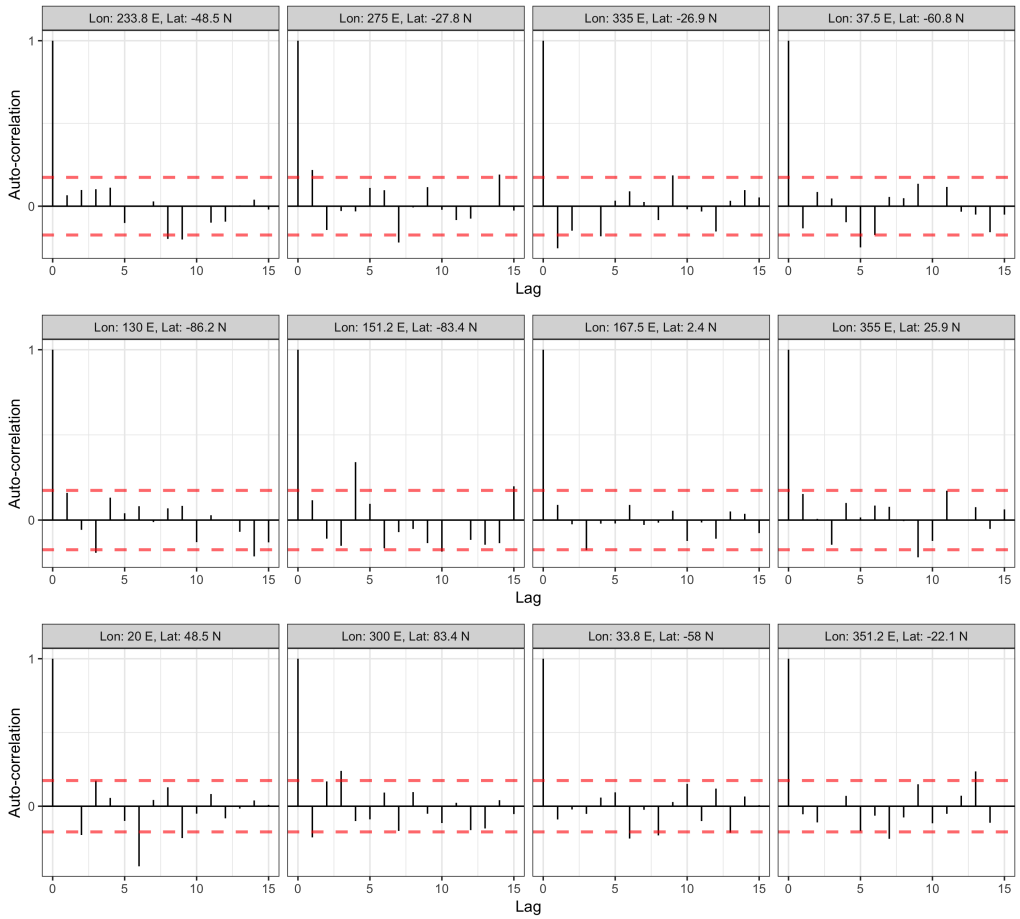


Figure 5.8: Auto-correlation plots of the ARMA model residuals for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables at four random spatio-temporal locations.

over latitude for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables along with loess smoothing (red curves). The left panels of Figure 5.9 suggest that values of $\beta < 0$ are important in the mid-latitudes for the TMQ climate variable, 60 degrees south for the TS climate variable and 70 degrees north for the U10 climate variable. The middle panels of Figure 5.9 demonstrate how the smoothness of the longitudinal bands vary over latitude and climate variables. Since a large value of κ corresponds to a smoother longitudinal band, the plots suggest that the longitudinal bands are smoother in the mid-latitudes for the TMQ climate variable, very smooth 60 degrees south for the TS climate variable and smoother 70 degrees south and 80 degrees north for the U10 climate variable.

Furthermore, the left panels of Figure 5.9 demonstrate that the range of dependence for all the climate variables decreases towards the poles. This is a result of the converging latitudinal bands towards the poles. Overall, Figure 5.9 suggests that the spectral mass functions have captured the stationary properties of the longitudinal bands over latitudes and climate variables.

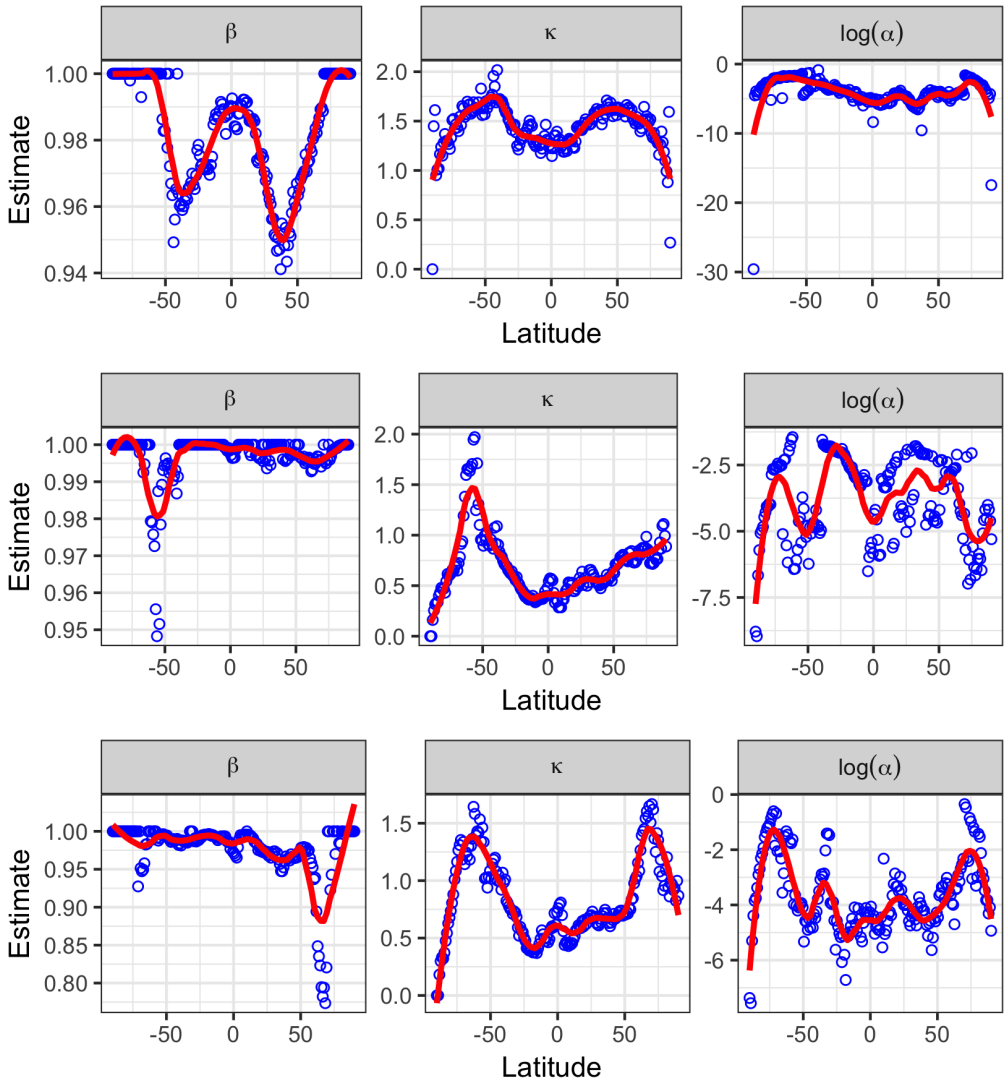


Figure 5.9: Estimates of the parameters β, κ and $\log(\alpha)$ (blue circles) with loess smoothing (red curves) over latitude for the TMQ (upper panel), TS (middle panel) and U10 (lower panel) climate variables.

Since the longitudinal and latitudinal models are specified with spectral mass func-

tions and cross-spectral mass functions, mean periodograms and mean cross-periodograms, i.e. empirical spectral and cross-spectral mass functions, can be used to gauge the quality of the trained models. Figure 5.10 displays mean periodograms and mean cross-periodograms (blue circles) with the trained spectral mass functions and cross-spectral mass functions (red lines) for four latitudinal bands at latitudes 50 degrees south, 10 degrees south, 30 degrees north and 60 degrees north. A spectral mass function displays the variance (auto-spectra) of all the different frequency signals that compose the process. The rate at which auto-spectra decays with frequency is related to the smoothness of the process. If a spectral mass function decays at a high rate, the process is composed of many low frequency (smooth) signals. The spectral mass functions (auto-spectra rows) in Figure 5.10 suggest that the fitted longitudinal model has captured the stationary properties of the longitudinal bands over latitudes and climate variables.

5.5 Latitudinal Training

In the third SMLE stage three complex valued VAR models are trained (one for each climate variable). For each complex valued VAR model non-stationarity is selected against stationarity ($\delta_{m_1,v} = \delta_{m_2,v}$ and $\tau_{m_1,v} = \tau_{m_2,v}$ for all $m_1 \neq m_2$). In all three cases the non-stationary complex valued VAR model is selected. Mean cross-periodogram plots of the residual longitudinal bands are used to check if the coherence mass functions have captured all the latitudinal stationary dependence. Note that mean cross-periodograms are the empirical versions of the cross-spectral mass functions. Figure 5.10 displays mean periodograms and mean cross-periodograms (blue circles) with the trained spectral mass functions and cross-spectral mass functions (red lines) for four latitudinal bands at latitudes 50 degrees south, 10 degrees south, 30 degrees north and 60 degrees north. These plots suggest that the complex valued non-stationary VAR models have accurately captured the latitudinal dependence given the assumptions.

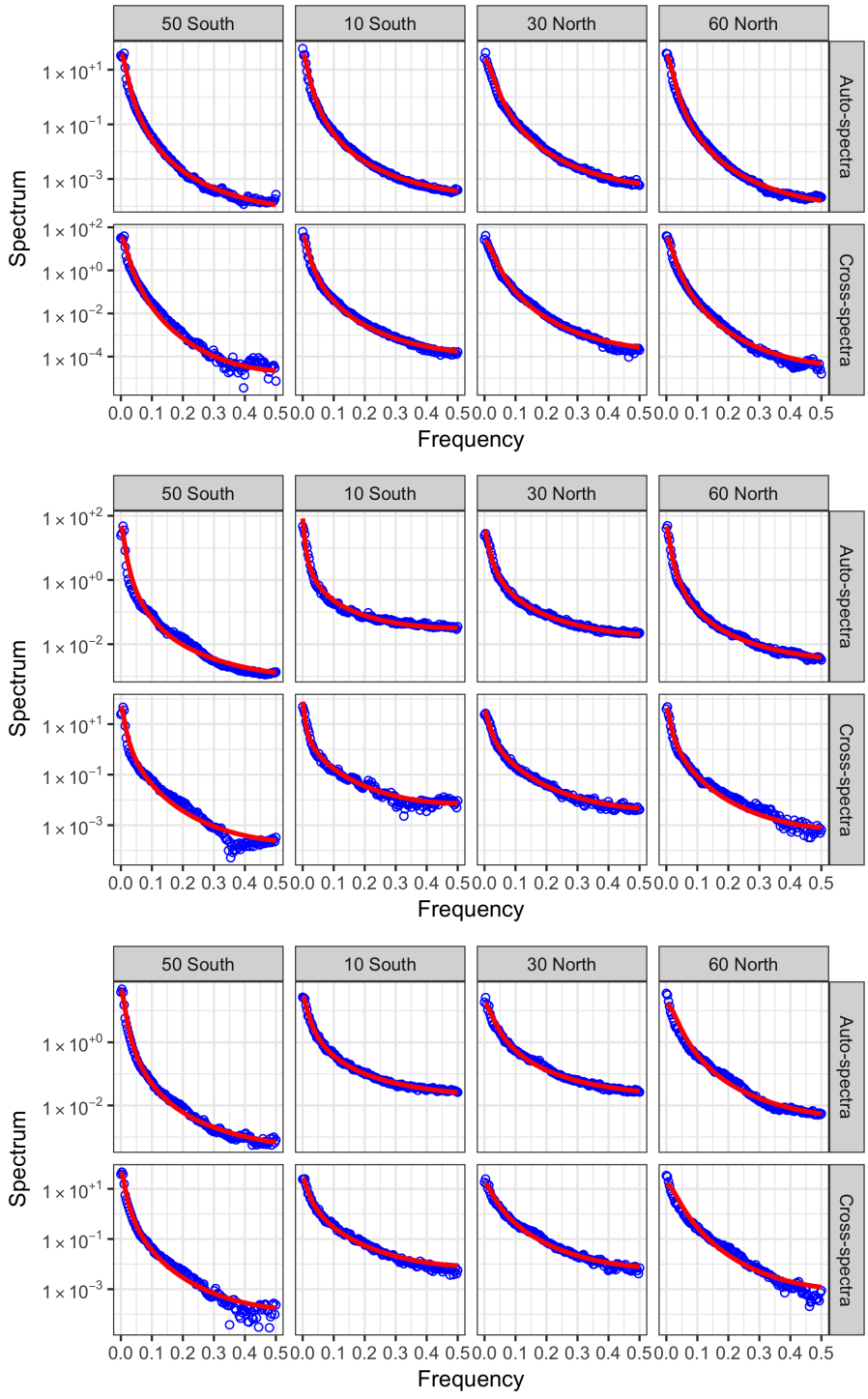


Figure 5.10: The mean periodograms and mean cross-periodograms (blue circles) with the trained spectral mass functions and cross-spectral mass functions (red lines) for four latitudinal bands at latitudes 50 degrees south, 10 degrees south, 30 degrees north and 60 degrees north.

