UNIVERSITY OF SOUTHAMPTON

Faculty of Natural and Environmental Sciences

Ocean and Earth Science

Developing a Multi-level Gaussian Process Emulator of an Atmospheric General Circulation Model for Palaeoclimate Modelling

by

Giang T. Tran

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

March 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
Ocean and Earth Science

Doctor of Philosophy

DEVELOPING A MULTI-LEVEL GAUSSIAN PROCESS EMULATOR OF AN ATMOSPHERIC GENERAL CIRCULATION MODEL FOR PALAEOCLIMATE MODELLING

by Giang T. Tran

The study of past climates provides a unique opportunity to test our understanding of the Earth system and our confidence in climate models. The nature of this subject requires a fine balance between complexity and efficiency. While comprehensive models can capture the system's behaviour more realistically, fast but less accurate models are capable of integrating on the long timescale associated with Palaeoclimatology. In this thesis, a statistical approach is proposed to address the limitation of our simple atmospheric module in simulating glacial climates by incorporating a statistical surrogate of a general circulation model of the atmosphere into our Earth system modelling framework, GENIE.

To utilise the available model spectrum of different complexities, a multi-level Gaussian Process (GP) emulation technique is proposed to established the link between a computationally expensive atmospheric model, PLASIM (Planet Simulator), and a cheaper model, EMBM (energy-moisture balance model). The method is first demonstrated by emulating a scalar summary quantity. A dimensional reduction technique is then introduced, allowing the high-dimensional model outputs to be emulated as functions of high-dimensional boundary forcing inputs. Even though the two atmospheric models chosen are structurally unrelated, GP emulators of PLASIM atmospheric variables are successfully constructed using EMBM as a fast approximation. With the extra information gained from the cheap model, the emulators of PLASIM's 2-D surface output fields, are built at a reduced computational cost. The emulated quantities are validated against simulated values, showing that the ensemble-wide behaviour of the spatial fields is well captured. Finally, the emulator of PLASIM's wind field is incorporated into GENIE, providing an interactive statistical wind field which responds to changes in the boundary condition described by the ocean module. While exhibiting certain limitation due to the structural bias in PLASIM's wind, the new hybrid model introduces additional variations to the over-diffusive spatial outputs of EMBM without incurring a substantial computational cost.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Giang T. Tran, declare that the thesis entitled *Developing a multi-level Gaussian Process emulator of an atmospheric general circulation emulator in the context of palaeoclimate* and the work presented in it are my own and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
  - Tran, G. T., Oliver, K. I. C., Sóbester, A., Toal, D. J. J., Holden, P. B., Marsh, R., Challenor, P., and Edwards, N. R.: Building a traceable climate model hierarchy with multi-level emulators, *Adv. Stat. Clim. Meteorol. Oceanogr.*, 2, 17-37, 2016.

Signed:

Date:

Acknowledgements

I would like to express my deepest appreciation to my supervisory team at Southampton, Kevin Oliver, András Sóbester, Peter Challenor and Bob Marsh for their support, guidance and invaluable knowledge, without which this work would not have been possible. Thanks to their encouragement and enthusiasm, I have learnt and come to love oceanography, climatology and statistics, subjects which were relatively unfamiliar to me just a few years ago. I am also grateful to my supervisors at the Open University, Neil Edwards and Phil Holden, whose technical input and insightful discussions have contributed to this work immensely.

I am indebted to Jeff and Simon, whose expertise and wisdom surely have shortened my struggle with GENIE, PLASIM and Mobilis, allowing me to conduct the experiments making up this work.

I am thankful to my friends at NOC, Despo, Glaucia and Nina for sharing with me the ups and downs of a PhD student's life and for making my time here so much fun. Thanks also go to my housemates, Ryan and Dave for helping me with my writing and for being a great source of knowledge and humour. I would also like to thank Heidi, Yun and Alfie, for their constant support despite being hundreds or thousands of miles away.

And last but not least, I am grateful to my parents and my sister for giving me the freedom and the opportunities to pursue further education and for fostering and indulging my passion for science since a young age.

List of Acronyms and Abbreviations

Acronym	Description
AABW	Antarctic bottom water
ACC	Antarctic circumpolar current
AGCM	atmospheric general circulation model
AMOC	Atlantic meridional overturning circulation
BP	before present
CMIP	Coupled Modelling Intercomparison Project
D-O	Dansgaard-Oeschger
EMBM	energy moisture balance model
EMIC	Earth system model of intermediate complexity
ENSO	El Niño Southern Oscillation
ENTS	efficient numerical terrestrial scheme
EOF	Empirical Orthogonal Function
GCM	General Circulation Model
GENIE	Grid ENable Integrated Earth system modelling
GOLDSTEIN	Global Ocean-Linear Drag Salt and Temperature Equation Integrator
GP	Gaussian Process
HUM	humidity
IAM	integrated assessment model
LGM	Last Glacial Maximum
MLH	maximin Latin hypercube
MOC	meridional overturning circulation
NADW	North Atlantic deep water
PC	principle components
PCA	Principle component analysis
PLASIM	Planet Simulator
PMIP	Palaeoclimate Modelling Intercomparison Project
PPTN	precipitation
PUMA	portable university model of the atmosphere

RMSE	root mean square error
SAT	surface air temperature
SIC	sea-ice cover
SIH	sea-ice height
SST	sea-surface temperature
SVD	Singular Value Decomposition
UWN	zonal wind component
VWN	meridional wind component

Chapter 1

Introduction

1.1 Thesis outline

> The climate system is a complex and interactive system consisting of the atmosphere, land surface, snow and ice, oceans and other bodies of water, and living things. The atmospheric component of the climate system most obviously characterises climate; climate is often defined as *average weather*. Climate is usually described in terms of the mean and variability of temperature, precipitation and wind over a period of time, ranging from months to millions of years (the classical period is 30 years).
>
> *IPCC Fourth Assessment report: Climate change 2007*

The Earth's climate system, as defined above in the IPCC report (Treut et al., 2007), is not constant but fluctuates on different timescales. Its impact on human life has been documented throughout history, from the Medieval Warm Period about 1000 years ago that permitted Viking exploration of the North Atlantic region to the Little Ice Age several hundred years later. Climate change in modern time, arises from not only natural causes but also from anthropogenic forcing, is likely to affect more people now than ever before in history and so has been an active topic of discussion since the early '70s (Schneider and Dickinson, 1974). Scientists over the years have contributed to the field through various means, including theoretical studies, observations of the current climate, reconstruction of past climates from proxies and numerical modelling. The nature of the climate system makes it hard to observe long-term trends or reproduce in the labs and hence, computer experiments have been immensely useful in testing climate theory and reconciling our theoretical understanding with real world observations.

The use of computer experiments in the form of numerical models or simulators is vital in our study of the ongoing and future global environmental changes. They represent

our understanding of the Earth system and are used to produce projections of the future climate changes under different greenhouse gas emission scenarios. Our confidence in these projections depends strongly on the reliability of the models. Therefore, it is important to verify that the models being used are capable of simulating climate states which differ significantly from that of the present. The study of past climates offers a unique opportunity to achieve this (Braconnot et al., 2012; Masson-Delmotte et al., 2013). The field of palaeoclimatology challenges our theoretical understanding of the climate system as well as our ability to identify and incorporate the most important physical processes into numerical models. The study of palaeoclimate also provides an opportunity to test systematic uncertainty in the structure of models. Significant effort has been put into analysing multi-model ensembles to identify inter-model differences in both future and past climates through the Coupled Modelling Intercomparison Project (CMIP) and the Palaeoclimate Modelling Intercomparison Project (PMIP), respectively.

The last ice age is a favourable test bed for models due to the availability of good proxies (Waelbroeck et al., 2009; Bartlein et al., 2011) and well known radiative forcing and boundary conditions. Also, the changes that happened during this period, for example, the changes from the Last Glacial Maximum (LGM: 21000 years before present, 21 kyr BP) to the pre-industrial climate, are of a comparable scale to the expected future changes by the end of the 21st century (Braconnot et al., 2012). This period is the focus of the three phases of the PMIP. Furthermore, the LGM is of particular interest to climate scientists owing to the series of abrupt events observed in Greenland temperature record (10 - 60 kyr BP), namely the Dansgaard-Oeschger (D-O) oscillations (Dansgaard et al., 1993) and the Heinrich events (Hemming, 2004). It is desirable for the models to be able to capture rapid and localised changes of comparable magnitude in the system.

Different models have been designed over the years to address different scientific questions which led to the development of a model spectrum of different complexities, ranging from the simplest conceptual box models to the coupled general circulation models (GCMs) of the atmosphere and the ocean. These comprehensive coupled GCMs have high resolution and explicitly represent physical processes. While they are the standard models for making projections based on future emission pathways, their high computational cost limits their applications. Most obviously, they are not practical for multi-millennial scale studies to understand the glacial-interglacial cycles. They are also not suitable for providing a large number of runs needed to explore a wide range of future scenarios or to explore the parameter and structural uncertainties within the model. A class of models called the Earth system models of intermediate complexity (EMICs) is developed to populate the region between a box model and a coupled GCM (Claussen et al., 2002).

An EMIC, apart from having a physical core that comprises of the atmosphere and the ocean, often include several Earth system components describing, for example, the biogeochemical cycles, ice sheets and ocean sediments. They are simple enough to allow

for long-term integration and exploring sensitivities. However, they have their drawbacks due to the low spatial resolution and the use of highly parameterized processes. As a result, they sometimes fail to reproduce climatic features observed in reconstructions. In this work, an EMIC called GENIE-1 is employed. It has been shown in the past that the glacial-interglacial Greenland temperature change is underestimated by GENIE-1's simple energy moisture balanced atmosphere (Holden et al., 2010). The lack of a dynamic atmospheric circulation restricts the range of feedbacks that can be simulated in the model. A discussion on the importance of having a 3-dimensional dynamical circulation in modelling glacial climates is given in Section 1.2.

By replacing GENIE-1's atmosphere with an atmospheric GCM (AGCM), the desired feedbacks might be better simulated. However, this would lead to a significant increase in the computational expense which is undesirable. Motivated by this, we set off to create a hybrid model of the atmosphere which includes a statistical component that serves as a surrogate for processes which are captured in an AGCM called PLASIM but are missing in the current simplified atmospheric component of GENIE-1. Detailed descriptions of the two models with their respective places in the model hierarchy will be covered Section 1.3.

In this work, a technique called emulation is used to construct the surrogate for PLASIM's atmospheric processes. The resulting surrogate model, or an emulator, is a statistical approximation to the full model that is fast and computationally cheap to run. The emulator is trained on a small number of carefully designed simulation runs, and once validated, provides estimates of the output of the model at previously untested parameter combinations. Here, we focus on a specific multi-level emulation technique that utilises the available climate model hierarchy. This method is combined with a dimensional reduction technique, allowing not only scalar but also high-dimensional variables to be emulated. Chapter 2 provides a general description of all the statistical techniques employed in this work. Specific details of which will be discussed as they are used throughout the thesis.

By using a combination of these statistical techniques, emulators are constructed and validated, giving estimations of prognostic atmospheric variables which can be used for the exchange of fluxes between the atmospheric and the oceanic components of GENIE-1. In other words, processes which are missing in EMBM but resolved or parameterized in the more complex model, PLASIM, can be estimated and supplied back to GENIE-1's climate components. The new hybrid model is a combination of the energy and moisture balance atmosphere with the statistical approximation of the dynamics of PLASIM's atmosphere. This new setup brings the complexity of the framework closer to that of a fully coupled GENIE-PLASIM model without incurring a large cost in computational power and time.

This work demonstrates the emulation process, first, on a simple example of a scalar climate variable. The method is then extended to cover 2-dimensional (2-D) spatial outputs (Chapter 3). Following the result of Chapter 3, Chapter 4 expands the method further to take into account not only high-dimensional outputs but also inputs. The emulation technique is also improved upon to capture variables in the two models which are more distinctive and hence, harder to emulate. Chapter 5 demonstrates the implementation of the statistical emulators constructed in Chapter 3 and Chapter 4 into the GENIE framework. A general assessment of the impact of the newly integrated component is discussed. The conclusions and future work are presented in Chapter 6.

1.2 The effect of a dynamic atmosphere in simulating past climates

1.2.1 The last glacial cycle

During the past nearly three million years, the Earth has undergone regular alternating episodes of cool and temperate periods known as glacials and interglacials, respectively. Evidence suggested that these cycles are linked to the variation in solar insolation modulated by the Earth's orbital changes (Milankovitch cycles; Berger, 1978). The most recent glacial period began approximately 116 kyr BP following a relative minimum in the Northern Hemisphere summer insolation (Stirling et al., 1998). The cold summer in Northern continents prevented snow and ice from melting away, promoting the expansion of ice sheets. During this cold period, the land ice volume increased and reached a maximum at around 21 kyr BP when global temperature was 3 to 5°C lower than today (Masson-Delmotte et al., 2013). The onset of the Northern Hemisphere deglaciation was initiated by an increase in summer insolation at high latitudes around 19 kyr. The deglaciation of the West Antarctic Ice Sheet occurred between 14 and 15 kyr (Clark et al., 2009).

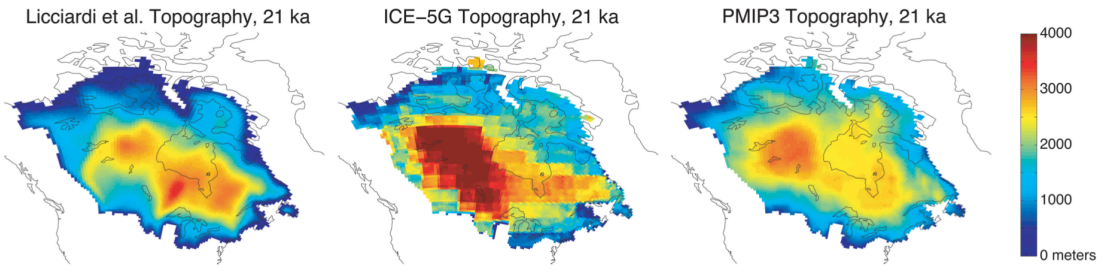


FIGURE 1.1: A comparison of different LIS reconstructions from (left to right) Licciardi et al. (1998), Peltier (2004) and PMIP3 (Abe-Ouchi et al., 2015). Figure is adapted from (Ullman et al., 2014).

Data from radiocarbon-dated moraines and other glacial deposits (e.g. Mickelson and Colgan, 2003) and results from physical models (e.g. Peltier, 2004) showed that large parts of northern Eurasia and North America were glaciated. Much of northern Europe was covered by an extensive ice sheet known as the Fennoscandian Ice Sheet, which extended longitudinally from Scandinavia in the west to central Siberia in the east and latitudinally as far south as the British Isles (Kleman et al., 2013). However, the dominant continental ice sheet at the LGM is the Laurentide Ice Sheet (LIS), which covered the entire North America continent poleward of 40°N (Clague, 1989). The spatial extent of the Laurentide is relatively well constrained from observation (Kleman et al., 2013). However, the thickness of the ice sheet remains uncertain. Several attempts at reconstruction placed the maximum elevation range between around 3000 to 4500 m (Licciardi et al., 1998; Peltier, 2004). Figure 1.1 shows the reconstructed extent and height of the Laurentide ice sheet at 21kyr BP from Licciardi et al. (1998) and Peltier (2004). Also included in this figure is the composite topography used in CMIP5/PMIP3 (Abe-Ouchi et al., 2015).

The LGM is often the target of the palaeoclimate community, e.g., the three phases PMIP1, 2 and 3 which aim at assessing the ability of a multi-model ensemble to simulate climate states that are very different from the present. Masson-Delmotte et al. (2006) have shown that while the general cooling extent and magnitude are consistently reproduced in PMIP models, the cooling in Greenland, an indication of the polar amplification, is not well captured.

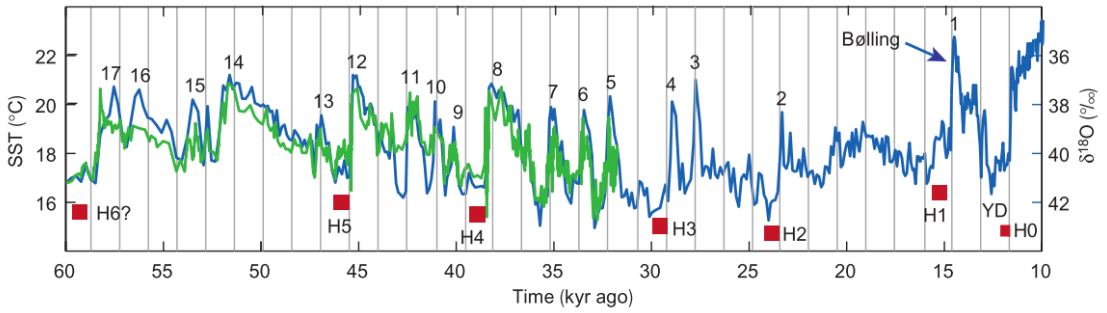


FIGURE 1.2: Temperature reconstruction from subtropical Atlantic ocean sediments (green) (Sachs and Lehman, 1999) and from the Greenland ice core GISP2 (blue) (Grootes et al., 1993) show several abrupt events between 10-60 kyr BP. 17 D-O events are numbered and the Heinrich events are marked by red squares. The vertical lines mark intervals of 1470 years. Figure is taken from (Rahmstorf, 2002)

Apart from the LGM, the last glacial cycle was also of great interest due to the many abrupt events observed. Throughout this glacial period, there were over 20 D-O events (17 of which are shown in Figure 1.2), during which, the North Atlantic region and Greenland experienced rapid warming with magnitude of 8 to 16°C over a few decades (Severinghaus et al., 1998; Stocker, 2000). Each of these warming events was followed by a more gradual cooling which occurred over several centuries creating an oscillation with

approximately 1470-year spacings or multiples thereof (Schulz, 2002). While these events are commonly associated with the North Atlantic and Greenland, evidence suggested that their effects may have been felt globally (Broecker, 2002; Voelker, 2002) with an anti-phase signal in the Southern Ocean (SO) and Antarctica (Stocker, 2000).

Also shown in Figure 1.2 are the occurrences of Heinrich events, which are characterised by distinct layers of sediments in the North Atlantic. They were thought to be transported there by a sudden discharge of icebergs from the Laurentide ice sheet. Heinrich events manifested as cold intervals with even larger amplitudes than the D-O warmings (Bond et al., 1993; Hemming, 2004). A trigger for these events needs to be able to explain the abruptness as well as the large amplitude changes. It was generally thought that variations in the strength of the Atlantic meridional overturning circulation (AMOC) are linked to these events. This hypothesis is based on both proxy data and model experiments suggesting that the AMOC exhibited threshold-like behaviour in the past and that it is a major contributor to the ocean heat transport in the North Atlantic.

It is believed that a freshwater input to high latitude North Atlantic would cause a decrease in deep-water formation there, which in turn can weaken or even shut down the AMOC. The reduced heat transport in this region means a colder climate in North Atlantic and Greenland. A stronger AMOC, on the other hand, would then lead to a warmer North Atlantic region (Broecker and Denton, 1989). Therefore a freshwater influx, thought to be from the melting icebergs, was the likely cause of the Heinrich cold events. Sediment data shows evidence of reduced North Atlantic deep water (NADW) formation during these events as would be expected (Elliot et al., 2002). However, such direct linkage has not been found for D-O events (Yu et al., 1996). A summary of the evidence of the D-O events and Heinrich events and the leading mechanisms are given by Clement and Peterson (2008).

Computer experiments with perturbations in the freshwater flux into the North Atlantic can cause a broad range of variations in the AMOC strength (Schiller et al., 1997). However, several groups of researchers have argued that a fluctuation in the AMOC alone is not sufficient to cause changes of the observed magnitude nor the suggested global extent (Elliot et al., 2002; Wunsch, 2006). Studies on the global heat transport (Trenberth and Solomon, 1994; Trenberth and Caron, 2001) showed that the atmosphere contributed the larger part of the heat budget except in the tropical region. It was demonstrated that the oceanic contribution to heat transport poleward of about 25°N falls very rapidly and by 40°N the input from the North Atlantic is less than 25% of the atmospheric contribution (Ganachaud and Wunsch, 2000). The atmosphere is also considered to be a major driver of the ocean circulation through direct input of energy in the forms of heat and momentum. Huang (1993) and Wunsch and Ferrari (2004) argued that buoyancy forcing alone cannot drive the circulation, and so the wind field must be considered if the ocean circulation is to change rapidly.

Apart from that, Wunsch (2006) and Zhang et al. (2014) both pointed out the disappearance of abrupt events after the removal of the Northern Hemisphere ice sheets, indicating a possible link between them. Mountain chains such as the Andes or the Rocky Mountains are formidable barriers to the westerly winds that influence the climate on a continental scale. The Laurentide and the Fennoscandian ice sheets are believed to have initiated a reorganisation of the atmospheric circulation pattern.

Motivated by these reasonings, we will discuss the importance of the atmospheric circulation and its interaction with the ice sheets during the last glacial period. It had been proposed that the changing wind stress in the Southern Ocean can affect the AMOC strength through the Drake passage effect (Toggweiler and Samuels, 1995). Studies of this effect have shown different results which are highly model dependent. While some models produce an apparent shift in the AMOC as wind stress in the Southern Ocean is increased or decreased (Delworth and Zeng, 2008), other studies by Wolfe and Cessi (2010) and Rahmstorf and England (1997) have shown similar behaviour but to a smaller extent. Recently, however, a review by Gent (2015) demonstrated that this is likely a result of the lack of eddy compensation in the ocean models used. Models that are eddy-resolving, eddy-permitting or having a non-constant eddy parameterization simulated eddy compensation and no effect on the AMOC. Muglia and Schmittner (2015) also showed that wind stress anomalies in Southern Hemisphere play insignificant role compared to Northern Hemisphere wind. Thus, our focus in the next section is on the reorganisation of the Northern Hemisphere wind system.

1.2.2 The influence of orographic forcing from Northern Hemisphere ice sheets on simulated climate

In meteorology, planetary waves (or Rossby waves) are planetary-scale horizontal disturbances associated with the jet stream. They are the result of the conservation of potential vorticity and depend on rotation in the presence of spherical geometry (Rossby, 1939). The phase of planetary waves propagates westward relative to the background flow. Thus, in a westerly regime, it is possible to have stationary flow with respect to the surface. Large-scale topographic features such as mountain ranges can force stationary waves through mechanical and thermal forcing. Mechanical forcing is the physical obstruction of the flow, forcing the currents to ascent over the topography. Thermal forcing refers to the anomalies in diabatic heating due to the mountains. The diabatic heating can also indirectly modify the mechanical forcing and thus alter the stationary wave patterns (Chen and Trenberth, 1988). A detailed review of the role of orographically induced stationary waves is given by Held et al. (2002). Other studies on the contributions from diabatic and orographic forcing in modern climate can be found in Chang (2009) and Sinha et al. (2012).

Apart from the cooling effect associated with high albedo and lapse rate, due to their large extent and high elevation, the Laurentide and the Fennoscandian ice sheets at the LGM represented significant obstacles to Northern Hemisphere mid-latitude westerlies. Studies using different models and ice sheet reconstructions have repeatedly demonstrated that the Northern Hemisphere ice sheets are one of the dominant factors in altering the large-scale atmospheric circulation (Justino et al., 2006; Li and Battisti, 2008; Langen and Vinther, 2009; Pausata et al., 2011; Hofer et al., 2012).

The ice sheet orography produced significant effects on not only stationary waves but also transient eddies. Both of them are relevant to the poleward fluxes of temperature, moisture, energy and angular momentum. Therefore, they must have induced significant changes in the baroclinicity of the mean flow, causing deviation from zonal mean climate (Lindemanm and Oerlemans, 1987; Roe and Lindzen, 2001; Liakka and Nilsson, 2010). The resulting topographically forced vertical winds may reorganise precipitation patterns (Roe and Lindzen, 2001; Roe, 2005; Hofer et al., 2012) which in turn affect the snow accumulation and ablation of the ice sheets.

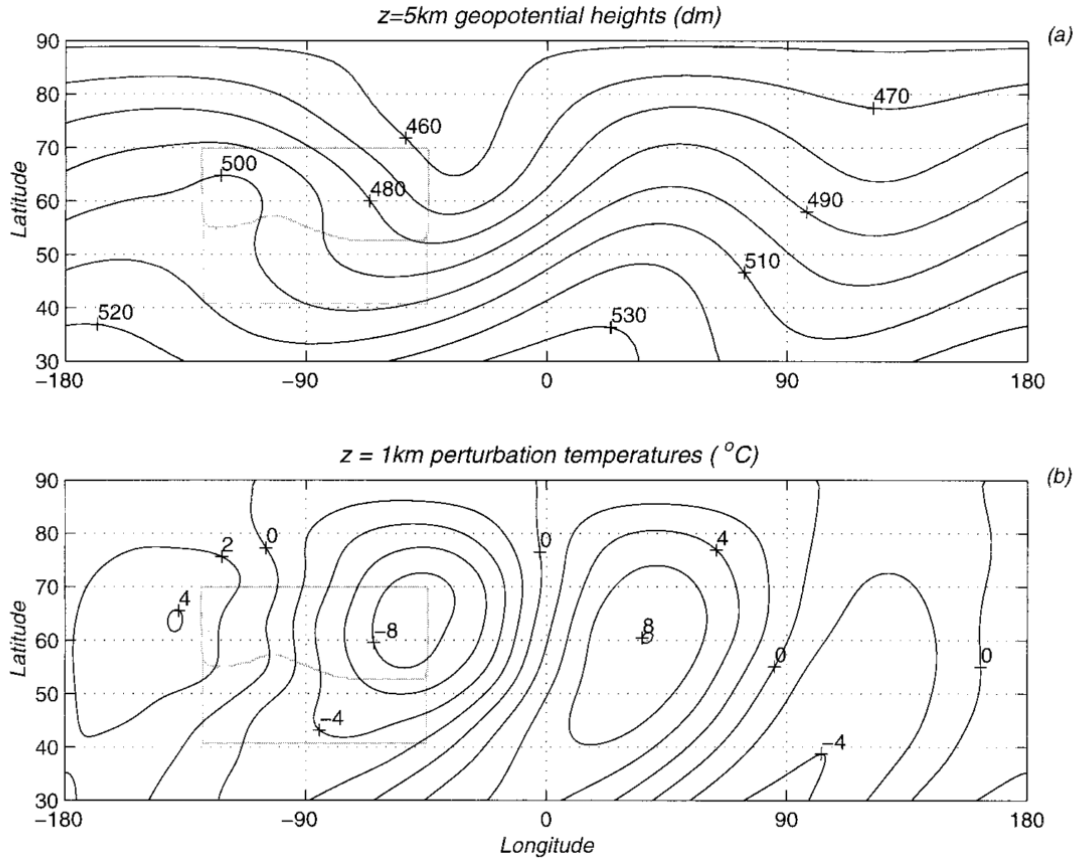


FIGURE 1.3: The perturbations from zonal mean in a) geopotential heights and b) temperature due to the stationary wave fields induced by an ice sheet (outlined in grey). Both fields show significant deviations from the zonal mean in the vicinity and downstream of the forcing region. This figure is taken from (Roe and Lindzen, 2001).

Using a simplified coupled ice sheet-stationary wave model, Roe and Lindzen (2001) ran several simulations with ice sheet-atmosphere feedback in order to study the coupling between them. By including the atmospheric feedback, a high-pressure anticyclonic circulation was observed over the western portion of the ice sheet (Figure 1.3). This anticyclonic circulation brought warm southerly air to the west bank of the ice sheet. The rest of the ice sheet, however, was still dominated by a northerly flow of air, which lowers the temperature to well below the zonal average. This differential warming/cooling pattern leads to a dramatic change in the equilibrium shape of the ice sheet. Another consequence of this circulation is the large East-West temperature difference of about 10°C. The large-scale influence due to ice sheet orography extended to North Western Europe, affecting the evolution of the Fennoscandian. A subsequent study by Liakka et al. (2012) obtained similar feedback between the ice-sheet induced stationary wave and the ice accumulation but with a more local temperature anomaly due to shorter stationary wavelength.

Another important effect of the reorganised circulation is its effect on the characteristics of the jet stream. The strength and orientation of the Atlantic jet guide the mid-latitude storm tracks and thus has a significant impact on the climate of the North Atlantic and Europe. Model results so far have demonstrated significant discrepancies on the intensity and the main tilt of the Atlantic jet. For example, while in PMIP3, the models generally produced a meridionally tilted Atlantic jet, PMIP2 simulations, on the other hand, showed an increased and more zonal Atlantic jet (Ullman et al., 2014). The most likely explanation for this discrepancy is the different ice sheet reconstructions used in these studies. Various studies comparing ice sheets of different heights found that a strengthened and nearly zonally oriented atmospheric jet stream is observed as the elevation of the LIS increases (Kageyama et al., 2006; Li and Battisti, 2008; Pausata et al., 2011; Löfverström et al., 2014; Merz et al., 2015; Löfverström et al., 2016). A stronger and more zonal jet resulted in a southward displacement of the storm track, leading to increased precipitation across eastern North Atlantic (Hofer et al., 2012). Consistently, the same models forced by lower ice sheets simulate a weaker and more meridionally tilted Atlantic jet (Kageyama et al., 1999; Li and Battisti, 2008; Ullman et al., 2014; Merz et al., 2015).

Among studies showing a southward shift and zonalisation of the jet stream under LGM condition, different conclusions are arrived at as to which mechanism drives this process. Laîné et al. (2009) and Rivière et al. (2010) attributed this feature to more frequent cyclonic wave breaking events as a result of stronger baroclinicity at LGM. During cyclonic wave breaking events, the eddy momentum flux convergence is shifted equatorward, displacing the jet in the same direction. Löfverström et al. (2016) observed the same behaviour of the Atlantic jet but proposed a different mechanism involving nonlinear Rossby wave reflection in the Atlantic. When the wave breaks, it can be absorbed to the mean flow or reflected back into the midlatitudes. In this study, it is

found that increasing LGM ice sheet height favours planetary wave reflection, which becomes more common in the subtropical North Atlantic leading to the zonalisation of the jet.

Through its impact on the wind field, surface temperature and the hydrological cycle, it's possible for the Northern Hemisphere ice sheets to affect the ocean circulation significantly. Experiments using coupled ocean-atmospheric models have shown that a higher LIS can result in a stronger AMOC (Ullman et al., 2014; Zhang et al., 2014; Muglia and Schmittner, 2015). The surface wind stress directly drives strong currents in the oceanic surface layer, transporting momentum from the atmosphere to the ocean by friction. The horizontal Ekman flow induced by wind stress carries the saline water from the lower latitudes (Timmermann and Goosse, 2004). Oka et al. (2001) examined the wind-stress driven horizontal transports (between 30°N and 60°N) and the vertical transport at high latitudes and concluded that the resulting salt transport is vital to the maintenance of the NADW by preventing the formation of the halocline. With increased wind stress forcings, the vertical salt transport increases leading to a stronger convection in the mid- to high-latitudes of the Atlantic and thus a more vigorous AMOC is expected. In a study by Yang et al. (2016), the effect of wind stress scaling on the strength of the AMOC is examined, showing this mechanism in the opposite case in which the AMOC is weakened by reducing wind stress forcing. Muglia and Schmittner (2015) simulates a stronger AMOC response under LGM boundary conditions as a result of the enhanced horizontal northward salt transport due to intensified gyre circulation. The wind also influences ocean transport in deeper layers through Ekman pumping. In regions where the Ekman transports converge (or diverge), vertical flow develops to remove (or replace) water masses resulting in downwelling (or upwelling). This process ventilates the deeper layers and thus affects deep-water formation (Killworth, 1983; Schiller et al., 1997).

It is clear from the evidence discussed above that the interaction between the atmospheric circulation and the Northern Hemisphere ice sheets can play a central role in controlling the Northern Hemisphere climate, especially in the North Atlantic. Studies focusing on its link to the D-O events have drawn connections between a stable zonal Atlantic LGM jet to the suppression of climatic feedbacks underlying the millennial-scale warming events (Seager and Battisti, 2007).

Sea ice plays an important role at high-latitudes through various feedbacks with the ocean and the atmosphere. Apart from its direct influence on the radiation budget through high albedo, sea-ice affects the transfer of heat, momentum and freshwater between the atmosphere and the ocean (Lohmann and Gerdes, 1998). It has been speculated that rapid variations in sea-ice cover in the North Atlantic played an important role in explaining abrupt events through feedbacks with the atmosphere and the ocean (Gildor and Tziperman, 2003). Because sea ice export is accomplished through near surface Ekman drift, changes in wind stress due to the ice sheets may modify the response of sea ice distribution over North Atlantic. Zhang et al. (2014) simulated weaker zonal wind

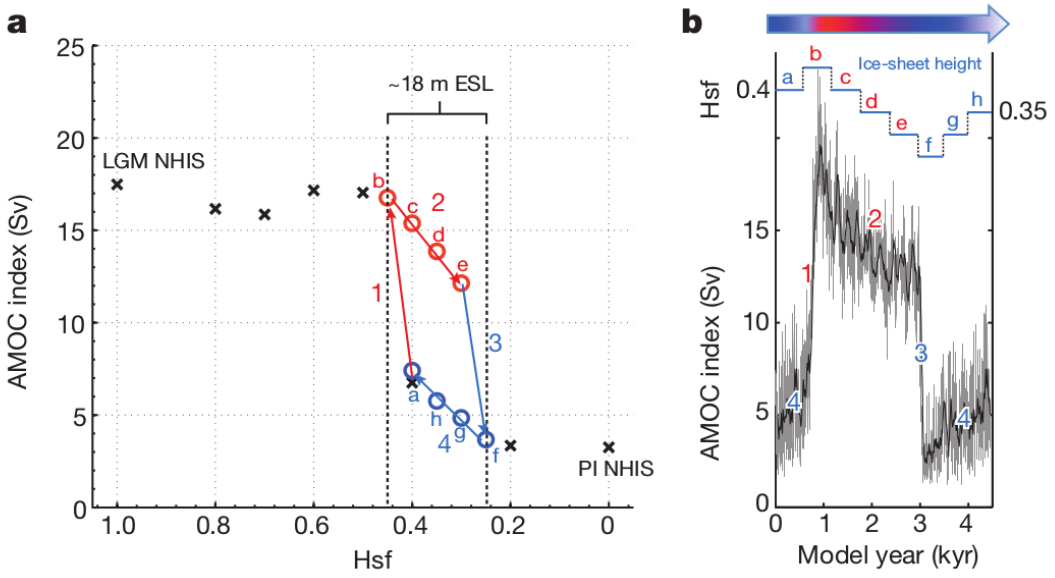


FIGURE 1.4: The figure shows a) the response of the glacial AMOC strength to changing ice sheet heights and b) The evolution of AMOC index in time and the corresponding stages of ice sheet heights observed in Zhang et al. (2014). Hsf is the orography scaling factor with 0 and 1 corresponding to the preindustrial and LGM ice sheets respectively. The hysteresis occurs at intermediate heights of the Northern Hemisphere ice sheets.

stress over the South Labrador Sea, weakening the export of sea ice to the Northeastern North Atlantic. This reduces the sea-ice cover and results in surface warming over the region, permitting open ocean convection and formation of NADW. At the same time, a strengthened wind-stress curl accelerate the North Atlantic gyres thus intensifying the northwards advection of heat and salinity. Increased ventilation of the warm subsurface waters encourages convection, further reduces sea ice in the region. Figure 1.4 shows the AMOC hysteresis in response to variations in the Northern Hemisphere ice sheet heights. The abrupt increase (and decrease) in AMOC associated with branch 1 (and 3) are attributed to the positive atmosphere-ocean-sea ice feedback under an intermediate height of the ice sheets.

This is by no mean an exhaustive discussion of all the possible interactions and feedbacks involved or initiated by the presence of the Northern Hemisphere ice sheets. Nevertheless, it is undeniable that in order to improve our understand of the last glacial/deglacial climates and to more accurately simulate the local responses in Greenland and the North Atlantic, the dynamical response of the atmosphere to ice sheets orography must be taken into consideration. With this in mind, we now move on to the discussion of the existence of models of various complexities and how our two models of interest, GENIE-1 and PLASIM fit in the bigger picture.

1.3 Model hierarchy

1.3.1 Climate models of different complexities

When working with a complex system such as the Earth's climate system, it is often beneficial to relate its behaviour to that of simplified versions of the same system. Each member of the model hierarchy of increasing complexity was designed for a particular purpose and hence, often focuses on different key aspects of the system. Ideally, by understanding how the dynamics of the system can change as each source of complexity is added or removed, we can improve our knowledge of the system as a whole. If the model hierarchy can be utilised effectively, we can learn more about the system by relating the different levels of model complexity to identify common and robust features as well as discrepancies in their structures that need addressing. However, this is not straightforward in real life due to the inherently chaotic nature of the climate system and the lack of a coherent hierarchy. Held (2005) discussed the need for model hierarchies in the field of climate theory and the fact that gaps between comprehensive and more idealised models have not been successfully closed.

The complexity of a model is a combination of many factors. Henderson-Sellers and McGuffie (1987) classified the model spectrum pictorially as a pyramid with complexity depending on the interaction between major processes represented in the atmospheric model (Figure 1.5). With the increases in model resolution and the number of interactive components accompanying advances in computational power, an updated description of the spectrum was needed. Claussen et al. (2002) proposed a crude representation of the model spectrum (Figure 1.5) in an attempt to define the complexity of an EMIC with respect to a conceptual and a GCM (comprehensive) model. The complexity is now described in terms of the number of interacting components, the cumulative dimension of the atmosphere and ocean, and the order of magnitude of grid cells.

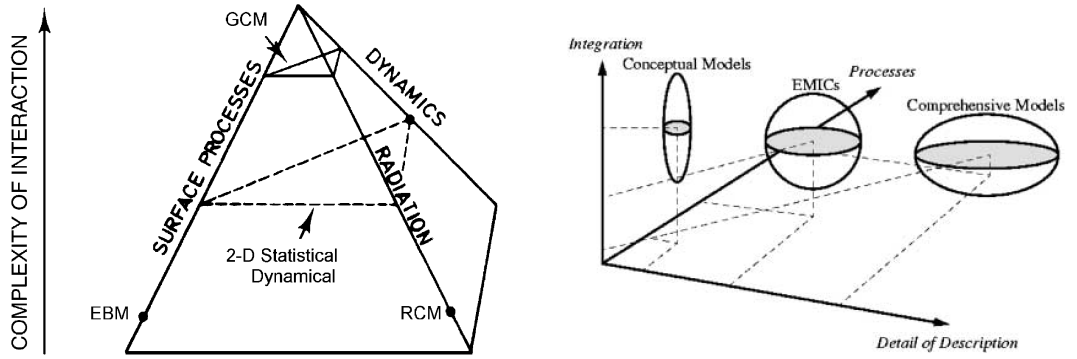


FIGURE 1.5: Two views on the spectrum of models of different complexities as proposed by Henderson-Sellers and McGuffie (1987) (left) and Claussen et al. (2002) (right). The diagram is taken from Claussen et al. (2002).

A typical idealised box model is located at one corner of the spectrum, characterised by the simplified processes at low resolution and the large possible number of interacting components. On the opposite end, a comprehensive coupled atmosphere-ocean-land GCM (AOL-GCM) is located. This type of model tries to simulate the system by capturing as much of the dynamics as possible, such as those used in the CMIP5 (Taylor et al., 2012). GCMs solve the fundamental equations of motion, together with the evolution of temperature, moisture (atmosphere) or salinity (ocean). They also contain parameterizations of unresolved subgrid-scale processes. Historically, due to their high resolution and explicitly resolved processes, a GCM was often limited to only having the core physical components. They have become increasingly more sophisticated and started to include more components representing the Earth system, such as the carbon cycle, vegetation or the ice sheet.

Despite the tendency for comprehensive models to grow towards Earth system models, simple EMICs remain an important part of the spectrum. Typically run at lower spatial and temporal resolutions and have more parameterized processes, EMICs are much faster to integrate and often do not require the use of expensive supercomputers. They are suitable for simulations to study glacial-interglacial cycles even with a large number of interacting components. Processes which operate on very long time scales such as chemical weathering can only be integrated in EMICs. Also, being faster and cheaper to run, EMIC frameworks are often used to develop and test new subcomponents such as continental ice sheets or carbon cycles before they are implemented into more complex models. Weber (2010) provided a general review of the several approaches to building an EMIC and highlighted the need to address research questions with a wide range of models to investigate the shortcoming or bias in a certain model.

Most of the studies discussed in the previous section employed comprehensive AGCMs or AOGCMs. When such a comprehensive coupled GCM model is used, it is often run using a specific parameter set, calibrated to match observations of modern climate or reconstruction of past climate. Their computational cost inhibits studies to explore the parametric uncertainty caused by non-optimal calibration of model parameters. An EMIC, on the other hand, can be used to generate large ensembles to assess parametric uncertainty thoroughly. Potentially, with a carefully chosen hierarchy, the commonality between different levels can be utilised to infer the parametric uncertainty in the computationally expensive models (Williamson et al., 2012).

Several intercomparison projects have been conducted using EMICs (Plattner et al., 2008; Eby et al., 2013) and both EMICs and AOGCMs (Stouffer et al., 2006). Figure 1.6, adapted from Stouffer et al. (2006) shows the SAT anomalies over Greenland and the northern North Atlantic of an AOGCM and an EMIC ensemble as a result of a hosing experiment. The magnitude and the distribution of the cooling due to a weakened AMOC vary considerably among different models. However, in general, AOGCMs simulate a more pronounced cooling and sharper features compared to EMICs. The

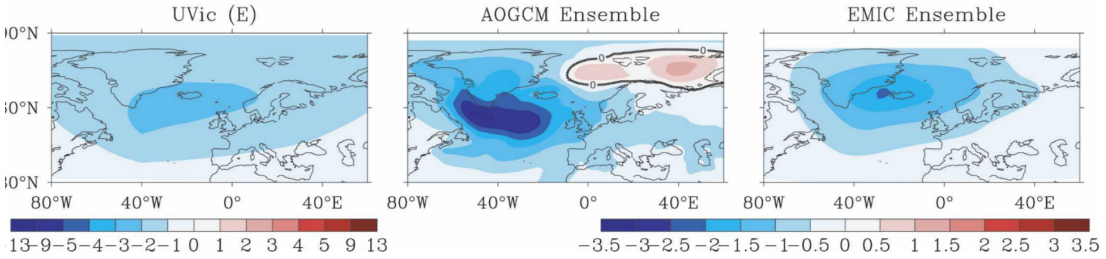


FIGURE 1.6: The cooling in Greenland and North Atlantic due to a weakened AMOC as simulated by UVic, an AOGCM ensemble and an EMIC ensemble, taken from Stouffer et al. (2006).

SAT anomalies south of Greenland as a function of AMOC strength anomalies display two linear trends with a smaller gradient fit corresponding to the EMIC ensemble. Furthermore, the AOGCM ensembles simulate a warming over the Barents and Nordic Seas due to a shift in the location of deep water formation and feedback with sea ice. The cooling signals over the North Atlantic and Pacific hinted at possible effect from an altered planetary wave pattern. Figure 1.6 also shows the cooling pattern from a single EMIC, the University of Victoria (UVic) Earth system climate model (Weaver et al., 2001).

An EMIC represents a trade-off between efficiency and fidelity. The simplified components used are sometimes structurally inadequate to simulate important aspects of the system. As discussed in the previous section, the interaction between a dynamic atmospheric circulation and the Northern Hemisphere ice sheets during the last glacial period and their feedbacks with the rest of the system might play an essential role in explaining past climatic events.

In this work, the two models GENIE-1 and PLASIM are chosen, influenced by our aim to simulate glacial climate, especially the last glacial period. GENIE-1 and PLASIM are both low-resolution models and the fast integration time allows large ensembles to be produced. The models are described in the next section and will be revisited in Section 3.2.

1.3.2 GENIE-1 and PLASIM

The Grid-ENabled Integrated Earth system modelling (GENIE) framework (Lenton et al., 2006, 2007) is designed to be flexible, with the option to run at various levels of complexity. This includes i) the number of interacting components, e.g. ocean biogeochemistry, atmospheric chemistry and ocean sediments; ii) several horizontal and vertical resolutions and iii) two levels of atmospheric model fidelity. The scientific drivers for the construction of GENIE are to study climate changes on long timescales, especially the last glacial-interglacial cycle. By employing components with low spatial resolution

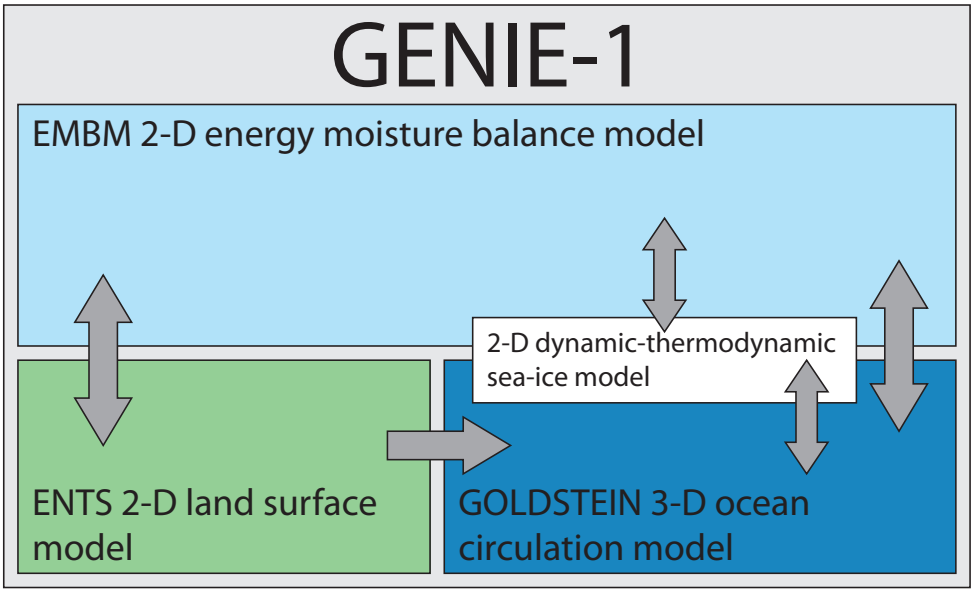


FIGURE 1.7: Schematic of the GENIE-1 model employed in this work. The arrows indicate the direction of the flux exchanges between the different modules.

and simplified physics, the framework is fast enough to perform simulation on 10^3-10^6 year timescales.

In this work, the GENIE configuration used is shown in Figure 1.7, which includes a seasonal 2-D energy-moisture balance atmosphere (EMBM), a 3-D frictional geostrophic ocean (GOLDSTEIN), a dynamic-thermodynamic sea ice and a terrestrial carbon and land surface model (ENTS) run on a 64×32 horizontal mesh. A detailed description of the atmosphere-ocean-sea-ice core components can be found in Edwards and Marsh (2005); Marsh et al. (2011) and the terrestrial scheme can be found in Williamson et al. (2006). This particular configuration is often referred to as GENIE-1 and will be covered again in Section 3.2.

The energy and moisture balance model (EMBM) of the atmosphere is based on that of the UVic (Fanning and Weaver, 1996; Weaver et al., 2001). Similar to UVic and many other models in the class of EMICs, GENIE-1 has a relatively low level of internal variability. Figure 1.8 shows typical behaviour of GENIE-1's annual mean Northern Hemisphere surface air temperature and the strength of the Atlantic overturning circulation. The 2-D atmosphere does not resolve synoptic activity and hence does not generate 'noise' to excite multidecadal oscillations in the ocean circulation. As a result, the quasi-steady-state time series for both quantities appear to be smooth.

Studies conducted using GENIE-1 have illuminated drawbacks from the use of a simple atmosphere such as EMBM. Similar to the behaviour of UVic and the EMIC ensemble shown in Figure 1.6, the spatial distribution of climate fields in GENIE-1 tends to

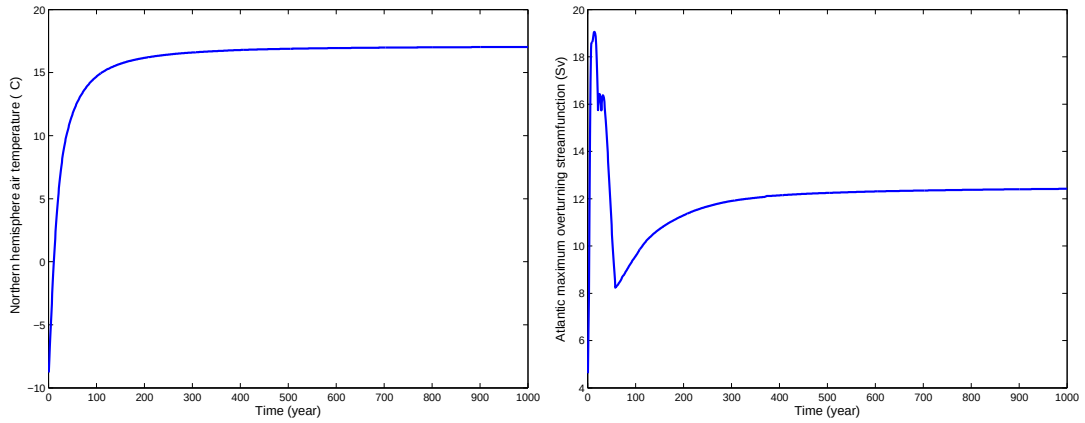


FIGURE 1.8: The evolution of the mean Northern Hemisphere surface air temperature (left) and the strength of the Atlantic overturning circulation (right) in GENIE-1. Both timeseries show annual mean values over a 1000 year integration from a uniform initial state.

be more diffusive than those obtained by an AOGCM. This is the case for both surface temperature fields and the precipitation field. The over-diffusive atmosphere can lead to a lack of distinct desert regions due to restricted transport of moisture to the continental interior and Northerly locations (Marsh et al., 2011). The equator-to-pole temperature difference tends to be small, and the tropics are too cold because too much heat is diffused away from the equator (Lenton et al., 2006). Holden et al. (2010) found similar characteristics between a GENIE-1 ensemble and the PMIP2 multimodel ensemble under LGM forcing (Masson-Delmotte et al., 2006). However, GENIE-1 exhibits lower variability, most likely due to the lack of dynamic 3-D atmospheric structure and the use of a globally constant outgoing longwave radiation parameterization. Lunt et al. (2006) conducted a series of equilibrium and transient integrations with GENIE-1 over a period from 30 kyr BP to pre-industrial period and showed that the first-order response of GENIE-1 to large-scale boundary condition changes (orography and ice sheets) is in good agreement with that from an AGCM coupled to a slab ocean (HadSM3). However, when it comes to comparing models to palaeo data, the spatial distribution of climatic variables is more relevant than the global annual mean values and GENIE-1 cannot simulate complex and localised changes related to atmospheric circulation, cloud cover, or precipitation.

