

**Annual minimum Arctic sea ice extent projections and  
uncertainty quantifications under different greenhouse gas  
emission scenarios**

**Candidate Number:** HMSB1

**Supervisor:** Professor Richard Chandler



University College London  
Department of Mathematics

Course code: STAT0035

Word count: 13986 words

Date: April 21, 2024

## **ACKNOWLEDGEMENT**

The author acknowledge the World Climate Research Programme which coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, and the Earth System Grid Federation for archiving the data and providing access.

The author would also like to take this opportunity to express their deepest gratitude to Professor Richard Chandler. The accomplishments of this project would not have been possible without his numerous supports, for example transforming the CMIP6 simulation files into organised time series files with maps and separate description documents, which is a laborious and painful process. His professional guidance and patient assistance have made this dissertation experience a remarkably fruitful, inspiring, and unforgettable experience throughout the author's undergraduate studies.

## Abstract

The Arctic sea ice serves significant role in global climate system, and annual minimum sea ice extent is an indicator of climate change. The loss of Arctic sea ice has been seen over decades, and can lead to considerable societal, economic, and environmental consequences. Different future development pathways have been introduced in climate research. In our report, we aim to project the annual minimum Arctic sea ice extent to the end of the 21<sup>st</sup> century under four different Shared Socio-economic Pathways (SSPs), based on the historical satellite measurements of Arctic sea ice and the simulation outputs from multiple climate models. With an EBM-inspired state-space formulation of the ensemble of climate model outputs, we utilise Kalman filter and Kalman smoother and Bayesian methods to estimate the parameters in the model. Uncertainties could be quantified by posterior sampling, and the trend and prediction intervals could be estimated. From the results, it is suggested that the annual minimum Arctic sea ice extent will fall below 2 million  $km^2$  almost surely by 2080 under high greenhouse gas emissions scenarios. The annual minimum Arctic sea ice may remain stable about the current level under a very sustainable global development. The methodology in this report is new to sea ice projection problems, and it improves the uncertainty quantification in the current approaches, although some future work can be done for improvements. The Arctic sea ice projections and associated uncertainties could be helpful in environmental policy making.

*Keywords:* multi-model ensemble; time series; state-space models; uncertainty quantification; Arctic sea ice projection.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Background . . . . .	5
1.2	Research Question . . . . .	6
1.3	Literature Review . . . . .	6
1.4	Structure of This Report . . . . .	8
<b>2</b>	<b>Data</b>	<b>8</b>
2.1	Satellite Measurements of Historical Arctic Sea Ice . . . . .	9
2.1.1	Data sources and processing . . . . .	9
2.1.2	Exploratory analysis and visualisation . . . . .	10
2.2	Ensemble of CMIP6 Climate Model Simulations . . . . .	11
2.2.1	Data sources and processing . . . . .	12
2.2.2	Exploratory analysis and visualisation . . . . .	13
2.3	Effective Radiative Forcings . . . . .	14
2.4	Log-transformation to Sea Ice Data . . . . .	16
2.5	Summary of Chapter 2 . . . . .	16
<b>3</b>	<b>Methodology</b>	<b>17</b>
3.1	Modelling the Real-world Annual Sea Ice Minima . . . . .	17
3.2	Kalman Filter and Kalman Smoother . . . . .	19
3.3	Parameter Estimation . . . . .	21
3.4	Trend Estimation and Prediction . . . . .	23
3.5	Multi-model Ensemble . . . . .	23
3.6	Uncertainty . . . . .	27
3.7	Summary of Chapter 3 . . . . .	27
<b>4</b>	<b>Application to Annual Minimum Arctic Sea Ice Extent</b>	<b>28</b>
4.1	Choices of Prior . . . . .	28
4.1.1	Prior choices for descriptors . . . . .	28
4.1.2	Initial settings for the state vector . . . . .	29
4.2	Model Results . . . . .	30
4.2.1	Parameter estimates and trend projection . . . . .	30
4.2.2	Uncertainty quantification . . . . .	33
4.3	Summary of Chapter 4 . . . . .	35
<b>5</b>	<b>Conclusion and Discussion</b>	<b>36</b>
5.1	Summary . . . . .	36

5.2 Discussion and Future Work . . . . .	36
--	----

## Appendices

<b>A1Comments on CMIP6 model runs and decisions</b>	<b>38</b>
---	-----------

<b>A2Diagnostic plots</b>	<b>39</b>
---------------------------	-----------

# 1 Introduction

## 1.1 Background

Sea ice in both the Arctic and Antarctic serves as a significant gauge of the global climate change and polar amplification effects. Arctic sea ice is not only an irreplaceable habitat to many endemic species, but also a crucial part in regulating the worldwide climate system by affecting polar atmospheric and oceanic circulation. It is suggested that the loss of sea ice and snow cover increases the likelihood of anomalous summer and winter weathers, for example the increase of summer heat extremes and droughts in North America and Europe (Tang et al., 2014; Francis and Skific, 2015). Moreover, the loss of sea ice also affects biodiversity in Arctic regions, with impacts such as the deprivation of available habitats and increasing prevalence of disease and infections (Kovacs et al., 2011). There is therefore widespread interest in the extent, variability, and projections of Arctic sea ice.

Over the past decades, there has been a remarkable decline in minimal Arctic sea ice extent, at approximately 9.4–13.6% decrease per decade (Barry and Gan, 2022). The lowest annual minimum sea ice extent is recorded in 2012, at approximately 3.61 million  $km^2$  (NSIDC, 2012). The inherent variability in summer ice extent is anticipated to escalate in the 21st century, due to the thinning of ice cover (Holland et al., 2008). Some proposed that under moderate future greenhouse gas (GHG) emissions, an ice-free Arctic summer is likely to be seen before the end of this century (Boé et al., 2009).

The reduction of Arctic sea ice can be largely attributed to both the natural fluctuations in the climate system, and the warming that results from rising levels of GHG emissions, which mainly consist of CO<sub>2</sub>, Methane(CH<sub>4</sub>), and Nitrous oxide(N<sub>2</sub>O) (Stroeve et al., 2012; Barry and Gan, 2022). It is generally acknowledged that the increase in GHG concentrations is predominantly attributed to human activities (IPCC, 2023).

In climate research, the Shared Socio-Economic Pathways (SSPs) explore plausible future socio-economic conditions and their impacts on GHG emissions and climate. Under different SSP scenarios, future GHG emission projections are created and widely applied to climate studies. These SSP-based scenarios are referred to as “SSPxyz”, where “x” refers to the basic storyline of future development, and “yz” represents the final net radiative forcing (in  $Wm^{-2}$ ) by the year 2100 in that storyline (O’Neill et al., 2014). The high and very high GHGs emission scenarios are from SSP370 and SSP585, which in turn correspond to “regional-rivalry” and “fossil-fuelled” development pathways. They project that the CO<sub>2</sub> emissions will roughly be the double of the current levels by 2100 and 2050, respectively. In “moderate” development (SSP245), CO<sub>2</sub> emissions are expected to remain approximately steady until 2150s. SSP119 and SSP126 are “sustainable” pathways, and CO<sub>2</sub> emissions would reach net zero by around 2050 and 2070, respectively, followed by varying levels of net negative CO<sub>2</sub> emissions (IPCC,

2023). To conclude, the SSPs 119, 126, 245, 370, and 585 correspond to “very low”, “low”, “intermediate”, “high” and “very high” GHG emission scenarios (O’Neill et al., 2014).

To understand how the climate system may respond to alternative future socio-economic development pathways, the Coupled Model Intercomparison Project (CMIP), currently in its sixth phase (CMIP6), integrates extensive climate models over the world that perform common sets of experiments. CMIP6 provides the most recent climate simulations up to the end of the 21<sup>st</sup> century under different future emission scenarios (Eyring et al., 2016). These simulations enable us to investigate the implications of alternative socio-economic scenarios for future changes in Arctic sea ice extent. However, projections provided by CMIP6 are from various climate models, and can vary substantially between models, even under the same emission scenario. Also, the simulations in historical periods reveal a variety of biases and discrepancies comparing with the satellite measured sea ice extent. This complicates the direct interpretation of the projections, and leads to non-negligible uncertainties when utilizing the CMIP6 simulations.

## 1.2 Research Question

The aim of this paper is to explore the future projection of annual minimum Arctic sea ice area under four different greenhouse gas emission scenarios, based on satellite measured monthly time series of sea ice area from January 1979 to December 2022, and multiple global climate models’ simulations in CMIP6. Then we aim to quantify the associated uncertainties.

The main research question is:

- What can we say about the future measured Arctic sea ice under different greenhouse gas emission scenarios, and can we characterise the associated uncertainty?

The future “measured” Arctic sea ice in the question refers to the satellite measurements that are expected to be obtained in the future. This “measured” sea ice is taken as a representation of the true real-world quantity of interest in our investigation.

## 1.3 Literature Review

In this section, we will first evaluate some studies on Arctic sea ice projections, then discuss some literature about state-of-art climate projection and uncertainty quantification techniques. For sea ice projection, one of the most straightforward methods is to apply regression on sea ice data and other relevant variables. For example, Ionita et al. (2019) propose a multiple linear regression model to predict the September Arctic sea ice with explanatory variables including historical data at certain lags, oceanic and atmospheric variables such as surface wind and long-wave radiations. Although their approach demonstrates a high predictive skill, the reliable prediction window is limited within 4 months ahead, and the prediction uncertainties are not characterised. Another approach using linear regression addresses the problem of uncertainty by

supplementing the satellite measurement with 18 climate model outputs from the third phase of CMIP project (CMIP3) (Boé et al., 2009). By computing a regression line between the current trend of satellite measured data and the mean percentages of remaining sea ice in every climate model during each 20-year segment in the future, the authors constrain future changes by current observations. From the regression results, estimates of sea ice evolution and confidence intervals are generated (Boé et al., 2009). Although linear regression can be reasonable as the future change of remaining September sea ice is correlated with the current trend, the linearity is not justified, and extreme climate model outputs might imperil the linear fit performance. Some other attempts were made to integrate multiple climate model outputs for Arctic sea ice projection. For example, Stroeve et al. (2012) applied a model selection method on a large ensemble of climate models from CMIP3 and CMIP5, under a stabilised GHG emission scenario. By comparing the model simulations and observations in historical period, the models whose majority distributions fall outside the observation range are removed, and the mean projection from the remaining models is used for subsequent analysis. However, the model rejection criterion lacks proper justification, and this approach does not make full use of all the available information. Also, a large uncertainty related to future GHG emissions is pointed out in the paper, which could be improved by integrating the GHG forcings into the analysis.

We discussed some heuristic approaches in Arctic sea ice projection and their imperfections, particularly in quantifying uncertainties and combining observations with multiple climate models. Also, there has been no developed investigation on the future trends of Arctic sea ice under different GHG emission scenarios. Below we will discuss some more principled statistical multi-model methods in climate projections, which have not previously been applied to Arctic sea ice projections but are inspiring for addressing the current problems.

A Bayesian framework is suggested to be particularly powerful when multiple models are involved. For example, Buser et al. (2009) developed a Bayesian framework which assumes that the observations and climate model data follow parametrised normal distributions with different means and variances settings. By combining the likelihood function with prior distributions of parameters, posterior distributions and subsequent information of interest can be derived. This could be extended to a wide range of climate variables of interest.

The climate simulations are not daily predictions, but instead attempt to reproduce the statistical properties of the climate system over decades. Chandler et al. (2023) suggested that the statistical properties of both the climate models outputs and the observations can be characterised by the parameters in a statistical model, which are called “descriptors”. They also discussed the uncertainty quantification using posterior samplings that could account for the uncertainties in the “descriptors” estimates. A “shared discrepancy” structure between the properties of the climate models and that of the real climate is introduced, which suggests that all climate model individuals share a common discrepancy between themselves and the

real-world system (Chandler, 2013). This is explainable because all climate models originate from common ancestral models, and are built under shared limited understanding of the Earth system. This inspires later studies, for example in Sansom et al. (2021), by incorporating this “shared discrepancy” structure, a shared inadequacy between climate models and the real-world is removed in model uncertainty calculation.

A further improvement to these methodologies is to account for the forcings from GHG emissions, which could limit the trend behaviour in climate projections (Chandler et al., 2023). To achieve this, some more complex model structures have been proposed, such as introducing energy balance models (EBMs) into the current frameworks. Qasmi and Ribes (2022) embed an EBM model to deduce the temperature response to natural forcings. Chandler et al. (2023) incorporate the EBM with the state-space models to capture and constrain the non-linear trends in temperature and precipitation projections. This is particularly inspiring for our projection of annual minimum Arctic sea ice extent, as a connection between the trend of sea ice extent and the forcings that represent GHG emission scenarios can be constructed in a similar way. Overall, several researchers discussed the problem of Arctic sea ice projection, and approaches analysing the observations and multiple climate models were proposed (Boé et al., 2009; Stroeve et al., 2012; Ionita et al., 2019). Some statistical methods developed by Buser et al. (2009), Chandler (2013), Sansom et al. (2021), and Chandler et al. (2023) in various climate projection studies provide insights into Arctic sea ice projection and uncertainty quantification.

## 1.4 Structure of This Report

In Chapter 1, we discussed the background and aim of this paper, followed by some reviews on relevant literature. In subsequent chapters, we will first explore the datasets used in this report in Chapter 2, and then we will introduce the methodology in Chapter 3. Chapter 4 discusses the application of these methods to sea ice projection, addressing prior choices and model results. Finally, we will conclude with summaries and discussions in Chapter 5.

## 2 Data

This report utilises two sources of Arctic sea ice area data, one is the satellite measurements of Arctic sea ice data, and the other one is the ensemble of climate model projections to the end of 21<sup>st</sup> century, under different greenhouse gas emission scenarios. The projections are derived from CMIP6 simulations under four hard socio-economic pathways, as introduced in Section 1.1. In addition, the effective radiative forcings corresponding to these GHG emission scenarios are also introduced. In this chapter, data sources and processing will be discussed in detail, followed by exploratory analyses and visualisation.

## 2.1 Satellite Measurements of Historical Arctic Sea Ice

### 2.1.1 Data sources and processing

The NASA National Snow and Ice Data Center (NSIDC) records time series data of Arctic sea ice concentration at monthly and daily resolutions from 26th October 1978 to early 2023, with the help of microwave radiometers (DiGirolamo et al., 2022).

In this report, both monthly and daily data have been downloaded for the period from 1st December 1978 to 31st December 2022. Throughout this period, there is an circular sector around the north pole that is not measured due to satellite trajectory inclination, which is referred as the “pole hole”. However, as technology improves, sensors were updated to reduce the area of this pole hole, resulting in differences in pole hole sizes as shown in Table 1. Hence, to avoid jumps and inconsistency in scales of sea ice areas for analysis, we apply the  $84.5^\circ$  N pole hole mask to all data files, which corresponds to a circular region of approximately 1.19 million  $km^2$  centered at the north pole. In the subsequent analysis, the sea ice area refer to the area outside this  $84.5^\circ$  N pole hole.

Dates Used	Pole hole area	Latitude
11/1978 -06/1987	1.19 million $km^2$	$84.5^\circ$ N
07/1987-12/2007	0.31 million $km^2$	$87.2^\circ$ N
01/2008-present	0.029 million $km^2$	$89.18^\circ$ N

Table 1: Pole hole sizes and dates used (DiGirolamo et al., 2022).

