Foundational basis for optimal climate change detection from energy-balance and cointegration models

Submitted by

Donald P. Cummins

to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, September 2021

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

. . . . . .

Donald P. Cummins

Abstract

This thesis has critically examined the validity of optimal fingerprinting methods for the detection and attribution (D&A) of climate change trends. The validity is called into question because optimal fingerprinting involves a linear regression of non-stationary time series. Such non-stationary regressions are in general statistically inconsistent, meaning they are liable to produce spurious results. This thesis has investigated, using an idealized linear-response-model framework motivated by energy-balance considerations, whether the standard assumptions of optimal fingerprinting are sufficient to guarantee consistency, and hence whether detected climate trends are likely to be genuine or artefacts of spurious correlation.

The principal reasoning tool in the thesis is the linear impulse-response model, familiar to many climatologists when parameterized as an energy-balance model (EBM), a simplified representation of global climate. A rigorous and efficient maximum likelihood method has been developed for estimating parameters of EBMs with any $k > 0$ number of boxes from CO2-quadrupling general circulation model (GCM) experiments and the method implemented as a free software package. It has been found that a three-box ocean is optimal for emulating the global mean surface temperature (GMST) impulse responses of GCMs in the Coupled Model Intercomparison Project Phase 5 (CMIP5).

A new linear-filtering method has also been developed for estimating historical effective radiative forcing (ERF) from time series of GMST. It has been shown that the response of any k-box EBM can be represented as an ARMA(k, $k-1$) autoregressive moving-average filter and that, by inverting the ARMA filter, time

series of surface temperature may be converted into radiative forcing. A comparison with an established method ("ERF_trans"), using historical simulations from HadGEM3-GC31-LL, found that the new method gives an ERF time series that closely matches published results (correlation of 0.83). Applying the new method to historical temperature observations, in combination with HadGEM3, produces evidence of a significant increase in ERF over the historical period with an estimated forcing in 2018 of 1.45 ± 0.504 W m^{-2}.

It has been proved, using an idealized linear-response-model framework where forcing is represented as an integrated process, that if standard assumptions hold then the optimal fingerprinting estimator is consistent, and hence robust against spurious regression. Hypothesis tests, conducted using historical GMST observations and simulation output from 13 GCMs of the CMIP6 generation, have produced no evidence that these assumptions are violated in practice. The historical trends in GMST which are detected and attributed using these GCMs are therefore very likely not spurious.

Consistency of the fingerprinting estimator was found to depend on "cointegration" between historical observations and GCM output. Detection of such a cointegration for the GMST variable indicates that the least-squares estimator is "superconsistent", with better convergence properties than might previously have been assumed. Furthermore, a new method has been developed for quantifying D&A uncertainty, which exploits the connection between cointegration and error-correction time series models to eliminate the need for pre-industrial control simulations.

Acknowledgements

I wish to thank Olivier Geoffroy for kindly providing the $4 \times CO_2$ climate model datasets used in Chapter 3, and Timothy Andrews of the Met Office Hadley Centre for providing the HadGEM3-GC31-LL forcing data used in Chapter 4. I acknowledge the World Climate Research Program's Working Group on Coupled Modelling, which is responsible for CMIP, and I thank the climate modeling groups (listed in Table 3.2) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

I am grateful to Stefan Siegert for helpful discussions regarding the Kalman filter, and to Philip G. Sansom for kindly sharing his R script which was used for batch processing of CMIP6 gridded data.

I would like to thank my parents and Kenneth J. Falconer at the University of St Andrews for encouraging me to pursue a PhD. And last, but not least, I wish to thank Peter A. Stott and David B. Stephenson at the University of Exeter, whose advice and encouragements over these years have been invaluable. I consider myself very fortunate to have had them as my supervisors.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Detection and attribution of climate change

Statistical methods for detection and attribution (D&A) of climate change trends have been widely used in climate change studies over the last two decades, and the resulting inferences have informed assessment reports from the Intergovernmental Panel on Climate Change (IPCC) (Hegerl et al., 2007; Bindoff et al., 2013; Eyring et al., 2021). Formal D&A studies commonly employ some variant of the method known as "optimal fingerprinting", introduced by Hasselmann (1979, 1997). Optimal fingerprinting frames D&A as the problem of separating the forced component of historical climate observations (i.e. the signal) from internal climate variability (the noise) (Hegerl and Zwiers, 2011). In practice, this separation of signal and noise is performed by projecting climate observations onto corresponding simulation output from general circulation models (GCMs), in a procedure analogous to a multivariate linear regression (Allen and Tett, 1999). Optimal fingerprinting assumes a regression model of the form

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{1.1}$$

where $\boldsymbol{y}$ denotes historical climate observations; X is a matrix of predicted climate-change signals, typically consisting of simulation output from a GCM; and $\boldsymbol{e}$ is a composite error term containing internal climate variability noise as well as other

sources of uncertainty. Regression coefficients $\boldsymbol{\beta}$ are known in D&A as "scaling factors". Detection and attribution inferences depend on obtaining reliable estimates of these scaling factors and establishing their statistical significance. Over the years, fingerprinting methods have become increasingly sophisticated due to a succession of proposed refinements (see Chapter 2 for a review of methodological developments).

While originally conceived as a multivariate method for use with gridded spatio-temporal datasets, typically requiring application of dimension-reduction techniques, simplified variants of optimal fingerprinting have more recently been applied to time series data, specifically to observations of global mean surface temperature (GMST) (Otto et al., 2015; Rypdal, 2015; Haustein et al., 2017). Global mean surface temperature is an important climate variable, both as a predictor of changes in local climate (Sutton et al., 2015), and as the metric of global warming used in communication with policymakers, e.g. the 1.5 and 2.0 degrees Celsius targets of the 2015 Paris Agreement. In their sixth assessment report, the IPCC make extensive use of observationally constrained GMST estimates for their global and regional climate change projections (IPCC, 2021). There is also evidence that the additivity assumption (see Chapters 2 and 5) implicit in optimal fingerprinting is more likely to hold for GMST than for other variables such as precipitation (Good et al., 2011). For these reasons, GMST is the climate variable of primary interest in this thesis.

1.2 Climate models in detection and attribution

The field of D&A is above all motivated by the need to understand how the Earth's climate responds to external "forcing factors". Detection and attribution methodologies like optimal fingerprinting rely on scientists' ability to simulate the climate system's response under different forcing scenarios, with the most powerful tool for this purpose being the GCM (Hegerl and Zwiers, 2011). Within the hierarchy of GCMs, the GCM variants used in D&A of climate change trends are often the most complicated and hence computationally demanding, featuring fully coupled oceans (AOGCMs), and cryo- and biosphere components (so-called Earth system

models or ESMs). An interesting property of the climate models from the various modelling centres around the world (see Table 3.2) is the fact that, despite sharing ancestors and model components, and despite having a common basis in the laws of physics, there exists great heterogeneity in how the models respond to external perturbations. This is particularly true in the case of the GMST climate variable: equilibrium climate sensitivity (ECS), the predicted increase in GMST associated with a doubling of atmospheric carbon dioxide concentration, varies from 2.1 K to 4.6 K across the Coupled Model Intercomparison Project Phase 5 (CMIP5) generation of GCMs (Flato et al., 2013). Representing or "capturing" the diversity of GCM behaviours with respect to GMST, through the calibration of simple climate models, is a key research theme of this thesis, and forms the basis of the material presented in Chapters 3 and 4 (discussed briefly below). The principal motivation for emulating GCMs in this way is to obtain idealized response functions, which tell us how surface temperature in a GCM would respond under a particular forcing scenario, without having to incur the computational costs of running a full simulation. Ideally, this approach would provide the best of both worlds: all of our physical knowledge goes into deriving the GCMs and hence the response functions, which can then be used extensively at modest computational expense. Whether and how far this idea holds in practice is tested in Chapter 4.

1.3 Thesis aims and questions

1.3.1 Aims

This thesis is motivated by an apparent failure in the canonical D&A literature (reviewed in Chapter 2) to explicitly recognize the time-indexed nature of climate datasets, an omission with potentially catastrophic implications. It has long been known in the field of time series analysis that the practice of regressing variables containing time trends comes with its own set of pitfalls, the most serious of which being the "spurious regression" phenomenon, whereby statistical significance is erro-

neously assigned to chance correlations between unrelated trending variables (Yule, 1926). The primary aim of this thesis is to establish a foundational basis for the validity of D&A, using a combination of theoretical argument and empirical evidence, and hence assess the threat to D&A inferences posed by spurious regression.

Addressing this primary aim has also led to the formulation of secondary research aims in the thesis. Early in the research process it was realized that an idealized, analytically tractable representation of climate would be a valuable tool for reasoning about D&A. What was needed was a "simple" climate model: simple enough for its macroscopic behaviour to permit description via manipulable analytical formulas; but sufficiently flexible to faithfully capture the dynamics of GCMs' GMST responses. The class of simple climate model chosen for this purpose was the k-box energy-balance model (EBM). Thus two secondary research aims emerged. The first was to determine if and how this class of model could be fitted to GCMs in a statistically rigorous way. The second was to identify the degree of model complexity (indexed by number of boxes k) required to optimally emulate the GMST responses of modern GCMs.

Further secondary research aims arose from the work on statistical estimation of EBMs. It was realized that the linear time-invariance property, which underlies its analytical tractability, allows the k-box model to be related to the wider class of linear time series models known as ARMA or autoregressive moving-average models. Finding the ARMA representation of an EBM would have the advantage of allowing dependence of GMST on unobserved deep-ocean temperatures to be substituted for dependence on past values of GMST. Thus, for the purpose of computation, GMST could be considered as a separate stochastic process in its own right. Of particular relevance to this thesis is the fact the ARMA representation of an EBM permits a straightforward inversion, allowing (a noisy estimate of) radiative forcing to computed recursively from only a time series of GMST. Three research aims were formulated: to show that a k-box EBM's ARMA representation exists and determine its order; to investigate the ARMA filter's performance as an estimator of historical

radiative forcing in GCM simulations; and to construct estimates of radiative forcing in the real world from time series of historical GMST observations.

1.3.2 Questions

This thesis addresses the following questions:

1. Is the optimal fingerprinting estimator statistically consistent under standard D&A assumptions and hence robust against spurious regression?

2. How likely is it that these assumptions hold in practice for the current generation of climate models?

3. Is it possible to formulate a new method for attributing observed changes in global temperatures to anthropogenic greenhouse gas emissions including uncertainties?

Secondary research questions include:

1. How might maximum likelihood estimates of physical parameters of k-box stochastic energy-balance models (EBMs) be obtained from surface temperatures and top-of-atmosphere net radiative fluxes in GCM experiments?

2. What number of boxes $k > 0$ represents the optimal EBM complexity for emulating temperature responses of modern GCMs?

3. Does a k-box EBM have a unique representation as an autoregressive moving-average (ARMA) filter and, if so, what are the respective degrees of its AR and MA polynomials?

4. How effective is the inverted ARMA representation of a GCM-calibrated EBM, applied to time series of surface temperature, as an estimator of effective radiative forcing?

1.4 Thesis structure

The next chapter will present a review of the methodological D&A literature, with a focus on the development of optimal fingerprinting methods. Chapter 3 will then address secondary research questions 1 and 2. Chapter 4 will address secondary research questions 3 and 4. The primary research questions of this thesis will be addressed in Chapter 5, whose line of argument exploits results developed in the previous two chapters.

Chapter 2

Review of detection and attribution literature

2.1 Aim

This chapter reviews the development of statistical methods for detection and attribution (D&A) of climate change. Most of the chapter focuses on the model-based, regression approach known as "optimal fingerprinting", which is the most common method in applied D&A studies.

2.2 Introduction

Detection and attribution of climate change requires a decomposition of observed climate data into externally forced and residual components, where the residual component represents unforced or "natural" variability. Various methods have been developed for performing this decomposition using purely observational data (Hegerl and Zwiers, 2011). Observation-based methods include empirical spatial pattern decompositions (e.g. Kawamura, 1994; Wallace et al., 1996; Portmann et al., 2009; Thompson et al., 2009) and empirical time series decompositions (e.g. Thompson and Wallace, 1998; Venzke et al., 1999; Schneider and Held, 2001; Solomon et al., 2011). Use of climate models in D&A is however well established, due to important

advantages over purely observational approaches. The main problem with relying on observations alone is that the observational record is, in effect, a single realization of the noisy climate system. From this single realization we hope to infer the qualitative properties (spatial and/or temporal patterns) of the forced signal and natural variability, as well as their respective contributions to the observational record. Climate models offer a form of regularization to this otherwise ill-posed problem, by constraining estimated climate change signals so that they are consistent with physical knowledge.

This chapter will review the development of the model-based methodology "optimal fingerprinting", an approach combining historical observations with climate model simulations, which has been widely used in D&A studies and incrementally refined over the last four decades. Optimal fingerprinting has also become the basis of the more recent field of extreme event attribution (Stott et al., 2016), however in the following review attention is restricted to the original application to climate "trend" attribution.

2.3 The optimal fingerprinting methodology

Optimal fingerprinting, also known as optimal detection, is a model-based approach to D&A of climate change, first proposed by Hasselmann (1979) and extended in Hasselmann (1997). The central assumption of optimal fingerprinting is that a vector (spatial and/or temporal) of observed climate data $\boldsymbol{y}$ can be represented as the sum of a forced climate change signal $\boldsymbol{\psi}$ and an internal climate variability (residual) component $\boldsymbol{e}$,

$$\boldsymbol{y} = \boldsymbol{\psi} + \boldsymbol{e}. \tag{2.1}$$

By estimating the forced climate change signal $\boldsymbol{\psi}$ and internal climate variability covariance $\Sigma = \mathrm{cov}(\boldsymbol{e})$, typically using a climate model, linear regression may be used to test the null hypothesis, $H_0 : \boldsymbol{\psi} = \mathbf{0}$, that an apparent climate change is consistent with internal variability alone. Rejection of the null hypothesis indicates

that a climate change is detected.

The second assumption of optimal fingerprinting is that the forced climate change signal $\boldsymbol{\psi}$ is simply the sum of forced signals $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_p$ corresponding to each of p candidate forcing factors, as if each forcing had been applied individually. A multiple regression of the observed climate data $\boldsymbol{y}$ on model-predicted climate change signals $\hat{\boldsymbol{\psi}}_1, \ldots, \hat{\boldsymbol{\psi}}_p$ can then be used to perform attribution hypothesis tests. The multiple regression model is

$$\boldsymbol{y} = \beta_1 \hat{\boldsymbol{\psi}}_1 + \cdots + \beta_p \hat{\boldsymbol{\psi}}_p + \boldsymbol{e}. \tag{2.2}$$

If, for a given forcing factor i, the scaling factor β_i is found not to be significantly different from one, then we say that model-predicted signal $\hat{\boldsymbol{\psi}}_i$ is consistent with observation. The optimal fingerprinting method thus yields formal D&A results at a specified level of statistical significance.

In an early application of optimal fingerprinting, Hegerl et al. (1996) attempted to detect greenhouse-gas-induced climate change. They found that applying the method to available data raised practical problems which needed to be addressed. Predicted climate change signals and internal climate variability estimates were obtained using simulations from general circulation models (GCMs). The estimated natural variability covariance matrix $\hat{\Sigma}$ was calculated from GCM pre-industrial control simulations. Due to computational constraints, the control simulations were not of adequate length to estimate the full internal variability covariance matrix. In order to obtain a full-rank sample covariance matrix, the dimension of the data space was reduced using principal component analysis (PCA). After normalizing all data by the estimated inverse natural variability covariance matrix $\hat{\Sigma}^{-1}$, the regression model was fitted using ordinary least squares (OLS). Statistically significant climate change in the 30-year period prior to 1996 was detected, however attribution was not possible due to the omission of some potentially influential external forcing factors. In a follow-up paper, Hegerl et al. (1997) conducted a full D&A study, in which multiple candidate forcing factors were considered. The dimension reduction was a source of concern in both studies, as the number of retained principal components or

empirical orthogonal functions (EOFs) was found to affect the resulting inference. An ad-hoc method, based on visual inspection of diagnostic graphs, was used in practice to determine the number of retained EOFs.

Allen and Tett (1999) more rigorously established optimal fingerprinting as a form of generalised linear regression. They highlighted a number of issues with the optimal fingerprinting algorithm and proposed an objective method for deciding how many EOFs to retain. Firstly, they pointed out that a single covariance matrix could not be used both for the fitting of the regression model and for the subsequent hypothesis testing, otherwise bias would be introduced due to overfitting. Secondly, the GCMs used in 1999 were known to underestimate internal climate variability on small spatial scales. An example was presented using temperature data from HadCM2. The influence of such underestimation on study results was magnified by the use of PCA for dimension reduction and subsequent inversion of the estimated covariance matrix. Allen and Tett (1999) proposed a chi-squared test to establish an appropriate dimension for the reduced space. Thirdly, a problem was identified with the way in which confidence regions for regression parameters were calculated. Specifically it was difficult to estimate degrees of freedom for an F-statistic using short GCM control simulations. Allen and Tett (1999) advocated using non-parametric confidence regions based on return times, sidestepping parametric assumptions such as multivariate normality. A further innovation was their use of conventional matrix notation to describe statistical models, as opposed to the notation from quantum mechanics and general relativity used in Hasselmann's earlier optimal fingerprinting papers.

A major improvement to the optimal fingerprinting algorithm was made by Allen and Stott (2003), who found that an important source of uncertainty in the regression step had been overlooked. Predicted climate change signals in optimal fingerprinting are typically estimated by running ensembles of GCM simulations and taking the ensemble mean of the output data. In earlier studies, the ensemble mean was assumed to be deterministic, i.e. without noise, and OLS was used to fit the

regression model. However, due to the small sizes of GCM ensembles in practice, ensemble means are still subject to considerable sampling uncertainty. Regression coefficients estimated using OLS are biased low as a result. Allen and Stott (2003) proposed the use of total least squares (TLS) instead of OLS regression. Under the assumption that internal climate variability in the observations is equal to the sampling uncertainty across ensemble members, TLS delivers unbiased estimates of regression coefficients. The authors drew attention to the potential value of D&A studies as a tool for diagnosing GCM performance. The multiple regression step in an attribution study can indicate whether a GCM is underestimating or overestimating the influence of one or more external forcing factors on climate.

These and other earlier optimal fingerprinting studies relied on the frequentist principles of hypothesis testing. Lee et al. (2005) experimented with a hybrid frequentist-Bayesian implementation of optimal fingerprinting. They fitted the standard multiple regression model to climate observations and GCM simulations using OLS but then passed the estimated regression coefficients into a pseudo-Bayesian inference. Prior distributions on the regression coefficients were chosen using output from energy-balance models (EBMs). A likelihood function for the regression coefficients was approximated by the probability density function (pdf) of the asymptotic sampling distribution of the OLS maximum likelihood estimators (MLEs). Feeding the EBM-derived priors and approximate likelihood into Bayes' theorem produced posterior distributions on the regression coefficients. A range of priors were considered in a prior sensitivity analysis. Truncated priors, excluding regression coefficients less than minus one or greater than two, were used in the final analysis. Results were presented in terms of Bayes factors providing inference roughly analogous to frequentist hypothesis tests. Lee et al. (2005) was an early attempt to incorporate prior insight into an attribution analysis.

Huntingford et al. (2006) further developed the TLS implementation of optimal fingerprinting. Allen and Stott (2003) had implicitly assumed that sampling variation across simulations from the same GCM had the same covariance structure as

variation across simulations from different GCMs. Huntingford et al. (2006) considered a generalisation of the regression model where this assumption was relaxed and inter-GCM variance was explicitly included. Accounting for inter-GCM variance explicitly allowed data from multiple GCMs to be used in the same analysis. The greater volume of data employed led to a reduction in uncertainty and smaller confidence intervals about estimated regression coefficients. Huntingford et al. (2006) raised the question of how to weight simulations from different GCMs. They reasoned that, since GCMs from different modelling centres provide alternative but equally plausible representations of reality, the ensembles should all carry equal weighting. Therefore the number of ensemble members from each GCM was not allowed to influence model weighting in their optimal fingerprinting analysis.