5.6 Multivariate Training

In the fourth SMLE stage three bivariate Gaussian models are trained (one for each pair of climate variables). Figure 5.11 displays the arguments and moduli of the mean cross-periodograms (blue circles) with loess smoothing (red curves) over frequency between the TMQ, TS and U10 climate variables. Note that the arguments for the upper panels of Figure 5.11 are periodic. Hence, the smallest arguments on these panels are close to the largest arguments. These plots suggest that the bivariate Guassian models have accurately captured the bivariate dependence given the assumptions. However, the arguments and moduli in Figure 5.11 are extremely difficult to interpret.

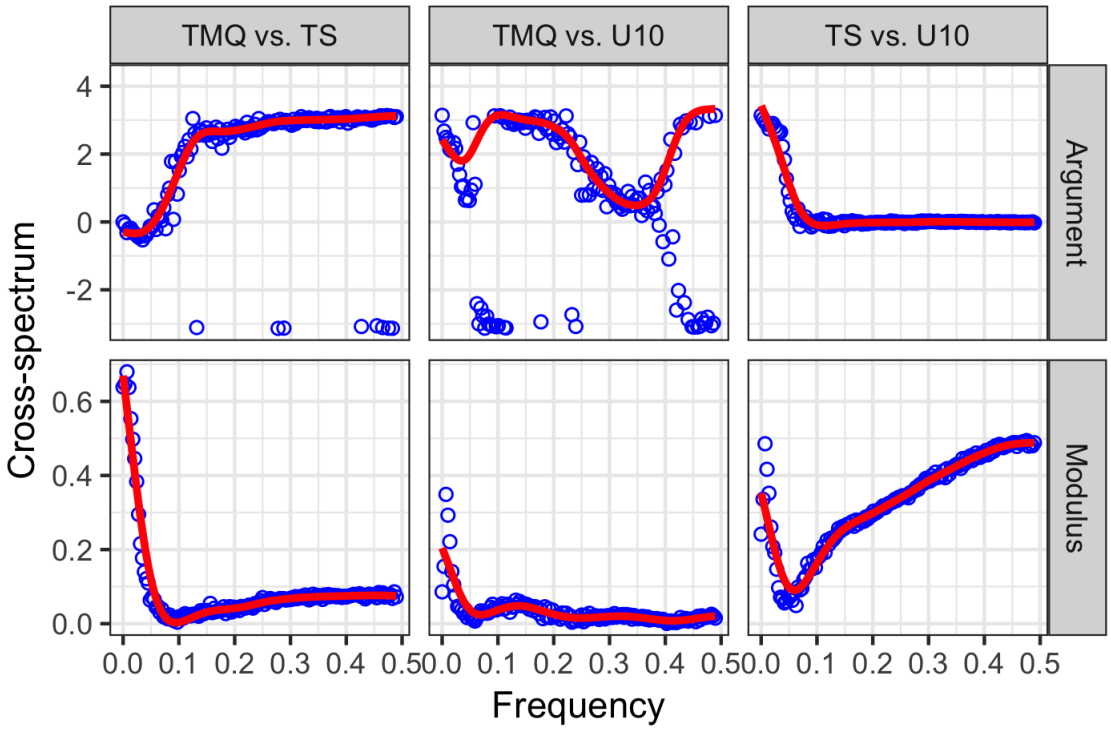


Figure 5.11: The arguments and moduli of the mean cross-periodograms (blue circles) with loess smoothing (red curves) over frequency between the TMQ, TS and U10 climate variables.

5.7 Training Summary

Some of the details of the four SMLE stages are included in Table 5.1. Table 5.1 includes the maximum number of parameters ($\max(P)$), the number of data points (N) and the cost of evaluation ($\mathcal{O}(N)$) for each marginal model and the number of marginal models in each stage (M). The number of data points does not include multiple realisations. In stage one, $288 \times 192 \times 3$ temporal models are trained, each with a maximum of three mean parameters, a standard deviation parameter and a maximum of three AR parameters. The computational cost of evaluating the temporal models was linear in the data space of dimension 95. In stage two, 192×3 longitudinal models are trained, each with an inverse range parameter, a maximum of one transition parameter and a smoothness parameters. The computational cost of evaluating the longitudinal models was log-linear in the data space of dimension 288. In stage three, 3 latitudinal models are trained, each with a scale parameters and a smoothness parameters. The computational cost of evaluating the latitudinal models was cubic in the data space of dimension 2×288. In stage four, 3 multivariate (bivariate) models are trained, each with a maximum of 20 natural cubic spline parameters. The computational cost of evaluating the multivariate models was cubic in the data space of dimension 2. Note that in stage four the multivariate model is not trained with all the data. This is advantageous since the cost of evaluating each multivariate (bivariate) model is independent of the number of variables. However, the number of bivariate models is dependent on the number of variables.

Stage	$\max(P)$	N	$\mathcal{O}(N)$	M
Temporal	7	95	$\mathcal{O}(N)$	$288 \times 192 \times 3$
Longitudinal	3	288	$\mathcal{O}(N \ln N)$	192×3
Latitudinal	2	2×288	$\mathcal{O}(N^3)$	191×3
Multivariate	20	2	$\mathcal{O}(N^3)$	3

Table 5.1: The maximum number of parameters ($\max(P)$), the number of data points (N) and the cost of evaluation ($\mathcal{O}(N)$) for each marginal model and the number of marginal models in each stage (M). The number of data points does not include multiple realisations.

Chapter 6

Diagnostics

The applicability of a stochastic generator depends on how accurately its simulations can represent climate variables from an ensemble. Hence, in this section, we compare the test ensemble from the fully-coupled CESM, consisting of three variables, to simulations from our univariate and multivariate global spatio-temporal models trained on the training ensemble, i.e. stochastic generator. The only difference between the univariate and multivariate global spatio-temporal models is that we set $\boldsymbol{\Xi}[v_1, v_2] \equiv 0$ for $v_1 \neq v_2$ for the univariate stochastic generator. This condition implies that the three climate variables are simulated independently rather than jointly. In the following sections the independent and joint stochastic generator ensembles correspond to simulations from the univariate and multivariate stochastic generators, respectively. Diagnostics are provided in two stages. First, we provide univariate diagnostics to assess the similarity of the stochastic generator simulations to climate model simulations using the test ensemble (members not in the training ensemble). We then provide multivariate diagnostics to assess the effect of simulating climate variables jointly rather than independently. Differences between the diagnostics are primarily gauged visually at this stage in the stochastic generator development (Castruccio et al., 2019). The main reason for gauging the differences between

the diagnostics visually at this stage is that climate scientists are primarily interested in whether the model can capture the phenomena that they are interested in or not. Identifying those phenomena that are not captured is therefore the primary focus. The aim is to continually address phenomena that are not captured until the climate science community is satisfied.

6.1 Univariate Diagnostics

We consider area-weighted statistics (to account for the non-regular latitude-longitude grid with more grid-points per unit area towards the poles) to compare the ensembles from the climate model and our stochastic generator (Table 6.1). The medians and means from the test ensemble and the joint stochastic generator ensemble are equal to two decimal places and the first and third quartiles are equal to one decimal place. This indicates that the joint stochastic generator accurately captured the body of the test ensemble distribution. The minimums and maximums are also well-represented for TS and TMQ. In the case of U10, the maximum is well-modelled, but the minimum is negative, which is not physically possible. While the difference is small, it nevertheless indicates the need for an improved model which can enforce non-negativity. A log transformation can remove the negative values. A transformation performed before modelling can result in biased simulations. Hence, it is important to incorporate any transformation into the model to account for biases. This is a potential improvement we postpone for future work.

We consider global and longitudinal ensemble member means to assess how accurately the joint stochastic generator captures variations in ensemble member mean over years and latitudes, respectively. There are ensemble member means for each of the 28 test and joint stochastic generator ensemble members. Figure 6.1 displays the means, maximums and minimums of the global (left panels) and longitudinal (right panels) ensemble member means over the test ensemble (red) and joint stochastic generator ensemble (blue) for the

Table 6.1: The mean (over years, longitudes and latitudes) and standard deviation (in parenthesis) of the weighted minimum, weighted first quantile, weighted median, weighted mean, weighted third quantile and weighted max over the test ensemble (left columns) and the joint stochastic generator (right columns) ensembles.

	TMQ (kg/m^2)		TS (K)		U10 (m/s)	
Ensemble	Test	Joint SG	Test	Joint SG	Test	Joint SG
Min.	0.23 (0.01)	0.18 (0.03)	214.60 (0.25)	214.50 (0.32)	0.65 (0.00)	-0.34 (0.13)
1st Qu.	15.04 (0.03)	15.03 (0.02)	282.56 (0.03)	282.56 (0.02)	3.40 (0.00)	3.40 (0.00)
Median	25.31 (0.04)	25.31 (0.04)	294.90 (0.02)	294.90 (0.02)	6.53 (0.01)	6.53 (0.01)
Mean	28.05 (0.03)	28.05 (0.05)	289.64 (0.02)	289.64 (0.02)	6.12 (0.00)	6.12 (0.00)
3rd Qu.	40.66 (0.05)	40.68 (0.09)	300.46 (0.02)	300.46 (0.02)	8.00 (0.01)	8.01 (0.01)
Max.	72.66 (0.70)	75.83 (2.33)	312.16 (0.23)	312.87 (0.49)	13.73 (0.09)	13.82 (0.16)

TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables. The maximum and minimum global ensemble member means, for each year, and longitudinal ensemble member means, for each latitude, are included as red and blue shaded regions, respectively. The longitudinal ensemble member means (right panels) suggest that the joint stochastic generator has captured the ensemble member means and their variations between ensemble members very accurately, since the ensemble member maximums, minimums and means are indistinguishable. The global ensemble member means (left panels) suggest that the joint stochastic generator has captured the ensemble member means and their variations between ensemble members accurately for the the TMQ (upper left panel) and TS (middle left panel) climate variable. However, the global TS ensemble member means for the joint stochastic generator ensemble have slightly larger variability between ensemble members and they are also slightly larger towards 2100. In contrast, the variation in global U10 ensemble member means (lower left panel) is larger for the test ensemble than the joint stochastic generator ensemble.