Sea ice concentrations, which is the fraction of ocean area covered by sea ice, takes integer values between 0 and 250, representing sea ice fractions from 0% to 100% within a grid cell, while lands and pole holes are marked as NA. Each measurement is recorded at a nominal grid cell size of  $25 \times 25 km^2$ , although the actual areas of some cells are not exactly this size, due to polar stereographic projection<sup>1</sup>. This is a projection method to minimise distortion at margins of ice zones by specifying projection planes tangent to Earth surface at  $70^\circ$  N. This projection method contributes to negligible distortion of the grids around the margins of ice zones, but it also introduces distortions of grid cell areas closer to the north pole.

Calculating the estimates of daily total sea ice area in Arctic region is not trivial. In each grid, the product of sea ice concentration (as a proportion) and the size of that grid cell gives the area covered by sea ice. Hence, taking the sum of this product over all the grid cells with available sea ice concentration data gives the estimated total area of sea ice in the Arctic region. To estimate the *true* area, we need to transform the areas of distorted grid cells back to their

<sup>1</sup>More details here: <https://nsidc.org/data/user-resources/help-center/guide-nsidcs-polar-stereographic-projection>

actual areas, which are caused by polar stereographic projection, by applying the associated transformation method provided in its corresponding polar projection coordinate system<sup>2</sup>, as guided by NSIDC.

As described in Section 1.2, the subsequent analysis focuses on a time series of annual minimum sea ice areas. These are obtained by first computing the monthly average sea ice area from the daily measurements, followed by finding the annual minimum of the monthly averaged sea ice area, after removing the months with recorded measurements fewer than 10 days.

### 2.1.2 Exploratory analysis and visualisation

Figure 1 displays the NSIDC satellite measured sea ice extent in the Arctic region in September 1979 and September 2020. The lightblue and white colors represent the area covered by sea ice, and the ocean is colored in darkblue. The green regions correspond to the lands, in which the one on the right is the Eurasian continent, and the bottom-left darkgrey area corresponds to north America. The black sector in the middle is the  $84.5^{\circ}$  N pole hole (see Section 2.1.1). For the sea ice, the whiter the color, the higher the sea ice concentration is in that region. As shown in the two maps, the annual minimum Arctic sea ice area in 2022 is almost half of that in 1979, especially in regions such as the Beaufort Sea and the East Siberian Sea.

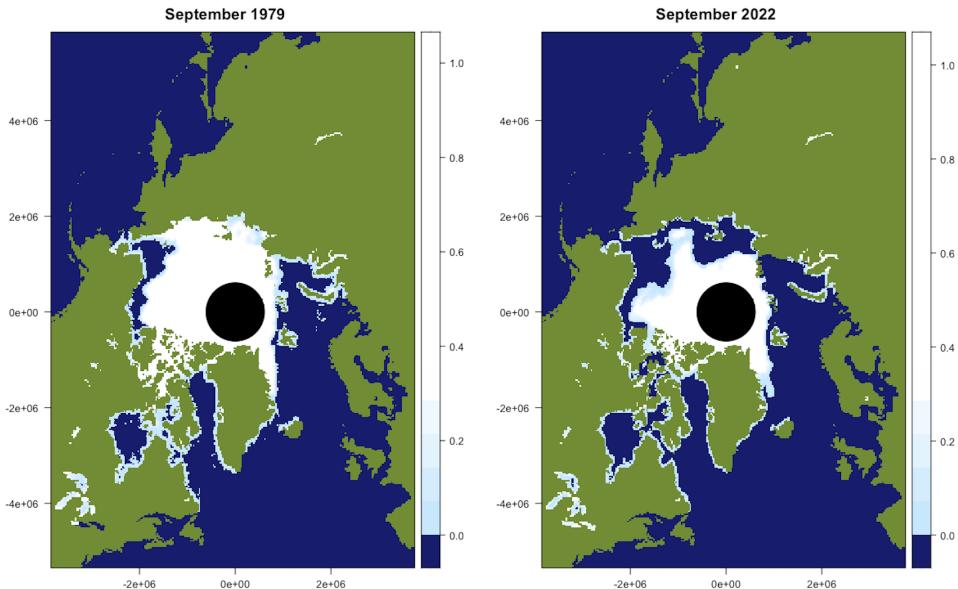


Figure 1: Distribution of Arctic sea ice fraction in September 1979(left) and September 2022(right) in polar stereographic projection.

In addition, Figure 2 presents the time series plot of the historical annual minimum Arctic sea ice (excluding pole hole region) from the satellite measurement data. The satellite measurement of Arctic sea ice area outside the pole hole declined from approximately 5.016 million  $km^2$  in

---

<sup>2</sup><https://epsg.io/3413>

1980 to around 2.5 million  $km^2$  in 2007, and then dropped to historical minimum at 1.66 million  $km^2$  in 2012. This sharp decline of annual minimum sea ice extent agrees with various remote sensing reports and analyses on sea ice observations from other data sources (Parkinson, 2022; Serreze and Stroeve, 2015; Stroeve et al., 2014). After 2012, the annual minimum sea ice seem to have rebounded and finally ended at around 2.5 million  $km^2$  in 2022. The mean of the measured area is approximately 3.56 million  $km^2$ , and a robustly fitted linear trend (dashed green line) is provided as additional support of the overall declining trend.

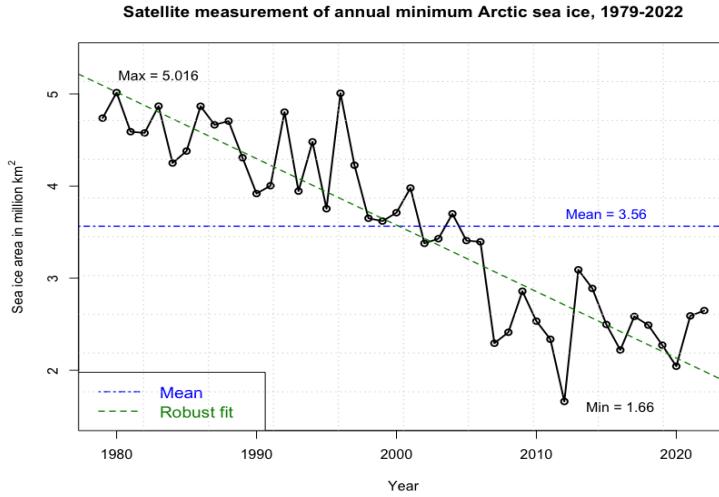


Figure 2: Satellite measured annual minimum Arctic sea ice area excluding pole hole.

## 2.2 Ensemble of CMIP6 Climate Model Simulations

Simulations from global climate models considered in this report are produced by different climate models from over 20 modelling centres around the world who participate in the CMIP6. These models are referred to as “simulators” in this report, and a collection of outputs from different simulators is the multi-model “ensemble”.

As described in Section 1.1, CMIP6 focuses on comparing and analyzing climate model simulations to assist the understanding of climate change. The Scenario Model Intercomparison Project (ScenarioMIP) is one of the main undertaking in CMIP6. This project aims to facilitate climate projections and uncertainty quantification by utilizing ensembles of outputs under various scenarios of future emissions and land use alteration settings (O’Neill et al., 2016; Eyring et al., 2016). The outputs consist of both historical simulations and future projections. The historical simulations, covering the period 1979 - 2014, are based on observations while being forced by evolving external factors, such as solar variability, volcanic aerosols, and alterations in atmospheric composition. The future simulations, typically run from 2015 to 2100, are driven by different future emissions scenarios outlined in the shared socio-economic pathways.

In the following section, we will describe the source and processing of the ensembles of simulations used in this paper and the associated exploratory analysis.

### 2.2.1 Data sources and processing

The simulation data in our analysis are obtained from CMIP6 Data Portal<sup>3</sup> under SSP 126, 245, 370 and 585 experiments in ScenarioMIP, along with their corresponding historical simulations. These four SSPs are selected because they are highly representative, and have relatively complete set of climate model simulations available, compared to others, according to the summary of ScenarioMIP Data Holdings<sup>4</sup>.

The obtained simulation files are in monthly resolutions, while each modelling centre might use different grid labels and at different spatial resolutions, as illustrated by Figure 3. From the histogram, we can see that there is a significant division (the dashed line) of the numbers of grid cells among the simulators. Most of the resolutions are similar but a few simulators are run at a much higher resolution. This becomes relevant in the methodology to be discussed in Section 3.5 and that the high-resolution simulators, which are the four models with more than  $6 \times 10^5$  grid cells, will be discarded from the subsequent analysis.

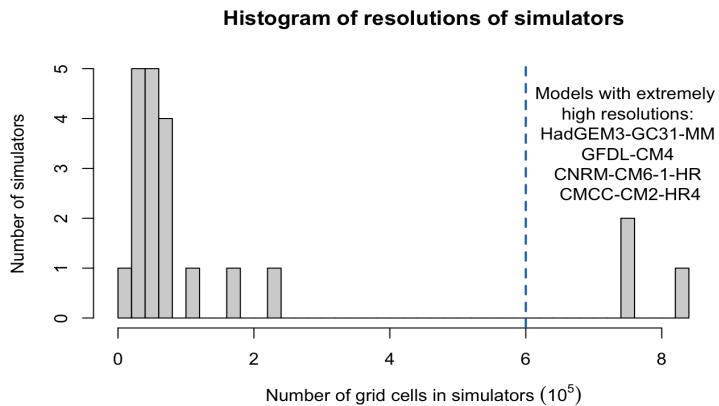


Figure 3: Histogram of number of grid cells used in simulators from CMIP6.

Since these simulation data are in similar format to the historical satellite measurement data, following similar area calculation procedures described in Section 2.1.1, the monthly time series of simulated Arctic sea ice area are obtained, in million  $km^2$ . For consistency with the satellite measurements, the simulated sea ice areas were computed excluding contributions from within the  $84.5^\circ$  N pole hole shown in Table 1. After this, annual minimum series are constructed by finding the minimum of monthly sea ice area in each year.

In some climate models, multiple runs of historical and future simulations are provided, labelled with different run IDs. These runs are generated under different climate models inputs, such as different initial values and perturbations in the system. The same run ID in each simulator represent identical input settings. Therefore, in each SSP experiment, we concatenate the historical runs with their future runs by matching their model names and run IDs. The runs which lack

<sup>3</sup><https://esgf-node.llnl.gov/projects/cmip6>

<sup>4</sup>[https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf\\_data\\_holdings/ScenarioMIP/index.html](https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/ScenarioMIP/index.html)

either historical or future simulations, or are discontinuous in time because of missing values, are considered incomplete, and hence discarded. More details about the omitted simulators and runs can be found in the Appendix A1.

### 2.2.2 Exploratory analysis and visualisation

We summarise all the simulators included in our report under each SSP experiment and their numbers of available runs in Table 2. We can see from the table that the numbers of runs differ between simulators and SSP experiments. Some of the simulators have more than one run under every experiment, such as the simulators **CanESM5** and **MPI-ESM1-2-LR**.

Simulator name	Historical + SSP126	Historical + SSP245	Historical + SSP370	Historical + SSP585
ACCESS-CM2	0	1	1	0
ACCESS-ESM1-5	0	2	0	0
CAMS-CSM1-0	2	2	2	2
CanESM5	20	50	44	29
CanESM5-CanOE	3	3	3	3
CESM2	0	3	0	3
CESM2-WACCM	0	2	2	3
CIESM	1	1	0	1
CMCC-CM2-SR5	1	1	1	1
CMCC-ESM2	1	1	1	1
CNRM-CM6-1	4	6	6	6
CNRM-ESM2-1	0	1	0	0
E3SM-1-1	0	0	0	1
EC-Earth3	0	1	0	12
EC-Earth3-AerChem	0	0	1	0
EC-Earth3-CC	0	1	0	1
EC-Earth3-Veg-LR	1	3	1	1
FIO-ESM-2-0	3	3	0	3
FGOALS-f3-L	1	1	1	1
GFDL-ESM4	0	2	0	0
HadGEM3-GC31-LL	1	1	0	3
INM-CM4-8	1	1	1	1
INM-CM5-0	1	1	5	1
IPSL-CM5A2-INCA	0	0	1	0
IPSL-CM6A-LR	3	11	10	6
MIROC-ES2L	1	15	2	9
MIROC6	3	3	3	3
MPI-ESM1-2-HR	2	2	10	2
MPI-ESM1-2-LR	30	30	30	30
MPI-ESM1-2-HAM	0	0	3	0
MRI-ESM2-0	1	1	5	1
NESM3	2	2	0	2
NorESM2-LM	1	3	3	1
NorESM2-MM	1	2	1	1
TaiESM1	1	1	1	1
UKESM1-0-LL	6	5	5	4
36 simulators	91 members	162 members	143 members	133 members

Table 2: CMIP6 simulators and the associated numbers of historical and SSP runs.

Figure 4 shows the time series plots of the ensembles of simulator outputs in the four SSP experiments, with the historical satellite measured sea ice area in thick black line and the ensemble means in thick red line. For multiple runs in one simulator, we consider them as different realisations of the same simulated trend from that model, and hence the runs from the same simulator are considered being in the same ‘group’. When we calculate the mean of the ensembles, we first average over the runs in every individual simulator, and then average over the group means. The simulations are the thinner coloured lines, and runs from the same simulator (i.e. in the same group) share the same color. From the plot, we can see a general pattern that both the simulations and the ensemble mean show faster decline towards zero in scenarios which represent higher greenhouse gas emissions.

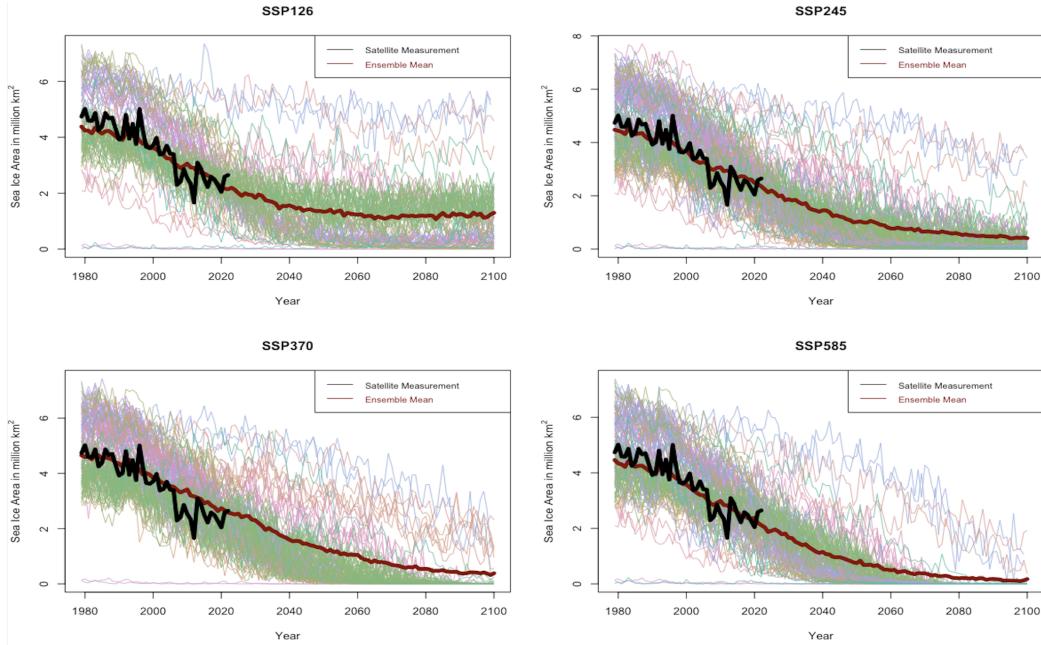


Figure 4: Time series of ensemble members with historical satellite measurement and ensemble mean, under SSP126, SSP245, SSP370, and SSP585.