The chi-squared test for choosing a suitable dimension reduction was widely used in D&A studies over the following decade. Ribes et al. (2009) showed that the dimension reduction step, formerly necessary for estimating internal climate variability, could be avoided entirely using a new method called regularized optimal fingerprinting (ROF). Regularized optimal fingerprinting uses the Ledoit-Wolf well-conditioned estimator for high-dimensional covariance matrices to estimate the internal climate variability. Ledoit-Wolf is a shrinkage estimator which uses the identity matrix as a target. A simulation study by Ribes et al. (2009) showed that ROF estimators of regression coefficients were subject to a lower mean square error (MSE) than previous methods. Ribes et al. (2009) were interested in D&A of climate change on regional and sub-regional scales. On small spatial scales, GCM-derived estimates of internal climate variability were known to be poor and such GCM deficiences are magnified by the use of PCA followed by inversion of the covariance matrix. The ROF method allows estimation of internal climate variability using early-20th-century observations due to its reduced requirement for many control data. Less computation is necessary and the problem of how many EOFs to retain is avoided.

Ribes et al. (2013) conducted a more thorough evaluation of the ROF method. They explained that the identity matrix used as a target for the shrinkage estimator

may be interpreted as a Bayesian prior on the internal climate variability. Ribes et al. (2013) discussed the possibility of using physical knowledge about internal climate variability to choose a more suitable matrix than simply the identity. It was also pointed out that some dimension reduction might still be necessary since Ledoit-Wolf still requires the sample covariance matrix to have full rank. A small dimension reduction was shown to have little impact on D&A results.

The extension of the optimal fingerprinting regression model, proposed by Huntingford et al. (2006), was revived by Hannart et al. (2014). They advocated a further extension of the model whereby each source of uncertainty considered would have its own covariance structure, including observational error which was previously unaccounted for. The extended model belongs to the class of models known as errors in variables (EIV), however the TLS procedure of Allen and Stott (2003) is unable to fit the extended model. Hannart et al. (2014) developed a so-called partial iterative maximisation approach where partial likelihoods for each model parameter are maximised in turn until an overall maximum is reached. The new approach was shown to perform better than TLS and OLS in a simulation study. The simulation study revealed a vulnerability common to all regression methods tested: asymptotic results used to generate confidence regions were shown to fail for small sample sizes, with the severity of the failure aggravated by a lower signal-to-noise ratio.

Hannart (2016) introduced a new method called integrated optimal fingerprinting, a generalisation of ROF. Integrated optimal fingerprinting replaces the three steps of the traditional fingerprinting algorithm with a single estimation step. The internal climate variability covariance matrix is treated as a nuisance parameter and removed by integration. Point estimates and confidence intervals are obtained from the integrated likelihood. Hannart (2016) explained that integrated optimal fingerprinting is equivalent to an empirical Bayesian analysis. No preliminary dimension reduction is necessary. Furthermore, the authors showed in a simulation study that performing a dimension reduction led to reduced precision of regression coefficient estimates. Performance of integrated optimal fingerprinting was shown

to exceed that of both traditional optimal fingerprinting and ROF. However, confidence region coverage probabilities were still vulnerable to small sample sizes and low signal-to-noise ratio.

Jones et al. (2016) carried out a multi-model D&A study using GCM simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5). Their results facilitated a detailed discussion of the performance of D&A methodology. The authors concluded that optimal fingerprinting is robust to gross errors in the magnitude of GCM-predicted climate change signals. They concluded that examining net anthropogenic influence gives consistent results across different GCMs. However, partitioning of anthropogenic influence into individual forcing factors was found to produce inconsistent results across different GCMs. The (still widely used) TLS fingerprinting methodology was found to be unable to reconcile differences in the spatial/spatio-temporal structure of climate response patterns. Jones et al. (2016) recommended that explicit modelling of inter-GCM variation, previously treated in Huntingford et al. (2006) and Hannart et al. (2014), should be examined further.

The first fully Bayesian implementation of optimal fingerprinting was published by Katzfuss et al. (2017), who formulated a hierarchical Bayesian model based on the model introduced by Allen and Stott (2003). Uninformative prior distributions were used to account for uncertainty in observations, GCM simulations and the internal climate variability covariance matrix, including the number of retained EOFs. Bayesian model averaging applies probabilistic weights to different dimension reductions. The authors produced a custom software package for the practical application of their method. Unlike the more recent frequentist implementations of optimal fingerprinting, the method proposed by Katzfuss et al. (2017) still relies on a dimension reduction and suffers from the associated loss in precision of regression coefficient estimates.

Ribes et al. (2017) proposed a more symmetric treatment of uncertainty in amplitudes and patterns of forced climate signals. The authors noted that the widely used TLS optimal fingerprinting regression model is appropriate only when forced signal

amplitude is the dominant source of uncertainty, i.e. the signal patterns themselves are largely known and do not vary between climate models. More sophisticated EIV approaches, e.g. Hannart (2016), which account for pattern uncertainty do not in general yield closed-form solutions and require the use of iterative procedures. It was also noted that the regression formulation of optimal fingerprinting considers signal amplitudes as free parameters, when in reality signal amplitudes are constrained by physical knowledge. Ribes et al. (2017) developed a new model, fitted using closed-form maximum likelihood estimation, whereby the scaling factors are removed from the EIV fingerprinting model. Removal of the scaling factors anchors signal amplitudes close to their model-predicted values. Closed-form expressions were also provided for D&A hypothesis testing.

Hannart (2019) considered the problem of dimension reduction which has existed since the early D&A studies. Optimal fingerprinting regression is typically performed in the reduced space spanned by a few leading eigenvectors (EOFs) of the natural variability covariance matrix Σ. This is in part since only the leading eigenvectors of Σ can be reliably estimated from limited GCM control runs. The authors noted that this reduced subspace maximizes the noise amplitude leading to an increase in uncertainty. They therefore proposed a new estimator of scaling factors $\boldsymbol{\beta}$ whereby the data are projected onto a new subspace which is orthogonal to the retained EOFs. This projection gives a better signal-to-noise ratio and reduced uncertainty in parameter estimates. The authors noted that, when a very small number of EOFs are used, their estimator has a MSE orders of magnitude lower than the conventional procedure.

2.4 Summary

Statistical methods for D&A of climate change can be divided into two categories: purely observational and model-based. The model-based methodology known as optimal fingerprinting has become the basis of many D&A studies. Optimal fingerprinting uses generalized linear regression and hypothesis testing to make formal

statements about the significance of, and causal factors behind, observed changes in climate. Physical knowledge encoded in general circulation models is used to constrain predicted climate change signals and estimate properties of natural climate variability. Inferences are drawn from a regression of observations on the predicted climate change signals.

Due to its reliance on small numbers of expensive climate model simulations and the short (and often sparse) historical record, optimal fingerprinting presents a number of technical challenges. For example, dimension reduction techniques are often used to make linear regression feasible. Such dimension reductions are known to influence results in ways which are not always predictable. Secondly, predicted climate change signals are typically subject to sampling uncertainty, stemming from perturbation of ensemble initial conditions and also inter-climate model variation. Sophisticated total least squares and errors in variables models have been developed to take into account uncertainty in predicted signals. More recently, methods based on regularized shrinkage estimators, integrated likelihood models and hierarchical Bayesian inference have been introduced to synthesize multiple sources of uncertainty.

A particular concern noted in recent studies is the apparent difficulty with which optimal fingerprinting, as a regression procedure, accounts for qualitative differences in predicted climate change signals from different climate models. Specifically, optimal fingerprinting can easily reconcile disagreements between climate models in the expected magnitude of responses to forcing factors, but not disagreements in the shape of the (spatial and/or temporal) response patterns. This is a natural consequence of the fact that observations and model simulation output are treated as static signals in the regression, rather than outputs of an input-output system evolving over time. Given that reponse patterns differ in shape between pairs of climate models, as well as between models and observations, this raises questions about what the scaling factors are supposed to measure in practice, and what exactly optimal fingerprinting is attempting to estimate.

Finally, it has been noted that the methodological studies reviewed in this section have not acknowledged certain properties which are specific to time-series data, or assessed the potential threat to D&A studies posed by the statistical phenomenon known as "spurious regression" (see Chapter 5). For example, in optimal fingerprinting the spatial and temporal correlations between variables are treated similarly through the use of combined space-time covariance matrices. A weakness of this approach is that notions pertaining to time, such as ordinality and causality, are not given adequate consideration. Furthermore, it is a long-accepted principle of time series analysis that regressing variables containing time trends is inherently more dangerous than static or "cross-sectional" regression, mainly due to the additional threat of spurious regression (Yule, 1926).

Chapter 3

Stochastic energy-balance models and parameter estimation

3.1 Aim

This chapter is based on material from Cummins et al. (2020b) and introduces the k-box stochastic energy-balance model as a tool for parameterizing global mean surface temperature in general circulation model simulations. A rigorous maximum likelihood fitting procedure for this class of model is developed, with an accompanying software implementation.

3.2 Introduction

The previous chapter reviewed the development of statistical methods for D&A of historical climate trends, focusing on the commonly used family of methods derived from Klaus Hasselmann's "optimal fingerprinting" algorithm (Hasselmann, 1979). Fingerprinting methods are based on a linear regression model

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e} \tag{3.1}$$

relating historical climate observations $\boldsymbol{y}$ to climate model output X (Hegerl and Zwiers, 2011). In the case of time series data, vector $\boldsymbol{y}$ is a sequence of historical observations at consecutive time steps, while each column of X is a corresponding sequence of simulation output from a forced climate model run (or the mean of an ensemble of such runs). The number of rows in X is the number of observations, and the number of columns in X is the number of candidate forcing factors considered. In the case of spatio-temporal data, where a spatial field is observed at each timestep, eqn (3.1) is obtained by applying the vec operator to observed and simulated datasets at each timestep. Optimal fingerprinting methods vary in how they estimate the so-called "scaling factors" $\boldsymbol{\beta}$ as well as how they parameterize the covariance of the error term $\boldsymbol{e}$, but the assumption of a linear relationship between $\boldsymbol{y}$ and X is a defining feature of the methodology common to all its implementations. The principal aim of this thesis (see Chapter 1) is to assess the reliability of optimal fingerprinting methods for detection and attribution (D&A): to determine necessary conditions for inferences to be valid; as well as to assess what evidence there is for such conditions holding in practice. Unfortunately, the terms in equation (3.1) are not amenable to analytic manipulation in their raw form. The exact underlying equations which give rise to true climate response $\boldsymbol{y}$ are unknown. While the general circulation models (GCMs), intended to approximate reality and typically used to produce the predicted signals X, are based on well-defined systems of partial differential equations (PDEs), these equations lack explicit solutions and are solved numerically on supercomputers. In order to reason mathematically about the regression model in equation (3.1), some form of simplification or parameterization is required. In this chapter, an example of a simplifying parameterization is introduced: the k-box stochastic energy-balance model.

3.2.1 Energy-balance models

An energy-balance model (EBM) is a simplified representation of climate where changes in global temperature are explained by imbalances in the Earth's energy

budget. Energy-balance models are simpler than atmosphere-ocean general circulation models (AOGCMs), which explicitly describe the fluid dynamics of Earth's atmosphere and oceans. Their simplicity means that EBMs are both analytically tractable and inexpensive to simulate. Compared with purely empirical statistical models, EBMs have two distinct advantages: (i) the choice of model structure is motivated by physical reasoning; (ii) certain model parameters and functions thereof have physical interpretability. Energy-balance models are therefore useful not only for climate forecasting but for physical inference about the climate system. As will be seen in Chapter 4, EBMs can be used as standalone tools for analysing time series of historical climate observations.

Energy-balance models in the literature vary in complexity. The class of EBM considered in this chapter is the k-box model (sometimes called k-layer), which represents the atmosphere and ocean as a set of vertically stacked boxes. The simplest k-box model is the so-called one-box model, obtained by a linearization of the zero-dimensional variant of the Budyko-Sellers model (Budyko, 1969; Sellers, 1969). The standard Budyko-Sellers EBM is based on the Stefan-Boltzmann Law, which relates outgoing radiation to the fourth power of surface temperature. Linearization of this model about pre-industrial surface temperature is a suitable approximation for small surface temperature perturbations. The one-box model continues to be used in climate change studies (e.g. Cox et al., 2018), but is known to insufficiently capture thermal inertia in the climate response on timescales of decades or more, and has been superseded in longer-timescale studies by the two-box model (Gregory, 2000; Held et al., 2010; Geoffroy et al., 2013a). The two-box model includes a representation of the deep ocean, which acts as a heat reservoir and provides the thermal inertia. Some recent studies have employed three-box models (Caldeira and Myhrvold, 2013; Proistosescu and Huybers, 2017; Fredriksen and Rypdal, 2017; Tsutsui, 2020). By taking the limit as $k \to \infty$ it is possible to approximate continuous vertical heat diffusion.

3.2.2 The stochastic k-box energy-balance model

The k-box energy-balance model used in this chapter is defined by the system of k linear differential equations

$$C_1 \frac{\mathrm{d}T_1}{\mathrm{d}t} = F(t) - \kappa_1 T_1 - \kappa_2(T_1 - T_2) + \xi(t) \tag{3.2}$$

$$C_2 \frac{\mathrm{d}T_2}{\mathrm{d}t} = \kappa_2(T_1 - T_2) - \kappa_3(T_2 - T_3) \tag{3.3}$$

$$\vdots$$

$$C_{k-1} \frac{\mathrm{d}T_{k-1}}{\mathrm{d}t} = \kappa_{k-1}(T_{k-2} - T_{k-1}) - \varepsilon\kappa_k(T_{k-1} - T_k) \tag{3.4}$$

$$C_k \frac{\mathrm{d}T_k}{\mathrm{d}t} = \kappa_k(T_{k-1} - T_k). \tag{3.5}$$

The first box represents the atmosphere, land surface and uppermost layer of the ocean, while boxes two to k together represent the deep ocean. Each box i has a temperature T_i and heat capacity C_i and is coupled to adjacent boxes above and below. T_1 is defined to be global mean surface temperature (GMST) anomaly relative to pre-industrial conditions. Heat transfer coefficients $\kappa_i > 0$ determine the strength of thermal coupling between boxes i and $i-1$. In the literature κ_1 is often written as λ and is referred to as the climate feedback parameter (e.g. Geoffroy et al., 2013a). We follow the convention of Fredriksen and Rypdal (2017) and use the letter κ for both climate feedback and heat uptake by the deep ocean. The heat transfer coefficient κ_k in the equation for box $k-1$ is multiplied by a so-called efficacy factor $\varepsilon > 0$, introduced by Held et al. (2010), to simulate variation in the effective strength of κ_1 during periods of transient (non-equilibrium) warming. The term $F(t)$ denotes radiative forcing measured at the top of the atmosphere and $\xi(t)$ is a stochastic temperature disturbance (see below). Table 3.1 contains physical units and a brief description of each parameter. Figure 3.1 gives a graphical representation of the k-box model.

Table 3.1: Parameters of k-box model with physical units and description.

Parameter	Unit	Description
γ	dimensionless	Stochastic forcing continuous-time autocorrelation parameter.
C_i	W yr m^{-2} K^{-1}	Total heat capacity of box i.
κ_i	W m^{-2} K^{-1}	Heat transfer coefficient. Controls heat flux across upper boundary of box i.
ε	dimensionless	Deep ocean heat uptake efficacy factor.
σ_η	W m^{-2}	Standard deviation of TOA stochastic forcing component.
σ_ξ	W m^{-2}	Standard deviation of stochastic disturbance applied to surface box.
$F_{4\times CO_2}$	W m^{-2}	Effective radiative forcing after quadrupling pre-industrial atmospheric CO_2.
τ_i	yr	The ith characteristic timescale of the k-box model.
a_i	dimensionless	Weighting of ith exponential basis function in step response of surface temperature.
ECS	K	Equilibrium climate sensitivity. Final temperature after doubling atmospheric CO_2.
TCR	K	Transient climate response. Surface temperature after 70 yrs of 1 % per yr CO_2 increase.

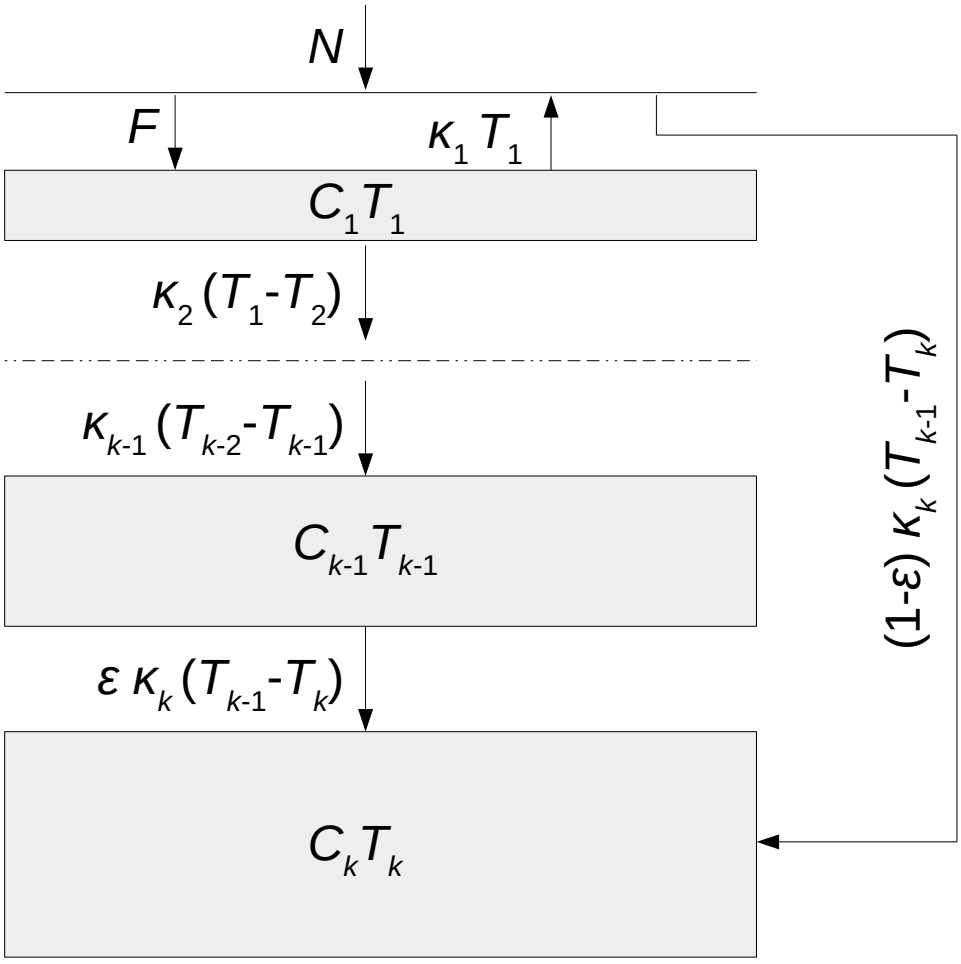

Figure 3.1: Vertical layout of the boxes in the k-box energy-balance model. The thickness of each box indicates its heat capacity and the arrows represent the flow of heat between adjacent boxes. The top of the atmosphere has no heat capacity and so is represented by a horizontal line. The dashed line in the middle is an abbreviation of the intervening boxes.

3.2.3 Stochastic forcing scheme

Natural variability in GMST can be partially explained within the EBM framework using a stochastic process in the radiative forcing term (Hasselmann, 1976). To enforce continuity of F in time we model $F(t)$ as a first-order red noise:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = -\gamma(F - F_{\mathrm{det}}(t)) + \eta(t), \tag{3.6}$$

where $F_{\mathrm{det}}(t)$ and $\eta(t)$ are the respective deterministic and stochastic forcing components. Formally, eqn (3.6) is the Langevin equation of an Ornstein-Uhlenbeck process, the continuous-time analogue of an autoregressive process in discrete time (Uhlenbeck and Ornstein, 1930). Here we assume $\eta(t)$ to be a Gaussian white-noise (WN) process with mean zero and standard deviation σ_η, in which case eqn (3.6) may be referred to as the Hasselmann (1976) model. In the limit as $\gamma \to \infty$ the stochastic forcing becomes white noise, whereas if $\gamma = 0$ we have a Wiener process. Inter-annual variation in radiative forcing is insufficient to explain all of the natural variability in surface temperature. Residual surface temperature variability is explained here by a Gaussian WN temperature disturbance $\xi(t)$ with mean zero and standard deviation σ_ξ. The term $\xi(t)$ functions like an external forcing but is not measurable at the top of the atmosphere since it represents dynamic variability which is generated internally.