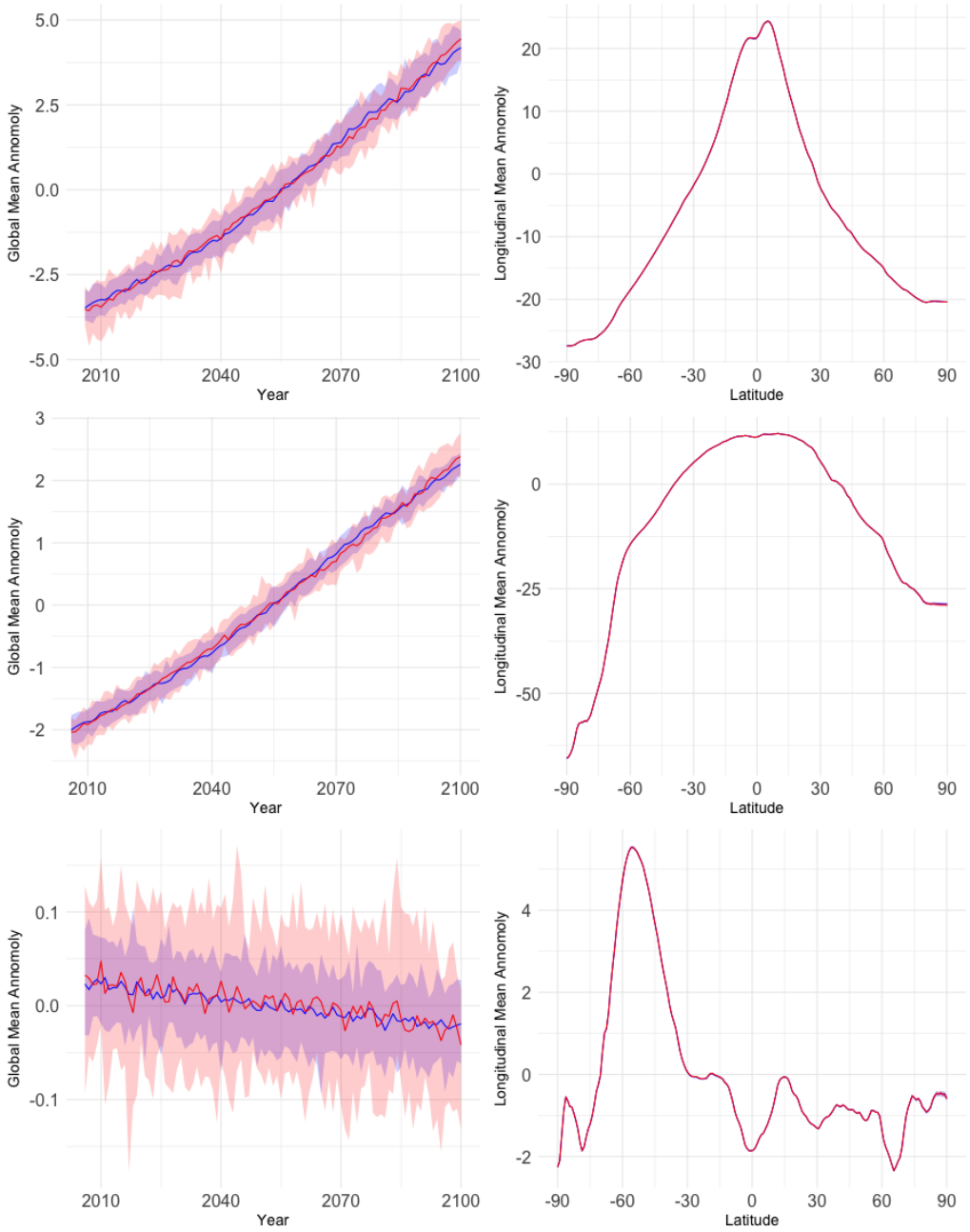


Figure 6.1: The means, maximums and minimums of the global (left panels) and longitudinal (right panels) ensemble member means over the test ensemble (red) and joint stochastic generator ensemble (blue) for the TMQ (upper panels), TS (middle panels) and U10 (lower panels) climate variables. The maximum and minimum global ensemble member means, for each year, and longitudinal ensemble member means for each latitude, are included as red and blue shaded regions, respectively. Note that the longitudinal ensemble member maximums, minimums and means are indistinguishable in the right panels.

The remaining univariate and multivariate diagnostics are based on a simple linear regression model, with year as the regressor, trained at each spatial location for each variable over the test ensemble and joint stochastic generator ensemble. Figures 6.2, 6.3 and 6.4 display the intercepts, slopes, residual standard deviations and residual lag-one auto-covariances over the spatial locations for the TMQ, TS and U10 climate variables, respectively. The intercepts, slopes, residual standard deviations and residual lag-one auto-covariances demonstrate that the multivariate global spatio-temporal model can capture complex variations in means, trends, standard deviations and auto-covariances over the globe.

In Figure 6.3, the intercepts and slopes demonstrate how the multivariate global spatio-temporal model can capture complex variation in means and trends over the globe accurately, even though they are captured independently with ARMA models over the globe. There are some differences between the residual standard deviations. First, the standard deviations for the joint stochastic generator ensemble are slightly smaller over parts of the Arctic Ocean and the Antarctic coast than for the test ensemble. Second, in the central Pacific Ocean, the variations in standard deviations for the joint stochastic generator ensemble are less smooth than for the test ensemble. The latter suggests that the multivariate global spatio-temporal model could benefit from training the ARMA models dependently (e.g. with a Gaussian process over the standard deviation parameters) over the central Pacific Ocean. This region is associated with the non-periodic El-Niño-Southern Oscillation, which is notoriously difficult to model. Third, there are some differences between the residual lag-one auto-covariances. The residual lag-one auto-covariances for the joint stochastic generator ensemble over the Arctic Ocean above Russia and across the Antarctic coast below the Pacific Ocean are substantially smaller than for the test ensemble. Both the smaller residual standard deviations and residual lag-one auto-covariances occur towards the poles. This suggests the multivariate global spatio-temporal model is not capturing the poles as accurately as the low and middle latitudes. Although there are some differences between the left and right panels of Figure 6.3 these differences are very

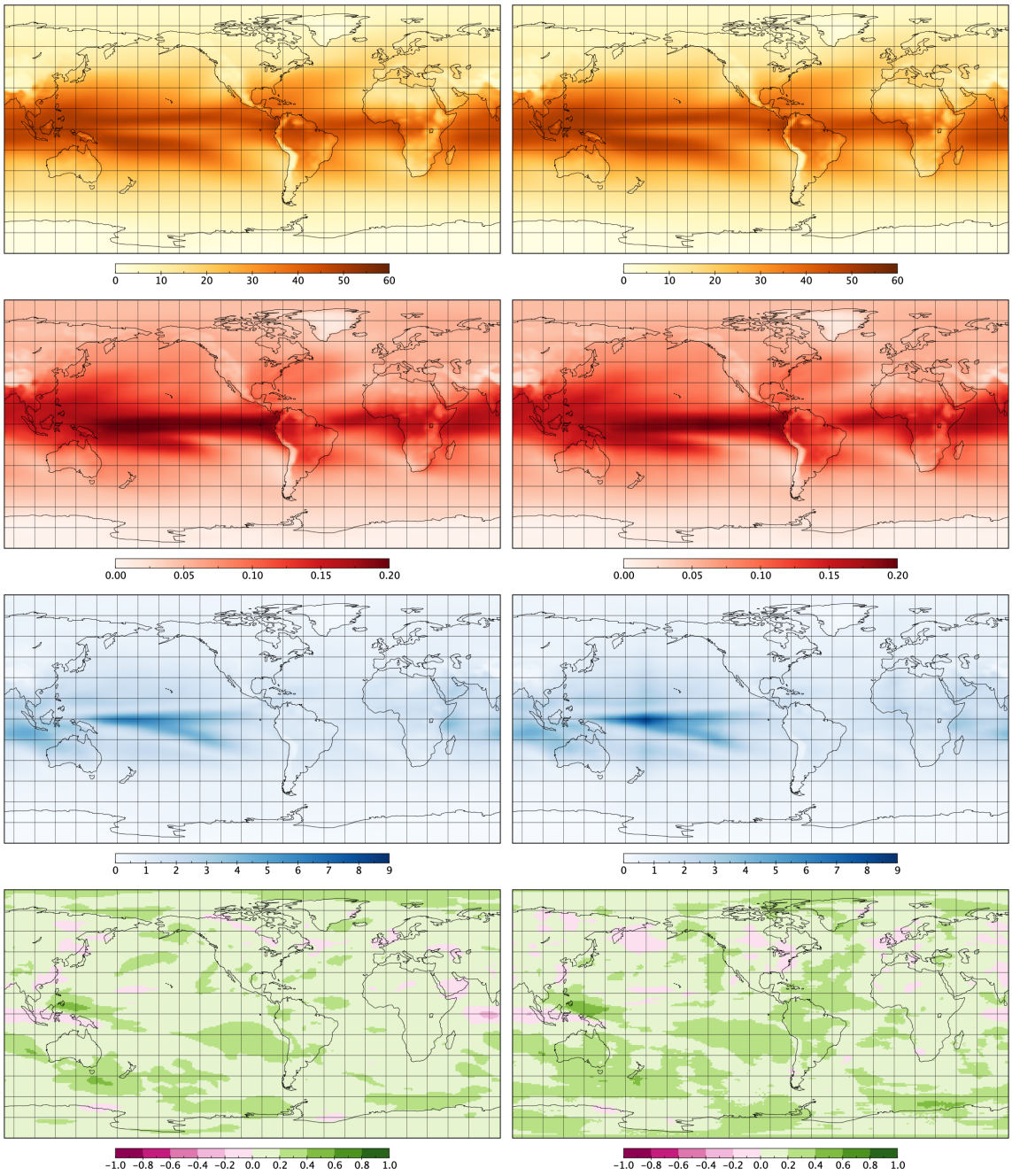


Figure 6.2: The intercepts (first row), the slopes (second row), the residual standard deviations (third row) and the residual lag-one auto-covariances (fourth row) from simple linear regression models trained at each spatial location for TMQ (kg/m^2) over the test ensemble (left column) and the joint stochastic generator ensemble (right column).

small.

From the residuals, lag-one longitudinal auto-correlations and lag-one latitudinal auto-correlations can be used to explore the variation in longitudinal and latitudinal dependence over the globe. Figures 6.5, 6.6 and 6.7, display the lag-one residual longitudinal auto-correlations (upper panels) and lag-one residual latitudinal auto-correlations (lower panels) at each spatial location over the test ensemble (left panels) and the joint stochastic generator ensemble (right panels) for the TMQ, TS, and U10 climate variables, respectively. A coarse colour scale has been selected so that the features discussed are more clear. The plots indicate that the joint stochastic generator ensemble has captured the strong longitudinal and latitudinal auto-correlations over most regions. However, the joint stochastic generator ensemble has not captured the weak longitudinal and latitudinal auto-correlations over some coastal and mountainous regions. This is expected, since the multivariate and global spatio-temporal model assumes that the VARMA innovations are axially symmetric (i.e. longitudinally stationary). In Castruccio and Guinness (2017) evolutionary processes are used to relax this assumption, however, the evolutionary process only remove the assumption that the stationary processes over land and ocean are the same. It is clear from Figure 6.5, 6.6 and 6.7 that this approach is not sufficient to capture the non-stationary spatial structures of the three climate variables. This issue is discussed in more detail in Chapter 7.