### 2.3 Effective Radiative Forcings

The atmospheric forcings used to drive the CMIP6 model simulations under SSP126, 245, 370, and 585 are the historical (1979-2014) and future (2015-2100) estimates of greenhouse gas concentrations and other forcing factors, such as volcanic and solar activities. The “effective radiative forcing” (ERF) represents a combined effect of those forcing agents, and is acknowledged as a valuable indicator of the eventual temperature response (Oshima et al., 2020). The net effective radiative forcings are calculated in watts per square meter. The ERF series from year 1979 to 2014 are the estimates of actual historical forcings and are identical across the four scenarios, while from 2015 onwards, these are different estimates of future ERFs under the four emission scenarios. The four ERF time series applied in our analysis are of annual frequency

and are available from Github (Smith et al., 2021). The forcing variable of interest can be obtained under the ‘total’ column in the files, which is the crude approximation to net ERF. As described in Section 1.1, experiments SSP 126, 245, 370, and 585 represent low, intermediate, high, and very high greenhouse gas emission scenarios. Figure 5 shows the corresponding estimated effective radiative forcings under these four scenarios. The downward spikes correspond to volcanic activities. This plot exhibits a positive relationship between the forcing index and the greenhouse gas emission level, which can be seen from the steeper trends and the higher maximum values of the four forcing series as the SSP scenarios vary from SSP126 to SSP585.

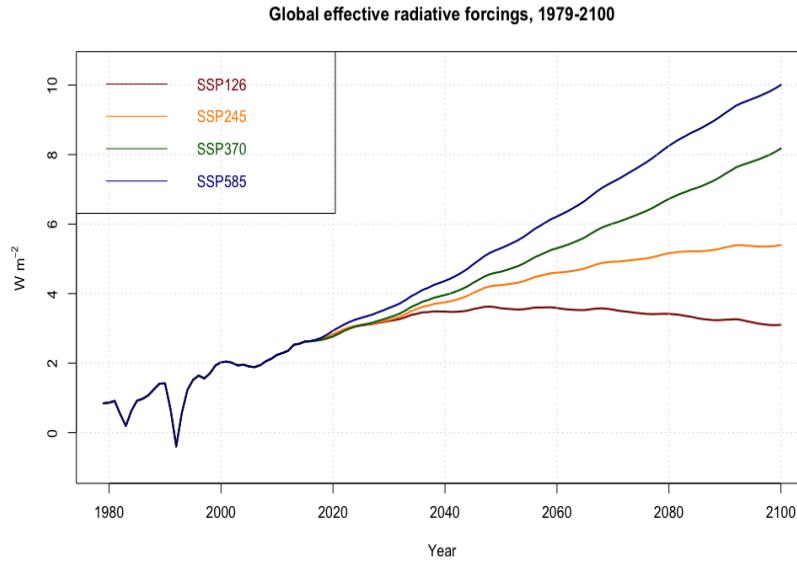


Figure 5: Annual time series of effective radiative forcings corresponding to SSP126, SSP245, SSP370, and SSP585.

As for historical forcings, Figure 6 shows the negative relationship between the annual minimum sea ice area and the estimated effective radiative forcings between 1979 and 2014. As the ERF increases, the minimum Arctic sea ice area tends to decrease.

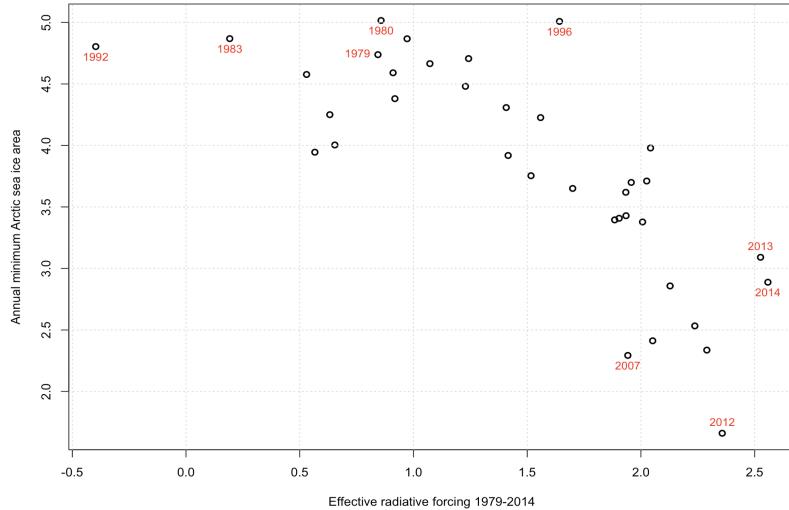


Figure 6: Annual minimum Arctic sea ice area against historical forcings 1979-2014.

## 2.4 Log-transformation to Sea Ice Data

One property of sea ice data is that the area cannot be negative. However, this could restrict the range of approaches that can be used for the subsequent analysis, for example, methods relying on normal distributions might not be appropriate for non-negative quantities. To address this limitation, one option is to apply data transformation. Log transformations are applied to both the satellite measurements and the simulator outputs, as described in equations (2.4.1). The  $Y_0^{original}$  represents the satellite measured data, and  $Y_{ij}^{original}$  represents the  $j^{th}$  run in the  $i^{th}$  simulator from the CMIP6 models. The lower bound  $10^{-5}$  attempts to minimise distortion from transformation as values tend to zero while ensuring finiteness. In the subsequent analysis,  $Y_0$  and  $Y_{ij}$  represent the log-transformed satellite measurements and simulations of Arctic sea ice area, respectively. Figure 7 displays the satellite measurements and simulation data after log-transformation.

$$\begin{cases} Y_0 &= \log(Y_0^{original} + 10^{-5}) \\ Y_{ij} &= \log(Y_{ij}^{original} + 10^{-5}) \end{cases} \quad (2.4.1)$$

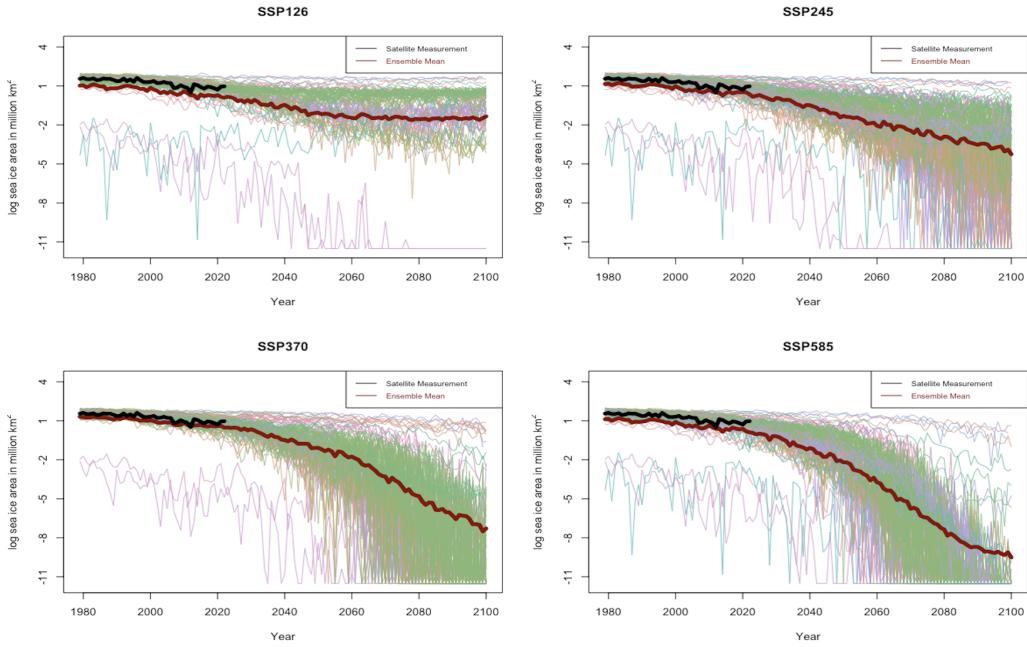


Figure 7: Time series of log-transformed ensemble members with historical satellite measurement and ensemble mean, under SSP126, SSP245, SSP370, and SSP585.

## 2.5 Summary of Chapter 2

In Chapter 2, we discussed the source of the data used in our paper, the subsequent data processing, data visualisation and exploratory analysis. We mainly focused on:

- Sources of the historical satellite measurement data, climate model simulations, and effective radiative forcings data.

- Based on the original data, we calculated the satellite measured annual minimum Arctic sea ice area, and following similar calculations, time series of annual minimum sea ice area are obtained from the climate model simulations.
- High resolution simulators are discarded and log transformation is applied to sea ice data.
- Exploring maps of satellite measurements, summary table of the climate model simulators, time series plots of the forcings, climate model simulations, and satellite measured sea ice data, as well as the log-transformed sea ice time series.

## 3 Methodology

### 3.1 Modelling the Real-world Annual Sea Ice Minima

This section introduces the state-space formulation for the real-world annual minima sea ice extent, based on the ideas developed by Chandler et al. (2023).

Let  $Y_0(t)$  be the real world annual minimum sea ice area at time  $t$  in log scale, as defined in equation (2.4.1). We can decompose  $Y_0(t)$  into the underlying trend and an “error” component:

$$Y_0(t) = \mu_0(t) + \epsilon_0(t) \quad (3.1.1)$$

where the ( $\epsilon_0(t)$ ) is mutually independent with mean 0 and constant variance  $\sigma_0^2$ . This constant variance assumption seems to be generally reasonable by observing that the annual variation in the log-transformed historical time series in Figure 7 is approximately constant.

In equation (3.1.1), a plausible representation of the trend  $\mu_0(\cdot)$  is needed, and the energy balance models (EBMs) can be helpful by connecting the trend with the ERF series in Section 2.3 which depict the total effect of GHG emissions. The EBM is a representation of the an idealised climate system where the energy in the atmosphere is always balanced (Bates, 2007). As mentioned in Section 1.3, EBMs have been applied in various climate research, assisting the analysis of climate responses to various external forcings and perturbations. In this report, we use a simple zero-dimensional EBM that describes the Earth as an uniform globe with global mean temperature being the only variable, which is expressed as an ordinary differential equation (ODE) (Chandler et al., 2023):

$$c_0 \frac{dT}{dt} = -bT(t) + f(t), \quad (3.1.2)$$

where  $T(t)$  is the perturbation about the global average temperature at its equilibrium in this climate system, and  $f(t)$  is the external forcing perturbation, which is the ERF series. Constants  $c_0$  and  $b$  are parameters related to the physical characteristics of the planet and are positive (Bates, 2007). This equation displays a trade-off between the effect of external forcings and the effect of global temperature returning to its equilibrium. When the effect of forcing is stronger,

the right-hand-side of equation (3.1.2) will be positive, and an increase in the global temperature will be seen. The EBM dynamics in equation (3.1.2) are for global temperature, however, the quantity of interest in our analysis is the sea ice extent, which tends to be negatively associated with temperature.

Given initial value  $T(0)$ , the solution to the linear first-order ODE (3.1.2) is

$$T(t) = T(0)e^{-\frac{b}{c_0}t} + c_0^{-1} \int_0^t f(t-u)e^{-\frac{b}{c_0}u} du \quad (3.1.3)$$

From this equation we can see the reason for not relating the forcing to our real-world quantity of interest directly. If we let  $T(t)$  to be the real-life climate variable, for a smooth forcing time series, the solution  $T(t)$  to the EBM equation (3.1.2) will also be smooth, which will result in a loss of erratic annual variations in the real-world series. Conversely, if the EBM is taken as a model for the underlying trend  $\mu_0$  rather than  $Y_0$ , then the trend will be smooth, but the error term in (3.1.1) will potentially produce inter-annual variations in the real-world quantity of interest  $Y_0$ .

Additionally, to handle the fact that our time series of interest are of annual frequency, the continuous functions and integrations in the result (3.1.3) need to be adapted to a discrete representation. First the  $T(t)$  is substituted by  $\mu_0(t)$ , which is the quantity of interest. Then we approximate the derivative in (3.1.2) with differences between successive times. As stated by Chandler et al. (2023), the discrete representation should be:

$$\mu_0(t) = \phi_0 \mu_0(t-1) + \gamma_0 f(t) \quad (3.1.4)$$

where  $\phi_0 = 1 - \frac{b}{c_0} < 1$ ,  $\gamma_0 = \frac{1}{c_0}$ ,  $\mu_0(t)$  and  $f(t)$  are the trend and forcing at time  $t$ . However, we notice that although the connection built by EBM between the trend and forcings provides a guidance on the trend structure, we should also allow deviations from this relationship which evolves with time. By adding a “drift” element  $\beta_0(t)$  to the above expression, (3.1.4) becomes:

$$\mu_0(t) = \phi_0 \mu_0(t-1) + \gamma_0 f(t) + \beta_0(t-1) \quad (3.1.5)$$

where the  $\beta_0(\cdot)$  is a random walk process:

$$\beta_0(t) = \beta_0(t-1) + Z_0(t), \quad Z_0(\cdot) \stackrel{iid}{\sim} N(0, \tau_0^2). \quad (3.1.6)$$

This  $\beta_0(t)$  can be interpreted as the amount that EBM approximation of the trend fail for the year  $t$ , and being a random walk allows more deviation from the EBM approximation over time, compared to the start. A smaller  $\tau_0^2$  indicates a more stable drift process and hence a more consistent relationship between the EBM evolution and the observations, as the relationship between them is not driven by the massive changes in the drift term. On the other hand, this drift process enables an adaptation when there is a significant inconsistency between the trend of the data and the EBM evolution.

In Chandler et al. (2023), it can be further shown that (3.1.5) contributes to a representation of the trend

$$\mu_0(t) = \phi_0^t \mu_0(0) + \gamma_0 \sum_{u=0}^{t-1} \phi_0^u f(t-u) + \mathcal{J}_0(t) \quad (3.1.7)$$

where  $\mathcal{J}_0(t)$  is associated with the drift term and has a complex structure. Equation (3.1.7) gives the solution to the discrete time process (3.1.5), and it is equivalent to the solution (3.1.3) in the continuous case (3.1.2) when  $\phi_0 > 0$ . So  $(0, 1)$  is a meaningful range for  $\phi_0$ .

Combining equations (3.1.1), (3.1.5), (3.1.6), we have three model equations which has the form of a linear state space model (Chandler et al., 2023)

$$Y_0(t) = \mu_0(t) + \epsilon_0(t), \quad \epsilon_0(\cdot) \stackrel{iid}{\sim} N(0, \sigma_0^2), \quad (3.1.8)$$

$$\mu_0(t) = \phi_0 \cdot \mu_0(t-1) + \gamma_0 \cdot f(t) + \beta_0(t-1), \quad (3.1.9)$$

$$\beta_0(t) = \beta_0(t-1) + Z_0(t), \quad Z_0(\cdot) \stackrel{iid}{\sim} N(0, \tau_0^2). \quad (3.1.10)$$

With the defined state vector  $S_0(t) = (\mu_0(t) \ \beta_0(t) \ \gamma_0)'$ , their corresponding state-space formulation is therefore:

$$S_0(t) = \begin{pmatrix} \phi_0 & 1 & f(t) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} S_0(t-1) + \begin{pmatrix} 0 \\ Z_0(t) \\ 0 \end{pmatrix}, \quad (3.1.11)$$

which opens up the possibility of using well-established approaches for state-space modelling.