Although the underlying mechanistic structure of the k-box EBM is derived from physical considerations, many of its parameters do not correspond to well-defined physical quantities in the real world. It is thus not generally possible to calculate realistic parameter values directly from first principles. Parameter values must instead be estimated empirically from data. In this chapter a maximum likelihood method is developed for estimating parameters of k-box models, the motivation for which having been set out in Chapter 1. The structure of the chapter is as follows: Section 3.3 provides a summary and critique of some methods previously employed to fit box models; Section 3.4 describes data requirements for successful parameter

estimation and the specific CMIP5 data used in this chapter; Section 3.5 outlines the maximum likelihood estimation framework; Section 3.6 describes a software tool created for applying the method described in this chapter; Section 3.7 evaluates the robustness of maximum likelihood estimation in a simulation study; Section 3.8 explains how the method was applied to climate model data from CMIP5 and presents an analysis of the results; the content of this chapter is summarized in Section 3.9.

3.3 Review of methods for fitting k-box energy-balance models

Maximum likelihood estimation is simple for one-box model parameters: given uniformly sampled data, estimation reduces to an ordinary least squares problem with a closed-form solution (Rypdal and Rypdal, 2014). When multiple boxes are included, latent variables are introduced and the estimation problem becomes non-linear and hence more difficult. Several methods have been proposed in the literature for estimating parameters of box-model EBMs with $k \geq 2$, including least-squares curve fitting (Geoffroy et al., 2013a; Caldeira and Myhrvold, 2013), frequency-domain regression (Fredriksen and Rypdal, 2017) and Bayesian estimation (Proistosescu and Huybers, 2017; Jonko et al., 2018). Box models have previously been fitted to the historical record, paleoclimate reconstructions and to data from general circulation model experiments. Three examples of existing methods are described below. The first method described, proposed by Geoffroy et al. (2013a), is compared in Section 3.8.3 with the maximum likelihood estimator proposed in this chapter.

Geoffroy et al. (2013a) derived explicit time-dependent solutions for the two-box model under idealized deterministic forcing scenarios. They proposed a procedure for estimating model parameters using measurements of GMST and top-of-the-atmosphere (TOA) net downward radiative flux (see Section 3.4) from the step responses of AOGCMs in CMIP5. Their method uses prior information about characteristic timescales to estimate the model parameters in sequence, with the

sum of squared residuals as the criterion to be minimized. The time-dependent solution of the two-box model is a sum of saturating exponentials and so estimating parameters in parallel by non-linear least squares can be a notoriously difficult problem (De Groen and De Moor, 1987; Kaufmann, 2003), which is avoided by estimating parameters sequentially. In a companion paper, Geoffroy et al. (2013b) added a deep ocean heat uptake efficacy factor ε to their model, requiring the use of iteration in their fitting procedure. Geoffroy et al. (2013a) did not specify an error model, however their least-squares fitting criterion would correspond to maximum likelihood estimation under an assumption of errors which are independent and identically distributed (i.i.d.) and Gaussian. We have found this assumption to be inconsistent with time series of residuals obtained by subtracting fitted two-box model trajectories from AOGCM step responses: such residual time series exhibit strong autocorrelation. Without specifying an error model it is also impossible to construct confidence intervals for parameter estimates.

Fredriksen and Rypdal (2017) estimated parameters of a three-box model with natural variability driven by a Gaussian WN process in the forcing term. They proposed an iterative least squares-based fitting algorithm to estimate the model parameters. Their method alternates between fitting the signal (expected temperature series for the first box) in the time domain and fitting the noise (time series of residuals) in the frequency domain. Fredriksen and Rypdal (2017) estimated model parameters using estimates of GMST from HadCRUT4 (Morice et al., 2012) and the Moberg et al. (2005) paleoclimate reconstructions, and forcing estimates from Crowley (2000) and Hansen et al. (2011). Unlike the other studies cited in this section, Fredriksen and Rypdal (2017) estimated parameters of box models ($k \geq 2$) without access to measurements of TOA net downward radiative flux, as they were fitting to historical datasets. Only a subset of the model parameters was estimated from data since, without radiative flux measurements, a wide range of possible values for the three characteristic timescales τ_1, τ_2, τ_3 were found to be equally compatible with the observations (see Section 3.4.2 for discussion). In their analysis three candidate

timescale configurations were chosen and the remaining parameters estimated. An important result of Fredriksen and Rypdal (2017) is that the stochastically forced three-box model produces a similar noise spectrum to so-called scale-invariant models, a related class of simple climate model. Parameters of scale-invariant models have been estimated by maximum likelihood (Rypdal and Rypdal, 2014) and more recently using Bayesian inference (Rypdal et al., 2018). The method of Rypdal et al. (2018) is generally applicable to linear response models, including box models, although the authors only present results for the scale-invariant model.

Jonko et al. (2018) estimated parameters of the two-box model using Bayesian hierarchical methods. In their model likelihood the variability in observed temperatures $T_1(t)$ and TOA net downward radiative flux $N(t)$ are jointly modelled as a vector autoregressive process of order one (VAR(1)). All VAR(1) correlations are considered free parameters, not constrained by the physical parameters of the EBM. Given prior distributions for parameters to be estimated, Markov Chain Monte Carlo (MCMC) is used to form an approximation to the posterior distribution. Jonko et al. (2018) used their method to pool information from 24 AOGCM step responses and produce a joint posterior for equilibrium climate sensitivity (ECS). They also included time series of historical temperature observations in their model likelihood to further constrain estimates of future warming. By opting not to include a stochastic forcing term Jonko et al. (2018) increase the number of parameters to estimate and lose physical motivation for the natural temperature variability in their model.

Of the approaches discussed above, none can be considered optimal in the sense of maximum likelihood or sampling from the posterior distribution of the full, stochastic k-box energy-balance model. In this chapter, a maximum likelihood method is therefore introduced for estimating stochastic k-box models with $k \geq 2$. Maximum likelihood estimators are widely used and have known asymptotic sampling properties allowing for simple quantification of parameter uncertainty. Furthermore, optimal complexity of maximum likelihood models can be identified using information criteria.

3.4 Step responses and radiative flux in CMIP5 experiments

3.4.1 The step reponse

From its definition as a set of linear differential equations, it may be seen that the k-box model is a linear time-invariant (LTI) system. It is therefore completely characterized by its impulse response or alternatively its step response, of which the impulse response is the time derivative (Ljung, 1987). The step response contains information about model behaviour on all relevant timescales. The Coupled Model Intercomparison Project Phase 5 (CMIP5) archive includes experiments (Taylor et al., 2012) designed to elicit the step response of AOGCMs by subjecting them to a step forcing of the form

$$F(t) = \begin{cases} F_{4\times \mathrm{CO}_2} & \text{if } t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

The forcing is achieved by an instant quadrupling of atmospheric carbon dioxide (CO_2) concentration. The reasoning behind this choice of forcing is that the amplitude should be large enough that the signal-to-noise ratio is high, but small enough not to induce strongly non-linear behaviour such as tipping points. Ideally the step-forcing experiment would be long enough for the system to stabilize at a new equilibrium temperature and multiple ensemble runs would be available for each AOGCM. However, since Earth system models (ESMs) are expensive to run, the step-forcing experiments in CMIP5 are typically 150 years in length and consist of a single ensemble member. These experiments nevertheless constitute the most information-rich datasets from which to infer the parameters of k-box models and simple climate models in general. The output of an AOGCM step-forcing experiment can even be used on its own to make climate predictions by convolving it with a forcing signal of interest (Good et al., 2011; Lucarini et al., 2017).

3.4.2 Radiative flux constraints

Models in CMIP5 can have equilibration times in the thousands of years (Rugenstein et al., 2020). It is therefore unreasonable to expect that 150-year time series of surface temperature would contain sufficient information to identify all EBM parameters. Firstly, if only surface temperatures are observed, then the forcing magnitude and heat capacity parameters are jointly unidentifiable unless one of them is set to an arbitrary value. Even then, attempting to fit multi-exponential response curves to such datasets in practice, e.g. using least-squares methods, results in highly correlated parameter estimates with correspondingly large uncertainty. This difficulty can be overcome by using measurements of net downward radiative flux at the top of the atmosphere (TOA) to constrain $F(t)$ and κ_1, thereby constraining the net heat uptake of the system at any point in time. Since the EBM parameters describe a heat storage mechanism, their estimation requires knowledge of the amount of heat being stored. Using equations (3.2) and (3.4) we extract the relation

$$N(t) = F(t) - \kappa_1 T_1(t) + (1-\varepsilon)\kappa_k (T_{k-1}(t) - T_k(t)) \tag{3.8}$$

where $N(t)$ denotes the TOA net downward radiative flux. If the system is in equilibrium at time t, i.e. $T_{k-1}(t) = T_k(t)$, and/or if $\varepsilon = 1$, equation (3.8) reduces to the traditional Gregory relation $N(t) = F(t) - \kappa_1 T_1(t)$ (Gregory et al., 2004). Note that, since fitting to $4 \times \mathrm{CO}_2$ experiments is essentially infeasible without measurements of $N(t)$, fitting to historical temperature observations with all parameters free is unlikely to produce meaningfully constrained estimates.

3.5 Maximum likelihood framework for parameter estimation

Computing the likelihood function for the k-box model is non-trivial. We typically observe the temperature of only the first box and hence for $k \geq 2$ at least half

of the model state variables are unobserved (latent). In this section we start by formulating the k-box EBM as a state-space model. We then show how the likelihood of this state-space representation can be evaluated recursively using the Kalman filter. Numerical maximization of the likelihood is briefly described and a method for constructing confidence intervals given. Finally, we explain how optimal model complexity can be identified using information criteria.

3.5.1 Matrix-vector equations

The purely deterministic, homogeneous (externally and internally unforced) k-box model with $\varepsilon = 1$ can be written in matrix-vector form

$$\dot{\boldsymbol{x}}_h(t) = A\boldsymbol{x}_h(t), \tag{3.9}$$

where

$$\boldsymbol{x}_h(t) = (T_1(t), \ldots, T_k(t))', \tag{3.10}$$

and

$$A_{i,j} = \begin{cases} -(\kappa_i + \kappa_{i+1})/C_i & \text{if } i = j \neq k, \\ \kappa_j/C_i & \text{if } j = i + 1, \\ \kappa_i/C_i & \text{if } j = i - 1, \\ -\kappa_i/C_i & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

For $\varepsilon \neq 1$ two entries in the penultimate row of A must be changed to match equation (3.4). The matrix A is tridiagonal because each box is coupled only to its immediate neighbours above and below. The ith eigenvalue of A is $-1/\tau_i$ where τ_i is the ith characteristic timescale of the linear system. An analytical expression for τ_i is given in Geoffroy et al. (2013a) for the case $k = 2$. See Appendix A.1 for a proof that A has real and non-positive eigenvalues for any k when $\varepsilon = 1$.

Analysis of the full inhomogeneous, stochastic k-box model is simplified by the

inclusion of radiative forcing F as a state variable. Defining the state vector

$$\boldsymbol{x}(t) = (F(t), T_1(t), \ldots, T_k(t))', \tag{3.12}$$

we can write the full model

$$\dot{\boldsymbol{x}}(t) = A^+\boldsymbol{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{w}(t), \tag{3.13}$$

where A^+ is simply matrix A augmented with one additional row-column pair (above and to the left) to account for equation (3.6):

$$A^+_{1,1} = -\gamma, \tag{3.14}$$

$$A^+_{2,1} = 1/C_1; \tag{3.15}$$

and where

$$\boldsymbol{b} = (\gamma, 0, \ldots, 0)', \tag{3.16}$$

$$u(t) = F_{\text{det}}(t), \tag{3.17}$$

and

$$\boldsymbol{w}(t) \sim \text{N}(\boldsymbol{0}, Q) \tag{3.18}$$

with

$$Q_{i,j} = \begin{cases} \sigma_\eta^2 & \text{if } i = j = 1, \\ (\sigma_\xi/C_1)^2 & \text{if } i = j = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.19}$$

3.5.2 Discretization scheme

The continuous-time model is a system of stochastic differential equations and may be analysed using the tools of stochastic calculus. However if observations consist of uniformly spaced discrete samples then it makes sense to discretize the model (see Section 3.8.1 for details of sampled data used in this chapter). Assuming constancy

of the deterministic forcing input $u(t) = F_{\text{det}}(t)$ between samples, the model can be discretized exactly (see Appendix A.2 for details)

$$\boldsymbol{x}(t) = A_d\boldsymbol{x}(t-1) + \boldsymbol{b}_d u(t-1) + \boldsymbol{w}_d(t) \tag{3.20}$$

where

$$A_d = \mathrm{e}^{A^+}, \tag{3.21}$$

$$\boldsymbol{b}_d = (A^+)^{-1}(A_d - I)\boldsymbol{b}, \tag{3.22}$$

$$\boldsymbol{w}_d(t) \sim \mathrm{N}(\boldsymbol{0}, Q_d), \tag{3.23}$$

$$Q_d = \int_{s=0}^{1} \mathrm{e}^{A^+ s} Q \mathrm{e}^{A^{+\prime} s}\,\mathrm{d}s, \tag{3.24}$$

with subscript d denoting discretization. The integral in Equation (3.24) can be evaluated via the matrix exponential method described in Section 1 of Van Loan (1978). The above scheme is commonly used in control engineering, where it is known as a "zero-order hold" discretization (Ljung, 1987).

3.5.3 State-space representation

As an LTI system the k-box model is amenable to powerful numerical techniques from control engineering, in particular the Kalman filter (Kalman, 1960). By choosing a model for observations $\boldsymbol{y}(t)$, the k-box model can be written in state-space form

$$\boldsymbol{x}(t) = A_d\boldsymbol{x}(t-1) + \boldsymbol{b}_d u(t-1) + \boldsymbol{w}_d(t) \tag{3.25}$$

$$\boldsymbol{y}(t) = C_d\boldsymbol{x}(t) + \boldsymbol{v}_d(t), \tag{3.26}$$

where matrix C_d is our observation operator and $\boldsymbol{v}_d(t)$ is an (optional) additive observation error. If we observe TOA net downward radiative flux $N(t)$ and surface

temperature $T_1(t)$ at each time t, both without error, then

$$\boldsymbol{y}(t) = (T_1(t), N(t))' = C_d \boldsymbol{x}(t) \tag{3.27}$$

where entries of C_d are determined by equation (3.8). In the general case (e.g. the historical record), observations might be contaminated by errors $\boldsymbol{v}_d(t)$ such that

$$\boldsymbol{v}_d(t) \sim \mathrm{N}(\mathbf{0}, \Sigma_t), \tag{3.28}$$

but for climate model experiments we assume $\boldsymbol{v}_d(t) = \mathbf{0}$ for all t.

3.5.4 Kalman filter

The Kalman filter was originally developed as a minimum mean-square-error (MMSE) estimator of state variables in a noisy linear dynamic system (Kalman, 1960). It may also be used to recursively calculate the likelihood of a time series of observations from this class of model (Tusell, 2011). The Kalman filter estimates the system state at time t, using the information contained in all previous observations up to and including time t, through a recursive procedure iterating over two steps: a prediction step and an update step. For the k-box model in state-space form we can write the Kalman recursions as follows, using the hat/subscript notation of Reid (2001).

Prediction step

Given $\hat{\boldsymbol{x}}_{t-1|t-1}$, our best estimate of the system state at time $t-1$ given data, the predicted state $\hat{\boldsymbol{x}}_{t|t-1}$ at time t is

$$\hat{\boldsymbol{x}}_{t|t-1} = A_d \hat{\boldsymbol{x}}_{t-1|t-1} + \boldsymbol{b}_d u_{t-1}. \tag{3.29}$$

The predicted error covariance of this *a priori* estimate is

$$P_{t|t-1} = A_d P_{t-1|t-1} A_d' + Q_d, \tag{3.30}$$

where $P_{t-1|t-1}$ is the covariance of the estimated state at time $t-1$.

Update step

Having then observed $\boldsymbol{y}_t$ we update our *a priori* estimate of $\boldsymbol{x}_t$ with this new information to obtain an *a posteriori* state estimate $\hat{\boldsymbol{x}}_{t|t}$ with corresponding covariance $P_{t|t}$. Our measurement pre-fit residual is

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - C_d \hat{\boldsymbol{x}}_{t|t-1} \tag{3.31}$$

which has covariance

$$S_t = \Sigma_t + C_d P_{t|t-1} C_d'. \tag{3.32}$$

Our *a posteriori* state estimate is simply our *a priori* estimate $\hat{\boldsymbol{x}}_{t|t-1}$ shrunk towards the observation $\boldsymbol{y}_t$

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + K_t \tilde{\boldsymbol{y}}_t \tag{3.33}$$

where the shrinkage amplitude is the optimal Kalman gain

$$K_t = P_{t|t-1} C_d' S_t^{-1}. \tag{3.34}$$

The covariance of the *a posteriori* estimate is

$$P_{t|t} = (I - K_t C_d) P_{t|t-1} (I - K_t C_d)' + K_t \Sigma_t K_t'. \tag{3.35}$$

The measurement post-fit residual is

$$\tilde{\boldsymbol{y}}_{t|t} = \boldsymbol{y}_t - C_d \hat{\boldsymbol{x}}_{t|t}. \tag{3.36}$$

In the complete absence of observational noise the recursions may still be computed by setting Σ_t equal to a diagonal matrix with each diagonal element a very small number.

3.5.5 Model likelihood

Since the k-box model is a causal linear filter, i.e. system states depend on past states and past inputs but not on future states and future inputs, we can factorize the likelihood function of the temperature observations

$$\mathcal{L}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n; \boldsymbol{\theta}) = \prod_{t=1}^{n} \mathcal{L}(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1; \boldsymbol{\theta}) \tag{3.37}$$

where $\boldsymbol{\theta}$ denotes the vector of model parameters. For numerical stability it is preferable to compute the log-likelihood

$$\ell(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n; \boldsymbol{\theta}) = \sum_{t=1}^{n} \ell(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1; \boldsymbol{\theta}) \tag{3.38}$$

which can be calculated recursively using the pre-fit residuals and their corresponding covariances from the Kalman filter

$$\ell(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t; \boldsymbol{\theta}) = \ell(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}; \boldsymbol{\theta}) - \frac{1}{2}\left(2\log(2\pi) + \log(|S_t|) + \tilde{\boldsymbol{y}}_t' S_t^{-1} \tilde{\boldsymbol{y}}_t\right). \tag{3.39}$$

Evaluation of (3.39) requires the distribution of $\boldsymbol{x}_0|\boldsymbol{\theta}$, upon which $\boldsymbol{y}_1$ depends, to be known. If we assume that at the beginning of a dataset the system is in a state of pre-industrial equilibrium then $\mathrm{E}(\boldsymbol{x}_0) = \mathbf{0}$ with some covariance matrix to be derived from the model parameters (see Appendix A.3). For an abrupt $4 \times \mathrm{CO}_2$ climate model experiment, the first element of $\boldsymbol{x}_0$ (corresponding to radiative forcing) has an expected value, given the model parameters, of $F_{4\times\mathrm{CO}_2}$.

The Kalman filter log-likelihood is essentially a weighted least-squares objective function which penalizes squared one-step-ahead prediction errors (pre-fit residuals). The weighting applied to each prediction error is determined by its corresponding

uncertainty (covariance).