Motivated to improve the model's ability to capture a more diverse range of climate state under glacial boundary conditions and to reduce some of the biases caused by the diffusive nature of EMBM, this work is focused on developing a hybrid model with embedded statistical components that emulate the behaviour of a 3-D dynamical atmosphere. Here we employ the multi-layer spectral model of the Planet Simulator (PLASIM) developed at the University of Hamburg (Fraedrich et al., 2005; Fraedrich, 2012). PLASIM consists of a fully dynamical 3-D atmosphere based on the Portable University Model of

the Atmosphere (PUMA), coupled with a 2-D mixed layer ocean. PUMA is based on the multi-layer spectral model by Hoskins and Simmons (1975). The atmosphere of PLASIM, referred to throughout this work simply as PLASIM, is a model of intermediate complexity designed to be a simple AGCM which is fast enough to run on a typical workstation.

Compared to GENIE-1, the PLASIM atmosphere has ten levels in a terrain-following coordinate system. The model solves the set of moist nonlinear primitive equations consisting of prognostic equations for horizontal divergence, vorticity, temperature, specific humidity and logarithm of surface pressure. The parameterizations for unresolved processes include long and shortwave radiation, interactive clouds, moist and dry convection, large-scale precipitation, boundary layer fluxes of latent and sensible heat and vertical and horizontal diffusion. Just like GENIE-1, PLASIM can be run with prescribed orography, allowing different ice sheets reconstructions to be employed as boundary conditions. PLASIM exhibits more internal variability than EMBM and can simulate the interactions between the ice sheets and the atmospheric circulation. For example, the atmospheric core PUMA has been used to study the nonlinear effect of the ice sheets have on stationary waves (Liakka et al., 2012) in a coupled experiment with an ice sheet model.

We can achieve 35 years of model integration using PLASIM with given boundary conditions or 5000 years of simulation with GENIE-1 in 8 CPU hours on a 64-bit processor. Therefore, simulations using GENIE-1 or PLASIM individually can be performed on a typical workstation, and large ensembles can be obtained by distributing the members on a computer cluster or a supercomputer. A full coupling between GENIE-1's ocean and sea ice with PLASIM's atmosphere, however, would require 1 to 2 months to perform a 5000-year run on the same machine. Therefore, we adopted the approach of using a fast surrogate model instead of PLASIM.

The next chapter will introduce the statistical techniques employed, namely the Gaussian Process emulator and principal component analysis. The multi-level emulation approach is also proposed as a mean to exploit information gained from EMBM to aid the emulation of PLASIM. Hence this work provides an example where models of different complexities can be used to inform one another.

Chapter 2

Predicting model results using emulators

2.1 Emulators

2.1.1 Gaussian process emulator

Computer models or simulators are employed in many scientific and technological fields. They represent our understanding of a complex system through sophisticated mathematical models, written in large computer codes. In areas like climatology or oceanography, the complexity of the system often leads to the development of increasingly sophisticated simulators. This can become a problem when many runs at different input values are required to assess a model's behaviour. This happens when we want to learn how the simulated climate changes as one or more parameters are perturbed (sensitivity analysis) or to determine an optimal set of parameters, out of the vast number of possibilities, which allows the model to reproduce physical observations closely (calibration). Historically, standard Monte Carlo-based methods have been employed in the case of sensitivity analysis (reviewed by Saltelli et al., 2000). This method typically requires thousands of model runs. Due to the large computational cost and the amount of time taken to perform one simulation, Monte Carlo-based methods are unsuitable for a majority of climate models including GENIE-1 and PLASIM.

Although the simulators are often deterministic such as in our case, the Bayesian paradigm treats the outputs as random variables, drawn from a distribution that is conditioned by our prior knowledge. Due to the complexity of the code, before the computational task is completed, the outcome is uncertain (de Finetti, 1974). As more simulations are performed, at different values of inputs, we learn more and more about the simulator's behaviour. If a sufficiently large amount of runs can be achieved, the uncertainty about the model's response becomes arbitrarily small. Since in practice, we

are often limited to a certain number of runs, emulators are employed to make inferences about the simulator. Statistical emulation, also referred to as meta-modelling, offers a significant gain in terms of accuracy and computational time over standard Monte Carlo-based methods. An early overview of the application of emulators in climate science can be found in Hankin (2005).

Quite often, computer codes are assumed to be Gaussian Processes. A Gaussian Process is a generalisation of the Gaussian distribution. While the Gaussian distribution is over vectors, a GP is over functions. Just like a Gaussian distribution can be defined by a mean and a variance, a GP is fully specified by a vector of mean function and a matrix of covariance function. In this thesis, our interest is focused on the GP emulator, also known as kriging (Rasmussen and Williams, 2006; Forrester et al., 2008), and a multi-level extension to this method, referred to as co-kriging (Kennedy and O'Hagan, 2000; Forrester et al., 2007; Cumming and Goldstein, 2008). An emulator, once constructed and validated, can be used as a surrogate of the simulator to produce estimates of the output at new parameter combinations (input points). The advantage of using the GP emulator is that the curve fits through the known points (training points from model runs at predefined sets of input parameters) and an estimated uncertainty is obtained for each emulated point.

To emulate a single summary quantity of the simulation outputs, for example, the global mean SAT, the assumptions made are as follows:

- *The model output is a smooth function of its inputs.*
- *The model can be represented as a GP.*
- *Each emulator is concerned with a single deterministic scalar output.*

While the model output is treated as a random variable and is uncertain before the run is completed, if the code has been run before at a slightly different point in the input space, we assume that this particular run output will not be too different from that of the other run. The 'smooth' assumption is valid when this the case. In climate models, this is a reasonable assumption in most situations, perhaps, except in the region of a bifurcation.

The climate model, $f(\cdot)$, can be viewed as a function of a set of inputs, $\mathbf{x} = [x_1, \cdots, x_d]$, where d is the number of perturbed model parameters. This number is commonly referred to as the number of dimensions of the emulator. The output of each model run is a scalar value y. Supposed we have n simulation runs, providing n realisations $\mathbf{y} = [y_1 = f(\mathbf{x}_1), \cdots, y_n = f(\mathbf{x}_n)]$. These comprise the training set used to train an emulator.

First, the function $f(\cdot)$ is represented by a GP prior described by a mean function $m(\cdot)$ and a covariance function $V(\cdot,\cdot)$

$$f(\cdot)|\beta,\sigma^2,\delta \sim \mathcal{N}(m(\cdot),V(\cdot,\cdot)). \tag{2.1}$$

This GP is used as a prior for Bayesian inference. The prior does not depend on the training data but specifies the assumptions we've made above. Then, the outputs from a selected number of simulations are incorporated, allowing us to update the prior to the posterior GP. This process is called training the GP model. Following Kennedy and O'Hagan (2001), $m(\cdot)$ and $V(\cdot,\cdot)$ are modelled hierarchically, meaning that they are parameterised in terms of hyperparameters. The mean function is given by:

$$m(\mathbf{x}) = h^T(\mathbf{x})\beta, \tag{2.2}$$

where $h(\mathbf{x})$ is a vector of known regression functions of the inputs, describing a class of shapes of the function $f(\cdot)$. β is an unknown vector of coefficients. In the case of ordinary kriging, $h^T(\mathbf{x}) = \mathbf{1}$, making β the unknown overall mean. A variation of kriging, called universal kriging, uses a linear mean function:

$$h^T(\mathbf{x}) = (1,\mathbf{x})^T, \tag{2.3}$$

where $h^T(\mathbf{x})$ is a $(s \times 1)$ vector with $s = d+1$. Then

$$m(\mathbf{x}) = \beta_1 + \beta_2 x_1 + \cdots + \beta_{d+1} x_d, \tag{2.4}$$

The coefficients $[\beta_2, \beta_{d+1}]$ now describe the expected trend of the simulator in response to each input.

The covariance function is given by:

$$V(\mathbf{x},\mathbf{x}') = \sigma^2 \Psi(\mathbf{x},\mathbf{x}'). \tag{2.5}$$

in which, σ^2 is an unknown variance of the GP and $\Psi(\cdot,\cdot)$ is the assumed correlation function:

$$\Psi(\mathbf{x},\mathbf{x}') = \exp\left[-\sum_{j=1}^{d} 10^{\theta_j}\left|x_j - x_j'\right|^{p_j}\right]. \tag{2.6}$$

The function Ψ represents the correlation between pairs of points, which is assumed to be stationary and continuous, that is, it only depends on the distance between the pair of inputs, $(\mathbf{x}-\mathbf{x}')$. This power exponential form of covariance structure is a popular choice due to its flexibility.

The value of Ψ depends on the correlation parameters p and θ, referred to as hyperparameters. Figure 2.1 demonstrates how varying these hyperparameters could affect the shape of the correlation function. In general, this correlation structure depends on the absolute distance between the sample points $x_j^{(i)} - x_j$, points which are closer together are more strongly correlated. The hyperparameters indicate the manner in which this correlation falls with distance.

The first hyperparameter, p, is sometimes referred to as a 'smoothness' parameter of the correlation function. The value of p can be anything between 0 and 2 with the most commonly used values being 1 and 2. The effect of p on the correlation function can be seen in Figure 2.1. At $p = 2$, Ψ is very smooth and is referred to as the squared exponential (or Gaussian) correlation function (Diggle and Ribeiro Jr, 2007). As p decreases, the correlation drops faster with increasing distance, and there is a discontinuity between $y(x_j^{(i)})$ and $y(x_j)$.

The second hyperparameter, θ is the inverse correlation length parameter, defining how quickly the correlation between the simulator outputs at two input points declines as the distance between them increases (Figure 2.1). This hyperparameter indicates the activity of the function in the corresponding dimension.

Both p and θ can be estimated for each dimension. For simplicity and to reduce computational cost, p is assumed to be the same for all dimensions. The value of $p = 2$ can lead to very ill-conditioned covariance structure and hence numerical instability in computing the matrix inversion. For this reason, we fix the value of p to 1.99.

The exponential power form of covariance structure used here is a common choice due to its flexibility. Its assumption of stationarity might fail, for example, when there is a bifurcation in the system. The covariance specified, however, provides a weak prior and as more training points are used, it contributes less to the final emulator. An independent value of θ is obtained for each dimension by maximising the likelihood of $\mathbf{y}$.

The specified GP is used as a prior for Bayesian inference and is parameterised in terms of the hyperparameters β, σ^2, θ and p. Given that the prior is Gaussian, by analytically marginalising β and σ^2, the marginal likelihood of the observed outputs at n training points, $\mathbf{y}$, given θ and p can then be computed (estimated by maximising the likelihood of y). A more detailed description of the derivations and formulations can be found in Mardia and Marshall (1984). The estimated θ_j in kriging and β_{j+1} in universal kriging indicate the relative activity in the j^{th} corresponding dimension. Very low values of these hyperparameters imply inactive inputs. The dummy parameter, FFX, is included to verify that the emulator is doing a good job at identifying inactive inputs.

Prior beliefs about the model behaviour are combined with observations from training points to produce a posterior distribution for the model. Having obtained estimates

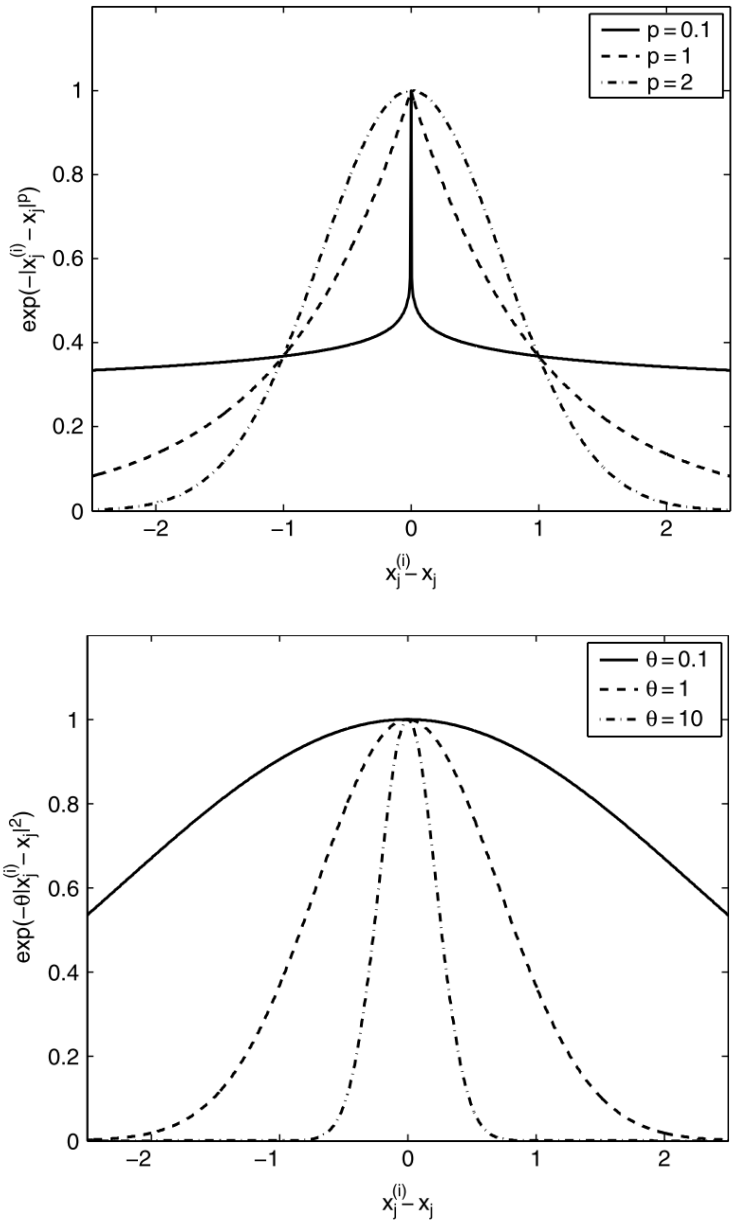


FIGURE 2.1: The effect of varying p (upper) and θ (lower) on correlations. In both plots, $x_j^{(i)} - x_j$ indicates the separation between a pair of points. The plots are taken from (Forrester et al., 2008)

for θ and p, the posterior distribution found can be used to make predictions about the model's outputs at unsampled inputs. The predictive distribution is a Student's t-distribution, with $n - s$ degrees of freedom

$$p(f(\mathbf{x})|\mathbf{y}, \theta) \sim t_{n-s}(m_1(\mathbf{x}), V_1(\mathbf{x}, \mathbf{x}')), \tag{2.7}$$

with

$$m_1(\mathbf{x}) = h^T(\mathbf{x})\hat{\beta} + T(\mathbf{x})A^{-1}(\mathbf{y} - H\hat{\beta}) \tag{2.8}$$

and

$$V_1(\mathbf{x}, \mathbf{x}') = \frac{\hat{\sigma}^2}{n-s-2}[c(\mathbf{x}, \mathbf{x}') - T(\mathbf{x})A^{-1}T^T(\mathbf{x}') + P(\mathbf{x})(H^T A^{-1} H)^{-1} P^T(\mathbf{x}')], \quad (2.9)$$

where H is the regression matrix of the design points, $H = h^T(\mathbf{x})$, and A is the design points correlation matrix, $A = \Psi(\mathbf{x}, \mathbf{x}')$; $T(\mathbf{x})$ is the correlaton vector between $\mathbf{x}$ and the training set, i.e. $(T(\mathbf{x}))_i = \Psi(\mathbf{x}, \mathbf{x}_i)$ and $P(\mathbf{x}) = h^T(\mathbf{x}) - T(\mathbf{x})A^{-1}H$. The estimated values of σ^2 and β are indicated as $\hat{\sigma}^2$ and $\hat{\beta}$, respectively.

Full description of the derivation of the posterior distribution are available in Rasmussen and Williams (2006).

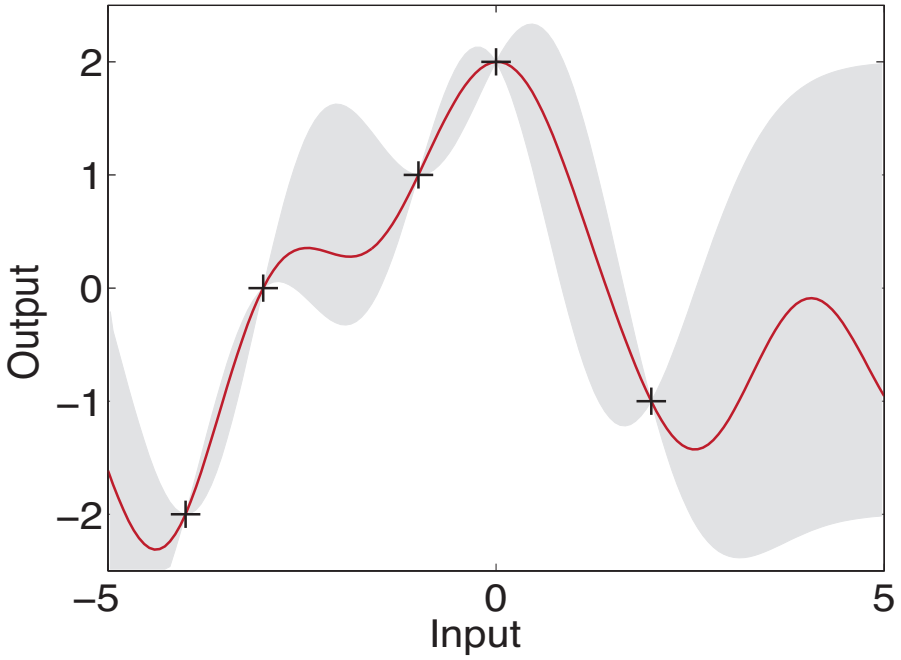


FIGURE 2.2: A 1-D scalar emulator conditioned on 5 training points. The estimated mean (red line) and variance (the shaded two standard deviation range) of the function can be obtained at an untried input. Adapted from Rasmussen and Williams (2006).

Figure 2.2 demonstrates a 1-D scalar example, in which the scalar output is a function of a single input. Conditioned on five training points, the prior is updated to a posterior which gives us an estimation of the function over the input space (red line) and uncertainty in that prediction (given as the shaded two standard deviation range). The simulator is deterministic and hence, once a training point is given, the simulator's behaviour at that point is known. There is no longer any uncertainty at this point and, thus the variance falls to zero. Since the function is assumed to be smooth, the predicted values around training points have smaller uncertainties while further away, the uncertainty range grows. Outside the range of the training points, this uncertainty grows rapidly and starts to reflect the prior standard deviation as we now extrapolate the function.

2.1.2 Multi-level emulators

Co-kriging is an extension to the previously described technique, which is applicable when a fast approximation of the primary simulator is available. For this method to work, the primary simulator and its approximation need to fulfil an additional assumption:

- *The different levels of code are correlated and contain information about one another.*

When only a small number of expensive runs are available, it has been shown that by combining these with cheaper runs from a simplified code, an emulator of the expensive model can be built at a lower cost (Forrester et al., 2007). Potentially, this method can be extended to more code levels (Kennedy and O'Hagan, 2000), including the conceptual 'reified' model (Goldstein and Rougier, 2009).

We make a simplification that the expensive and cheap models, f_e and f_c respectively, can be represented by GP emulators of the same smoothness p. The cheap model is first emulated and then linked to the expensive one using the single multiplier approach:

$$f_e(\mathbf{x}) = \rho f_c(\mathbf{x}) + f_d(\mathbf{x}), \tag{2.10}$$

The expensive function is modelled as the cheap GP multiplied by a scaling factor ρ, plus a separate GP, f_d, modelling the stochastic residual of the expensive model (Kennedy and O'Hagan, 2000; Forrester et al., 2007). This approximation is chosen for its simplicity as well as the assumption that the main difference between the two models is a matter of scale, rather than changes in the shape or the location of the output. This assumption is made based on the fact that both EMBM and PLASIM are driven by the boundary conditions specified by GENIE-1's ocean. They essentially share similar inputs but have the ability to respond differently.

Two sets of training points are required for the construction of a co-kriging emulator, a cheap set $\mathbf{y}_c = f_c(\mathbf{x}_c)$, which finely samples the input space, and a small, sparse set $\mathbf{y}_e = f_e(\mathbf{x}_e)$ of expensive points. Let the number of cheap and expensive points be n_c and n_e, respectively.

When the number of PLASIM training points is small, such that a kriging emulator cannot be built with high accuracy, co-kriging employing a large additional number of training points from GENIE-1's EMBM can be used instead. The number of points required depends on the size of the problem as well as the smoothness of the function being emulated. A general rule of thumb for the number of training points for kriging is 10 times the number of parameters (Loeppky et al., 2009). The inputs at which the

expensive training set is obtained, $\mathbf{x}_e$, is a subset of the cheap set, $\mathbf{x}_c$. These expensive points are chosen using an exchange algorithm described by Cook and Nachtsheim (1980). A short description of this algorithm is provided in Section 3.3.2 together with details on the design of training and validating points.

The covariance matrix for co-kriging, Ψ_{ck}, can be written in block form as

$$\Psi_{ck} = \begin{pmatrix} \sigma_c^2 A_c(x_c) & \rho\sigma_c^2 A_c(x_c, x_e) \\ \rho\sigma_c^2 A_c(x_e, x_c) & \rho\sigma_c^2 A_c(x_e) + \sigma_e^2 A_d(x_e) \end{pmatrix}, \tag{2.11}$$

with $A_c = \Psi(\mathbf{x}, \mathbf{x}'; \theta_c)$ and $A_d = \Psi(\mathbf{x}, \mathbf{x}'; \theta_d)$. This covariance matrix encompasses the correlation between cheap points ($A_c(x_c)$), expensive points ($A_c(x_e)$ and $A_d(x_e)$) and the cross-correlation between the cheap and expensive points ($A_c(x_c, x_e)$ and $A_c(x_e, x_c)$). Details on the formulation and derivation of this equation can be found in Kennedy and O'Hagan (2000) and Forrester et al. (2007).

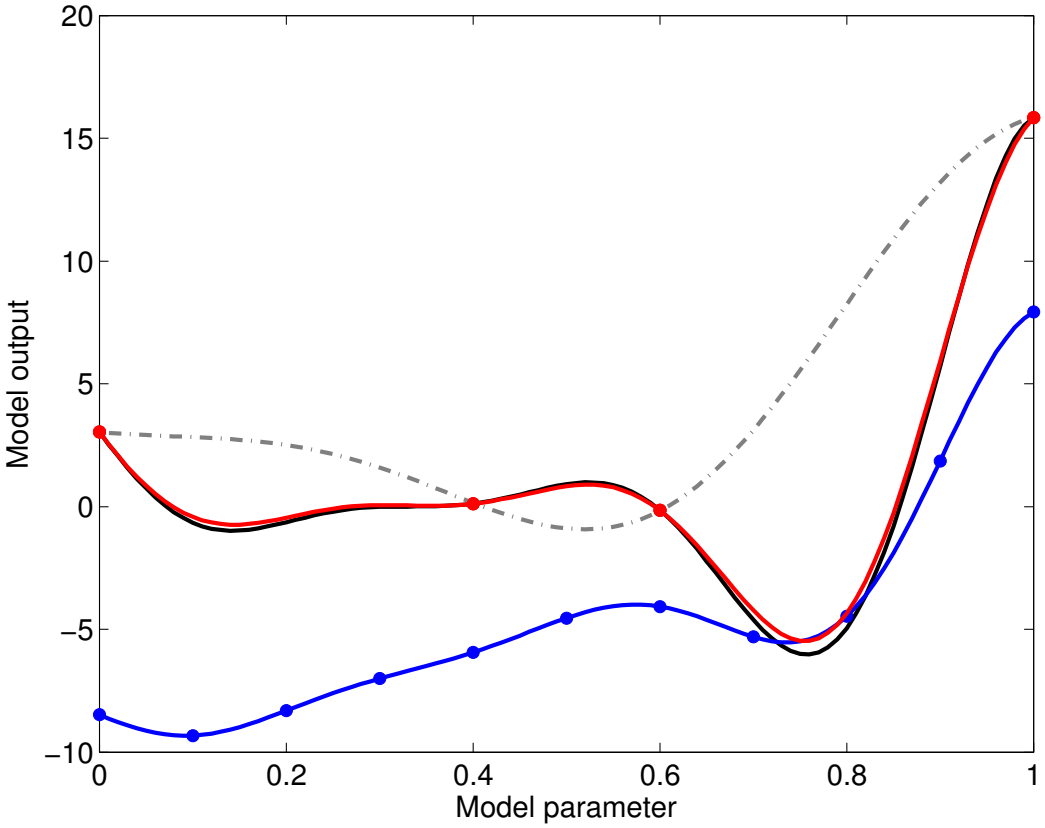


FIGURE 2.3: A 1-D scalar example of co-kriging adapted from Forrester et al. (2008). The figure shows two sets of training points, the expensive (red) and cheap (blue) sets. The kriging emulator using only the expensive points are shown as the grey dash-dotted line. The cheap and expensive emulators constructed using co-kriging are represented by the blue and red solid lines, respectively. The co-kriging emulator correctly capture the behaviour of the real expensive function (black solid line).

Figure 2.3 demonstrates a 1-D scalar example of co-kriging compared to kriging, taken from Forrester et al. (2008). Given a set of 4 expensive training points (red dots), the kriging technique (kriging emulator prediction is represented by the dash-dotted grey line) is unable to capture the true form of the expensive function (solid black line). The co-kriging emulator (solid red line) combines the information from the expensive function with more data from a cheap approximation (blue dots) to accurately predict the behaviour of the expensive function (red line). First, the emulator of the cheap function is constructed using the same procedure described in Section 2.1.1. The co-kriging emulator then estimate the scaling factor, ρ and the GP representing the difference between the cheap and the expensive functions by studying the pairs of training points evaluated at the same input from both functions. Finally, the expensive emulator is constructed as described in Eq 2.10. Hence, the construction of the cheap and expensive emulators are done seperately as opposed to jointly.

Both kriging and co-kriging emulators constructed in this work use readily available software from Forrester et al. (2008).

2.2 Dimensional reduction using principal component analysis

So far, we have only discussed the use of GP emulators for single outputs. This can be a summary value of outputs such as the strength of the AMOC or the global average SAT (Hankin, 2005). The relevant output is, however, usually a high dimensional array, containing fields and/or time-series of many climate variables (e.g. SST, SAT or precipitation).

Climate variables at different spatial or temporal locations can be emulated independently (Lee et al., 2012). This method, however, requires significant computational power and ignores the covariance between outputs close to one another (Rougier, 2007). Other extension techniques using multivariate approaches that can capture the correlations between the outputs have been developed (Rougier, 2008; Conti and O'Hagan, 2010). However, these methods are not well suited for high-dimensional output.

Many studies have been done on the extension from univariate GP emulation to handle multivariate outputs, most commonly through a dimension-reducing technique. There are a large number of linear and non-linear dimensional reduction techniques. These techniques reduce the dimensionality of a data set by embedding this data into a subspace of lower dimensionality. Among these methods, PCA is the most commonly used as this method is quick and has the advantage that the first few components explain the majority of the variance across the ensemble (Higdon et al., 2008; Holden and Edwards, 2010; Wilkinson, 2010; Challenor et al., 2010).

PCA can be applied using either the eigendecomposition technique or the singular value decomposition (SVD) technique. While we will be using SVD in this work, it is necessary to introduce both of them in order to clarify the terminology used.

The mapping from the input space to the lower dimensional output space becomes the function being emulated instead of the direct input-output relationship. This method has been applied successfully in emulating time series (Boukouvalas and Cornford, 2009; Challenor et al., 2010) and temporally evolving spatial patterns of climate variables (Holden et al., 2014, 2015).

For each ensemble member, our 2-D field of interest, with 64×32 grid cells is reshaped into a 2048×1 column vector $\mathbf{u} = [u_1, \cdots, u_m]$, with $m = 2048$. By concatenating the surface outputs obtained from running the simulations, we obtain a state vector $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n]$, with n being the number of simulation. When only a fraction of the ensemble is used, n will take the value of the size of that subset. The state vector $\mathbf{U}$, therefore, has a dimension of $m \times n$. In all the work in this thesis, $\mathbf{U}$ is demeaned (has its mean removed) and in some cases is also standardised (has the standard deviation removed). More details on how this is done can be found in the following chapters.

PCA via eigendecomposition is done on the dispersion or the covariance matrix of the state vector. There are two possible covariance matrices of $\mathbf{U}$, an $m \times m$ matrix

$$\mathbf{C} = \frac{1}{n}\mathbf{U}\mathbf{U}^{\mathrm{T}}, \tag{2.12}$$

and an $n \times n$ matrix

$$\mathbf{C}' = \frac{1}{m}\mathbf{U}^{\mathrm{T}}\mathbf{U}. \tag{2.13}$$

Our dataset has two dimensions, a spatial and an indexing dimension. In both cases, the matrix is projected onto itself, and one of the dimensions gets removed, leaving us with a measure of the covariance of the data with itself across the removed dimension. Both of these can be meaningful depending on the purpose of the analysis. In the first case, $\mathbf{C}$ is the covariance matrix of the spatial locations with each other, as determined by their variance across the ensemble. In the second case, the covariance at different parameter sets is obtained. Since we are interested in emulating the spatial variations of an output field, the first case is suitable. Quite often, the second dimension is time instead of an indexing dimension. If this is the case, the second covariance matrix $\mathbf{C}'$ allows us to study the covariance at different times.

Eigenvalue analysis can then be applied to the symmetric matrix $\mathbf{C}$

$$\mathbf{C}\mathbf{E} = \mathbf{E}\mathbf{\Lambda}, \tag{2.14}$$

where $\mathbf{E}$ is the eigenvectors matrix with columns $\mathbf{e}_i$ and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues λ_i. These eigenvectors are called the empirical orthogonal functions because

they are derived from the data and they are orthogonal,

$$\mathbf{E}\mathbf{E}^{\mathrm{T}} = \mathbf{I}. \tag{2.15}$$

If we define

$$\mathbf{Z} = \mathbf{E}^{\mathrm{T}}\mathbf{U}, \tag{2.16}$$

the original data can be expressed as

$$\mathbf{U} = \mathbf{E}\mathbf{Z}, \tag{2.17}$$

where the matrix $\mathbf{Z}$ contains principal component vectors.

The percentage of total variance explained by the j^{th} mode is

$$V_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k} \times 100. \tag{2.18}$$

SVD, unlike eigenanalysis, can be applied directly on any $m \times n$ matrix $\mathbf{U}$, giving

$$\mathbf{U} = \mathbf{L}\mathbf{S}\mathbf{R}^{\mathrm{T}}, \tag{2.19}$$

where $\mathbf{L}$ and $\mathbf{R}$ are the matrices of left and right singular vectors, respectively; $\mathbf{S}$ is the diagonal matrix of singular values (Figure 2.4). The columns of $\mathbf{L}$ are the eigenvectors of the covariance matrix $\mathbf{C}$, while columns of $\mathbf{R}$ are the eigenvectors of $\mathbf{C}'$. Therefore,

$$\mathbf{L} = \mathbf{E}, \tag{2.20}$$

is the matrix of EOFs. The right singular vectors are also often called the component scores and the principal components can be obtained from these scores as followed

$$\mathbf{Z} = \mathbf{S}\mathbf{R}^{\mathrm{T}}. \tag{2.21}$$

Any of the simulated fields can be constructed as a linear combination of the EOFs, weighted by their respective series of principal components. Each $m \times 1$ column of $\mathbf{L}$ is an eigenvector describing a map or a mode of variation in the ensemble. Its principal component is the corresponding $n \times 1$ column of $\mathbf{Z}$. Either the component scores or the principal components can be emulated and used to reconstruct the simulated data. Each pair of component scores and PCs differ by a known constant specified by their corresponding singular value. A pictoral representation of this decomposition is shown in Figure 2.4. In the following work, it's the principal components that are emulated. The n elements (or indices) of each component score correspond to the n simulations used as training points. The component scores obtained can be considered as scalar functions of the input parameters and can be emulated using kriging or co-kriging. The

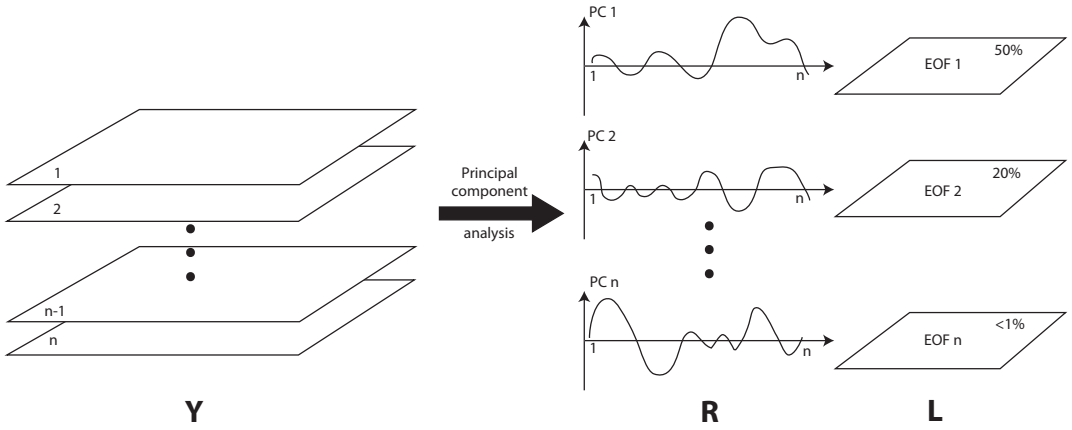


FIGURE 2.4: A pictoral description of PCA. The n ensemble surface fields are gathered into a state vector $\mathbf{Y}$. The results of the SVD are shown on the right: i) the EOFs describe spatial patterns of the n different modes, $\mathbf{L}$; ii) the component scores, $\mathbf{R}$ which specify the scaling of each EOF for each of the n simulations and iii) the percentage of variance explained derived from the singular values.

number of training points, n, would be n_c and n_e for the cheap and expensive emulator, respectively.

The singular values are related to the eigenvalues by

$$\mathbf{S}\mathbf{S}^{\mathrm{T}} = n\mathbf{\Lambda}, \tag{2.22}$$

and therefore, the percentage of total varianced explained of the j^{th} mode is

$$V_j = \frac{\sigma_j^2}{\sum_{k=1}^{m} \sigma_k^2} \times 100. \tag{2.23}$$

Each EOF displays the structure of spatial variations across the ensemble. However, these are unit vectors and need to be scaled if we are interested in their amplitude corresponding to the amplitude in the data set. This can be obtained by multiplying the EOFs with the corresponding eigenvalues

$$\mathbf{D} = \frac{1}{\sqrt{n}}\mathbf{L}\mathbf{S}. \tag{2.24}$$

The EOF maps plotted in later chapters are scaled this way.

Finally, with all the tools explained, we can use SVD or eigenanalysis for dimensional reduction of a 2D data set. The i^{th} grid cell of the j^{th} field from $\mathbf{U}$ can be written as

$$\hat{U}_{ij} = \sum_{k=1}^{q} L_{ik} S_{kk} R_{kj}^{T}. \tag{2.25}$$

where q is the number of modes retained. When $q = m$, $\hat{U}_{ij} \equiv U_{ij}$, otherwise, $\hat{U}_{ij}$ is an approximation of U_{ij}.

The top (or high order) EOFs often explain most of the variance in the data such that the dimension of $\mathbf{U}$ can be reduced by keeping only the first q components ($q < m$). The elements of the PCs are now used as training data instead of the direct climate variable. We assume that these PCs also fulfil the same assumptions made for the climate variables. Emulators are built for the first q component scores, providing an estimation of $\hat{\mathbf{Z}}$, for an untried input combination. They are then used to reconstruct the final prediction, $\hat{\mathbf{U}}$ of the emulated field.

The prediction, $\hat{\mathbf{U}}$, is different from the simulated value of $\mathbf{U}$ by an error component, which can be decomposed into truncation error and component error. Truncation error is due to dimensional reduction. This is kept low by making sure that enough EOFs are retained to explain most of the variance in the ensemble. Although there is no definite rule on how much percentage explained would be sufficient, a high value such as 90% for SAT should be satisfactory. The component error is a result of imperfect estimation by the emulator, i.e. error in estimating the correct hyperparameters. This can be minimised by making sure enough training data is used to ensure the emulator can capture the real trend of the ensemble. The GP emulator also provides an estimate of this error.

Chapter 3

Emulating surface air temperature

3.1 Introduction

Sophisticated computer simulations are used in climate research to improve our understanding of the climate system. They are often used to project future changes in global temperature, corresponding to different emission scenarios. Our confidence in these projections is highly dependent on how reliable the simulations are. For example, the study of palaeoclimate offers an insight into the Earth's past climate system and also provides valuable out-of-sample data to validate our simulations. However, this requires running complex simulations on multi-millennial timescales, which is computationally demanding. For most coupled atmosphere-ocean general circulation models (AOGCMs), this is currently not feasible. Other studies such as uncertainty and sensitivity analysis or history matching require a thorough exploration of the input parameter space. The class of fast models, known as Earth system models of intermediate complexity (EMICs) is suitable for these types of studies. Their efficiency is achieved by a combination of lower spatial and/or temporal resolution and the use of simplified parameterizations. However, depending on the nature of the questions asked, these lower fidelity models might be insufficient.

To address this issue, an emulator is often employed to provide a statistical estimation of the expensive model's response without the need to perform a new simulation. Even then, this approach becomes impractical when the models of interest are very computationally intensive. In order to build a reliable emulator, a certain number of simulations are needed to provide the basis upon which the emulator is built. This number can be large, especially when multiple model parameters are varied or when the model's climate response exhibits non-linear behaviours. For a computationally expensive GCM, a sufficient number of simulations are often not affordable. This chapter describes an efficient

emulation process that utilises the connection between models of different complexities. The idea is to establish a traceable hierarchy, using an emulator for the simple model to construct an emulator of the more complex one (Kennedy and O'Hagan, 2000; Cumming and Goldstein, 2008).

While the high-fidelity (complex) model is computationally expensive, the low-fidelity (simple) model is cheaper to evaluate and can be sampled more finely across the input space, providing extra information where expensive data are sparse. The models forming this hierarchy can be structurally related or structurally unrelated. Models are referred to as structurally related when they are from the same family of code but have different resolutions. These models might have other differences resulting from the change in mesh resolution. Examples of such models are the HadCM3 (Hadley Centre Coupled Model version 3; Pope et al., 2000) and FAMOUS (Fast Met Office/U.K. Universities Simulator; Jones et al., 2005) of the MET Office. Multi-level emulation has been employed before to link such models (Forrester et al., 2007; Cumming and Goldstein, 2008; Williamson et al., 2012). Here, our focus is on structurally unrelated atmospheric models, which solve different sets of equations. Since both the cheap and expensive codes model the same physical system, it is reasonable to expect qualitative similarities between the two. This argument is supported by studies showing no systematic difference in model behaviour between EMICs and AOGCMs (Stouffer et al., 2006; Plattner et al., 2008).

The following work illustrates the use of a method that combines multi-level emulation with a dimensional reduction technique through an example study using GENIE-1, from the Grid ENabled Integrated Earth system modelling framework (GENIE), and PLASIM (Planet Simulator). GENIE-1 and PLASIM are chosen in this case since they are both suitable for Earth system modelling for long timescales, but are structurally different. PLASIM's atmosphere is also substantially more complex and thus, computationally more expensive than GENIE-1's energy-moisture balance model, EMBM, of the atmosphere. EMBM incorporates the vertically integrated energy-moisture balance equations while PLASIM is based on the moist primitive equations representing the conservation of momentum, mass and energy. EMBM, therefore, is not capable of producing air temperature and pressure at different altitudes or an interactive cloud and wind field. The hierarchy formed by these two models is exploited using the multi-level technique, allowing us to construct an emulator of PLASIM atmospheric variables at a reduced cost. Specifically, Gaussian process emulators are used to obtain the statistical relationship between the response of the EMBM atmosphere and the PLASIM atmosphere to changes in their boundary conditions (sea surface temperature, longwave and shortwave radiative forcing). This ability of this relationship to predict the behaviour of PLASIM atmosphere, in the absence of feedbacks on other climate system components, is then assessed. The dimensional reduction technique is employed to extend the emulation method for prediction of high-dimensional outputs in addition to scalar summary quantities.

Once constructed, the emulators provide estimates of simulation results, at untried combinations of the inputs, as finely as needed, at a low cost. This enables statistical methods such as history matching (Holden et al., 2010; Edwards et al., 2011) and sensitivity/uncertainty analysis (Rougier et al., 2009). Information from the cheap code can also be used to inform future designs of experiments using the expensive code. Apart from above, the emulators of 2-D surface fields similar to the one constructed here can potentially be used to provide the fields needed for coupling with other climate models or components of climate models.

3.2 Model configurations

In this study, we utilise the atmospheric component of GENIE-1 (version 2.7.8) (Lenton et al., 2006), an EMIC, as the cheap model. GENIE-1 was formerly known as C-GOLDSTEIN in Edwards and Marsh (2005) and has since been modified for incorporation into the GENIE framework (Lenton et al., 2006). It is most recently described in Marsh et al. (2011). GENIE-1 is designed with scalable spatial resolution and high efficiency, suitable for long integrations (10^3 to 10^6 years) to study past climate and large ensembles to explore the uncertain input parameter space (Holden et al., 2010).

The configuration of GENIE-1 employed here couples a single layer EMBM atmosphere to a 3-D frictional geostrophic ocean model with linear drag (GOLDSTEIN) and a thermodynamic, advection-diffusion sea-ice model (GOLDSTEIN sea ice). The ocean component is run at 64×32 horizontal resolution and 16 vertical layers. Also incorporated in this configuration is the efficient numerical terrestrial scheme (ENTS) designed for long simulations (Williamson et al., 2006). ENTS represents a hybrid of a simple bucket model with an explicit but simplified carbon cycle. The effect of orography is applied to surface processes in ENTS by applying a constant lapse rate (Holden et al., 2010). Orography, therefore, has an effect on the land surface temperature and so indirectly influences the atmosphere. The atmospheric processes such as heat and moisture transport do not interact with the orography.

The parameterization of atmospheric transport of heat and moisture in EMBM is done by diffusion. Moisture can also be advected by a prescribed monthly climatological wind field. This wind field is fixed and is the same for all simulations in EMBM. The effect of cloud cover on incoming shortwave radiation is captured by a prescribed albedo field, diagnosed from reanalysis data (Lenton et al., 2006). The effect of cloud cover on outgoing long-wave radiation is parameterised as perturbations to the unmodified clear-skies outgoing long-wave radiation. Precipitation is assumed to occur whenever the relative humidity is above a certain adjustable threshold.

The atmosphere of PLASIM-ENTS (Holden et al., 2014), driven by boundary conditions specified by GOLDSTEIN ocean and sea ice, is chosen as the expensive model. PLASIM

(Fraedrich et al., 2005) consists of an atmospheric GCM of intermediate complexity, which can interact with reduced sub-models of the ocean, sea ice and land (Fraedrich et al., 2005). Hereafter, we refer to the atmospheric component of PLASIM-ENTS as simply PLASIM. PLASIM solves the primitive equations for vorticity, divergence, temperature and the logarithm of surface pressure. It includes a hydrological cycle, interactive clouds, and a simple radiation scheme. Coupling between the PLASIM atmosphere and an ocean model other than its own has been used before to study the effects of mountains and ice sheets on ocean circulation (Schmittner et al., 2011). An emulator of PLASIM-ENTS has been employed in a range of integrated assessment modelling couplings with various technico-economic models (Mercure et al., 2014; Labriet et al., 2015).

PLASIM is run at T21 resolution, which corresponds to a triangular truncation applied at wave number 21. It is almost an exact match of GENIE-1's 64×32 mesh except for negligible differences at the highest latitudes. While EMBM has only one layer, the atmosphere of PLASIM is represented by ten vertical layers in terrain following σ-coordinates. Most importantly, EMBM uses prescribed wind fields, which means that feedbacks due to changing atmospheric circulation patterns are not captured, while PLASIM's interactive wind field can change according to the different specified boundary conditions, leading to more diverse climate states. While the cloud albedo in EMBM is prescribed, the cloud albedo in PLASIM is a function of height and area of coverage.

For our study, monthly surface output fields of GENIE-1, namely, sea surface temperature (SST), fractional sea-ice coverage (SIC) and sea-ice thickness (SIH) are used to drive PLASIM. This means that the atmospheric circulation can change according to the underlying sea surface temperature and sea-ice condition but cannot influence the ocean or sea-ice physical state. This constrains PLASIM responses to a certain extent. The atmospheric responses of EMBM and PLASIM to the same set of physically plausible boundary conditions are compared and emulated. The surface air temperature (SAT) from EMBM atmosphere is treated as a fast approximation of PLASIM SAT when multi-level emulation is applied.

3.3 Ensemble design

3.3.1 Model parameters

To explore emulator performance in situations where the climate states are very different from modern conditions, an ensemble is designed to fill a large input space; 12 model parameters and one dummy variable are varied, either linearly or logarithmically, over the ranges indicated in Table 3.1. In this experiment, we are primarily interested in the effects introduced by the dynamical atmosphere of PLASIM, and so the parameters were

chosen according to their influence on SAT. Parameters with significant contributions to SST and the strength of the AMOC, and hence indirectly influencing SAT, are also included. This judgment is based on previous studies using large GENIE-1 ensembles (Lenton et al., 2006; Holden et al., 2010).

The first parameter (ICF) represents the boundary condition of the glacier coverage as well as the corresponding orography at different snapshots in time extending from the present (0 kyr before present) to the Last Glacial Maximum (LGM) (21 kyr before present) with steps of 1 kyr. Each value of ICF corresponds to a spatial distribution of land ice at a certain period according to the Peltier reconstruction ICE-5G (Peltier, 2004). Both ice area and ice volume are non-linear functions of ICF. Independent from ICF, the atmospheric CO_2 concentration (RFC) is varied from 150 ppm to 1400 ppm to include the glacial-interglacial variations, pre-industrial and modern climate as well as future responses to rising greenhouse gas emissions. The upper limit is chosen to include the CO_2-equivalent concentration for all greenhouse gases by 2100 according to the high emission pathway RCP8.5 (Riahi et al., 2011; Meinshausen et al., 2011). The equivalents of these two parameters are also varied accordingly in PLASIM. Other PLASIM parameters are kept at default values, which are listed in Haberkorn et al. (2009).

TABLE 3.1: Ten of the chosen parameters, with the exception of ICF and RFC, are taken from an ensemble design used in Holden et al. (2010). The ranges were initially based on those used in the same study. However, adjustments are needed since the model is run at 64×32 horizontal resolution here compared to the previously used 36×36 mesh. The ranges shown below are obtained after an initial exploratory ensemble. The distribution specifies whether their values (Lin) or the log of their values to base 10 (Log) are used to generate the sampling plan in section 3.3.2.

	Code	Parameter	Min	Max	Dist.
1	ICF	Ice sheet and orography configuration	0	21	Lin
2	OHD	Ocean isopycnal diffusivity ($m^2\ s^{-1}$)	300	4000	Log
3	OVD	Ocean diapycnal diffusivity ($m^2\ s^{-1}$)	5×10^{-6}	2×10^{-4}	Log
4	ODC	Ocean friction coefficient ($days^{-1}$)	0.5	3	Lin
5	WSF	Wind scale coefficient	1	3	Lin
6	AHD	Atmospheric heat diffusivity ($m^2\ s^{-1}$)	4×10^6	7.0×10^6	Log
7	AMD	Atmospheric moisture diffusivity ($m^2\ s^{-1}$)	5×10^4	6×10^6	Log
8	APM	Atlantic-Pacific freshwater flux (Sv)	0.032	0.640	Lin
9	RMX	Relative humidity threshold for precipitation	0.6	0.9	Lin
10	OL0	Clear skies OLR reduction ($W\ m^{-2}$)	0	10	Lin
11	OL1	OLR feedback ($W\ m^{-2}\ K^{-1}$)	−0.5	0.5	Lin
12	RFC	CO_2 forcing (ppm)	150	1400	Lin
13	FFX	Dummy varable	0	1	Lin

Mixing and transport in the ocean are controlled by the isopycnal and diapycnal diffusivity parameters (OHD and OVD, respectively), a momentum drag coefficient (ODC)

and a wind scaling factor (WSF) (Edwards and Marsh, 2005). These parameters affect the ocean boundary conditions, which are seen by both EMBM and PLASIM directly.

APM is a flux correction responsible for transporting fresh water from the Atlantic to Pacific, affecting deep-water sinking in the North Atlantic and hence the strength of the AMOC (Edwards and Marsh, 2005). The uncertain impact of atmospheric transport is captured through atmospheric heat and moisture diffusivity parameters (AHD and AMD, respectively) (Edwards and Marsh, 2005). OL0 and OL1 modify the outgoing long-wave radiation and are included to allow for uncertainty due to cloud coverage and its dependence on a change in the global average SAT (Thompson and Warren, 1982; Matthews and Caldeira, 2007). RMX is the threshold value of relative humidity for precipitation, capturing the uncertainty in water vapour feedbacks (Lenton et al., 2006). Except for APM, these atmospheric parameters directly control the behaviour of EMBM, but only affect PLASIM indirectly through their influence on the ocean boundary conditions.

In addition to these 12 model parameters, a dummy parameter is included for statistical validation purposes, which will be discussed in more detail in section 3.4.3.

3.3.2 Statistical design

First, all input parameters are normalised to $[0, 1]$ from their original ranges in Table 3.1. An approximate maximin Latin hypercube (MLH; Morris and Mitchell, 1995 sampling plan is then generated, producing 660 combinations of the 13 chosen parameters to form a GENIE-1 perturbed physics ensemble. The maximin criterion, also known as the Morris-Mitchell criterion, is applied since a randomly generated Latin hypercube does not ensure good space-filling properties, which are desirable to explore the input space evenly using a small number of points. An MLH sample is a Latin hypercube sample that maximises the minimal separation between pairs of design points. Here, the separation is simply the Euclidean distance between the points.

The number of cheap and expensive training points are subsets of the MLH designs above. They are chosen using an exchange algorithm described by Cook and Nachtsheim (1980). The expensive points, which are a subset of the cheap training points are selected in the same manner. First, a random subset $\mathbf{x}_e$ is selected at random and the Morris-Mitchell criterion is calculated. The first point $x_e^{(1)}$ is then exchanged with each of the remaining points in x_c. The exchange which gives the best Morris-Mitchell criterion is chosen. By repeating the same procedure for the remaining points $x_e^{(2)}, \cdots, x_e^{(n_e)}$, the 'best' subset is obtained.

Each member simulation of this ensemble is run for 5000 years to reach a steady state; 600 simulations were completed successfully, producing a broad range of climate responses, which are summarised in Table 3.2. The 60 failures are located at the end of

one or more parameter ranges, where numerical instability occurs. Failure is most commonly due to low values of AHD and AMD. Although the design space can be narrowed down to reduce the failure rate, this would also restrict the range of the resulting climate states. Since we wish to build emulators, which can predict a broad range of climate responses without having to extrapolate beyond the designed range, this ensemble design is appropriate.

TABLE 3.2: A summary of the simulated climate states from the 600-member ensembles of GENIE-1 with EMBM and PLASIM. The SAT used to compute these values are the annual average of the final year of a 5000-year simulation for GENIE-1 and annual average calculated from the last 30 years of a 35-year simulation for PLASIM.