### 3.2 Kalman Filter and Kalman Smoother

When a model is in state space form, we are able to apply Kalman filter, which works recursively to estimate the state vector given all the information available up to time  $t$ , and enables the maximum likelihood estimation of model parameters (Harvey, 1989). In our report, the Kalman filter is applied to construct the likelihood function of unknown parameters  $\sigma_0^2, \tau_0^2$  and  $\phi_0$  under the assumption of normality in (3.1.8) and (3.1.10). The Kalman smoother works backwards from final time  $T$ , and updates optimal estimates of the past state vectors  $\alpha_1, \dots, \alpha_T$ , by using all the available data (Chandler and Scott, 2011). Both Kalman smoother and Kalman filter work under the same principles, but with different focuses.

The rest of this section will focus on the explanations of Kalman filter and Kalman smoother in the general case of state-space models, which will be mainly based on Chapter 5.5.3 in the book by Chandler and Scott (2011). The general state-space model can be defined as

$$Y_t = Z_t \alpha_t + \epsilon_t \quad (3.2.1)$$

$$\alpha_t = M_t \alpha_{t-1} + \eta_t. \quad (3.2.2)$$

Equation (3.2.1) is the *measurement equation* that describes the relationship between the  $p \times 1$  observation vector  $Y_t$  and an usually non-observable  $m \times 1$  state vector  $\alpha_t$  at time  $t$ . And

(3.2.2) is the *transition equation* that illustrates how the state vector evolve over time. Below we include some further details about the elements in the two equations:

- $Z_t$ : an  $p \times m$  matrix that connects the states and observations.
- $M_t$ : an  $m \times m$  matrix containing the coefficients connecting successive state vectors.
- $(\epsilon_t)$ : a sequence of uncorrelated vectors such that means are 0 and common  $p \times p$  covariance matrix is  $H_t$ . This describes the variations about the state vector.
- $(\eta_t)$ : an uncorrelated sequence of vectors such that the means are 0 and common  $m \times m$  covariance matrix is  $Q_t$ . Also,  $(\eta_t)$  is uncorrelated with  $(\epsilon_t)$ .

Suppose we are given the observations  $y_1, y_2, \dots, y_T$ , we denote the estimator of state vector at time  $t_1$  given all information up to and including  $t_2$  as  $\hat{\alpha}_{t_1|t_2}$ . When  $t_1 > t_2$ , the prediction of  $Y_{t_1}$  at time  $t_2$  is  $\hat{Y}_{t_1|t_2}$ . For the estimation error  $\alpha_{t_1} - \hat{\alpha}_{t_1|t_2}$ , we denote its error covariance matrix as  $P_{t_1|t_2}$ . Similarly, define the error covariance matrix of  $Y_{t_1} - \hat{Y}_{t_1|t_2}$  as  $F_{t_1|t_2}$ .

Suppose the state vector at time  $T$  is estimated as  $\hat{\alpha}_{T|T} = \mathbb{E}(\alpha_T | Y_1, \dots, Y_T)$ , then given this estimate and its error covariance  $P_{T|T}$ , the optimal estimator of  $\alpha_{T+1}$  is

$$\begin{aligned}\hat{\alpha}_{T+1|T} &= \mathbb{E}[M_{T+1}\alpha_T + \eta_{T+1}|Y_T, Y_{T-1}, \dots, Y_1], \text{ by transition equation} \\ &= M_{T+1}\mathbb{E}[\alpha_T|Y_T, Y_{T-1}, \dots, Y_1] + \mathbb{E}[\eta_{T+1}|Y_T, Y_{T-1}, \dots, Y_1] \\ &= M_{T+1}\hat{\alpha}_{T|T}.\end{aligned}\tag{3.2.3}$$

By induction, we have  $\hat{\alpha}_{T+h|T} = M_{T+h}M_{T+h-1}\cdots M_{T+1}\hat{\alpha}_{T|T}$ .

Since the one-step ahead estimator is unbiased,  $E[(\alpha_T - \hat{\alpha}_{T|T})(\alpha_T - \hat{\alpha}_{T|T})'] = P_{T|T}$ , and because  $Q_T$  is the covariance matrix of  $\eta_T$ , it can be shown that the related error covariance is

$$P_{T+1|T} = M_{T+1}P_{T|T}M'_{T+1} + Q_{T+1}.\tag{3.2.4}$$

Recursively applying the calculations in (3.2.3) and (3.2.4), we are able to obtain predictions of state vectors up to  $h$  steps ahead, along with the covariance matrices  $P_{T+1|T}, \dots, P_{T+h|T}$ . Then by using the measurement equation (3.2.1), we can obtain up to  $h$  step ahead predictions of the observation vectors, which is given by  $\hat{Y}_{T+i|T} = \mathbb{E}[Z_{T+i}\alpha_{T+i|T} + \epsilon_{T+i}] = Z_{T+i}\hat{\alpha}_{T+i|T}$ ,  $i = 1, 2, \dots, h$ . The corresponding prediction error is

$$Y_{T+i} - \hat{Y}_{T+i|T} = Z_{T+i}(\alpha_{T+i} - \hat{\alpha}_{T+i|T}) + \epsilon_{T+i}$$

Following similar processes deriving (3.2.4), we can get the error covariance matrix  $F_{T+i|T}$ :

$$F_{T+i|T} = Z_{T+i}P_{T+i|T}Z'_{T+i} + H_{T+i}\tag{3.2.5}$$

where the  $H_{T+i}$  is the covariance matrix of  $\epsilon_{T+i}$ .

Given the above results, if  $\hat{\alpha}_{T_1|T-1}$  and  $P_{T_1|T-1}$  are calculated, then  $\hat{\alpha}_{T|T-1} = \mathbb{E}[\alpha_T | Y_{T-1}, \dots, Y_1]$  and  $\hat{Y}_{T|T-1} = \mathbb{E}[Y_T | Y_{T-1}, \dots, Y_1]$  and their corresponding error covariance matrices can be

obtained. By adding the assumption that  $(\epsilon_t)$  and  $(\eta_t)$  in equations (3.2.1) and (3.2.2) follow multivariate normal distributions, we arrive at the conditional joint multivariate distribution of  $(\alpha'_T \quad Y'_T)':$

$$\begin{pmatrix} \alpha_{T|T-1} \\ Y_{T|T-1} \end{pmatrix} \sim \mathcal{N} \left( \mathbb{E} \left[ \begin{pmatrix} \alpha_T \\ Y_T \end{pmatrix} \middle| Y_{T-1}, \dots, Y_1 \right], \begin{pmatrix} P_{T|T-1} \\ F_{T|T-1} \end{pmatrix} \right). \quad (3.2.6)$$

In Harvey (1989), when the observation  $Y_T$  is available, we can update the estimator of  $\alpha_T$  to (3.2.7), and we update the corresponding covariance matrix to (3.2.8)

$$\hat{\alpha}_{T|T} = \hat{\alpha}_{T|T-1} + P_{T|T-1} Z'_T F'_{T|T-1} (y_T - \hat{Y}_{T|T-1}) \quad (3.2.7)$$

$$P_{T|T} = P_{T|T-1} - P_{T|T-1} Z'_T F_{T|T-1}^{-1} Z_T P_{T|T-1}. \quad (3.2.8)$$

To conclude, equations (3.2.3) and (3.2.4) are called *prediction* equations, as they predict the state vector and observation vector at time  $T$ , using past information up to time  $T-1$ . They provide expressions for the mean and covariance matrix of the one-step-ahead predictive distribution which, under the model stated in Section 3.1, is multivariate normal. On the other hand, equations (3.2.7) and (3.2.8) are the *updating* equations, which update the estimation of  $\alpha_T$  and  $Y_T$  when  $Y_T$  is observed. The combination of these four equations makes up the **Kalman filter**. If given the initial values of  $\alpha_{0|0}$  and  $P_{0|0}$ , Kalman filter estimates the state vector when each new observation becomes available. In this report, the initial values of  $\alpha_{0|0}$  and  $P_{0|0}$  are chosen to be the plausible values in the spirit of setting prior distributions in a Bayesian framework, which will be discussed in detail in Section 4.1.2.

Having observed all  $\{Y_1, \dots, Y_T\}$ , Kalman filter provides real-time state estimation  $\{\hat{\alpha}_{t|t} : t = 1, 2, \dots, T\}$  and the corresponding  $\{P_{t|t} : t = 1, 2, \dots, T\}$ . The **Kalman smoother**, which uses all the information  $\{Y_1, \dots, Y_T\}$ , works backwards to update our estimates of state vectors at the past time points  $t = 1, 2, \dots, T-1$ . Relevant calculations are shown below (Harvey, 1989)

$$\begin{aligned} \hat{\alpha}_{t|T} &= \hat{\alpha}_{t|t} + P_{t|t} M'_{t+1} P_{t+1|t}^{-1} (\hat{\alpha}_{t+1|t} - M_{t+1} \hat{\alpha}_{t|t}), \\ P_{t|T} &= P_{t|t} + P_{t|t} M'_{t+1} P_{t+1|t}^{-1} (P_{t+1|T} - P_{t+1|t}) P_{t+1|t}^{-1} M_{t+1} P_{t|t}. \end{aligned}$$

By applying these recursively, starting from the final time point  $T$ , the Kalman smoother returns updated estimators of the state vectors  $\{\hat{\alpha}_{t|T} : t = 1, 2, \dots, T\}$ .

### 3.3 Parameter Estimation

From Section 3.1, there are three unknown parameters that we need to estimate in the model of the real-world sea ice minima, which together make up the *descriptor* vector which we have discussed in Section 1.3. So the descriptor vector for real-world quantity is  $\theta_0 = (\sigma_0^2 \quad \tau_0^2 \quad \phi_0)'$ . The complete collection of all the descriptors will be described in Section 3.5, where we fully develop the state space model that account for the multi-model structure.

Given observations  $\{y_1, \dots, y_T\}$ , the Kalman Filter can be used to form a likelihood function for the unknown descriptors under the assumption of normality (as stated in the final part of Section 3.1). Assume we have the initial conditions that  $\alpha_1 \sim \mathcal{N}(a_1, P_1)$ , where the  $a_1$  and  $P_1$  are prior guesses that will be discussed in Section 4.1.2, and let  $Y_t = \{y_1, \dots, y_t\}$ , then we get the likelihood and the log-likelihood functions (Durbin and Koopman, 2001)

$$L(\theta; y) = p_\theta(y_1, \dots, y_n) = p_\theta(y_1) \prod_{t=2}^n p_\theta(y_t | Y_{t-1}),$$

$$\log L(y) = \log p_\theta(y_1) + \sum_{t=2}^n \log p_\theta(y_t | Y_{t-1}) \quad (3.3.1)$$

where  $p(y_t | Y_{t-1})$  is the density of the one-step-ahead prediction, which is multivariate normal distribution with mean and covariance matrix calculated using the Kalman Filter, as discussed in Section 3.2.

Although the log-likelihood can be calculated, its analytical expression is not simple due to the recursive nature of the Kalman Filter calculations, hence maximisation could not be done in an analytical way. Using numerical methods, we maximise over the transformed descriptor vector  $\theta_0^l = (\log \sigma_0^2 \quad \log \tau_0^2 \quad \log[\phi_0/(1 - \phi_0)])'$ . These transformations are taken to 1) ensure the estimated variances are non-negative, 2) constrain estimated  $\phi_0$  within its range  $(0, 1)$ , 3) assist the accuracy in uncertainty measurements which will be discussed in Section 3.6 (Chandler et al., 2023). However, when the variances are very close to zero so that the log-transformed variances are extremely negative, the log-likelihood function  $\ell(\theta_0^l; y)$  could become very flat, resulting in a difficulty in numerical maximisation (Chandler et al., 2023).

To address this problem, we introduce a *penalised* log-likelihood function,  $\ell^*(\theta_0^l; y)$ , which adds a small concave quadratic penalty to increase the curvature at regions where  $\ell(\theta_0^l; y)$  is flat, as described below

$$\ell^*(\theta_0^l; y) = \ell(\theta_0^l; y) - a(\theta_0^l - \theta^l)^2,$$

where  $\theta^l$  is some reasonable prior guesses for the values of descriptors in  $\theta_0^l$ . The resulting *penalised* likelihood function will be

$$L^*(\theta_0^l; y) = \exp[\ell^*(\theta_0^l; y)] = L(\theta_0^l; y) \cdot \exp[-a(\theta_0^l - \theta^l)^2]$$

where the exponential term at the end is proportional to Gaussian probability density functions (pdfs). Thus, the penalty term could be viewed as a sum of logarithm of Gaussian pdfs. Since in Bayesian inference, we have  $\log \text{posterior} = \log \text{likelihood} + \log \text{prior} + \text{constant}$ , we could view this penalised log-likelihood as a log posterior density in the Bayesian framework, where the prior part is Gaussian. Hence the penalised MLE can alternatively be considered as a **maximum a posteriori** (MAP) estimator (Chandler et al., 2023). Also, if the log posterior is roughly quadratic around the MAP estimators, the posterior approximately follows multivariate normal distribution. This approximation is relevant to uncertainty quantification in

Section 3.6. The prior choices of parameters the initial values will be discussed in Section 4.1, and the maximisation work is done using the `nlm` functions (R Core Team, 2023).

### 3.4 Trend Estimation and Prediction

After estimating the  $\theta_0$ , we utilise the Kalman smoother (see Section 3.2) to estimate the state vectors, for example the trend  $\{\mu_0(t) : t = 1, 2, \dots, T\}$ , and associated prediction intervals for the observation vector  $Y_0(t)$  could also be calculated. From the Kalman smoother outputs, we can obtain the estimations  $\hat{\mu}_0$ , and the variance matrix of the estimation errors  $P_{t|T}$  (Durbin and Koopman, 2001). The remainder of this section discusses the derivation of a 95% prediction interval for  $Y_0(t)$ .

At each time  $t$ ,  $Y_0(t) = \mu_0(t) + \epsilon_0(t)$  by equation (3.1.1), so we have

$$\begin{aligned} Y_0(t) &= \hat{\mu}_0(t) - [\hat{\mu}_0(t) - \mu_0(t)] + \epsilon_0(t) \\ \therefore \text{Var}[Y_0(t) - \hat{\mu}_0(t)] &= \text{Var}[\hat{\mu}_0(t) - \mu_0(t)] + \text{Var}[\epsilon_0(t)]. \end{aligned}$$

Therefore the 95% confidence interval for  $Y_0(t)$  can be expressed as

$$\hat{\mu}_0(t) \pm z_{0.975} \sqrt{\text{Var}[Y_0(t) - \hat{\mu}_0(t)]} = \hat{\mu}_0(t) \pm 1.96 \sqrt{\text{Var}[Y_0(t) - \hat{\mu}_0(t)]},$$

where the  $\text{Var}[\hat{\mu}_0(t) - \mu_0(t)]$  is obtainable from Kalman smoother.

### 3.5 Multi-model Ensemble

Having discussed the formulation of EBM-inspired state space model, the Kalman filter, and parameter estimation, we now consider the full model structure accounting for the ensemble of simulators, to integrate their information for a better projection of the real-world quantity of interest. In this section, we mainly discuss the multi-model ensemble's relationship with the real climate system and the shared discrepancy framework which was mentioned in Section 1.3. At the end, we will arrive at a full state-space formulation of the ensemble members and real-world measurements, with an updated descriptor vector.