3.5.6 Maximum likelihood estimation

The maximum likelihood estimator (MLE) of the model parameters $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[-\ell(\boldsymbol{y};\boldsymbol{\theta})\right] \tag{3.40}$$

where $\ell(\boldsymbol{y};\boldsymbol{\theta})$ denotes the k-box model log-likelihood function. We minimize the negative log-likelihood numerically: a modern derivative-free algorithm such as BOBYQA (Powell, 2009) is well suited to this task. Standard errors and confidence intervals can be obtained using asymptotic properties of the MLE (for proofs see Schervish (1995)). In the limit as sample size tends to infinity the MLE $\hat{\boldsymbol{\theta}}$ is normally distributed with mean vector $\boldsymbol{\theta}$ and covariance matrix I^{-1} where I denotes the Fisher information matrix

$$I_{jk} = -\mathrm{E}\left[\frac{\partial^2 \ell(\boldsymbol{y};\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k}\right]. \tag{3.41}$$

Although the Fisher information I depends on the values of the true parameters $\boldsymbol{\theta}$, we can obtain a consistent estimator $\hat{I}$ by plugging the MLE $\hat{\boldsymbol{\theta}}$ into equation (3.41). We calculate $\hat{I}$ using a numerical estimate of the Hessian of the negative log-likelihood at the MLE $\hat{\boldsymbol{\theta}}$. The estimated asymptotic sampling distribution of the MLE is then used to calculate standard errors and confidence intervals.

3.5.7 Optimal model complexity

The number of boxes, k, offers a natural parameterization of model complexity. When emulating an AOGCM with an EBM it is desirable to fit the most parsimonious model which does not significantly underperform compared to more complex models. Models with different numbers of boxes k can be compared (e.g. Caldeira and Myhrvold, 2013) using Akaike's Information Criterion (AIC). The AIC score

for a fitted model m is defined

$$\mathrm{AIC}(m) = -2\ell(m) + 2p(m) \tag{3.42}$$

where ℓ is the log-likelihood and p is the number of parameters (Akaike, 1974). The k-box model m_k has $p(m_k) = 2k + 5$ since we have k heat capacities C_i, k heat transfer coefficients κ_i, a radiative forcing $F_{4\times\mathrm{CO}_2}$, two standard deviations σ_η and σ_ξ, and two dimensionless parameters γ and ε. We have

$$\mathrm{AIC}(m_k) = -2\ell(m_k) + 4k + 10. \tag{3.43}$$

Competing models can be compared using the decision rule whereby for a given AOGCM we choose the number of boxes, k, which minimizes $\mathrm{AIC}(m_k)$.

3.6 Software implementation

I have developed a package for the R software environment (R Core Team, 2021) for simulation, fitting, filtering and predicting with k-box EBMs. The package estimates parameters of k-box models from time series of GMST and TOA net downward radiative flux by numerically maximizing the likelihood function. The model likelihood is evaluated using modern implementations of the matrix exponential (Goulet et al., 2021) and the Kalman filter (Luethi et al., 2020). The likelihood is maximized using an implementation (Johnson, 2020; Ypma et al., 2020) of the BOBYQA optimization algorithm (Powell, 2009). Confidence intervals for parameter estimates are obtained using the Fisher information, as described in Section 3.5.6, where the Hessian of the likelihood function is evaluated numerically using an implementation of Richardson's extrapolation (Gilbert and Varadhan, 2019). The R package, which includes the datasets used in this chapter, is available for download at `https://github.com/donaldcummins/EBM`.

3.7 Simulation study

3.7.1 Methods

A simulation study was performed to investigate the feasibility of fitting k-box models to AOGCM step response data via the proposed maximum likelihood method. The step response of HadGEM2-ES from CMIP5 was used to fit a two-box model and a three-box model (optimal under AIC). HadGEM2-ES was chosen as this model has been used extensively for climate change studies. Data from HadGEM2-ES consisted of annually averaged values (see Section 3.8.1 for details of CMIP5 data used). Estimated two-box model parameters were: $\gamma = 1.58$; C_1, $C_2 = 7.73$, 89.3 W yr m^{-2} K^{-1}; κ_1, $\kappa_2 = 0.632$, 0.522 W m^{-2} K^{-1}; $\varepsilon = 1.52$; σ_η, σ_ξ, $F_{4\times\mathrm{CO}_2} = 0.428$, 0.643, 6.86 W m^{-2}. Estimated parameters for the three-box model were: $\gamma = 1.73$; C_1, C_2, C_3 = 3.62, 9.47, 98.7 W yr m^{-2} K^{-1}; κ_1, κ_2, $\kappa_3 = 0.536$, 2.39, 0.634 W m^{-2} K^{-1}; ε = 1.59; σ_η, σ_ξ, $F_{4\times\mathrm{CO}_2} = 0.434$, 0.323, 6.35 W m^{-2}. Each of the two fitted models was used to generate 1000 simulated step responses (see Figure 3.2). Parameters were then estimated for each of the simulated datasets using the same maximum likelihood methodology. The resulting sets of parameter estimates form a Monte Carlo approximation to the estimator sampling distributions (see Figures 3.3 and 3.4).

3.7.2 Results

Estimator sampling distributions for the two-box and three-box models were examined for excessive bias, variance and pairwise correlations. Results for the two-box model simulations are discussed below. Analysis of results for the three-box model leads to analogous conclusions.

Pairwise parameter correlations are visible in the estimated sampling distribution of the two-box model estimator (see Figure 3.3). The strongest correlation (positive) is between the parameters controlling the stochastic forcing, γ and σ_η, i.e. the whiter the noise the greater the disturbance needed at each time step to

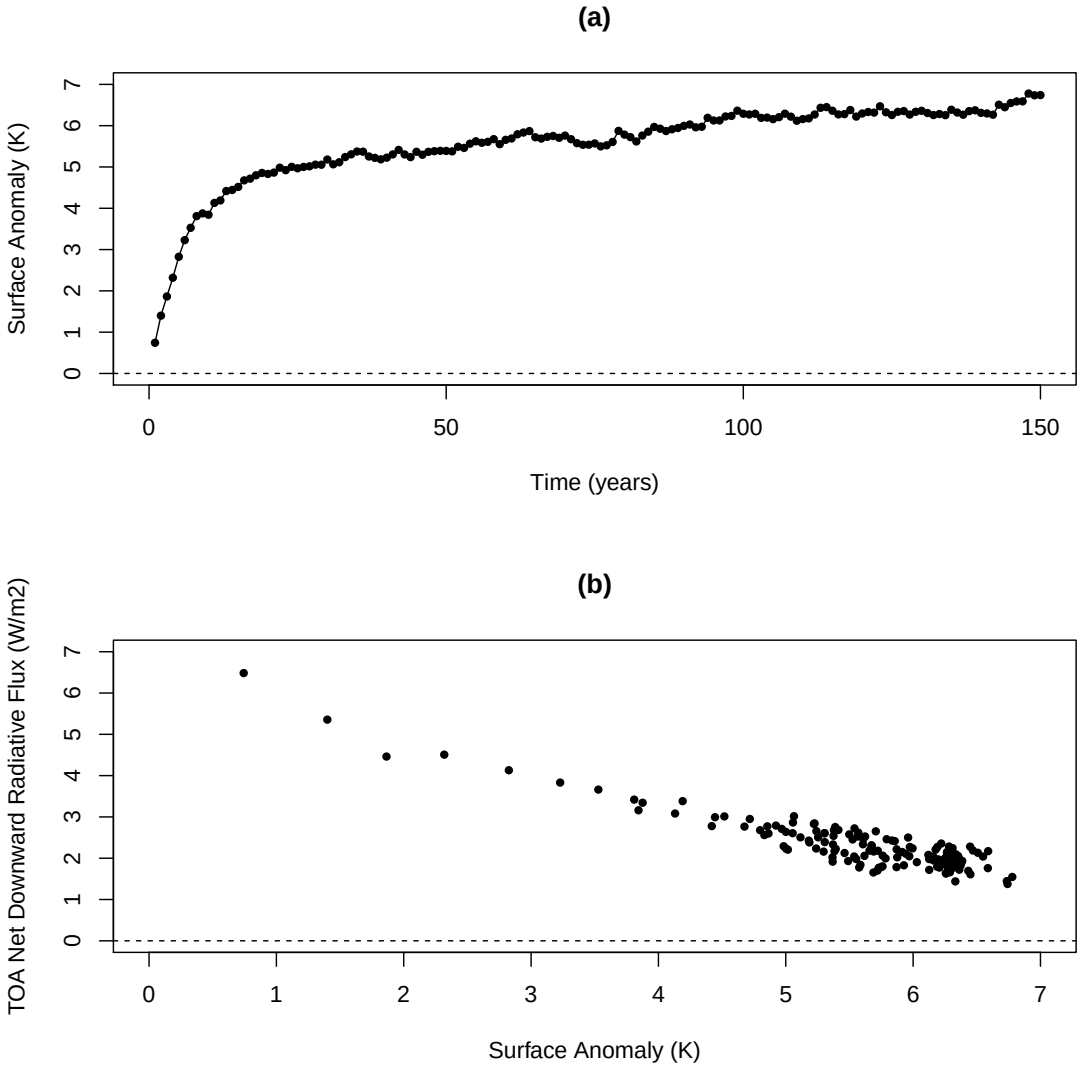

Figure 3.2: Example simulated dataset from a two-box model with parameters: γ = 1.58; C_1, C_2 = 7.73, 89.3 W yr m^{-2} K^{-1}; κ_1, κ_2 = 0.632, 0.522 W m^{-2} K^{-1}; ε = 1.52; σ_η, σ_ξ, $F_{4\times CO_2}$ = 0.428, 0.643, 6.86 W m^{-2}. (a) shows increasing surface temperatures during the first 150 years after CO_2 quadrupling. (b) shows the values of TOA net downward radiative flux in each year plotted against the corresponding surface temperature.

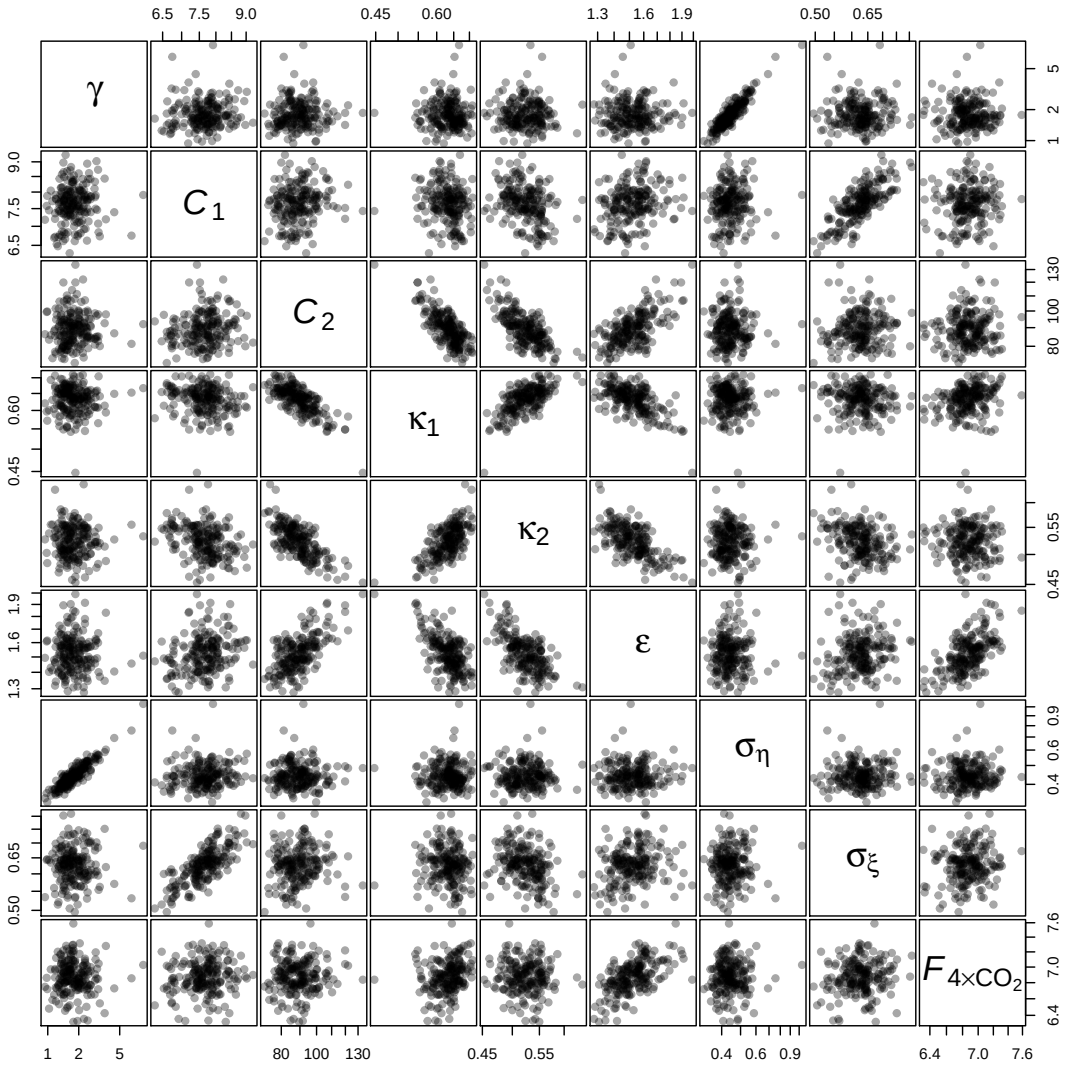

Figure 3.3: Pairs plot showing approximate sampling distribution of the maximum likelihood estimator. Each point represents a model fitted to a simulated dataset. Simulated datasets are from a two-box model with parameters: $\gamma = 1.58$; C_1, C_2 = 7.73, 89.3 W yr m^{-2} K^{-1}; κ_1, κ_2 = 0.632, 0.522 W m^{-2} K^{-1}; $\varepsilon = 1.52$; σ_η, σ_ξ, $F_{4\times CO_2}$ = 0.428, 0.643, 6.86 W m^{-2}. Plot axes are logarithmic to increase visibility of parameter correlations.

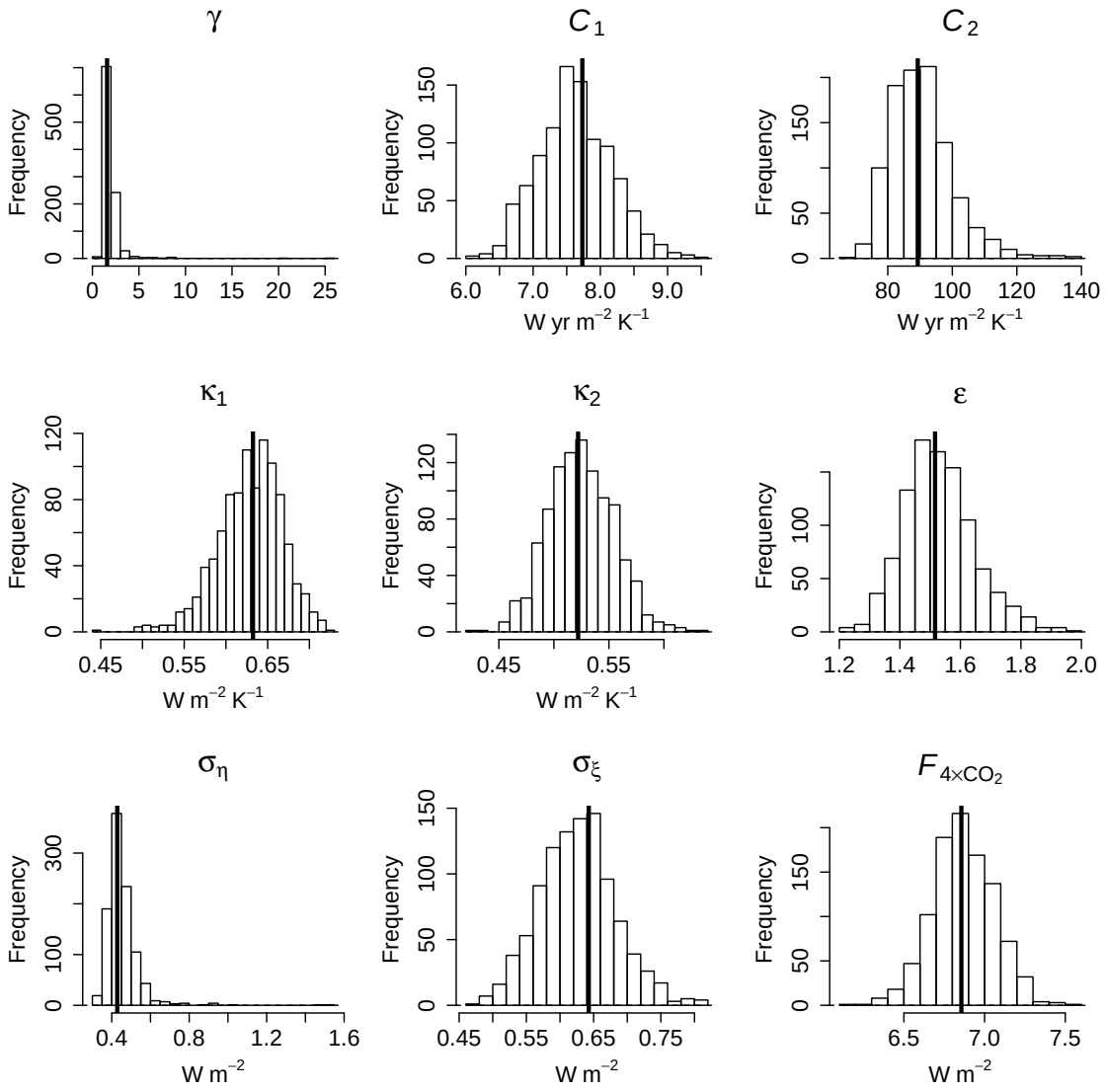

Figure 3.4: Histograms showing approximate sampling distribution of the maximum likelihood estimator. The thick vertical lines indicate the true value of each parameter. Simulated datasets are from a two-box model with parameters: $\gamma = 1.58$; C_1, $C_2 = 7.73$, 89.3 W yr m^{-2} K^{-1}; κ_1, $\kappa_2 = 0.632$, 0.522 W m^{-2} K^{-1}; $\varepsilon = 1.52$; σ_η, σ_ξ, $F_{4\times\mathrm{CO}_2} = 0.428$, 0.643, 6.86 W m^{-2}.

obtain the same overall level of variability. The second strongest correlation (negative) is between the climate feedback parameter κ_1 and deep ocean heat capacity C_2. The third strongest correlation (positive) is between C_1 and σ_ξ. A natural explanation for this is that when the heat capacity of the first box C_1 is increased the corresponding temperature T_1 has more inertia and hence requires a stronger stochastic disturbance amplitude σ_ξ to maintain the same level of variability. The fourth strongest correlation (negative) is between κ_2 and C_2. This correlation is related to the time taken for relaxation of the system on the longer timescale τ_2. A longer relaxation time can be achieved either by increasing the heat capacity of the second box C_2 or by reducing the heat transfer coefficient κ_2 between boxes one and two.

Model parameters can be divided, by correlation, into two disjoint sets: set (i), stochastic forcing parameters γ and σ_η; and set (ii), all remaining parameters. Neither γ nor σ_η is correlated with any parameter in set (ii), nor does either parameter appear well constrained by the simulated datasets, the consequence being a mutual inflation of uncertainty. The correlations between parameters in set (ii) appear to act favourably: individual parameter uncertainty in set (ii) is uniformly low with coefficients of variation mostly less than 10 percent. If at least one parameter in set (ii) is well constrained by observations, as appears to be the case, then uncertainty in the other parameters decreases as a result.

Estimated marginal distributions of the two-box model parameters resemble unimodal bell curves (see Figure 3.4), with the notable exception of γ and σ_η. The parameter γ appears poorly bounded from above (hard to rule out very white stochastic forcing) and this uncertainty propagates into σ_η. The maximum likelihood estimator is asymptotically unbiased but in general has a finite sample bias. Estimates of all two-box model parameters display some bias. Parameters γ and σ_η have positive relative biases of 21 and six percent respectively, which is unsurprising given the skewness of their marginal distributions. For parameters in set (ii) the bias is in all cases less than two percent of the parameter's true value.