	Min	Max	Mean	S.d.
PLASIM global mean SAT (°C)	−6.05	23.33	11.25	4.63
EMBM global mean SAT (°C)	−2.62	24.43	12.56	4.25
GENIE-1 global mean SST (°C)	7.77	27.10	17.01	3.24
GENIE-1 maximum strength of the AMOC (Sv)	0.82	36.59	15.31	5.94
GENIE-1 Antarctic sea-ice area ($\times 10^6$ km^2)	0.00	23.09	1.73	2.60

A second MLH design is generated in the same parameter space, producing 214 successful simulations, for validation purposes. The emulator predictions at these points are compared against the simulated values to assess the performance of the emulators.

For each successful GENIE-1 simulation, surface output fields are extracted and used to force PLASIM for another 35 years. Each sampling plan, therefore, produces two equivalent ensembles of EMBM and PLASIM outputs. The fields used to initiate PLASIM simulations are SST, SIC and SIH as mentioned in section 3.2. The 600-member ensemble mean and standard deviation of GOLDSTEIN SST and ice area are shown in Figure 3.1. The ice coverage plotted is a combination of the fractional sea-ice cover from GOLDSTEIN sea ice (SIC) and the glacier mask described by ICF. The change in elevation corresponding to each glacier mask is applied for both GENIE-1 and PLASIM.

Both ensemble designs are larger than needed in this case. On average, ten simulations are needed for each parameter being varied (Loeppky et al., 2009). Since 13 parameters are perturbed, a 130-member ensemble would be sufficient. There are several reasons why a 600-member ensemble was used. First, the number of simulations required ultimately depends on the variations of the variable of interest within the specified parameter space. If this variable behaves non-linearly and exhibits a bifurcation, more simulations would be required to capture such behaviour accurately. Second, the required number of simulations of cheap and expensive models are unknown. Different combinations of subsets with varying sizes are used and compared in the following section. It is ideal to generate a new design separately for each case, but this is highly inefficient and will result in a large incoherent ensemble with low reusability. Therefore, it is preferable to start with a large design from which different subsets can be chosen. These subsets are

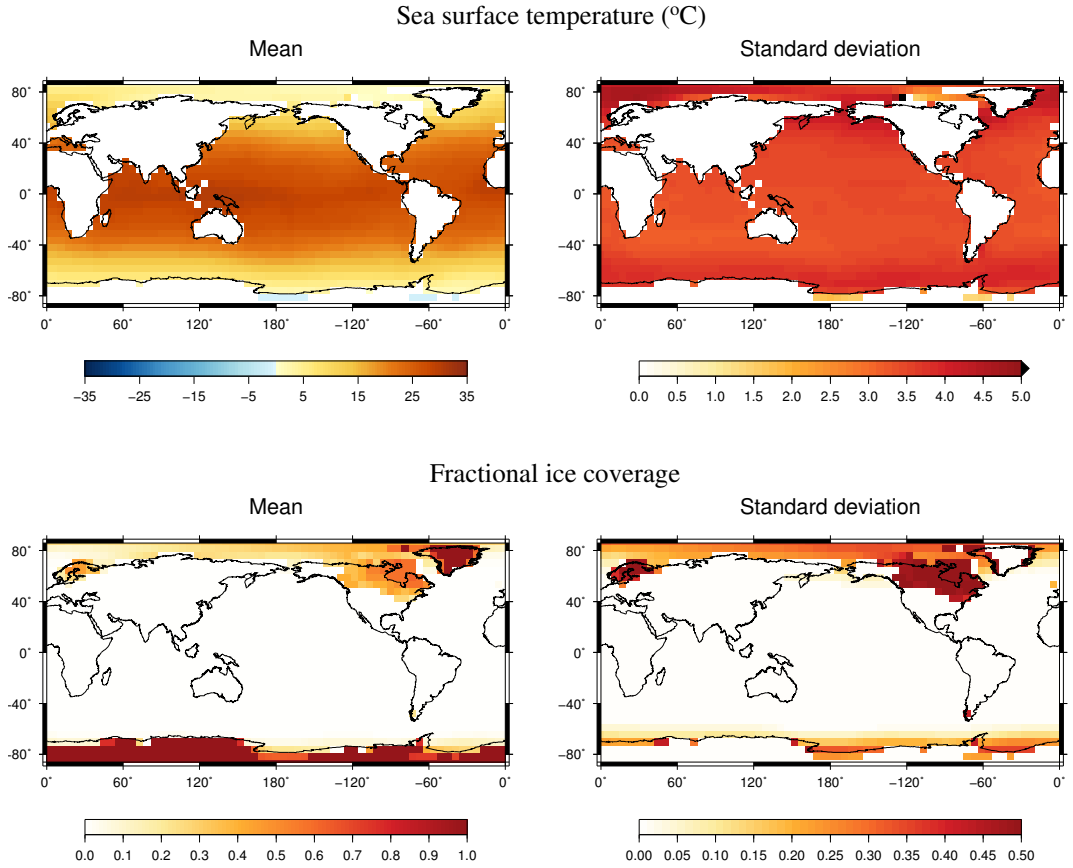


FIGURE 3.1: The mean and standard deviation of SST and fractional ice coverage across the 600-member ensemble. The SST and sea-ice coverage are prognostic output of GENIE-1 while the land ice coverage is regridded from Peltier ICE-5G. These fields, among others, are applied as surface boundary conditions to drive PLASIM atmosphere.

all subjected to the same maximin criteria mentioned above using the exchange algorithm. While this ensemble will be more reusable, a subset of it will most likely have a worse space-filling property than an independent MLH design of the same size. This is minimised by starting from a very large ensemble like the one employed here.

3.4 Results

3.4.1 Simulated climates

The EMBM output SATs are averaged over the final year of the 5000-year simulations while PLASIM output fields are averaged over the last 30 years. The ranges of some output variables obtained from the 600-member ensembles of GENIE-1 and PLASIM simulations are summarised in Table 3.2. The diversity of the output climate states is demonstrated by the considerable variation in SST, SAT, Antarctic sea-ice area and

strength of the AMOC, which is weakened or shut down in some simulations. Because of the large upper limit of atmospheric CO_2 concentration and GENIE-1's general bias towards low Antarctic sea ice, in some simulations, the Southern Ocean appears to be completely ice free. The SAT in PLASIM is lower in general and exhibits a slightly larger variation compared to EMBM's value.

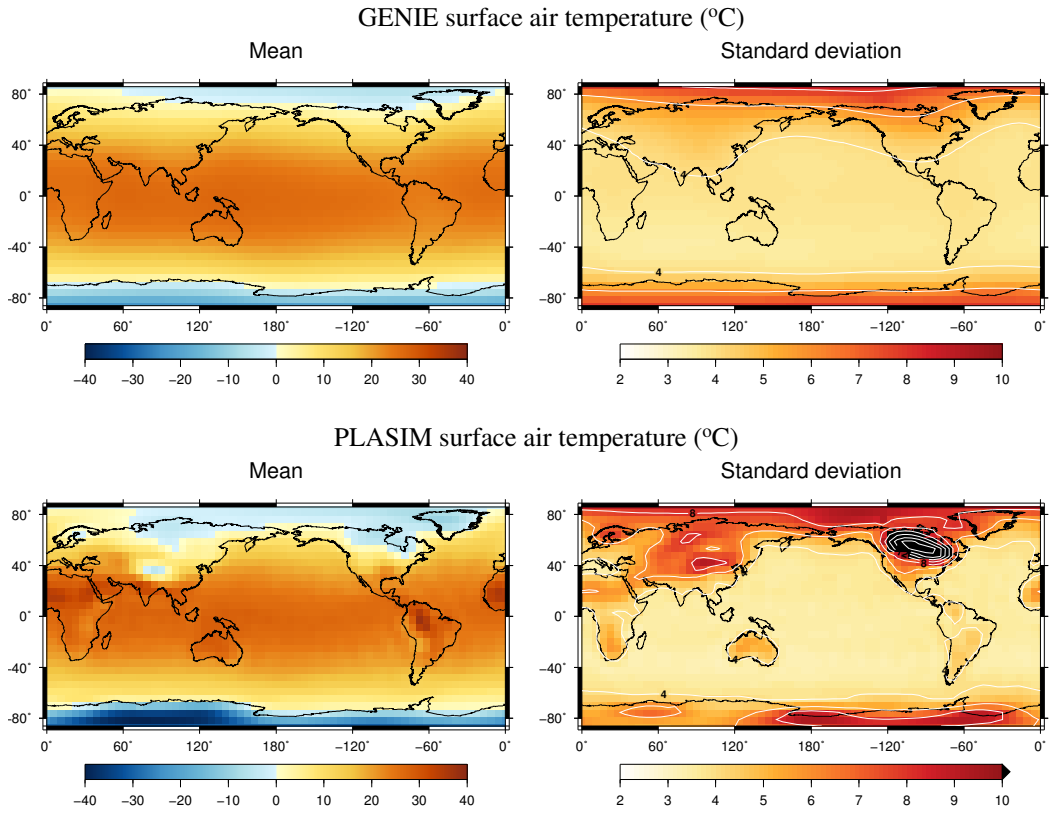


FIGURE 3.2: The mean and standard deviation of SAT across the 600-member ensembles of GENIE-1 and PLASIM. There are white cells on the PLASIM SD plot where the outputs go beyond the plotted range. The largest standard deviation in PLASIM is 17.5°C. The contours on the mean and SD plots are shown every 10 and 4°C, respectively.

Figure 3.2 shows the ensemble mean and standard deviation of PLASIM and EMBM SAT. Although similar spatial patterns are seen in both, PLASIM exhibits a larger variation spatially and across the ensemble, particularly at high elevations. The comparison between the two models also shows that EMBM climate is much more zonal, with little land-sea difference. This is one of the known weaknesses of the energy-moisture balance model of the atmosphere, which is too diffusive (Lenton et al., 2006). A clearer distinction between the ocean and the continents is modelled in PLASIM as shown in the standard deviation plot of Figure 3.2.

The differences seen in Table 3.2 and Figure 3.2 are partly due to the nature of EMBM and PLASIM SAT fields. In contrast to PLASIM, EMBM does not take into account the effect of the elevation when calculating SAT. EMBM's and PLASIM's global mean SAT can be compared to the annual global mean SAT at the 1000 mb pressure surface and the 2 m surface from National Centers for Environmental Prediction - Department of Energy (NCEP-DOE) Reanalysis 2 (1979–2013), respectively (Kanamitsu et al., 2002). The two climatologies have global mean SAT of 8.5°C (1000 mb) and 6.9°C (2 m), respectively, correspond to a difference of 2.6°C. The difference in each PLASIM-EMBM pair ranges from −1.3°C to 6.1°C, with a mean of 1.32°C. Among simulations with modern glacier configurations and atmospheric CO_2 within 340–400 ppm, the average difference is 1.51°C, lower than the climatological value by approximately 1°C. The large difference between the two ensembles can be attributed to the large parameter range and the difference in climate sensitivity. With dynamic wind and interactive cloud, PLASIM is expected to produce a more realistic precipitation pattern, especially over the continents. Interactions between the atmosphere and the ice sheets can also lead to larger variations due to orography or precipitation feedbacks. Their climate sensitivities will be explored later on with the help of the GP emulators constructed.

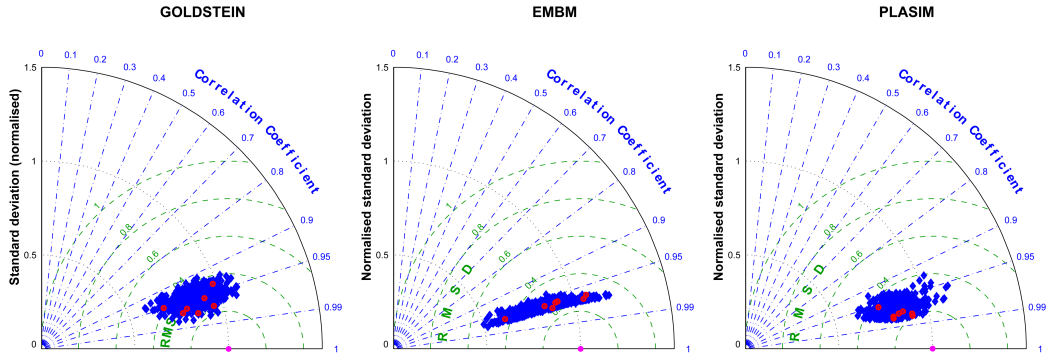


FIGURE 3.3: Taylor diagrams showing a comparison between model runs with climatology: GOLDSTEIN SST (left), EMBM SAT (middle) and PLASIM SAT (right). The magenta dots represent reanalysis taken from Locarnini climatology (1900–2005) (Locarnini et al., 2006) (left), NCEP-DOE reanalysis 2 annual mean SAT (1979–2013) at 1000 mb (Kanamitsu et al., 2002) (middle) and NCEP-DOE reanalysis 2 annual mean SAT (1979–2013) at 2m Kanamitsu et al. (2002) (right). The points highlighted in red represent runs with ICF $\in \{0, 1, 2, 3, 4\}$ and 340 ppm $<$ RFC $<$400 ppm.

The resulting SATs from both models are compared against climatology in Figure 3.3 using Taylor diagrams. These plots demonstrate the range of output obtained with respect to modern climate. The modern climate states here serve as reference points to better demonstrate the spread of the simulated ensembles as well as their differences. Both the standard deviations (SD) and root mean square differences (RMSD) are normalised (and non-dimensionalised) by dividing them by the SD of the observations. GOLDSTEIN SSTs from all simulation runs are compared with annual mean SST (1900-2005) from NOAA World Ocean Atlas (Locarnini et al., 2006). The SATs from the single-layer atmosphere EMBM are compared with annual mean surface air temperature over the

period from 1979 to 2013 at the 1000 mb pressure surface from NCEP-DOE Reanalysis 2. The SATs from PLASIM are compared with the air temperature at 2 m from the same reanalysis. The simulation runs with ice sheet configuration and CO_2 concentration similar to those within the 1979–2013 period, ICF $\in \{0, 1, 2, 3, 4\}$ and 340 ppm $<$ RFC $<$400 ppm, are highlighted in red. A plot showing the difference between the mean surface temperatures over this group of simulations and climatology is included in Appendix A.

The simulated pattern of SST correlates well with observation (average correlation coefficient of 0.95), while the majority of the ensemble exhibits smaller spatial variability than climatology (average normalised SD of 0.85). The spread in these modern GOLDSTEIN SST points is due to the large range of the perturbed GENIE-1 parameters. The standard deviations of SAT are also underestimated in EMBM (average normalised SD of 0.83). PLASIM SATs correlate well with the climatology (average correlation coefficient of 0.97). The spatial variation in PLASIM SAT has a similar mean to EMBM but has a larger range (both ensembles have average normalised SD of 0.83).

3.4.2 Scalar emulation

An emulator is first constructed for EMBM global mean SAT with a starting subset of 30 training points. The coefficients of determination (r^2) and the root mean square error (RMSE) between the simulated and emulated validation points are computed and then used as indications of the validity of the emulator. The coefficient of determination, r^2, is the square of the sample correlation coefficient:

$$r^2 = \left(\frac{\mathrm{cov}(\mathbf{y}, \widehat{\mathbf{y}})}{\sqrt{\mathrm{var}(\mathbf{y})\mathrm{var}(\widehat{\mathbf{y}}))} \right)^2 . \tag{3.1}$$

More training points are gradually added to produce more accurate emulators with decreasing RMSE and increasing r^2. At approximately 200 points ($n_c = 200$), adding more training data no longer significantly reduces the RMSE value (Figure 3.4). It is concluded that approximately 200 cheap points are sufficient to capture the variation over the EMBM output space. We then attempt to build co-kriging emulators for global mean SAT in PLASIM using 200 cheap points and additional expensive data points. Again, 30 expensive points are chosen for initial training. It is found that 50 PLASIM points ($n_e = 50$) are enough to construct a good emulator with RMSE = 0.51°C and $r^2 = 0.98$.

The number of training points required varies from one emulator to another since it depends strongly on the function being emulated. As the number of parameters increases, the dimension of the emulator also increases and hence more training points are required. Typically an average of 10 points per dimension is assumed. This, however,

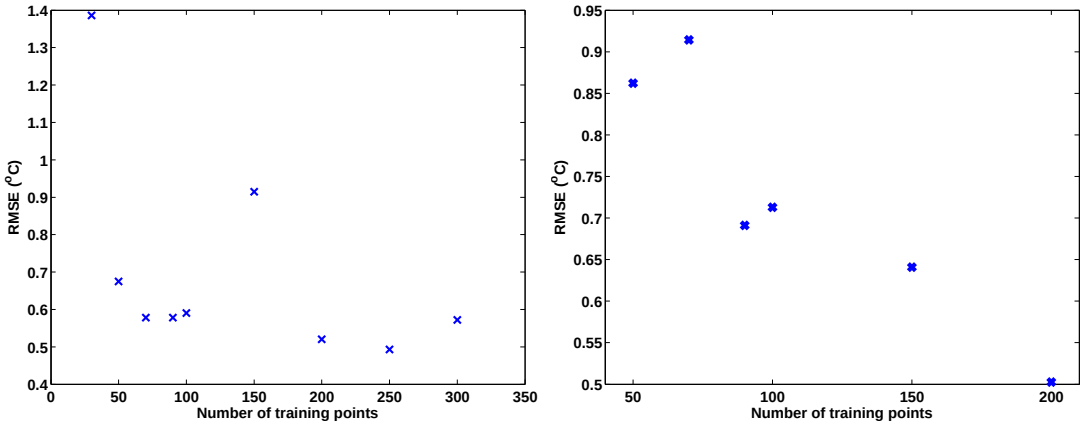


FIGURE 3.4: Plots of the emulator RMSE against the number of training points for kriging emulator of GENIE global mean SAT (left) and PLASIM global mean SAT (right). In general, the error decreases as the number of training points increase. However, this is not the case in some instances due to the particular distribution of the training points used. Further exploration would be required to identify the precise reason why, for example, the 150-member training set leads to an abnormally poor emulator of GENIE SAT.

depends on how non-linear or how 'active' the function is. A highly nonlinear function might require many more points while a more linear function might not need as many as ten points per dimension.

Kriging emulators using only expensive points are also constructed to provide a comparison between the two techniques. When the same amount of training data is used, co-kriging outperforms kriging. More expensive points are then added to improve the kriging emulator until a similar value of RMSE is obtained. In this case, the kriging emulator using $n_e = 200$ PLASIM training points gives RMSE = 0.50°C and $r^2 = 0.98$ (Figure 3.4). Therefore, co-kriging achieves the same level of accuracy with only 25% as many expensive data.

Using our experimental setup, a 30-year expensive simulation using PLASIM takes approximately the same time as a 5000-year cheap simulation using GENIE. Because of our specific configuration, however, an expensive simulation requires the respective cheap simulation to be obtained first, making an expensive training point twice as computationally expensive as a cheap point. By obtaining 50 expensive points, we already have 50 cheap points and only need 150 further cheap points to construct the co-kriging emulator presented above. The total cost of this emulator is equivalent to the cost of 125 expensive training points, lower than the number required to construct an equivalent kriging emulator of PLASIM global mean SAT (200 expensive points). When the number of extra cheap simulations required are over 300, it is no longer cheaper to use co-kriging because the cost of PLASIM simulations required would be lower. The decision to invest in a small number of expensive simulations or a larger number of cheap

simulations can be problematic as the necessary number of simulations are not known in advance and depend on various factors (the variable being emulated and the complexity of the model(s) used). This is a topic which deserves further attention in future studies.

Another pair of emulators is produced for the global SAT anomaly from SST (global annual mean SAT minus SST). In this case, the component of the SAT response that is a trivial function of the boundary conditions is removed. Following the procedure described above, a co-kriging emulator using 70 expensive points and 250 cheap points was constructed and compared to a kriging emulator using only 70 expensive points. The RMSE and r^2 are included in Table 3.4. The co-kriging emulator obtains RMSE $= 0.31$°C and $r^2 = 0.95$. This time, a kriging emulator using 100 expensive points gives similar validation result, RMSE $= 0.33$°C and $r^2 = 0.92$. The co-kriging emulator still manages to utilise meaningful information from EMBM, albeit not as well as in the previous example, and reduces the expensive points needed by approximately 30%.

TABLE 3.3: Validation results for kriging and co-kriging emulators of PLASIM global mean SAT. The co-kriging emulator uses 50 expensive points and 200 cheap points while the kriging emulator here uses the same 50 expensive points.

	Kriging emulator	Co-kriging emulator	Simulated ensemble
RMSE (°C)	0.93	0.51	N/A
r^2	0.94	0.98	N/A
Ensemble mean (°C)	10.96	11.30	11.40
Ensemble s.d. (°C)	4.89	4.73	4.57

TABLE 3.4: Validation results for kriging and co-kriging emulators of PLASIM global mean SAT−SST. The co-kriging emulator uses 70 expensive points and 250 cheap points while the kriging emulator here uses the same 70 expensive points.

	Kriging emulator	Co-kriging emulator	Simulated ensemble
RMSE (°C)	0.42	0.31	N/A
r^2	0.91	0.95	N/A
Ensemble mean (°C)	−5.81	−5.77	−5.72
Ensemble s.d. (°C)	1.65	1.70	1.50

For both kriging and co-kriging emulators using the same expensive training points, the emulated global mean SATs at the 214 validation points are plotted against their simulated values (Figure 3.5). The corresponding RMSE and r^2 values are shown in Table 3.3. Table 3.3, 3.4 and Figure 3.5 show that the co-kriging emulators reproduce the simulated values more accurately. Table 3.3 and 3.4 also contains the ensemble mean and standard deviation from both co-kriging and kriging emulators, compared with the true values obtained from the simulated ensemble.

While co-kriging outperformed kriging in both cases, multi-level emulation does a much better job at predicting SAT than SAT minus SST. Nevertheless, the r^2 scores between

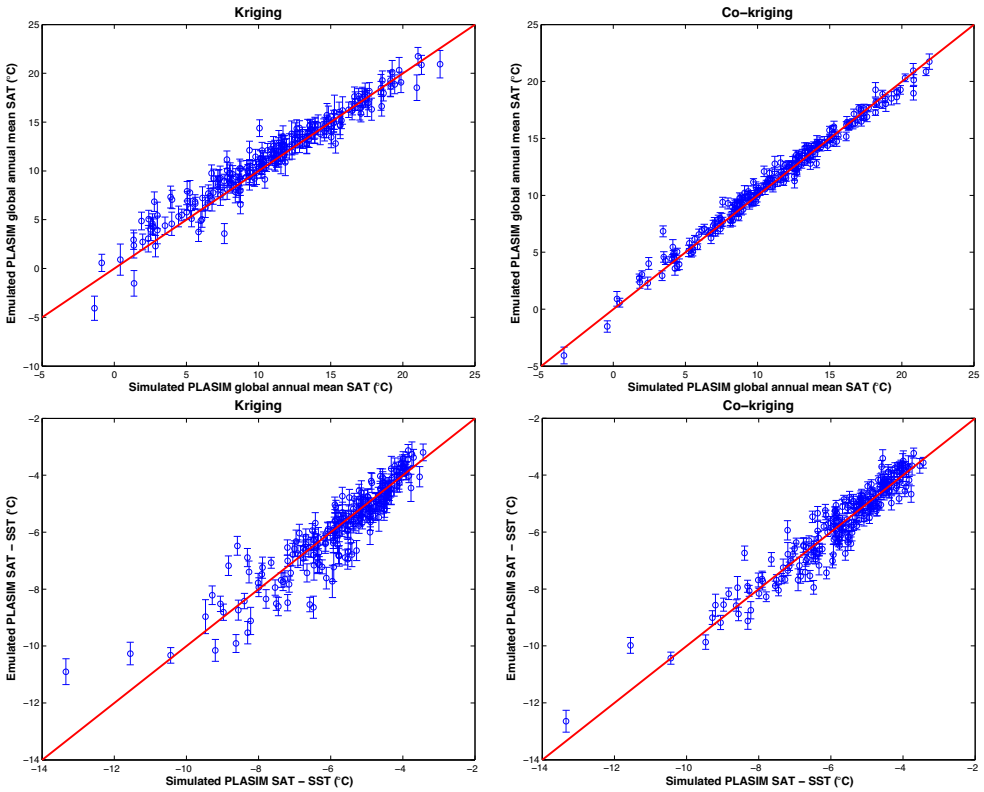


FIGURE 3.5: The upper panels show PLASIM simulated global mean SAT at the 214 validation points plotted against their emulated values from both kriging (left) and co-kriging (right) emulators. The error bars indicate a 2 standard deviation interval at each point. The lower panels show the results of the global mean SAT-SST emulators.

simulated and emulated values from the co-kriging emulators are over 0.90 for both. The standard deviations across the ensembles are slightly overestimated in both emulators. From the figure, the emulated values can be seen to deviate more for larger anomalies.

The uncertainty in the emulator predictions, arising from not having evaluated the model at untried input configurations, is called the 'code uncertainty' (O'Hagan, 2006). An advantage of the GP emulator employed is that we can quantify this uncertainty, which is represented as the error bar at each prediction in Figure 3.5. The additional information from the cheap training data appears to reduce this estimated uncertainty for the co-kriging emulator.

The error bars in all the emulators (both kriging and co-kriging) constructed here appear to be underestimated. A possible cause for this is the use of the smooth correlation structure described in Eq 2.6. As discussed in the previous section, the emulator's uncertainty where a training point exists is zero and increases away from said point. When the correlation between that point and its neighbours are strong, the error bar grows slowly. Hence, an overly-smooth correlation function can lead to the underestimation of the uncertainty range between training points. Further work employing a coarser covariance structure such as the exponential covariance function with $p = 1$ or a Matérn

class covariance function (Rasmussen and Williams, 2006; Diggle and Ribeiro Jr, 2007) are desirable to investigate this issue.

3.4.3 EOF decomposition

The following analysis attempts to explain the processes and parameters that determine the spatial distributions of SAT in GENIE-1 and PLASIM using PCA. SVD was applied to two $2048 \times n$ matrices of EMBM and PLASIM SAT fields, where $n = n_c = n_e = 660$. Over 99% of the variance across the ensemble in these fields can be explained by the top ten EOFs, as shown in Table 3.5. This indicates that they are sufficient to generate a good approximation to the simulated responses. As suggested from the emulator for global mean SAT, less than 600 points would be enough for the emulators. To ensure that the decomposition is robust, SVD is applied to smaller subsets ($n = 30$ to $n = 250$). The EOFs appear to be qualitatively the same. Only minor quantitative differences are obtained; therefore, the EOFs and PCs are judged as robust and representative of the ensemble behaviour. These subsets are chosen using the same exchange algorithm mentioned in Section 3.3 to obtain designs that give the best space-filling Morris-Mitchell criterion (Morris and Mitchell, 1995).

TABLE 3.5: Percentage of variance in SAT, explained by the first 10 EOFs for GENIE-1 with EMBM and with PLASIM. The 150-member ensembles are used to obtain these values.

	EMBM	PLASIM
EOF 1	86.33%	79.53%
EOF 2	11.27%	8.62%
EOF 3	1.55%	6.85%
EOF 4	0.47%	2.61%
EOF 5	0.10%	0.43%
EOF 6	0.07%	0.57%
EOF 7	0.05%	0.30%
EOF 8	0.03%	0.21%
EOF 9	0.03%	0.16%
EOF 10	0.03%	0.07%
Total	99.93%	99.35%

The high percentage of variance explained by the retained EOFs means that by successfully emulating them, the SAT field of PLASIM can be accurately estimated. For EMBM data to be useful, its EOFs and PCs need to carry meaningful information about PLASIM's modes. To verify this, an analysis of the EOFs and PCs of the two models are carried out.

The first EOFs of SAT in both models are illustrated in Figure 3.6. Their corresponding PCs are emulated as functions of the model parameters using universal kriging

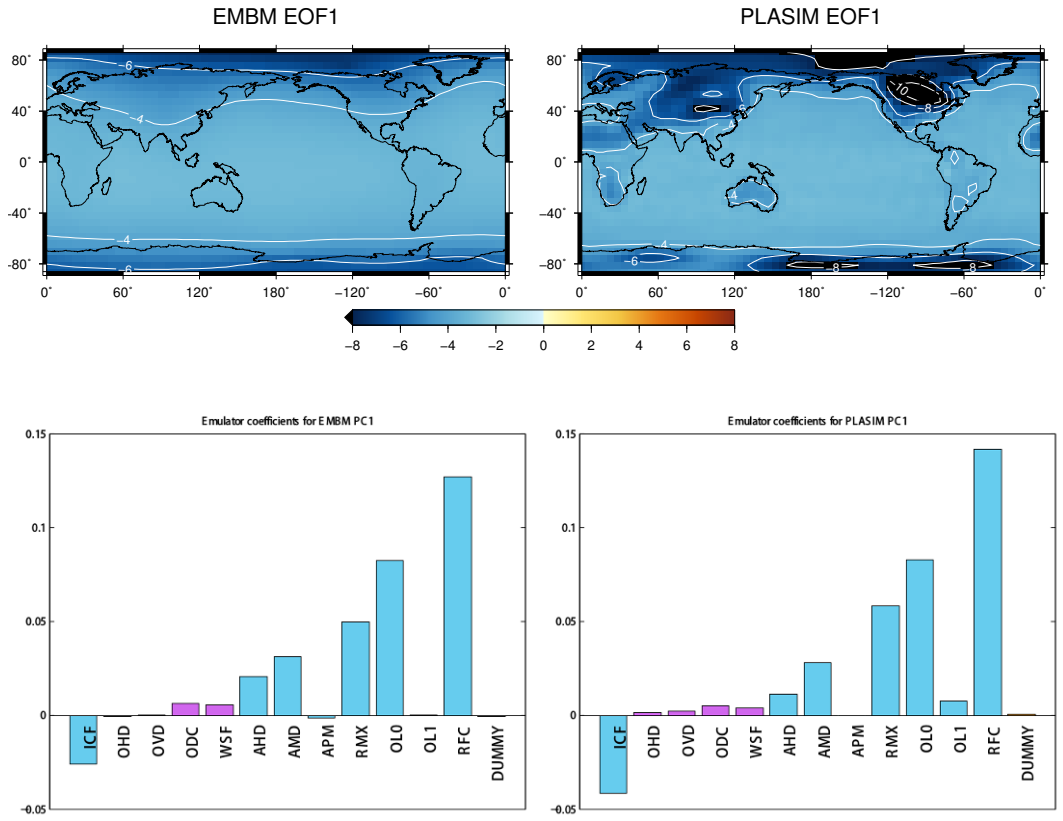


FIGURE 3.6: The first EOFs of EMBM and PLASIM SAT (upper) and the universal kriging emulator coefficients of their corresponding PCs (lower). All 600 data points are used to train each of these emulators. The black cells in PLASIM EOF1 indicate values lower than the plotted range. Contours are drawn over both plots at a 2°C interval.

(Section 2.1.1). Also shown in this figure are the coefficients of the prior mean, β (as described in Eq. 2.4), which reflect the relative importance of the parameters in determining each PC. These coefficients are the gradients of the linear mean function fitted to the data. Each coefficient corresponds to a dimension or an input parameter. They are not purely objective measures since their values depend on the ranges over which the parameters were varied. Also, the mean function is linear, so they do not contain information on the non-linearity of the emulated function. They also inherit uncertainties from the statistical estimation.

The first EOF for both models is the same sign globally, suggesting a change in the radiation budget due to the greenhouse gas and the albedo effects. The effects due to changing glacier condition and atmospheric CO_2 concentration are accentuated in PLASIM because corresponding changes are taken into account in PLASIM. According to the emulator coefficients, the largest contributions are due to RFC, OL0, RMX and ICF in both PLASIM and EMBM. Large values of ICF result in a lower global mean SAT due to higher albedo. Large values of RFC, OL0 and RMX, on the other hand, have the opposite effect on global mean temperature due to more heat being absorbed

by the increased greenhouse gas content in the atmosphere. Hence, ICF has the opposite sign to RFC, OL0 and RMX.

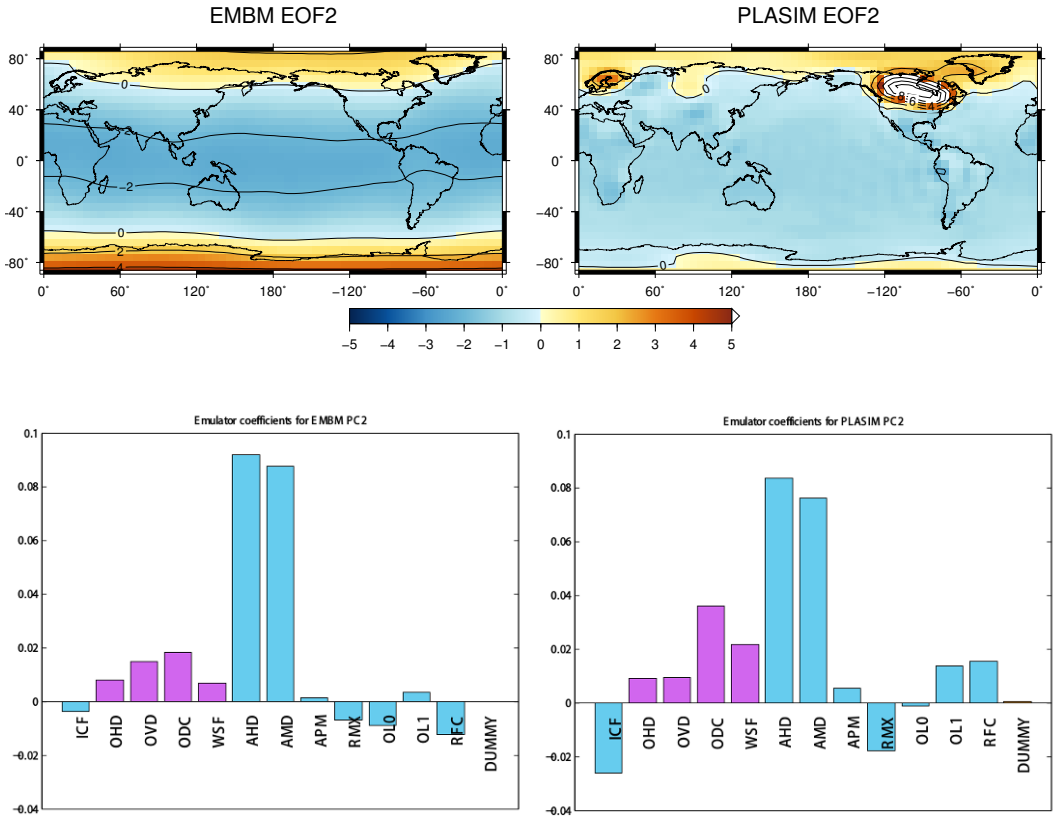


FIGURE 3.7: The second EOFs of EMBM and PLASIM SAT (upper) and the universal kriging emulator coefficients of their corresponding PCs (lower). All 600 data points are used to train each of these emulators. The white cells in PLASIM EOF2 indicate values higher than the plotted range. Contours are drawn over both plots at a 2°C interval.

The second EOFs in EMBM and PLASIM exhibit changes of opposite sign at Equator and polar regions, reflecting a redistribution of the heat budget due to changes in the efficiency of poleward heat transport (Figure 3.7). The parameters controlling heat diffusivity in the atmosphere (AHD and AMD) play the largest role in this process. While they dominate the signals, there are smaller contributions from the ocean heat diffusivity parameters (OHD and OVD), which have similar but smaller effects compared to AHD and AMD. Other small signals do not necessarily agree with each other; i.e., RFC has opposite signs in the two models.

With emulator coefficients of approximately 0, the dummy variable (FFX) is correctly identified as an inactive parameter in all cases (Figure 3.6 and 3.7), giving us more confidence in using the coefficients. Any parameter with coefficients of comparable magnitude to FFX is also assumed to be inactive, such as OHD and OVD for EMBM and PLASIM's first EOF.

These EOFs indicate similar modes of variability in GENIE and PLASIM, fulfilling the assumption made for co-kriging. The extra training points from EMBM, therefore, are expected to provide inference on PLASIM's behaviour. Each pair of PCs from EMBM and PLASIM form a set of cheap and expensive training data for the corresponding emulator. Even though this is applied to all ten PCs, according to Table 3.5, only the first four modes contribute significantly to the total variance. Lower-order modes appear indistinguishable from noise. It is difficult to emulate them independently, and so it is unlikely that any meaningful relationship between them can be found by co-kriging. Although all 600 data points are used to train each of these emulators, results obtained from smaller subsets (the smallest subset tested contained 50 simulations) show no systematic differences.

The assumptions made for Eq. 2.10 are expected to hold in the case of emulating PLASIM's PCs. The emulator coefficients in Figure 3.6 and 3.7 show that the PCs of the two models exhibit similar trends due to the varying input parameters. The difference in the magnitude of the contributions from these parameters should be sufficiently approximated using a scaling constant, ρ, and a stochastic process, f_d. The spatial pattern in PLASIM, however, depends on the EOFs and so different regional responses compared to EMBM can still be emulated using this method.

3.4.4 Emulation of 2-D output fields

We retained the first 10 EOFs of EMBM and PLASIM SAT, which describe 99.93 and 99.35% of the simulated ensemble variance, respectively (Table 3.5). Each field can be approximated as a linear combination of these ten EOFs, scaled by their respective PCs according to Eq. 2.25. By using this method of dimensional reduction, only ten emulators or fewer are needed instead of 2048 emulators if individual grid point is emulated. Both kriging and co-kriging emulators are then constructed for each of these PCs.

Using the same procedure as described in Section 3.4.2, exploratory exercises show that approximately $n_c = 150$ training points are needed to obtain a good emulator of the EMBM SAT fields. The cheap data are, therefore, the 150 indices of each of the first ten rotated PCs of the 2048×150 matrix of GENIE data. It is found that at $n_e = 50$, we obtain a co-kriging emulator that validates well against simulated values.

Kriging emulators using only the expensive data from PLASIM are also constructed for comparison. Again, co-kriging outperforms kriging when the same 50 expensive training points are used. More expensive points are then added to the kriging emulators and for approximately 150 points, similar RMSE and r^2 are obtained. Therefore, co-kriging reduces the required expensive training data to one-third of the amount needed when using kriging.

TABLE 3.6: Validation of each PC emulator using the 214-member ensemble. The correlation coefficients show how well matched the emulated PCs are compared with the simulated values. The co-kriging emulator uses 50 expensive points and 150 cheap points while the kriging emulator here uses the same 50 expensive points.

	1	2	3	4	5	6	7	8	9	10
Kriging r^2	0.91	0.75	0.84	0.50	0.15	0.04	0.00	0.09	0.24	0.00
Co-kriging r^2	0.97	0.83	0.84	0.64	0.18	0.05	0.02	0.10	0.24	0.00

The co-kriging (trained with 50 expensive and 150 cheap points) and kriging (trained with 50 expensive points) are validated using the 214-member validation set. Both the individual PCs and the final reconstructed SAT are validated against true values. First, to test the emulator's ability to reproduce PC values, each emulated PC is validated against those decomposed from the simulated ensemble (Table 3.6). For the first score, the co-kriging emulator validated very well with an r^2 value of 0.97. Lower-order PC coefficients are generally harder to emulate; hence, the value of r^2 decreases down the list. It is possible that they reflect physical processes that are more difficult to represent as simple functions of the input parameters or simply represent stochastic processes.With a low value of r^2, the emulator does little more than adding some random noise, e.g. from the 6^{th} to the 10^{th} PCs, with the exception of the 9^{th}. There are several reasons for this. First, the PCs of EMBM might reflect random noise and so cannot be emulated. Since the cheap emulators are not meaningful, the expensive ones can gain no useful information. Second, PLASIM's PCs might be noise and co-kriging fails to work for the same reason. Finally, the relationships between EMBM and PLASIM PCs might not have been successfully determined. This either means that EMBM did not contain the information on these PLASIM's modes, or the emulator fails to determine it. Even though the signal from the 9th mode is very small, it was emulated with some success. Despite the fact that mode 6, 7, 8 and 10 were not emulated successfully, co-kriging still performs either comparable to or better than kriging.

The ten co-kriging emulators of PLASIM PCs are then used to reconstruct the SAT fields at each validation point. To validate the simulated SAT fields, the quality of the individual emulations and the spatial pattern of the emulated field are tested. In order to test the proportion of the total ensemble variance captured by the emulator:

$$V_T = 1 - \sum_{n=1}^{214}\sum_{i=1}^{2048}(S_{n,i} - E_{n,i})^2 \Big/ \sum_{n=1}^{59}\sum_{i=1}^{2048}(S_{n,i} - \bar{S}_i)^2 \tag{3.2}$$

where $S_{n,i}$ is the simulated output at grid cell i in the n^{th} member of the validating ensemble, $E_{n,i}$ is the corresponding emulated output and $\bar{S}_i$ the ensemble mean simulated output at grid cell i. V_T assesses the error in the emulator for each simulation, averaged over the 214 simulations, and measures the degree to which individual simulations can be

regarded as accurate. The RMSE values between each emulated and simulated surface fields are calculated and averaged across the whole validation set. V_T and RMSE are used in combination to assess the emulator validity.

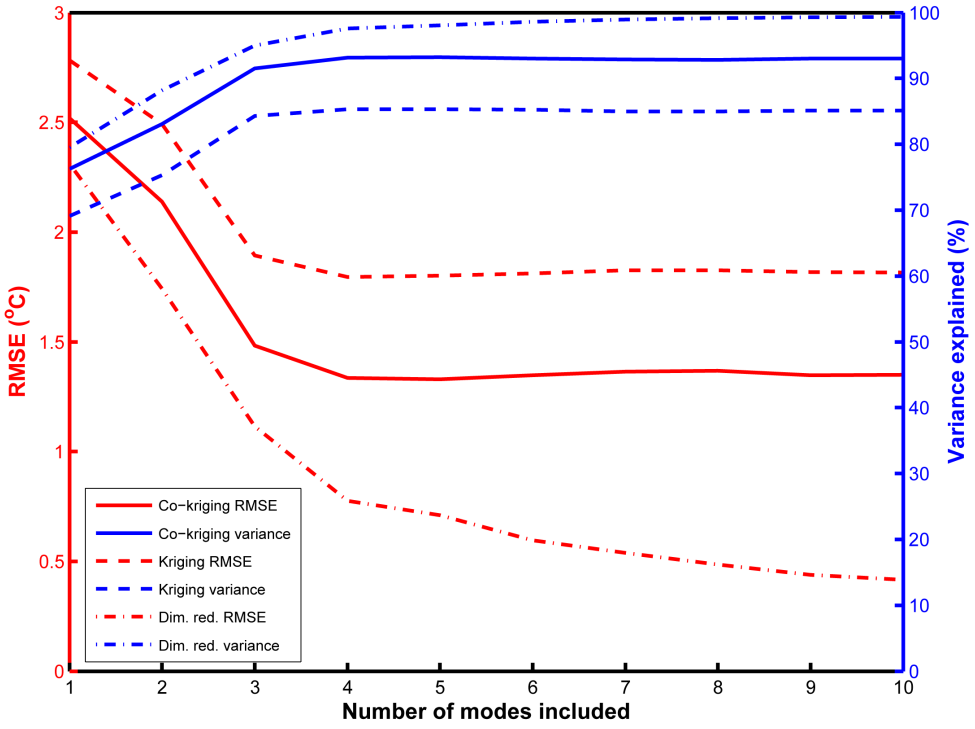


FIGURE 3.8: Comparison between kriging (dashed line) and co-kriging (solid line) emulators. The variance explained (blue) when each PC is added is shown together with the RMSE (red) of the corresponding reconstructed validation SAT fields. The dot-dashed lines represents the same values obtained if the emulator were perfect. The deviations of these line from RMSE= 0°C and V_T = 100% are errors introduced by dimensional reduction.

Figure 3.8 demonstrates the effect of each added PC to the value of V_T and RMSE. When only the first emulated component is considered, the co-kriging emulator reproduces 76.2% of the simulated variance (averaged over all space and all ensemble members), which is close to the 79.5% variance explained by the first EOF (Table 3.5). This is also reflected by the high degree of accuracy of the PC 1 emulator (Table 3.6). The addition of the next four emulated components brings the percentage of simulated variance being captured, V_T, to 93.2%, close to the total amount of 98.0% explained by the first five EOFs (Table 3.5). The average RMSE is 1.33°C, which is approximately 1.7% of the average spatial variation in temperature or 4.8% of the average variation across the whole ensemble at each grid point. The last five emulated PCs have a negligible effect on both V_T and RMSE. Among these, only the 9^{th} PC improves the overall result while the others worsen it. For the kriging emulators, the same behaviour is observed but with lower accuracies. The maximum variance explained by the kriging emulators is 85.3%. Also included in Figure 3.8 are lines corresponding to the validation results if

the emulators were perfect. These demonstrate the errors introduced by the dimensional reduction process.

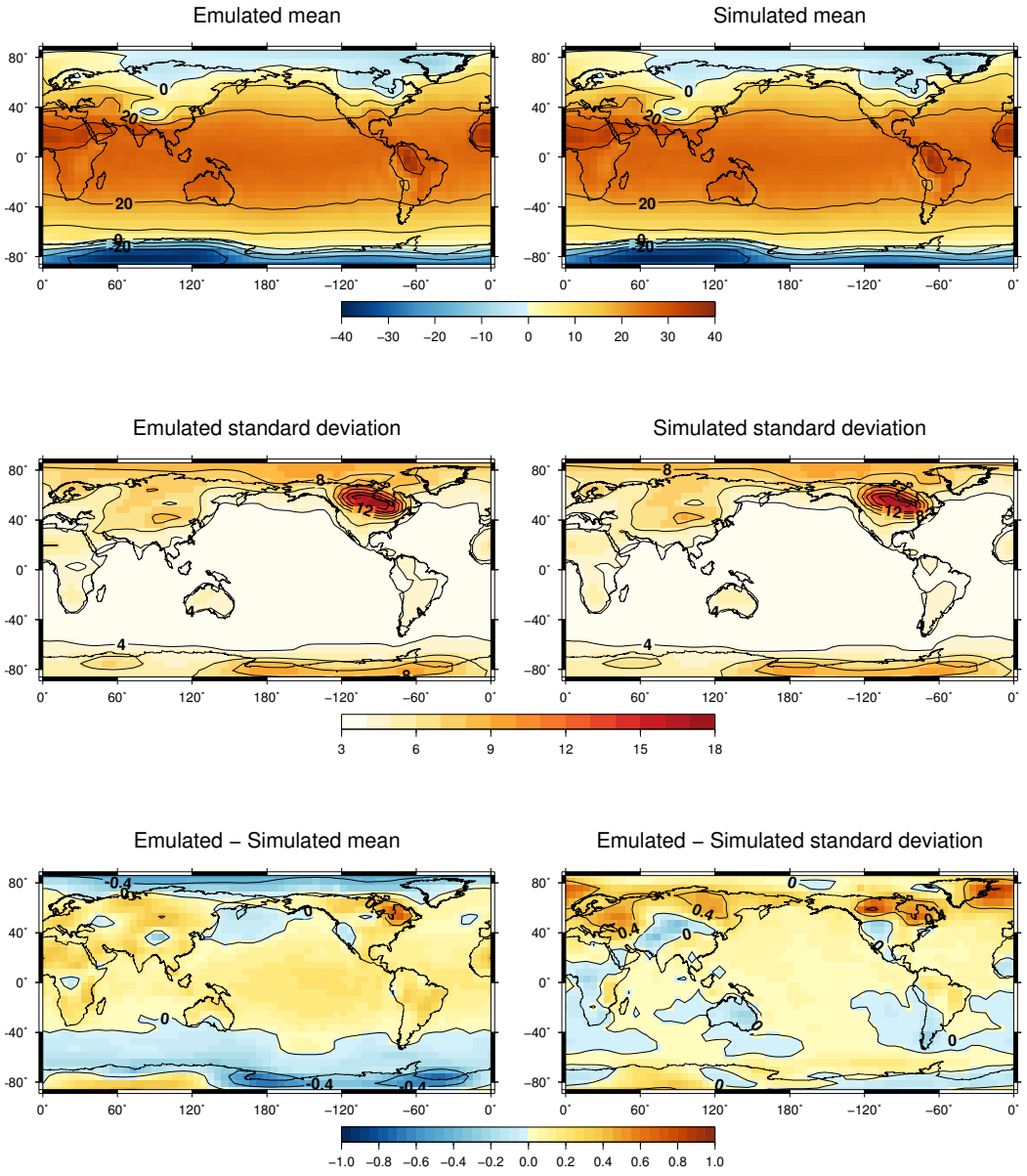


FIGURE 3.9: Mean and standard deviation of the emulated (upper and middle left) and simulated (upper and middle right) validating ensembles. The emulated - simulated differences in mean (lower left) and standard deviation (lower right) are also shown.

Figure 3.8 shows the emulated and simulated spatial pattern of the ensemble mean and standard deviation. The differences between these emulated and the simulated fields are within 1°C. Therefore, the ensemble behaviour is well reproduced. There is, however, a slight underestimation of the SD over the North American continent where the glacier mask is applied. The 2-D SAT emulator appears to underestimate the ensemble variance by a small amount. The error seen is a combination of the two types of errors introduced

in Section 2.2. Despite having very different outputs (Figure 3.3), the method proposed successfully utilises GENIE-1's EMBM output to aid the construction of PLASIM SAT emulator.

In the work presented here, only annually averaged fields are considered. The generalisation to emulate monthly average fields or seasonal cycles is straight forward. We simply have to replace the current 2048×1 annual-averaged maps with a 24576×1 map of the 12 monthly averaged fields.

3.4.5 Relationship with the coupled system

We have demonstrated that information from a cheap atmospheric model (EMBM) can be used to improve predictions of the steady-state behaviour of an expensive atmospheric model (PLASIM) in unsampled parts of parameter-/boundary-forcing space. This behaviour is a function of the boundary conditions on the atmospheric model (SST, longwave and shortwave radiative forcing), as represented in this statistical study by the 13 parameters. This technique has advantages when attempting to understand or project the decoupled response of individual climate system components to their boundary conditions. For example, in the context of impact assessment models, the spatial pattern of changes in SAT and precipitation is often needed to study the impact of climate change on areas such as health, land use and energy production. These spatial temperature and precipitation response patterns are obtained from climate models forced by arbitrary CO_2 concentrations resulting from particular policy decisions. Different statistical emulation techniques have been employed to reproduce the output of AOGCMs under a large range of forcing scenarios (Holden et al., 2014; Castruccio et al., 2014). Our multi-level emulation technique offers an alternative method to reproduce the key characteristics of an AOGCM using only a small training set, given a larger ensemble of a cheaper model of the same system, covering unsampled CO_2 concentrations. Another example where our technique can be applied is in emulating a carbon cycle model to provide an estimation of the atmospheric CO_2 concentration as a function of a time series of anthropogenic CO_2 emissions and non-CO_2 radiative forcing (Foley et al., 2015). CO_2 concentration from coupled climate-carbon cycle models can be emulated and replace the simple carbon cycle component often used in integrated assessment models.