Suppose there are  $M$  simulators in an ensemble, since we may have multiple runs in one simulator, let  $Y_{ij}$  denote vector of outputs from the  $j^{th}$  run in the  $i^{th}$  simulator, where  $i \in \{1, 2, \dots, M\}$ . The runs from one simulator are considered to share the same underlying trend, as they are generated from the same climate model. Then the former EBM-inspired formulation given by equations (3.1.8), (3.1.9), (3.1.10) becomes

$$Y_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t), \quad \mathbb{E}[\epsilon_{ij}(t)] = 0, \text{Var}(\epsilon_{ij}) = \sigma_i^2 I \quad (3.5.1)$$

$$\mu_i(t) = \phi_0 \cdot \mu_i(t-1) + \gamma_i \cdot f(t) + \beta_i(t-1) \quad (3.5.2)$$

$$\beta_i(t) = \beta_i(t-1) + Z_i(t), \quad \mathbb{E}[Z_i(t)] = 0, \text{Var}(Z_i(t)) = \tau_i^2 \quad (3.5.3)$$

for individual simulator, where  $(\epsilon_{ij}(\cdot))$  and  $(Z_i(\cdot))$  are uncorrelated sequences (Chandler et al., 2023). According to the above equations, individual simulator has its own descriptors, trend, drift process and the coefficient of forcing ( $\gamma_i$ ).

The differences between simulators, and between individual simulators and the Earth system are recognised in the multi-model ensemble (Sansom et al., 2021). The facts that the climate models originate from common ancestral models and that they share similar limitations suggest that a *shared discrepancy* exists between the models and the real world. This shared discrepancy is discussed in detail in Chandler (2013). As a result, the trends  $\mu_i(\cdot)$  and the drift processes  $\beta_i(\cdot)$  are potentially not independent within the ensemble members. To account for this, initially we write the simulator-specific drifts and trends as

$$\beta_i(t) = \beta_\omega(t) + \tilde{\beta}_i(t) \quad (3.5.4)$$

$$\mu_i(t) = \mu_\omega(t) + \tilde{\mu}_i(t) \quad (3.5.5)$$

where the  $\mu_\omega(t)$  and  $\beta_\omega(t)$  are called the *consensus trend* and *consensus drift* among the simulators. All the quantities with  $\tilde{\cdot}$  on top denote the mutually independent discrepancies between the consensus and each simulator, or between the consensus and the real-world quantity. It is convenient to define the difference between the simulator consensus and the corresponding real-world quantity as the *shared discrepancy*, such as  $\tilde{\beta}_\omega(t) = \beta_\omega(t) - \beta_0(t)$  and  $\tilde{\mu}_\omega(t) = \mu_\omega(t) - \mu_0(t)$ . Figure 8 provides a general illustration of this framework, where  $\theta$ 's are descriptor vectors.

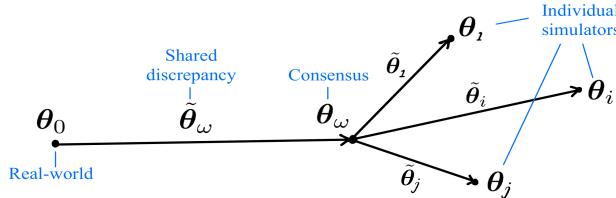


Figure 8: An illustration of the shared discrepancy framework.  $\theta_0$  denotes the descriptor for the real-world quantity;  $\{\theta_i : i = 1, \dots, M\}$  are descriptors for simulators; and  $\theta_\omega$  is the consensus among simulator descriptors, where  $\tilde{\theta}_\omega = \theta_\omega - \theta_0$  is the shared discrepancy.

According to the definitions of  $\mu_i$  and  $\beta_i$  in equations (3.5.2) and (3.5.3), the consensus and discrepancy structure must also apply to the  $\{Z_i\}$  (Chandler and Barnes, 2024). So

$$Z_i(t) = Z_0(t) + \tilde{Z}_\omega(t) + \tilde{Z}_i(t) \quad (3.5.6)$$

where  $Z_0(t)$ ,  $\tilde{Z}_\omega(t)$ ,  $\tilde{Z}_i(t)$  are mutually uncorrelated, and  $\tilde{Z}_i(t)$  represent the deviation from the consensus of the  $i^{th}$  simulator. Then the variance of the right-hand-side will be  $\tau_0^2 + \tilde{\tau}_\omega^2 + \tau_i^2$ , which is strictly greater than  $\tau_0^2$ . This constraint is hard to justify, as it says that every ensemble member deviates from the EBM approximation faster than does the real-world quantity of

interest. Hence, to remove this constraint, we modify equation (3.5.6) to

$$Z_i(t) = \alpha Z_0(t) + \tilde{Z}_\omega(t) + \tilde{Z}_i(t), \quad (3.5.7)$$

where  $\alpha$  is some real number between -1 and 1, because this enables  $\alpha^2 \tau_0^2 + \tilde{\tau}_\omega^2 + \tilde{\tau}_i^2 < \tau_0^2$ , so that the constraint is removed. This reveals that modifications to (3.5.4) and (3.5.5) are needed consequently. By relationships (3.5.1- 3.5.3), equations (3.5.4) and (3.5.5) are modified as

$$\beta_i(t) = \alpha \beta_0(t) + \tilde{\beta}_\omega(t) + \tilde{\beta}_i(t), \quad (3.5.8)$$

$$\mu_i(t) = \alpha \mu_0(t) + \tilde{\mu}_\omega(t) + \tilde{\mu}_i(t). \quad (3.5.9)$$

The expressions  $\gamma_i = \alpha \gamma_0 + \tilde{\gamma}_\omega + \tilde{\gamma}_i$  and  $\phi_i = \phi_\omega + \tilde{\phi}_i$  can also be written (Chandler et al., 2023). Given the above, a large number of descriptors  $\{\alpha, \sigma_0^2, \tau_0^2, \phi_0, \tilde{\tau}_\omega^2, \sigma_i^2, \tilde{\tau}_i^2, \tilde{\phi}_i\}$  for  $i = 1, 2, \dots, M$  are involved, which pose difficulties in maximising the log-likelihood function using numerical methods, as stated in Section 3.3. The complexity of numerical optimisation can be reduced if we could assume common values of  $\sigma_i^2, \tilde{\tau}_i^2$ , and  $\tilde{\phi}_i$  among all ensemble members, which is similar to the assumption of exchangeability among simulators that is widely used in the Bayesian literature on multimodel ensembles (Chandler et al., 2023; Sansom et al., 2021). For models with similar levels of complexity, this assumption is reasonable. However, some possible aspects may imperil this assumption, and one of them is that the resolution used in different simulators might be significantly different. As displayed in Section 2.2.1, Figure 3 shows a significant division (the dashed line) of the resolutions between the simulators. Hence the simulators with extremely high resolutions are removed. Given the above, we arrive at 8 parameters in the descriptor vector for estimation, which are  $\alpha, \sigma_0^2, \tau_0^2, \phi_0, \tilde{\tau}_\omega^2, \sigma_i^2, \tilde{\tau}_1^2$  and  $\phi_1$ .

Additionally, for relations (3.5.2) and (3.5.3) to hold, the corresponding components of (3.5.7- 3.5.9) must follow similar dynamics, and equations (3.5.10-3.5.13) are the dynamics required (Chandler et al., 2023).

$$\tilde{\mu}_i(t) = \phi_1 \tilde{\mu}_i(t-1) + \tilde{\gamma}_i f(t) + \tilde{\beta}_i(t-1) \quad (3.5.10)$$

$$\tilde{\mu}_\omega(t) = \alpha(\phi_\omega - \phi_0) \mu_0(t-1) + \phi_\omega \tilde{\mu}_\omega(t-1) + \tilde{\gamma}_\omega f(t) + \tilde{\beta}_\omega(t-1). \quad (3.5.11)$$

$$\tilde{\beta}_i(t) = \tilde{\beta}_i(t-1) + \tilde{Z}_i(t); \quad \mathbb{E}(\tilde{Z}_i(t)) = 0, \text{Var}(\tilde{Z}_i(t)) = \tilde{\gamma}_i^2 \quad (3.5.12)$$

$$\tilde{\beta}_\omega(t) = \tilde{\beta}_\omega(t-1) + \tilde{Z}_\omega(t); \quad \mathbb{E}(\tilde{Z}_\omega(t)) = 0, \text{Var}(\tilde{Z}_\omega(t)) = \tilde{\gamma}_\omega^2 \quad (3.5.13)$$

Also we notice that there are  $M + 1$  sources of series to estimate  $M + 2$  trends, which are  $\mu_0, \tilde{\mu}_\omega, \tilde{\mu}_1, \dots, \tilde{\mu}_M$ . This will result in a non-uniqueness in the trends. To demonstrate this, the trend  $\mu_i(t) = \alpha \mu_0(t) + \tilde{\mu}_\omega(t) + \tilde{\mu}_i(t) = \alpha \mu_0(t) + [\tilde{\mu}_\omega(t) - K] + [\tilde{\mu}_i(t) + K]$  holds for any  $K$ . However, this non-identifiability could be solved by adding a “sum to zero” constraint  $\sum_{i=1}^M \tilde{\mu}_i(t) = 0$ , which gives  $\tilde{\mu}_M(t) = -\sum_{i=1}^{M-1} \tilde{\mu}_i(t)$  (Chandler et al., 2023).

Given the above relationships, we now want to construct the full state-space formulation. In

our multi-model ensemble, define  $n_1, n_2, \dots, n_M$  as the number of runs in the  $M^{th}$  simulator. Let  $\mathbf{Y}(t) = (Y_0(t) \ Y_{11}(t) \cdots Y_{1n_1}(t) \ Y_{21}(t) \cdots Y_{Mn_M}(t))'$ . The state vector is expressed as

$$\mathbf{S}(t) = (\mu_0(t) \ \beta_0(t) \ \gamma_0 \ \tilde{\mu}_\omega(t) \ \tilde{\beta}_\omega(t) \ \tilde{\gamma}_\omega \ \tilde{\mu}_1(t) \ \tilde{\beta}_1(t) \ \tilde{\gamma}_1 \ \cdots \ \tilde{\mu}_{M-1}(t) \ \tilde{\beta}_{M-1}(t) \ \tilde{\gamma}_{M-1})' \quad (3.5.14)$$

Hence a full state-space formulation of the dynamic linear model (DLM) will be

$$\mathbf{Y}(t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \alpha & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \alpha & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & \cdots & -1 & 0 & 0 \end{pmatrix} \mathbf{S}(t) + \begin{pmatrix} \epsilon_0(t) \\ \epsilon_{11}(t) \\ \vdots \\ \epsilon_{1n_1}(t) \\ \epsilon_{21}(t) \\ \vdots \\ \epsilon_{Mn_M}(t) \end{pmatrix} \quad (3.5.15)$$

According to equations (3.1.9) and (3.5.10-3.5.13), the transition equation is given by

$$\mathbf{S}(t) = \begin{pmatrix} \phi_0 & 1 & f(t) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \alpha(\phi_1 - \phi_0) & 0 & 0 & \phi_1 & 1 & f(t) & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_1 & 1 & f(t) & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \phi_1 & 1 & f(t) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \mathbf{S}(t-1) + \begin{pmatrix} 0 \\ Z_0(t) \\ 0 \\ 0 \\ \tilde{Z}_\omega(t) \\ 0 \\ \tilde{Z}_1(t) \\ 0 \\ \vdots \\ 0 \\ \tilde{Z}_{M-1}(t) \\ 0 \end{pmatrix} \quad (3.5.16)$$

In this expression, under the “sum to zero” constraint, the covariance matrix of the vector ( $\tilde{Z}(t)$ ) on the right-hand-side is no longer diagonal, and its form is given by equation (19) in Chandler et al. (2023).

Under these representations, we can apply the Kalman filter to estimate descriptors  $\sigma_0^2, \sigma_1^2, \tau_0^2, \tilde{\tau}_\omega^2, \tilde{\tau}_1^2, \phi_0, \phi_1$  and  $\alpha$ , which jointly make up the full descriptor vector  $\theta$ . As discussed in Section 3.3, given prior distributions on the transformed descriptors, we will be able to obtain posterior distributions of quantities of interest. Uncertainties could also be characterised, which will be discussed in the next section.

### 3.6 Uncertainty

In the state-space formulation, we are able to obtain prediction intervals for  $Y_0$  using methods described in Section 3.4, but this does not consider the estimation uncertainty of the descriptor vector  $\theta$ . The Bayesian approach to the analysis and the prior distributions of transformed descriptors discussed in Sections 1.3, 3.1, and 3.3 allow us to quantify the uncertainty in future projections by calculating prediction intervals for  $Y_0(t)$  that account for the uncertainty in estimating the descriptor vector  $\theta_0$ .

Let  $\mathbf{Y}$  be the set of all observations up to time  $T$  that were used to estimate  $\theta$ . And let  $\mathbf{S}$  be the entire set of all the state vectors over that time. Denote  $\mathbf{Y}^{new}$  as the future observations. We apply the following procedure of posterior sampling to obtain the uncertainties associated with the real-world quantity of interest, with the help of the associated functions in the `TimSPEC` package in R which achieve these in practice (Chandler, 2024).

1. First we sample  $\theta$  from the posterior distribution  $\pi(\theta|\mathbf{Y})$ . By default the samples are drawn from the large-sample multivariate normal approximation to the posterior, which has been discussed in Section 3.3, although importance sampling can be used if necessary to improve this approximation (Chandler et al., 2023).
2. Next we sample the collection of state vectors  $\mathbf{S}$  from  $\pi(\mathbf{S}|\theta, \mathbf{Y})$ , given the sampled  $\theta$  from
  1. The “forward-filtering-backward-sampling” (FFBS) approach is one of the mostly used method to sample state vectors given  $\theta$  and observations.
3. Finally, we use these sampled state vectors in conjunction with the sampled values of  $\theta$  to sample  $(Y_0(t))$  from distribution  $\pi(\mathbf{Y}^{new}|\theta, \mathbf{S})$ , which are typically normal with  $\mathbb{E}[Y_i(t)|\mathbf{S}] = \mu_i(t)$  and  $Var[Y_i(t)|\mathbf{S}] = \sigma_i^2$ .

### 3.7 Summary of Chapter 3

In Chapter 3, we mainly focused on the derivation and explanations of the EBM-inspired state space formulation for the real-world quantity and then extended it to the full formulation including the ensemble of simulators. We started by developing a state-space model for the real-world observations in Section 3.1. We then explained the uses of Kalman filter and Kalman smoother for constructing likelihood functions of unknown descriptors, and for estimating state vectors, respectively. Kalman filter enables parameter estimation using a penalised log-likelihood in Section 3.3, and the Kalman smoother output assists the trend estimation and finding prediction intervals in Section 3.4. Then in Section 3.5, we described the shared-discrepancy framework, and arrived at a full state-space formulation of the multi-model ensemble, and finalised 8 descriptors. Finally in Section 3.6, we explained the posterior sampling of  $\mathbf{Y}^{new}$ , which can be used to characterise uncertainty.

## 4 Application to Annual Minimum Arctic Sea Ice Extent

In this section, we will be addressing the problems about prior choices of descriptors and the initial values of state vector, as mentioned in Sections 3.3 and 3.5. Then we will apply the presented methods to our ensemble of annual minimum Arctic sea ice extent projections to estimate parameters, generate projections, and quantify uncertainties.

### 4.1 Choices of Prior

#### 4.1.1 Prior choices for descriptors

As described in Section 3.5, there are eight parameters that make up the descriptor vector we aim to estimate, and Table 3 summarises their meanings, which are similar to the interpretations from Chandler et al. (2023).