3.7.3 Conclusions

The simulation study demonstrates that the proposed maximum likelihood method reliably estimates parameters of two-box and three-box box models from the step response of a typical AOGCM from CMIP5. Pairwise correlation and estimator bias were found to influence estimates of stochastic forcing parameters γ and σ_η, however other model parameters were not adversely affected.

3.8 Fitting to CMIP5 climate model simulations

The R package was used to fit two-box and three-box models to the step responses of 16 ESMs from CMIP5 (see Table 3.2), using the same data as Geoffroy et al. (2013b).

3.8.1 Data and fitting procedure

The step response data consist of values of GMST T_1 and TOA net downward radiative flux N averaged over each of 150 years in the experiment. While it is possible, in practice, to fit four-box models using the methodology described in this chapter, it was decided that the upper limit for these data should be $k = 3$. It was found, through numerical experimentation, that fitting a fourth box typically yields an estimated characteristic timescale substantially shorter than one year. While the inclusion of a fourth box sometimes led to an improvement in AIC score (due to better fitting to the first year of the step response), the physical parameters of four-box models were typically degenerate, e.g. there would be a very small heat capacity in one of the boxes, causing matrix A in eqn (3.11) to have a large condition number.

For each ESM the fitted box model with lower AIC (see Section 3.5.7) was chosen as the optimal k-box emulator. The same procedure was applied to the multi-model mean (MMM) of the 16 step-response datasets. Maximum likelihood parameter estimates are reported for these optimal fits (see Table 3.3), with corresponding es-

Table 3.2: CMIP5 climate model expansions (Geoffroy et al., 2013a).

Model	Expansion
BCC-CSM1-1	Beijing Climate Center, Climate System Model, 1-1
BNU-ESM	Beijing Normal University - Earth System Model
CanESM2	Canadian Earth System Model, version 2
CCSM4	Community Climate System Model, version 4
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5
CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation Mark, version 3.6.0
FGOALS-s2	Flexible Global Ocean-Atmosphere-Land System Model gridpoint, second spectral version
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth Science Model 2M
GISS-E2-R	Goddard Institute for Space Studies Model E, coupled with Russell ocean model
HadGEM2-ES	Hadley Centre Global Environmental Model 2, Earth System
INM-CM4	Institute of Numerical Mathematics Coupled Model, version 4.0
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO, low resolution
MIROC5	Model for Interdisciplinary Research on Climate, version 5
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution
MRI-CGCM3	Meteorological Research Institute Coupled General Circulation Model, version 3
NorESM1-M	Norwegian Earth System Model, intermediate resolution

timates (see Table 3.4) of characteristic timescales τ_i, surface temperature response coefficients a_i, equilibrium climate sensitivity (ECS) and transient climate response (TCR). It should be noted that, as the shortest timescale of the three-box model is of the order of one year, estimated parameters of the first box will be affected by changes in radiative forcing due to stratospheric and tropospheric "rapid adjustments" (Chung and Soden, 2015). We refer to the fits chosen using AIC as optimal k-box emulators for the remainder of this chapter.

3.8.2 Results

From Table 3.4 it can be seen that, for all 16 fitted ESMs, three boxes are required for optimal emulation under AIC. According to AIC the multi-model mean requires three boxes. Figure 3.5 shows three examples of fitted step responses for optimal k-box emulators. In all fitted models the heat capacities of the boxes increase with depth while, with the exception of GISS-E2-R, the heat transfer coefficients decrease with depth (excluding the feedback parameter κ_1). The approximate signal-to-noise ratio, calculated $F_{4\times CO_2}/\sqrt{\sigma_\eta^2 + \sigma_\xi^2}$ for the step-forcing experiment, ranges from 7.1 to 19 with a median of 9.9. This high ratio allows us to fit models with many parameters and without excessive parameter uncertainty (see Table 3.5). The improved fit to the step response moving from a two-box to a three-box model is often clearly visible (e.g. Figure 3.6).

The number of boxes, k, influences the impulse responses of the fitted box models, sometimes strongly (see Figure 3.7). The mathematical definition of the impulse response is given in Appendix A.4. For all 16 ESMs from CMIP5 the impulse response of the optimal k-box emulator runs hotter in the first few years than that of the corresponding two-box model. Moving from two to three boxes increases the instantaneous sensitivity by between 17 and 174 percent (see Table 3.6). This suggests that when modelling the GMST response to impulse-like forcing events such as volcanic eruptions the greater flexibility of a three-box model might prove valuable.

Table 3.3: Estimated parameters for optimal k-box emulators of ESMs in CMIP5. In all cases $k = 3$. For physical units and descriptions of parameters see Table 3.1. MMM refers to the k-box model fitted to the average of the datasets from all 16 ESMs.

Model	γ	C_1	C_2	C_3	κ_1	κ_2	κ_3	ε	σ_η	σ_ξ	$F_{4\times\mathrm{CO}_2}$
BCC-CSM1-1	2.9	5.3	12.3	49	1.21	1.7	0.79	1.28	0.46	0.40	7.1
BNU-ESM	2.3	4.0	9.9	85	0.94	1.6	0.71	0.98	0.60	0.66	7.4
CanESM2	2.5	4.6	11.1	66	1.01	1.8	0.81	1.24	0.53	0.52	7.9
CCSM4	2.1	4.4	13.0	70	1.28	2.3	1.05	1.44	0.49	0.49	8.0
CNRM-CM5.1	11.5	4.0	9.6	90	1.14	2.4	0.60	0.90	0.83	0.41	7.2
CSIRO-Mk3.6.0	1.7	3.6	16.0	63	0.59	2.4	1.15	1.73	0.70	0.50	6.1
FGOALS-s2	2.3	4.3	8.1	135	0.86	2.2	1.11	1.19	0.82	0.66	7.9
GFDL-ESM2M	3.3	4.8	10.2	114	1.34	2.6	1.13	1.19	0.77	0.56	6.9
GISS-E2-R	1.6	4.9	31.6	107	1.82	1.7	4.66	1.46	0.32	0.30	8.3
HadGEM2-ES	1.7	3.6	9.5	99	0.54	2.4	0.63	1.59	0.43	0.32	6.4
INM-CM4	1.6	4.3	7.9	275	1.66	2.7	0.81	0.78	0.33	0.32	6.3
IPSL-CM5A-LR	1.9	2.7	16.7	101	0.73	2.4	0.63	1.21	0.50	0.38	6.5
MIROC5	1.8	4.7	17.9	139	1.55	1.7	1.33	1.18	0.54	0.89	8.7
MPI-ESM-LR	2.5	4.4	13.7	70	1.12	2.0	0.91	1.44	0.68	0.71	8.9
MRI-CGCM3	2.6	4.5	14.5	61	1.26	2.2	0.71	1.22	0.56	0.40	6.8
NorESM1-M	2.2	5.2	13.4	105	1.08	2.6	1.29	1.50	0.52	0.47	7.0
MMM	1.9	5.1	11.2	89	1.03	2.0	0.99	1.29	0.15	0.15	7.2

Table 3.4: Characteristic timescales τ_i, surface temperature response coefficients a_i, equilibrium climate sensitivity (ECS) and transient climate response (TCR) of optimal k-box emulators fitted to ESMs in CMIP5. Column ΔAIC shows the decrease in AIC moving from two to three boxes. For physical units and descriptions of parameters see Table 3.1. MMM refers to the k-box model fitted to the average of the datasets from all 16 ESMs.

Model	τ_1	τ_2	τ_3	a_1	a_2	ECS	TCR	ΔAIC
BCC-CSM1-1	1.54	7.8	162	0.28	0.33	2.9	1.9	21.0
BNU-ESM	1.32	8.8	272	0.25	0.38	3.9	2.5	17.1
CanESM2	1.34	7.6	220	0.23	0.34	3.9	2.3	21.0
CCSM4	1.05	6.1	201	0.25	0.30	3.1	1.9	29.0
CNRM-CM5.1	0.91	8.6	259	0.21	0.49	3.2	2.1	40.2
CSIRO-Mk3.6.0	1.03	6.8	315	0.14	0.18	5.2	1.9	32.0
FGOALS-s2	1.03	5.5	393	0.14	0.36	4.6	2.3	8.9
GFDL-ESM2M	0.96	5.6	262	0.20	0.38	2.6	1.5	11.2
GISS-E2-R	1.34	3.7	235	0.46	0.10	2.3	1.4	21.3
HadGEM2-ES	0.95	8.2	532	0.10	0.31	5.9	2.4	43.1
INM-CM4	0.78	5.9	551	0.23	0.52	1.9	1.4	33.0
IPSL-CM5A-LR	0.78	13.2	394	0.19	0.33	4.4	2.2	75.7
MIROC5	1.31	7.8	321	0.39	0.24	2.8	1.8	5.5
MPI-ESM-LR	1.23	7.4	231	0.26	0.29	4.0	2.3	16.4
MRI-CGCM3	1.12	9.4	190	0.27	0.36	2.7	1.7	38.5
NorESM1-M	1.12	5.9	302	0.17	0.29	3.2	1.6	13.9
MMM	1.35	6.9	273	0.2	0.34	3.5	2.0	68.8

Table 3.5: Example approximate 95% confidence intervals for parameters of k-box models fitted to HadGEM2-ES. For physical units and descriptions of parameters see Table 3.1.

	$k=2$			$k=3$		
Parameter	MLE	2.5%	97.5%	MLE	2.5%	97.5%
γ	1.58	1.04	2.41	1.73	1.15	2.60
C_1	7.73	6.64	9.01	3.62	2.98	4.39
C_2	89.29	73.02	109.18	9.47	7.61	11.80
C_3				98.66	84.10	115.74
κ_1	0.63	0.56	0.71	0.54	0.46	0.63
κ_2	0.52	0.46	0.59	2.39	1.82	3.12
κ_3				0.63	0.57	0.71
ε	1.52	1.30	1.77	1.59	1.38	1.83
σ_η	0.43	0.35	0.52	0.43	0.35	0.53
σ_ξ	0.64	0.53	0.77	0.32	0.27	0.39
$F_{4\times CO_2}$	6.86	6.46	7.28	6.35	6.03	6.70

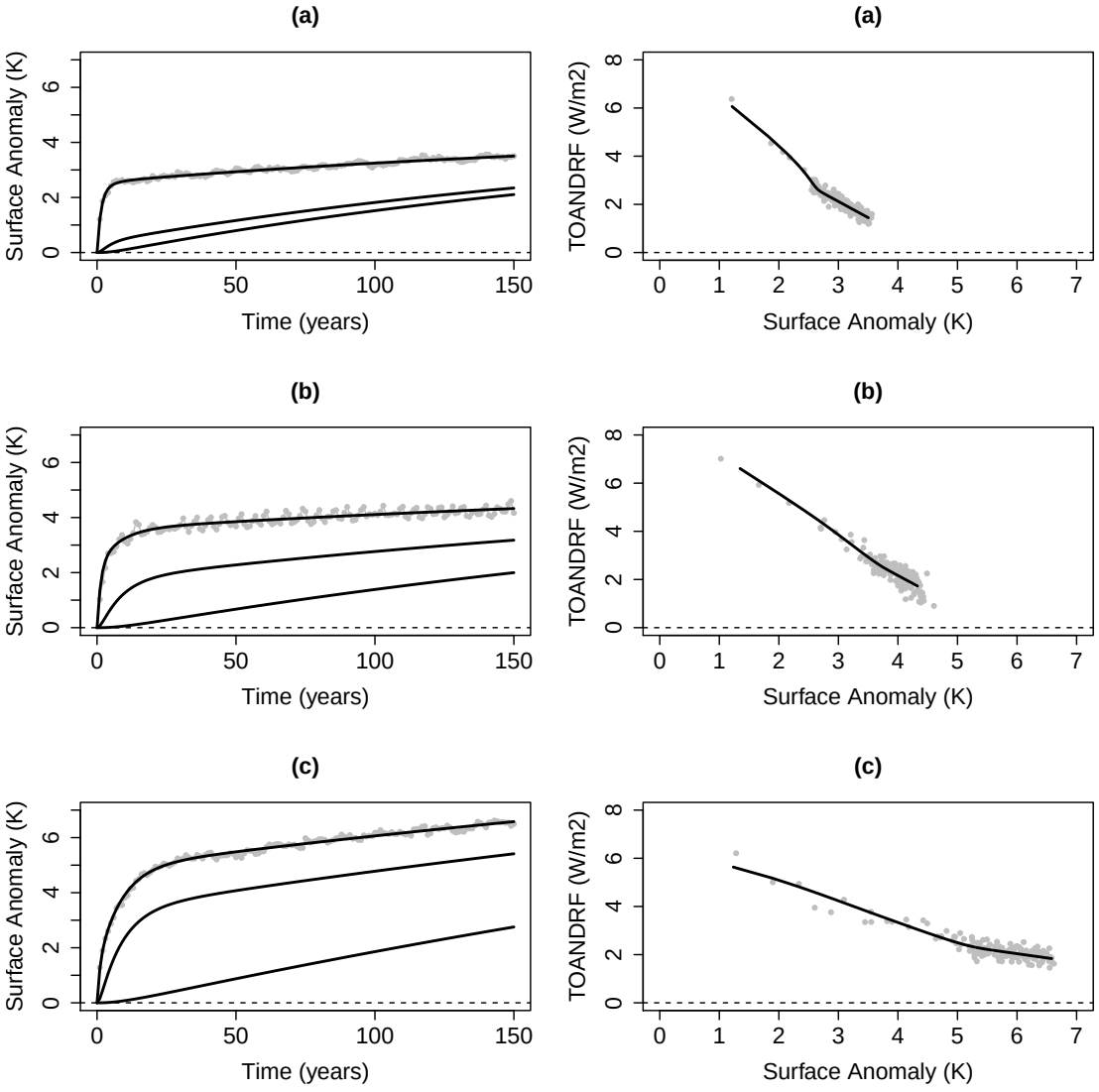

Figure 3.5: Observed and fitted three-box step responses of three ESMs from CMIP5. On the left are temperature trajectories for each box. On the right are TOA net downward radiative fluxes against surface temperature. Grey dots are observations while the black curves are expected box-model responses. Models are (a) GISS-E2-R, (b) MIROC5, (c) HadGEM2-ES.

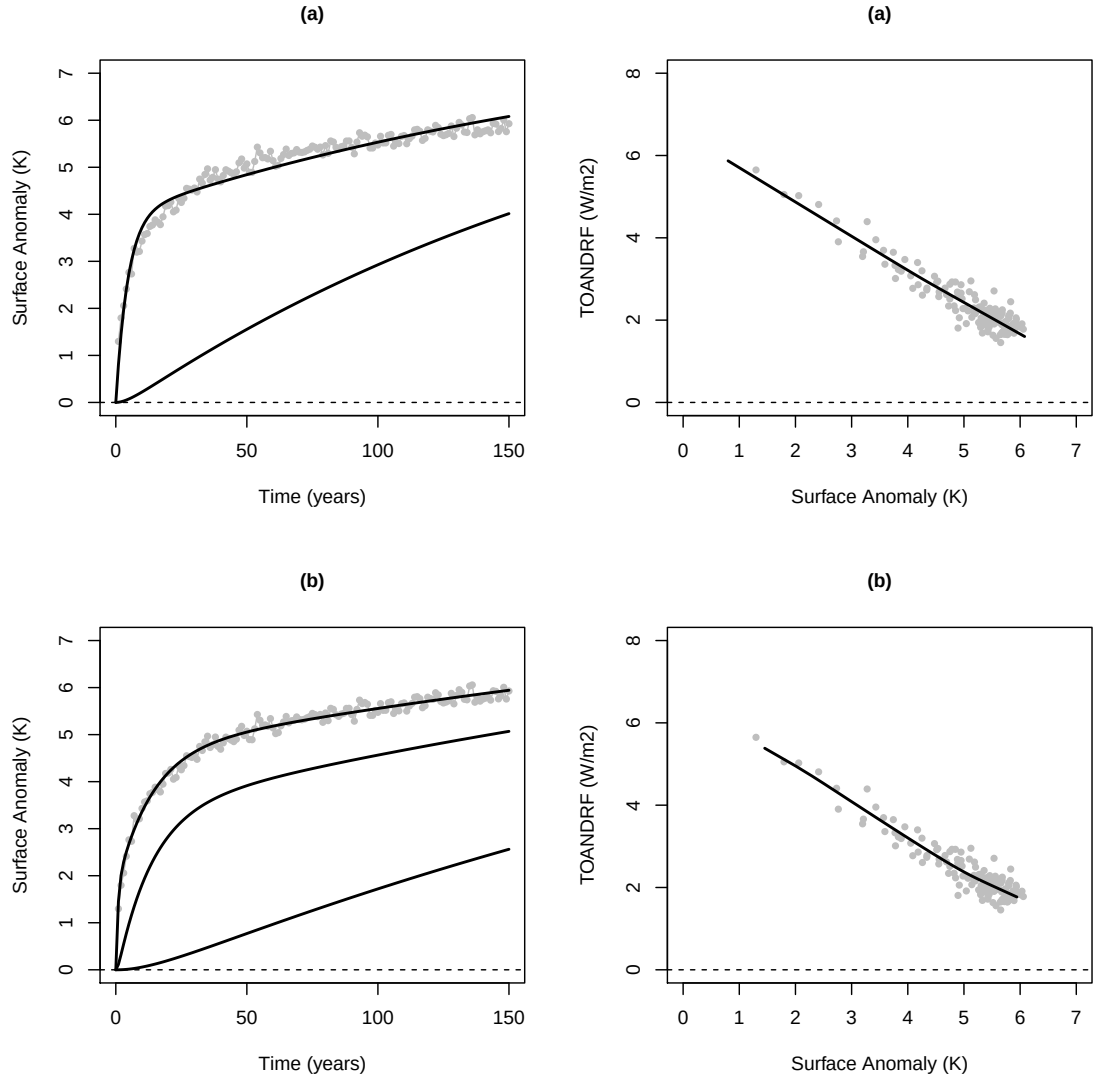

Figure 3.6: Two-box (a) and three-box (b) fits to the step response of IPSL-CM5A-LR. On the left are temperature trajectories for each box. On the right are TOA net downward radiative fluxes against surface temperature. Grey dots are observations while the black curves are expected box-model responses.