In reality, changes to the climate system components that are focused on will feedback on other climate system components; i.e., if the present study were extended to the fully coupled system, differences in SAT, wind stress and the hydrological cycle between PLASIM and the EMBM would feedback on SST and sea-ice distribution.

Within this context, we now explore the relationship between the 'climate sensitivities' of the EMBM and PLASIM atmospheres, both forced by GENIE-EMBM SSTs as discussed above, before considering how our approach could in future be extended to the fully

coupled system. Our 600-member ensemble design generated in Section 3.3 is used as the basis of two new designs. ICF is fixed at 0 for both sets. Climate sensitivity is defined as the warming response to a doubling of atmospheric CO_2 from the pre-industrial values. Hence, a control set (CTRL) has RFC set to 278 ppm and another set ($2 \times CO_2$) has RFC set to 556 ppm. The emulators constructed in the previous section are used to predict the SAT fields resulting from these two designs. This process can be done within seconds, at almost no additional computational cost.

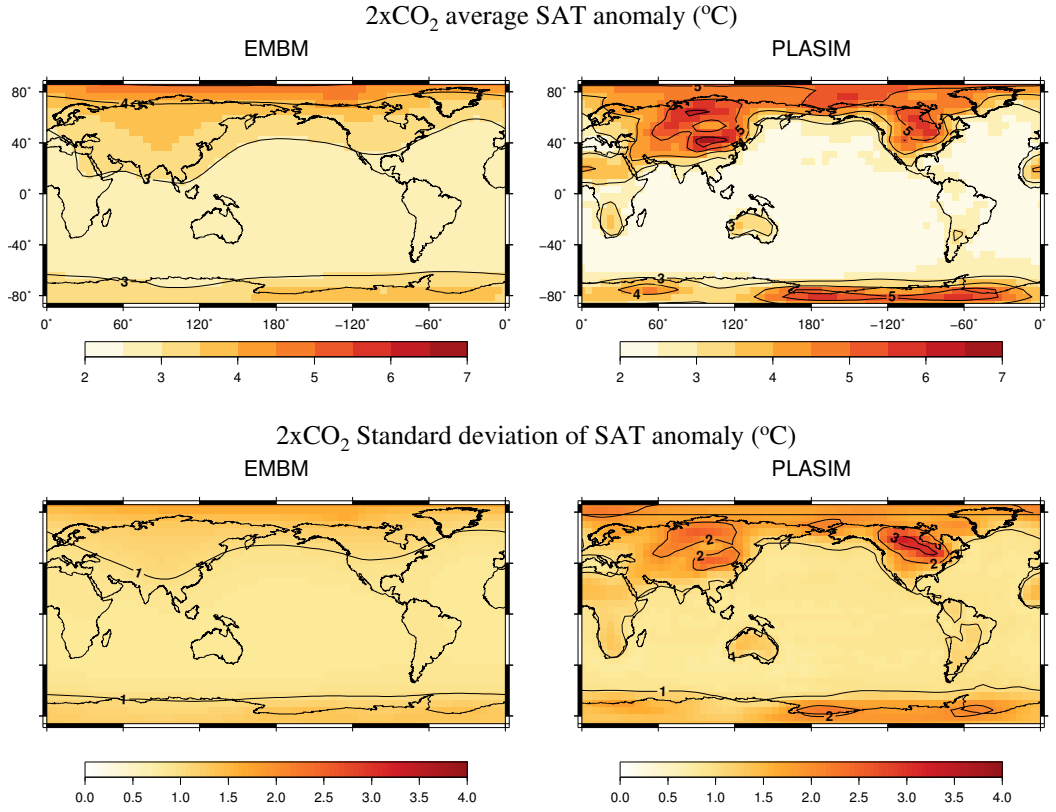


FIGURE 3.10: Mean (upper panel) and standard deviation (lower panel) of the SAT anomaly corresponding to a double in atmospheric CO_2 concentration in EMBM and PLASIM.

The average SAT anomalies due to a doubling of atmospheric CO_2 concentration for both models, $2 \times CO_2 -$ CTRL, are shown in Figure 3.10. The area-weighted global mean SST is used to calculate the probability distribution of climate sensitivity for the two models, shown in the upper panel of Figure 3.11. The means of the two distributions are ΔTCO_2 of 2.99 ± 0.91°C for EMBM and 3.37 ± 0.95°C for PLASIM. Figure 3.11 shows that the climate sensitivities in the two models have similar distributions with means differing by approximately 0.38°C. The range is broad due to the parameters varied. PLASIM displays larger changes in temperature over the continent in general and especially over high elevation areas (Figure 3.10). Because of this, the average anomaly ΔTCO_2 in a PLASIM simulation is larger than the corresponding value in EMBM. The relationship

between the two distributions is approximately linear, as shown in the lower panel of Figure 3.11. Since no PLASIM parameter is varied apart from ICF and RFC (which are both held constant in this experiment), PLASIM climate sensitivity is heavily influenced by the GOLDSTEIN surface conditions.

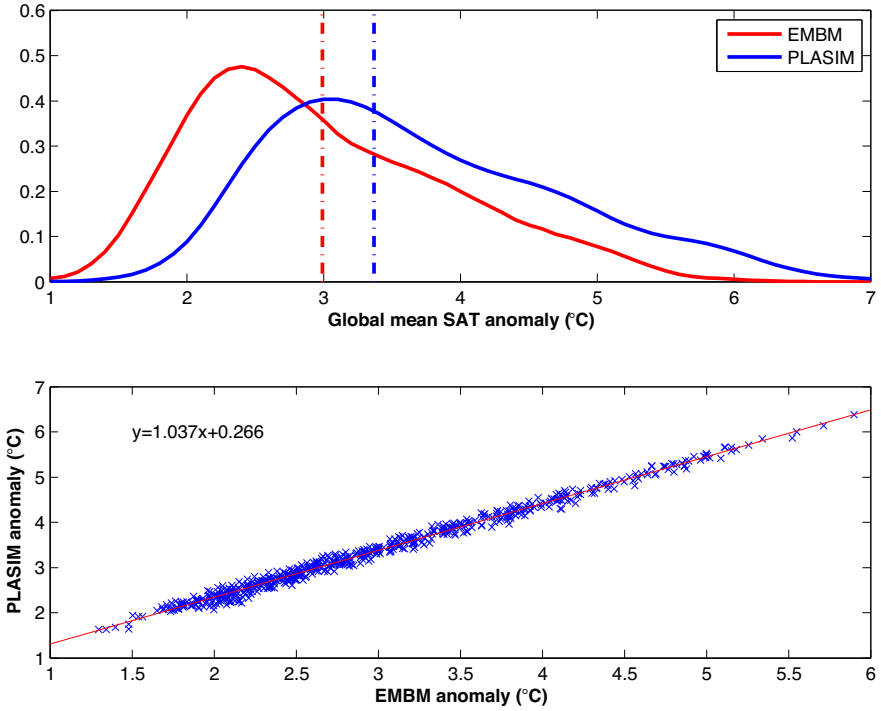


FIGURE 3.11: The upper panel shows the probability distributions of EMBM (red) and PLASIM (blue) climate sensitivities. The mean of each distribution is denoted by the dot-dashed line of the same colour. The lower panel shows a plot of PLASIM anomalies against EMBM anomalies. The coefficients of the linear function fitted through the data are included in the figure.

In a hypothetical coupled experiment, it is reasonable to speculate that the generally larger response of SAT to CO_2 in PLASIM than the EMBM would yield a broader range of SSTs in the GOLDSTEIN ocean, amplifying the difference in climate sensitivity between the two models. There are two alternative approaches that could be used to extend the technique described here to this fully coupled system. The first or 'direct' approach sees PLASIM fully coupled to GENIE-1's subcomponents, allowing for two-way interaction between the atmosphere and the ocean/sea ice. In this case, the current statistical technique can be applied directly to emulate atmospheric variables from PLASIM as functions of the ocean's parameters, using EMBM as the cheap approximation. How beneficial EMBM's information is in this set-up compared to the results presented in this chapter is uncertain without further work. The 'indirect' approach involves the coupling of PLASIM's steady-state emulators with GOLDSTEIN ocean and sea-ice components. The atmospheric outputs from PLASIM, such as SAT, precipitation and wind stress

can be emulated as a function of the prescribed SST and used, in return, as boundary conditions for the ocean. This framework would be able to capture some processes, which are currently not adequately modelled or not represented at all in EMBM. There are certain implications for when such a framework would be useful since the emulators are built upon a collection of steady-state simulations where only one-way interaction between PLASIM and the ocean component is available. This type of framework would not be suitable in the context of processes such as ENSO (El Niño-Southern Oscillation) in which the atmosphere and ocean vary together on interannual timescales. However, it may be useful when events with much longer timescales, where the atmosphere can be regarded as being in equilibrium with the ocean, are considered. While information on chaotic higher-frequency atmospheric variability is lost, extra information from the higher-fidelity atmospheric model is gained without incurring a large computational cost.

3.5 Summary and conclusions

We have described in this chapter the development and evaluation of large ensembles of GENIE-1 and PLASIM simulations for application in statistical emulation.

For this work, we employ the non-parametric fitting method of Gaussian process emulation. Two variations of this well-established method, kriging and universal kriging, are briefly described in Section 2.1.1. Compared to polynomial fitting techniques, such as the one employed by Holden et al. (2014), this approach provides an estimate of the uncertainty introduced by the emulation process, also referred to as 'code uncertainty'.

To efficiently extend this method from emulating scalar output to emulating high-dimensional output, e.g. the 2-D SAT fields, principal component analysis is used. This powerful technique decomposes the surface output fields of both EMBM and PLASIM models into EOFs, scaled by the respective PCs. However, the EOFs are statistical modes, and direct connection to physical processes cannot always be drawn directly. Emulator coefficients of the PCs corresponding to these modes, however, can provide a link between them and the varying model parameters, allowing for better interpretation of the model behaviour. It also allows us to identify and preserve the correlation between grid cells.

Here, the first five PCA modes are emulated instead of individual grid cell values, reducing the computational cost significantly. Although not explored in this work, the links between different model outputs may also be exploited to allow for further reduction of dimensions when emulating multivariate output.

A multi-level emulation technique, co-kriging, is used to build both scalar and high-dimensional output emulators for PLASIM with additional information from EMBM.

The constructed co-kriging emulators successfully estimate both the global mean SAT and the 2-D array of SAT fields of PLASIM as functions of the 13 GENIE-1 parameters. Being cheaper to evaluate, EMBM can be used to sample GENIE-1's parameter space more finely, providing information where PLASIM data are sparse. Despite being structurally unrelated, the link between EMBM and PLASIM is successfully established, resulting in PLASIM emulators being built using a smaller amount of expensive data. The combination of PCA with co-kriging allows us to emulate the spatial pattern of PLASIM SAT accurately despite the model having a different response to EMBM's. Emulated outputs are validated against simulated values using a separate validation ensemble. Both spatial pattern and magnitude of SAT are well reproduced across the ensemble. Apart from the ensemble mean and standard deviation, individual simulations are also successfully emulated with high accuracy. The emulators, however, show a tendency to underestimate the variance spatially and across the ensemble. This is unavoidable because of the dimensional reduction process. The quantification of the emulator uncertainties are beyond the scope of this work and should be explored in further studies in order to improve the emulators' performance. We will revisit this point briefly in Chapter 6.

Here, we have focused only on SAT, but this method can be applied to other variables of the atmosphere, such as precipitation (PPTN) or wind fields. In the case of PLASIM, co-kriging emulation of PPTN using GENIE's PPTN field as a fast approximation is not likely since the description of this field in the two models differed quite significantly. The same goes for other PLASIM quantities, which have no equivalent in EMBM. However, it is possible that other GENIE-1 fields might be more suitable as the fast approximation to PLASIM's PPTN, e.g. SST or elevation. Work has been done in the past using elevation as a fast approximation for PPTN (Hevesi et al., 1992).

This work establishes the technique for emulating the equilibrium response of the model. Compared to available efficient frameworks such as the MIT IGSM-CAM (Massachusetts Institute of Technology - Integrated Global System Model linked with the National Center for Atmospheric Research (NCAR) Community Atmosphere Model, Monier et al., 2013, a present limitation of this technique is in the scope for two-way coupling (e.g. in the present study the PLASIM atmosphere passively responds to the ocean). However, a future study will show that it is possible to emulate the atmospheric fields (precipitation, surface winds, etc.) that directly influence other model components and use these as boundary conditions. This technique has the limitation that the atmosphere is treated as being in a steady state with the ocean, so that the effect of interannual variability cannot be explicitly represented, but would nevertheless be of value for modelling long-timescale phenomena such as glacial-interglacial cycles.

We have demonstrated that multi-level emulation across structurally unrelated models provides useful information more efficiently than using either model in isolation. Several challenges remain before a coupled model making use of such an emulator can be

constructed, and the steady-state vs. transient issue is one of them. The seasonality, which is currently lacking, will also be included by the modification described in Section 3.4.4. PLASIM's parameters, which do not have an equivalent in EMBM, are not yet considered. The current experiment design does not allow for the effect of aerosols, sea ice or vegetation to be studied. It simply attempts to improve the current simulated climate in GENIE-1 by incorporating the dynamics of the PLASIM atmosphere. The role of these parameters will likely be explored in future studies.

The advantage of the emulation technique used here is that it does not depend on a fixed set of models and can be applied to a wide range of models for different applications. It also provides a useful tool in coupling models of different fidelity and resolutions. The emulators, however, are built for specific applications and so care should be taken to avoid extrapolating beyond the emulated space.

In conclusion, the work presented here demonstrates a concept with applications in not only climate research but extending to a wide range of problems where multi-level computer models are available.

Chapter 4

Cross-variable emulation

4.1 Introduction

In the previous chapter, we presented the emulation process of a scalar quantity and a 2-D surface field: the global annual mean SAT and the 2-D annual mean SAT field. The high-dimensional output of an atmospheric model, PLASIM, was emulated as a function of scalar model inputs (GENIE-1's parameters) using the dimensional reduction technique employed by Holden et al. (2010). In this chapter, we would like to extend the application of this method to emulate monthly atmospheric quantities, namely the zonal and meridional components of surface wind velocity, specific humidity and precipitation rate by relating them to the prescribed boundary forcing conditions given by GENIE-1's SST and sea-ice.

Together with SAT, these atmospheric variables are often desirable in studies of the climate system. In general, information on precipitation is vital to the understanding of the hydrological cycle and hence the climate system as a whole. In the context of future climate change, the spatial and temporal distribution of these prognostic variables are often needed in integrated assessment models (IAMs), where they are critical to policy advice on a wide range of human activities such as agriculture, power generation and flood/drought monitoring. The responses of precipitation and atmospheric circulation to changes in boundary condition, such as the presence of mountains or continental ice-sheets, are also likely to play a major role in understanding past climate. For example, Löfverström et al. (2016) and Ullman et al. (2014) studied how the different configurations and elevations of the continental ice-sheets affected the glacial climate through changes in the large-scale circulation. The changes in stationary waves induced by the ice sheet topography may alter the precipitation field in the surrounding area (Roe and Lindzen, 2001). Being able to capture these responses could significantly improve the simulated glacial climate in an EMIC with a low complexity atmosphere like GENIE-1.

It is vital, however, for the EMIC to remain an efficient framework capable of simulating the climate system on multi-millennial timescales or producing large perturbed ensembles for uncertainty quantification.

Motivated by this, we aim to construct an AGCM emulator using PLASIM that is capable of coupling with GENIE-1's subcomponents. The variables to be emulated are the surface winds, surface specific humidity and precipitation rate since they control the fluxes exchanged between the atmosphere and the ocean and sea-ice. The calculation of outgoing longwave radiation depends on the surface relative humidity and surface temperature. Sensible heat flux depends on the air-surface temperature difference and the surface wind speed. Latent heat release is proportional to the precipitation rate (Weaver et al., 2001; Edwards and Marsh, 2005).

These variables are more challenging to emulate compared to SAT since, unlike SAT, these fields, with the exception of humidity, are not as strongly constrained by the provided SST and sea-ice. Moreover, the method introduced in the previous chapter is not immediately applicable in this case since the fields we are interested in might not have a direct fast approximation in EMBM. Here we consider three different situations: i) Humidity is simulated in EMBM and is likely to be suitable as the fast approximation of PLASIM's humidity. ii) While present in both models, precipitation is parameterized differently in EMBM (Section 3.2) and it's unlikely that EMBM's precipitation contains very useful information on PLASIM's field. iii) Unlike these two, surface winds are not simulated in EMBM. Therefore, there is no direct fast approximation of PLASIM winds in EMBM. Our aim here is to introduce modifications to our previously established multi-level emulation technique so that it can be applied to all the three cases stated above.

First, to couple PLASIM emulators with GOLDSTEIN, the monthly exchange fluxes are required. To capture the seasonal variations, monthly surface fields are used instead of the mean annual field. The spatio-temporal behaviour of each emulated variable is captured using PCA as a data reduction technique. Second, instead of being emulated as a function of GENIE-1 inputs, PLASIM's output fields are estimated as a function of the prescribed boundary forcing conditions specified by GOLDSTEIN ocean and sea-ice. This modification, which will be described in Section 4.2, enables PLASIM responses to any particular SST and sea-ice configuration to be obtained, allowing exchanges from the ocean to the atmosphere at any desirable timestep. The dimensional reduction of the high-dimensional inputs has been demonstrated in Holden et al. (2015) by relating the vegetation field to the temperature and precipitation fields. This approach was shown to perform comparably well to the case where the vegetation field was emulated directly from the parameters controlling the temperature and precipitation fields. This approach also provides a tool for understanding the processes connecting the input and output fields. Finally, the last modification is to use combined PCA (Kutzbach and Guetter, 1986; Bretherton et al., 1992) to identify coupled patterns between two fields.

This technique allows the common modes of variability between, for example, SAT and precipitation rate to be obtained. The climate response of PLASIM and the relationships between the variables of interest with SAT are examined in Section 4.3. By linking each PLASIM variable with the SAT field, information from EMBM SAT can once again be incorporated into our multi-level emulation technique. The constructed emulators for the four atmospheric variables are described and validated in Section 4.4.

4.2 Dimensional reduction of emulator inputs

The simulations used in this chapter for training and validating the emulators are from the same 600-member and 214-member ensembles described in Section 3.3.1. All the inputs (SST, SIC and SIH) and outputs (SAT, precipitation, wind speeds and humidity) concerned in this chapter are 2-D fields. As seen in the previous chapter, each annual mean surface field contains 64×32 grid cells and can be reshaped into a 2048×1 state vector. Since we are interested in the capturing the seasonal variation, for each variable, there are now 12 state vectors corresponding to 12 calendar months for each simulation run. PCA has been used in the previous chapter to reduce the spatial dimension of the SAT. It has also been used in various studies to reduce the time dimension of an evolving timeseries. Challenor et al. (2010) used PCA to reduce the time dimension of AMOC timeseries while Holden et al. (2015) employed PCA to reduce both the spatial and temporal dimension of their SAT field, allowing them to emulate the response of the climate system at different time slice. Other methods for reducing the dimensionality of simulator's inputs prior to emulation are discussed in Maniyar et al. (2007). A decision tree designed to help modellers choose between several different dimensionality reduction approaches was also presented in their work.

Following Holden et al. (2015), for each atmospheric variable of an ensemble member, a new state vector is constructed by concatenating together all 12 state vectors, giving a spatio-temporal field $\mathbf{U} = [\mathbf{U}_1, \cdots, \mathbf{U}_{12}]^{\mathrm{T}}$, which has the dimension of 24576×1. A training set of n members can be represented as a $24576 \times n$ matrix. PCA is then performed, via SVD, on this new matrix in the same way as before.

The initial conditions driving each PLASIM simulation are provided by GOLDSTEIN's three prognostic variables, SST, sea-ice fractional cover (SIC) and sea-ice height (SIH). Therefore, the total dimension of our new input fields is $64 \times 32 \times 12 \times 3 = 73728$. Clearly, without reducing the dimensionality of the input fields, emulation would be infeasible. Fortunately, these input are likely to contain correlation structures within each field and between the three fields. Similar to the PLASIM outputs, these fields can be reduced using PCA into a smaller and manageable set of inputs. The importance of the inputs, currently spread among the 73728 dimensions, are redistributed within a new set of inputs in which the first few components explain the majority of the variation across the

ensemble, allowing redundant information to be identified and removed. Since this PCA step is applied to both the inputs and outputs of PLASIM, we shall distinguish them by referring to the output to be emulated as $\mathbf{U}$ and the input for emulator as $\mathbf{W}^i$, with $i = 1, 2, 3$ corresponding to SST, SIC and SIH respectively. Since these quantities only exist in oceanic grid cells, the land mask is removed from these fields, leaving the matrix $\mathbf{W}^i$ with the dimension of $16332 \times n$. PCA is then applied to each field independently (equation 2.19), giving

$$\mathbf{W^i} = \mathbf{L^i S^i R^i}^{\mathrm{T}}, \tag{4.1}$$

with $i = 1, 2, 3$. The columns of $\mathbf{L}^i$ are the EOFs of SST, SIC and SIH when $i = 1, 2, 3$ respectively. These are the spatial patterns of different statistical modes identified within these fields, ordered in decreasing importance. The columns of $\mathbf{R}^i$ give the coefficient scores while the diagonal of $\mathbf{S}^i$ gives the singular values. The new inputs which we'll use for the emulators are the principle components, $\mathbf{Z}^i$:

$$\mathbf{Z^i} = \mathbf{S^i R^i}^{\mathrm{T}}. \tag{4.2}$$

Since the emulator inputs become the PCs obtained from the SVD instead of the model parameters, to provide a prediction corresponding to a new surface field, the PCs of that new forcing field need to be obtained. We can assume that the addition of a single surface field does not modify the ensemble behaviour significantly and simply add the extra field in with the training ensemble. The updated ensemble is then decomposed, giving the PC of the new surface field. This method should work relatively well when the training ensemble is large enough that the addition of a single field does not have a large impact on the decomposition result. However, this is not ideal because technically the new PCs correspond to a slightly different set of EOFs than the training one. Also, this requires repeated decompositions of a large matrix and thus is computationally wasteful. Therefore, we decided to simply project the new boundary forcing field onto the EOFs obtained from the decomposition of the training ensemble. This is a simple matrix multiplication and can be done relatively efficiently. There is a possibility that the obtained PCs are outside the range of the original training values. Thus, care should be taken that such instances do not go unnoticed. The two methods are compared by emulating the 2-D SAT, and the results are shown in Table 4.2. We will discuss this in more detail at the end of this section.

By retaining only high-order principle components $\mathbf{Z}$, the dimension of the input is reduced significantly. The variance explained by the first ten EOFs for these three fields are displayed in Table 4.1. Evidently, the top EOFs capture the majority of the variance across the ensemble for all three variables. One might decide to keep the first ten EOFs of each field which results in 30 emulator inputs. While this is a significant

reduction from 73728, it is still a large number, and further reduction is desirable. A quick inspection shows that only the first 3 EOFs of SST explain more than 1% of the ensemble variance. Moreover, sea-ice cover and thickness are both well correlated to SST. The correlation between the first PC pair of SST-SIC and SST-SIH are found to be 0.77 and 0.67, respectively. It is, therefore, highly likely that we can reduce the dimension of the inputs further.

TABLE 4.1: The variance explained by the first 10 EOFs of SST, SIC and SIH. The 600-member GENIE-1 ensemble was used to obtain these values.

	SST	SIC	SIH
1	85.04	59.89	67.13
2	8.94	12.01	22.50
3	3.43	4.17	4.57
4	0.79	3.37	1.84
5	0.53	2.16	0.85
6	0.25	1.75	0.75
7	0.22	1.29	0.45
8	0.12	1.11	0.39
9	0.09	1.01	0.24
10	0.07	0.83	0.15
Total	99.49	87.60	98.87

To identify redundant inputs for further reduction, we conducted some preliminary tests by, first, emulating SST as a function of the original 12 model parameters to study the ensemble climate and its relation to these parameters. This step also aids the intepretation of the PLASIM climate in later sections. Universal kriging (Section 2.1.1) is used since it provides the coefficients of the linear part of the emulator, allowing us to compare the relative contribution of each parameter. Figure 4.1 shows the first 3 SST EOFs and their corresponding universal co-kriging emulator coefficients.

The first mode, which explains 87.60% of the total variance across the ensemble, is a radiative forcing mode, dominated by the atmospheric CO_2 concentration, RFC. Other parameters with a significant contribution to this EOF are OL0, RMX, OL1 and ICF. A larger atmospheric CO_2 concentration causes the global surface air temperature to rise, which in turn, warms the ocean. Large values of OL0 and RMX have a similar effect since they also increase the effective greenhouse gas concentration in the atmosphere by allowing more water vapour to remain there. The larger coefficient of RFC is due to the wide range of CO_2 concentrations being varied and does not necessarily mean that the model is more sensitive to RFC. The effect of ICF is of the opposite sign to RFC since larger values of ICF corresponds to larger glacier coverage. Thus, increasing ICF causes a cooling signature globally through the albedo effect. The largest variations in this EOF are seen in regions at high latitudes where sea-ice coverage changes can amplify the temperature change.

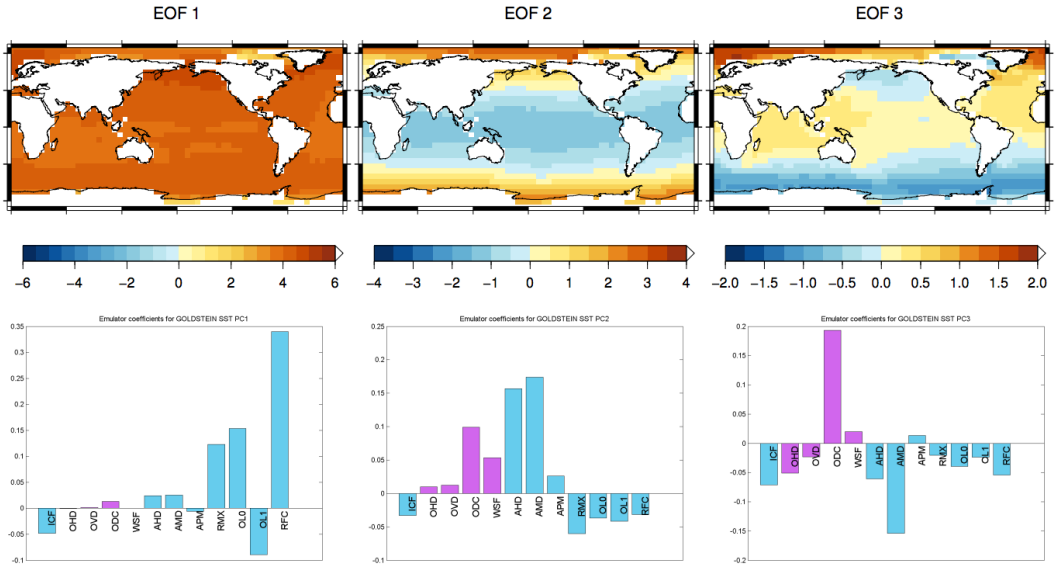


FIGURE 4.1: The spatial structure of the first 3 EOFs of SST (top panel). The corresponding emulator coefficients of the universal kriging emulators of these modes are shown in the lower panel. All 600 cheap simulations were used to train these emulators.

The second mode (8.24%) displays an increase or decrease in the equator to pole temperature gradient. A smaller gradient, seen in Figure 4.1 as a warming at high latitudes and cooling in the tropics, is induced by increasing the atmospheric moisture and heat diffusivity, AMD and AHD. As heat and moisture are carried towards the poles more efficiently, the meridional temperature difference is reduced. The inverse ocean drag coefficient, ODC, governs the parameterisation of the friction in the ocean and hence, the dissipation of momentum. An increase in ODC reduces the friction in momentum balance. This parameter also has a significant impact on the strength of the meridional ocean circulation in the North Atlantic. WSF is a wind-stress scaling factor, controlling the strength of the wind-driven gyres.

The third mode (2.02%) is dominated by ODC and AMD, parameters which influence the freshwater forcing and surface density in the North Atlantic. Large values of ODC lead to a stronger AMOC and hence more heat being transported northwards in the Atlantic. This can result in a warmer Arctic with less sea-ice coverage as seen in the spatial pattern of this EOF. Low AMD restricts moisture transport out of low latitudes, enhancing the development of the surface salinity at high latitudes, which leads to a stronger AMOC.

In general, similarities can be drawn between these SST modes and the SAT modes presented in Section 3.4.3. Differences such as the larger contributions from transport parameters and a smaller role from ICF are as expected. The role of the AMOC strength is also more prominent in SST.

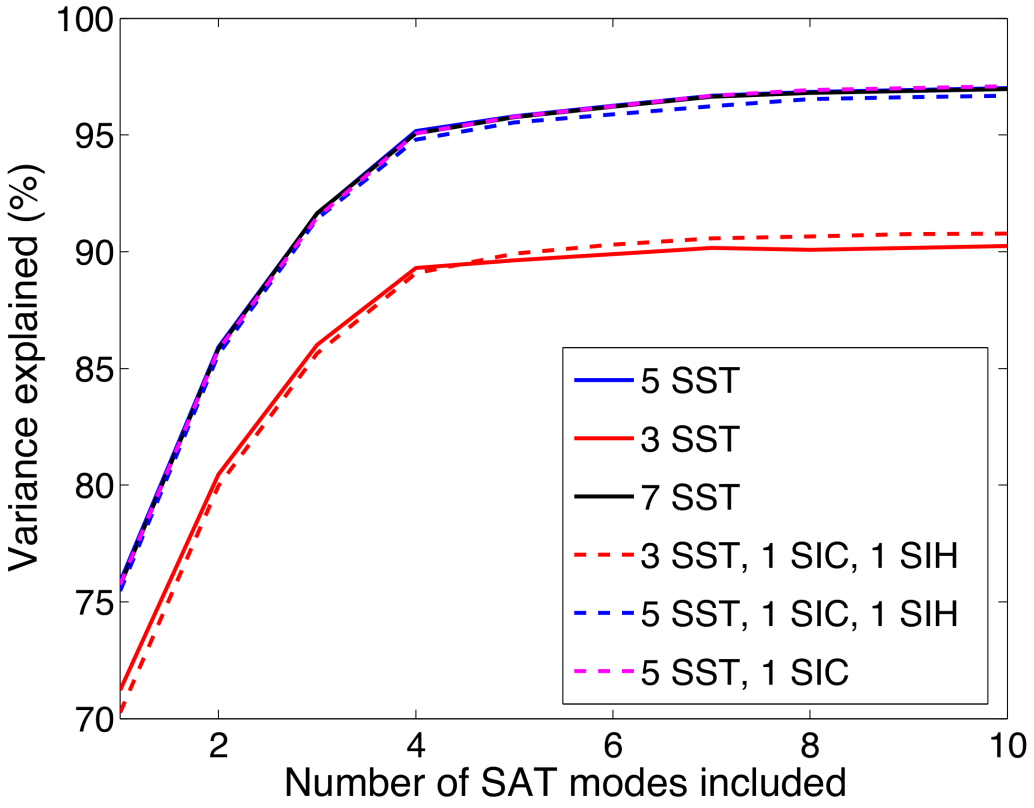


FIGURE 4.2: The variance explained by the SAT emulators using various combinations of inputs. The number of PCs used from each SST, SIC and SIH fields are indicated. All the emulators also include ICF and RFC as inputs and use the same amount of training points.

To better capture the effect of ICF and RFC in PLASIM, these two parameters are also introduced as emulator inputs. This is done because these two parameters are also varied in PLASIM. Their indirect effects through the GOLDSTEIN ocean are captured through the PCs while their direct effects will be captured from these two parameters. This will mean that the inputs are not independent of each other.

Considering the fact that only the first three modes of SST explain over 1% of the ensemble variance (Table 4.1), it is natural to assume that using them as inputs would be sufficient. Emulators of the EMBM 2-D SAT field similar to the one in Section 3.4.4 are constructed using the first 3, 5 and 7 SST PCs as inputs, in addition to ICF and RFC. Validation of the three emulators shows that the addition of the fourth and fifth PCs does improve the emulator performance significantly (total variance explained increased by 6.77%). The addition of even lower order PCs, in this case, the sixth and seventh components, do not contribute significantly nor positively to the final emulator. The variance explained by these emulators are shown in Figure 4.2. Both variance explained and the RMSE are shown in Table 4.2. While the 4th and 5th modes explain less than

1% of the total variance, it is possible that despite their small contributions to the global signal, they are important in explaining regional features.

TABLE 4.2: The variance explained by the first 10 EOFs of SST, SIC and SIH. The 600-member GENIE-1 ensemble was used to obtain these values.

Input combination	RMSE (°C)	Variance explained (%)
3 SST	1.59	90.23
5 SST	0.98	97.01
7 SST	0.98	96.96
3 SST, 1 SIC, 1 SIH	1.13	90.78
5 SST, 1 SIC, 1 SIH	0.96	96.67
5 SST, 1 SIC	0.96	97.08
5 SST (decomposition)	0.98	97.01

The two methods of obtaining PCs of a single new input field discussed earlier are also shown in Table 4.2. The results are practically the same, and so the more efficient method is adopted. The fact that the first method (indicated as the '5 SST decomposition' method in the table), is equally successful as the projection method (refered to as '5 SST') suggests that the EOFs obtained from the training ensemble is robust.

In a similar process, the addition of SIC or SIH PCs is tested. Three different combinations of inputs are used, i) The first three PCs of SST and the first PC of SIC and SIH, ii) The first five PCs of SST and the first PC of SIC and SIH, iii) The first five PCs of SST and the first PC of SIC. The addition of the sea-ice components has little effect on the performance of the emulator. In addition to these PCs, the two parameters ICF and RFC are included in all cases.

Ultimately, it is decided that the first five PCs from SST and the two PLASIM parameters are used as inputs for all emulators built in this chapter.. The dimensionality of the emulator input has been reduced from three 64×32 surface fields with two extra parameters to seven inputs.

4.3 Emulating multiple atmospheric outputs

4.3.1 Combined principle component analysis

In Section 3.4.4, PCA was used to reduce the dimension a single high-dimensional variable, the SAT. The link between two sets of PCs describing EMBM and PLASIM SAT was then determined using co-kriging. Since this convenient relationship is not guaranteed for other variables, we now explore the possibility of 'between variable' dimensional reduction. This idea stems from the fact that model outputs are not independent, and correlations exist between not only spatial fields and timeseries but also across output

types. For example, in an EMBM atmosphere, the SAT has a strong effect on both humidity and precipitation. Here, the relationship between the low-dimensional representations of different types of variables is examined. Figure 4.3 shows the scatter plots of the first 5 PCs of PLASIM SAT against the first 5 PCs of a) the zonal wind, b) the meridional wind, c) specific humidity and d) log-precipitation rate. It is common to use the logarithm of precipitation rather than precipitation, as the former is closer to being Gaussian distributed than the latter. Pairs with correlation over 0.5 are annotated in the figure. It is encouraging to see that there is a strong correlation between each pair, especially along the diagonal of each plot. In some cases, for example, the second PC of wind speed vs. the second PC of SAT, two nearly parallel branches can be seen in the scatter plot. These are the evidence that a bifurcation exists within our parameter space and the branches represents the two possible regimes. There are also indications of non-linear relationship, for example, between the first pair of PCs of humidity and SAT or precipitation and SAT.

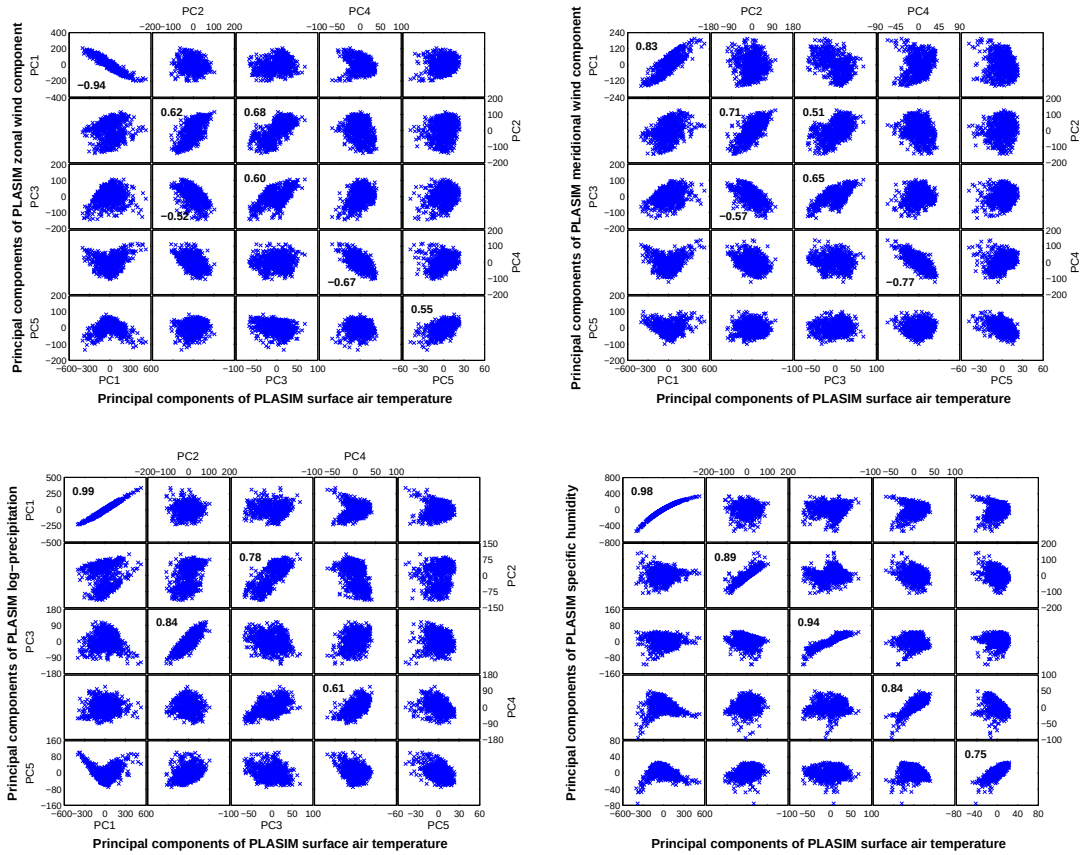


FIGURE 4.3: Correlation between the PCs of SAT and a) zonal wind velocity, b) meridional wind velocity, c) log-precipitation, d) specific humidity. The PCs shown are obtained using the 600-member ensemble of PLASIM simulations.

Given the strong correlation between these variables with SAT, 'between variables' dimensional reduction is applied on each pair of variables using combined PCA (also known as combined EOF). This method has been employed to emulate a combination of four

output time series from a simple climate model by Boukouvalas and Cornford (2009). This is a linear decomposition technique, and thus, nonlinear relationships between variables will be lost. However, by determining the common factor between each of these variables with PLASIM SAT, we can utilise EMBM's SAT as a fast approximation in the multi-level emulator. This technique would be advantageous over a single-level emulator of, says log-precipitation, using 'within variable' dimensional reduction, as long as the information gained from EMBM outweighs the nonlinear information lost from combined PCA.

While there is no limit on how many fields can be decomposed together, information on each individual field is lost when more are added for each variable; the SAT is decomposed together with only one other field each time. To do this, we simply concatenate PLASIM SAT and each field of interest together forming a new state vector

$$\mathbf{V^i} = [\mathbf{U}, \mathbf{U}^i]^{\mathrm{T}}, \tag{4.3}$$

where $\mathbf{U}$ is the $24576 \times n$ matrix of PLASIM SAT and $\mathbf{U}^i$ is the other atmospheric variable; $i = 1, 2, 3$ and 4 for zonal wind, meridional wind, log-precipitation and humidity, respectively. SVD is then applied to decompose the $49152 \times n$ matrix $\mathbf{V^i}$.

Since we are putting together two different quantities with different units, the standardised state vectors are used to avoid having one field dominating the decomposition result solely because of its large relative magnitude compared to the other field. For example, the range of SAT is about four times larger than that of the zonal and meridional wind speeds, and so it is likely that the resulting pattern from combined PCA will attribute a higher importance on SAT. The columns of the standardised state vector are given by

$$\tilde{U}_{ij} = \frac{\mathbf{U}_{ij} - \mu_i}{\sigma_i}, \tag{4.4}$$

where $\mu_i = \frac{1}{m}\sum_{j=1}^{m} \mathbf{U}_{ij}$ is the ensemble mean and $\sigma_i = (\frac{1}{m-1}\sum_{j=1}^{m}(U_{ij} - \mu_i)^2)^{1/2}$ is the ensemble standard deviation at the i^{th} grid cell. This transformation makes the data adimensional and treats the two field equally. It is possible to specify an arbitrary scaling constant to put more emphasis on one variable if desirable. However, we do not attempt this here. The standardisation leaves us with the same result as would be obtained using the eigen decomposition method (Section 2.2) on correlation matrices. Therefore, this is often referred to as combined correlation PCA (or combined correlation EOF).

This procedure is applied for each variable. The variance explained by the first ten modes averaged over 12 months are listed in Table 4.3. While these values are lower than their equivalence obtained from a simple EOF, they are still reasonable considering the nature of these fields. Unlike SAT and humidity, the proportion of variance explained by the first ten PCs is lower for the wind speed components and precipitation rate.

TABLE 4.3: The variance explained by the first 10 EOFs of zonal wind (UWN), meridional wind (VWN), precipitation (PTN) and humidity (HUM). The values listed are obtained using combined PCA of each variable with PLASIM SAT.

	UWN	VWN	PTN	HUM
1	29.47	24.86	27.03	88.76
2	20.46	25.91	14.33	5.48
3	11.65	9.49	9.73	2.05
4	6.07	4.77	7.38	0.60
5	4.96	3.85	5.66	0.80
6	3.52	3.23	3.42	0.34
7	2.19	2.92	3.49	0.32
8	1.88	2.49	2.54	0.08
9	1.38	1.49	1.92	0.15
10	1.09	1.11	0.98	0.01
Total	82.66	80.14	79.44	98.68

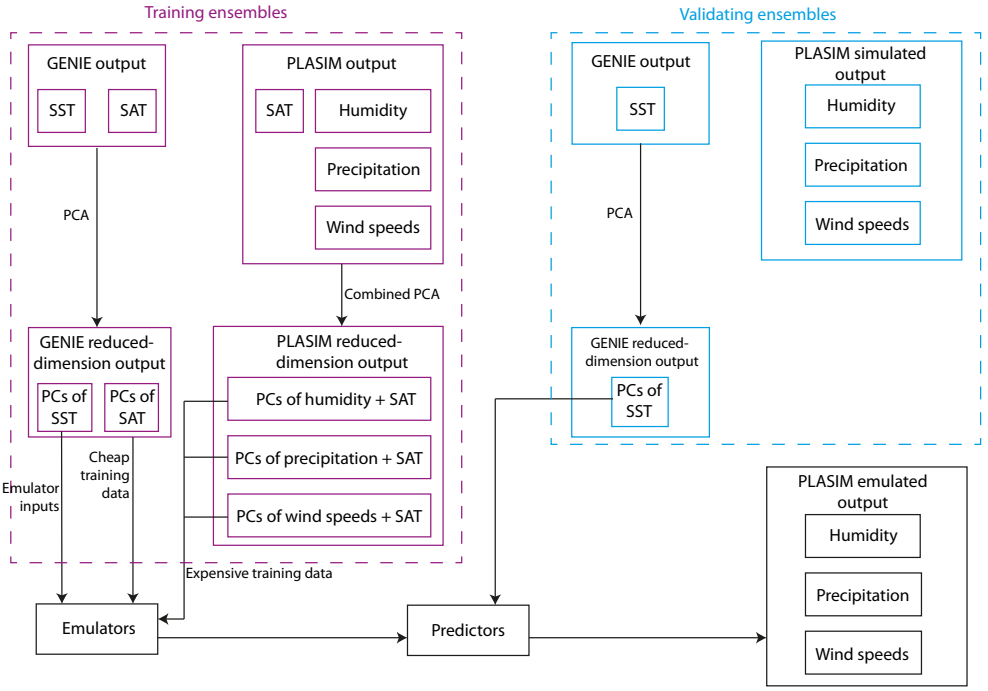


FIGURE 4.4: The training (purple) and validating (blue) ensembles are shown here with their respective sets of output. The dimensionally-reduced fields from GENIE and PLASIM training ensembles are used to construct the emulators. The emulator's predictor then takes the PCs of validating SST and produces the emulated fields. The emulators are validated by comparing the PLASIM?s simulated and emulated outputs.

Figure 4.4 summarises the dependencies between the different models and emulators. The simulation outputs from GENIE and PLASIM are decomposed using PCA (or combined PCA) and used to train the emulators. The GENIE SST from the validating ensemble is dimensionally reduced and then provided to the emulator's predictors. The

emulated and simulated PLASIM variables are compared in Section 4.4. It's worth noting that only subsets of the 600-member training ensembles are used while for validation, the whole 214 simulations are used.

We now analyse the model outputs and the EOFs of each field. Figure 4.5 shows the mean SAT, surface wind speed, specific humidity and precipitation rate averaged over the 600-member ensemble. The mean SAT fields display a typical pattern with high temperature over the tropics, especially in the Sahel and the tropical South America regions. The coldest temperatures are observed at high latitudes (the Arctic and Antarctica) and high elevation (Tibetan Plateau and North America ice sheet) areas.

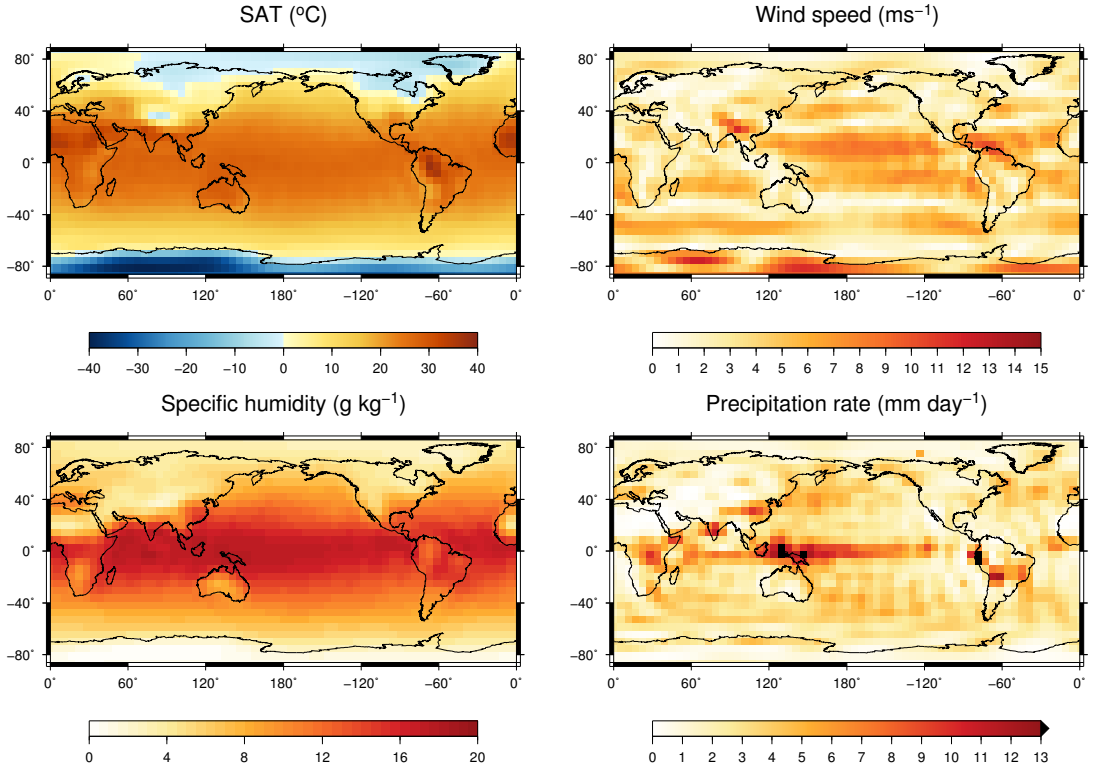


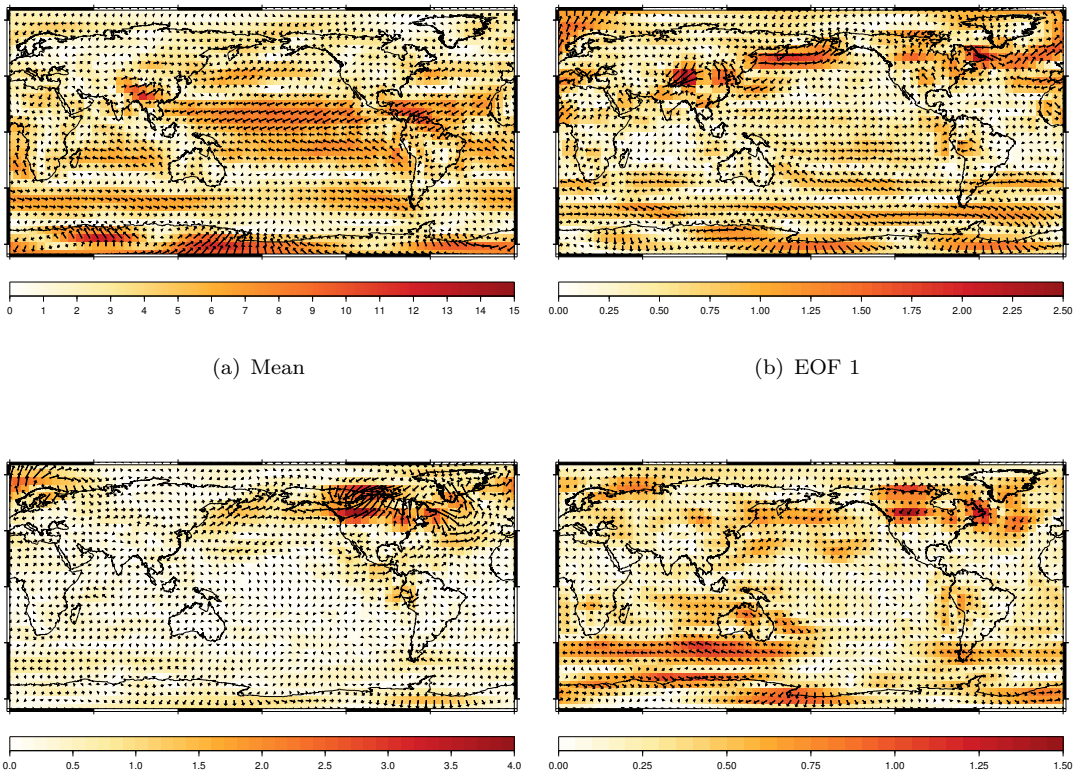
FIGURE 4.5: Ensemble mean surface air temperature, surface wind speed, specific humidity and precipitation rate. All 600 PLASIM simulations are used.

In general, the mean surface wind speed shows a weak resemblance to the modern day wind speed map. Surface westerlies are observed between 30°–60° in both hemispheres while north easterlies and south easterlies are seen around the equator in the Northern and Southern Hemisphere, respectively. The north easterlies appear to be stronger than the Southern Hemisphere westerlies in this ensemble mean. The strongest surface wind speed is seen over Antarctica.