Descriptors	Explanation
$\sigma_0^2$	variance of year-to-year fluctuations around the underlying trend for real-world quantity
$\sigma_1^2$	common variance of year-to-year fluctuations about the underlying trend for ensemble members
$\tau_0^2$	variance of the drift process in EBM approximation for real-world quantity
$\tilde{\tau}_\omega^2$	variance of the shared discrepancy component in the drift increment in multi-model ensemble
$\tilde{\tau}_1^2$	common variance of drift increments relative to ensemble consensus, for ensemble members
$\phi_0$	factor of trend inertia for real-world quantity of interest
$\phi_1$	common factor of trend inertia shared by simulators in the ensemble
$\alpha$	scaling factor of trends and drifts of real-world quantity, related to ensemble consensus

Table 3: Interpretations of descriptors from Section 3.5. The real-world quantity is the logged measured annual minimum sea ice area, and the ensemble members are also in log scale.

In order to find the MAP estimators for the transformed descriptors (see Section 3.3), we need some prior settings on  $\alpha$ ,  $\log \sigma_0^2$ ,  $\log \sigma_1^2$ ,  $\log \tau_0^2$ ,  $\log \tilde{\tau}_\omega^2$ ,  $\log \tilde{\tau}_1^2$ ,  $\text{logit } \phi_0$ , and  $\text{logit } \phi_1$ .

Since our annual minimum Arctic sea ice is in log scale, the prior for variance parameters are considered in terms of relative changes instead of absolute changes in corresponding variables (Chandler et al., 2023). For example, if we say the observation is within  $\pm k \ km^2$  about the trend on a log scale, then if  $k$  is fairly small, this translates into a range of approximately  $\pm 100k\%$  about the trend on the original scale. The prior choices are documented below.

- $\log \sigma_0^2$ : It is improbable that over 5% of annual minimum sea ice areas will be more than a factor of 5 different from, or fall within 1% of, the underlying trend. This implies the  $\sigma_0$  for the log annual minimum sea ice area is bounded by 0.005 and 0.8, from the reasoning

$$\log(0.99\mu_0^{original}) \leq \log Y_0^{original} \leq \log(1.01\mu_0^{original})$$

$$\therefore \log 0.99 + \mu_0 \leq Y_0 \leq \log 1.01 + \mu_0 \implies \sigma_0 = \log 1.01 \div 2 = 0.005$$

$$\text{similarly, } \log(0.2\mu_0^{original}) \leq \log Y_0^{original} \leq \log(5\mu_0^{original})$$

$$\therefore -\log 5 + \mu_0 \leq Y_0 \leq \log 5 + \mu_0 \implies \sigma_0 = \log 5 \div 2 = 0.8$$

where  $Y_0$ ,  $\mu_0$  are the sea ice area and trend in log scale, and the superscript *original* represents “in original scale”, as mentioned in (2.4.1). Hence, the standard deviation  $\sigma_0$  of the log annual minimum sea ice area  $Y_0$  is between 0.005 and 0.8, so  $\log \sigma_0^2$  is between  $\log(0.005^2) = -10.60$  and  $\log(0.8^2) = -0.4$ . We can interpret this as the 95% central interval covering 4 standard deviations, so the prior mean and standard deviation of  $\log \sigma_0^2$  are  $(-10.6 - 0.4)/2 = -5.5$  and  $| -10.6 - (-0.4)|/4 = 2.6$ , respectively.

- $\log \sigma_1^2$ : The same prior distribution is applied to  $\log \sigma_1^2$  as it has comparable meaning to  $\log \sigma_0^2$  regarding the scale of year-to-year fluctuations.
- Similarly,  $\log \tilde{\tau}_\omega^2$  and  $\log \tilde{\tau}_1^2$  are assumed to share the same prior as  $\log \tau_0^2$ . To find plausible prior values for the mean and standard deviation of  $\log \tau_0^2$ , we consider the cumulative change of the EBM approximation error over 100 years, which is  $\sum_{i=1}^{100} Z_0(t+i) = \beta_0(t+100) - \beta_0(t)$  according to equation (3.1.6). If converted to original scale, the cumulative change will be  $\exp[\sum_{i=1}^{100} Z_0(t+i)]$ , which is unlikely to vary more than a factor of 100 or less than a factor of 1.01 over the century (Chandler et al., 2023). So  $(\log 1.01, \log 100) = (0.01, 4.6)$  is a plausible range for  $\sum_{i=1}^{100} Z_0(t+i)$ . Following the same reasoning as for  $\sigma_0^2$  above, its variance is then between  $0.005^2$  and  $2.3^2$ . Averaging over 100 years,  $\tau_0^2$  lies within  $(0.005^2/100, 2.3^2/100)$ . So  $\log \tau_0^2$  is likely to be within  $(\log(0.005^2/100), \log(2.3^2/100)) = (-15.2, -2.9)$ . Thus,  $\log \tau_0^2$  has prior mean -9.1 and standard deviation 3, which are also applied to  $\log \tilde{\tau}_\omega^2$  and  $\log \tilde{\tau}_1^2$ .
- $\text{logit } \phi_0$ : From Section 3.3, this transformation to  $\phi_0$  constrains it within  $(0, 1)$ . A  $\mathcal{N}(0, 5^2)$  prior is set for  $\text{logit } \phi_0$ , because this gives more than 95% of probability that  $\phi_0$  lies between 0.0001 and 0.9999. The same prior is adopted for  $\text{logit } \phi_1$ .
- $\alpha$ : The scaling factor  $\alpha$  in equations (3.5.7-3.5.9) is not likely to be negative. To let the ensemble consensus trend  $\mu_\omega$  be roughly the same as the real-world  $\mu_0$ ,  $\alpha$  is expected to be close to 1. So a  $\mathcal{N}(1, 1)$  prior is chosen for  $\alpha$  (Chandler et al., 2023).

#### 4.1.2 Initial settings for the state vector

From Section 3.5 equation (3.5.14), we have the full state vector  $\mathbf{S}(t)$

$$\mathbf{S}(t) = (\mu_0(t) \ \beta_0(t) \ \gamma_0 \ \tilde{\mu}_\omega(t) \ \tilde{\beta}_\omega(t) \ \tilde{\gamma}_\omega \ \tilde{\mu}_1(t) \ \tilde{\beta}_1(t) \ \tilde{\gamma}_1 \ \cdots \ \tilde{\mu}_{M-1}(t) \ \tilde{\beta}_{M-1}(t) \ \tilde{\gamma}_{M-1})'.$$

which involves the expected value of logged measured annual minimum sea ice  $\mu_0(t)$ , the initial drift approximation  $\beta_0(0)$ , the coefficient of forcing  $\gamma_0$ , and their corresponding components in the shared and simulator-specific discrepancies.

In order to construct the likelihood function using (3.3.1) and apply Kalman filter to estimate parameters, we need initial values of the state vector and the initial covariance matrix of the associated estimation error, which are the  $\alpha_{0|0}$  and  $P_{0|0}$  in Section 3.2. We call them  $\mathbf{S}(0)$  and

$\mathbf{C}_0$  in our specified state-space formulation.

In  $\mathbf{S}(0)$ , except for  $\mu_0(0)$  and the  $\gamma$ 's, we set the prior means for all the other quantities to 0, as we lack prior reasons to decide the positiveness or negativeness for the drift and the trend components in discrepancies. As for the prior mean for  $\mu_0(0)$ , it is unlikely that the annual minimum Arctic sea ice area is more than 8 million  $km^2$  or below 0 million  $km^2$  in the regions outside the pole hole (see Section 2.1.1), so the prior mean of the logged sea ice extent is chosen to be  $\log[(8 + 0)/2] = \log 4$ . Regarding the prior means for  $\gamma$ 's, the coefficients of forcings, from prior knowledge we know that the evolution of Arctic annual minimum sea ice extent is negatively related to the GHG emissions and effective radiative forcings, which is  $f(t)$ . So we take -0.5, which is a negative value between -1 and 0, to be the prior of the  $\gamma$ 's.

As for the prior covariance matrix  $\mathbf{C}_0$ , we first construct a diagonal matrix whose diagonal elements are decided by some assessments of possible ranges for each quantity in the state vector. For the trend  $\mu_0(0)$ , it is unlikely that in original scale, the actual value departs from its prior expectation by more than a factor of five, so its prior variance is chosen to be  $[(\log 5)/2]^2 = 0.65$ . The variance of  $\beta_0(0)$  is unlikely to be larger than that of  $\mu_0(0)$ , by equation (3.1.10), so it is conservative to adopt the same prior variance for  $\beta_0(0)$ . However, we do not know much about the actual values of  $\gamma$ 's, so we allow for a larger prior variance to them, say  $[(\log 25)/2]^2 = 2.6$ . The prior variances for  $\tilde{\mu}_\omega(t), \tilde{\beta}_\omega(t)$  and  $\tilde{\gamma}_\omega$  are unlikely to be larger than those of the corresponding real-world quantities, so we apply the same set of prior variances to  $\tilde{\mu}_\omega(t), \tilde{\beta}_\omega(t)$  and  $\tilde{\gamma}_\omega$ . However, for the components associated to simulator-specific discrepancies, which are the  $\tilde{\mu}_i(t), \tilde{\beta}_i(t)$  and  $\tilde{\gamma}_i$ 's ( $i = 1, \dots, M - 1$ ), variations can be larger, because some simulators may have trends that deviates substantially from the consensus, so a larger value  $[(\log 25)/2]^2$  is assigned to their prior variances.

The below equations summarise the prior means and variances used for the state vector

$$\begin{aligned}\mathbf{S}(0) &= (\log 4 \quad 0 \quad -0.5 \quad 0 \quad 0 \quad -0.5 \quad 0 \quad 0 \quad -0.5 \quad \dots \quad 0 \quad 0 \quad -0.5)' \\ \mathbf{C}_0 &= \text{diag}(0.65, \quad 0.65, \quad 2.6, \quad 0.65, \quad 0.65, \quad 2.6, \quad 2.6, \quad 2.6, \quad 2.6, \dots, \quad 2.6, \quad 2.6, \quad 2.6)\end{aligned}$$

## 4.2 Model Results

### 4.2.1 Parameter estimates and trend projection

Under each SSP experiment, the estimates for transformed descriptors ( $\log \sigma_0^2, \log \sigma_1^2, \log \tau_0^2, \log \tilde{\tau}_\omega^2, \log \tilde{\tau}_1^2, \text{logit } \phi_0, \text{logit } \phi_1$  and  $\alpha$ ) and the state vectors of the real-world quantity can be computed by integrating all the methods and prior choices discussed above, and is done in practice using the TimSPEC package in R (Chandler, 2024). Tables 4 summarises the estimated values for the transformed descriptors and the time-invariant forcing coefficients (the  $\gamma$  values, estimated using Kalman smoother) in the full state-space formulation for each SSP experiment.

Descriptors	$\log \sigma_0^2$	$\log \sigma_1^2$	$\log \tau_0^2$	$\log \tilde{\tau}_\omega^2$	$\log \tilde{\tau}_1^2$	$\text{logit } \phi_0$	$\text{logit } \phi_1$	$\alpha$	$\gamma_0$	$\tilde{\gamma}_\omega$
Estimates (SSP126)	-3.617	-1.761	-8.623	-8.995	-5.290	-1.580	0.342	0.806	-0.194	0.033
Standard Errors	(0.144)	(0.015)	(0.177)	(0.197)	(0.129)	(0.173)	(0.100)	(0.161)	(0.036)	(0.031)
Estimates (SSP245)	-3.800	-0.116	-7.292	-9.489	-7.897	-1.611	1.747	0.914	-0.203	0.142
Standard Errors	(0.034)	(0.016)	(0.034)	(0.034)	(0.114)	(0.036)	(0.035)	(0.115)	(0.039)	(0.030)
Estimates (SSP370)	-4.378	0.412	-5.632	-7.396	-5.984	-1.333	1.215	0.520	-0.097	-0.042
Standard Errors	(0.122)	(0.011)	(0.153)	(0.133)	(0.108)	(0.139)	(0.095)	(0.104)	(0.056)	(0.038)
Estimates (SSP585)	-4.392	0.229	-5.505	-6.975	-5.145	-1.534	0.997	0.572	-0.108	-0.090
Standard Errors	(0.127)	(0.012)	(0.135)	(0.136)	(0.108)	(0.139)	(0.097)	(0.102)	(0.056)	(0.040)

Table 4: Estimates of transformed descriptors and  $\gamma_0$  and  $\tilde{\gamma}_\omega$  in the state-space model.

From the table, we can observe that the estimates of some descriptors vary between SSPs, which is contradicting our expectation that the descriptors are similar across all SSPs, as it seems the only difference between them is the forcing. We observe that most of the standard errors for the descriptors are relatively small, so that the differences of estimates between SSPs are not likely to be explained by sampling variations, which is further supported in Section 4.2.2. Alternatively, it is possible that the SSPs also differ in other ways, because the simulators and their numbers of runs in the ensembles are different between SSPs. If this explanation stands, then it suggests that the current model describing the ensemble and real-world climate is oversimplified to fully capture the structure.

In Table 4, in addition to the transformed descriptors, we also included the estimates of the  $\gamma_0$ 's and their associated shared discrepancies  $\tilde{\gamma}_\omega$ 's in the last two columns. We observe that in all SSPs,  $\gamma_0$  is larger than  $\tilde{\gamma}_\omega$  in absolute value, which indicates that the real-world measurements are more sensitive to the forcings than do the simulators.

With the estimated descriptors, we can obtain posterior distributions for the logged real-world Arctic sea ice minimum series, using the methodology described in Section 3.6. With the estimation for logged annual minimum Arctic sea ice extent, we can transform from the log scale to the original scale by taking exponential to the results, and obtain the estimates of measured annual minimum Arctic sea ice and the associated prediction intervals. Figure 9 shows the post-processed ensemble of Arctic sea ice extent, based on historical measurements from 1979-2022 (thick black line) and ensemble outputs from 1979-2100 (red thin lines). The blue thick lines in the four plots are the posterior mean trends of measured annual minimum sea ice extent, and the coloured regions are the prediction intervals for  $Y_0$  as described in Section 3.4.

We can see that the estimated trend  $\hat{\mu}_0$  generally captures the historical changes in annual minimum sea ice extent. In the projection period, the estimated trends in all scenarios are slightly above the ensemble means which only account for the ensemble outputs. This is reasonable because the projected trends both capture the simulation outputs from ensembles and inherit the trend from real-world measurements which becomes relatively flat in recent years, compared to the decreasing trends in the ensemble simulations.

Among the four socio-economic development scenarios, the projected trend under the lightest GHG emission (in SSP126) shows a flat pattern with no significant decline in the annual minimum sea ice extent, and a slight upward trend can be seen at the end of the century. However, in all the other three scenarios, the annual minimum Arctic sea ice is projected to decrease, especially in pathways SSP370 and SSP585. With high and very high future GHG concentrations, it is possible that by the end of this century or even earlier, the annual minimum Arctic sea ice will fall below 1.19 million  $km^2$ .