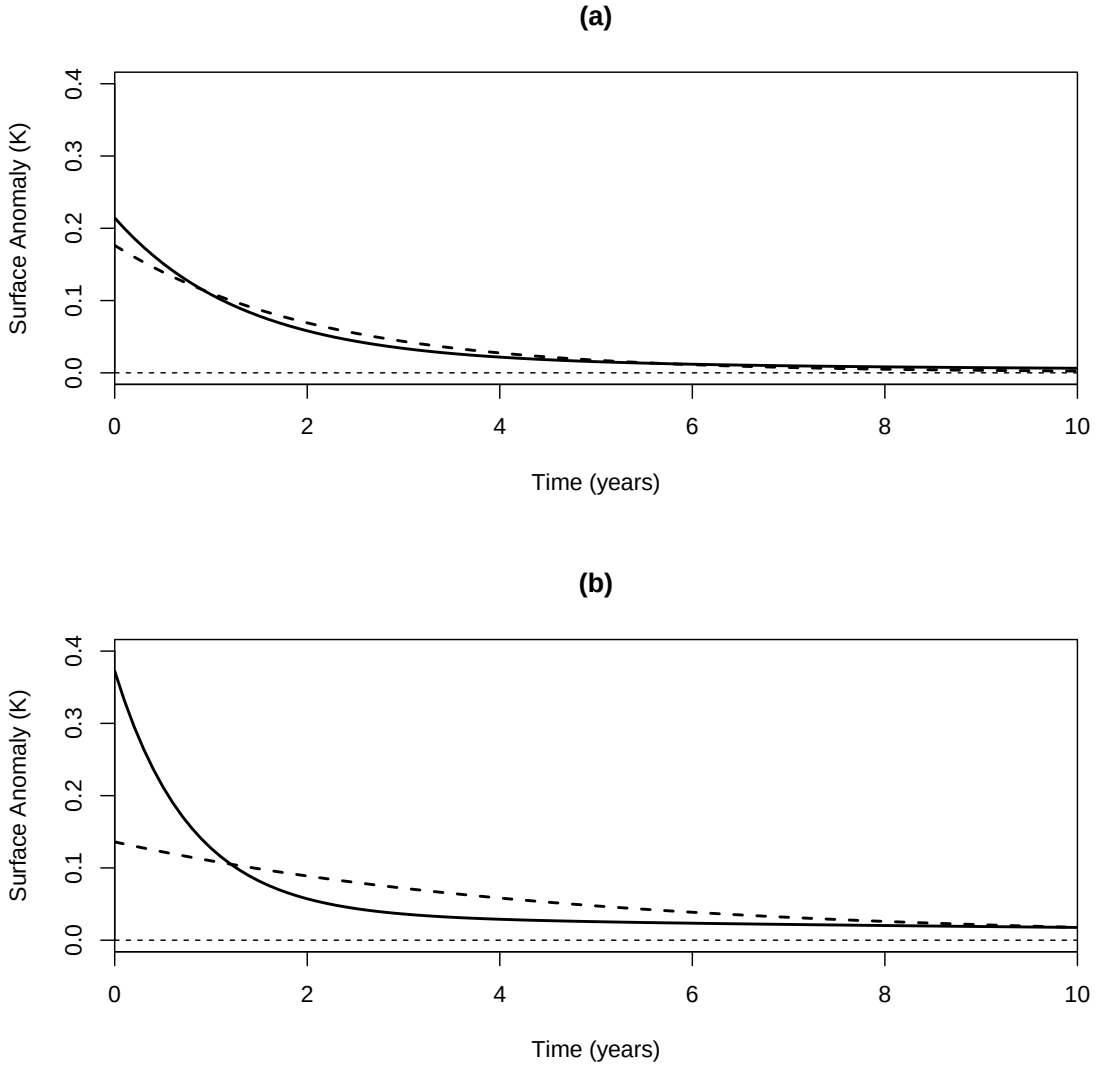

Figure 3.7: Impulse responses of fitted k-box models. The curves are the expected temperature trajectories of the first box of the fitted models in response to a unit-impulse forcing. The solid and dashed curves correspond to three-box and two-box fits respectively, fitted using maximum likelihood. Models are (a) MIROC5, (b) IPSL-CM5A-LR.

Table 3.6: Instantaneous increase in surface temperature (K) under a unit-impulse forcing scenario. Results are given for two-box and three-box maximum likelihood fits. Also given is the percentage increase moving from two to three boxes. MMM refers to the k-box model fitted to the average of the datasets from all 16 ESMs.

Model	MLE Two-Box	MLE Three-Box	% Increase
BCC-CSM1-1	0.13	0.19	42
BNU-ESM	0.17	0.25	45
CanESM2	0.15	0.22	48
CCSM4	0.14	0.23	58
CNRM-CM5.1	0.12	0.25	106
CSIRO-Mk3.6.0	0.19	0.28	50
FGOALS-s2	0.16	0.23	42
GFDL-ESM2M	0.14	0.21	46
GISS-E2-R	0.17	0.2	17
HadGEM2-ES	0.13	0.28	114
INM-CM4	0.14	0.23	72
IPSL-CM5A-LR	0.14	0.37	174
MIROC5	0.18	0.21	22
MPI-ESM-LR	0.16	0.23	46
MRI-CGCM3	0.13	0.22	74
NorESM1-M	0.12	0.19	54
MMM	0.12	0.19	62

3.8.3 Comparison with a least-squares estimator

Figure 3.8 compares two-box model parameter estimates obtained using maximum likelihood with those obtained by Geoffroy et al. (2013b). Maximum likelihood typically yields lower estimates of the heat capacities C_1 and C_2 but higher estimates of the heat transfer coefficient κ_2. This results in shorter estimated characteristic timescales τ_1, τ_2 when using maximum likelihood. Estimates of the radiative parameters κ_1, ε and $F_{4\times\mathrm{CO}_2}$ appear insensitive to the choice of fitting methodology in the case $k = 2$.

3.8.4 Hidden state estimation

Under the proposed observation model (see Section 3.5.3), a fitted k-box model can be combined with temperature and forcing data to filter the (possibly noisy) observations and estimate the temperatures of the unobserved boxes (see Figure 3.9). In this way we can see the attenuation of natural variability in temperature with increasing depth. The thermal inertia of the deep ocean boxes with their large heat capacity means that in the CO_2 quadrupling experiment the noise in these boxes is dwarfed by the signal. Given measurements of TOA net downward radiative flux, *a posteriori* state-variable uncertainty is negligible (invisible when plotted), so the reconstructed series in Figure 3.9 may be taken as known. Filtering and hidden state estimation with k-box models is not restricted to step responses or AOGCM experiments, but rather is applicable to any combination of global temperature and radiative forcing data, including the observational record.

3.9 Summary

The k-box energy-balance model described in this chapter offers a simple but flexible representation of the response of global mean surface temperature to radiative forcing, both deterministic and stochastic, over a range of timescales. Parameter estimation for this class of model is non-trivial: since we can typically observe the

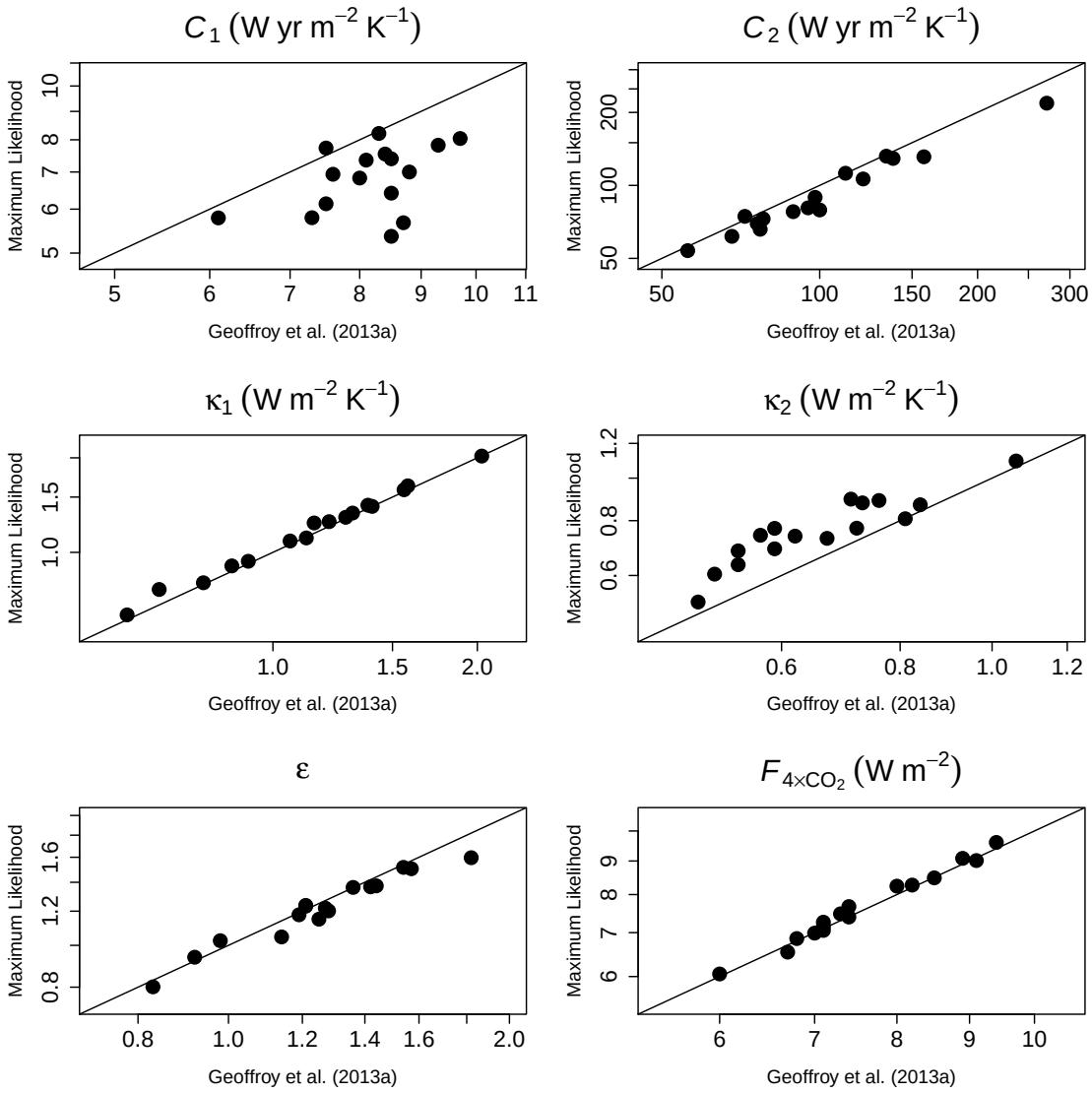

Figure 3.8: Maximum likelihood parameter estimates for two-box models compared with corresponding estimates from Geoffroy et al. (2013b). Each point is one of 16 ESMs from CMIP5. The solid lines have equation $y = x$ and show where estimates are the same for both fitting methodologies.

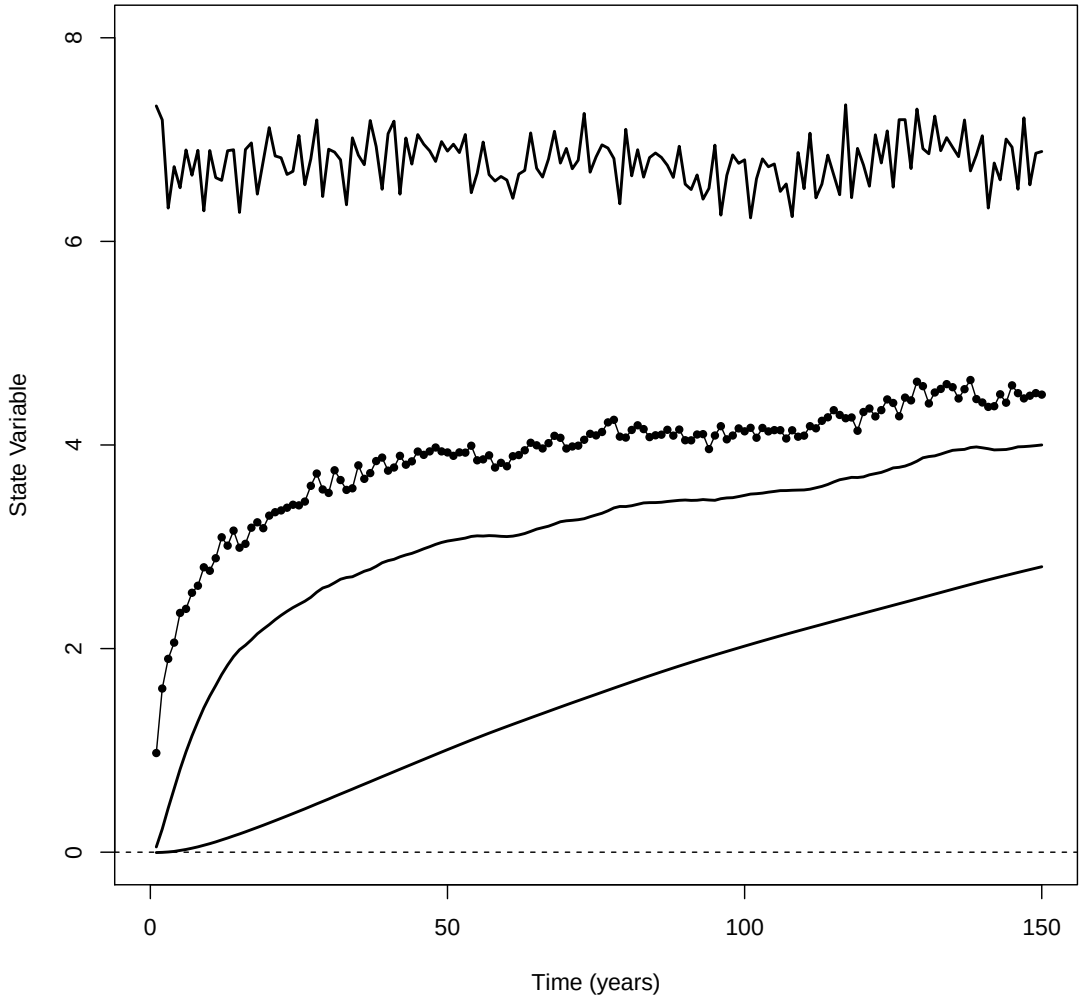

Figure 3.9: Reconstructed three-box model state variables in the MRI-CGCM3 step-forcing experiment. The dots are observed surface temperatures $T_1(t)$ while the solid curves are reconstructed time series of the latent variables in their respective units. Latent variables are, from top to bottom, radiative forcing $F(t)$ in W m^{-2}, deep ocean box temperatures $T_2(t)$ and $T_3(t)$ in K.

temperature of only the first box we have a situation where for $k \geq 2$ at least half of the model state variables are latent. We have shown how, by finding a state-space representation of the linear dynamic system and evaluating the likelihood recursively via the Kalman filter, maximum likelihood estimates of all model parameters may be obtained.

The k-box model is a linear time-invariant system and thus characterized by its response to a step forcing, a forcing scenario which has been simulated in AOGCM experiments. A simulation study has been carried out to investigate feasibility, reliability and performance of the proposed method when applied to step-response data. The proposed method has been found to reliably estimate the k-box model parameters.

An important advantage of maximum likelihood estimation is that optimal model complexity can be chosen using information criteria. To demonstrate this, two-box and three-box models were fitted to each of a set of 16 Earth system models from CMIP5 with the optimal number of boxes chosen by Akaike's information criterion. It was found that for all 16 AOGCMs three boxes are required for optimal k-box emulation. Results obtained via maximum likelihood estimation were compared with equivalent results from the method of Geoffroy et al. (2013b). It was found that estimates of some model parameters differ systematically depending on the choice of fitting method. The number of boxes, k, was found to influence the impulse responses of the fitted models, sometimes strongly. These results suggest that, under impulse-like forcing scenarios, AOGCM responses might be better emulated using three-box models.

Finally, an example has been presented showing how a fitted k-box model can be combined with temperature and forcing data to reconstruct the temperatures of unobserved boxes corresponding to the deep ocean. Noise filtering and hidden state estimation using k-box AOGCM emulators are possible wherever we have a combination of global temperature and radiative forcing data, including the observational record. In the next chapter, the concept of energy-balance models as linear

filters will be explored in more detail and a link will be made to a widely used class of time series models. The performance of such filters as estimators of radiative forcing will be investigated and their practical applicability to the historical record demonstrated. In Chapter 5, results from these two chapters will be used as part of an argument addressing the main research questions of this thesis.

Chapter 4

Estimation of historical radiative forcing using energy-balance model ARMA representations

4.1 Aim

This chapter is based on material from Cummins et al. (2020a) and introduces an alternative representation of k-box energy-balance models as autoregressive moving-average filters. Performance of the digital filter representation as a tool for computing radiative forcing from surface temperature is investigated, and the method applied to the historical temperature record.

4.2 Introduction

In the previous chapter, the k-box stochastic energy-balance model (EBM) was introduced as a simplified representation of global climate, and as a potential tool for analysing the regression model implicit in optimal fingerprinting methods for detection and attribution (D&A) of climate change trends. Many of the results employed in the previous chapter followed from the fact that the k-box model is a linear time-invariant (LTI) system, and this property will be exploited further in

the present chapter. Detection and attribution is by its nature an example of an "inverse problem": given historical climate observations (i.e. system outputs), what can be said about the relative contributions of the various forcing factors (system inputs)? In this chapter, we investigate the potential of the k-box model as a means of addressing this inverse problem directly, by estimating historical radiative forcing from time series of global surface temperature.

4.2.1 Estimation of historical radiative forcing

The estimation of historical radiative forcing, a measure of the net change in the energy balance of the climate system in response to an external perturbation, is a matter of strong scientific interest, as evidenced by the dedication of a whole chapter to this topic in the fifth assessment report from the Intergovernmental Panel on Climate Change (IPCC) (Myhre et al., 2013). Historical radiative forcing, alongside equilibrium climate sensitivity (ECS), the long-term increase in global mean surface temperature (GMST) caused by a doubling of atmospheric CO_2 concentration, and transient climate response (TCR), the increase in GMST after 70 years of one-percent-per-year increasing CO_2, are the three main quantities that inform long-range forecasts of global warming. Estimation of historical forcing is therefore of particular relevance to climate policy decision makers.

The term radiative forcing refers to a change in the Earth's energy balance relative to some pre-defined baseline value, usually chosen to represent pre-industrial conditions. In this chapter we use the effective radiative forcing (ERF), whose definition is given in Myhre et al. (2013): "the change in net [top-of-the-atmosphere] downward radiative flux after allowing for atmospheric temperatures, water vapour and clouds to adjust, but with surface temperature or a portion of surface conditions unchanged." The adjustments of atmospheric temperatures and other variables, included in the definition of ERF, are commonly referred to as "rapid adjustments" and take place over much shorter timescales than resultant changes in surface temperature. (Vial et al., 2013; Chung and Soden, 2015; Smith et al., 2018).

It is infeasible to calculate historical radiative forcing from observational data using the raw definition of ERF, as relevant climate variables, in particular top-of-the-atmosphere (TOA) net downward radiative flux, were unobserved for most of the historical period. Techniques have therefore been developed for diagnosing radiative forcing from general circulation model (GCM) experiments. Forster et al. (2016) describe methods, such as ERF_trans, which use GCMs to simulate the historical period with and without emissions of various forcing agents, and hence calculate associated changes in Earth's energy balance. Resulting series of estimated forcings are strictly conditional on the climate model used. This approach is computationally expensive and there is also some unavoidable noise contamination of results due to internal variability in model output. More recently, Andrews and Forster (2020) have combined GCM simulations with instrumental observations of historical GMST and global heat uptake to constrain historical ERF.

Alternative approaches, based on simple climate models, have been used to estimate historical radiative forcing from the observational record (e.g. Tanaka et al., 2009; Urban and Keller, 2010; Padilla et al., 2011; Aldrin et al., 2012; Urban et al., 2014; Johansson et al., 2015; Ljungqvist, 2015). These studies used Bayesian inference (or non-linear Kalman filtering in the case of Padilla et al. (2011)) to jointly estimate some or all of the parameters of a simple climate model together with corresponding series of historical forcing. A common aim of these studies is to constrain ECS, uncertainty in ECS being found dependent on corresponding uncertainty in historical forcing (Tanaka et al., 2009). As well as being computationally cheaper, methods based on simple climate models have the appeal of directly incorporating observational data, as opposed to GCMs whose dependence on the historical record (through parameter tuning) can be more subtle.

4.2.2 Application to detection and attribution of climate change trends

The line of investigation pursued in this chapter is motivated by the potential application of simple climate models and forcing estimation techniques to the D&A problem. Simple climate models have previously been used for D&A of changes in GMST, e.g. Otto et al. (2015); Rypdal (2015); Haustein et al. (2017). These studies regressed time series of historical temperature observations on predicted temperature series from linear impulse-response models, thus applying a simplified version of the widely used optimal fingerprinting methodology (Hasselmann, 1997; Allen and Tett, 1999; Allen and Stott, 2003), which is more typically applied to high-dimensional gridded datasets of observations and GCM output.

In the context of D&A, surface temperature is an observable proxy for radiative forcing, which is itself not directly observable. The suitability of this proxy for regression analysis is reduced by two artifacts of the climate system's thermal inertia: (i) a delayed temperature response to changes in radiative forcing; (ii) strong temporal autocorrelation in natural temperature variability. Within the framework of linear impulse-response models, instantaneous surface temperature is simply a convolution of previous changes in radiative forcing (Good et al., 2011). One might therefore propose that D&A of changes in GMST using simple climate models could be improved by first deconvolving temperature observations to obtain series of estimated radiative forcing, and then by performing the traditional regression step on radiative forcing time series. In this way one might reasonably expect to eliminate (or at least reduce) both the time delay in the regressed series and also the temporal autocorrelation in the regression residuals.