The mean humidity field shows the typical pattern with highest values along the equator, especially over the Western Pacific and decreasing poleward and over mountain areas. For precipitation, the rain band associated with the Intertropical Convergence Zone (ITCZ) is visible with the largest amount of rain observed on the west side of the

Pacific. The ITCZ appears to be significantly weaker in the Atlantic than in the Pacific The location of the ITCZ appears to be centred on the equator. The mid-latitude storm tracks are weak but can be seen in the Pacific. The South Pacific Convergence Zone is not visible here. The dry zones associated with the downwelling branch of the Hadley cells receive very little rainfall, especially in Northern Africa.

4.3.2 Surface wind velocity



(a) Mean

(b) EOF 1

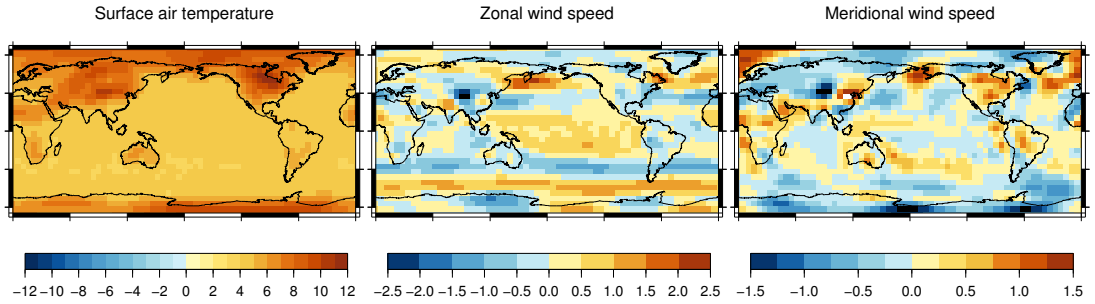
(c) EOF 2

(d) EOF 3

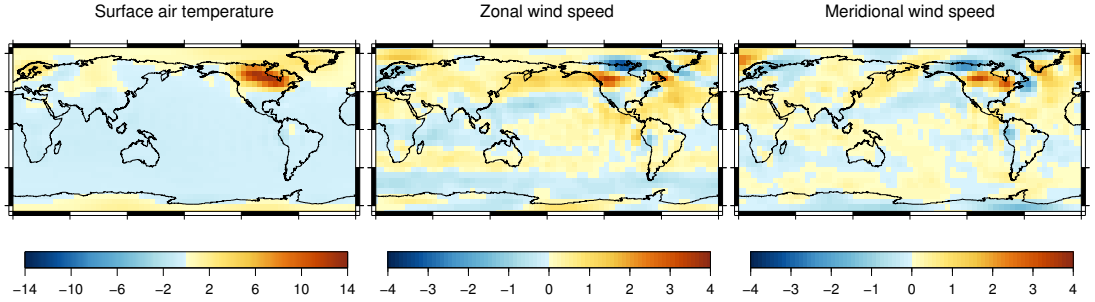
FIGURE 4.6: The surface wind patterns of (a) the mean surface wind velocity field across the ensemble; (b), (c) and (d) the surface wind anomaly corresponding to the first 3 EOFs. These fields are computed from the zonal and meridional wind shown in Figure 4.7. The wind vectors in the 3 EOF plots share the same scaling while the mean wind vectors are scaled down by a factor of 6.

While each wind speed component will be decomposed separately with SAT for the construction of their emulators, both zonal and meridional components are decomposed together with SAT here. This allows a direct comparison between the modes of the two components as well as the construction of the surface wind anomaly vectors from them.

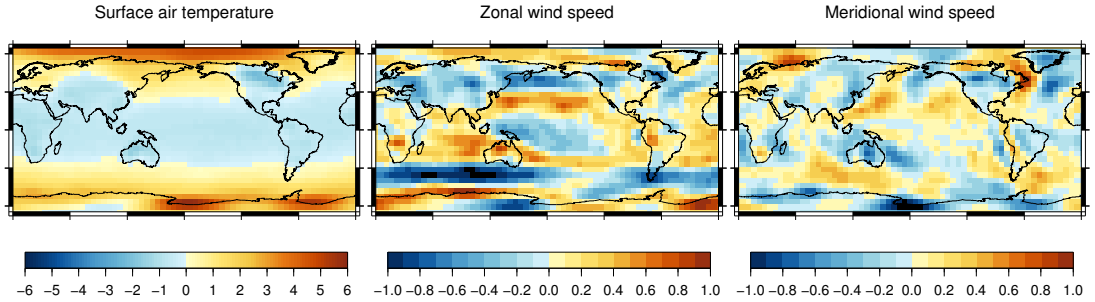
Figure 4.7 shows the first three modes of SAT, zonal and meridional wind components averaged over the 12 month. The wind velocity anomalies corresponding to each mode



(a) EOF 1



(b) EOF 2



(c) EOF 3

FIGURE 4.7: The first three EOFs of SAT, zonal and meridional wind components obtained from a combined correlation PCA decomposition using outputs from the whole 600-member ensemble.

are computed and shown in Figure 4.6. The EOFs of SAT here are similar but not identical to the modes obtained when SAT is decomposed alone (Section 3.4.3).

In the first modes of SAT and wind velocity components anomalies, a global cooling/warming trend is observed in SAT with largest changes over the high latitudes and high altitudes. Moderate changes are found in the tropical regions and over the ocean. Correspondingly, we see an overall increase in surface wind velocity in this mode. Stronger trade winds in both hemispheres are seen in the zonal and meridional wind components, indicating more active Hadley cells. The Westerlies are also strengthened and appear to shift equatorward. There are also significant increases in wind velocity

over the Arctic and Antarctica in this EOF. Universal kriging coefficients show that this mode is predominantly driven by the first EOF of SST, which, in turn, is controlled by RFC, OL0, RMX, OL1 and ICF as discussed in Section 4.2 and Figure 4.1.

Both the second and the third modes are driven mainly by the second EOF of SST, which reflect on the SAT patterns. The second mode also has an extra contribution from ICF, a parameter that has a stronger effect on atmospheric quantities than on oceanic ones. The anomalies due to these modes are weaker in general compared to the first mode. Changes in velocities at low latitudes are relatively small. However, significant local changes are seen due to the presence of the Laurentide and the Fennoscandian ice sheets. Strong winds can be seen flowing downslope of the Laurentide. The Southern Hemisphere westerlies are slightly weaker while in the Northern Hemisphere, they are disrupted by the continental ice sheets.

4.3.3 Humidity and precipitation

The behaviour seen in specific humidity (Figure 4.8) is relatively straightforward; humidity increases as surface temperature rises and vice versa. The humidity in the tropics appears to be more sensitive to a change in temperature. These are the results of the Clausius-Clapeyron relation. As the temperature decreases, the atmosphere can hold less water vapour and hence has a lower humidity. This relationship is non-linear and a warm atmosphere can hold a much higher moisture content than a cold atmosphere. A small change in temperature in the tropics leads to a larger change in humidity than a similar change at high latitudes. This is evident in all three modes. While they have very distinct spatial patterns compared to SAT, their responses to the changes in boundary forcing conditions, specified by the PCs are the same.

The precipitation pattern in the first EOF (Figure 4.9) is characterised by a large-scale drying resulting from the reduced evaporation associated with the global cooling. The largest drying is seen over Antarctica. The desert regions of both hemispheres experience an increase in precipitation. This is likely to be a result of the weaker Hadley cell associated with this mode of SAT anomaly. We've observed stronger zonal and meridional components of the easterlies, which is an indicator of a stronger Hadley cell as SAT increases (Figure 4.7 and 4.6). A weakened Hadley cell would lead to a wetter dry band and less precipitation over the ITCZ.

For both the second and third modes, we tend to see more rain in areas with a positive anomaly in SAT. Regional responses, however, do diverge from this. The presence of the continental ice sheet over North America and Greenland induces more complicated patterns over these areas as seen in the second mode. In the third mode, Australia and South Africa appear to become wetter when SAT decreases. In this EOF, the zonal

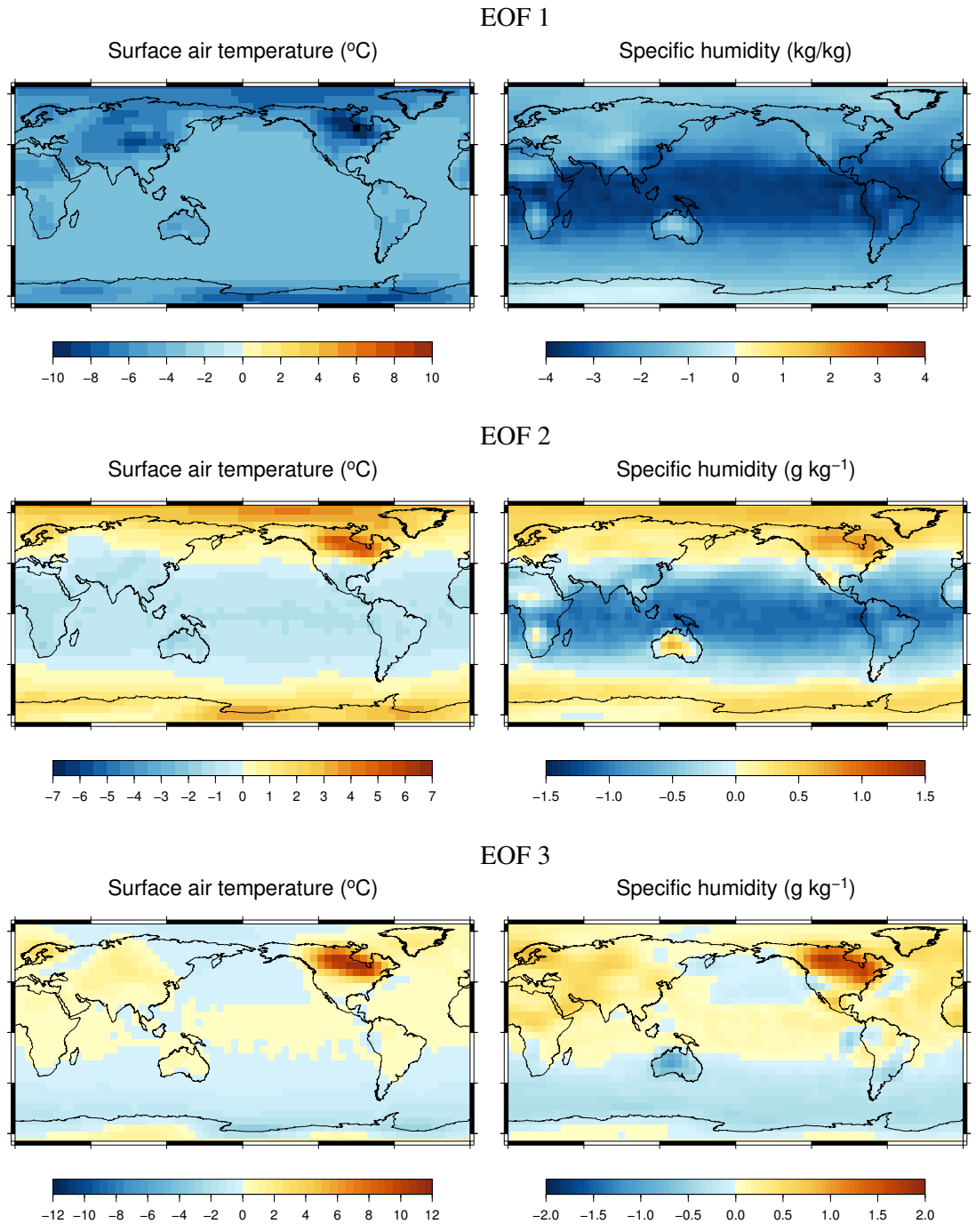


FIGURE 4.8: The first three EOFs of SAT and specific humidity obtained from a combined correlation PCA decomposition using outputs from the whole 600-member ensemble.

wind anomaly appears to bring more moisture from the ocean into the South African and Australian continents.

The seasonal variations of each set of EOFs are included in Appendix B.

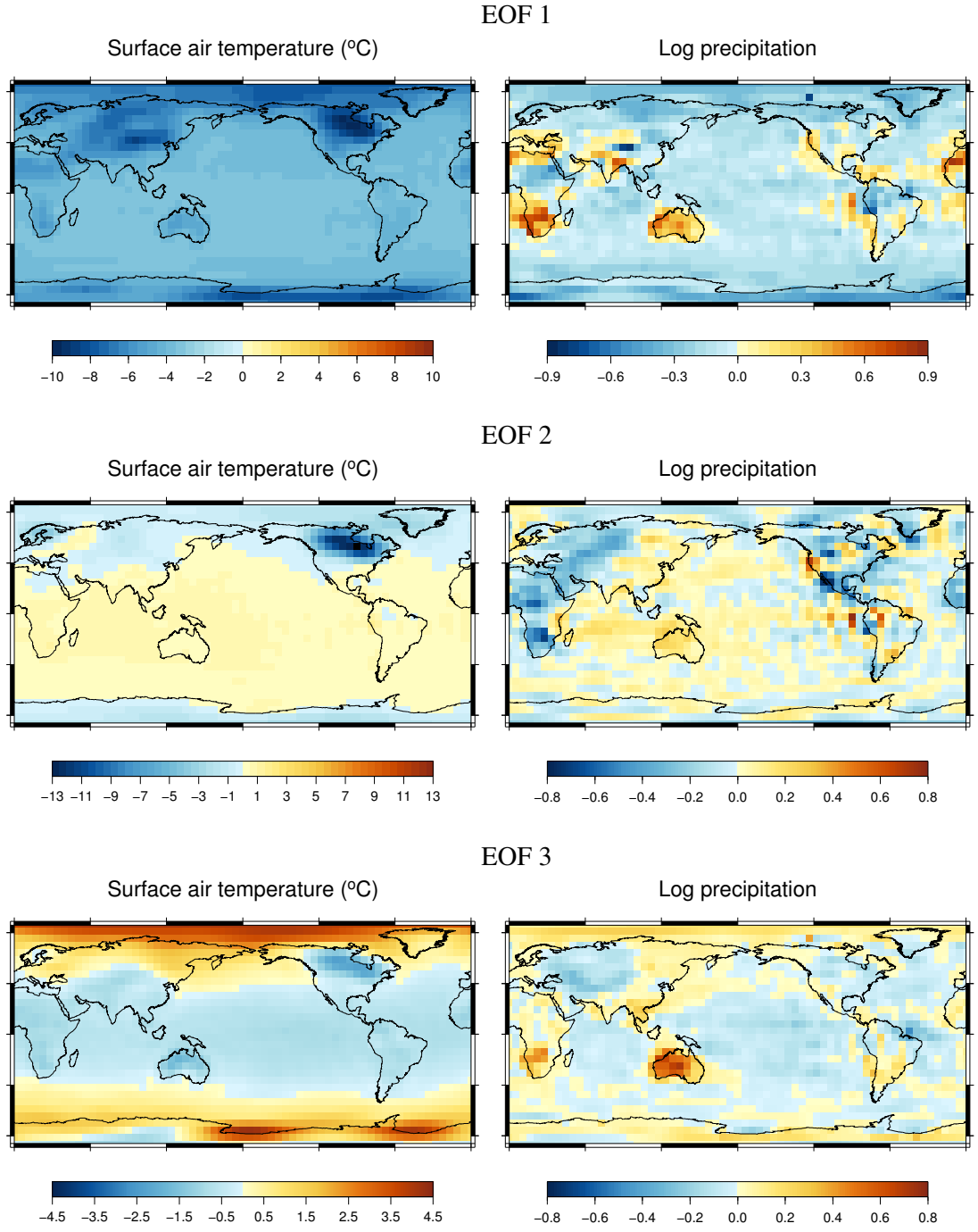


FIGURE 4.9: The first three EOFs of SAT and log-precipitation obtained from a combined correlation PCA decomposition using outputs from the whole 600-member ensemble.

4.4 Emulator validation

For each variable, the state vector is constructed using 90 expensive points and decomposed into pairs of 'expensive' EOFs and PCs. The expensive PCs are emulated using co-kriging. The 'cheap' PCs are obtained by projecting EMBM SAT state vectors

(constructed from 200 cheap points) onto the 'expensive' EOFs and used as the fast approximation of the expensive PCs. This is done for the first ten modes of each variable. Each mode is emulated, and added sequentially and the emulated surfaces are validated against the simulated ones using the validation ensemble. We attempt to evaluate how well the emulators do at predicting the ensemble behaviour and how well each simulation is reproduced.

Two measures are produced to evaluate the emulator performance across the training ensemble, the average normalised RMSE and the percentage of total variance explained, V_T. For each surface field, the normalised RMSE, $\hat{\epsilon}$ is calculated as

$$\hat{\epsilon}_j = \frac{\epsilon_j}{U_j^{max} - U_j^{min}} \times 100, \tag{4.5}$$

where ϵ_j is the RMSE between the j^{th} emulated and simulated field and U_j^{max} and U_j^{min} are the maximum and minimum values of that field. This quantity gives the error as a percentage of the spatial range in the corresponding field. The error plotted in Figure 4.10 is the normalised RMSE averaged over all 214 validation pairs. The normalised RMSE and the variance explained by the ten simulated EOFs are also shown (dashed line) in this figure. This is the variance that would be achieved by a perfect co-kriging emulator, limited by dimensional reduction of outputs into EOFs and PCs. Thus, the imperfections demonstrated by these lines indicate the truncation error. The departure from this line by the emulator's variance is a result of the component error.

As more modes are added, emulator performances improve. In all cases, the first three modes are emulated most successfully and captured over 50% of the total variance while the rest improves the emulator by small but not negligible amounts. The kriging emulator results for the same quantities are also included in the plot. These emulators use only 90 expensive training points for each PLASIM's variable. They also use covariance PCA as opposed to combined correlation PCA as for co-kriging.

For the wind speed emulators, 73.66% and 69.62% of the total variance for zonal and meridional components, respectively, are captured using co-kriging. They are approximately 10% lower than the total variance captured by the ten simulated modes but are 9.16% (zonal) and 7.64% (meridional) higher than the kriging emulator results. The final emulated fields for both components have an average normalised RMSE of 2%. For both components, the co-kriging emulators perform comparably or slightly underperform compared to kriging in the first two modes. The kriging emulators outperform even the perfect co-kriging case here. This is because the first two modes obtained using ordinary PCA of wind speed explain more of the total variance than the combined PCA do. The fact that the co-kriging emulators ultimately outperform kriging ones demonstrate the value of the added information from EMBM.

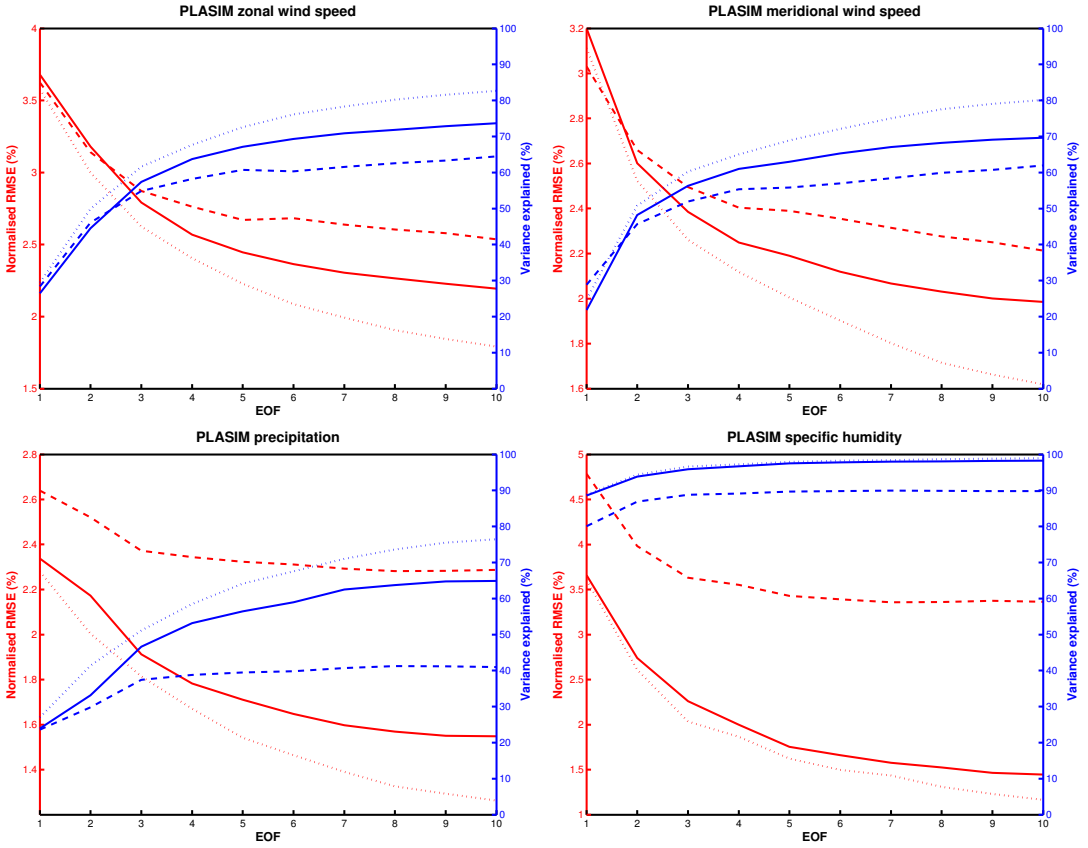


FIGURE 4.10: The normalised RMSE and variance explained by the first 10 emulated EOFs for zonal wind speed, meridional wind speed, specific humidity and precipitation. Co-kriging (solid) is compared against kriging (dashed). The RMSE/variance that would be achieved by a perfect emulator are also included for each plot (dotted line). Each co-kriging emulator uses 200 cheap and 90 expensive training points while a kriging emulator uses 90 expensive training points.

The precipitation emulator performs less well, capturing 64.88% of the total variance while the ten simulated modes capture 76.49%. Compared to the 40.97% achieved by the kriging emulator, the co-kriging step clearly adds more useful information than it takes away. The average normalised RMSE for precipitation is 1.55%. Precipitation tends to exhibit more internal variation than other variables, resulting in a smaller fraction of the total variance being explained by the top EOFs. Among the four variables, precipitation is also least strongly correlated to SAT.

The humidity emulator performs best thanks to its similarity to SAT. 98.24% out of 98.94% of the ensemble variance is captured. The average normalised RMSE is 1.51%. Co-kriging, in this case, manages to capture 8.40% of the total variance that were not emulated by the kriging emulator. It is not surprising that the emulator for humidity performs well considering the high correlation between SAT and the specific humidity.

The expensive emulator's inputs are members of the cheap input set when co-kriging is used. This means that the PCs corresponding to the 90 expensive training points were

a subset of the PCs obtained from the decomposition of 200 SST fields. When kriging is used, however, only 90 SST fields are used in trying to capture the whole ensemble behaviour. As a result, the EOFs and PCs computed in the co-kriging case are likely to be more reliable and robust.

The results shown in Figure 4.10 are averaged over the 12 calendar months. A breakdown result of the variance explained for each month is included in Appendix C.

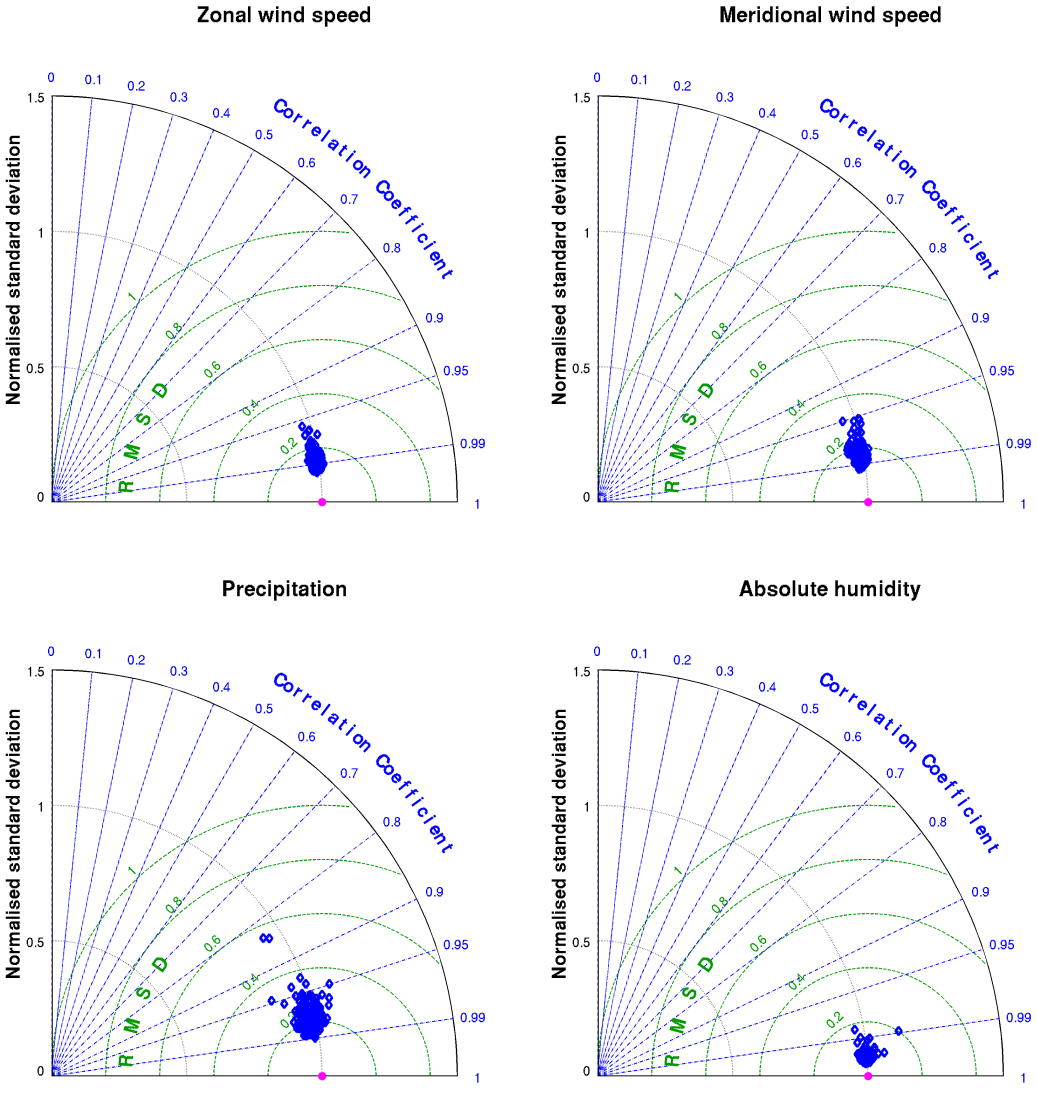


FIGURE 4.11: A comparison of the emulated (blue) and simulated (magenta) zonal and meridional wind components, precipitation and specific humidity. The simulated outputs are treated as observations and the emulated values are compared against them in pairs.

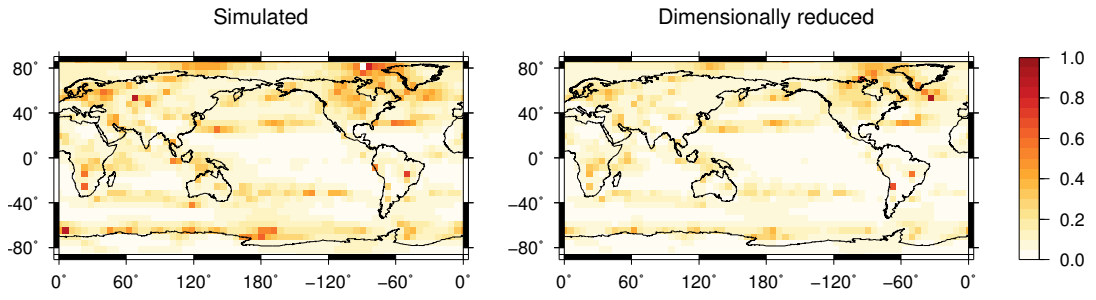
Taylor diagrams (Figure 4.11) are used to compare each 2-D surface field to its simulated one. The Taylor diagrams display, at the same time, the correlation coefficient, standard deviation and root mean square difference (RMSD) of the emulated field with

respect to their corresponding simulated field. Both standard deviation and RMSD are normalised by the simulated field's standard deviation. All four emulators can reproduce well the surface pattern, as demonstrated by the high average correlation. For all variables, except precipitation, the minimum correlation is approximately 95% or higher. For precipitation, while the vast majority of the emulated fields are well correlated to the simulated ones, there are two outliers with correlation below 85%. The standard deviations of the emulated ensembles are centred around 0.99 of the original field'ss value, indicating that the emulator very slightly underestimates the spatial variation in general. The average RMSD appears to be around 15-20% of the original field's standard deviation, except for humidity, where this value is much lower, around 5%. The emulated ensembles tend to cluster together on the diagrams indicating that the performance across the ensemble is mostly consistent. The two outlying points in the precipitation diagram are associated with very large scores of SST PC1, which depends on RFC, OL0, RMX and OL1. A quick verification shows that the high RMSDs correspond to very large values of RFC and OL0 at the same time. While the PC scores for these two points are still within the training range, it is important to note that certain combinations of parameters of the validation simulations can lead to PC scores beyond the training range. This is indeed observed within this validation ensemble. However, the performance at those points appears to be reasonable.

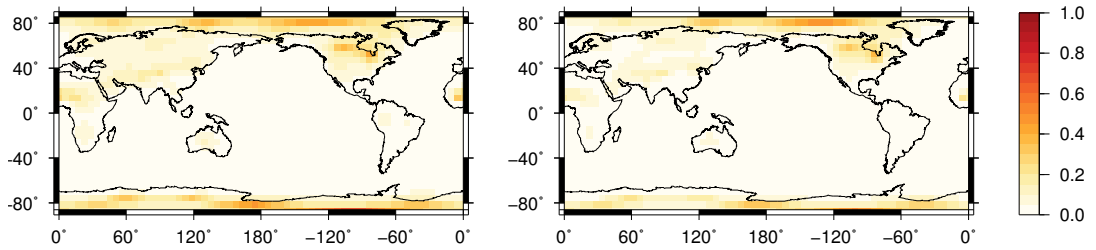
The emulators appear to be doing a good job at reproducing PLASIM's simulated fields in general. Now we want to examine the distribution of errors to identify weak areas in the emulator. The component errors are compared against the original and dimensionally reduced fields by evaluating the error at each grid cell as a fraction of the values at that cell with

$$\hat{P}_j = \frac{|\hat{U}_j - U_j|}{U_j}, \tag{4.6}$$

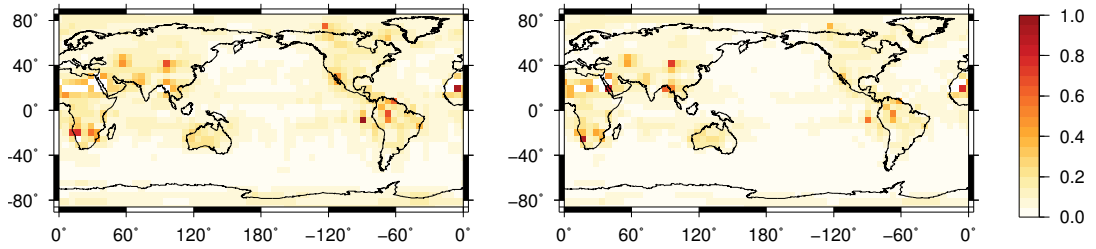
where $\hat{U}_j$ is the emulated value at the j^{th} grid cell, and U_j is either the simulated or the dimensionally reduced value (the simulated field described by the first ten EOFs only) in the same grid cell. By computing this error using the dimensionally reduced fields, we only look at the error introduced by the emulation process. Using the simulated values gives the total error, a combination of component error and the truncation error. Figure 4.12 display the geographical distributions of the total error (left) and component error (right). The errors are calculated for surface wind speed (top), humidity (middle) and log-precipitation rate (bottom). The white colour, which indicates small errors, dominates all three plots showing good agreement between the emulated and simulated values. The similarity between the total errors and the component errors suggests that component errors dominate, and truncation errors are less significant. In general, large fractional errors are associated with low values. This is clearest in precipitation, where fractional errors with a magnitude over 1 are seen in the Sahel, where very little rain is observed. The humidity emulator is valid everywhere. The errors are large over the Laurentide but they are still under 50%, and this area is expected to have low humidity.



(a) Wind speed



(b) Specific humidity



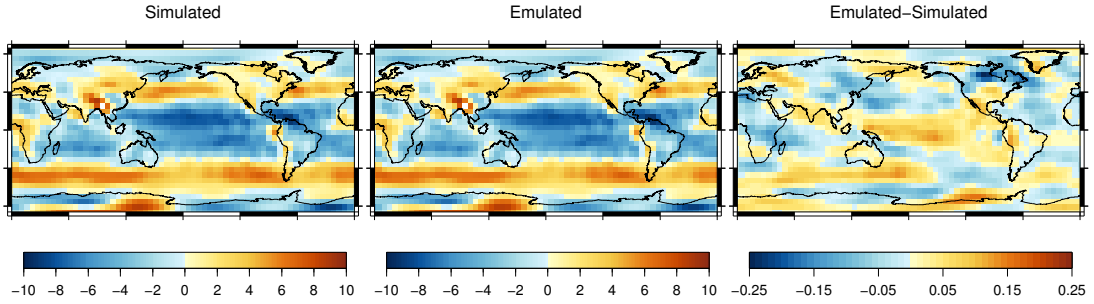
(c) Precipitation rate

FIGURE 4.12: The distribution of fractional error averaged over the 214-member validation ensemble for wind speed, humidity and precipitation. The total error is shown on the left while the emulator error is shown on the right.

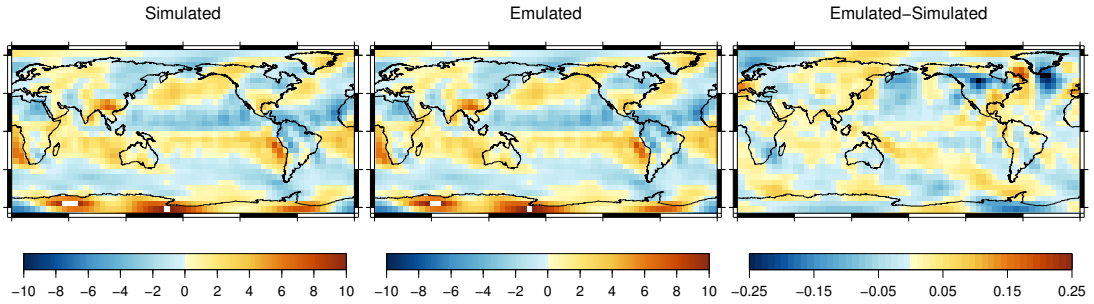
For wind speed, the area with the largest error is also in a location with very low winds to the west of Greenland.

The total error appears to be smaller than the component error over the Laurentide for humidity. The variance in this area, as described by the second and third EOFs, are very large. It is possible that one or more lower order EOFs have the opposite effect here. By truncating these modes, the dimensional reduction systematically overestimates the variation in this area. The lower total error could be the result of the tendency of the emulator to underestimate the spatial variation, which balances out the systematic error here.

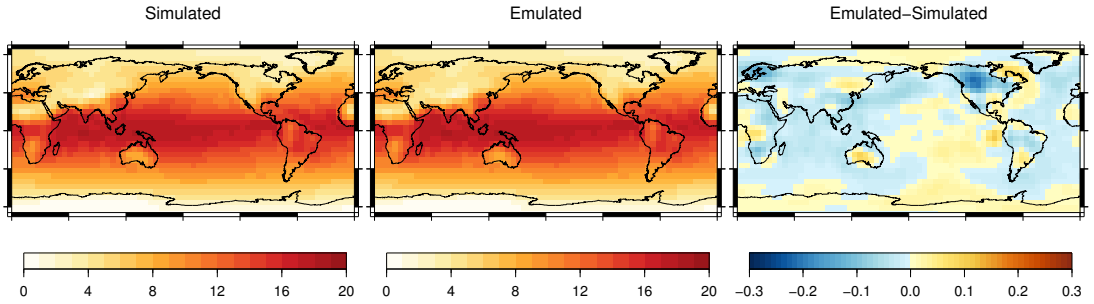
Figure 4.13 summarises the emulation result by comparing the simulated and emulated mean fields for all four variables. The emulated minus simulated difference is also shown. Since this is the difference in the mean fields across the emulated and simulated ensemble,



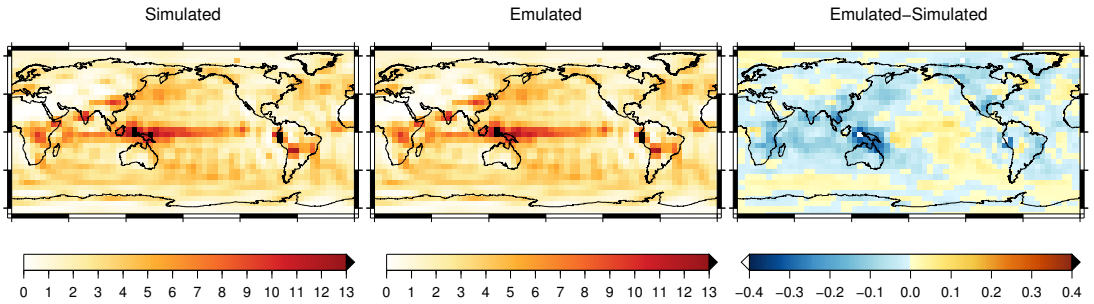
(a) Zonal wind component (m s^{-1})



(b) Meridional wind component (m s^{-1})



(c) Specific humidity (g kg^{-1})



(d) Precipitation rate (mm day^{-1})

FIGURE 4.13: Comparison between the simulated and emulated ensemble mean of the zonal wind speed, meridional wind speed, humidity and precipitation. The difference between the two are shown on the right.

it highlights the systematic error that prevails in all simulations. The emulated and

simulated fields are very similar in all four cases. Areas of large errors are seen for the easterlies and Southern Hemisphere westerlies and to the region south-east of Greenland. Precipitation over the maritime continent in the Pacific appears to be underestimated. The largest differences in humidity are seen in the southern hemisphere tropical Pacific. Overall, the magnitude of these differences is small compared to the actual values.

4.5 Summary and conclusions

We have successfully combined several statistical techniques to construct emulators relating four high-dimensional PLASIM's atmospheric variables to the high-dimensional boundary forcing fields provided by GENIE-1's ocean and sea-ice.

By emulating PLASIM output as a function of the climate forcing field, estimation of PLASIM's climate response to GENIE-1's SST at any timestep can be obtained. High-dimensional fields of exchange fluxes between the atmosphere and the ocean can be produced efficiently and replace those currently provided by the fixed climatology in EMBM. A setup which consists of GENIE-1's subcomponents and PLASIM emulators would be computationally cheaper to run than a fully coupled GENIE-PLASIM model. The new improved fields provided by PLASIM emulators allows the changes in atmospheric circulation to feedback onto the ocean. A more realistic wind and precipitation field can significantly affect the simulated climate as well as the ocean circulation in GENIE.

While demonstrated on this specific example, the technique is certainly applicable to a wide range of studies involving high-dimensional input and forcing fields. This technique has the advantage that while the number of model parameters increases, it's likely that the number of PCs required would remain relatively low. It is also independent of the spatial resolution of the high-dimensional input fields. This step can be seen as a screening step, in which redundant information from 'inactive' inputs is removed. A 'nugget' can be introduced to the emulators to account for the variability associated with these inputs (Boukouvalas and Cornford, 2009). While not observed in this work, Boukouvalas and Cornford (2009) also pointed out that the use of a linear projection method such as PCA can lead to negative values in a strictly non-negative quantity such as humidity or precipitation.

By utilising combined correlation PCA, the common structures of variability between each variable and SAT are identified and emulated. This step generalises our multi-level emulation framework, allowing us to use the SAT from EMBM to help improve the emulators of other variables in PLASIM. Information does not only travel between two levels of model fidelity but also between different variables. The emulators constructed are capable of reproducing all four variables, most successfully for humidity and less so for precipitation. Areas, where systematic errors occur as a result of the linear

decomposition and the fact that emulators tend to underestimate ensemble variations, are identified.

Chapter 5

GENIE-1 with interactive statistical wind fields

5.1 Introduction

We have, by this point, established the statistical methods used to emulate PLASIM's atmospheric variables efficiently. The emulated fields produced in Chapter 4 are now ready to be incorporated into the GENIE framework. Ultimately, we would like to utilise all five emulated variables, the surface air temperature, surface specific humidity, precipitation rate and the two surface wind velocity components. However, this is a substantial task which could lead to drastic changes in GENIE-1's behaviour. It is, therefore, broken down into three smaller steps: i) replace the prescribed winds in EMBM with PLASIM's emulated winds; ii) altering the moisture balance in EMBM using emulated precipitation and humidity fields; iii) directly replacing SAT field as well as other radiation terms in the atmosphere. The decision to separate this task into smaller steps stems from the desire to assess the impact of each additional source of complexity. Therefore, after each component is added, the model's outputs should be analysed and compared to the original model. In this thesis, we only concern ourselves with the wind fields, namely the zonal and meridional components of wind velocity and wind stress.

As discussed in Chapter 1, the lack of an interactive wind field and the inadequate representation of precipitation are among the main limitations of the current simple atmosphere. It is, therefore, most beneficial and natural to first consider the winds. Wind speed and wind stress play a major role in the climate system by influencing the global heat budget and ocean circulation. In the atmosphere, the winds contribute to the heat budget by transporting heat (sensible heat) and moisture (latent heat). The surface wind stress field also drives the ocean circulation through its input of momentum, its effect influences the meridional salt transport and the upwelling of deep water (Chen

et al., 1994). Studies have shown that they might play a decisive role in determining the regime of the AMOC through various mechanisms (Toggweiler and Samuels, 1995; Timmermann and Goosse, 2004), especially under glacial conditions (Arzel et al., 2010; Oka et al., 2012). Furthermore, the presence of the Laurentide ice sheet is believed to influenced the climate on a continental scale by significantly altering the structure of the atmospheric circulation (Justino et al., 2006; Li and Battisti, 2008; Langen and Vinther, 2009; Pausata et al., 2011; Hofer et al., 2012). The resulting anomalies induced by stationary waves forced by the ice sheets can be significant in not only the vicinity surrounding the ice sheets but also on a continental scale (Roe and Lindzen, 2001). This leads to deviations in the zonal mean climate at mid to high latitudes in the Northern Hemisphere. The wind stresses felt by the ocean are also altered, triggering further changes through feedbacks involving the AMOC state. The surface wind can influence the deep ocean circulation through its effect on the formation of deep water in the North Atlantic via its effect on the salt transport in the high latitude North Atlantic (Oka et al., 2001), possibly amplified by a shift in the sea-ice edge (Yang et al., 2016) or through the wind stress driven upwelling in the Southern Ocean (Toggweiler and Samuels, 1995; Kuhlbrodt et al., 2007).

Compared to the current fixed modern wind fields used in EMBM, the simulated winds from PLASIM are obtained from a simplified AGCM with a fully dynamical atmosphere. This wind field can interact with glacial orography boundary conditions prescribed. The atmosphere of PLASIM and its predecessor, PUMA, have been used to examine the atmospheric response to an ice sheet on an idealised North American continent (Liakka et al., 2012), the sensitivity of North Atlantic atmospheric temperature and oceanic circulation as a result of the LGM orography (Romanova et al., 2006) and the impact of global ice sheets of various heights on the global circulation (Schmittner et al., 2011).

The emulated winds encompass the effects induced by changing continental ice sheets as well as well as the large CO_2-equivalent radiation forcing. Chapter 4 presented emulators of the wind components as functions of the SST boundary forcing, allowing the emulated winds to be coupled with GENIE-1's ocean module interactively. The model setup shares some similar aspect to the regional hybrid coupled model presented in Barnett et al. (1993) where the assumption made is that near-surface wind speed and wind stress are determined by SST, and the atmosphere responds instantaneously to changes in the boundary condition described by the ocean. The statistical aspect of this model provides a link between the SST from an OGCM to the wind stress field based on the relationship obtained from observational SST and wind stress data. The two observational datasets are decomposed into separate EOFs whose PCs can be linked through a regression, which is then used to provide a prediction of the wind stress response to a new SST field. Other methods to provide the link between such pairs of spatial fields exist, such as the canonical correlation analysis of Barnett and Preisendorfer (1987). Another regional model which employed singular value decomposition of observational data to

supply wind stress anomaly is seen in Syu et al. (1995). Cimatoribus et al. (2012) presented a global hybrid model that captures the basic representation of feedbacks due to the ocean-atmosphere interaction derived from a local deviation of SST through the use of linear regressions.

In Section 5.2, we demonstrate how the emulated wind fields can be incorporated into the GENIE framework. The emulated fields are generated using the constructed emulators and supplied to EMBM and GOLDSTEIN at every ocean time step. The emulator receives from the ocean the monthly mean SST and returns vector components of the surface wind velocity and wind stress. The new model setup can be effectively viewed as an EMBM with an interactive statistical wind field. This new model configuration is then used to produce equivalent simulations of the ensembles described in Section 3.3.1. Section 5.3 compares the resulting interactive statistical wind fields to the fixed climatological wind fields previously used. The following sections show some general assessment of the model outputs to study the effects of the newly introduced wind field. This section provides a summary of the general trends of the new ensemble compared to the old ones without going into the effects of the interactive wind field on any specific region of the parameter space. The final section closes this chapter with a discussion on the performance of the new GENIE configuration and its implication for future work.

5.2 Incorporating PLASIM's emulated wind field

5.2.1 The current wind field in EMBM

Within the GENIE framework, the EMBM atmosphere requires prescribed wind velocity components, surface wind speed and surface wind stress components. Depending on the modules activated in a simulation, different combinations of winds have been used in the past. For example, when the baseline setup, which consists of the EMBM atmosphere and the GOLDSTEIN ocean and sea ice modules, was used in conjunction with the biogeochemical components BIOGEM and ATCHEM, the wind stress climatology from the Southampton Oceanography Centre (SOC) by Josey et al. (1998) was interpolated onto the model grid while zonal and meridional wind speed fields were interpolated from the annual mean NCEP-NCAR 40 year reanalysis project (Kalnay et al., 1996). Recently, the Josey et al. (1998) wind stresses have been replaced with a newer version obtained from the NCEP-DOE Reanalysis 2 (Kanamitsu et al., 2002). The wind speeds are also updated accordingly. However, when the ENTS module is used instead of BIOGEM and ATCHEM, the atmosphere calls surface wind speeds from a different source. In this case, which is also the setup used in this work, the two wind velocity components and the surface wind speed are the monthly fields derived from NCEP-NCAR reanalysis as described by Weaver et al. (2001). The wind stresses used are the

same as the annually average 1000 mb wind stresses from Kanamitsu et al. (2002). This situation is summarised in Table 5.1.

TABLE 5.1: Three different configurations of GENIE-1 and their corresponding wind climatologies. The SOC climatology is from Josey et al. (1998) while the NCEP-NCAR and NCEP-DOE reanalyses are from Kalnay et al. (1996) and Kanamitsu et al. (2002), respectively.

Cofiguration	Atmospheric winds	Ocean wind stress
BIOGEM (old)	Monthly mean from NCEP-NCAR reanalysis	Annual mean from SOC climatology
BIOGEM (new)	Monthly mean from NCEP-DOE reanalysis 2	Monthly mean from NCEP-DOE reanalysis 2
ENTS	Monthly mean from NCEP-NCAR reanalysis	Annual mean from NCEP-DOE reanalysis 2

In summary, there is a discrepancy between the winds seen by the atmosphere and the ocean in our current setup. Furthermore, these interpolated winds fields are generated prior to the model runs and remain fixed throughout a simulation, independent of the glacial boundary conditions or other perturbed parameters. The use of mean annual wind stresses also reduces the seasonality in the ocean. Apart from the scientifically driven motivation described earlier, by replacing these fixed wind fields with consistent interactive statistical winds, we also remove this inconsistency, making the model more coherent.

5.2.2 The wind field emulator

Using 90 expensive and 200 cheap simulation runs, the emulators for zonal and meridional wind velocity components are constructed using a multi-level GP emulation technique (Chapter 4). The validation showed that the emulators are capable of capturing approximately 74% and 70% of the ensemble variance for the zonal and meridional wind components, respectively. The emulation process produced a set of hyperparameters, θ (Equation 2.6), describing the correlation function relating the emulator's outputs in each dimension. Using these hyperparameters, the 'predictor' of an emulator can translate the inputs (PCs from GOLDSTEIN SST) into the outputs (PCs of the combined SAT and wind field). The description of this predictor can be found in Bastos and O'Hagan (2009) and Forrester et al. (2008).

The emulators require 12 months of SST as inputs to predict the PLASIM wind components of the corresponding months. Thus, 12 monthly-averaged SST fields are calculated and continually updated as the model runs. The update frequency is the same as the flux exchanging frequency, which is at every timestep. Figure 5.1 shows this step as a

part of the coupling scheme between the GOLDSTEIN ocean module and the emulator's predictors in EMBM.

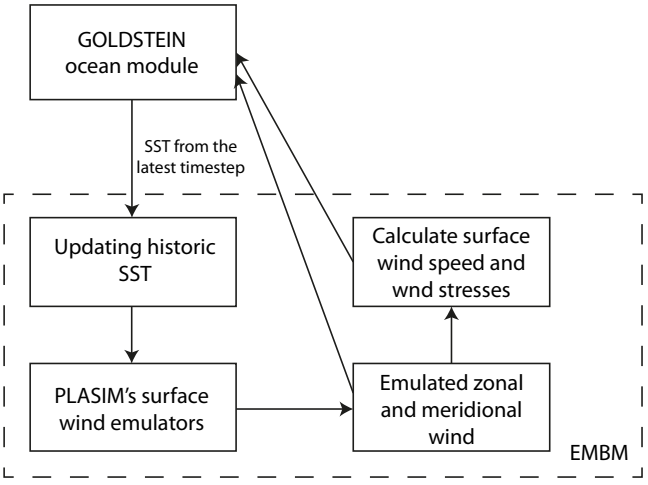


FIGURE 5.1: A schematic of the coupling between the GOLDSTEIN ocean module and the emulators (located in the SURFLUX routine in EMBM). The processes within the dashed box are all incorporated into the atmospheric module, EMBM.

The SST fields are saved as a 2-D vector $\mathbf{T}_{ij}$ with $i = 1, \ldots, 2048$ for the 2048 horizontal grid cells and $j = 1, \cdots, 12$ for the 12 calendar months. After each time step, the most recent SST field is used to update the values of $\mathbf{T}$ for that month. Therefore, it is not a snapshot of SST that drives the emulated wind field but the monthly average SST of the current and the previous 11 months. All the land grid cells are removed and the 12 months are then concatenated together to form the input field $\mathbf{W}$ with a dimension of 16332×1, similar to equation 4.1. As presented in section 4.2, the SST field is dimensionally reduced into a smaller set of variables (five variables) to use as the emulator's inputs. The PCs of $\mathbf{W}$ at each time step are obtained by projecting $\mathbf{W}$ onto the EOFs of the training ensemble's SST (three of which are displayed in Figure 4.1). This step is equivalent to obtaining the coordinates of the current SST state in the coordinate system defined by the EOFs of the training ensemble.