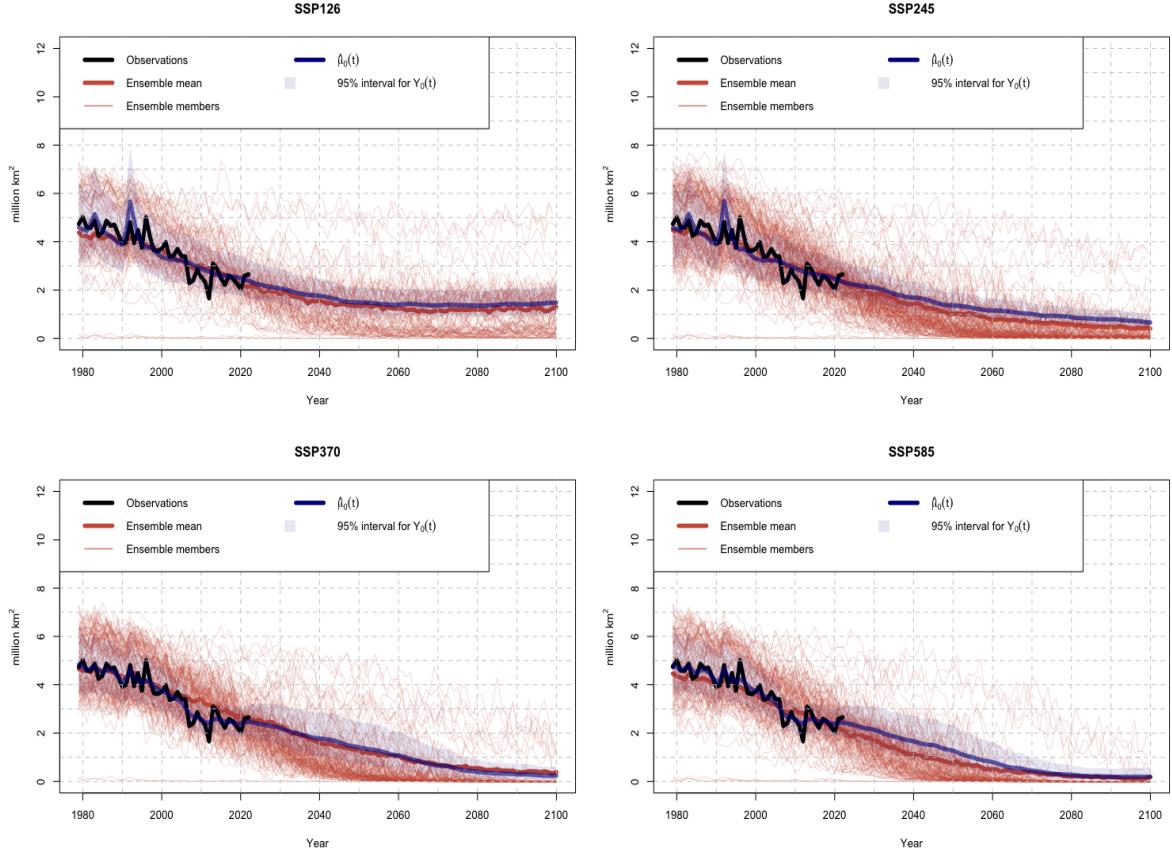


Figure 9: Postprocessed ensembles of annual minimum Arctic sea ice extent, under experiments SSP 126, 245, 370, and 585.

We can also extract the estimated trends, to get Figure 10 which presents the relationship between projections under different development pathways. Under all emission scenarios, the annual minimum sea ice extent is projected to decline, although in SSP126 there is a slight upward tendency from around 2050. We can see that the higher the SSP scenario, the faster the projected decline, and the year when annual minimum sea ice area falls below some threshold is earlier in higher emission scenarios. As for the historical period (1979-2022), we notice that the estimated trends are very similar in SSP126 and SSP245, while SSP370 and SSP585 share another trend. We also observe that the trends in SSP126 and SSP245 have two spikes in the early years, which coincide with the spikes in the ERF plot (see Figure 5). This can be

explained by the fact that in Table 4, the estimated values for descriptor  $\tau_0^2$  in SSP126 and 245 are smaller, and the  $\gamma_0$ 's in SSP126 and 245 have larger absolute values than that in SSP370 and 585. These facts suggest more stable drift processes in SSP126 and 245, so that the forcing series are able to contribute more patterns to their trends, according to equations (3.1.8-3.1.10).

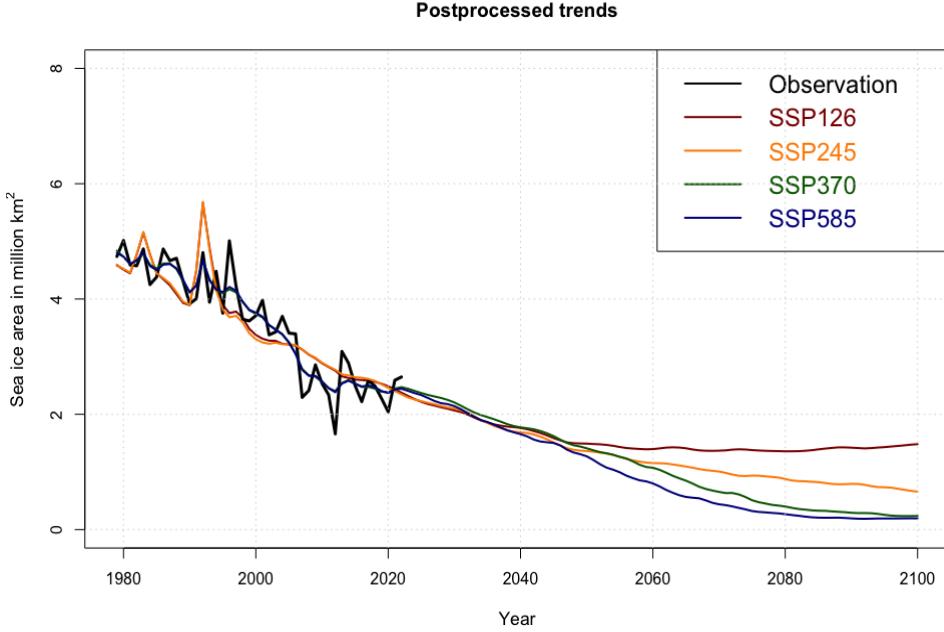


Figure 10: Kalman smoothed trend projections from postprocessed ensembles of annual minimum Arctic sea ice extent under SSPs 126, 245, 370, and 585.

#### 4.2.2 Uncertainty quantification

The projections presented so far and their associated prediction intervals do not account for the uncertainty in descriptors estimation. As discussed in Section 3.6, by drawing samples of  $\mathbf{Y}^{\text{new}}$  from approximate posterior distribution, we could estimate the trend  $\hat{\mu}_0(t)$  by averaging over 1000 posterior samples, and the associated 95% prediction interval can be obtained from the 2.5% and 97.5% quantiles in the samples. Figure 11 displays the resulting  $\hat{\mu}_0(t)$  and prediction intervals after the uncertainties in descriptors are taken into account, for the four SSP scenarios. These updates are not substantially different from the estimates and prediction intervals in Figure 9, indicating the role of the uncertainties in descriptor estimation is not significant in this situation. This also suggests that the sampling error is not the dominant reason for the differences in the descriptor estimates between SSP experiments, as presented in Table 4. Because if sampling error is the primary reason, the posterior mean estimates of historical trends for different SSPs in Figure 11 would be similar.

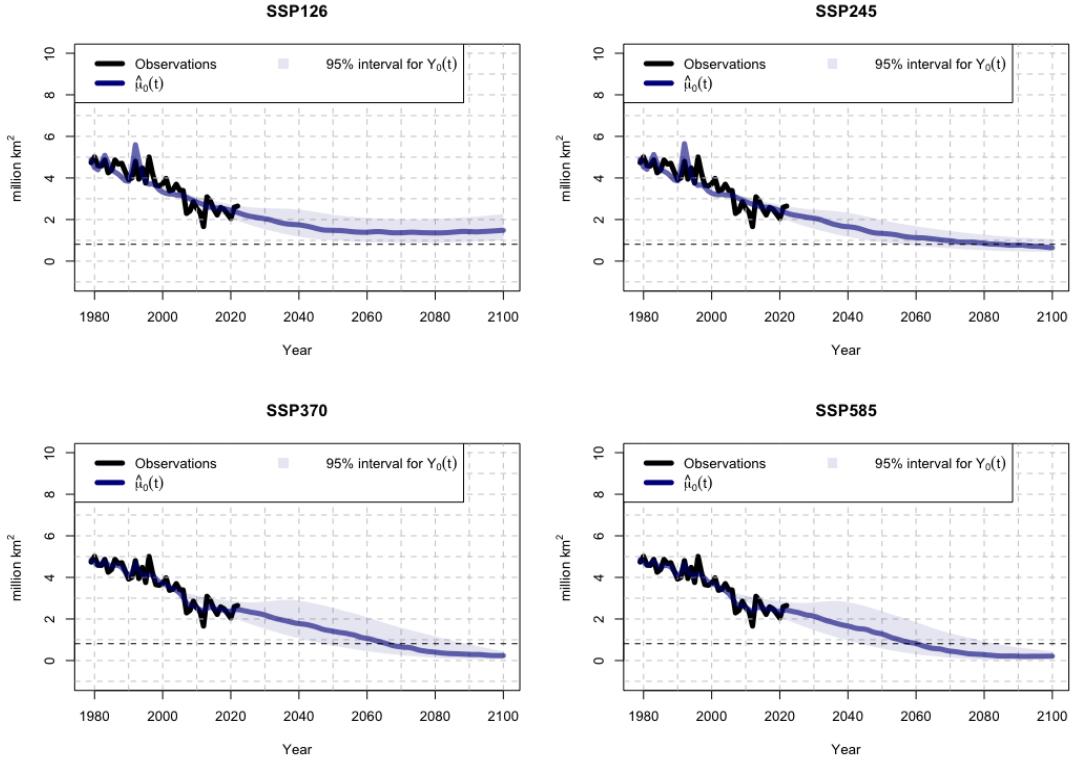


Figure 11: Post-processed ensembles of annual minimum Arctic sea ice extent, accounting for uncertainties in descriptor estimation. The dashed horizontal lines at 0.81 million  $km^2$  represent the 2 million  $km^2$  threshold, because in our analysis the pole hole area of 1.19 million  $km^2$  has been removed.

To further explore the associated uncertainty using the posterior samples, we explore the distribution of the year when annual minimum sea ice first falls below a certain threshold, say 2 million  $km^2$ . There are many alternative threshold values we can explore, but this threshold is chosen because it is approximately a half of the recently measured annual minimum data, and is substantially lower than the recorded minimum sea ice extent so far (3.61 million  $km^2$ ). In every posterior sample, we have both the sampled trend  $\mu_0$  and the sampled real-world measured sea ice  $Y_0$ . So we can get the years when the 1000 samples of underlying trend and real-world measurement first go below this threshold respectively. Figure 12 displays the empirical cumulative distributions of the years when the measured sea ice and the underlying trend respectively falls below 2 million  $km^2$ , obtained from the posterior samples.

The first plot in Figure 12 shows the cumulative probability of the year that the real-world measured annual minimum sea ice extent goes below 2 million  $km^2$ , which is likely to happen earlier and of higher probabilities in higher GHG emission scenarios. For example, in SSP370 and SSP585 development pathways, the measured Arctic sea ice would fall below 2 million  $km^2$  by the year 2080 almost surely, while this is less likely to happen with a posterior probability

of 0.8 and 0.7 if in SSP245 and SSP126 scenarios, respectively. The second plot shows the cumulative probability that the year when the underlying trend of sea ice extent meets the threshold, and we can see that under SSP126, the underlying trend is possible to remain above this threshold, as it is not reached by any posterior samples in this scenario. Also, according to the plots, the real-world measured sea ice is likely to approach the threshold earlier than does the underlying trend, because there are more variations in  $Y_0$  than in  $\mu_0$ .

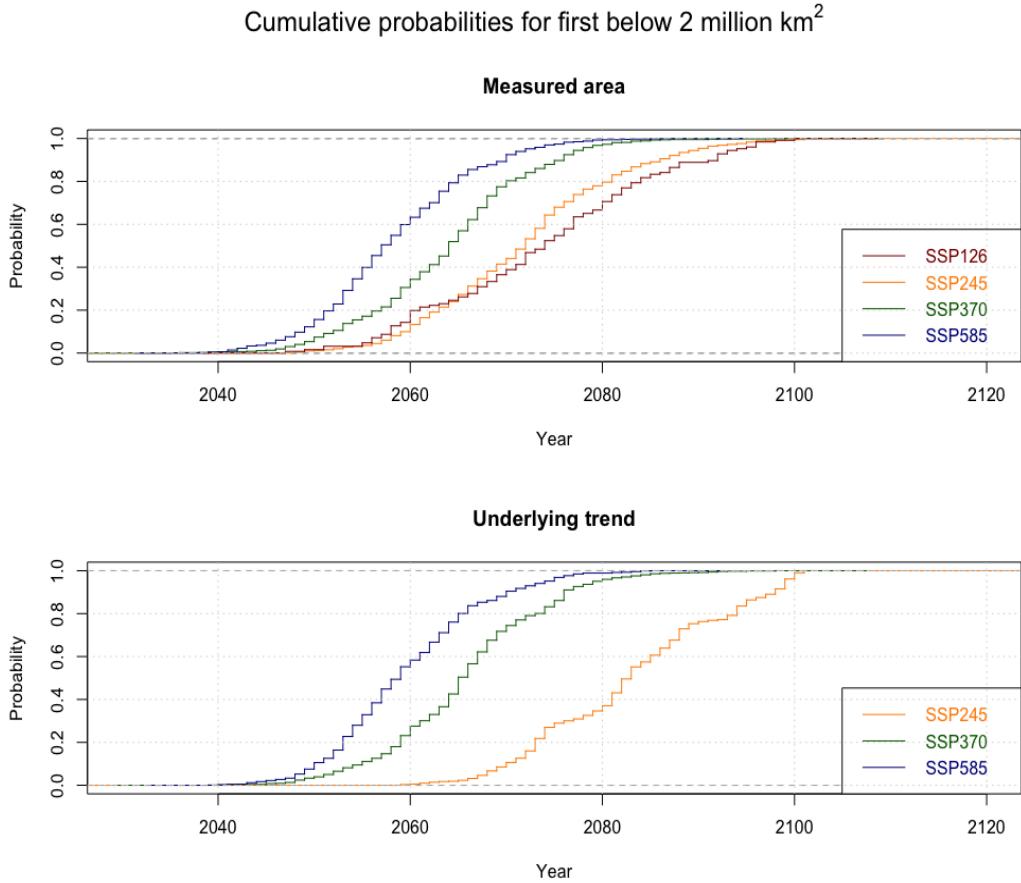


Figure 12: Cumulative probabilities for the year when real-world measured annual minimum sea ice extent (top plot) and its underlying trend (bottom plot) fall below 2 million  $km^2$ , under SSP experiments 126, 245, 370, and 585. These empirical cumulative probabilities are calculated by applying the step functions to the years when each posterior sample first attains the threshold. There is no SSP126 line in the bottom plot because no sampled trend falls below the threshold.

### 4.3 Summary of Chapter 4

In Chapter 4, we first discussed the choices of priors for the descriptors and state vector. Then we explored the fit results for the state-space model based on ensemble of simulations and measured annual minimum Arctic sea ice extent, and characterised the associated uncertainties.

- In Section 4.1, the plausible prior means and variances for the transformed descriptors are selected by considering the relative changes instead of absolute changes. The initial values

for the state vector and its initial covariance matrix arez decided by prior knowledge and reasoning.

- In Section 4.2, the estimated parameters and trend projections are discussed. The characteristics in the estimated trends in different SSPs can be explained by some features in the estimated values of the descriptors. Also, posterior sampling is applied to account for uncertainty of descriptors estimation. The sampling also help to obtain cumulative probabilities related to the year when annual minimum Artic sea ice extent falls below a threshold. Results suggest that in higher emission scenarios, the annual minimum sea ice extent is projected to decline rapidly with high probabilities to fall below 2 million  $km^2$  before 2080.