The remainder of this chapter is a proof-of-concept study in which a method is developed for performing the proposed deconvolution of temperature time series using k-box energy balance models (EBMs). To do this, we exploit a known correspondence between linear ordinary differential equations (ODEs) and discrete-time autoregressive moving-average (ARMA) time series models. Specifically, we use the

fact that a system of k first-order linear ODEs has a discrete-time representation as an ARMA(k, $k-1$) filter (Spolia and Chander, 1974; Chang et al., 1982). This correspondence has been used before in the context of three-box EBMs: by Grieser and Schönwiese (2001) as a convenient means of integrating the continuous-time model over a time series of discrete forcing inputs; and by Stern (2005) as a way to enforce energy balance constraints on estimated parameters of an ARMA model.

This chapter introduces a novel application of the three-box EBM filter in its inverted form. Li and Jarvis (2009) showed that the three-box model is simply a physically motivated parameterization of a causal, linear input-output system which maps a time series of radiative forcings onto a corresponding time series of surface temperatures. Here it is shown how the k-box model's equivalent ARMA filter representation may be derived, and we describe how, by inverting the ARMA filter, radiative forcing may be obtained from a temperature time series.

4.3 Energy-balance models as ARMA filters

For an LTI system (such as the k-box model), the surface temperature response to a general forcing $F(t)$ can be written as

$$T_1(t) = \sum_{v=0}^{\infty} R(v)F(t-v). \tag{4.1}$$

where $R(v)$ is the discrete-time impulse response function. Since an impulse is the derivative of a step change, the discrete-time impulse response function is equal to the once-differenced response to a step change in forcing (e.g. a CO_2-quadrupling experiment). Defining the "backshift" operator B such that $\mathrm{B}^i\, x(t) = x(t-i)$, we can write

$$\sum_{v=0}^{\infty} R(v)F(t-v) = \sum_{v=0}^{\infty} R(v)\mathrm{B}^v\, F(t). \tag{4.2}$$

Let

$$\mathbf{\Phi}(\mathrm{B}) = \sum_{v=0}^{\infty} R(v)\mathrm{B}^v. \tag{4.3}$$

It can be shown (see Appendix B) that for the k-box model

$$\boldsymbol{\Phi}(\mathrm{B}) = \frac{\boldsymbol{\theta}(\mathrm{B})}{\boldsymbol{\phi}(\mathrm{B})}, \tag{4.4}$$

with $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ polynomials in B of degree k and $k-1$ respectively. Thus we have

$$\boldsymbol{\phi}(\mathrm{B})\, T_1(t) = \boldsymbol{\theta}(\mathrm{B})\, F(t), \tag{4.5}$$

i.e. the linear map from $F(t)$ onto $T_1(t)$ is an ARMA(k, $k-1$) filter. The representation in (4.4) as a ratio of two polynomials gives the corresponding inverse filter

$$\boldsymbol{\Phi}^{-1}(\mathrm{B}) = \frac{\boldsymbol{\phi}(\mathrm{B})}{\boldsymbol{\theta}(\mathrm{B})}, \tag{4.6}$$

which is ARMA($k-1$, k). The AR and MA coefficients can be efficiently calculated from the discrete-time impulse response using software for computing Padé approximants such as the *Pade* package in R (Adler, 2020).

Note that the above result for k-box EBMs is a special case of a long-standing and more general result: that a system of k first-order linear ODEs has a discrete-time representation as an ARMA(k, $k-1$) filter (Spolia and Chander, 1974). EBM-derived ARMA filters have previously been used in the climate literature in the "forward" direction, i.e. to convert time series of annually averaged forcings into temperatures (Grieser and Schönwiese, 2001), and for the purpose of EBM parameter estimation (Stern, 2005). Here the ARMA filter is applied in the reverse direction, to find the forcing input required by an EBM to yield a given time series of temperatures.

If polynomials $\boldsymbol{\phi}(\mathrm{B})$ and $\boldsymbol{\theta}(\mathrm{B})$ are taken to be

$$\boldsymbol{\phi}(\mathrm{B}) = 1 - \sum_{i=1}^{k} \phi_i \mathrm{B}^i \tag{4.7}$$

and

$$\boldsymbol{\theta}(\mathrm{B}) = \sum_{i=0}^{k-1} \theta_i \mathrm{B}^i, \tag{4.8}$$

then, in the case $k = 3$, eqn (4.5) may be written explicitly as

$$T_1(t) - \phi_1 T_1(t-1) - \phi_2 T_1(t-2) - \phi_3 T_1(t-3) = \theta_0 F(t) + \theta_1 F(t-1) + \theta_2 F(t-2). \tag{4.9}$$

4.4 Three-box climate model fits

In the previous chapter, a maximum likelihood method was developed for estimating parameters of k-box EBMs from abrupt CO_2-quadrupling GCM experiments. Using the maximum likelihood estimator, three-box models have been fitted to the two most recent Earth system models (ESMs) from the UK Met Office: HadGEM2-ES from CMIP5 and HadGEM3-GC3.1-LL from CMIP6 (see Figure 4.1). Table 4.1 contains maximum likelihood estimates of model parameters, as well as estimates of equilibrium climate sensitivity (ECS), transient climate response (TCR), characteristic timescales τ_1, τ_2, τ_3 and the coefficients of the ARMA polynomials $\boldsymbol{\phi}(\mathrm{B})$ and $\boldsymbol{\theta}(\mathrm{B})$. The ARMA coefficients were calculated from the discrete-time impulse responses of the fitted models. Note that ARMA models can alternatively be estimated using Bayesian inference (e.g. Monahan, 1983).

4.5 ARMA filter validation using HadGEM3-GC3.1-LL

4.5.1 Methods

Numerical estimates of the ARMA coefficients enable, in theory, conversion of time series of surface temperatures into corresponding series of radiative forcings. The properties of this proposed temperature-forcing conversion were investigated using CMIP6 historical runs from the HadGEM3-GC3.1-LL climate model.

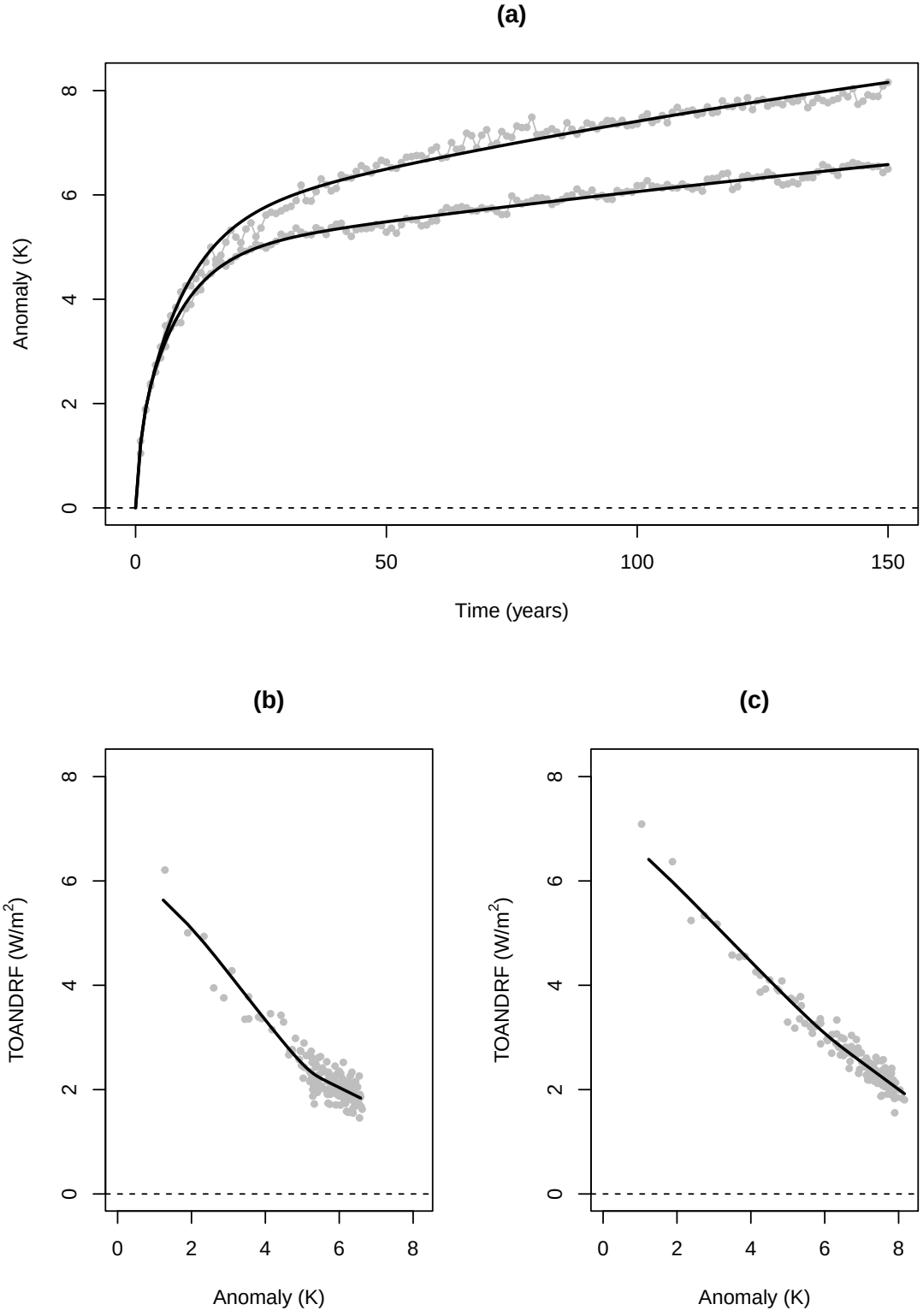

Figure 4.1: Three-box model fitted values. Panel (a) shows the abrupt $4 \times CO_2$ surface temperature responses of the HadGEM2-ES (cooler) and HadGEM3-GC3.1-LL (hotter) climate models. Anomalies are relative to the time average of surface temperature in the corresponding pre-industrial control (piControl) simulations. Panels (b) and (c) show the TOA net downward radiative flux change in the respective models as a function of surface temperature.

Table 4.1: Maximum likelihood parameter estimates. For descriptions of all model parameters see Table 3.1. The values of ECS and TCR in this table are derived directly from the box models' physical parameters.

	Model	
Parameter (Unit)	HadGEM2-ES	HadGEM3-GC3.1-LL
γ (dimensionless)	1.73	3.15
C_1 (W yr m^{-2} K^{-1})	3.62	3.97
C_2 (W yr m^{-2} K^{-1})	9.47	9.06
C_3 (W yr m^{-2} K^{-1})	98.7	65.9
C_{tot} (W yr m^{-2} K^{-1})	112	79
κ_1 (W m^{-2} K^{-1})	0.536	0.607
κ_2 (W m^{-2} K^{-1})	2.39	2.99
κ_3 (W m^{-2} K^{-1})	0.634	0.65
ε (dimensionless)	1.59	1.18
σ_η (W m^{-2})	0.434	0.458
σ_ξ (W m^{-2})	0.323	0.632
$F_{4\times\text{CO}_2}$ (W m^{-2})	6.35	7.19
τ_1 (yr)	0.953	0.822
τ_2 (yr)	8.21	9.04
τ_3 (yr)	532	270
ECS $= F_{4\times\text{CO}_2}/2\kappa_1$ (K)	5.92	5.92
TCR (K)	2.44	2.82
ϕ_1 (dimensionless)	2.23	2.19
ϕ_2 (dimensionless)	-1.54	-1.45
ϕ_3 (dimensionless)	0.31	0.264
θ_0 (K m^2 W^{-1})	0.194	0.172
θ_1 (K m^2 W^{-1})	-0.33	-0.284
θ_2 (K m^2 W^{-1})	0.136	0.113

Model surface temperatures from 1850-2014 were averaged annually (Jan-Dec), globally and over four ensemble members. Anomalies were calculated by subtracting from the whole series the mean absolute temperature in the first 50 years (1850-1899). The final temperature series was fed into the fitted HadGEM3-GC3.1-LL ARMA filter using the digital filter implementation in the *signal* package in R (signal developers, 2021).

The resulting filtered forcing series was compared (see Figure 4.2) with a time series of effective radiative forcing (ERF) diagnosed by Andrews et al. (2019) from a HadGEM3-GC3.1-LL RFMIP run using the ERF_trans method described in Forster et al. (2016). Note that Andrews et al. (2019) applied a time-mean historical volcanic forcing to the HadGEM3-GC3.1-LL piControl, meaning that they report an ERF of around 0.2 $\mathrm{W\,m^{-2}}$ in 1850. To avoid an offset due to the choice of temperature baseline, a constant 0.2 $\mathrm{W\,m^{-2}}$ was added to the forcing series computed using the ARMA filter.

Note that, in Section 4.3, no specific model for internal climate variability was assumed, in contrast with the stochastic EBM of the previous chapter, which has a parametric noise model. While the maximum likelihood estimator developed in Chapter 3 requires an explicit parameterization of internal climate variability, the present application of the ARMA filter as a recursive estimator of ERF does not. Nevertheless, the output of this procedure, when applied to GCM simulations, may be interpreted using the Chapter 3 parameterization as the sum of terms $F(t)$ and $\xi(t)$ in eqn (3.2). Thus there is some noise contamination of resulting estimates (see Section 4.6.3 for detailed discussion).

4.5.2 Results

Results from the two methods agree strongly and the filtered forcing series explains a majority of the variance in the ERF_trans series: the coefficient of determination is $R^2 = 0.69$ (correlation of 0.83). Reassuringly the ARMA filter correctly infers the rate of forcing increase in recent decades. However, there appears to be a systematic

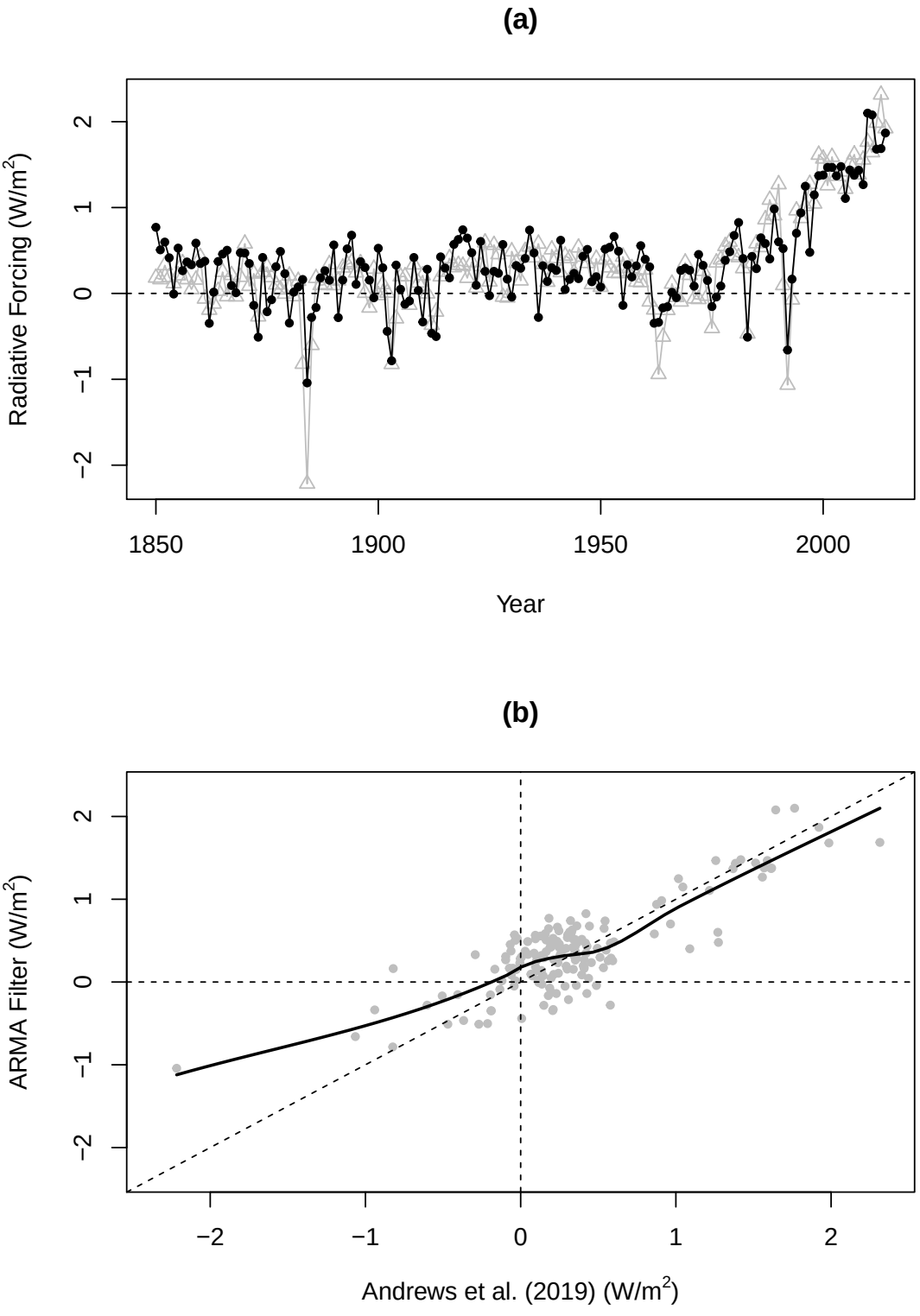

Figure 4.2: Comparison of forcing series estimated from HadGEM3-GC3.1-LL simulations of the historical period. In panel (a) grey triangles denote forcings estimated by Andrews et al. (2019) while black dots denote forcings obtained using the three-box ARMA filter. Panel (b) shows how estimates from the two methods vary over the range of forcing values using a non-parametric curve fit. The diagonal line has equation $y = x$.

disagreement in years with strong negative forcing. Negative forcing spikes associated with volanic eruptions are smaller when calculated using the ARMA filter. This is especially evident in the case of the Krakatoa explosion of 1883.

The observed discrepancy in inferred volcanic forcing is not entirely unexpected. The ARMA filter is derived from a three-box EBM which is known to be unable to resolve temperature responses on timescales significantly shorter than one year. Because the data used here are annual averages, there is also scope for error due to discretization, as a volcanic eruption might occur earlier or later in a given year. Finally, it may also be the case that the GCM's temperature response to volcanic forcing deviates from the linearity assumption of the ARMA filter.

4.6 Application to historical surface temperatures

4.6.1 Methods

The box-model ARMA filters can be applied to time series of historical surface temperatures to obtain series of estimated historical forcings. The Cowtan and Way 2.0 (CW2.0) historical temperature series is an updated version of the dataset described in Cowtan and Way (2014), which is a modification of the HadCRUT4 series (Morice et al., 2012), corrected for coverage bias. Because CW2.0 is a blend of surface air temperatures (SATs) and sea surface temperatures (SSTs), it is not directly comparable with the pure-SAT $4 \times CO_2$ datasets used to calibrate the EBMs. The CW2.0 temperatures series have therefore been scaled by 1.09 according to Richardson et al. (2016) for the present application. The HadGEM2-ES and HadGEM3-GC3.1-LL ARMA filters have been applied to the scaled CW2.0 temperatures yielding estimated historical forcing series for the period 1850-2018 (see Figure 4.3).