Apart from the five PCs of SST, the CO_2 concentration (RFC) and the ice sheet configuration (ICF) are also inputs of the emulators. Since we will be running steady-state simulations, the values of all these 'parameters' remain constant throughout each simulation. The emulators, however, are capable of coping with an evolving CO_2 concentration and possibly ice sheets also. This point requires further consideration if transient experiments are to be attempted. We will revisit this idea in Chapter 6.

Once given the five SST PCs, the predictors provide estimations of the first ten PCs of the combined fields (SAT + a wind component). The surface wind field is then reconstructed using the ten estimated PCs and the EOFs from the training ensemble.

The surface wind speed and wind stress components are calculated from these emulated winds. The wind stresses are also translated onto the edge-centred grid instead of the tracer grid. The surface wind components and surface wind speed are used in EMBM while the wind stresses are supplied to the GOLDSTEIN ocean module, influencing the SST at the next time step and complete the coupling. All the described procedures are contained within the EMBM module.

Since 12 months of SST are required for the emulators, the first model year uses the fixed climatology wind fields before the emulators replace them on the first timestep of the second simulation year. As the model runs, the PCs obtained from a timestep might exceed the range for which the emulators were designed. This is allowed to happen at the early stage of the simulation when there is a large amount of variability in the system. However, if the PCs are still outside the range by the end of the simulation, that particular run is deemed unsuccessful. Apart from that, simulation runs with certain combinations of model parameters that lead to numeric instabilities are also considered unsuccessful.

The same 660-member and 242-member maximin Latin hypercube sample plans in section 3.3.1 are used as designs for two new ensembles of 5000-year steady-state simulations using the new interactive winds. These new ensembles will be referred to as the interactive ensembles (INT) as opposed to the fixed ensembles (FIX) in previous chapters. The success rate for the INT ensemble (83%) is lower than that of the FIX ensemble (89%). In order to make a direct comparison between the two sets, only simulations that are successful in both configurations are used, leaving us with a 546-member and a 202-member ensemble. We now look at the differences between the INT and the FIX ensembles to assess the effect of the interactive statistical wind field.

5.3 The fixed and the interactive winds

5.3.1 Wind speed

In this section, we compare the INT and FIX wind velocity components and surface wind speeds. First, the fixed annual mean climatology wind field are shown in Figure 5.2. The zonal and meridional fields used for moisture and heat transport in the atmosphere were derived from the NCEP-NCAR 40 years reanalysis. Instead of using the surface wind components, the fields used are obtained by vertically integrating the wind velocity components, weighted by the moisture content at each level of the atmosphere (Weaver et al., 2001). The prescribed fields contain the monthly mean for 12 calendar months. At each model time step, these monthly winds are interpolated using the scheme by Killworth (1983). The two-highest-latitude grid rows are zonally-averaged to give smoother results in these regions (Edwards and Marsh, 2005).

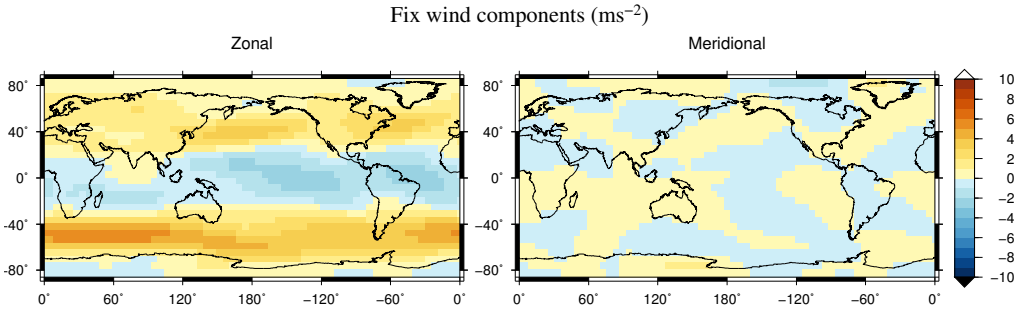


FIGURE 5.2: The fixed climatological zonal (left) and meridional (right) wind components derived from the NCEP-NCAR 40 years reanalysis (Kalnay et al., 1996). The fields are the same for all simulations in the FIX ensemble.

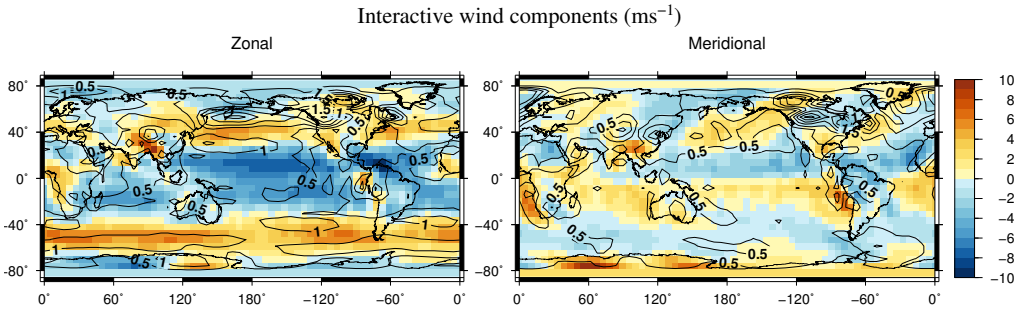


FIGURE 5.3: The annual zonal (left) and meridional (right) interactive wind components averaged over the entire 546-member ensemble. The ensemble standard deviations are denoted by the contour lines drawn at 0.5 m s^{-1} intervals.

Figure 5.3 show the ensemble mean of zonal and meridional interactive wind. The ensemble's standard deviations are shown as contours at 0.5 m s^{-1} intervals. The mean strength of the Northern Hemisphere and Southern Hemisphere westerlies are roughly the same while the easterlies appear to be the strongest zonal winds. Maximum zonal wind speeds occur in the tropical Atlantic and Pacific in the Northern Hemisphere. The meridional wind component shows strong Hadley circulations in the tropics and weaker circulation cells at mid and high latitudes. In both components, there are maximum and minimum wind speeds associated with mountain ranges and continental ice sheets. Large ensemble variations are seen over the Laurentide ice sheet while smaller but appreciable deviations from the ensemble mean are seen between the Greenland and the Fennoscandian ice sheets. The large standard deviations just north and south of the Pacific westerlies might be an indication of a shift in its position. There are little variations in the tropics for both components, indicating that the strong Easterlies is a persisting feature throughout the ensemble. In the Southern Hemisphere, there are moderate variations in the westerlies (zonal component) and over Antarctica (meridional component).

These are not equivalent to but share a similar structure with the winds in Figure 5.2.

This is because the INT winds are most strongly influenced by the surface layer where the majority of the atmosphere's moisture content is located. Similar to the FIX winds, these INT wind components represent the heat and moisture advection in the atmosphere that is not captured in the one layer diffusive EMBM atmosphere. While both components are used for moisture advection, only the zonal component is used to advect heat. For the purpose of representing moisture advection in the atmosphere, the use of the surface winds might not be a better choice compared to the vertically integrated winds. However, once we implemented the emulated humidity and precipitation fields from PLASIM, these wind components would no longer be required to transport moisture.

In general, the INT wind components are stronger than the FIX components, meaning that more heat and moisture is advected in an INT simulation compared to its FIX counterpart. They will affect both the heat and moisture distribution. In EMBM, while the wind patterns are prescribed by climatology, it is possible to change the amplitude of the wind by applying a scaling factor. Indeed, the scale of the zonal and the meridional winds are controlled separately by two model parameters. These parameters have been explored in a tuning exercise with GENIE-1 at a different horizontal resolution (Marsh et al., 2013). By changing these parameters, the difference in strength between the fixed and the interactive winds might be reconciled to a certain extent. These factors, however, are applied to the 2-D wind fields globally so anomalies due to the different wind structures will remain. This means that the resulting climate state of the INT ensemble is likely to be shifted from the FIX ensemble due to the different wind strength in addition to exhibiting different spatial features as a result of the different wind structures induced by the ice sheets.

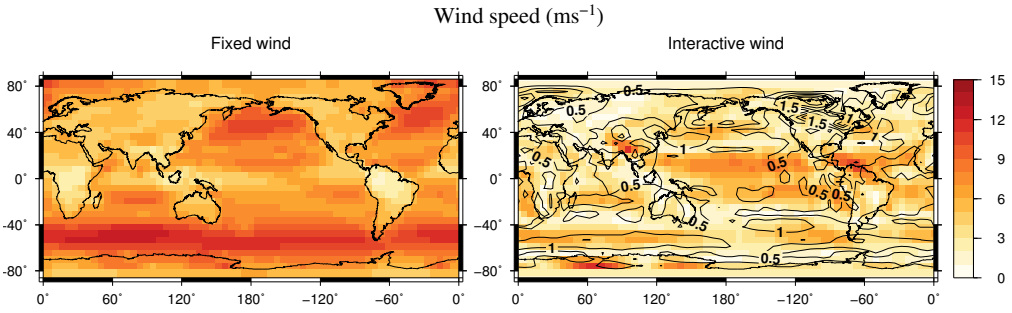


FIGURE 5.4: The annual mean fixed (left) and interactive (right) wind speeds. The interactive wind is averaged over the whole 546-member ensemble. The contour denotes the ensemble standard deviation with 0.5 m s^{-1} intervals.

Figure 5.4 shows FIX and INT surface wind speed used in calculations of the sensible and latent heat fluxes between the atmosphere and the ocean. It is likely that this field has a bigger impact on the model's climate compared to the wind components above. It is also used in the evaporation term thus affecting precipitation and humidity fields. Unlike the wind components, the surface wind speeds are of the same nature in both the FIX and the INT simulations. However, it is evident that they exhibit substantial

differences. Since the fixed wind field was taken from a reanalysis, the differences seen here highlight the inadequacy of the new emulated wind.

In Figure 5.4, the interactive wind is the ensemble average while the fixed wind is constant. The fixed wind appears to be much stronger in general and has main features resembling the zonal wind field. The most prominent features are the strong westerlies in both hemispheres. The interactive wind speed shows a very different pattern, with the strongest wind being the easterlies over North Pacific and North Atlantic. The interactive westerlies are too weak everywhere else. The Southern Hemisphere westerlies are only approximately half the strength of the fixed counterpart. The weaker wind speed can lead to a reduction in sensible heat exchange between the atmosphere and the ocean as well as lower evaporation, further influencing the water vapour content in the atmosphere and the salt transport in the ocean. For the interactive wind, the standard deviation across the ensemble is shown as contours at a 0.5 m s^{-2} interval. Similar to the wind components in Figure 5.3, there are small variations in the Southern Hemisphere westerlies and almost no variations in the tropics. Therefore the differences in these areas between the two wind fields are persistent throughout the ensemble.

5.3.2 Wind stress

The transfer of momentum between the atmosphere and ocean is parameterised as a stress applied to the ocean's surface. In this case, the wind stress is not directly emulated but is calculated using the bulk aerodynamic formula:

$$\boldsymbol{\tau}_r = \rho_a c_d |\mathbf{U}_s| \mathbf{U}_s. \tag{5.1}$$

where $\mathbf{U}_s$ is the surface wind velocity, $\rho_a = 1.25$ kg m^{-3} is the constant air density and $c_d = 1.3 \times 10^{-3}$ is a constant dimensionless drag coefficient. Here we use the resting approximation and assume that the ocean surface velocity is a lot weaker than $\mathbf{U}_s$.

The fixed and interactive wind stresses are shown side-by-side in Figure 5.5. Compared to the FIX stresses, the surface wind stresses calculated from emulated wind velocity are significantly weaker. The most noticeable difference is the weak midlatitude westerlies over the North Atlantic, North Pacific and the Southern Ocean. The meridional components have comparable strengths in general but have different features at high latitudes. Large differences in both components over the continents do not affect the ocean circulation and thus are not relevant here. The weaker westerlies, however, should have a notable impact on ocean circulation. The ensemble standard deviations are shown on the interactive component plots as contours at 0.01 N m^{-2} intervals. The standard deviations over the westerlies are small and cannot account for the discrepancies between the FIX and the INT wind stresses.

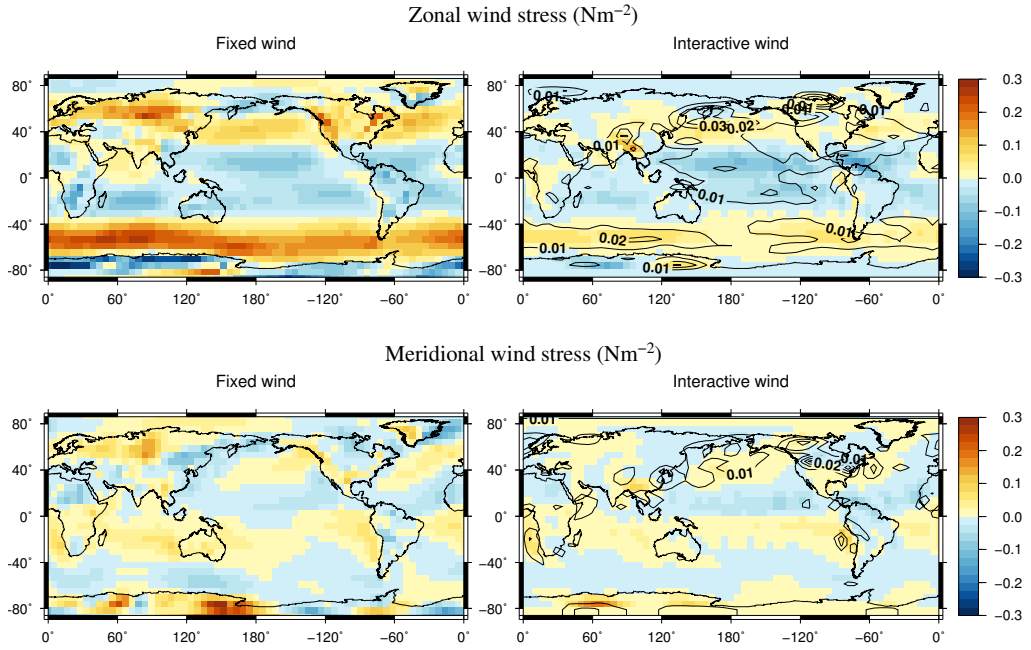


FIGURE 5.5: The fixed and interactive zonal and meridional wind stress components. The interactive components are the mean annual fields, averaged across the 546-member ensemble. The contours shown on the interactive component plots are the ensemble standard deviations. Contours are drawn ever 0.01 N m^{-2}.

Figure 5.6 shows the zonal-mean zonal wind stresses for INT (red) and FIX (black) simulations. The shaded area indicates the ensemble zonal mean standard deviation. While the easterlies are of similar strength, the westerlies in both hemispheres are severely underestimated, especially in the southern hemisphere. The interactive zonal wind stress over the Southern Hemisphere is less than half the strength of the climatological FIX stress. The location of the Northern Hemisphere westerlies is clearly shifted equatorwards in the INT ensemble mean. The large standard deviations in the Pacific from 40°N-60°N indicate more poleward Pacific westerlies in some simulations. Since the spread in the model ensemble cannot account for the discrepancy between FIX and INT wind stress, the most likely reasons for the differences are the emulators' error (unable to reproduce PLASIM's winds) or PLASIM's error (unable to reproduce realistic wind speed and wind stress). Furthermore, It's expected that the errors between GENIE's wind stresses and observation will be even larger compared to wind speeds considering how the stresses are calculated. This is due to the quadratic dependence of wind stress on the wind velocity. The zonal-mean zonal wind stress from PLASIM's simulations are also shown in Figure 5.6 for comparison. The strength of the PLASIM and INT zonal winds are very close to each other, suggesting that the poor comparison is most likely due to PLASIM's simulated wind.

Similar to the wind components, a scaling parameter is applied to the wind stress fields in order to obtain more realistic wind-driven gyres. This parameter scales up the wind

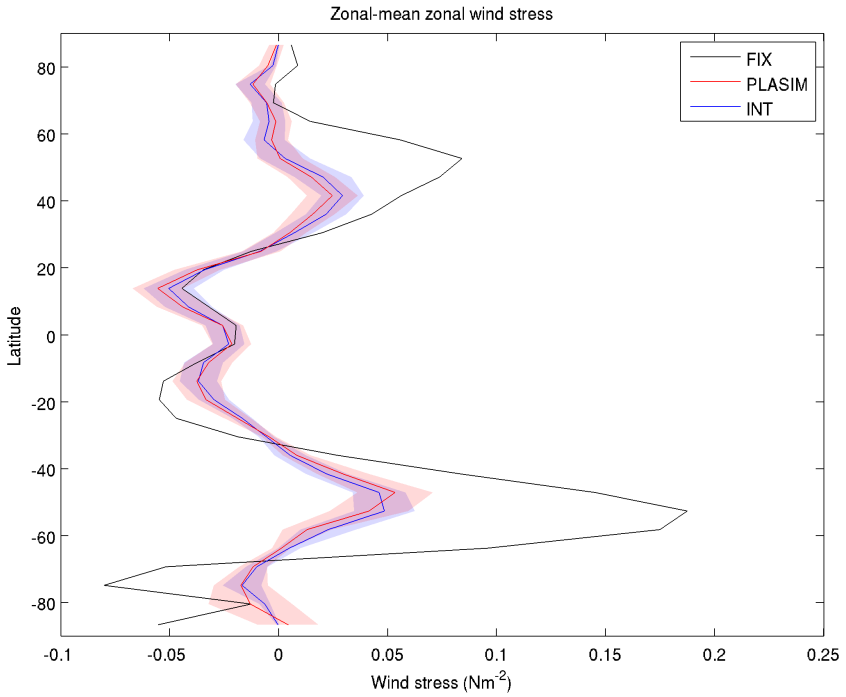


FIGURE 5.6: The zonal-mean zonal wind stress for the INT (red) and FIX (black) ensembles. The mean zonal wind stress of the PLASIM's simulated ensemble is also shown (blue). The ensemble means are accompanied by the standard deviations (shaded areas of corresponding colour).

stresses to compensate for the strong momentum damping in the physically simplified ocean. As discussed previously, while a large scaling factor can compensate for the weak wind stress in general, the structure of the INT stresses remain significantly different to the FIX fields. For example, by doubling the zonal wind stress, more realistic Northern Hemisphere westerlies can be achieved. However, that also leads to the easterlies being overestimated and the Southern Hemisphere westerlies will still be too weak. The interactive monthly wind stresses are likely to induce more seasonal variability within GENIE compared to when the fixed winds are used.

Normally, the strong westerly wind in the Southern Ocean induces upwelling of deep water masses towards the surface and diverts them northward by Ekman transport. A reduction in the zonal westerly wind stress is expected to reduce upwelling and ventilation of the Southern Ocean. This whole process can play a role in driving the AMOC through the 'Drake Passage' effect proposed by Toggweiler and Samuels (1995). While sea ice is not directly forced by wind stresses, it is advected with ocean surface currents that are thus forced by these fields. It is hard to estimate the possible effect of wind stress on sea-ice without further investigations.

5.3.3 GENIE's interactive winds compared to emulated PLASIM winds

The steady state interactive wind in GENIE-1 is expected to be different from PLASIM's simulated wind due to emulation error and the additional variability due to feedbacks with the ocean. Section 4.4 assessed the total emulation error which encompasses the truncation error and component error (Section 2.2). We now look at how the emulated wind components behave within GENIE, due to internal variability, by comparing the emulated PLASIM wind field, as a function of the SST (Section 4.4), with the statistical wind field evolved within GENIE-1. The emulated PLASIM winds are used instead of the simulated winds in order to eliminate the emulation error from the comparison. The 202-member validation ensembles are used in this section since we've already computed the emulated PLASIM wind fields for these simulations.

Figure 5.7 shows the INT ensemble winds (blue) compared to the equivalent emulated PLASIM wind (magenta). The differences seen are results of the interactions and feedbacks within GENIE's climate modules. The INT ensemble produces winds which, while reaching different steady states, are not too significantly difference from those reached using the one-way interaction between the SST boundary conditions and the winds. On average, the correlation between each pair is very high for both components (over 0.95%). The normalised RMSD between each pair is around 0.2, or 20% of the spatial standard deviation in the emulated PLASIM wind. The two-way interaction leads to INT ensemble winds having a slightly lower standard deviation. It is likely that with emulated precipitation and humidity field, the difference in the two ensembles will be larger due to additional degrees of freedom.

It's apparent that both the emulation error and the differences due to interval variability are small compared to the spatial variation over each 2-D wind field. Moreover, the zonal mean zonal interactive wind stress is similar to PLASIM's wind stress (Figure 5.6). Therefore, the poor comparison between INT and FIX wind stresses is most likely due to PLASIM's simulated wind being too weak.

This weak wind stress is consistent with Schmittner et al. (2011), in which the PLASIM wind stresses at T42 and T21 resolutions with different orographic scaling are compared to a reanalysis. The authors found that PLASIM significantly underestimated the Southern Hemisphere westerly wind stress at both resolutions. The T21 simulation had much weaker wind stresses in general compared to the T42 simulation, further underestimating the Southern Hemisphere westerly wind stress. Similar to our observation, the T21 wind stresses in both hemisphere are more symmetric than the reanalysis. The reanalysis used was the NCEP/NCAR reanalysis data from Kalnay et al. (1996), which is the same source as the FIX wind speeds (but different from the FIX wind stresses). The zonal-mean zonal wind stress from this reanalysis (0.16 $N\ m^{-2}$) is slightly weaker than the FIX wind stress (from NCEP-DOE reanalysis 2) shown in Figure 5.6. Based on Held and Phillipps (1993) and Tibaldi et al. (1990), Schmittner et al. (2011) concluded

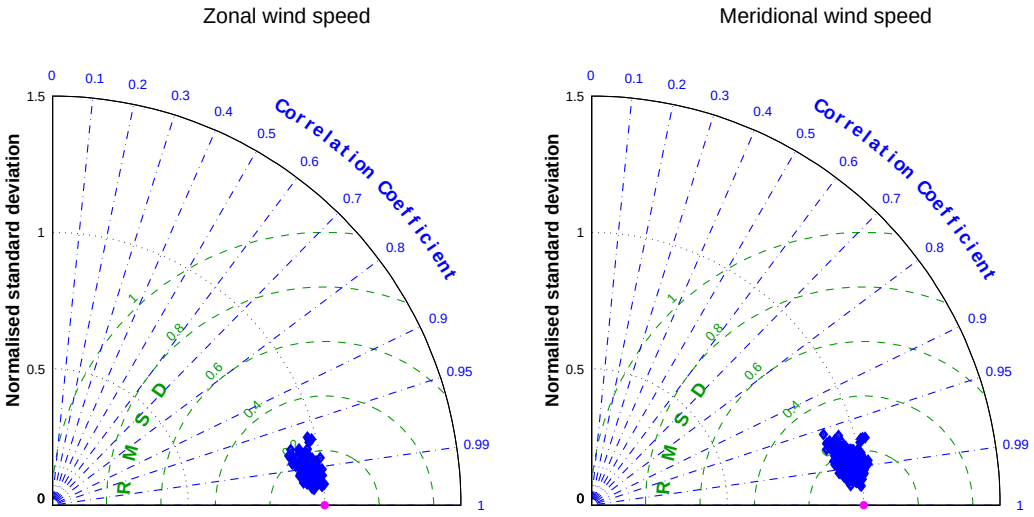


FIGURE 5.7: Taylor diagrams of the GENIE's interactive zonal and meridional wind components compared to the PLASIM's emulated values using the 202-member validation ensemble. The PLASIM's emulated values are treated as reality and are denoted by the magneta point. The GENIE's interactive winds are shown in blue.

that the differences in surface zonal wind stress between the T42- and T21- resolution versions are not due to the representation of orography but a result of the meridional resolution. Spectral atmospheric models with low horizontal resolution show a tendency to underestimate the horizontal eddy momentum flux. As a lower surface flux is required to balance the low momentum flux, this underestimation leads to weak simulated zonal winds. Tibaldi et al. (1990) and Schmittner et al. (2011) suggest that systematic errors associated with this feature converge around T42 resolution.

A Taylor plot of each INT wind component is produced to assess individual simulations, as well as the ensemble spread with respect to observations. Figure 5.8 displays a comparison of INT winds to long-term annual mean wind velocity components from NCEP-NCAR Reanalysis 2. The zonal components compare with the reanalysis more favourably than the meridional components. The average correlations are approximately 70% and 45% for the zonal and meridional components, respectively. The spatial standard deviations of both components are lower than observations, especially for the zonal component. This is likely due to the weak westerlies in both hemispheres. While there are errors introduced by the imperfect emulation process as demonstrated by the discrepancies between emulated and simulated PLASIM winds (Figure 4.11), the weak correlations seen here between GENIE's interactive winds and the reanalysis are largely due to the accuracy of simulated PLASIM winds. The GENIE interactive winds are obtained from emulators that treat PLASIM climate as the truth and hence inherit the same problems that PLASIM has.

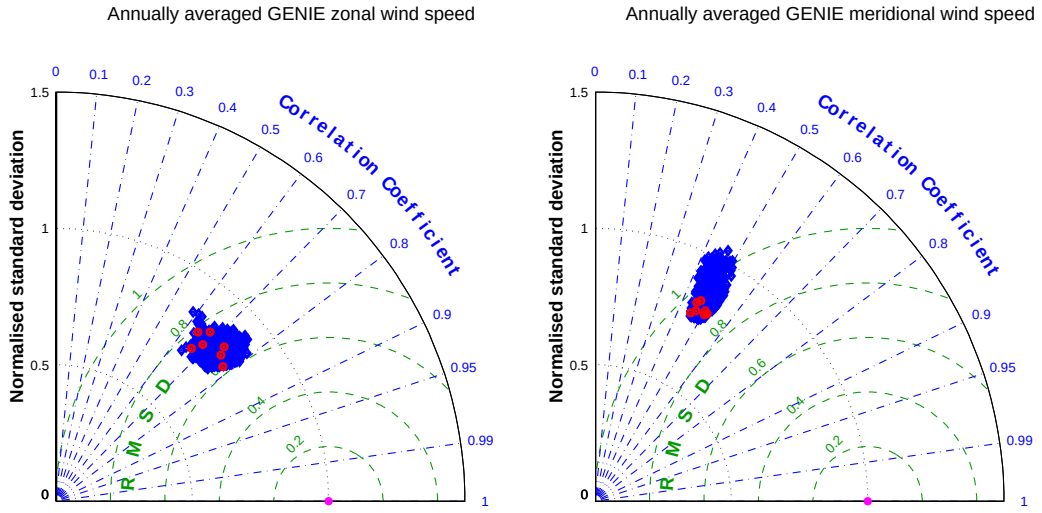


FIGURE 5.8: The GENIE's interactive zonal and meridional wind components compared to NCEP-DOE reanalysis 2. The red points denote simulations with ICF $\in \{0, 1, 2, 3, 4\}$ and 340 ppm $<$ RFC $<$400 ppm.

5.4 Assessing the new GENIE-1's climate

5.4.1 Compared to reanalysis

The GENIE-1's FIX and INT ensembles are compared against NCEP-DOE reanalysis 2. SAT and log of precipitation rate are shown in Figure 5.9. The new interactive statistical wind field does not significantly affect the ensemble comparison In the case of surface air temperature, there is little difference except for a small increase in the spread of correlation scores, meaning that the INT ensemble produces a more diverse spatial pattern compared to the reanalysis. A slightly lower normalised standard deviation is likely to be the result of weaker wind stresses and the lower standard deviation in the wind field as expected from the emulator. The effect of the wind field is more dramatic for precipitation where an increase in the ensemble spread in correlation is clear. Since the winds directly influence evaporation and the transport of moisture in the atmosphere, the interactive winds have a more apparent signal on precipitation. It is perhaps not too surprising that the new winds do not have a larger impact on both SAT and precipitation since the dominant process of heat and moisture transport in the EMBM atmosphere is still diffusion. While the interactive winds add extra degrees of freedom which work to increase ensemble variability, the weak emulated wind stresses counteract this to some extent.

While neither INT fields compare well with observation, they are also not significantly worse than the FIX fields, showing that the contribution to structural uncertainty in the global climate from the wind field in a diffusive atmospheric model like EMBM is

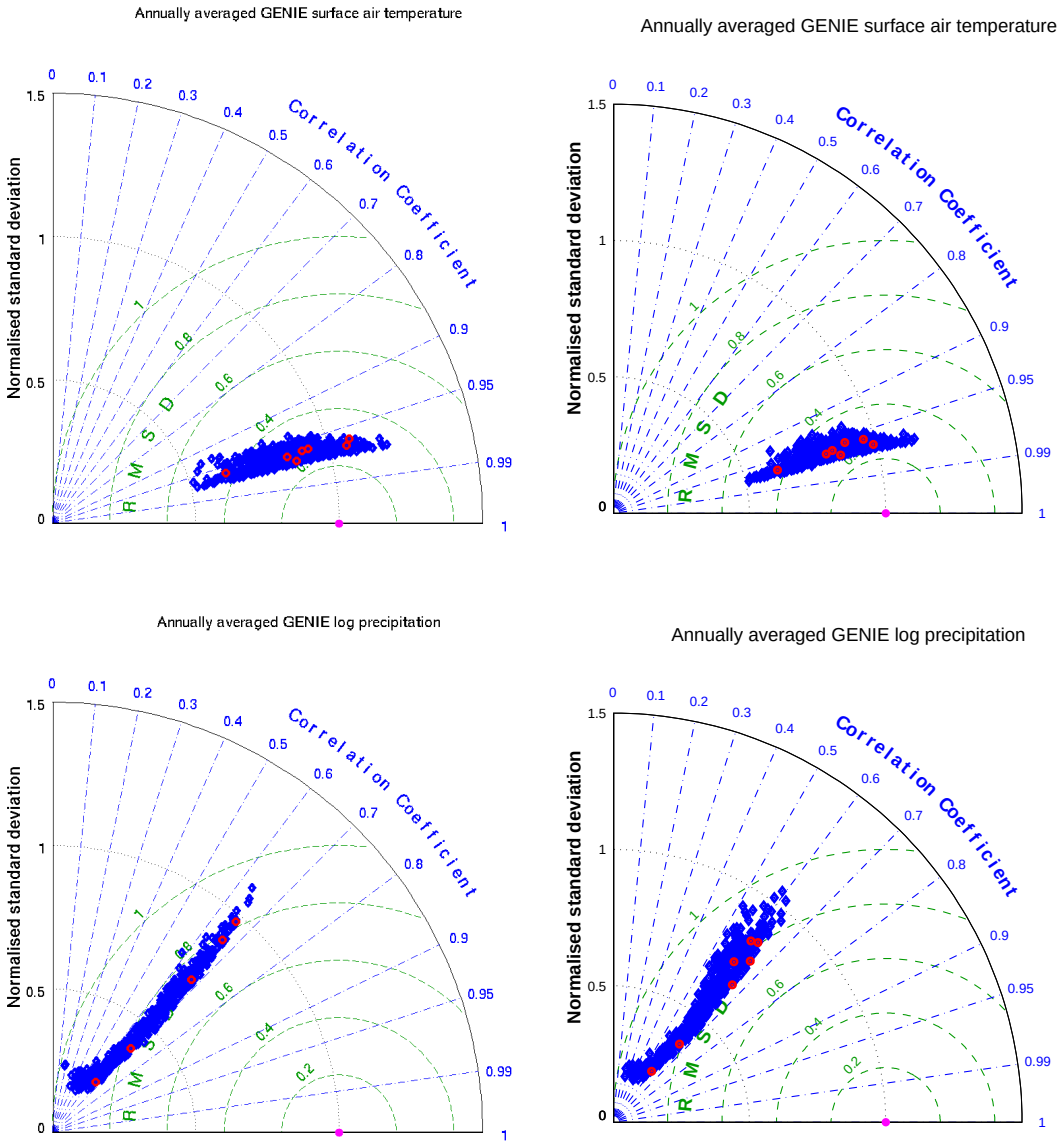


FIGURE 5.9: The steady state output fields from an ensemble of 5000-year integrations of GENIE-1 with fixed and interactive winds, compared to NCEP-DOE reanalysis 2: a) SAT with fixed winds, b) SAT with interactive winds, c) log of precipitation rate with fixed winds and d) log of precipitation rate with interactive winds.

limited. We now explore the differences in the steady-state climate in the INT and FIX ensembles in both the atmosphere and the ocean.

5.4.2 Compared to the FIX ensemble

To assess the differences between the climate in the FIX and INT ensembles, the global mean SST, SAT and the intensity of the Atlantic meridional overturning circulations (AMOC) are compared in Figure 5.10. Here, the strength of the AMOC is defined as

the maximum meridional overturning streamfunction in the North Atlantic below 500 m depth (ignoring the surface wind-driven circulation). While the trend relating the global mean SAT and global mean SST between the two ensembles is approximately linear, the AMOC clearly exhibits nonlinear behaviours. For SAT and SST, a linear fit is shown for each data set. The coefficients are also displayed in the Figure. The gradients for both fit lines (included in the figure) are close to 1, indicating a shift toward warmer climates across the whole INT ensemble, regardless of the input parameters. On the SST plot, the points are more concentrated around the fitted line at low values of global mean SST and diverge more at the higher end, indicating some weak dependency on the parameters.

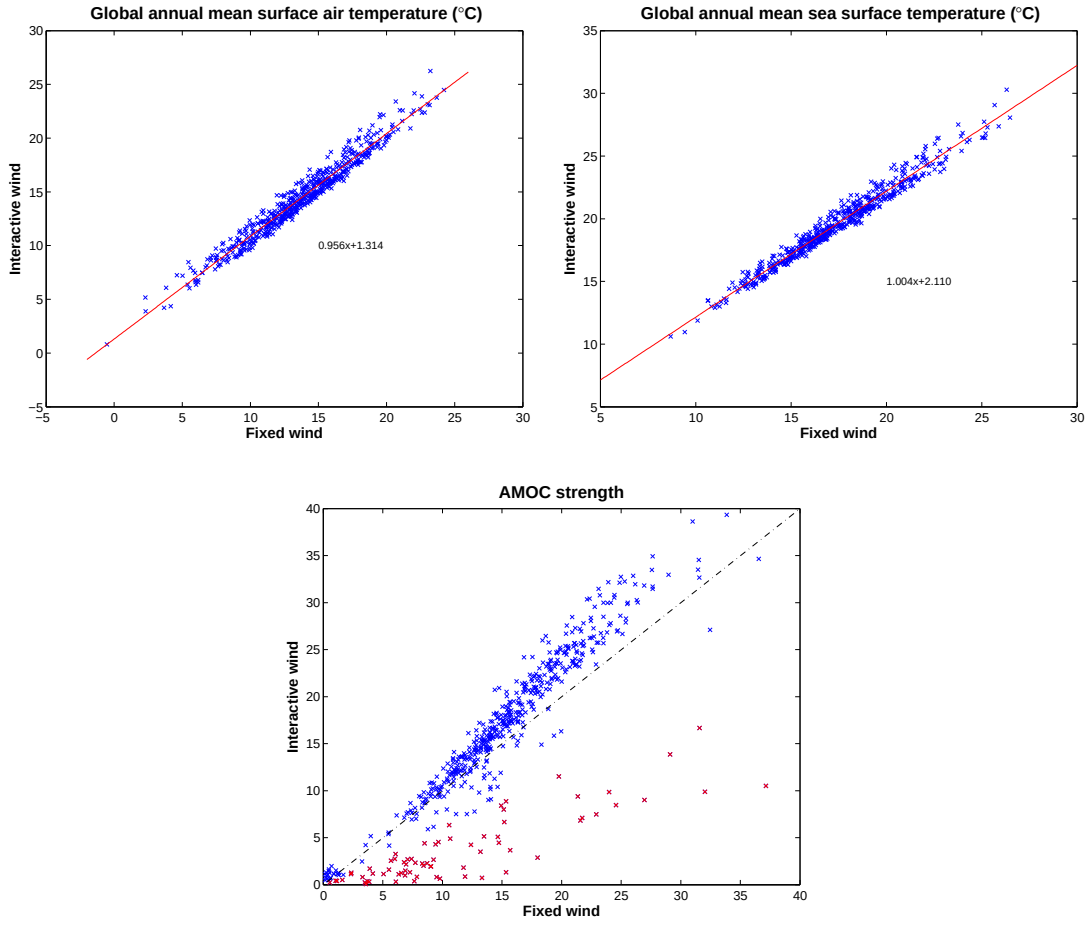


FIGURE 5.10: The interactive wind vs fixed wind outputs for: a) SAT, b) SST and c) AMOC strength. The best fit lines and the coefficients are also shown for SAT and SST. The dotted line in the AMOC plot is a reference line with a gradient of 1 and an intercept of 0.

For the North Atlantic meridional overturning streamfunction, the data points appear to follow two different branches. The use of the interactive wind fields can lead to a very different AMOC state given the same input parameters as before. The majority of the simulation pairs lie on the upper branch where INT simulations produce slightly stronger

AMOC. This branch might be approximated as linear, although it is less so compared to global mean SAT and SST. The lower branch shows that for some simulations, interactive winds lead to a significantly weaker AMOC compared to when fixed winds are used. The effect of this lower branch is not observed in the global mean SAT or SST plots suggesting that the effect on surface temperature is small and/or confined to a small region. Simulations with INT wind producing an AMOC of less than 60% the strength of its counterpart in the FIX simulations are highlighted in red.

Table 5.2 summarises the global mean of some selected quantities across the ensemble for comparison. Simulations on the collapsed branch are removed from the calculation of these quantities. Overall, the INT ensemble has higher SAT, SST and smaller sea-ice cover, as indicated by the linear coefficients in Figure 5.10. The difference in the ensemble means is not large, about 18% of the ensembles' standard deviation. The temperature over Greenland's grid cells is also higher in INT simulations. The INT ensemble exhibits smaller fluctuations in both global mean and Greenland SAT.

TABLE 5.2: A comparison of some global mean outputs from the FIX and the INT ensembles (upper half of the table). The second half of the table shows the equator-to-pole temperature difference computed in the two ensembles.

	FIX		INT	
	Mean	S.d.	Mean	S.d.
Global SAT (°C)	13.68	3.87	14.38	3.78
Global SST (°C)	17.39	3.04	19.57	3.10
Greenland SAT (°C)	1.25	5.43	1.40	5.23
Sea-ice cover (million km^2)	3.96	4.49	1.90	2.29
ΔT_{NH} (°C)	22.71	3.94	22.23	3.58
ΔT_{SH} (°C)	29.62	4.07	28.17	3.70
ΔT_{GL} (°C)	25.23	4.03	25.13	3.80

Also included in Table 5.2 are the equator-to-pole temperature differences ΔT. The Northern Hemisphere temperature difference is calculated as:

$$\Delta T_{NH} = T_e - T_n, \tag{5.2}$$

where T_e and T_n are the mean temperature at the equator (averaged across the region from 30°S to 30°S) and at high northern latitudes (averaged across the latitudinal band northward of 60°N), respectively. The Southern Hemisphere and equator-to-Greenland differences are calculated in the same way but with T_n replaced by T_s (averaged across the latitudinal band southward of 60°S) and Greenland mean SAT, respectively.

The INT ensemble has smaller equator-to-pole temperature differences in both hemispheres, indicating that the warming signal is stronger at high latitudes compared to

the tropics. The warming is also stronger in the Southern Hemisphere where temperature difference is smaller in the INT ensemble by over 1°C. Interestingly, the warming signal in Greenland, ΔT_{GL}, is smaller than the equator-to-pole temperature difference in the Northern Hemisphere, ΔT_{NH} and is approximately the same magnitude as in the FIX ensemble (only differs by 0.10°C on average). When the collapsed branch is taken into account, ΔT_{GL} is actually larger in the INT ensemble, indicating even less warming on average in Greenland.

Since the trend in temperature is persistent throughout the ensemble, it's most likely that the globally warmer climate is linked to the significantly weaker wind stresses. The wind stress is also a likely cause for the change in the equilibrium state of the ocean circulation. Figure 5.11 and Figure 5.12 show the ensemble mean Atlantic and global meridional overturning streamfunctions, respectively. The clockwise circulation is shown in red with solid contour lines drawn every 1 Sv. The blue colour and dashed contour indicate a flow in the opposite direction.

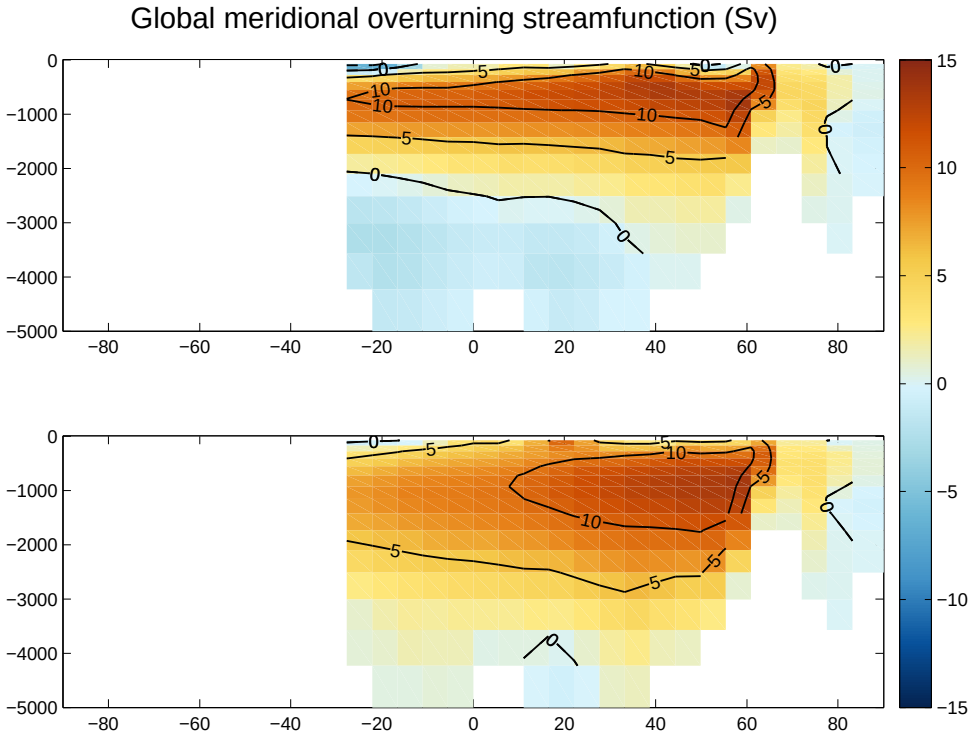


FIGURE 5.11: Ensemble mean North Atlantic overturning streamfunction for the FIX (left) and INT (right) ensemble. The colour scale denotes the strength of the overturning cell with red and blue indicating clockwise and anticlockwise circulations, respectively. The contours are drawn at 1 Sv intervals.

In both ensembles, we see an active North Atlantic meridional cell with an average strength of approximately 13 Sv. The clockwise circulation associated with the Atlantic overturning streamfunction is shallower in the FIX ensemble. In the INT ensemble, this

overturning cell extends deeper down to a greater depth while being more confined latitudinally. The bottom anticlockwise circulation associated with the Antarctic Bottom Water (AABW) can only be seen in the FIX ensemble.

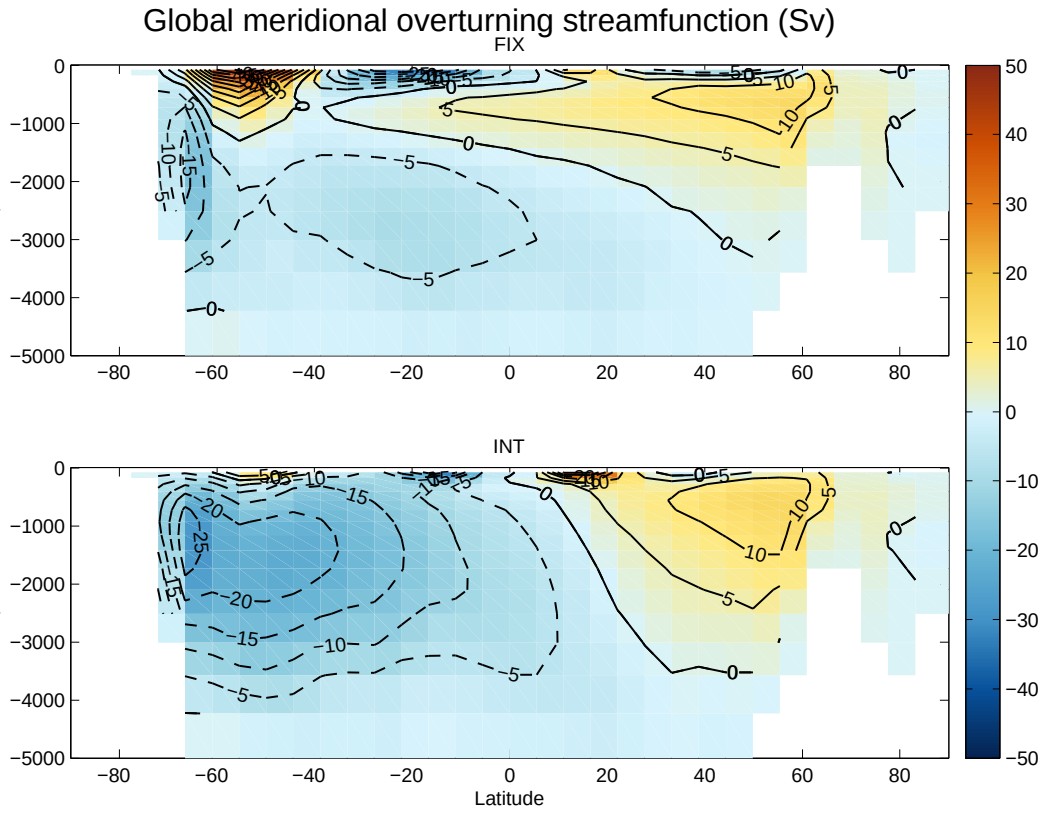


FIGURE 5.12: Ensemble mean global overturning streamfunction for the FIX (upper) and INT (lower) ensemble. The colour scale denotes the strength of the overturning cell with red and blue indicating clockwise and anticlockwise circulations, respectively. The contours are drawn at 5 Sv intervals.

Figure 5.12 shows the ensemble mean global overturning stream function. In the FIX ensemble, the northward flow in the upper layers of North Atlantic ocean transports warm and salty water to high latitudes where the water is cooled, becomes denser and sinks to form NADW. This dense water flows southward at 1000-2000 m depth and penetrates into the Southern Hemisphere where it mixes with water from different sources. Below this clockwise cell is an anticlockwise cell associated with the abyssal AABW, formed at high latitudes in the Southern Ocean. A strong Deacon cell driven by the westerlies is seen from 50–70°S. Other wind-driven cells are seen in the upper tropical and midlatitudes.

In the INT ensemble, the difference in the surface pattern strongly reflect the differences in the zonal mean zonal wind stress shown in Figure 5.6. The surface cells corresponding to the upwelling North of the equator is stronger in the INT ensemble while the upwelling south of the equator is slightly weaker. The weak westerlies wind stress in the Southern

Hemisphere results in a weak and shallow Deacon cell (50°S), most likely, associated with a weaker Antarctic circumpolar current (ACC). The weaker ACC and westerlies at high latitudes Southern Hemisphere would lead to more heat being exchanged between the equator and the South Pole, warming the Southern Ocean, reducing the sea-ice cover and decreasing the temperature difference.

Through the effects they have on the Deacon cell and the formation of the AABW, the weak westerlies significantly influence the structure of the global overturning streamfunction, making it more symmetric. With a reduced Deacon cell and strengthened surface upwelling at the equator, the NADW formed at high latitude Northern Atlantic which flows southward at intermediate levels does not penetrate into the Southern Hemisphere but instead returns to the surface in the tropical North Atlantic, resulting in a circulation confined to the Northern Hemisphere. This deep circulation reduces the extent of the AABW in the abyssal North Atlantic. This North Atlantic overturning cell also becomes slightly stronger and deeper indicating that deep water formation is largely unaffected by the weak wind stress in the North Atlantic. The small increase in strength of this cell might be a consequence of stronger upwelling at the equator, transporting more water northward and drawing more water upward in the returning flow.

The circulation cell in the Southern Hemisphere is almost a mirror of the Northern Hemisphere cell, with the deep water formed at high latitudes flowing at great depth and returning to the surface at the equator. This picture is consistent with the equator-to-pole temperature differences calculated in Table 5.2. The warming signal is stronger in the Southern Ocean while the different structure of the North Atlantic overturning circulation leads to a smaller warming signal in the Northern Hemisphere polar region.

Apart from the surface wind stress, other parameters such as the ocean friction coefficient and the surface freshwater flux (ODC and APM, respectively, in Table 3.1) also play an important role in controlling the state of the ocean circulation. APM is a scaling parameter for the surface moisture flux correction which transports fresh water from the Atlantic to the Pacific in three zonal bands (Marsh et al., 2004).

5.4.3 Accessing new climate states

We now compare the ensemble mean and standard deviations in the 2-D spatial fields. Figure 5.13 shows the ensemble mean (colour) and the ensemble standard deviation (contour) for SST and sea-ice cover for the FIX and INT ensembles. As a result of the warmer Southern Ocean, the INT ensemble has a lower fractional sea-ice cover. The warmer SST leads to a reduction in sea-ice cover, lowering the albedo, further warming the system. This persisted throughout the whole ensemble, as demonstrated by the low standard deviation in the Southern Ocean. The SST in the INT ensemble is also slightly warmer on average. Most visible differences apart from the sea-ice edge in the

Southern Ocean, are at high latitudes North Atlantic and North Pacific. The standard deviation shows much more variation and a less zonal structure of the SST field in the INT ensemble.