## 5 Conclusion and Discussion

### 5.1 Summary

In this report, we focused on the projection of annual minimum sea ice extent in the Arctic region and quantified associated uncertainties. We constructed an EBM-inspired state-space model to connect the trend with the forcings of GHG emissions and to integrate the ensemble of climate models outputs. It is inspired by the framework and methodology developed by Chandler et al. (2023), although the quantities of interest for analysis are different and projections under multiple socio-economic development pathways are explored as an extension.

Based on satellite measured sea ice extent from 1979 to 2022, the ensemble of CMIP6 simulators outputs from 1979 to 2100, and the four effective radiative forcings from 1979 to 2100, we fitted the constructed model and obtained trend projections under four socio-economic pathways. Associated uncertainties are quantified. The results suggested that the magnitude and speed of decline in annual minimum Arctic sea ice extent are positively related to the future GHG emissions. In pathways with high GHG emissions (SSP370 and SSP585), the minimum sea ice extent is projected to decline rapidly and fall below 2 million  $km^2$  by the year 2080 or earlier, with probability almost 1. Under a sustainable development pathway (i.e. SSP126) however, the minimal sea ice is possible to remain above this threshold by the end of this century.

### 5.2 Discussion and Future Work

There are some limitations in our report, and future work could be done for improvements. To begin with, the best approach we have at the moment to deal with non-negative data is log transformation. However, the selection of the small constant, which is used as the lower bound for log-transformation in equation (2.4.1), lacks sufficient justifications, although  $10^{-5}$  has been chosen because it does not bring significant distortion to the trends after log-transformation.

It has been observed that the model fitting results and uncertainty measurements are sensitive to the choices of this lower bound. For instance, the resulting differences in estimated state vectors could lead to various patterns in the projected trend. Hence, some alternative ways to handle non-negative data could be explored in future work to resolve the issue introduced by log transformation.

Another consideration is the assumption of normal distribution with constant variances for the error term in equation (3.1.1). This assumption may not be so appropriate, according to the diagnostic plots for the fitted real-world observations in Figure 13 in the Appendix A2, where some explanations are provided. Although the diagnostic plots suggest potential problems, it is probably not realistic to expect a statistical model that perfectly captures all the structures in over 100 time series, which is the number of series we have under every SSP experiment.

In addition, Figure 6 suggest that there might be a non-linear relationship between the effective radiative forcings and the annual minimum sea ice area. Although if we remove the two extreme values at year 1992 and 1983 (which correspond to volcanic eruptions), the remaining data points become arguably linear, some non-linear relationships can be explored to improve the model. Furthermore, the pole hole area removed in our report is relatively big, so that our projection of sea ice area is restricted to the regions outside the pole hole. But if we use a smaller pole hole, the historical data will be available from 1988, which will lead to a loss of almost 1/4 of annual records in the satellite measurements, and would arguably reduce the model fit performance. However, when more satellite measurements become available in the future, we may switch to a smaller pole hole to assist the sea ice projection.

# Appendices

## A1 Comments on CMIP6 model runs and decisions

Below is a summary of discarded CMIP6 simulators and the reasons (see Section 2.2.1).

- The simulator BCC-CSM2-MR is removed because the land-sea mask has errors in the Arctic region.
- The simulator BCC-ESM1 is removed because the land-sea mask has errors in the Arctic region.
- The simulator CNRM-CM6-1-HR is removed because of high resolution.
- The simulator HadGEM3-GC31-MM is removed because of high resolution.
- The simulator GFDL-CM4 is removed because of high resolution.
- The simulator CMCC-CM2-HR4 is removed because of high resolution.

Below is the summary of runs with irregular start and end dates (see Section 2.2.1), from data processing in R.

Historical runs
The EC-Earth-Consortium run EC-Earth3.r115i1p1f1.gn is omitted because data starts from 1979 but stops at year 2009
SSP126
The CAMS run CAMS-CSM1-0.r1i1p1f1.gn data start from 2015 but stops at year 2099
The CAMS run CAMS-CSM1-0.r2i1p1f1.gn data start from 2015 but stops at year 2099
SSP245
The CAMS run CAMS-CSM1-0.r1i1p1f1.gn data start from 2015 but stops at year 2099
The CAMS run CAMS-CSM1-0.r2i1p1f1.gn data start from 2015 but stops at year 2099
The EC-Earth-Consortium run EC-Earth3.r12i1p1f2.gn is omitted because data start from 2020
The EC-Earth-Consortium run EC-Earth3.r14i1p1f1.gn data start from 2015 but stops at year 2092
The EC-Earth-Consortium run EC-Earth3-Veg.r6i1p1f1.gn data start from 2015 but stops at year 2064
The MRI run MRI-ESM2-0.r1i3p1f1.gn data start from 2015 but stops at year 2050
The MRI run MRI-ESM2-0.r3i3p1f1.gn data start from 2015 but stops at year 2050
SSP370
The CAMS run CAMS-CSM1-0.r1i1p1f1.gn data start from 2015 but stops at year 2099
The CAMS run CAMS-CSM1-0.r2i1p1f1.gn data start from 2015 but stops at year 2099
The HAMMOZ-Consortium run MPI-ESM-1-2-HAM.r1i1p1f1.gn data start from 2015 but stops at year 2055
The HAMMOZ-Consortium run MPI-ESM-1-2-HAM.r2i1p1f1.gn data start from 2015 but stops at year 2055
The HAMMOZ-Consortium run MPI-ESM-1-2-HAM.r3i1p1f1.gn data start from 2015 but stops at year 2055
The NCAR run CESM2-WACCM.r2i1p1f1.gn data start from 2015 but stops at year 2055
The NCAR run CESM2-WACCM.r3i1p1f1.gn data start from 2015 but stops at year 2055
The NCC run NorESM2-LM.r2i1p1f1.gn data start from 2015 but stops at year 2054
The NCC run NorESM2-LM.r3i1p1f1.gn data start from 2015 but stops at year 2054
SSP585
The CAMS run CAMS-CSM1-0.r1i1p1f1.gn data starts from 2015 but stops at year 2099
The CAMS run CAMS-CSM1-0.r2i1p1f1.gn data starts from 2015 but stops at year 2099
The EC-Earth-Consortium run EC-Earth3.r113i1p1f1.gn data starts from 2015 but stops at year 2068

Table 5: Summary of irregular start and end dates in historical and SSP experiment runs.

## A2 Diagnostic plots

As mentioned in Section 5.2, we include some diagnostic plots to assess the model fitting to the real-world measurements series, and only SSP126 and 370 are presented here, as their diagnostic plots are similar to that of the other two. The first diagnostic plot is the time series plot of standardised residuals, and it checks our assumption that the standardised residuals follow normal distributions with zero mean and constant variance. The second plot presents the autocorrelations between standardised residuals. The last plot contains Ljung-Box statistics, and the p value checks the presence of autocorrelation up to lag k. From Figure 13 (a), we can see some trends in the residuals over time. From its ACF plot, autocorrelation is significant at lag 4, and p-value in the Ljung-Box test goes down to 0 at lag 4, which suggest some dependencies and autocorrelations between residuals at lag 4.

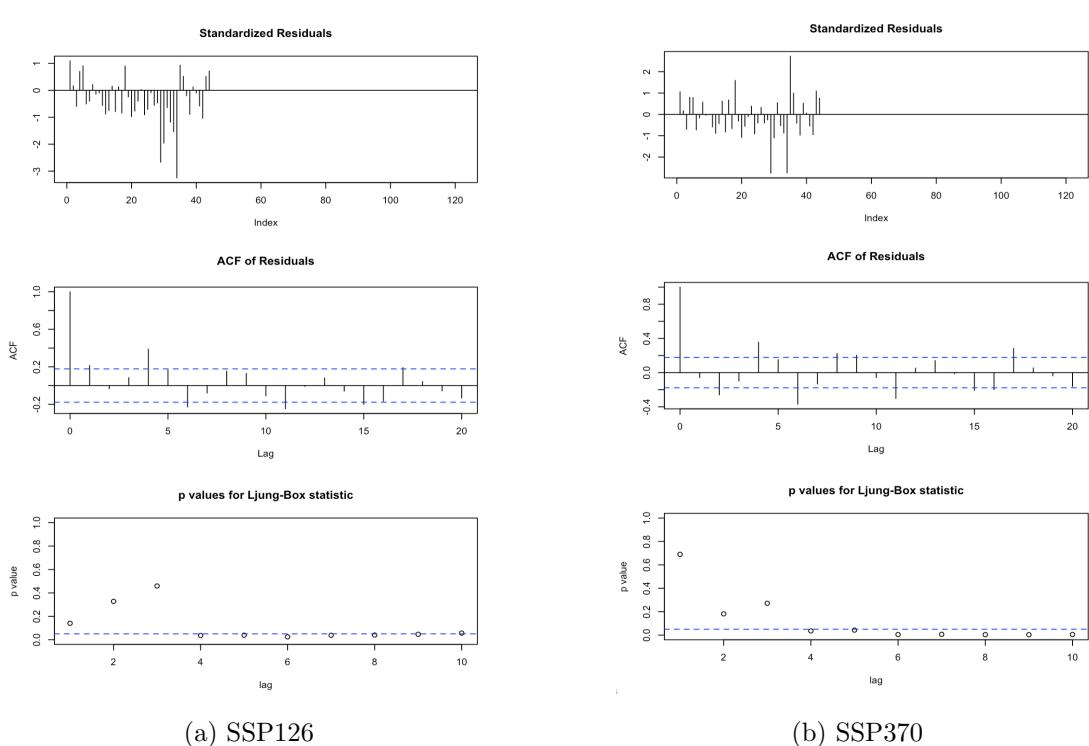


Figure 13: Diagnostic plots of model fit results for real-world measured annual minimum sea ice extent in experiments SSP126 and 370.

In plot (b), there are some significant autocorrelations at lags 2, 4, 6, ... that show oscillating pattern, and are hard to be explained by physical factors. There are no residual trends in SSP 370, probably because the forcing is stronger here, and the climate models can provide more precise information on the coefficients of the forcing in EBM, which is then inherited by the observations, compared to SSP126 and SSP245 where the forcings are weaker. Support for this explanation comes from the residual plots for SSP 585 (not shown), which show a similar absence of residual trend. Additionally, the ACF plots suggest that the lag 1 term in the EBM ( $\phi_0$ ) doesn't fully capture the memory in the system.

## References

- Barry, R. G. and Gan, T. Y. (2022). The future cryosphere: Impacts of global warming. In *The Global Cryosphere: Past, Present, and Future*, page 397–426. Cambridge University Press.
- Bates, J. R. (2007). Some considerations of the concept of climate feedback. *Quarterly journal of the Royal Meteorological Society*, 133(624):545–560.
- Boé, J., Hall, A., and Qu, X. (2009). September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience*, 2(5):341–343.
- Buser, C. M., Künsch, H., Lüthi, D., Wild, M., and Schär, C. (2009). Bayesian multi-model projection of climate: bias assumptions and interannual variability. *Climate Dynamics*, 33(6):849–868.
- Chandler, R. and Barnes, C. (2024). Working with Multiple Models. Invited chapter. In *Handbook of Statistical Methods for Computer Modeling*. (editors D. Bingham, M. Haran, B. Sanso, and D. Williamson).
- Chandler, R., Barnes, C., and Brierley, C. (2023). Decision-relevant characterisation of uncertainty in UK climate projections. Technical report, UK Climate Resilience Programme project CR20-3 Enabling the use and producing improved understanding of EuroCORDEX data over the UK. Available from <https://www.ucl.ac.uk/statistics/sites/statistics/files/projectionuncertainty.pdf>.
- Chandler, R. E. (2013). Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1991):20120388.
- Chandler, R. E. (2024). *TimSPEC: Time Series postprocessing of ensembles for climate*. R package version 0.0-7.
- Chandler, R. E. and Scott, E. M. (2011). *Statistical methods for trend detection and analysis in the environmental sciences / Richard E. Chandler, E. Marian Scott*. Statistics in practice. Wiley, Chichester.
- DiGirolamo, N. E., Parkinson, C. L., Cavalieri, D. J., Gloersen, P., and Zwally, H. J. (2022). Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 2. NASA National Snow and Ice Data Center Distributed Active Archive Center. Boulder, Colorado USA. Available at: <https://nsidc.org/data/NSIDC-0051/versions/2>.
- Durbin, J. R. and Koopman, S. J. (2001). *Time series analysis by state space methods / J. Durbin and S. J. Koopman*. Oxford statistical science series. Oxford Universiyy Press, Oxford.

- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.
- Francis, J. and Skific, N. (2015). Evidence linking rapid Arctic warming to mid-latitude weather patterns. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2045):20140170.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter / Andrew C. Harvey*. Cambridge University Press, Cambridge.
- Holland, M. M., Bitz, C. M., Tremblay, L.-B., and Bailey, D. A. (2008). The Role of Natural Versus Forced Change in Future Rapid Summer Arctic Ice Loss. In *Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications*, pages 133–150. American Geophysical Union (AGU).
- Ionita, M., Grosfeld, K., Scholz, P., Treffeisen, R., and Lohmann, G. (2019). September Arctic sea ice minimum prediction – a skillful new statistical approach. *Earth System Dynamics*, 10(1):189–203.
- IPCC (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *IPCC*, pages 1–180. Core Writing Team, H. Lee and J. Romero (eds.).
- Kovacs, K. M., Lydersen, C., Overland, J. E., and Moore, S. E. (2011). Impacts of changing sea-ice conditions on Arctic marine mammals. *Marine Biodiversity*, 41(1):181–194.
- NSIDC (2012). Arctic sea ice extent settles at record seasonal minimum. Technical report, National Snow and Ice Data Center. Arctic Sea Ice News and Analysis, Available at: <https://nsidc.org/arcticseainews/2012/09/>.
- O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., and van Vuuren, D. P. (2014). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, 122(3):387–400.
- O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9):3461–3482.
- Oshima, N., Yukimoto, S., Deushi, M., and Koshiro, T. (2020). Global and Arctic effective radiative forcing of anthropogenic gases and aerosols in MRI-ESM2.0. *Progress in Earth and Planetary Science*, 7(38). <https://doi.org/10.1186/s40645-020-00348-w>.

- Parkinson, C. L. (2022). Arctic sea ice coverage from 43 years of satellite passive-microwave observations. *Frontiers in Remote Sensing*, 3:1021781.
- Qasmi, S. and Ribes, A. (2022). Reducing uncertainty in local temperature projections. *Science Advances*, 8(41):eabo6872.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sansom, P. G., Stephenson, D. B., and Bracegirdle, T. J. (2021). On constraining projections of future climate using observations and simulations from multiple climate models. *Journal of the American Statistical Association*, 116(534):546–557.
- Serreze, M. C. and Stroeve, J. (2015). Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2045):20140159.
- Smith, C., Forster, P., Berger, S., Collins, W., Hall, B., Lunt, D., Palmer, M., Watanabe, M., Cain, M., Harris, G., Leach, N., Ringer, M., and Zelinka, M. (2021). Figure and data generation for Chapter 7 of the IPCC's Sixth Assessment Report, Working Group 1 (plus assorted other contributions). *Github*. Github repository, Version 1.0, Available at: <https://github.com/IPCC-WG1/Chapter-7/tree/v.1.0>.
- Stroeve, J., Barrett, A., Serreze, M., and Schweiger, A. (2014). Using records from submarine, aircraft and satellites to evaluate climate model simulations of Arctic sea ice thickness. *The Cryosphere*, 8:1839–1854.
- Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N. (2012). Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophysical Research Letters*, 39(16).
- Tang, Q., Zhang, X., and Francis, J. A. (2014). Extreme summer weather in northern mid-latitudes linked to a vanishing cryosphere. *Nature Climate Change*, 4(1):45–50.