6.2 Multivariate Diagnostics

To assess how accurately the univariate and multivariate stochastic generators capture the dependencies between the three climate variables, the cross-correlation between the residuals at each spatial location are displayed for each pair of variables for the test ensemble, the joint stochastic generator ensemble and the independent stochastic generator ensemble, see Figure 6.8. TMQ and TS (upper left panel) display strong positive and

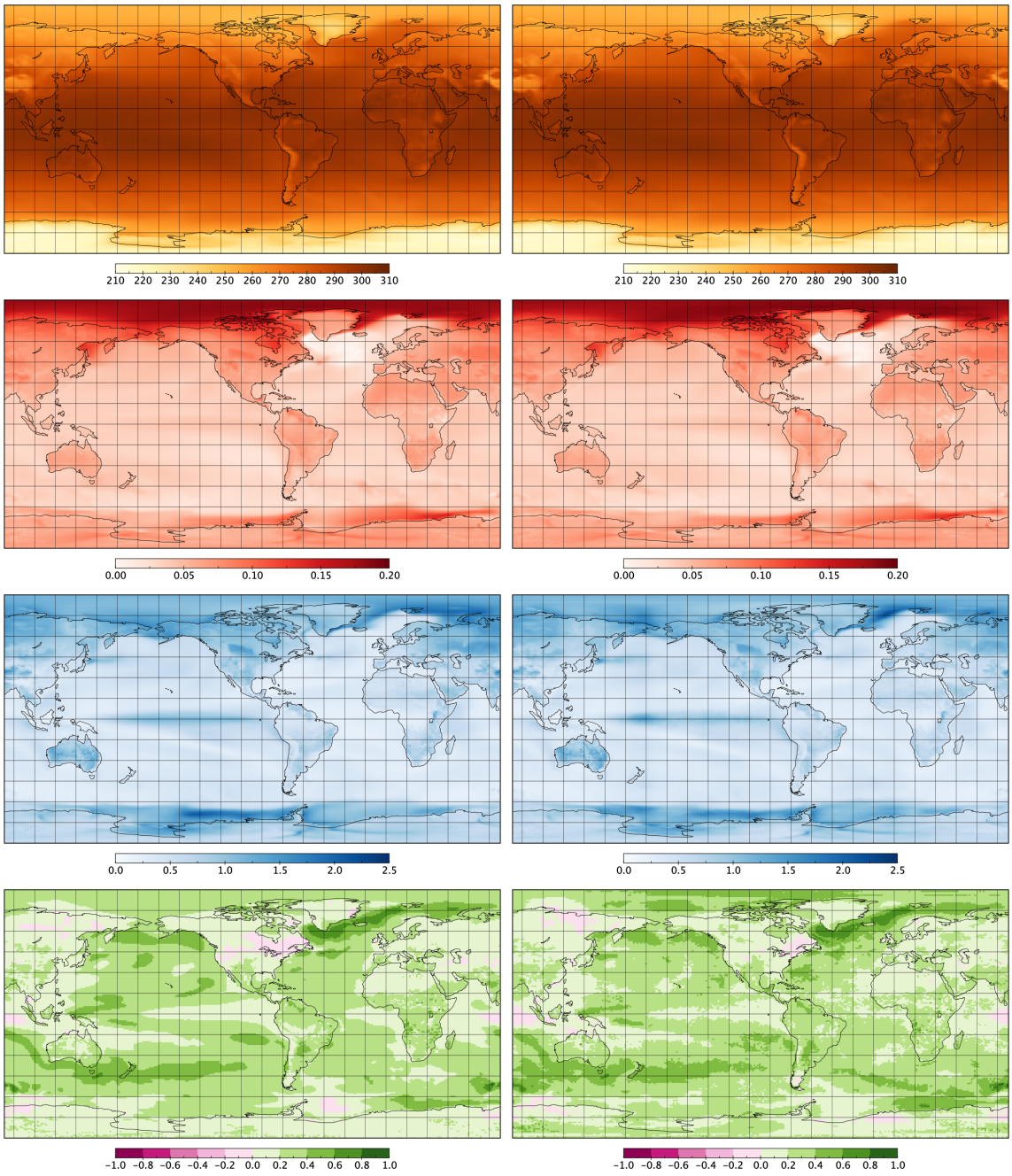


Figure 6.3: The intercepts (first row), the slopes (second row), the residual standard deviations (third row) and the residual lag-one auto-covariances (fourth row) from simple linear regression models trained at each spatial location for TS (K) over the test ensemble (left column) and the joint stochastic generator ensemble (right column).

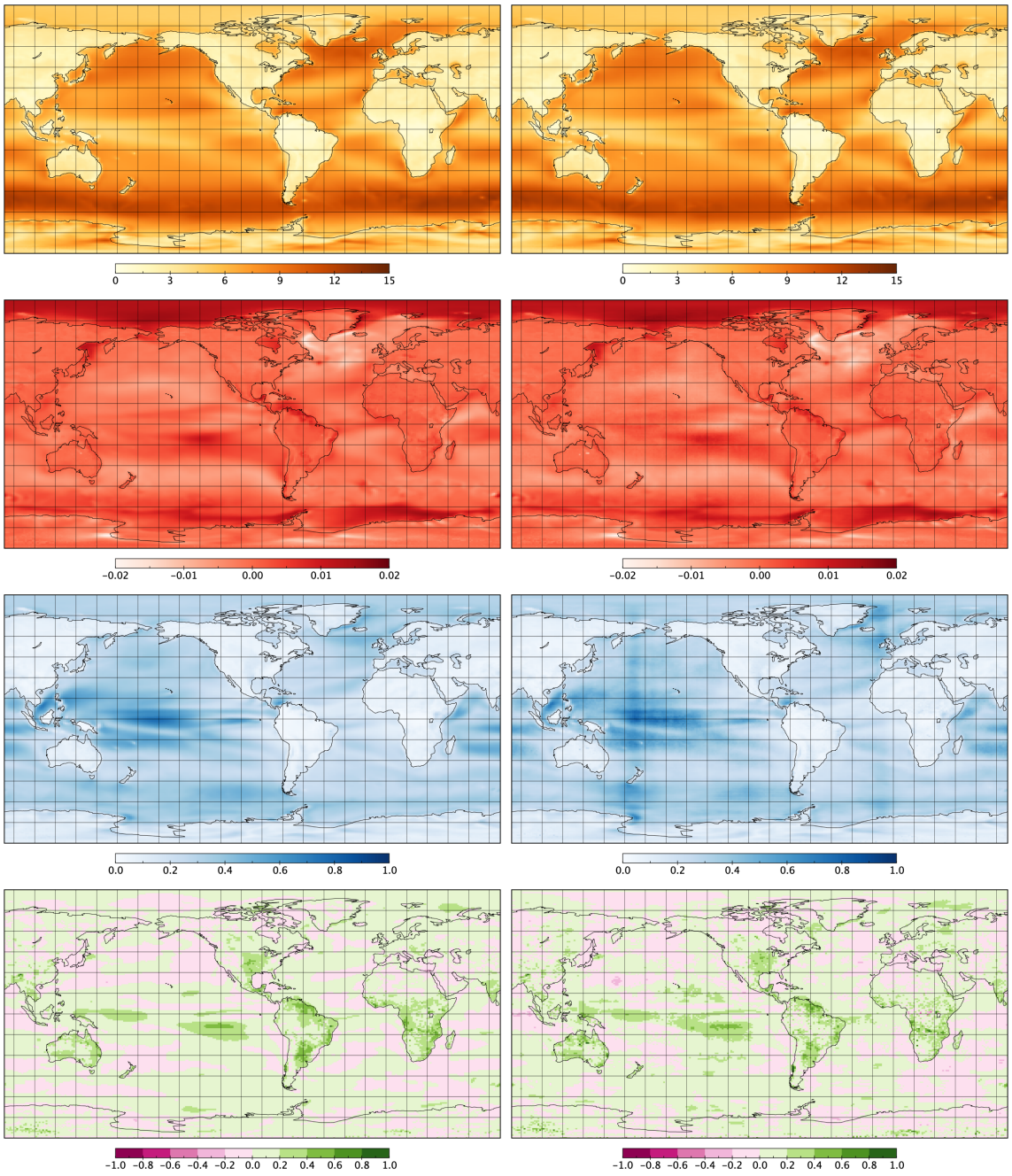


Figure 6.4: The intercepts (first row), the slopes (second row), the residual standard deviations (third row) and the residual lag-one auto-covariances (fourth row) from simple linear regression models trained at each spatial location for U10 (m/s) over the test ensemble (left column) and the joint stochastic generator ensemble (right column).

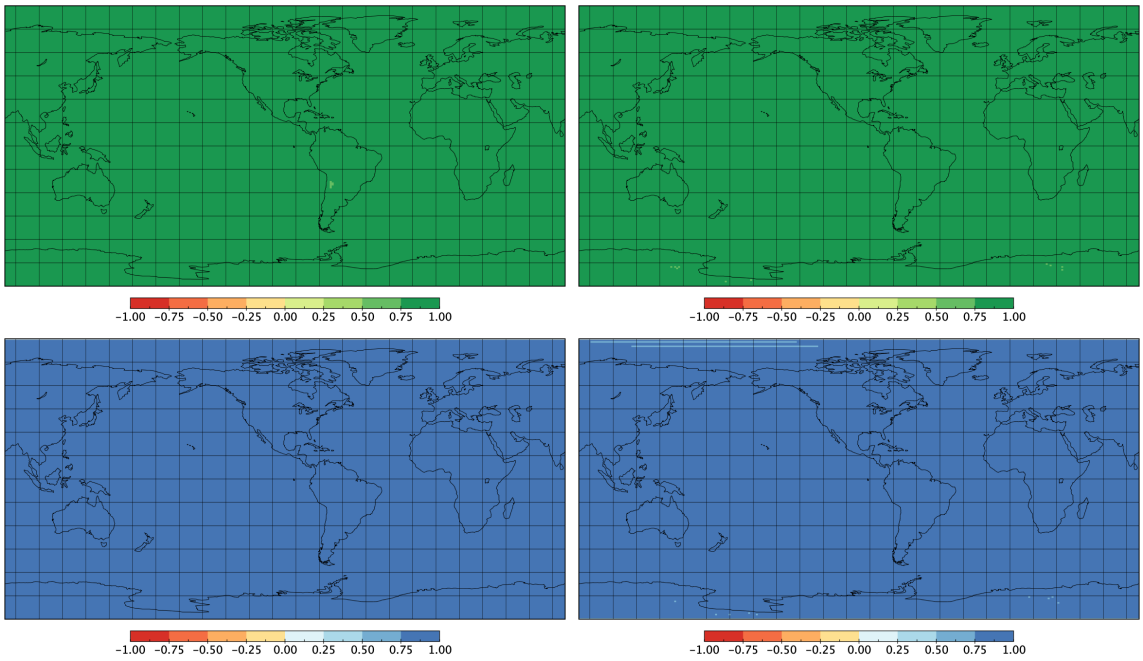


Figure 6.5: The lag-one residual longitudinal auto-correlations (upper panels) and lag-one residual latitudinal auto-correlations (lower panels) from the simple linear regression models trained at each spatial location for TMQ (kg/m^2) over the test ensemble (left panel) and the joint stochastic generator ensemble (right panel).

negative spatially varying cross-correlation. The lower left panel of Figure 6.8 displays approximately zero cross-correlation between the TMQ and TS residuals over space. This is expected, since the TMQ and TS climate variables were simulated independently. There appears to be some structure to the slight positive and negative cross-correlations between the TMQ and TS residuals; however, these do not correspond to the structures displayed in the upper left panel of Figure 6.8. The middle left panel (Figure 6.8) displays positive cross-correlation between the TMQ and U10 residuals over space. The left panels of Figure 6.8 suggest that the multivariate global spatio-temporal model cannot capture variation in cross-correlation over space, but if there is an average positive or negative cross-correlation over space, the model will capture it. The upper middle panel of Figure 6.8 suggests that the average cross-correlation between the TMQ and U10 residuals over space is approximately zero. As suggested from the left column of panels in Figure 6.8, the

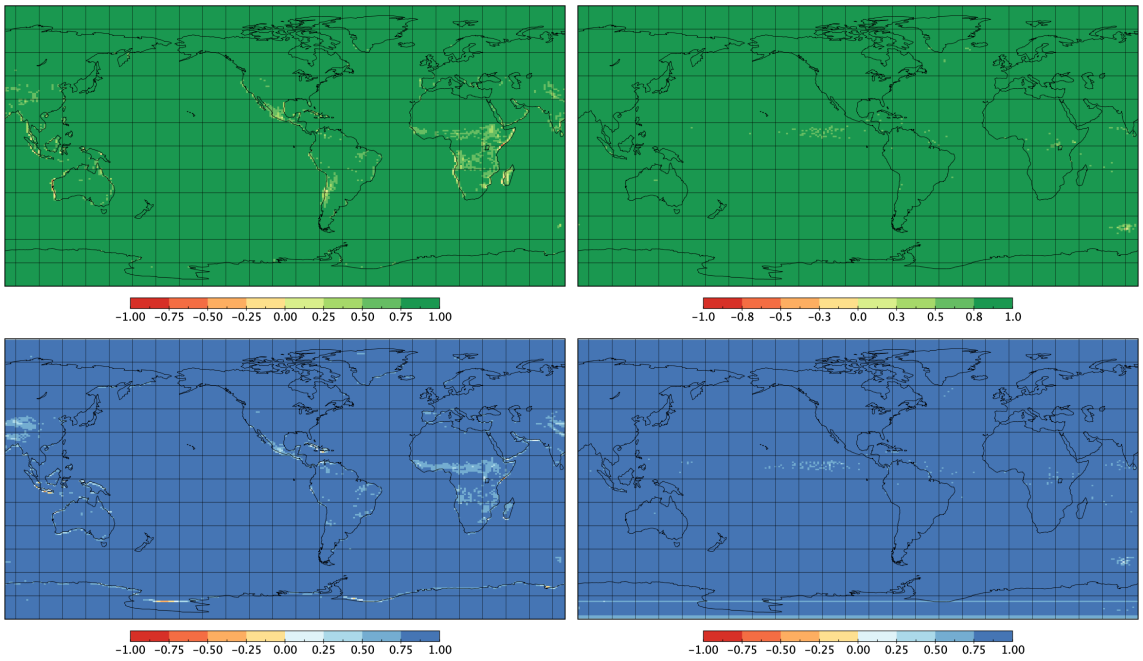


Figure 6.6: The lag-one residual longitudinal auto-correlations (upper panels) and lag-one residual latitudinal auto-correlations (lower panels) from the simple linear regression models trained at each spatial location for TS (K) over the test ensemble (left panel) and the joint stochastic generator ensemble (right panel).

model can only capture the average positive or negative cross-correlation over space and not the spatial variation. However, since the average is approximately zero in the upper middle panel of Figure 6.8, there is no substantial improvement to the joint stochastic generator ensemble over the independent stochastic generator ensemble. In contrast to the middle column of panels of Figure 6.8, the average cross-correlation between the TS and U10 residuals is negative for the middle right panel of Figure 6.8. Hence, the model has captured the average negative cross-correlation between the TS and U10 residuals over space but not the structure displayed in the upper right panel of Figure 6.8.