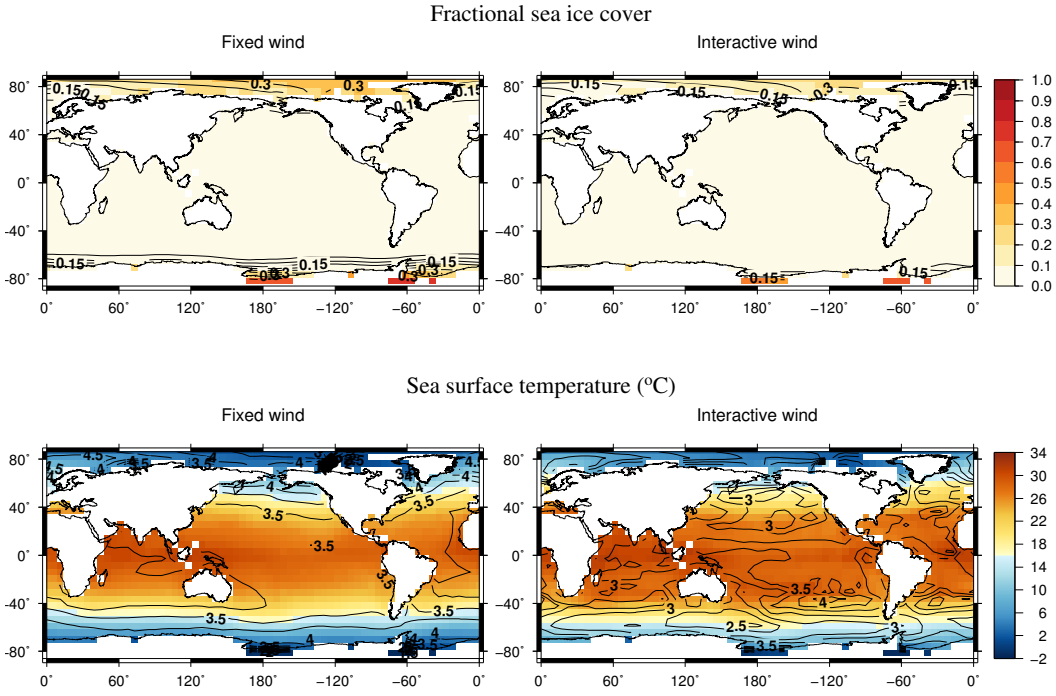


FIGURE 5.13: A comparison of the ensemble mean and standard deviation of fractional sea-ice cover and sea surface temperature from the two ensembles. The colour scale denotes the ensemble mean values and the contours denote standard deviations. Contours are drawn at 0.03 and 0.25°C for sea-ice cover and sea temperature, respectively.

The mean and standard deviation of SAT, HUM and PPTN in Figure 5.14 show similar features. In both hemispheres, the air temperature at high latitudes exhibits more zonal variations across the INT ensemble as a result of the wind fields affected by the elevations. The humidity in the tropics increases with the use of INT due to more evaporation as a result of the surface warming. The larger standard deviations of SST and SAT in the tropics and the increase in the wind speed components translate to larger variabilities in both humidity and precipitation in the same regions. For both of these fields, the continental areas which are normally too arid in GENIE now see larger changes in precipitation on average in the INT ensemble. There is also more variation in precipitation over a thinner ITCZ. Both the high humidity and high precipitation bands around the equator are more confined to the equator in the Pacific. The precipitation pattern in the FIX ensemble shows almost no variation over the ocean and at higher latitudes despite the large parameter range used. INT winds, with stronger meridional components which transport more moisture polewards, bring larger fluctuations to higher latitudes in both the Atlantic and Pacific. There are now variations seen just South of Greenland that were not simulated before. A previous study employing GENIE-1 with the prescribed

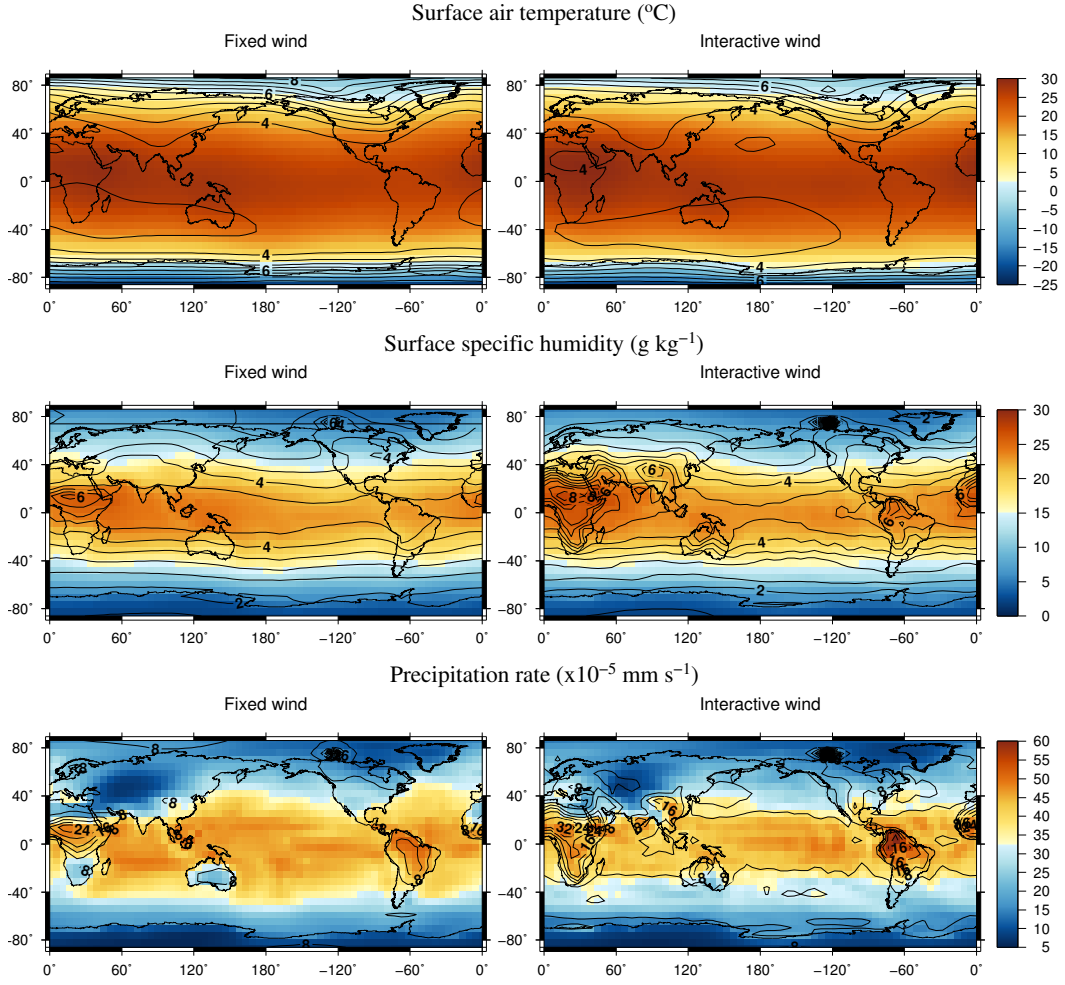


FIGURE 5.14: A comparison of the ensemble mean and standard deviation of SAT, specific humidity and precipitation rate from the two ensembles. The colour scale denotes the ensemble mean values and the contours denote standard deviations. Contours are drawn at 0.5°C, 0.5 g kg^{-1} and 4×10^{-5}mm s^{-1} for SAT, humidity and precipitation rate, respectively.

ice sheets by Lunt et al. (2006) has shown that the simulated LGM temperature over the Fennoscandian ice sheet is underestimated when compared to the HadSM3 model. This is a result of the pre-industrial land-sea mask used which leads to underestimation of the LGM Fennoscandian ice sheet extent. We have the same pre-industrial continental mask in all our simulations and so the same bias is expected. This also leads to an underestimation of the sea-ice cover in the Arctic.

In general, while the introduction of the interactive wind has a small contribution to the mean spatial climate patterns, more variations are seen in all variables but sea ice. Despite the weak amplitude of the surface wind stresses, the use of interactive winds works to reduce zonality and bring more variation to higher latitudes. Significant changes to the system can be observed in certain regions of the parameter space where the AMOC reaches a different equilibrium state. While exciting, this might be a result

of the structural bias of the wind field rather than a robust feature from the interactive aspect of the winds.

5.5 Summary and conclusions

This chapter presented the steps taken to implement PLASIM's wind emulators as a component of the EMBM atmosphere. We will now summarise the main impacts of the new interactive winds, the advantages, the drawbacks as well as some possible improvements.

While the setup is currently implemented for the basic GENIE-1 configuration with ENTS at 64×32 mesh, an interpolation routine can translate between the different meshes of GENIE-1 and PLASIM. Technically, since the emulator's inputs are the PCs of the SST forcing field, as long as the PCs derived from the new SST are within the range of the training set, the emulator would be valid, even if the model parameters are outside the range of the original training set. However, this is only true if the model still behaves in a similar manner and has the same climate sensitivity in extended parameter region. When an extra component is activated, however, it is likely that the model's behaviour and sensitivity will be different, and thus the emulators constructed are no longer valid.

The model becomes more computationally expensive due to the extra calculations resulting in each 5000-year integration taking more than twice as long to run. With each addition of an emulated component, the model will certainly require more computational power. However, considering that the PLASIM atmosphere driven by GENIE's SST took roughly 100 times longer to integrate the same amount of simulation time, this new configuration of GENIE-1 is still significantly more efficient and the model remains suitable for ensemble runs or long-term integrations.

A whole annual cycle of historical SST forcing is taken into account when generating the new statistical wind field at each time step. By doing this, the emulators are, effectively, emulating the dynamical response of the wind fields given a memory of the SST forcing. This also constrains the 'smoothness' of the wind field's evolution in time. It is unlikely that the winds can significantly shift between a short period of time considering its autocorrelation.

The impact of the weak wind stresses on the AMOC as well as the global overturning stream function is largely in agreement with the study by Saenkon et al. (2002), where a more symmetric global pattern is seen with wind stress removed. Unlike their case, however, our Easterly wind stresses are not removed but enhanced, leading to a stronger North Atlantic overturning instead. Unlike experiments with wind stress weakened

evenly or completely removed Toggweiler and Samuels (1995); Rahmstorf and England (1997), the 'Drake Passage effect' is not observed in the simulations of the INT ensemble.

The most significant impact of the new interactive statistical wind field is to shift the model's climate toward a warmer state compared to when the fixed climatological wind field is used. It also leads to broader model response within the same parameter space that is most evident in the AMOC strength and the amount of rainfall over continents. This is encouraging since the precipitation in EMBM is known to be too arid in substantial areas of continental interiors.

This increase in ensemble variability is small partly due to the weak emulated winds. While the emulators using PCA tend to underestimate spatial variability in the emulated fields, the large discrepancy seen between the interactive statistical winds and climatology is believed to be a result of PLASIM's limitations. Studies conducted using PLASIM and other atmospheric models in the past have shown that representation of wind stress over the Southern Ocean is insufficient at T21 resolution.

The task of implementing an emulator of the wind field of an atmospheric generation model as a statistical component in the EMBM atmosphere of GENIE-1 has been achieved in this work, demonstrating another possible application for emulators of high dimensional output fields. While only the wind field is discussed here, it is straightforward to introduce the emulators of other atmospheric variables, namely the humidity, precipitation and surface air temperature fields. However, considering the structural inaccuracy of the simulated PLASIM wind fields, it might be more sensible to address this issue before introducing further sources of complexity to EMBM.

By only replacing the wind field, we have the chance to investigate the effect of having an interactive wind field with a different structure on the global climate. The original motivation for wanting to replace the simple and diffusive EMBM atmosphere with emulators of a GCM is to improve the simulated climate in Greenland and the North Atlantic. With the new wind fields taking into account the change in ice sheet mask and interaction between the ocean and the atmosphere, the new steady state climate exhibits more variations across the ensemble, across the 2-D spatial pattern both zonally and meridionally. This effect, however, is small and is more visible at low latitudes. This can be attributed, in part, to the structure of the emulated winds and in part, to the diffusive manner by which the heat and moisture are transported in the atmosphere. Therefore, simply changing the wind field in an energy and moisture balance model like GENIE-1's atmosphere is unlikely to alter the climate state significantly. This is especially true when considering the relative magnitude of the changes induced by the wind field to the parametric uncertainty associated with the other 12 varying parameters.

While the changes in wind field as a result of the presence of the Northern Hemisphere ice sheets can have a significant impact on the high latitude climate and possibly the global climate through various feedback processes, they cannot be captured using a

simple EMBM atmosphere even when an interactive wind is introduced. This analysis suggests that in order to significantly modify the climate response at high latitude and in Greenland, apart from addressing the issue with wind speed patterns, both the heat and moisture components of the model require improvement. This leads to the next step which is to alter the moisture content and the hydrological cycle by introducing emulated surface humidity and precipitation rate. Similar to the work presented here, the effect of the added complexity should be assessed in terms of contribution to the resulting climate state.

Based on the comparison between model outputs and climatology as well as the known limitations of the wind field, the applicability of this particular setup should be regarded with reservation. The importance of the Southern Hemisphere westerly wind stress in driving the ocean circulation and biogeochemical cycling are well known (Sarmiento et al. 2004). Apart from directly driving the ocean currents, winds also play an important role in controlling the heat and moisture content of the atmosphere. At this point, it is not obvious that the benefit of having access to a broader range of climate states can outweigh these drawbacks. A potential solution to this structural problem is to utilise the emulators as estimators of the wind variabilities in response to the perturbed parameters and forcings. Climatological winds can be used as the mean field. By doing this, a more realistic wind structure is ensured while extra variabilities can still be achieved. This issue, of course, in no way diminishes the advantages of the type of statistical coupling introduced in this work. At this point, we have successfully constructed and tested emulators of various atmospheric variables from the PLASIM atmosphere model. These emulators have been implemented into the GENIE framework, creating a new functioning model that allows feedbacks between a statistical wind field and other components of the EMIC's climates system. While introduced at an extra computational cost, the model continues to be efficient which is central to this work.

Initially, a one-at-a-time approach was planned for the incorporation of the remaining emulators (e.g. SAT, precipitation and humidity). However, it is likely that this can lead to an imbalance in the heat and moisture budget of the system. In the case of the wind field alone, other processes remain consistent within the system. When introducing only emulated precipitation and humidity, however, a potential mismatch between the precipitation received and the evaporated content from a certain grid cell and in the system as a whole can occur. Over time, this can lead to a persistent net gain or loss of the moisture content of the atmosphere and results in unphysical climate states. The same situation can happen to the heat transport between the ocean/sea-ice and the atmosphere. This means that for the next stage, to avoid unnecessary complications, all other aspects of the atmosphere should be included at once.

The equilibration time for the coupled system is of the order of centuries or longer when deep ocean circulation is concerned. Compared to the ocean, the atmospheric response

time to climate forcing is relatively short due to its lower heat capacity. Thus, the assumption that the atmosphere responds instantaneously to ocean forcing is a reasonable assumption. This allows the use of asynchronous time-integration in which the oceanic and the atmospheric modules are integrated concurrently over two different time periods. In this case, the ocean was integrated for 5000 years, continuously exchanging the boundary conditions with a statistical wind module which provides estimations of the PLASIM's winds integrated over 35 years. This coupling strategy, demonstrated here between the oceanic and atmospheric components of a climate model, can apply to any system in which such an asynchronous coupling scheme is valid. For example, the coupling between an AGCM or a coupled GCM to an ice sheet model (Pritchard et al., 2008; Herrington and Poulsen, 2012).

Chapter 6

Conclusions

6.1 Summary and discussion

This thesis describes the construction of a statistical surrogate model of various atmospheric quantities from an AGCM of intermediate complexity using a combination of GP emulator and dimensional reduction techniques. The emulators are then incorporated into the efficient EMIC framework of GENIE, allowing the ocean module to communicate with a statistical wind field.

Over the past several decades, both theoretical studies into dynamic response of the atmospheric circulation to ice sheet orography (e.g. Lindemanm and Oerlemans, 1987; Roe and Lindzen, 2001; Liakka and Nilsson, 2010) and studies of the LGM climate using AOGCMs (e.g. Otto-Bliesner et al., 2006; Laîné et al., 2009; Rivière et al., 2010) have highlighted the importance of capturing a wide range of feedbacks induced by the continental-scale Northern Hemisphere ice sheets, especially the Laurentide. While comprehensive coupled GCMs are ideal for this task in terms of the physics described in the models, their computational cost does not allow simulations on a long time scale to be performed. It also restricts the ability to assess model errors due to uncertainty in the input parameters or the reconstructed boundary conditions. Our GENIE framework, on the other hand, is suitable for multi-millennial scale simulations and can be used to produce large ensembles to investigate model uncertainties. While generally comparing well with AOGCMs, EMICs with simplified physics often produce a less prominent and more diffusive signal compared to comprehensive models (Stouffer et al., 2006; Plattner et al., 2008). The 2-D EMBM atmosphere of GENIE-1 with its simplified physics is not adequate at capturing the wide range of feedbacks due to changing ice sheet orography. Motivated to improve the simulated glacial climate in GENIE-1, especially in the North Atlantic and Greenland region, aspects of a comprehensive atmospheric model, PLASIM were incorporated into GENIE-1 without sacrificing its efficiency by the use of a combination of statical techniques.

Driven by the desire to employ the climate model hierarchy in an effective way to inform our understanding of the models at different levels of complexity, we proposed the use of co-kriging to emulate PLASIM. While emulators have been employed in climate research to efficiently provide predictions of model outputs at unsampled inputs for a broad range of applications: uncertainty (Lee et al., 2012) and sensitivity analysis (Bounceur et al., 2015), calibration or history matching (Edwards et al., 2011) and for applications in Integrated Assessment Models Holden and Edwards (2010), these emulators tend to use only information from the model being emulated. There have been other instances where a computer code at a lower level of complexity was employed to make inference on the more complex code to aid the emulation process (Forrester et al., 2007; Cumming and Goldstein, 2008). Using co-kriging, we have successfully utilised the extra information from the EMBM output to efficiently capture PLASIM's global annual mean SAT in Chapter 3. The result shows that co-kriging reduces the number of expensive simulations used by 25% and 30% for the two examples presented. PCA was then introduced to extend the application to the case of high-dimensional output, which is the 2-D SAT field. PCA allows us to decompose the ensemble climate into meaningful spatial patterns which, with the help of universal kriging, helps illuminate the role of certain model parameters with respect to each mode of variations. It has been demonstrated in this work that while the two models exhibit different features and variability across the 2-D output fields, the response from EMBM and PLASIM for the same inputs are highly correlated. With the link between the PCs of the two ensembles successfully identified, the 2-D SAT as a function of 13 model parameters is emulated using half the expensive simulations compared to when kriging is used. This approach offers a possible way to tackle cases in which the climate model is too computationally intensive to produce enough simulations to construct a one-level emulator. Kennedy and O'Hagan (2000) discussed the possibility to extend the multi-level method to beyond two levels of complexity.

This work demonstrates the use of a model hierarchy to improve our understanding of the individual models and the commonality and differences between the two models in response to related inputs. Cumming and Goldstein (2008) demonstrated the use of structurally related models (the coarse simulator has two layers compared to a 25-level fine simulator) using a different multi-level emulation technique. Structurally unrelated models are considered to be more challenging since it is not guaranteed that a meaningful connection can be established due to either the large gap in their complexities or to the lack of commonality in the input/output space. In our experiment, the use of GOLDSTEIN's variables as boundary conditions for PLASIM in combination with the same boundary forcing (CO_2 and ice sheets), allows the two models to be linked through the shared input space. In this work, the EOF modes between models share similar features and driving mechanisms. This is the first time the PCs have been used in combination with a multi-level emulation technique on structurally unrelated models.

While PCA utilises the correlation in space between simulations of each model, the co-kriging emulator exploits the connection between the same variable simulated by two different models.

Chapter 4 focuses on developing the method further to combine the spatial correlation between various variables in the same model with co-kriging so that SAT from EMBM can be informative to the emulators of PLASIM's precipitation, specific humidity and wind fields. Furthermore, PCA was also used to take into account the seasonal cycle by treating the monthly surface fields as a single field. This approach is based on the work done by Holden et al. (2014). The emulator constructed using this method has a knowledge of how the atmosphere behaves within an average model year as a function of the SST boundary forcing. Another important feature introduced in this chapter is the dimensional reduction of high-dimensional input fields following the work from Holden et al. (2015). In this case, the original input consists of several fields (SST, SIC and SIH) which evolve in time (12 months). Similar to the output, PCA was used to perform this step. An ad hoc screening procedure was then carried out to reduce the number of inputs further to the final number of 7. This was possible due to the correlation between the SST and the sea-ice fields. As a result, GENIE-1's oceanic fields can be used as inputs, enabling the possibility to emulate PLASIM's atmospheric variables as a function of SST directly. This way, fluxes between the atmosphere and the ocean can be emulated, allowing the exchange between the ocean and the statistical components. The coupling of a climate model with a statistical emulator of a complex subcomponent can be achieved efficiently using this approach. Indeed, an example where our technique can be applied is in emulating a carbon cycle model to provide an estimation of the atmospheric CO_2 concentration as a function of a time series of anthropogenic CO_2 emissions and non-CO_2 radiative forcing (Foley et al., 2015). CO_2 concentration from coupled climate-carbon cycle models can be emulated and replace the simple carbon cycle component often used in integrated assessment models.

Another advantage of the dimensionally reduced inputs is their use as a screening method. While traditional model parameters are treated as independent, high dimensional spatial input fields have structures which, if reduced effectively, can be represented by a much smaller number of variables than the dimension of the fields. Since the computational cost of constructing an emulator depends strongly on the number of inputs, having a manageable number of inputs is desirable. Large spatio-temporal boundary condition fields can be reduced to a manageable set of inputs allowing quantitative assessment of their uncertainty to the model's simulated climate. An example where this technique can be valuable is in studying the effect of using different ice sheet reconstructions has on the atmospheric circulation and ocean circulation. As discussed in Chapter 1, studies using ice sheets of different heights concluded that the range of uncertainty in existing ice sheets reconstructions is enough to result in different behaviour

of the atmospheric jet through downstream effects on planetary wave generation (Otto-Bliesner et al., 2006; Ullman et al., 2014), which in turn contributes to large differences in temperature, the storm track or even the ocean circulation. Little has been done so far to identify the consequences of changing ice sheet heights, the location of maximum height or the ice extent. Emulators that treat the high-dimensional ice sheet reconstructions as inputs offer an opportunity to explore the uncertainty introduced from the boundary conditions and how sensitive the model's simulated climate is to these boundaries.

With the statistical techniques developed in Chapter 3 and 4, the surface wind components, humidity and precipitation are emulated. The emulator for specific humidity is the most successful one, capturing 97.88% of the ensemble variance. The wind and precipitation rate are more challenging to emulate. Among wind components, the zonal component was more successfully emulated (73.66%) while the meridional component achieves 69.62%. Only 64.88% of the total ensemble variance is captured by the log precipitation emulator. This field is the most challenging to emulate due to the large internal variabilities. Nevertheless, the emulated fields resemble PLASIM's precipitation fields, which can simulate features that are not seen in EMBM, e.g. a defined ITCZ. Each emulator's performance was assessed by comparing predictions with the simulated values of a separate validation ensemble. Two sources of errors are considered, the truncation error and the component error. In general, the linear decomposition of combined fields leads to a reduction in the variance explained by the top EOFs. As a result, the emulators tend to underestimate the variance across the ensemble, failing to capture extreme values. The lower the correlation between the emulated field with SAT, the poorer this decomposition technique works. This is evident from the fact that the truncation error is lowest for humidity and highest for precipitation. The second source of error is the emulator's inability to correctly emulate the PCs. In all cases, the emulators perform well for the first few components but start to level off at around the fifth mode. It's possible that other dimensional reduction techniques could improve the emulator performance.

With the emulators validated, Chapter 5 demonstrated the implementation of the wind emulators into the existing EMBM atmospheric module. This is similar to the hybrid coupled model from Barnett and Preisendorfer (1987) and Barnett et al. (1993) in which the relationship between EOFs of the SST and surface winds are modelled. While Barnett and Preisendorfer (1987) used canonical correlation analysis to infer the spatial pattern of the wind field from the given SST, we offered an alternative which employs statistical emulators in combination with PCA instead. The results of GENIE-1 with the newly implemented interactive winds are compared to a previous ensemble with fixed climatological winds. The results show two main findings: i) the inclusion of an interactive wind field to an energy-moisture balance model has little effect on the spatial distribution of atmospheric quantities, especially for SAT (precipitation is more

affected). However, it does lead to more variations to be observed across the ensemble; ii) the wind simulated in PLASIM at T21 shows some significant structural bias with weak westerlies and strong easterlies. As a result, a different equilibrium structure of the overturning circulation is observed across the interactive wind ensemble. Unlike studies of the effect of changing Southern Hemisphere wind stress on ocean circulation, where only the Southern Hemisphere winds are changed (Toggweiler and Samuels, 1995; Delworth and Zeng, 2008) or the wind structure is scaled up and down globally (Montoya and Levermann, 2008; Yang et al., 2016), in our ensemble, the winds in the Northern Hemisphere are also changed and the easterlies are strengthened. Thus, a weakening of the AMOC is not observed in response to the reduced upwelling in the Southern Ocean in most of the simulations. Several simulations with interactive winds, however, do end up with a significantly weaker AMOC. This suggests that a threshold is crossed in the system, but the mechanism is yet to be explored at this stage.

Despite the drawback from the biased wind structure, this thesis demonstrated the introduction of an emulator as a statistical component within an EMIC framework. This raises the possibility for emulators to be used in similar ways, for example, to replace super-parameterisation models of subgrid scale processes within a climate model or to provide dynamic forcing to drive an OGCM. Both the statistical tools and concepts discussed in this work have wide reaching applications in climate research and beyond.

6.2 Future work

The result presented Chapter 3 has been published in a peer reviewed journal (Tran et al., 2016) and the work done on the other two result chapters can potentially be developed into manuscripts for publication. The newly constructed statistical component of the atmosphere can be made accessible to the research community through the GENIE's code repository. There remains much to be done to improve the quality and broaden the scope of the research presented in this thesis.

Due to time constraint, only the wind components have been implemented so far. As previously discussed, ultimately, the specific humidity, precipitation rate and SAT should also be implemented. Apart from these fields, there are also radiation terms that need to be considered. Based on the comparison between the FIX and INT ensembles, simply replacing the wind field has small effects on both the global model behaviour and the regional response in the North Atlantic and Greenland. In general, the spatial structure of atmospheric variables such as SAT, humidity and precipitation remained largely controlled by diffusivity. Therefore, we'd like to investigate whether the situation can be improved when the moisture and heat terms are also introduced, completing the AGCM emulator. Unfortunately, the structure of PLASIM's atmosphere appears to be a significant limiting factor to the application of the current statistical model. Ideally, the work

presented should be repeated with a higher horizontal resolution atmospheric model. The most likely choice would be PLASIM at T42 resolution, based on the conclusion by Schmittner et al. (2011). In that case, additional simulations are needed to train the new emulators. Both the GENIE-1 and PLASIM ensembles produced here can be reused (as ensembles of opportunity) to emulate PLASIM T42. The multilevel technique can utilise either or both available GENIE-1 and PLASIM simulations. While only two levels of complexity are concerned in this work, it is theoretically possible to extend the emulation framework to more than two levels (Kennedy and O'Hagan, 2000), utilising both available ensembles.

An important point which was not addressed in the work is the model-data comparison. Throughout the thesis, we have treated PLASIM output as the 'truth' and the output from GENIE-1's EMBM as a fast approximation of that truth. Thus the emulator performance is assessed based on its ability to reproduce PLASIM's outcome. Ultimately the model is expected to reproduce a climate that compares well with reality or in the case of climate research, with reanalysis climatologies or reconstructions from proxies. Therefore, the behaviour of both models requires further assessment. For example, the ability of GENIE-1 and PLASIM with boundary conditions prescribed by GOLDSTEIN's ocean and sea-ice outputs to simulate the LGM or pre-industrial climate should be tested. While GENIE-1 has undergone various tuning exercise in the past, none of them was performed on the same model mesh with the same module configuration and hence are not necessarily applicable. While a large number of GENIE-1's parameters were perturbed in this work, the PLASIM's parameters (apart from the ice sheets and greenhouse gas concentration), are kept at default values in all the simulations conducted here. These parameters are tuned with an annual climatological SST from AMIP (Atmospheric Model Intercomparison Project) II-period (1979-1996) (Haberkorn et al., 2009). A commonly faced problem when coupling models which are independently tuned is that the interaction between the coupled GENIE-PLASIM is statistically different from the forcing used to tune PLASIM alone. Thus, it is not unexpected that they drift toward a new climate state at equilibrium. It is important to note that this does not mean we would like to determine a fixed set of parameters tuned specifically to reproduce climatology when driven by GENIE-1 but instead to make sure that the range of the ensemble outcome encompasses climate states which are in good agreement with climatology. This concept is referred to as 'precalibration' or history matching (Holden et al., 2010), whereby the model is required to reproduce the main features of the climate system but is not constrained by detailed observation. In our case, the aim would be to obtain a more realistic structure of the wind field in PLASIM.

Throughout the work conducted, there are areas in which further investigation can be beneficial. We have managed to employ co-kriging to emulate PLASIM's high-dimensional outputs. We have investigated the discrepancy between simulated and emulated quantities both in terms of reproducing the ensemble mean and the spatial

distribution of large errors. While not explored in this work, examining the emulator's estimated uncertainty and how it propagates with the dimensional reduction of combined fields can inform us of the emulator's shortcomings and possibly diagnose areas which can be improved. Successful emulation relies on the fact that while having distinctive spatial patterns, the first-order response in both models to boundary forcing conditions, described by their PCs, are well correlated and can be linked via the single multiplier formulation of co-kriging. While this connection works particularly well for the first few SVD modes, the emulators of low-order PCs perform less well. This might be because the lower order processes in the two models are less correlated. Since the decomposition determines purely statistical modes, they often don't correspond to an actual physical process in the models. It is possible that if the dimensionally-reduced modes of variation represent meaningful behaviours of the system, the correlation between the PCs from the two models can be improved. A possibility which deserves consideration in this case is the rotation of principal components. In cases where the PCs are used purely for dimensional reduction purpose, unrotated EOFs/PCs are good solutions. However, we are interested in understanding the model behaviour as well as identify meaningful relationships between two levels of model complexity through EOFs/PCs and rotated EOF offers the ability to isolate specific modes of variation. The idea is to transform the EOF to another system of coordinates by applying a rotation matrix which fulfils a specific criterion. There are various rotation criteria in the literature as well as methodologies to analyse the data to identify the optimal choice of an optimal solution. Examples of both can be found in Richman (1986). Other possibilities including the use of factor analysis, unrotated or rotated (Osborne and Costello, 2009) instead of PCA or the use Canonical Correlation Analysis (CCA) to related, for example, the wind components or precipitation rate of PLASIM to the SAT of PLASIM or directly to the SST of GOLDSTEIN. CCA defines coordinate systems such that the correlation between the projections of two different datasets onto these coordinate systems are mutually maximised (Barnett and Preisendorfer, 1987; Hardoon et al., 2004).

The treatment of temporal variation can also benefit from further work. In this work, we followed the approach of Holden et al. (2014), which uses PCA to reduce the dimensionality of spatio-temporal data. This method treats the temporal behaviour of each quantity the same way as spatial variations. There exist several alternative approaches to emulate a timeseries output such as those discussed in Conti and O'Hagan (2010). The first and most simple method is the 'many single-output' emulators which emulate the outputs in time separately. In this situation, we would need to build 12 separate emulators, one for each month. Another possibility is the 'time input' emulator which treats time as another parameter and each month can be considered as a training point. This approach would lead to 12 times as many training points, which increases the computational expense required to build an emulator significantly. Finally, Conti and O'Hagan (2010) proposed a 'multi-output' emulator which generalised the univariate GP emulator work with a multivariate output. Each of these methods

has some advantages over the others. Holden et al. (2014) pointed out that the use of spatio-temporal data allows for the possibility of abrupt transitions because spatial fields are not forced to be similar through time. The 'time input' method is more restrictive because it would impose a form in time. The 'many single-output' approach, on the other hand, fits separate emulators to each month and hence no fixed structure in time is assumed. This flexibility can be advantageous, depending on the application. This is less of an advantage in our case since we are not dealing with a long time series but with a seasonal cycle and spatial fields are expected to behave in a known fashion. A 'time input' emulator is a more rigid and computationally intensive special case of the 'multi-output' emulator. These methods, among other multivariate techniques, often deal with either high-dimensional spatial outputs or scalar outputs that evolve in time. Future research on techniques which incorporate both of these factors seems worth investigating considering its applicability in works involving time-varying high-dimensional boundary conditions.

This leads us to another possible direction for future work, which is to ensure that our hybrid model with a statistical atmospheric component is capable of performing transient simulations. The statistical model would then need to work with boundary-forcing inputs which evolve in time, e.g. the greenhouse gas concentration and the ice sheet reconstructions. The current dimensional reduction method for inputs can be applied directly on a CO_2 timeseries, decomposing it into a set of parameters. Apart from PCA, several other techniques can be used to decompose a time series. For example, Holden et al. (2014) used a linear decomposition into Chebyshev polynomials. The situation is more challenging for ice sheet configurations since it involves 2-D fields which change in time. So far, we treated each time slide of the reconstruction as a parameter, encompassing not only the ice sheet heights but also the extents. The changes in ice sheet configurations between different time steps is nonlinear. Hence, the model's behaviour is most likely to be non-stationary with respect to this parameter. This treatment, while satisfactory for the current steady-state simulations, requires improvement to work in transient experiments. Instead of treating the ice sheet mask as a numerical parameter corresponding to the time slice, a dimensional reduction technique could be used to classify the structure of the ice sheet in terms of a set of parameters (PCs corresponding to certain structural modes). This allows flexibility in the boundary conditions in transient simulation as well as enabling uncertainty/sensitivity analysis on these inputs.

Another important aspect in extending the technique to transient simulations is the assumption made about the equilibrium response of the atmosphere. This means that the hybrid model is not designed for applications in which the atmosphere and the ocean vary together on interannual timescales (e.g. the ENSO). On the long timescales considered in palaeoclimate research, however, the transient behaviour of the Earth system appears to be very close to equilibrium to the large scale boundary forcings during the period from 30 kyr BP to the pre-industrial, e.g. changes in orbital parameters,

CO_2 concentration and ice sheets when fresh water from melting is ignored (Lunt et al., 2006). The difference between long-term transient and steady-state simulations in which a collapse in the ocean circulation occurs is yet to be tested.

Appendix A

Comparison of the annual mean simulated climates to observation

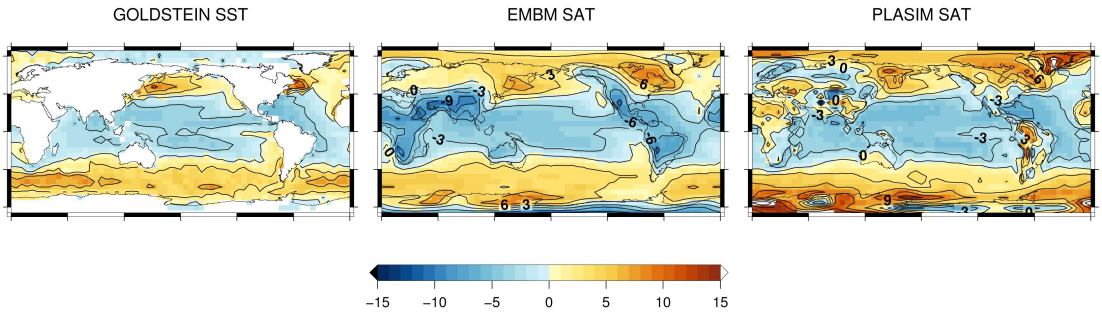


FIGURE A.1: The annual mean 'model minus observation' differences of GOLDSTEIN SST, EMBM SAT and PLASIM SAT. Observation-based reanalysis fields are long-term averages of NOAA World Ocean Atlas (Locarnini et al., 2006) and the 1000 mb and 2m surface air temperature fields from the NCEP-DOE 2 reanalysis (Kanamitsu et al., 2002).

The simulated temperature fields shown here are averaged over a subset of simulations which fulfil the 'modern' boundary forcing conditions, i.e., ICF $\in \{0, 1, 2, 3, 4\}$ and 340 ppm $<$ RFC $<$400 ppm. They correspond to the red points in Figure 3.3. In all three fields, the temperature in the tropics is lower than observation while the opposite is true at high latitudes, except for the ice covered SST in the Arctic. The negative differences over the continents in EMBM have been seen in previous studies using various mesh resolution by Marsh et al. (2011) who concluded that this effect was due to the neglect of orographic effects on air mass trajectories, associated condensation and latent heating. This anomaly is not observed in PLASIM. The warm anomalies around Eastern Eurasia and eastern North America in EMBM was linked to the strong diffusive transport of heat released over the warm western boundary currents. This feature persisted in PLASIM. It's important to note that, unlike the simulations used in Marsh et al. (2011), the subset

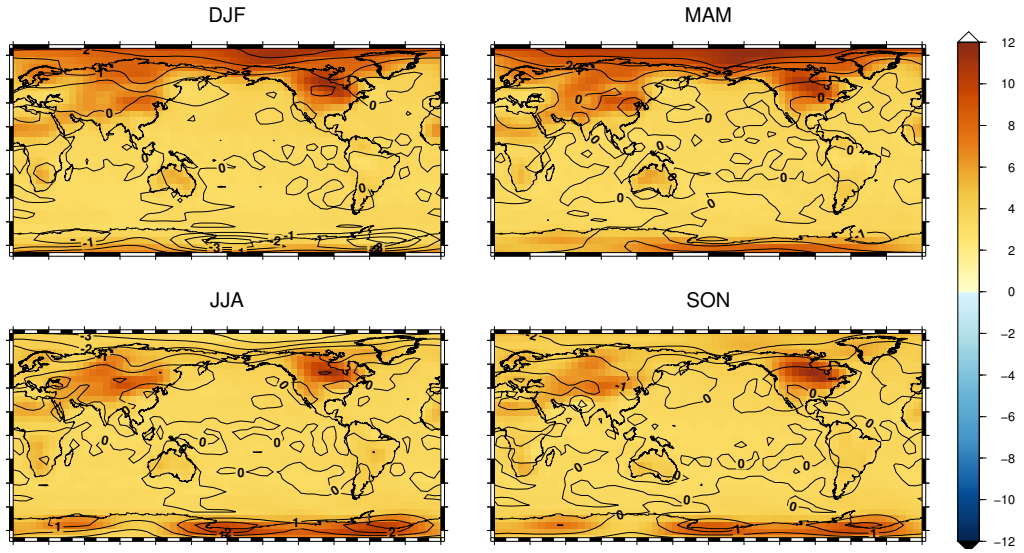
presented here employed 12 varying model parameters and is expected to produce larger anomalies in the annual mean 'model minus observation' plot.

Appendix B

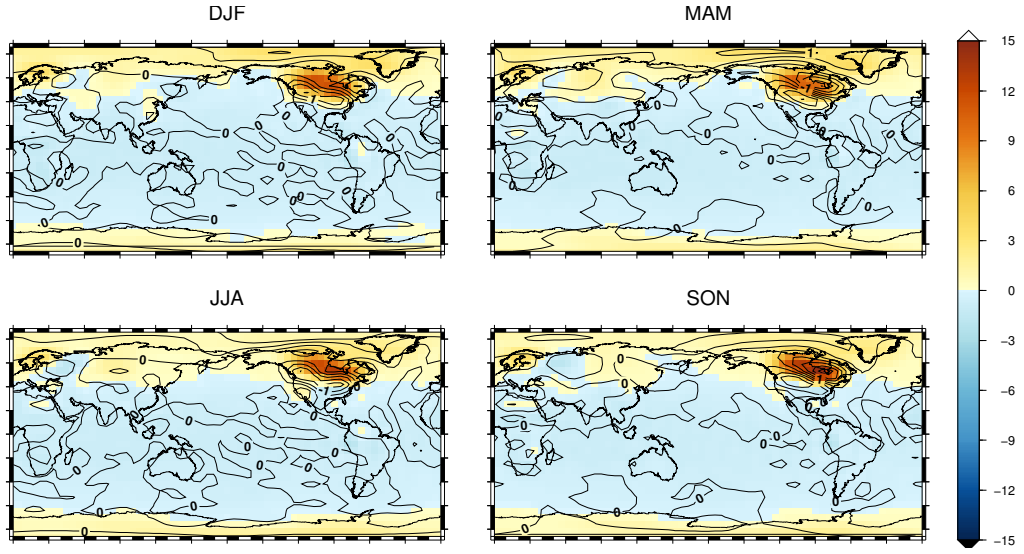
Seasonal EOFs

The seasonal spatial patterns of the first 3 EOFs for SAT and zonal wind component are shown in Figure B.1 and B.2. Abbreviations used are as followed: DJF, December-January-February; MAM, March-April-May; JJA, June-July-August; SON, September-October-December.The annual mean corresponds to these fields can be found in Figure 4.7. The contours on each figure denote the anomaly compared to the annual mean field (the 'seasonal minus annual mean'). The four seasonal EOFs share the same features as the annual mean for all these 3 EOFs. The seasonal cycle is evident in the first SAT EOF with the strongest signals seen in the winter hemisphere. The much larger variation in the Arctic during the Northern Hemisphere winter is due to changes in sea-ice cover. The surface winds of this mode show the same behaviour with stronger winds in the winter and weaker during the summer. However, the seasonal variations in the Northern Hemisphere zonal wind are more significant due to the larger seasonal variation in temperature. The Southern Hemisphere westerlies remain weak throughout the year in the first mode. Unlike the second EOF from the decomposition of SAT alone, this mode has a dominant feature associated with the Laurentide. This is due to the large changes in the surface winds induced by the orography and albedo (Figure B.2). Interestingly, the dipole wind pattern north of the Laurentide in this mode is most prominent during the Northern Hemisphere winter while the temperature anomalies directly over the North American ice sheet are largest in SON and DJF.

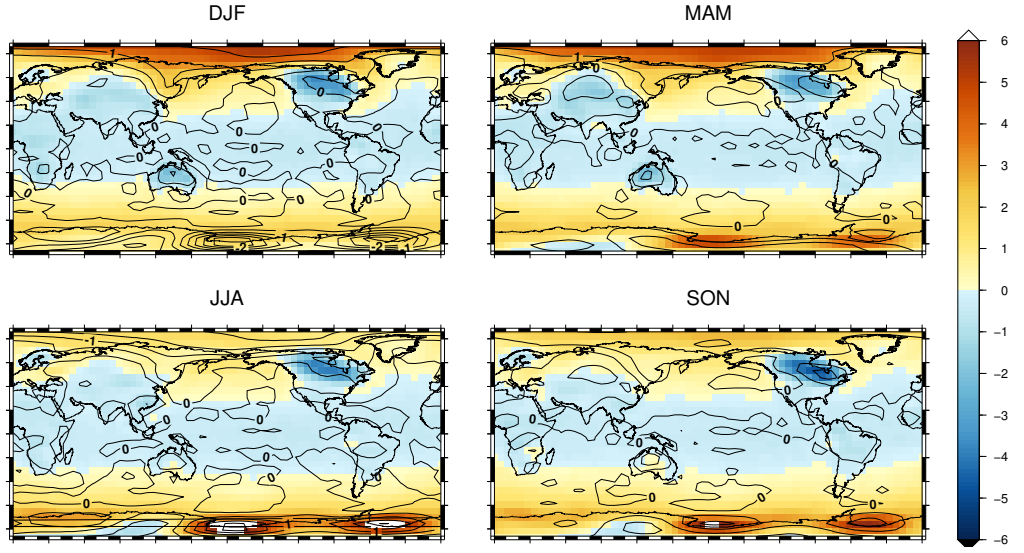
In contrast with the wind EOFs, the humidity modes (Figure B.3) have stronger signals in the summer in general. The wind is stronger in the winter due to the larger equator-to-pole temperature difference, driving stronger surface wind. The surface humidity, on the other hand, increases in the summer as higher temperature enhances evaporation. The EOFs of precipitation (Figure B.4) show similar trend but with less seasonality on average compared to humidity since the largest changes are between the equator and the mid-latitudes (around 40°) in both hemispheres, where seasonal signals are not prominent.



(a) EOF 1

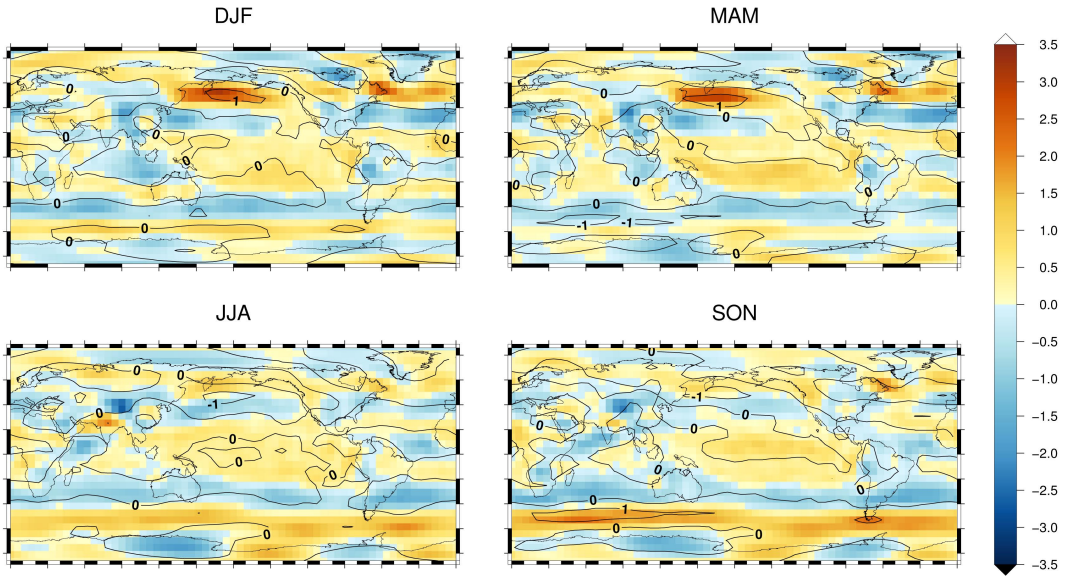


(b) EOF 2

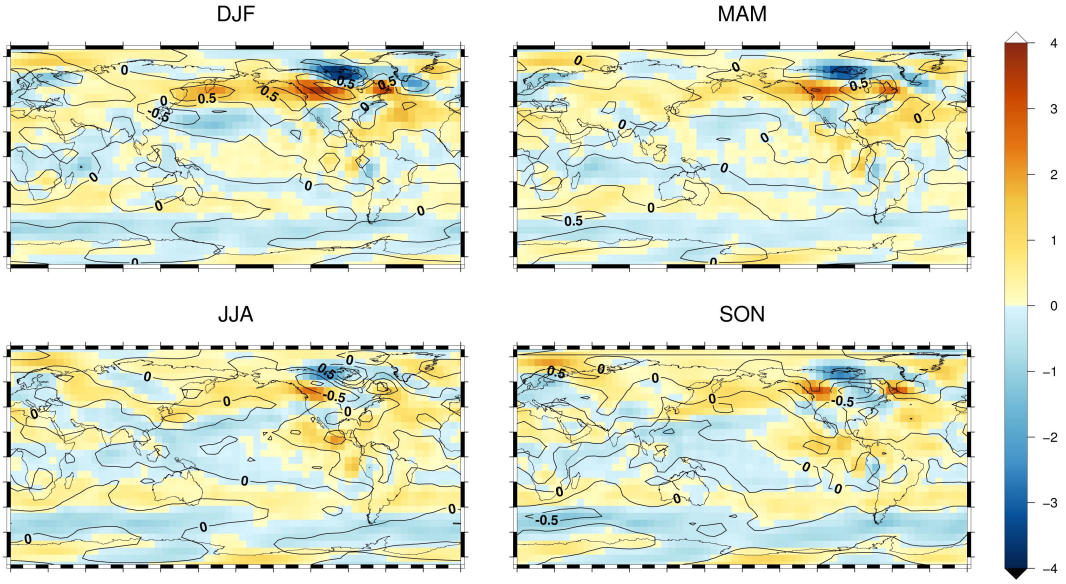


(c) EOF 3

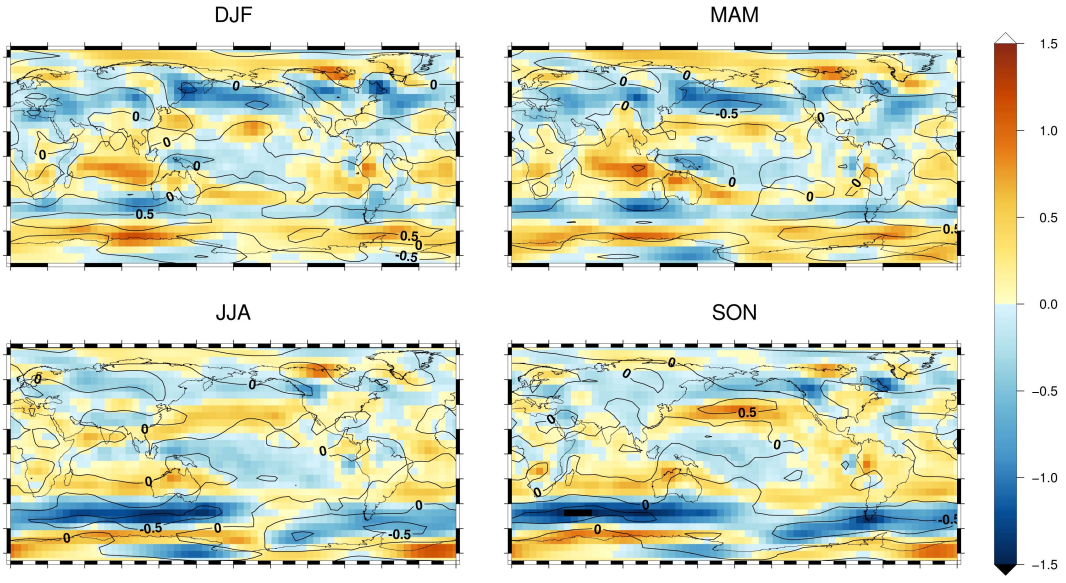
FIGURE B.1: The seasonal cycle in the first 3 EOF of SAT, corresponding to the annual mean EOFs shown in Figure 4.7.