To assess the variation in cross-correlation over space between ensemble members, cross-correlations for each member are displayed at nine different longitudes and all latitudes in Figure 6.9. These plots demonstrate how the cross-correlations of the joint stochastic

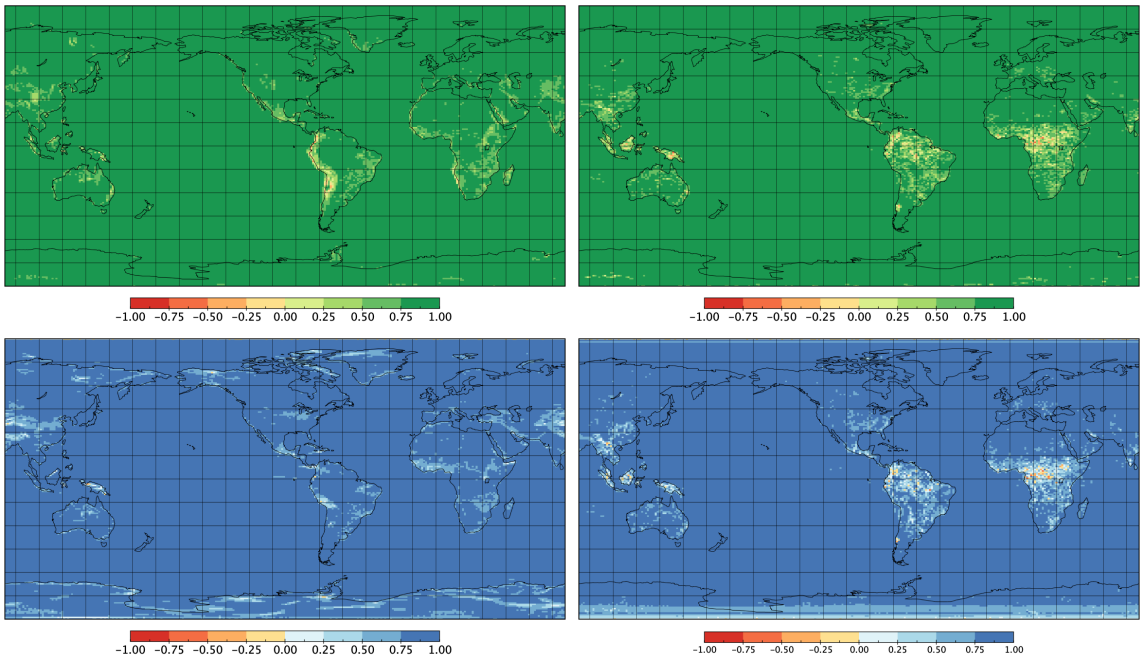


Figure 6.7: The lag-one residual longitudinal auto-correlations (upper panels) and lag-one residual latitudinal auto-correlations (lower panels) from the simple linear regression models trained at each spatial location for U10 (m/s) over the test ensemble (left panel) and the joint stochastic generator ensemble (right panel).

generator are close to the average cross-correlation of the test ensemble, while the cross-correlation of the univariate stochastic generator is approximately zero as expected. These figures also suggest that the variation in cross-correlation between ensemble members is smaller for the test ensemble than for the joint or independent stochastic generator ensembles.

The diagnostics in this chapter have highlighted various strengths of the model. These include the model's capacity to capture complex variations in mean, trend, standard deviation and lag-one temporal auto-correlation as well as lag-one longitudinal and latitudinal auto-correlation. The model can capture average cross-correlation across the globe but not variation in cross-correlation across the globe. The limitation of this model to capture variations in cross-correlation across the globe is in the model for the innovation. Modelling

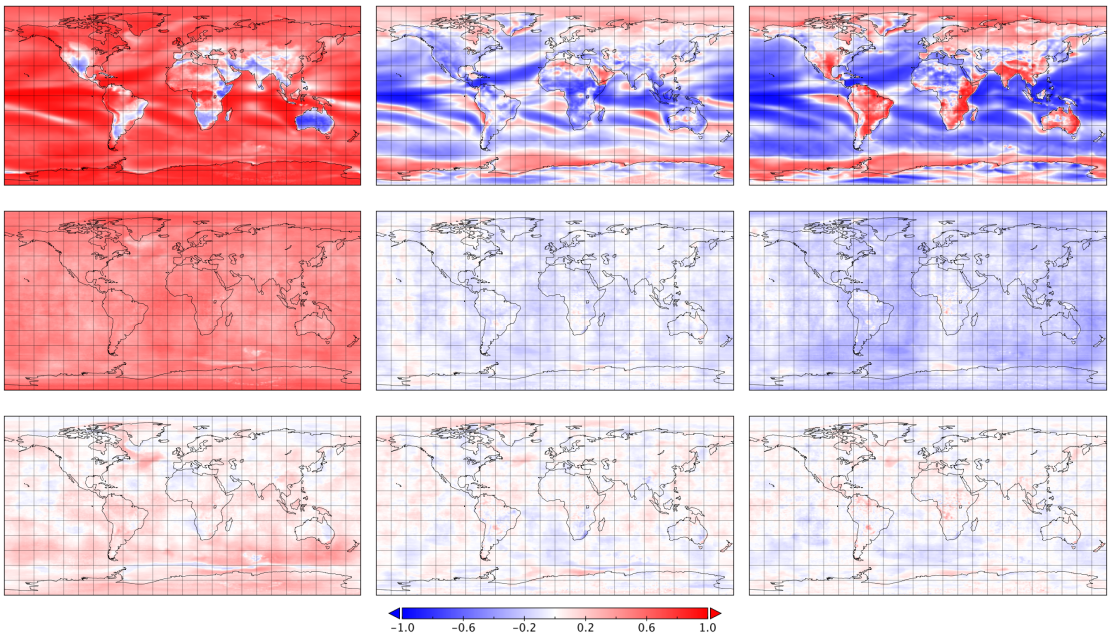


Figure 6.8: The cross-correlation between the TMQ and TS residuals (left panels), the TMQ and U10 residuals (middle panels) and the TS and U10 residuals (right panels) from simple linear regression model fits at each spatial location over the test ensemble (upper panels), the joint stochastic generator ensemble (middle panels) and the independent stochastic generator ensemble (lower panels).

the innovations in the spectral domain across longitude, assuming stationary longitudinal bands, has the effect of reducing the capacity of the model to capture phenomena that are longitudinally localised. This can be improved with an evolutionary spectral approach Castruccio and Guinness (2017) however, this model can only capture phenomena that are localised to either land or ocean. As Figures 6.5, 6.6 and 6.7 suggest that the phenomena that the model can not capture is localized to coastal regions rather than to ocean and land regions, the evolutionary spectral approach is still to demonstrate its effectiveness in this regard. An ideal model would require the capacity to capture lag-one longitudinal and latitudinal auto-correlations at each spatial location. Similarly to how the current model can capture means, trends, standard deviations and temporal auto-correlations at each spatial location. A possible direction for a model of this kind is introduced in Chapter 7

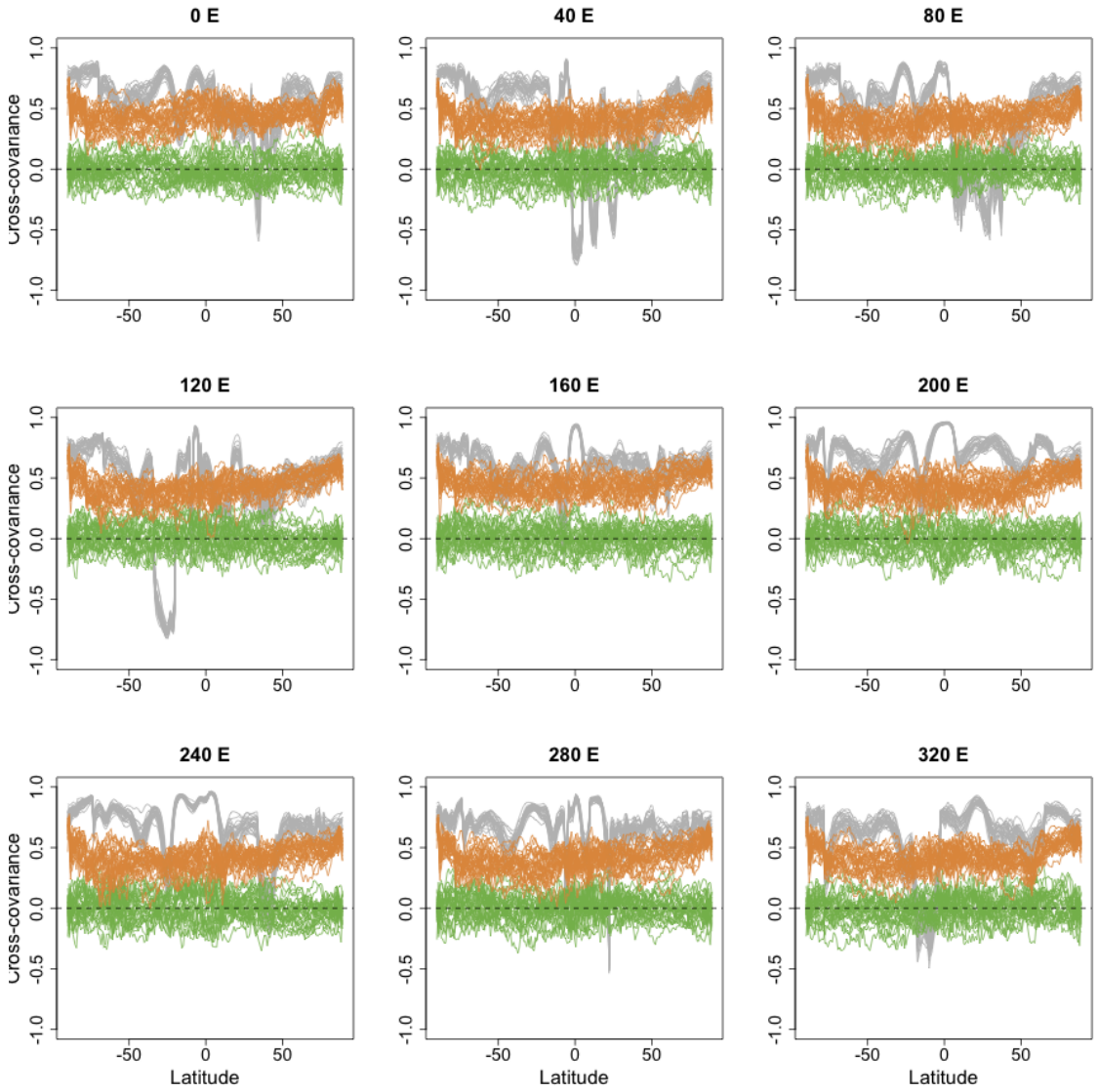


Figure 6.9: The cross-correlation between the TMQ and TS residuals at nine different longitudes and all latitudes for each member of the test ensemble (grey), the independent stochastic generator ensemble (green) and the joint stochastic generator ensemble (orange).

covering further work.

Chapter 7

Further Work

Chapter 6 demonstrates how the proposed multivariate global spatio-temporal model can capture complex features of the CESM-LE that are important to climate scientists. However, there are also limitations of the current model. Three of these limitations, which include: the stationary longitudinal model, Gaussian model and polar model, are considered in detail. A series of possible solutions are also provided.

7.1 Non-stationary Longitudinal Model

It is assumed in Section 4.3 that the multivariate spatio-temporal innovations are multivariate axially symmetric. Hence, although the longitudinal bands of the innovations are non-stationary with respect to mean and variance they are jointly stationary with respect to correlation. This is a limitation of the model since Figures 6.5, 6.6 and 6.7 suggest that the longitudinal bands are non-stationary with respect to correlation. The longitudinal model that results from the multivariate axially symmetric assumption is a product of Whittle models, each of which are fully specified by a spectral mass function. Hence, a

natural extension of this model could use an evolutionary spectral process (Priestley, 1981, Section 11.2).

An evolutionary spectral process approach was introduced in Castruccio and Guinness (2017) where the spectra could be different over land and ocean. This model implicitly assumes that the longitudinal bands are separately stationary over land and ocean. Hence, the comparison of mean periodograms calculated from the land and ocean separately suggested a significant improvement over the axially symmetric assumption. Despite this significant improvement, the lag-one residual longitudinal cross-correlations in Figures 6.5, 6.6 and 6.7 suggest that the non-stationary correlation structures are substantially more complicated. These figures suggest that the most substantial non-stationary correlation structures results from coastal and mountainous regions. As a consequence a more complex longitudinal model is required.

The temporal and latitudinal models are dynamically specified, however, the longitudinal model is not. Hence, a natural approach would be to also dynamically specify the longitudinal model. For example, a non-stationary AR order one model like the latitudinal model. The only complication, regarding this particular specification, arises from the circular geometry of the longitudinal process. Consider the random vector $\mathbf{Y} = (Y_0, \ldots, Y_{n-1})$ with a circular geometry with the following non-stationary circular AR order one model

$$Y_k = \phi_k Y_{k-1(\text{mod } n)} + \epsilon_k \quad \text{where} \quad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2).$$

Since, the AR dependence decays exponentially around the circle *ad infinitum* the condition on each σ_k^2 that results a unit-variance random vector is complicated to derive, however, there is a simple closed form solution. First, note that the random variable Y_k in terms of Y_{n-1} is

$$Y_k = \phi_{0,k} Y_{n-1} + \sum_{j=0}^{k} \phi_{j+1,k} \epsilon_j \tag{7.1}$$

where $\phi_{a,b}$ is equal to the product of ϕ_i from a to b (inclusive) if $a \leq b$ and 1 for $a > b$.

Second, note that the random variable Y_{n-1} in terms of Y_k is

$$Y_{n-1} = \phi_{k+1,n-1} Y_k + \sum_{j=k+1}^{n-1} \phi_{j+1,n-1} \epsilon_j. \tag{7.2}$$

The first equation (7.2) can be substituted into the second equation (7.1) to result in an equation in terms of the independent innovations (i.e. ϵ_k for all k). Since this equation (7.3) is in terms of the independent innovations the calculation of variances and correlations is simplified:

$$Y_k = \frac{1}{1 - \phi_{0,n-1}} \left(\sum_{j=0}^{k} \phi_{j+1,k} \epsilon_j + \sum_{j=k+1}^{n-1} \phi_{j+1,n-1} \phi_{0,k} \epsilon_j \right). \tag{7.3}$$

It then follows that if and only if

$$\sigma_k^2 = (1 - \phi_k^2) \frac{1 - \phi_{0,n-1}}{1 + \phi_{0,n-1}} \tag{7.4}$$

the random vector is unit-variance; that is the variance of Y_k is one for all k. It is reasonable to assume that this model can capture complex longitudinal structures and that it can provide efficient simulation. However, it is not obvious how estimation will be performed with the SMLE method. A preliminary exploration of this model suggests that a method of moments approach would be possible, however, a significant amount of further work is required to solve this problem.

7.2 Trans-Gaussian Model

It is assumed in Section 4.1 that the multivariate global spatio-temporal model is Gaussian. This assumption has numerous advantages: model specification only requires a mean vector and covariance matrix, the model is mathematically tractable and the model is closed under marginalisation. It is suspected that a necessary condition for a parametric model (e.g. Gaussian) to be marginally parameterised is for that model to be closed

under marginalisation for some data subsets as no marginally parameterised model that is not closed under marginalisation for some data subsets is currently known. Hence, the Gaussian assumption is fundamental in this regard.

The Gaussian assumption appears to provide an accurate approximation for the annually averaged ensemble data. However, at a higher temporal frequency (e.g. monthly averaged) this approximation, from the central limit theorem, will be less accurate. One possible solution to this problem comes from transformed Gaussian models. Consider the random vector $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and assume that the transformation $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ where $\mathbf{Y}_k = g_k^{-1}(\mathbf{X}_k)$ for all k is Gaussian. The likelihood function for $\mathbf{X}$ is

$$\frac{||\mathbf{J}_g||}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(g(\mathbf{X}) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-\mathbf{1}}(g(\mathbf{X}) - \boldsymbol{\mu})\right)$$

where $||\mathbf{J}_g||$ is the absolute value of the determinant of the Jacobian of g. Since $||\mathbf{J}_g|| = ||\mathbf{J}_{g_1}|| \cdot ||\mathbf{J}_{g_2}||$ the model for $\mathbf{X}$ is closed under marginalisation for the data subsets $\mathbf{X}_1$ and $\mathbf{X}_2$. Hence, this approach could be used to overcome the Gaussian assumption without compromising much of the numerous advantages. As a consequence to the multivariate global spatio-temporal model a transformation of each of the time series would increase the flexibility of the model towards skewness and kurtosis but not effect the stages two to four of the SMLE method. What transformations and how they might be incorporated into the time series model is suggested for further work.

Chapter 8

Conclusion

For a small number of variables, a multivariate stochastic generator can be trained on a climate model ensemble consisting of only a few members to simulate a large ensemble. As a consequence, if only a small number of climate variables are of interest, then a large ensemble of these variables can be obtained with a stochastic generator with reduced computational and storage expenses. The computational expense is reduced since only a small ensemble is required to train the stochastic generator and the storage expense is reduced since only the stochastic generator requires storage; both important considerations for climate modeling centers.

However, the applicability of a stochastic generator depends on how accurately its simulations can represent ensemble features of interest. For example, the univariate diagnostics suggest that extreme value analysis and polar region analysis would not be appropriate for the proposed stochastic generator, since extremes and polar regions are less accurately captured. While analyses that require the accurate representation of means, trends, variances, and temporal auto-correlations over space would be suitable. The multivariate diagnostics demonstrate that the multivariate stochastic generator can capture cross-correlations,

but it is limited to an average over space. While this is a substantial improvement over univariate stochastic generators, it still limits the application of multivariate stochastic generators to variables that have spatially homogeneous cross-covariance, which is rare in practice.

The application of stochastic generators to large ensembles is relatively new. A stochastic generator requires a complex multivariate global spatio-temporal model for its simulations to accurately represent the ensemble distribution of multiple variables. Although developing such a model is a difficult task, the advantages of a stochastic generator are such that attempts are valuable. The multivariate global spatio-temporal model underlying the proposed stochastic generator can capture complex structures in mean, trend, variance and temporal auto-correlation over space and longitudinal auto-correlation over latitudes. However, there are limitations that require model improvements, namely: increased flexibility in modeling tails (e.g. non-Gaussian assumptions), increased flexibility in modeling polar regions, and a model specification that can capture spatially-heterogeneous cross-correlation structures between variables.

Appendix A

Appendix

A.1 Marginal DVARMA Model

Proof. If the DVARMA model (3.2) is stable then it can be represented as $\mathbf{Y}_t = \boldsymbol{\mu}_t + \sum_{i=0}^{\infty} \Psi_i \Sigma \mathbf{U}_{t-i}$ (Lütkepohl, 2005, p. 421) where Ψ_i is a linear combination of the diagonal matrices of AR and MA parameters. Consequently, $\Psi_i = \text{diag}(\psi_{i,\mathbf{x}_s})$ are diagonal matrices of infinite order MA parameters. The mean vector (at time t) and lag-h auto-covariance matrix-valued function of this model are $\boldsymbol{\mu}_t$ and

$$\Gamma(h) := \sum_{i=0}^{\infty} \Psi_{i+h} \Sigma \mathrm{R}(\boldsymbol{\nu}) \Sigma \Psi_i, \tag{A.1}$$

respectively. Therefore, the DVARMA spatio-temporal covariance function is

$$C(\mathbf{x}_1, \mathbf{x}_2, h) := \sigma^2_{\mathbf{x}_1} \sigma^2_{\mathbf{x}_2} \rho_{\mathbf{x}_1,\mathbf{x}_2} \sum_{i=0}^{\infty} \psi_{i+h,\mathbf{x}_1} \psi_{i,\mathbf{x}_2} \tag{A.2}$$

where $\rho_{\mathbf{x}_1,\mathbf{x}_2}$ is the element of $\mathbf{R}$ in row $\mathbf{x}_1$ and column $\mathbf{x}_2$. □

A.2 SMLE Consistency

Proof. Base case ($k = 2$): Since $\widehat{\boldsymbol{\eta}}_2(\mathbf{Y}) \subseteq \widehat{\boldsymbol{\theta}}_1(\mathbf{Y})$ by definition, $\widehat{\boldsymbol{\eta}}_2(\mathbf{Y})$ is consistent by assumption and $\widehat{\boldsymbol{\theta}}_2(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_2(\mathbf{Y}))$ is consistent by the Spall consistency theorem (Spall, 1989). Inductive hypothesis ($k > 2$): Suppose the theorem holds for k from 2 to $n < K$. Since $\widehat{\boldsymbol{\eta}}_{n+1}(\mathbf{Y}) \subseteq \{\widehat{\boldsymbol{\theta}}_1(\mathbf{Y}), \ldots, \widehat{\boldsymbol{\theta}}_n(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_n(\mathbf{Y}))\}$ by definition, $\widehat{\boldsymbol{\eta}}_{n+1}(\mathbf{Y})$ is consistent by the inductive hypothesis and $\widehat{\boldsymbol{\theta}}_{n+1}(\mathbf{Y}, \widehat{\boldsymbol{\eta}}_{n+1}(\mathbf{Y}))$ is consistent by the Spall consistency theorem.(Spall, 1989). □

A.3 Residual Transformation

Proof. Let B denote the back-shift operator such that $\mathrm{B}\mathbf{Y}_r = \mathbf{Y}_{t-1}$. With the back-shift operator the DVARMA model (3.2) can be transformed as

$$\begin{aligned}
\mathbf{Y}_t - \boldsymbol{\mu}_t &= \Sigma\mathbf{U}_t + \sum_{i=1}^{p} \Phi_i(\mathbf{Y}_{t-i} - \boldsymbol{\mu}_{t-i}) + \sum_{i=1}^{q} \Pi_i \Sigma \mathbf{U}_{t-i} \\
\mathbf{Y}_t - \boldsymbol{\mu}_t + \sum_{i=1}^{p} \Phi_i \mathrm{B}^i(\mathbf{Y}_t - \boldsymbol{\mu}_t) &= \Sigma\mathbf{U}_t + \sum_{i=1}^{q} \Pi_i \mathrm{B}^i \Sigma \mathbf{U}_t \\
\left(I - \sum_{i=1}^{p} \Phi_i \mathrm{B}^i\right)(\mathbf{Y}_t - \boldsymbol{\mu}_t) &= \left(I + \sum_{i=1}^{q} \Pi_i \mathrm{B}^i\right) \Sigma \mathbf{U}_t \\
\Sigma^{-1}\left(I + \sum_{i=1}^{q} \Pi_i \mathrm{B}^i\right)^{-1}\left(I - \sum_{i=1}^{p} \Phi_i \mathrm{B}^i\right)(\mathbf{Y}_t - \boldsymbol{\mu}_t) &= \mathbf{U}_t.
\end{aligned}$$