(a) EOF 1

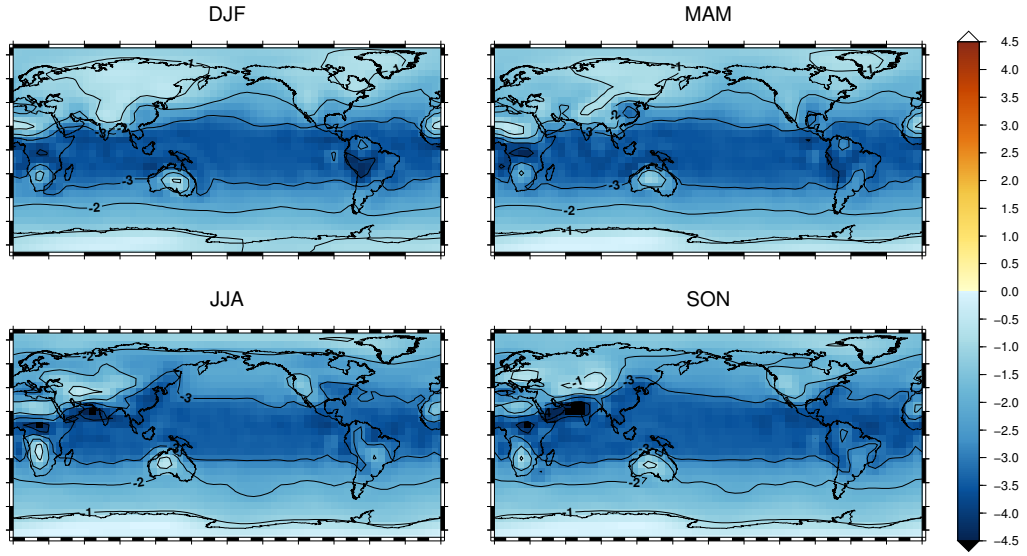


(b) EOF 2

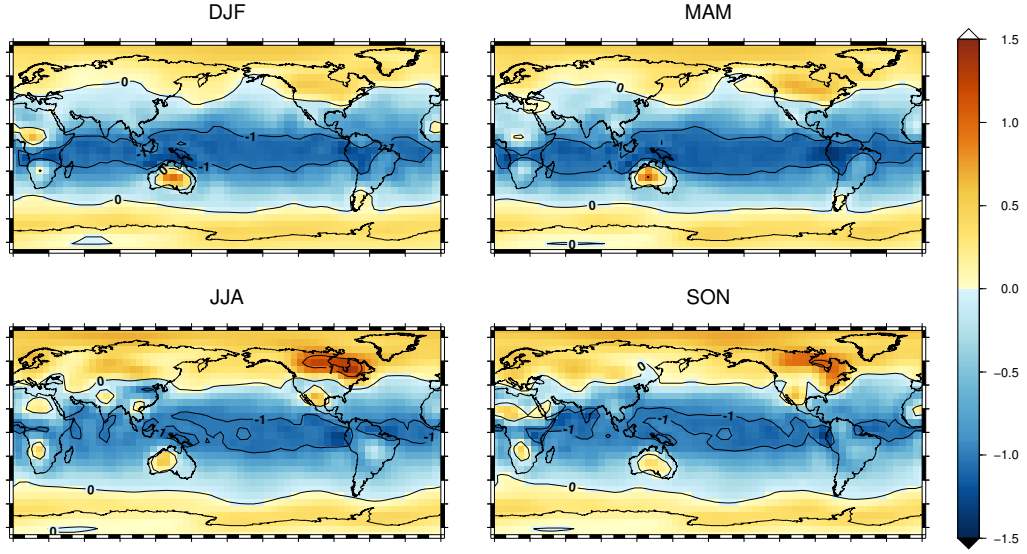


(c) EOF 3

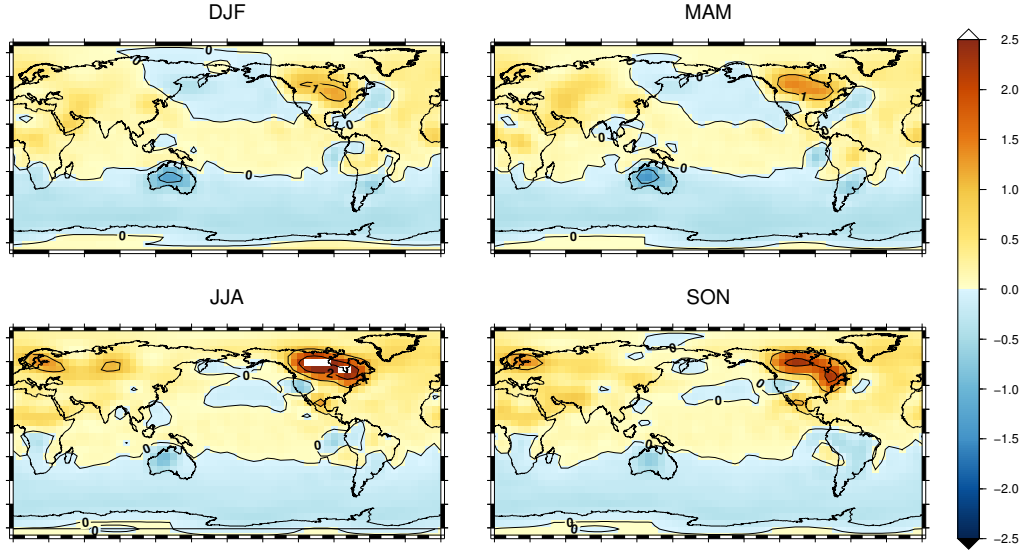
FIGURE B.2: The seasonal cycle in the first 3 EOF of zonal wind, corresponding to the annual mean EOFs shown in Figure 4.7.



(a) EOF 1

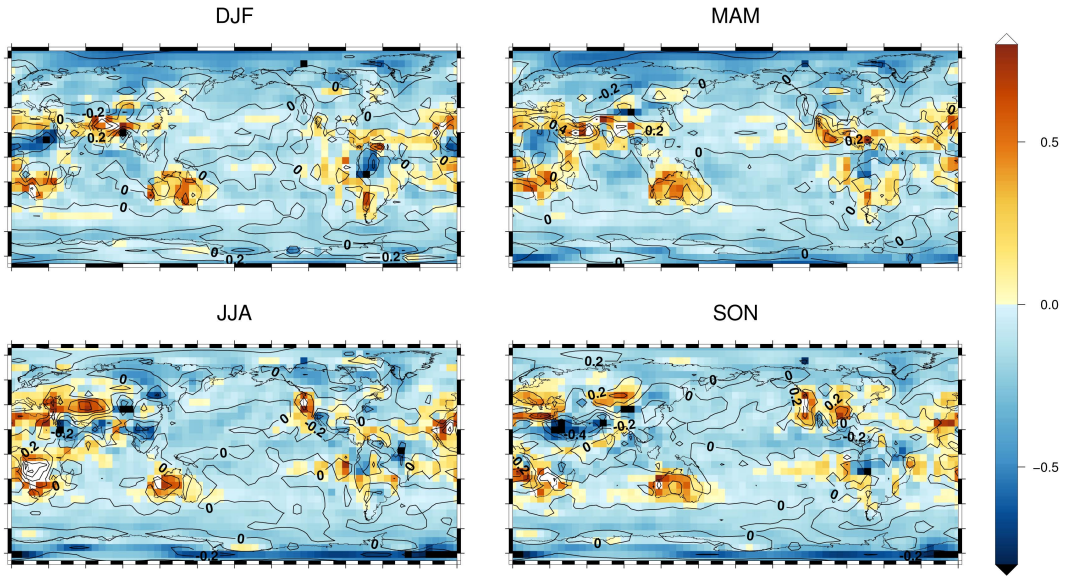


(b) EOF 2

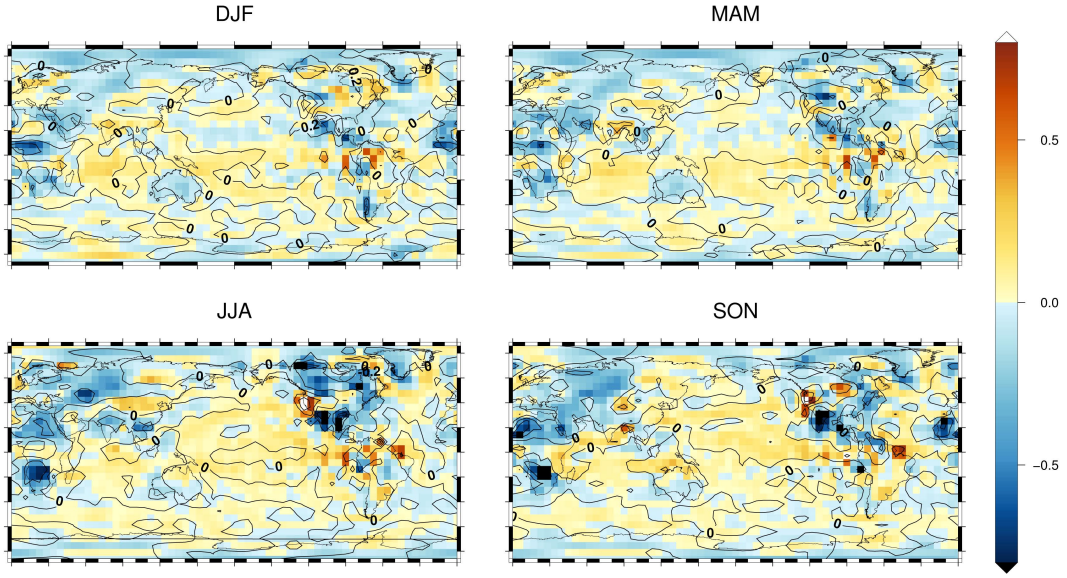


(c) EOF 3

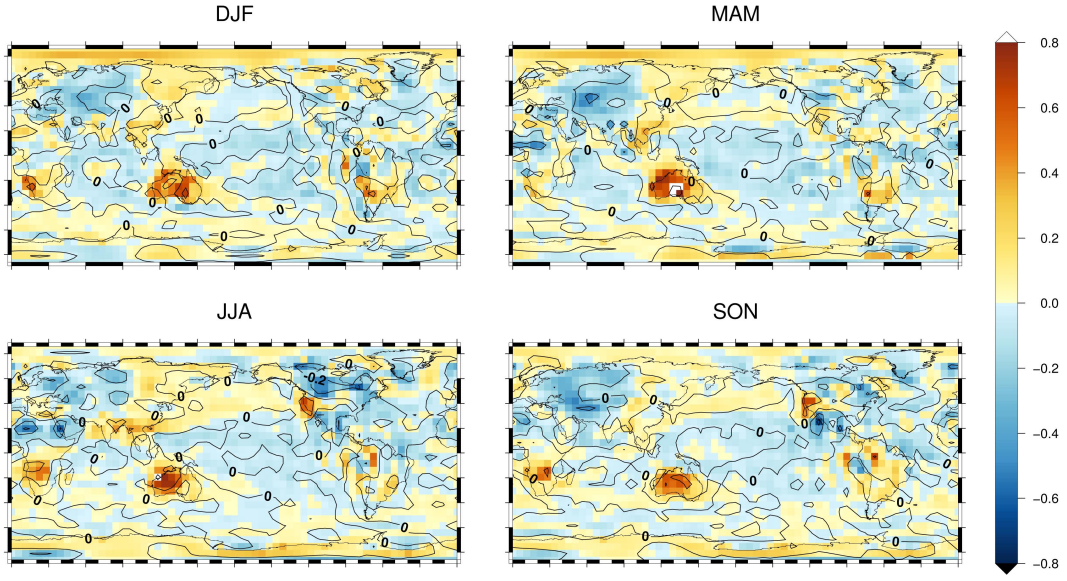
FIGURE B.3: The seasonal cycle in the first 3 EOF of specific humidity, corresponding to the annual mean EOFs shown in Figure 4.7.



(a) EOF 1



(b) EOF 2



(c) EOF 3

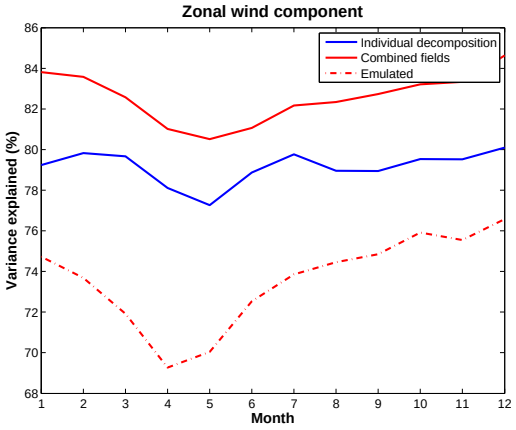
FIGURE B.4: The seasonal cycle in the first 3 EOF of log-precipitation, corresponding to the annual mean EOFs shown in Figure 4.7.

Appendix C

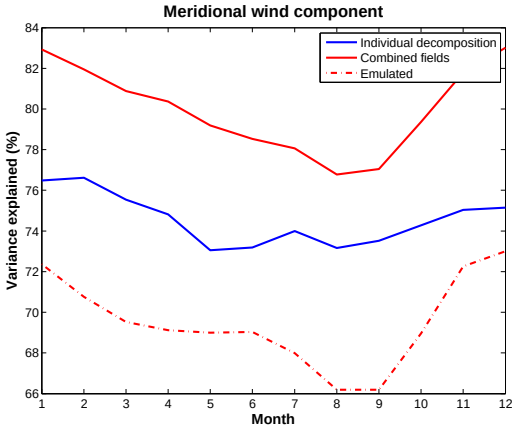
The variance explained for each calendar month

The variance explained by the first emulated 10 EOFs for each month (red dashed line) is compared against the variance explained by the same 10 modes if the emulators were perfect (red solid line). For all variables, the plots also include the ensemble variance captured if each month were to be decomposed independently (blue solid line). For wind components and precipitation rates, the decomposition of the combined spatio-temporal fields lead to a higher amount of variability being captured by the first 10 EOFs compared to when individual decomposition of each month is used. The opposite is true for humidity, however, the differences are relatively small (less than 1% of the total variance).

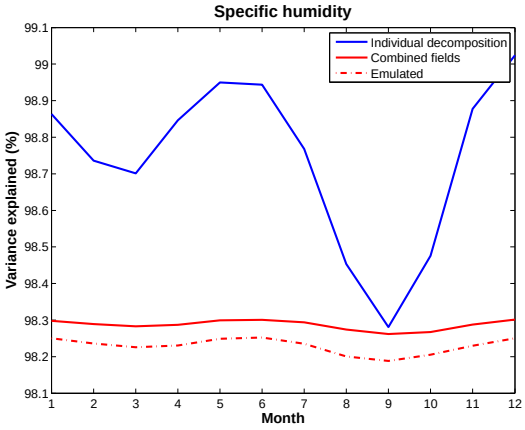
The performance of the emulator for each month differs widely and are not the same for the different fields. Because the variance explained for each month is computed as the percentage of the total variation across the ensemble in that month, a low value here might be a consequence of the truncation error. This is likely to be the case since both the real and emulated variance share the same trend, certain months are less well captured by the dimensional reduction step.



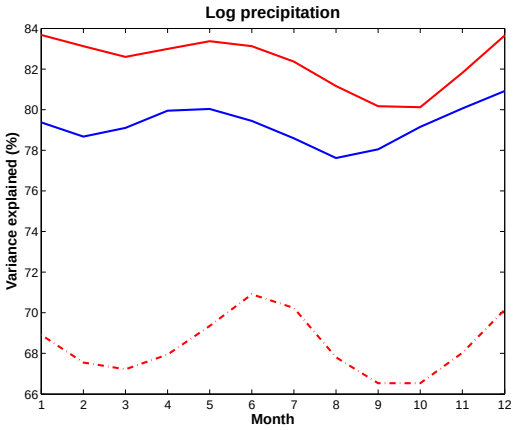
(a) Zonal wind



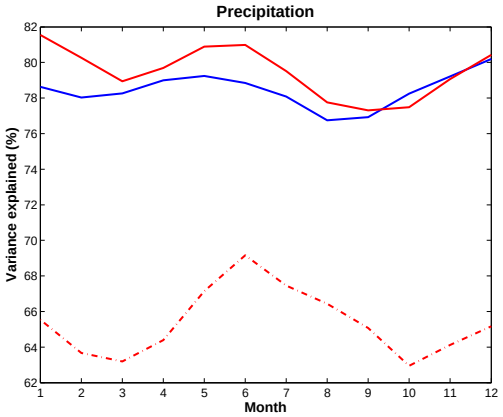
(b) Meridional wind



(c) Specific humidity



(d) log precipitation



(e) Precipitation

FIGURE C.1: The variance explained for each month by the multi-level emulators constructed in Chapter 4.

Bibliography

Abe-Ouchi, A., Saito, F., Kageyama, M., Braconnot, P., Harrison, S. P., Lambeck, K., Otto-Bliesner, B. L., Peltier, W. R., Tarasov, L., Peterschmitt, J. Y., and Takahashi, K.: Ice-sheet configuration in the CMIP5/PMIP3 Last Glacial Maximum experiments, *Geoscientific Model Development*, 8, 3621–3637, 2015.

Arzel, O., Colin de Verdière, A., and England, M. H.: The Role of Oceanic Heat Transport and Wind Stress Forcing in Abrupt Millennial-Scale Climate Transitions, *Journal of Climate*, 23, 2233–2256, 2010.

Barnett, T. P. and Preisendorfer, R.: Origins and levels of monthly and seasonal forecast skill for United states furface air temperature determined by canonical correlation analysis, *Monthly Weather Review*, 115, 1987.

Barnett, T. P., Latif, M., Graham, N., Flugel, M., Pazan, S., and White, W.: ENSO and ENSO-related predictability. Part I: prediction of equatorial Pacific Sea surface temperature with a hybrid coupled ocean-atmosphere model, 1993.

Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis, B. A. S., Gajewski, K., Guiot, J., Harrison-Prentice, T. I., Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppa, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A. E., Williams, J., and Wu, H.: Pollen-based continental climate reconstructions at 6 and 21 ka: A global synthesis, *Climate Dynamics*, 37, 775–802, 2011.

Bastos, L. S. and O'Hagan, A.: Diagnostics for Gaussian Process Emulators, *Technometrics*, 51, 425–438, 2009.

Berger, A.: Long-Term Variations of Daily Insolation and Quaternary Climatic Changes, 1978.

Bond, G., Broecker, W., Johnsen, S., McManus, J., Labeyrie, L., Jouzel, J., and Bonani, G.: Correlations between climate records from North Atlantic sediments and Greendland ice, *Nature*, 365, 143–147, 1993.

Boukouvalas, A. and Cornford, D.: Dimension Reduction for Multivariate Emulation, Tech. Rep. November 2009, Aston University, Birmingham, 2009.

Bounceur, N., Crucifix, M., and Wilkinson, R. D.: Global sensitivity analysis of the climate-vegetation system to astronomical forcing: An emulator-based approach, *Earth System Dynamics*, 6, 205–224, 2015.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, *Nature Climate Change*, 2, 417–424, 2012.

Bretherton, C. S., Smith, C., and Wallace, J. M.: An Intercomparison of Methods for Finding Coupled Patterns in Climate Data, 1992.

Broecker, W. S.: The Glacial World According To Wally, Eldigio Press, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, 3 edn., 2002.

Broecker, W. S. and Denton, G. H.: The role of ocean-atmosphere reorganizations in glacial cycles, *Geochimica et Cosmochimica Acta*, 53, 2465–2501, 1989.

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., and Moyer, E. J.: Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs*, *Journal of Climate*, 27, 1829–1844, 2014.

Challenor, P. G., McNeall, D., and Gattiker, J.: Assessing the probability of rare climate events, in: The Oxford handbook of applied bayesian analysis, edited by O'Hagan, A. and West, M., chap. 16, pp. 403–430, Oxford University Press, New York, 2010.

Chang, E. K. M.: Diabatic and Orographic Forcing of Northern Winter Stationary Waves and Storm Tracks, *Journal of Climate*, 22, 670–688, 2009.

Chen, D., Busalacchi, A. J., and Rothstein, L. M.: The roles of vertical mixing, solar radiation, and wind stress in a model simulation of the sea surface temperature seasonal cycle in the tropical Pacific Ocean, *Journal of Geophysical Research*, 99, 20 345–20 359, 1994.

Chen, S.-C. and Trenberth, K. E.: Orographically Forced Planetary Waves in the Northern Hemisphere Winter: Steady State Model with Wave-Coupled Lower Boundary Formulation, 1988.

Cimatoribus, A. A., Drijfhout, S. S., and Dijkstra, H. A.: A global hybrid coupled model based on atmosphere-SST feedbacks, *Climate Dynamics*, 38, 745–760, 2012.

Clague, J. J.: Quaternary geology of the Canadian Cordillera, 1989.

Clark, P., Dyke, A., Shakun, J., Carlson, A., Clark, J., Wohlfarth, B., Mitrovica, J., Hostetler, S., and McCabe, M.: The Last Glacial Maximum, *Science*, 325, 710–714, doi:doi:10.1126/science.1172873, URL citeulike-article-id:5391015{%}5Cnhttp://dx.doi.org/10.1126/science.1172873, 2009.

Claussen, M., Mysak, L. A., Weaver, A. J., Crucifix, M., Fichefet, T., Loutre, M.-F., Weber, S. L., Alcamo, J., Alexeev, V. A., Berger, A., Calov, R., Ganopolski, A., Goosse, H., Lohmann, G., Lunkeit, F., Mokhov, I. I., Petoukhov, V., Stone, P., and Wang, Z.: Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models, *Climate Dynamics*, 18, 579–586, 2002.

Clement, A. C. and Peterson, L. C.: Mechanisms of abrupt climate change of the last glacial period, *Reviews of Geophysics*, 46, 1–39, 2008.

Conti, S. and O'Hagan, A.: Bayesian emulation of complex multi-output and dynamic computer models, *Journal of Statistical Planning and Inference*, 140, 640–651, 2010.

Cook, R. D. and Nachtsheim, C. J.: A comparison of algorithms for constructing exact D-optimal designs, *Technometrics*, 22, 315–324, 1980.

Cumming, J. and Goldstein, M.: Small Sample Designs for Complex High-Dimensional Models Based on Fast Approximations, *Technometrics*, 51, 377–388, 2008.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdóttir, Á. E., Jouzel, J., and Bond, G. C.: Evidence for general instability of past climate from a 250-kyr ice-core record, *Nature*, 364, 218–220, 1993.

de Finetti, B.: Theory of Probability, John Wiley & Sons, Ltd, New York, 1974.

Delworth, T. L. and Zeng, F.: Simulated impact of altered Southern Hemisphere winds on the Atlantic Meridional Overturning Circulation, *Geophysical Research Letters*, 35, 1–5, 2008.

Diggle, P. J. and Ribeiro Jr, P. J.: Model-based geostatistics, Springer, New York, doi:10.1007/978-0-387-98135-2, URL http://www.springer.com/statistics/statistical+theory+and+methods/book/978-0-387-98134-5?cm{_}mmc=AD-{_}-Enews-{_}-ECS12245{_}V1-{_}-978-0-387-98134-5, 2007.

Eby, M., Weaver, a. J., Alexander, K., Zickfeld, K., Abe-Ouchi, A., Cimatoribus, a. a., Crespin, E., Drijfhout, S. S., Edwards, N. R., Eliseev, a. V., Feulner, G., Fichefet, T., Forest, C. E., Goosse, H., Holden, P. B., Joos, F., Kawamiya, M., Kicklighter, D., Kienert, H., Matsumoto, K., Mokhov, I. I., Monier, E., Olsen, S. M., Pedersen, J. O. P., Perrette, M., Philippon-Berthier, G., Ridgwell, A., Schlosser, A., Schneider von Deimling, T., Shaffer, G., Smith, R. S., Spahni, R., Sokolov, a. P., Steinacher, M., Tachiiri, K., Tokos, K., Yoshimori, M., Zeng, N., and Zhao, F.: Historical and idealized climate model experiments: an intercomparison of Earth system models of intermediate complexity, *Climate of the Past*, 9, 1111–1140, 2013.

Edwards, N. R. and Marsh, R.: Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model, *Climate Dynamics*, 24, 415–433, 2005.

Edwards, N. R., Cameron, D., and Rougier, J.: Precalibrating an intermediate complexity climate model, *Climate Dynamics*, 37, 1469–1482, 2011.

Elliot, M., Labeyrie, L., and Duplessy, J.-c.: Changes in North Atlantic deep-water formation associated with the Dansgaard-Oeschger temperature oscillations (60-10 ka), *Quaternary Science Reviews*, 21, 1153–1165, 2002.

Fanning, A. F. and Weaver, A. J.: An atmospheric energy-moisture balance model: Climatology, interpentadal climate change, and coupling to an ocean general circulation model, *Journal of Geophysical Research*, 101, 15 111, 1996.

Foley, A. M., Holden, P. B., Edwards, N. R., Mercure, J.-F., Salas, P., Pollitt, H., and Chewpreecha, U.: Climate model emulation in an integrated assessment framework: a case study for mitigation policies in the electricity sector, *Earth System Dynamics Discussions*, 6, 2015.

Forrester, A., Sobester, A., and Kean, A.: Engineering design via surrogate modelling, vol. 1, John Wiley & Sons, Ltd, Chichester, 2008.

Forrester, A. I., Sóbester, A., and Keane, A. J.: Multi-fidelity optimization via surrogate modelling, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463, 3251–3269, 2007.

Fraedrich, K.: A suite of user-friendly global climate models: Hysteresis experiments, *The European Physical Journal Plus*, 127, 53, 2012.

Fraedrich, K., Jansen, H., Kirk, E., Luksch, U., and Lunkeit, F.: The Planet Simulator: Towards a user friendly model, *Meteorologische Zeitschrift*, 14, 299–304, 2005.

Ganachaud, A. and Wunsch, C.: Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data., *Nature*, 408, 453–457, 2000.

Gent, P. R.: Effects of Southern Hemisphere Wind Changes on the Meridional Overturning Circulation in Ocean Models, *Annual Review of Marine Science*, pp. 1–34, 2015.

Gildor, H. and Tziperman, E.: Sea-ice switches and abrupt climate change., *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 361, 1935–42; discussion 1942–4, 2003.

Goldstein, M. and Rougier, J.: Reified Bayesian modelling and inference for physical systems, *Journal of Statistical Planning and Inference*, 44, 2009.

Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S., and Jouzel, J.: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores, *Nature*, 366, 552–554, 1993.

Haberkorn, K., Sielmann, F., Lunkeit, F., Kirk, E., Schneidereit, A., and Fraedrich, K.: Planet Simulator Climate, Tech. rep., Meteorologisches Institut, Universität Hamburg, 2009.

Hankin, R. K. S.: Analysis of Computer Code Output, *Journal of Statistical Software*, 14, 1–21, 2005.

Hardoon, D. R., Szedmak, S., and Shawe-taylor, J.: Canonical correlation analysis ; An overview with application to learning methods, *Neural Computation*, 16, 2639–2664, 2004.

Held, I. M.: The Gap between Simulation and Understanding in Climate Modeling, *Bulletin of the American Meteorological Society*, 86, 1609–1614, 2005.

Held, I. M. and Phillipps, P. J.: Sensitivity of the eddy momentum flux to meridional resolution in atmospheric GCMs, 1993.

Held, I. M., Ting, M., and Wang, H.: Northern Winter Stationary Waves : Theory and Modeling, *Journal of Climate*, 15, 2125–2144, 2002.

Hemming, S. R.: Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint, *Reviews of Geophysics*, 42, 2004.

Henderson-Sellers, A. and McGuffie, K.: A Climate Modelling Primer, Wiley, Chichester, 1987.

Herrington, A. R. and Poulsen, C. J.: Terminating the last interglacial: The role of ice sheet-climate feedbacks in a GCM asynchronously coupled to an ice sheet model, *Journal of Climate*, 25, 1871–1882, 2012.

Hevesi, J. A., Flint, A. L., and Istok, J. D.: Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: Isohyetal maps, *Journal of Applied Meteorology*, 31, 677–688, 1992.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M.: Computer Model Calibration Using High-Dimensional Output, *Journal of the American Statistical Association*, 103, 570–583, 2008.

Hofer, D., Raible, C. C., Dehnert, A., and Kuhlemann, J.: The impact of different glacial boundary conditions on atmospheric dynamics and precipitation in the North Atlantic region, *Climate of the Past*, 8, 935–949, 2012.

Holden, P. B. and Edwards, N. R.: Dimensionally reduced emulation of an AOGCM for application to integrated assessment modelling, *Geophysical Research Letters*, 37, 2010.

Holden, P. B., Edwards, N. R., Oliver, K. I. C., Lenton, T. M., and Wilkinson, R. D.: A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, *Climate Dynamics*, 35, 785–806, 2010.

Holden, P. B., Edwards, N. R., Garthwaite, P. H., Fraedrich, K., Lunkeit, F., Kirk, E., Labriet, M., Kanudia, A., and Babonneau, F.: PLASIM-ENTSem v1.0: a spatio-temporal emulator of future climate change for impacts assessment, *Geoscientific Model Development*, 7, 433–451, 2014.

Holden, P. B., Edwards, N. R., Garthwaite, P. H., and Wilkinson, R. D.: Emulation and interpretation of high-dimensional climate model outputs, *Journal of Applied Statistics*, 4763, 2015.

Hoskins, B. and Simmons, A.: A multi-layer spectral model and the semi-implicit method, *Quarterly Journal of the Royal Meteorological Society*, 101, 637–655, 1975.

Huang, R. X.: Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation, *Journal of Physical Oceanography*, 23, 2428, 1993.

Jones, C., Gregory, J., Thorpe, R., Cox, P., Murphy, J., Sexton, D., and Valdes, P.: Systematic optimisation and climate simulation of FAMOUS, a fast version of HadCM3, *Climate Dynamics*, 25, 189–204, 2005.

Josey, S. A., Kent, E. C., and Taylor, P. K.: The Southampton Oceanography Centre (SOC) Ocean - Atmosphere heat, momentum and freshwater flux atlas, *Report of the Southampton Oceanography Centre*, pp. 1–30, 1998.

Justino, F., Timmermann, A., Merkel, U., and Peltier, W. R.: An initial intercomparison of atmospheric and oceanic climatology for the ICE-5G and ICE-4G models of LGM paleotopography, *Journal of Climate*, 19, 3–14, 2006.

Kageyama, M., Valdes, P. J., Ramstein, G., Hewitt, C., and Wyputta, U.: Northern hemisphere storm tracks in present day and last glacial maximum climate simulations: A comparison of the European PMIP models, *Journal of Climate*, 12, 742–760, 1999.

Kageyama, M., Laîné, A., Abe-Ouchi, A., Braconnot, P., Cortijo, E., Crucifix, M., de Vernal, A., Guiot, J., Hewitt, C. D., Kitoh, A., Kucera, M., Marti, O., Ohgaito, R., Otto-Bliesner, B., Peltier, W. R., Rosell-Melé, A., Vettoretti, G., Weber, S. L., and Yu, Y.: Last Glacial Maximum temperatures over the North Atlantic, Europe and western Siberia: a comparison between PMIP models, MARGO sea-surface temperatures and pollen-based reconstructions, *Quaternary Science Reviews*, 25, 2082–2102, 2006.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, 1996.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP-DOE AMIP-II Reanalysis (R-2), *Bulletin of the American Meteorological Society*, 83, 1631–1643, 2002.

Kennedy, M. C. and O'Hagan, A.: Predicting the Output from a Complex Computer Code When Fast Approximations Are Available, *Biometrika*, 87, 1–13, 2000.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63, 425–464, 2001.

Killworth, P. D.: Deep convection in the world ocean, *Rev. Geophys. Space Phys.*, 21, 1–26, 1983.

Kleman, J., Fastook, J., Ebert, K., Nilsson, J., and Caballero, R.: Pre-LGM northern hemisphere ice sheet topography, *Climate of the Past*, 9, 2365–2378, 2013.

Kuhlbrodt, T., Griesel, A., Montoya, M., Levermann, A., Hofmann, M., and Rahmstorf, S.: On the driving processes of the Atlantic meridional overturning circulation, *Reviews of Geophysics*, 45, 2007.

Kutzbach, J. and Guetter, P.: The influence of changing orbital parameters and surface boundary conditions on climate simulations for the past 18000 years, *Journal of the atmospheric sciences*, 43, 1726, 1986.

Labriet, M., Joshi, S. R., Vielle, M., Holden, P. B., Edwards, N. R., Kanudia, A., Loulou, R., and Babonneau, F.: Worldwide impacts of climate change on energy for heating and cooling, *Mitigation and Adaptation Strategies for Global Change*, 20, 1111–1136, 2015.

Laîné, A., Kageyama, M., Salas-Mélia, D., Voldoire, A., Rivière, G., Ramstein, G., Planton, S., Tyteca, S., and Peterschmitt, J. Y.: Northern hemisphere storm tracks during the last glacial maximum in the PMIP2 ocean-atmosphere coupled models: energetic study, seasonal cycle, precipitation, *Climate Dynamics*, 32, 593–614, 2009.

Langen, P. L. and Vinther, B. M.: Response in atmospheric circulation and sources of Greenland precipitation to glacial boundary conditions, *Climate Dynamics*, 32, 1035–1054, 2009.

Lee, L. a., Carslaw, K. S., Pringle, K. J., and Mann, G. W.: Mapping the uncertainty in global CCN using emulation, *Atmospheric Chemistry and Physics*, 12, 9739–9751, 2012.

Lenton, T. M., Williamson, M. S., Edwards, N. R., Marsh, R., Price, a. R., Ridgwell, a. J., Shepherd, J. G., and Cox, S. J.: Millennial timescale carbon cycle and climate change in an efficient Earth system model, *Climate Dynamics*, 26, 687–711, 2006.

Lenton, T. M., Marsh, R., Price, a. R., Lunt, D. J., Aksenov, Y., Annan, J. D., Cooper-Chadwick, T., Cox, S. J., Edwards, N. R., Goswami, S., Hargreaves, J. C., Harris, P. P., Jiao, Z., Livina, V. N., Payne, a. J., Rutt, I. C., Shepherd, J. G., Valdes, P. J., Williams, G., Williamson, M. S., and Yool, A.: Effects of atmospheric dynamics and ocean resolution on bi-stability of the thermohaline circulation examined using the Grid ENabled Integrated Earth system modelling (GENIE) framework, *Climate Dynamics*, 29, 591–613, 2007.

Li, C. and Battisti, D. S.: Reduced Atlantic storminess during last glacial maximum: Evidence from a coupled climate model, *Journal of Climate*, 21, 3561–3579, 2008.

Liakka, J. and Nilsson, J.: The impact of topographically forced stationary waves on local ice-sheet climate, *Journal of Glaciology*, 56, 534–544, 2010.

Liakka, J., Nilsson, J., and Lofverstrom, M.: Interactions between stationary waves and ice sheets: linear versus nonlinear atmospheric response, *Climate Dynamics*, 38, 1249–1262, 2012.

Licciardi, J. M., Clark, P. U., Jenson, J. W., and Macayeal, D. R.: Deglaciation of a soft-bedded Laurentide Ice Sheet, *Quaternary Science Reviews*, 17, 427–448, 1998.

Lindemanm, M. and Oerlemans, J.: Northern hemisphere ice sheets and planetary waves: A strong feedback mechanism, *Journal of Climatology*, 7, 109–117, 1987.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., and Garcia, H. E.: Volume 1 : Temperature, in: World ocean atlas 2005, edited by Levitus, S., vol. 1, chap. Volume 1 :, NOAA Atlas NESDIS 61, U.S. Gov. Printing Office, Washington, D.C., 2006.

Loeppky, J. L., Sacks, J., and Welch, W. J.: Choosing the Sample Size of a Computer Experiment: A Practical Guide, *Technometrics*, 51, 366–376, 2009.

Löfverström, M., Caballero, R., Nilsson, J., and Kleman, J.: Evolution of the large-scale atmospheric circulation in response to changing ice sheets over the last glacial cycle, *Climate of the Past*, 10, 1453–1471, 2014.

Löfverström, M., Caballero, R., and Nilsson, J.: Nonlinear stationary wave reflection as a mechanism for zonalising the Atlantic winter jet at the LGM, *Geophysical Research Letters*, pp. 3329–3342, 2016.

Lohmann, G. and Gerdes, R.: Sea Ice Effects on the Sensitivity of the Thermohaline Circulation*, *Journal of Climate*, 11, 2789–2803, 1998.

Lunt, D. J., Williamson, M. S., Valdes, P. J., Lenton, T. M., and Marsh, R.: Comparing transient, accelerated, and equilibrium simulations of the last 30 000 years with the GENIE-1 model, *Climate of the Past*, 2, 221–235, 2006.

Maniyar, D. M., Cornford, D., and Boukouvalas, A.: Dimensionality Reduction in the Emulator Setting, Tech. Rep. October 2007, Neural Computing research group, Aston University, Birmingham, 2007.

Mardia, K. V. and Marshall, R. J.: Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, 71, 135–146, 1984.

Marsh, R., Yool, A., Lenton, T. M., Gulamali, M. Y., Edwards, N. R., Shepherd, J. G., Krznaric, M., Newhouse, S., and Cox, S. J.: Bistability of the thermohaline circulation identified through comprehensive 2-parameter sweeps of an efficient climate model, *Climate Dynamics*, 23, 761–777, 2004.

Marsh, R., Müller, S. a., Yool, A., and Edwards, N. R.: Incorporation of the C-GOLDSTEIN efficient climate model into the GENIE framework: "eb_go_gs" configurations of GENIE, *Geoscientific Model Development*, 4, 957–992, 2011.

Marsh, R., Sóbester, A., Hart, E. E., Oliver, K. I. C., Edwards, N. R., and Cox, S. J.: An optimally tuned ensemble of the "eb_go_gs" configuration of GENIE: parameter sensitivity and bifurcations in the Atlantic overturning circulation, *Geoscientific Model Development*, 6, 1729–1744, 2013.

Masson-Delmotte, V., Kageyama, M., Braconnot, P., Charbit, S., Krinner, G., Ritz, C., Guilyardi, E., Jouzel, J., Abe-Ouchi, A., Crucifix, M., Gladstone, R. M., Hewitt, C. D., Kitoh, A., LeGrande, A. N., Marti, O., Merkel, U., Motoi, T., Ohgaito, R., Otto-Bliesner, B., Peltier, W. R., Ross, I., Valdes, P. J., Vettoretti, G., Weber, S. L., Wolk, F., and Yu, Y.: Past and future polar amplification of climate change: Climate model intercomparisons and ice-core constraints, *Climate Dynamics*, 26, 513–529, 2006.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González Rouco, J. F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., Timmermann, A., and Rouco, J. F. G.: Information from Paleoclimate Archives, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 383–464, 2013.

Matthews, H. D. and Caldeira, K.: Transient climate-carbon simulations of planetary geoengineering., *Proceedings of the National Academy of Sciences of the United States of America*, 104, 9949–54, 2007.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, 109, 213–241, 2011.

Mercure, J., Pollitt, H., Chewpreecha, U., Salas, P., Foley, A. M., Holden, P. B., and Edwards, N. R.: The dynamics of technology diffusion and the impacts of climate policy instruments in the decarbonisation of the global electricity sector, *Global Environmental Change*, in review, 2014.

Merz, N., Raible, C. C., and Woollings, T.: North Atlantic eddy-driven jet in interglacial and glacial winter climates, *Journal of Climate*, 28, 3977–3997, 2015.

Mickelson, D. M. and Colgan, P. M.: The southern Laurentide Ice Sheet, *Developments in Quaternary Science*, 1, 1–16, 2003.

Monier, E., Scott, J. R., Forest, C. E., and Schlosser, C. A.: An integrated assessment modeling framework for uncertainty studies in global and regional climate change : the MIT IGSM-CAM, *Geoscientific Model Development*, pp. 2063–2085, 2013.

Montoya, M. and Levermann, A.: Surface wind-stress threshold for glacial Atlantic overturning, *Geophysical Research Letters*, 35, 1–5, 2008.

Morris, M. D. and Mitchell, T. J.: Exploratory designs for computational experiments, *Journal of Statistical Planning and Inference*, 43, 381–402, 1995.

Muglia, J. and Schmittner, A.: Glacial Atlantic overturning increased by wind stress in climate models, *Geophysical Research Letters*, 42, 9862–9869, 2015.

O'Hagan, A.: Bayesian Analysis of Computer Code Outputs : A Tutorial, *Reliability Engineering and System Safety*, 91, 1290–1300, 2006.

Oka, A., Hasumi, H., and Suginohara, N.: Stabilization of thermohaline circulation by wind-driven and vertical diffusive salt transport, *Climate Dynamics*, 18, 71–83, 2001.

Oka, A., Hasumi, H., and Abe-Ouchi, A.: The thermal threshold of the Atlantic meridional overturning circulation and its control by wind stress forcing during glacial climate, *Geophysical Research Letters*, 39, 1–6, 2012.

Osborne, J. W. and Costello, A. B.: Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis, *Pan-Pacific Management Review*, 12.2, 131–146, 2009.

Otto-Bliesner, B. L., Brady, E. C., Clauzet, G., Tomas, R., Levis, S., and Kothavala, Z.: Last Glacial Maximum and Holocene Climate in CCSM3, *Journal of Climate*, 19, 2526–2544, 2006.

Pausata, F. S. R., Li, C., Wettstein, J. J., Kageyama, M., and Nisancioglu, K. H.: The key role of topography in altering North Atlantic atmospheric circulation during the last glacial period, *Climate of the Past*, 7, 1089–1101, 2011.

Peltier, W.: Global glacial isostasy and the surface of the ice-age earth: The ICE-5G (VM2) model and GRACE, *Annual Review of Earth and Planetary Sciences*, 32, 111–149, 2004.

Plattner, G. K., Knutti, R., Joos, F., Stocker, T. F., von Bloh, W., Brovkin, V., Cameron, D., Driesschaert, E., Dutkiewicz, S., Eby, M., Edwards, N. R., Fichefet, T., Hargreaves, J. C., Jones, C. D., Loutre, M. F., Matthews, H. D., Mouchet, A., Muller, S. a., Nawrath, S., Price, A., Sokolov, A., Strassmann, K. M., and Weaver, a. J.: Long-Term Climate Commitments Projected with ClimateCarbon Cycle Models, *Journal of Climate*, 21, 2721–2751, 2008.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. a.: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, *Climate Dynamics*, 16, 123–146, 2000.

Pritchard, M. S., Bush, A. B. G., and Marshall, S. J.: Neglecting ice-atmosphere interactions underestimates ice sheet melt in millennial-scale deglaciation simulations, *Geophysical Research Letters*, 35, L01 503, 2008.

Rahmstorf, S.: Ocean circulation and climate during the past 120,000 years., *Nature*, 419, 207–14, 2002.

Rahmstorf, S. and England, M. H.: Influence of Southern Hemisphere Winds on North Atlantic Deep Water Flow, *Journal of Physical Oceanography*, 27, 2040–2054, 1997.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian processes for machine learning, The MIT Press, Cambridge, 2006.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP 8.5-A scenario of comparatively high greenhouse gas emissions, *Climatic Change*, 109, 33–57, 2011.

Richman, M. B.: Rotation of Principal Components, *Journal of Climatology*, 6, 293–335, 1986.

Rivière, G., Laîné, A., Lapeyre, G., Salas-Mélia, D., and Kageyama, M.: Links between Rossby Wave Breaking and the North Atlantic OscillationArctic Oscillation in Present-Day and Last Glacial Maximum Climate Simulations, *Journal of Climate*, 23, 2987–3008, 2010.

Roe, G. H.: Orographic Precipitation, *Annual Review of Earth and Planetary Sciences*, 33, 645–671, 2005.

Roe, G. H. and Lindzen, R. S.: The Mutual Interaction between Continental-Scale Ice Sheets and Atmospheric Stationary Waves, *Journal of Climate*, 14, 1450–1465, 2001.

Romanova, V., Lohmann, G., Grosfeld, K., and Butzin, M.: The relative role of oceanic heat transport and orography on glacial climate, *Quaternary Science Reviews*, 25, 832–845, 2006.

Rossby, C.-G.: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action, 1939.

Rougier, J.: Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations, *Climatic Change*, 81, 247–264, 2007.

Rougier, J.: Efficient Emulators for Multivariate Deterministic Functions, *Journal of Computational and Graphical Statistics*, 17, 827–843, 2008.

Rougier, J., Sexton, D. M. H., Murphy, J. M., and Stainforth, D.: Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments, *Journal of Climate*, 22, 3540–3557, 2009.

Sachs, J. P. and Lehman, S. J.: Subtropical North Atlantic Temperatures 60,000 to 30,000 Years Ago, *Science*, 286, 756–759, 1999.

Saenkon, O. A., Gregory, J. M., Weaver, A. J., and Eby, M.: Distinguishing the Influence of Heat, Freshwater, and Momentum Fluxes on Ocean Circulation and Climate, *Journal of Climate*, 15, 3686–3697, 2002.

Saltelli, A., Chan, K., and Scott, E. M.: Sensitivity Analysis, Wiley, Chichester, 2000.

Schiller, a., Mikolajewicz, U., and Voss, R.: The stability of the North Atlantic thermohaline circulation in a coupled ocean-atmosphere general circulation model, *Climate Dynamics*, 13, 325–347, 1997.

Schmittner, A., Silva, T. A. M., Fraedrich, K., Kirk, E., and Lunkeit, F.: Effects of Mountains and Ice Sheets on Global Ocean Circulation, *Journal of Climate*, 24, 2814–2829, 2011.

Schneider, S. H. and Dickinson, R. E.: Climate modeling, *Reviews of Geophysics*, 12, 447, 1974.

Schulz, M.: On the 1470-year pacing of Dansgaard-Oeschger warm events, *Paleoceanography*, 17, 4–1–4–9, doi:10.1029/2000PA000571, URL http://doi.wiley.com/10.1029/2000PA000571, 2002.

Seager, R. and Battisti, D. S.: Challenges to Our Understanding of the General Circulation : Abrupt Climate Change, in: The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges, edited by Schneider, T. and Sobel, A. S., chap. 12, pp. 331–371, Princeton University Press, 2007.

Severinghaus, J. P., Sowerst, T., and Alleyt, R. B.: Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice, *Nature*, 39, 141–146, 1998.

Sinha, B., Blaker, A. T., Hirschi, J. J. M., Bonham, S., Brand, M., Josey, S., Smith, R. S., and Marotzke, J.: Mountain ranges favour vigorous Atlantic meridional overturning, *Geophysical Research Letters*, 39, 1–7, doi:10.1029/2011GL050485, 2012.

Stirling, C. H., Esat, T. M., Lambeck, K., and McCulloch, M. T.: Timing and duration of the Last Interglacial: Evidence for a restricted interval of widespread coral reef growth, *Earth and Planetary Science Letters*, 160, 745–762, 1998.

Stocker, T. F.: Past and future reorganizations in the climate system, *Quaternary Science Reviews*, 19, 301–319, 2000.

Stouffer, R. J., Yin, J., and Gregory, J.: Investigating the Causes of the Response of the Thermohaline Circulation to Past and Future Climate Changes, *Journal of Climate*, 19, 1365, 2006.

Syu, H.-H., Neelin, D. J., and Gutzler, D.: Seasonal and interannual variability in a Hybrid Coupled GCM, *Journal of Climate*, 8, 2121–2143, 1995.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, 2012.

Thompson, S. L. and Warren, S. G.: Parameterization of outgoing infrared radiation derived from detailed radiative calculations, *Journal of the Atmospheric Sciences*, 39, 2667–2680, 1982.

Tibaldi, S., Palmer, T. N., Branković, Č., and Cubasch, U.: Extended-range predictions with ecmwf models: Influence of horizontal resolution on systematic error and forecast skill, *Quarterly Journal of the Royal Meteorological Society*, 116, 835–866, 1990.

Timmermann, A. and Goosse, H.: Is the wind stress forcing essential for the meridional overturning circulation?, *Geophysical Research Letters*, 31, 1–4, 2004.

Toggweiler, J. R. and Samuels, B.: Effect of drake passage on the global thermohaline circulation, *Deep-Sea Research Part I*, 42, 477–500, 1995.

Tran, G. T., Oliver, K. I. C., Sobester, A., Toal, D. J. J., Holden, P. B., Marsh, R., Challenor, P. G., and Edwards, N. R.: Building a traceable climate model hierarchy with multi-level emulators, *Advances in Statistical Climatology, Meteorology and Oceanography*, 2, 1–21, 2016.

Trenberth, K. E. and Caron, J. M.: Estimates of Meridional Atmosphere and Ocean Heat Transports, *Journal of Climate*, 14, 3433–3443, 2001.

Trenberth, K. E. and Solomon, A.: The global heat balance: heat transports in the atmosphere and ocean, *Climate Dynamics*, 10, 107–134, 1994.

Treut, L., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., Prather, M., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., and Tignor, M.: Historical Overview of Climate Change Science, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., Quin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., vol. Chapter 1, pp. 93–127, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.

Ullman, D. J., Legrande, A. N., Carlson, A. E., Anslow, F. S., and Licciardi, J. M.: Assessing the impact of Laurentide Ice Sheet topography on glacial climate, *Climate of the Past*, 10, 487–507, 2014.

Voelker, A. H.: Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: a database, *Quaternary Science Reviews*, 21, 1185–1212, 2002.

Waelbroeck, C., Paul, A., Kucera, M., Rosell-Melé, A., Weinelt, M., Schneider, R., Mix, a. C., Abelmann, A., Armand, L., Bard, E., Barker, S., Barrows, T. T., Benway, H., Cacho, I., Chen, M.-T., Cortijo, E., Crosta, X., de Vernal, A., Dokken, T., Duprat, J., Elderfield, H., Eynaud, F., Gersonde, R., Hayes, A., Henry, M., Hillaire-Marcel, C., Huang, C.-C., Jansen, E., Juggins, S., Kallel, N., Kiefer, T., Kienast, M., Labeyrie, L., Leclaire, H., Londeix, L., Mangin, S., Matthiessen, J., Marret, F., Meland, M., Morey, a. E., Mulitza, S., Pflaumann, U., Pisias, N. G., Radi, T., Rochon, A., Rohling, E. J., Sbaffi, L., Schäfer-Neth, C., Solignac, S., Spero, H., Tachikawa, K., and Turon, J.-L.: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nature Geoscience*, 2, 127–132, 2009.

Weaver, A. J., Eby, M., Wiebe, E. C., Bitz, C. M., Duffy, P. B., Ewen, T. L., Fanning, A. F., Holland, M. M., Macfadyen, A., Matthews, H. D., Meissner, K. J., Saenko, O., Schmittner, A., Wang, H., and Masakazu, Y.: The UVic earth system climate model : Model description , climatology , and applications to past , present and future climates, *Atmosphere-Ocean*, 39, 361–428, 2001.

Weber, S. L.: The utility of earth system models of intermediate complexity (EMICs), *Wiley Interdisciplinary Reviews: Climate Change*, 1, 243–252, 2010.

Wilkinson, R. D.: Bayesian Calibration of Expensive Multivariate Computer Experiments, in: Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainity, edited by Biegler, L., Biros, G., Ghattas, O., Heinkenschloss, M., Keyes, D., Mallick, B., Marzouk, Y., Tenorio, L., Waanders, B. v. B., and Willcox, K., chap. 10, John Wiley & Sons, Ltd, Chichester, 2010.

Williamson, D., Goldstein, M., and Blaker, A.: Fast linked analyses for scenario-based hierarchies, *Journal of the Royal Statistical Society. Appl. Statist.*, 61, 665–691, 2012.

Williamson, M., Lenton, T., Shepherd, J., and Edwards, N.: An efficient numerical terrestrial scheme (ENTS) for Earth system modelling, *Ecological Modelling*, 198, 362–374, 2006.

Wolfe, C. L. and Cessi, P.: What Sets the Strength of the Middepth Stratification and Overturning Circulation in Eddying Ocean Models?, *Journal of Physical Oceanography*, 40, 1520–1538, 2010.

Wunsch, C.: Abrupt climate change: An alternative view, *Quaternary Research*, 65, 191–203, 2006.

Wunsch, C. and Ferrari, R.: Vertical Mixing, Energy, and the General Circulation of the Oceans, *Annual Review of Fluid Mechanics*, 36, 281–314, 2004.

Yang, H., Wang, K., Dai, H., Wang, Y., and Li, Q.: Wind effect on the Atlantic meridional overturning circulation via sea ice and vertical diffusion, *Climate Dynamics*, 46, 3387–3403, 2016.

Yu, E., Francois, R., and Bacon, M. P.: Similar rates of modern and last-glacial ocean thermohaline circulation inferred from radiochemical data, *Nature*, 379, 689–694, 1996.

Zhang, X., Lohmann, G., Knorr, G., and Purcell, C.: Abrupt glacial climate shifts controlled by ice sheet changes, *Nature*, 512, 290–294, 2014.