□

A.4 Spectral Mass Function Sum

Proof. From (4.7) the trace of $\mathbf{R}_{m,v}$ is

$$\begin{aligned}
\operatorname{trace}(\mathbf{R}_{m,v}) &= \operatorname{trace}(\mathrm{W}^{-1}\operatorname{diag}(\mathrm{W}\mathbf{r}_{m,v})\mathrm{W}) \\
&= \operatorname{trace}(\operatorname{diag}(\mathrm{W}\mathbf{r}_{m,v})\mathrm{W}\mathrm{W}^{-1}) \\
&= \operatorname{trace}(\operatorname{diag}(\mathrm{W}\mathbf{r}_{m,v})) \\
&= \sum_{c\in\mathbb{Z}_L} f_{m,v}(c).
\end{aligned}$$

Since $\mathbf{R}_{m,v}$ is a correlation matrix the trace is also L. □

A.5 Estimated Whittle Likelihood Function

Proof.

$$\begin{aligned}
-2\ln\mathcal{N}(\mathbf{0}, \mathbf{R}_{m,v} \mid \widehat{\mathbf{u}}_{r,t}[m,v]) &= \ln|2\pi\mathbf{R}_{m,v}| + \widehat{\mathbf{u}}^*_{r,t}[v,m]\mathbf{R}^{-1}_{m,v}\widehat{\mathbf{u}}_{r,t}[m,v] \\
&= \ln|2\pi\mathrm{W}\mathbf{F}_{m,v}\mathrm{W}^*| + (\mathrm{W}\widehat{\mathbf{u}}_{r,t}[m,v])^*\mathbf{F}^{-1}_{m,v}\mathrm{W}\widehat{\mathbf{u}}_{r,t}[m,v] \\
&= \ln|2\pi\mathbf{F}_{m,v}| + \widetilde{\mathbf{u}}^*_{r,t}[m,v]\mathbf{F}^{-1}_{m,v}\widetilde{\mathbf{u}}_{r,t}[m,v] \\
&= \ln(2\pi) + \sum_{c=0}^{L-1}\ln f_{m,v}(c) + \sum_{c=0}^{L-1}\frac{|\widetilde{\mathbf{u}}_{r,t}[c,m,v]|^2}{f_{m,v}(c)}
\end{aligned}$$

where $\mathbf{F}_{m,v} \coloneqq \operatorname{diag}(\mathrm{W}\mathbf{r}_{m,m,v,v})$ and $\widetilde{\mathbf{u}}_{r,t}[m,v] \coloneqq \mathrm{W}\widehat{\mathbf{u}}_{r,t}[m,v]$ and the index l is now c. □

A.6 Cross-spectral Mass Function

Proof. The diagonal vector AR model of order one in section 4.5 has the following vector MA model of order m representation

$$\widetilde{\mathbf{V}}_{r,t}[c,m] = \sum_{i=1}^{m} \left(\prod_{j=i+1}^{m} \Psi_{c,j} \right) \mathbf{W}_{r,t}[c,i], \tag{A.3}$$

where the product from $m+1$ to m is defined to be one. Therefore the element in the row v_1 and column v_2 of

$$\begin{aligned} \mathrm{E}(\widetilde{\mathbf{U}}_{r,t}[c,m_1]\widetilde{\mathbf{U}}^{*}_{r,t}[c,m_2]) &= \sum_{i=2}^{m_1 \wedge m_2} \left(\prod_{j=i+1}^{m_1} \Psi_{c,j} \right) \mathbf{\Xi}_{c,i} \left(\prod_{j=i+1}^{m_2} \Psi_{c,j} \right) \\ &+ \left(\prod_{j=2}^{m_1} \Psi_{c,j} \right) \mathbf{\Xi}_{c} \left(\prod_{j=2}^{m_2} \Psi_{c,j} \right) \end{aligned}$$

is

$$\begin{aligned} f_{m_1,m_2,v_1,v_2}(c) &= f^{1/2}_{m_1,v_1}(c) \cdot f^{1/2}_{m_2,v_2}(c) \cdot \left\{ \sum_{i=2}^{m_1 \wedge m_2} \left(\prod_{j=i+1}^{m_1} \psi_{c,j,v_1} \right) \cdot \mathbf{\Xi}_{c,i}[v_1,v_2] \cdot \left(\prod_{j=i+1}^{m_2} \psi_{c,j,v_2} \right) \right. \\ &+ \left. \left(\prod_{j=2}^{m_1} \psi_{c,j,v_1} \right) \cdot \mathbf{\Xi}_{c}[v_1,v_2] \cdot \left(\prod_{j=2}^{m_2} \psi_{c,j,v_2} \right) \right\} \\ &= f^{1/2}_{m_1,v_1}(c) \cdot f^{1/2}_{m_2,v_2}(c) \cdot \left\{ \mathbf{\Xi}_{c}[v_1,v_2] \cdot \sum_{i=1}^{m_1 \wedge m_2} \left(\prod_{j=i+1}^{m_1} \psi_{c,j,v_1} \right) \cdot \left(\prod_{j=i+1}^{m_2} \psi_{c,j,v_2} \right) \right. \\ &- \left. \mathbf{\Xi}_{c}[v_1,v_2] \cdot \sum_{i=1}^{(m_1 \wedge m_2)-1} \left(\prod_{j=i+1}^{m_1} \psi_{c,j,v_1} \right) \cdot \left(\prod_{j=i+1}^{m_2} \psi_{c,j,v_2} \right) \right\} \\ &= f^{1/2}_{m_1,v_1}(c) \cdot f^{1/2}_{m_1,v_1}(c) \cdot \mathbf{\Xi}_{c}[v_1,v_2] \cdot \left(\prod_{j=(m_1 \wedge m_2)+1}^{m_1} \psi_{c,j,v_1} \right) \cdot \left(\prod_{j=(m_1 \wedge m_2)+1}^{m_2} \psi_{c,j,v_2} \right) \end{aligned}$$

□

A.7 Estimated Multivariate Whittle Likelihood Function

Proof. Since $\mathbf{R}_v$ is a block circulant matrix it can be expressed as

$$\mathbf{R}_v = [\mathbf{R}_{m_1,m_2,v,v}] = (\mathrm{I}_M \otimes \mathrm{W})[\mathbf{F}_{m_1,m_2,v,v}](\mathrm{I}_M \otimes \mathrm{W}^*) = (\mathrm{I}_M \otimes \mathrm{W})\mathbf{F}_v(\mathrm{I}_M \otimes \mathrm{W}^*).$$

where $\mathbf{F}_{m_1,m_2,v,v} \coloneqq \mathrm{diag}(\mathrm{W}\mathbf{r}_{m_1,m_2,v,v})$ and $\mathbf{F}_v \coloneqq [\mathbf{F}_{m_1,m_2,v,v}]$. Consequently,

$$\begin{aligned}
-2\ln\mathcal{N}(\mathbf{0}, \mathbf{R}_v \mid \widehat{\mathbf{u}}_{r,t}[v]) &= \ln|2\pi\mathbf{R}_v| + \widehat{\mathbf{u}}_{r,t}[v]^*\mathbf{R}_v^{-1}\widehat{\mathbf{u}}_{r,t}[v] \\
&= \ln|2\pi\mathbf{F}_v| + \widetilde{\mathbf{u}}_{r,t}^*[v]\mathbf{F}_v^{-1}\widetilde{\mathbf{u}}_{r,t}[v] \\
&= LM\ln(2\pi) + \ln|\mathrm{P}\mathbf{F}_v\mathrm{P}^*| + (\mathrm{P}\widetilde{\mathbf{u}}_{r,t}[v])^*(\mathrm{P}\mathbf{F}_v\mathrm{P}^*)^{-1}(\mathrm{P}\widetilde{\mathbf{u}}_{r,t}[v]) \\
&= LM\ln(2\pi) + \ln|\mathbf{D}_v| + (\mathrm{P}\widetilde{\mathbf{u}}_{r,t}[v])^*\mathbf{D}_v^{-1}(\mathrm{P}\widetilde{\mathbf{u}}_{r,t}[v]) \\
&= LM\ln(2\pi) + \sum_{c=0}^{L-1}\ln|\mathbf{D}_{v,c}| + \sum_{c=0}^{L-1}\widetilde{\mathbf{u}}_{r,t}[v,c]^*\mathbf{D}_{v,c}^{-1}\widetilde{\mathbf{u}}_{r,t}[v,c]
\end{aligned}$$

where P is a permutation matrix such that $\mathbf{D}_v = \mathrm{P}\mathbf{F}_v\mathrm{P}^*$ is a block diagonal matrix with blocks $\mathrm{D}_{v,c}$ for $c = 0, \ldots, L-1$ (Rue and Held, 2005, Section 2.6.2) where the element in row m_1 and column m_2 of $\mathbf{D}_{v,c}$ is

$$f_{m_1,m_2,v,v}(c) = f_{m_1,v}^{1/2}(c) \cdot f_{m_2,v}^{1/2}(c) \cdot \left(\prod_{j=m_1+1}^{m_2} \upsilon_{v,j,c}\right)$$

for $m_1 < m_2$. □

Bibliography

Arkell, B. and Darch, G. (2006), 'Impacts of climate change on london's transport systems', *Proceedings of the ICE–Municipal Engineer* **159**(4), 231–7.

Baker, A. H., Hammerling, D. M., Mickelson, S. A., Xu, H., Stolpe, M. B., Naveau, P., Sanderson, B., Ebert-Uphoff, I., Samarasinghe, S., De Simone, F. et al. (2016), 'Evaluating lossy data compression on climate simulation data within a large ensemble', *Geoscientific Model Development* **9**(12), 4381–4403.

Baker, A. H., Xu, H., Dennis, J. M., Levy, M. N., Nychka, D., Mickelson, S. A., Edwards, J., Vertenstein, M. and Wegener, A. (2014), A methodology for evaluating the impact of data compression on climate simulation data, *in* 'Proceedings of the 23rd international symposium on High-performance parallel and distributed computing', ACM, pp. 203–214.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008), 'Gaussian predictive process models for large spatial data sets', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(4), 825–848.

Bellman, R. (1961), *Adaptive Control Process: A Guided Tour*, Princeton University Press.

Berliner, L. M. (1992), 'Statistics, probability and chaos', *Statistical Science* pp. 69–90.

Bevilacqua, M., Gaetan, C., Mateu, J. and Porcu, E. (2012), 'Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach', *Journal of the American Statistical Association* **107**(497), 268–280.

Branstator, G. and Teng, H. (2010), 'Two limits of initial-value decadal predictability in a cgcm', *Journal of climate* **23**(23), 6292–6311.

Byrne, M. P. and Ogorman, P. A. (2013), 'Land–ocean warming contrast over a wide range of climates: Convective quasi-equilibrium theory and idealized simulations', *Journal of Climate* **26**(12), 4000–4016.

Cameletti, M., Ignaccolo, R. and Bande, S. (2011), 'Comparing spatio-temporal models for particulate matter in piemonte', *Environmetrics* **22**(8), 985–996.

Castruccio, S. (2016), 'Assessing the spatio-temporal structure of annual and seasonal surface temperature for cmip5 and reanalysis', *Spatial Statistics* **18**, 179–193.

Castruccio, S. and Genton, M. (2014), 'Beyond axial symmetry: An improved class of models for global data', *Stat* **3**, 48–55.

Castruccio, S. and Genton, M. G. (2016), 'Compressing an ensemble with statistical models: an algorithm for global 3d spatio-temporal temperature', *Technometrics* **58**(3), 319–328.

Castruccio, S., Genton, M. and Sun, Y. (2019), 'Visualising spatio-temporal models with virtual reality: From fully immersive environments to apps in stereoscopic view', *Journal of the Royal Statistical Society - Series A (with discussion)* . in press, read before the Royal Statistical Society.

Castruccio, S. and Guinness, J. (2017), 'An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **66**(2), 329–344.

Castruccio, S., Huser, R. and Genton, M. G. (2016), 'High-order composite likelihood inference for max-stable distributions and processes', *Journal of Computational and Graphical Statistics* **25**(4), 1212–1229.

Castruccio, S., Ombao, H. and Genton, M. G. (2018), 'A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fmri data', *Biometrics* .

Castruccio, S., Stein, M. L. et al. (2013), 'Global space–time models for climate ensembles', *The Annals of Applied Statistics* **7**(3), 1593–1611.

Collins, M. (2002), 'Climate predictability on interannual to decadal time scales: the initial value problem', *Climate Dynamics* **19**(8), 671–692.

Collins, M. and Allen, M. R. (2002), 'Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability', *Journal of Climate* **15**(21), 3104–3109.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G. et al. (2013), 'Long-term climate change: projections, commitments and irreversibility'.

Cressie, N. and Johannesson, G. (2008), 'Fixed rank kriging for very large spatial data sets', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(1), 209–226.

Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016), 'Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets', *Journal of the American Statistical Association* **111**(514), 800–812.

Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A., Schaap, M. et al. (2016), 'Nonseparable dynamic nearest neighbor gaussian process models for large spatio-temporal data with an application to particulate matter analysis', *The Annals of Applied Statistics* **10**(3), 1286–1316.

Davis, P. J. (2012), *Circulant Matrices*, American Mathematical Soc.

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlöme, S., von Stechow, C., Zwickel, T. and Minx, J. C. (2014), *PCC, 2014: Summary for Policymakers. In: Climate Change 2014: Mitigation of Climate Change. Contribution*

of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press.

Edwards, M., Castruccio, S. and Hammerling, D. (2018), 'Marginally parametrized spatio-temporal models and stepwise maximum likelihood estimation', *arXiv preprint arXiv:1806.11388* .

Edwards, M., Castruccio, S. and Hammerling, D. (2019), 'A multivariate global spatiotemporal stochastic generator for climate ensembles', *Journal of Agricultural, Biological and Environmental Statistics* pp. 1–20.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E. (2016), 'Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization', *Geoscientific Model Development* **9**(5), 1937–1958.

Farmer, G. T. and Cook, J. (2013), *Climate Change Science: A Modern Synthesis: Volume 1-The Physical Climate*, Vol. 1, Springer Science & Business Media.

Field, C. B., Barros, V. R., Mach, K. and Mastrandrea, M. (2014), *Climate change 2014: impacts, adaptation, and vulnerability*, Vol. 1, Cambridge University Press Cambridge and New York.

Friedman, J., Hastie, T. and Tibshirani, R. (2001), *The Elements of Statistical Learning*, Springer.

Geiger, R. (1954), 'Klassifikation der klimate nach w. köppen', *Landolt-Börnstein, Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie Vol* **3**.

Gelfand, A. E., Banerjee, S. and Gamerman, D. (2005), 'Spatial process modelling for univariate and multivariate dynamic spatial data', *Environmetrics* **16**(5), 465–479.

Gillis, J. (2015), 'Short answers to hard questions about climate change', *The New York Times* **25**.

Giltinan, D. and Davidian, M. (1995), 'Nonlinear models for repeated measurement data', *Monographs on Statistics and Applied Probability (Chapman and Hall, 1995), ISBN* **772450420**.

Glenn, M., Kim, S.-H., Ramirez-Villegas, J. and Laderach, P. (2013), 'Response of perennial horticultural crops to climate change', *Horticultural Reviews* **41**, 47–130.

Golub, G. H. and Van Loan, C. F. (2012), *Matrix computations*, 3 edn, JHU Press.

Gonzalez, P., Neilson, R. P., Lenihan, J. M. and Drapek, R. J. (2010), 'Global patterns in the vulnerability of ecosystems to vegetation shifts due to climate change', *Global Ecology and Biogeography* **19**(6), 755–768.

Guinness, J. and Hammerling, D. (2018), 'Compression and conditional emulation of climate model output', *Journal of the American Statistical Association* **113**(521), 56–67.

Hamilton, J. D. (1994), *Time Series Analysis*, Vol. 2, Princeton University Press.

Hardy, Y. and Steeb, W.-H. (2010), 'Vec-operator, kronecker product and entanglement', *International Journal of Algebra and Computation* **20**(1), 71–76.

Hestenes, M. R. and Stiefel, E. (1952), 'Methods of conjugate gradients for solving linear systems', *Journal of Research of the National Bureau of Standards* **49**(6).

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K. et al. (2013), 'The community earth system model: a framework for collaborative research', *Bulletin of the American Meteorological Society* **94**(9), 1339–1360.

Jackson, J. E., Yost, M. G., Karr, C., Fitzpatrick, C., Lamb, B. K., Chung, S. H., Chen, J., Avise, J., Rosenblatt, R. A. and Fenske, R. A. (2010), 'Public health impacts of climate change in washington state: projected mortality risks due to heat events and air pollution', *Climatic Change* **102**(1), 159–186.

Jeong, J., Castruccio, S., Crippa, P., Genton, M. G. et al. (2018), 'Reducing storage of global wind ensembles with stochastic generators', *The Annals of Applied Statistics* **12**(1), 490–509.

Jeong, J., Yan, Y., Castruccio, S. and Genton, M. G. (2017), 'A stochastic generator of global monthly wind energy with tukey g-and-h autoregressive processes', *arXiv preprint arXiv:1711.03930* .

Johannesson, G., Cressie, N. and Huang, H.-C. (2007), 'Dynamic multi-resolution spatial models', *Environmental and Ecological Statistics* **14**(1), 5–25.

Jones, R. H. (1963), 'Stochastic processes on a sphere', *The Annals of mathematical statistics* **34**(1), 213–218.

Katzfuss, M. (2017), 'A multi-resolution approximation for massive spatial datasets', *Journal of the American Statistical Association* **112**(517), 201–214.

Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008), 'Covariance tapering for likelihood-based estimation in large spatial data sets', *Journal of the American Statistical Association* **103**(484), 1545–1555.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S. C., Danabasoglu, G., Edwards, J. et al. (2015), 'The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability', *Bulletin of the American Meteorological Society* **96**(8), 1333–1349.

Knuth, D. E. (1997), *The Art of Computer Programming. Volume 1: Fundamental Algorithms*, 3 edn, Addison-Wesley.

Koopmans, L. H. (1995), *The Spectral Analysis of Time Series*, Elsevier.

Lindgren, F., Rue, H. and Lindström, J. (2011), 'An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation ap-

proach', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.

Lindstrom, P. and Isenburg, M. (2006), 'Fast and efficient compression of floating-point data', *IEEE transactions on visualization and computer graphics* **12**(5), 1245–1250.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

MacKay, D. J. (2003), *Information theory, inference and learning algorithms*, Cambridge university press.

McGranahan, G., Balk, D. and Anderson, B. (2007), 'The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones', *Environment and urbanization* **19**(1), 17–37.

McIlveen, R. (1991), *Fundamentals of Weather and Climate*, Psychology Press.

Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S. and Stevens, B. (2014), 'Climate model intercomparisons: preparing for the next phase', *Eos, Transactions American Geophysical Union* **95**(9), 77–78.

Mendelsohn, R., Dinar, A. and Williams, L. (2006), 'The distributional impact of climate change on rich and poor countries', *Environment and Development Economics* **11**(2), 159–178.

Moss, R., Babiker, W., Brinkman, S., Calvo, E., Carter, T., Edmonds, J., Elgizouli, I., Emori, S., Erda, L., Hibbard, K. et al. (2008), 'Towards new scenarios for the analysis of emissions: Climate change, impacts and response strategies'.

Nelder, J. A. and Mead, R. (1965), 'A simplex method for function minimization', *The computer journal* **7**(4), 308–313.

Nocedal, J. and Wright, S. J. (1999), *Numerical Optimization*, Springer.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. and Sain, S. (2015), 'A multiresolution gaussian process model for the analysis of large spatial datasets', *Journal of Computational and Graphical Statistics* **24**(2), 579–599.

ONeill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R. and van Vuuren, D. P. (2014), 'A new scenario framework for climate change research: the concept of shared socioeconomic pathways', *Climatic Change* **122**(3), 387–400.

Patterson, H. D. and Thompson, R. (1971), 'Recovery of inter-block information when block sizes are unequal', *Biometrika* **58**(3), 545–554.

Paul, K., Mickelson, S., Dennis, J. M., Xu, H. and Brown, D. (2015), Light-weight parallel python tools for earth system modeling workflows, *in* 'Big Data (Big Data), 2015 IEEE International Conference on', IEEE, pp. 1985–1994.

Pawitan, Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press.

Pednekar, A. M., Grant, S. B., Jeong, Y., Poon, Y. and Oancea, C. (2005), 'Influence of climate change, tidal mixing, and watershed urbanization on historical water quality in newport bay, a saltwater wetland and tidal embayment in southern california', *Environmental science & technology* **39**(23), 9071–9082.

Peel, M. C., Finlayson, B. L. and McMahon, T. A. (2007), 'Updated world map of the köppen-geiger climate classification', *Hydrology and earth system sciences discussions* **4**(2), 439–473.

Perry, A. L., Low, P. J., Ellis, J. R. and Reynolds, J. D. (2005), 'Climate change and distribution shifts in marine fishes', *science* **308**(5730), 1912–1915.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic press.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., Oneill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O. et al. (2017), 'The shared socioeconomic

pathways and their energy, land use, and greenhouse gas emissions implications: an overview', *Global Environmental Change* **42**, 153–168.

Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, CRC press.

Sayood, K. (2017), *Introduction to Data Compression*, Morgan Kaufmann.

Schabenberger, O. and Gotway, C. A. (2017), *Statistical Methods for Spatial Data Analysis*, CRC press.

Serreze, M. C. and Francis, J. A. (2006), 'The arctic amplification debate', *Climatic change* **76**(3-4), 241–264.

Spall, J. C. (1989), 'Effect of imprecisely known nuisance parameters on estimates of primary parameters', *Communications in Statistics-Theory and Methods* **18**(1), 219–237.

Stein, M. L. (2012), *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M. (2013), *IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Strand, G. and Baker, A. (2018), Private Communication.

Su, H., Jiang, J. H., Neelin, J. D., Shen, T. J., Zhai, C., Yue, Q., Wang, Z., Huang, L., Choi, Y.-S., Stephens, G. L. et al. (2017), 'Tightening of tropical ascent and high clouds key to precipitation change in a warmer climate', *Nature communications* **8**, 15771.

Sun, S. and Hansen, J. E. (2003), 'Climate simulations for 1951–2050 with a coupled atmosphere–ocean model', *Journal of Climate* **16**(17), 2807–2826.

Tagle, F., Castruccio, S., Crippa, P. and Genton, M. (2017), Statistical compression of wind speed data, *in* 'AGU Fall Meeting Abstracts'.

Thornton, P. K., Jones, P. G., Alagarswamy, G. and Andresen, J. (2009), 'Spatial variation of crop yield response to climate change in east africa', *Global Environmental Change* **19**(1), 54–65.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F. et al. (2011), 'The representative concentration pathways: an overview', *Climatic change* **109**(1-2), 5.

Vavasis, S. A. (1991), *Nonlinear Optimization: Complexity Issues*, Oxford University Press, Inc.

Washington, W. M. and Parkinson, C. (2005), *Introduction to Three-Dimensional Climate Modeling*, University science books.

Webersik, C. (2010), *Climate Change and Security: A Gathering Storm of Global Challenges: A Gathering Storm of Global Challenges*, Abc-Clio.

Whittle, P. (1954), 'On stationary processes in the plane', *Biometrika* pp. 434–449.

Willems, P., Olsson, J., Arnbjerg-Nielsen, K., Beecham, S., Pathirana, A., Gregersen, I. B., Madsen, H. et al. (2012), *Impacts of climate change on rainfall extremes and urban drainage systems*, IWA publishing.

Yudin, D. B. and Nemirovskiĭ, A. S. (1977), 'Evaluation of the informational complexity of mathematical programming problems', *Matekon* **13**(3), 25–45.

Yudin, D. B. and Nemirovskiĭ, A. S. (1983), *Problem complexity and method efficiency in optimization*, Wiley.

Zhang, B., Sang, H. and Huang, J. Z. (2015), 'Full-scale approximations of spatio-temporal covariance models for large datasets', *Statistica Sinica* pp. 99–114.