4.6.2 Results

It can be seen from panel (d) of Figure 4.3 that the forcing series generated by the two ARMA filters are very similar. The HadGEM3-GC3.1-LL ARMA filter reports

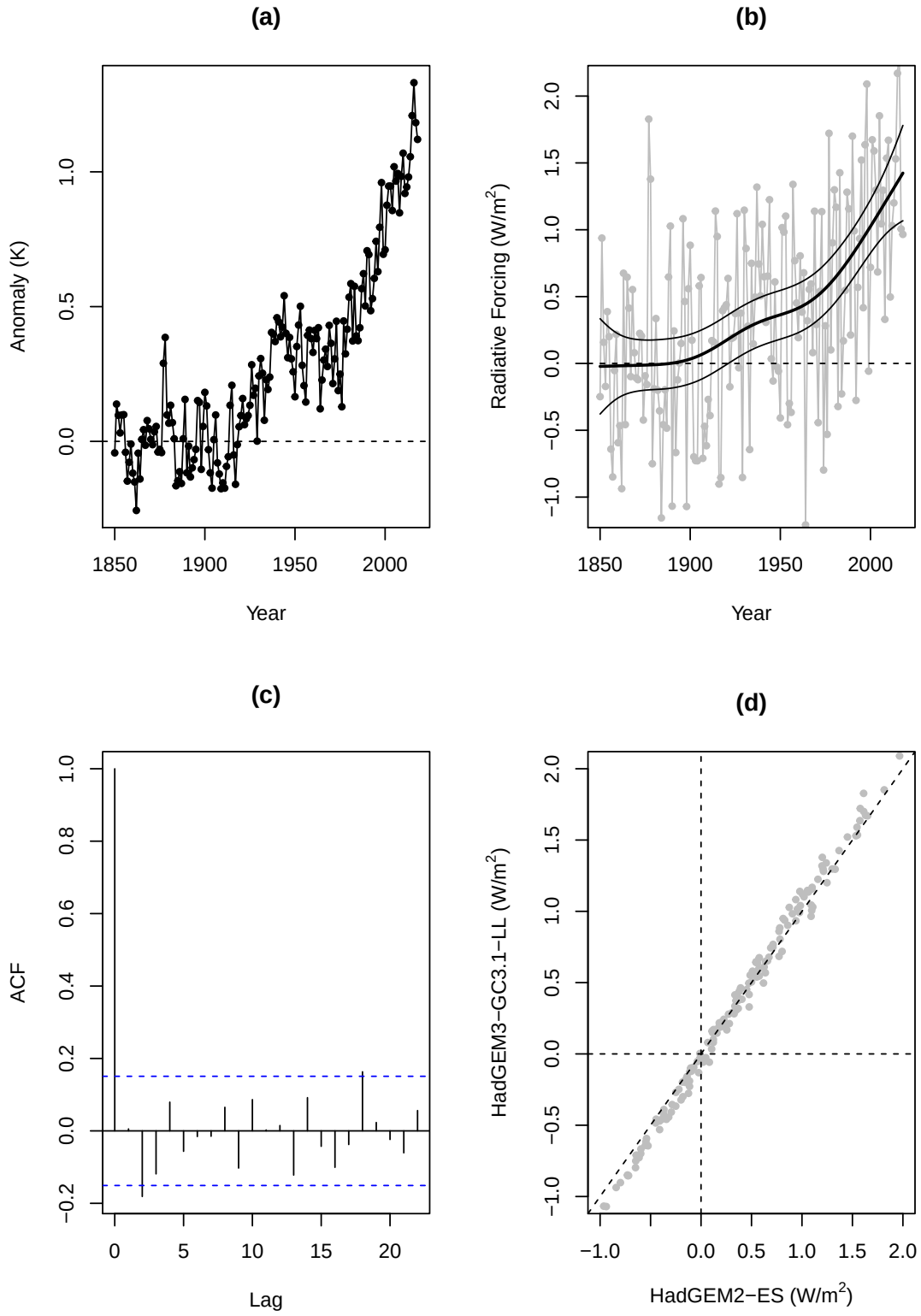

Figure 4.3: HadGEM3-GC3.1-LL three-box ARMA filter reconstruction of historical radiative forcing. Panel (a) is the Cowtan and Way 2.0 temperature series; (b) is the corresponding filtered forcing series, with estimated trend and 95 % confidence intervals from a generalized additive model (GAM) fit; (c) is the sample autocorrelation function of the GAM residuals; (d) shows how estimates from the two models vary over the range of forcing values.

slightly higher forcing in the latter years of the historical period and slightly more extreme negative forcing in years surrounding major volcanic eruptions.

The reconstructed forcing series are very noisy since the natural variability which contaminates historical surface temperatures is amplified by the ARMA filter. However, unlike noise in the temperature series, the filtered noise is essentially uncorrelated in time. This follows from the fact that the three-box EBM successfully accounts for the thermal inertia (memory) of the system. The white noise-like properties of natural variability in the filtered forcing series mean that the long-term trend can be extracted using regression techniques. A generalized additive model (GAM) was fitted to the estimated forcing series using the *mgcv* package in R (Wood, 2011, 2021). Approximate 95 % confidence intervals (± 1.96 standard errors) for the estimated trend indicate a dramatic acceleration in radiative forcing in the second half of the twentieth century, with an estimated forcing increase of 1.45 ± 0.504 $\mathrm{W\,m^{-2}}$ between 1850 and 2018.

4.6.3 Quantification of uncertainty

The filtered forcing series and its subsequent decomposition into signal and noise are subject to multiple sources of uncertainty which must be taken into consideration. As well as the (substantial) internal climate variability, there is observational uncertainty in the historical temperature series and parameter uncertainty in the fitted impulse responses. Given a standard assumption of zero-mean errors, the aggregate noise from observational error and internal climate variability should be well accounted for by GAM regression smoothing. Uncertainty in the fitted impulse responses is also of limited concern: since the EBMs were fitted to abrupt CO_2-quadrupling experiments, which have a good signal-to-noise ratio, uncertainty in the fitted impulse responses is in fact quite small (see Figure 4.1). It should however be noted that, under forcing scenarios less drastic than $4 \times CO_2$, fitting of exponential response functions can suffer from ill-conditioning (De Groen and De Moor, 1987; Kaufmann, 2003). Of arguably greater concern is uncertainty due

to inter-model variation between GCMs, which can be seen as a measure of confidence in a particular GCM's ability to represent the true climate. Inter-model variation in GCM output can arise through the use of different parameter values and/or model structures in GCMs (Flato et al., 2013). A full multi-model ensemble, perhaps across all the GCMs in CMIP6, would be required to quantify this satisfactorily, as HadGEM2-ES and HadGEM3-GC3.1-LL are clearly not independent samples. Numerical results in this chapter should therefore be seen as conditional on those two climate models.

While one might argue that the use of a post-hoc GAM regression is reasonable for the reasons given above, a more integrated approach to uncertainty quantification in future analyses could be achieved using Bayesian methods. A Bayesian alternative to GAM smoothing of the filtered forcing series is the "latent force model" approach (Alvarez et al., 2009; Särkkä et al., 2019). In a latent force model, uncertainty in the unobserved historical forcing input to a system of ODEs is represented in continuous time using a Gaussian process. Another alternative is the use of sequential methods based on the Kalman filter (Kalman, 1960). ARMA models have natural representations as linear Kalman filters (de Jong and Penzer, 2004), and sequential filtering methods can account for uncertainty in both model parameters and unknown inputs simultaneously, by augmenting the system state vector with estimates of the uncertain parameters (Lourens et al., 2012; Yu and Chakravorty, 2015). Similar techniques have previously been used in the climate literature: for example, by Annan et al. (2005) and Padilla et al. (2011), who used Kalman filtering to sequentially tune parameters of an Earth model of intermediate complexity (EMIC) and a two-box EBM respectively; and by Cohen and Wang (2014), who estimated historical time series of global black carbon emissions.

4.7 Summary

A method has been developed for estimating radiative forcing from time series of surface temperatures using GCM-calibrated, k-box EBMs. There is a known cor-

respondence between k-box EBMs and ARMA models. In this chapter it has been shown how, by inverting a k-box EBM's equivalent ARMA filter representation, a convenient mapping from temperatures onto forcings may be obtained. The method has been validated using historical simulations from the HadGEM3-GC3.1-LL climate model and found to generally perform well, the notable exception being negative forcing due to volcanic eruptions, which was underestimated. Three-box EBMs fitted to the HadGEM2-ES and HadGEM3-GC3.1-LL climate models have been combined with the Cowtan and Way 2.0 temperature dataset to produce estimates of radiative forcing for the historical period (1850-2018). Forcing estimates calculated using the two models' ARMA filters are very similar. A significant increase in radiative forcing over the historical period has been detected at the 5% level of significance.

The method developed in this chapter is a direct application of energy-balance model theory to the D&A of past temperature changes, by examining patterns of forcings rather than temperatures. Future work will examine how performance of the method generalizes to the other GCMs in CMIP6, and will address in more detail the question of uncertainty quantification. By combining ARMA filter estimates of historical radiative forcing with known observational constraints it may be possible to further constrain climate sensitivity metrics, such as ECS and TCR, and hence constrain projections of future warming.

In the next chapter, the principal research questions of this thesis will be addressed. It will be determined whether, and if so under what circumstances, optimal fingerprinting methods for D&A of climate change trends are valid. The k-box-model ARMA representation, explored in this chapter, will be found to have considerable value as an algebraic reasoning device, not just as a numerical tool as has been shown here. The material developed in this and the previous chapter will provide physical interpretability to the next chapter's results and their respective proofs.

Chapter 5

Could detection and attribution of climate change trends be spurious regression?

5.1 Aim

This chapter investigates, by modelling radiative forcing as an integrated process within an idealized linear-response-model framework, whether the optimal fingerprinting estimator is consistent under standard assumptions, and hence robust against spurious regression. Hypothesis tests are conducted using observations of historical GMST and simulation output from 13 CMIP6 general circulation models (GCMs) to determine whether these assumptions hold in practice.

5.2 Introduction

The previous chapter established a link between energy-balance models (EBMs), simple models of global climate parameterized in terms of real-world physical quantities, and general autoregressive moving-average (ARMA) linear time series models. In this chapter, it will be shown how representing the forced responses of climate models as general ARMA filters (a much wider class of causal time-series mod-

els than just EBMs) provides a useful framework for analyzing how fingerprinting methods work when applied to trending variables. Recall that optimal fingerprinting assumes a regression model of the form

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{5.1}$$

where $\boldsymbol{y}$ denotes historical climate observations; X is a matrix of predicted climate-change signals, typically consisting of simulation output from a GCM; coefficients $\boldsymbol{\beta}$ are the "scaling factors"; and $\boldsymbol{e}$ is a composite error term containing internal climate variability noise as well as other sources of uncertainty. Detection and attribution (D&A) involves obtaining reliable estimates of these scaling factors and establishing their statistical significance, typically through the use of least-squares estimators (see Chapter 2).

A key feature of the optimal fingerprinting studies reviewed in Chapter 2 is the way in which datasets of climate observations and model simulation output are treated as static signals in the regression, rather than outputs of an input-output system evolving over time. For example, in spatio-temporal D&A studies the spatial and temporal correlations are treated similarly through the use of combined space-time covariance matrices. One problem with this approach is that properties unique to time-indexed data, such as notions of ordinality and causality, are not given adequate consideration. It has long been known in the field of time series analysis that the practice of regressing variables containing time trends comes with a unique set of pitfalls, the most serious of which being the "spurious regression" phenomenon (discussed in detail in the next section). In this chapter, the threat to D&A inferences posed by spurious regression will be assessed, using a combination of theoretical argument (Section 5.4) and empirical evidence (Section 5.5). The statistical consistency of least-squares estimators used in optimal fingerprinting studies will be found to depend on a property known as "cointegration", which may be described informally as a stricter form of correlation arising between time series. Section 5.5 will also introduce and demonstrate a new method for estimating

uncertainty in D&A results, which exploits the notion of cointegration to obviate dependence on pre-industrial control (piControl) simulations (and the strong assumptions therewith).

5.3 Regression of non-stationary variables

When discussing optimal fingerprinting as applied to time-indexed climate variables, it is necessary to introduce some definitions from time series analysis, the most important of which being the notion of "stationarity". In this chapter, a time series variable $u(t)$ is said to be stationary if and only if its mean and variance are finite and do not depend on time t. Such a time series exhibits a mean-reverting behaviour. An example of a stationary time series is the first-order autoregressive or AR(1) model

$$u(t) = \rho u(t-1) + \varepsilon(t), \tag{5.2}$$

where $-1 < \rho < 1$ is a correlation and $\varepsilon(t)$ a white-noise process, commonly called a "shock" or "innovation". The AR(1) model in eqn (5.2) has mean zero and constant variance $\sigma^2/(1-\rho^2)$, where $\sigma^2 = \mathrm{Var}(\varepsilon)$. Stationary autoregressive processes have been proposed as simple models of internal climate variability (Hasselmann, 1976). If a trend in a time series can be described as a change in the mean (deterministic trend) or variance (stochastic trend) over time then, by their definition, stationary time series do not exhibit trends.

Detection and attribution is concerned with "trending" (non-stationary) climate variables. While the above definition of stationarity is quite prescriptive, the corresponding class of non-stationary time series, i.e. those violating the conditions, is too broad to be practically useful for the present purposes. Instead, attention will be restricted to the class of non-stationary time series known as "integrated" or "difference-stationary". For a time series $v(t)$ to be difference-stationary, the differenced series $\Delta v(t) = v(t) - v(t-1)$ must be stationary. The equivalent term "integrated" comes from the fact that $v(t)$ can be constructed by integrating (taking

partial sums of) the stationary time series $\Delta v(t)$. In this chapter, integrated will be abbreviated to I(1), where 1, the "order of integration", denotes the number of times a series must be differenced to achieve stationarity. An example of an I(1) time series, the simple random walk, can be obtained from eqn (5.2) by setting $\rho = 1$:

$$v(t) = v(t-1) + \varepsilon(t). \tag{5.3}$$

The variance of $v(t)$ grows linearly in time, so the series is non-stationary. It may be said that $v(t)$ contains a "unit root" stochastic trend (see Section 5.4).

The climate variables in D&A studies are hypothesized to be non-stationary, due to the presence of externally forced trends, both natural and anthropogenic. Radiative forcing due to greenhouse gas (GHG) emissions has a natural representation as an integrated process, where the integration is the accumulation of gases in the atmosphere over time. The idea of unit-root stochastic trends has a long history in climate change studies, e.g. Kaufmann and Stern (1997); Stern and Kaufmann (2000); Kaufmann and Stern (2002); Kaufmann et al. (2006); Mills (2008); Kaufmann et al. (2011, 2013), although it has been disputed (Gay-Garcia et al., 2009). It might further be argued that the processes driving anthropogenic GHG emissions are themselves integrated, where the integration represents accumulation of industrial capacity, however there is no numerical evidence for significant higher-order integration in annual records of historical GMST (see Section 5.4). In optimal fingerprinting, observed and simulated realizations of non-stationary variables are commonly regressed on one another using classical estimators such as ordinary least squares (OLS) (Allen and Tett, 1999) or total least squares (TLS) (Allen and Stott, 2003), depending on the size of the GCM ensemble. However, it has long been known that regressions involving non-stationary variables are susceptible to a phenomenon called "spurious regression", whereby statistically significant linear relationships are found between completely unrelated time series (e.g. Yule, 1926). Granger and Newbold (1974) showed that regressing two independent random walks produces inflated t-statistics and often leads to the detection of a statistically significant relationship

when in reality none exists. In general, OLS regressions of I(1) time series are statistically inconsistent, i.e. the coefficient estimates do not converge in the limit of infinite data, except in the special case where the series are "cointegrated" (Engle and Granger, 1987). Two or more I(1) time series are said to cointegrate when there exists a linear combination of the series which is itself stationary. Regressing cointegrated time series using OLS yields coefficient estimates which are not only consistent but "superconsistent", meaning they converge in probability to the coefficients' true values at a rate proportional to the length of the series (Engle and Granger, 1987). Thus the question of whether the regressors in optimal fingerprinting are cointegrated is critical for evaluating the reliability of D&A of climate change trends.

The risk of spurious regression in D&A pertains specifically to the attribution problem. This is because, in the case of detection, p-values are calculated "under the null", i.e. under an assumption of no climate change. In the absence of climate change, the left-hand side of eqn (5.1) would be stationary by definition, and there would be no risk of spurious regression. In the case of attribution, where climate change is taken as given, spurious regression refers to the misattribution of climate trends to one or more candidate factors, meaning that the resulting allocation of blame is inaccurate. Such misattribution does not require the presence of an "exogenous" trend (e.g. caused by a hidden forcing mechanism), but may instead be caused by flawed representation of forced trends included in the climate model. In particular, any discrepancy between true and modelled forcing which accumulates over time (e.g. due to inaccurate data / incomplete understanding of physical processes) has the potential to induce non-stationarity in the error term of the regression equation, leading to inconsistent scaling factor estimates and invalid confidence intervals. Given that climate models are known to differ in their representation of radiative forcings, there is a prima facie case for investigating this possibility (Myhre et al., 2013).

Methods based on the notion of cointegration have been used previously in anal-

yses of climatic time series (Bindoff et al., 2013). Much effort has gone into studying cointegrations between groups of real-world variables, such as temperatures and forcings (Stern, 2006; Turasie, 2012; Beenstock et al., 2012; Stern and Kaufmann, 2014; Pretis et al., 2015; Storelvmo et al., 2016; Estrada and Perron, 2017; Bruns et al., 2020) or temperatures and sea level (Schmith et al., 2012). However, little effort has gone into discussing the presence or lack of cointegration between observed and model-simulated realizations of the same climate variable.

This chapter addresses the two primary aims of the thesis: firstly, to determine mathematically whether the experimental design and model assumptions of optimal fingerprinting together imply cointegration of the regression and therefore consistency of the least squares estimator; secondly, to investigate whether there is empirical evidence of such a cointegration arising in practice for the GMST variable. The first aim will be addressed in Section 5.4 and key results proved within an idealized linear-response-model framework. It will be shown how, by parameterizing the impulse response as an EBM, the formulas in the proof can assume physical interpretability in terms of real-world quantities. Section 5.5 deals with the second aim by means of hypothesis testing, applied to historical observations and output from the latest generation of GCMs. A new method for calculating confidence regions without recourse to piControl simulations will also be introduced. The content of the chapter is summarized in Section 5.6.

5.4 Theoretical arguments for cointegration

This section will assess the consistency of optimal fingerprinting regression in the presence of I(1) non-stationary forcings. To begin with, some definitions are required.

5.4.1 Impulse-response model definition

Let y denote a climate variable of interest for which historical observations are available. Assuming that a change in y in response to an externally imposed effective radiative forcing (ERF) F may be adequately described by a linear and time-invariant (LTI) impulse-response function, the time series of observations $y(t)$ may be written as an autoregressive moving-average (ARMA) model of arbitrary order $p, q \geq 0$,

$$y(t) - \sum_{i=1}^{p} \phi_i y(t-i) = c + \sum_{i=0}^{q} \theta_i F(t-i) + \xi(t), \tag{5.4}$$

where $\mu = c/(1 - \sum_i \phi_i)$ is variable y's pre-industrial baseline, i.e. its mean value in the absence of any forcing F; coefficients ϕ_i and θ_i are sequences of weights determining the autoregressive (AR) and moving-average (MA) parts of the impulse-response function; and $\xi(t)$ is a stationary zero-mean stochastic process representing internal climate variability, plus other sources of noise/uncertainty reasonably considered stationary, such as observational error. The definition of ERF is given in Myhre et al. (2013) and is such that the climate system's response to ERF should be indifferent to the particular forcing agent responsible. The ARMA model in eqn (5.4) is very general, incorporating all finite-impulse-response (FIR) models, as well as all infinite-impulse-response (IIR) models of exponential type. Defining the "backshift operator" B such that $\mathrm{B}^i x(t) = x(t-i)$, eqn (5.4) may be written

$$y(t) = \mu + \mathbf{\Phi}(\mathrm{B}) F(t) + \varepsilon(t), \tag{5.5}$$

where the rational function

$$\mathbf{\Phi}(\mathrm{B}) = \frac{\boldsymbol{\theta}(\mathrm{B})}{\boldsymbol{\phi}(\mathrm{B})} = \frac{\sum_{i=0}^{q} \theta_i \mathrm{B}^i}{1 - \sum_{i=1}^{p} \phi_i \mathrm{B}^i}, \tag{5.6}$$

is known as the "transfer function". The time series of radiative forcings $F(t)$ is assumed to be non-stationary with non-stationarity modelled as I(1),

$$F(t) = F(t-1) + \Delta F(t), \tag{5.7}$$

where the series of forcing increments $\Delta F(t)$ is a stationary stochastic process. The rationale for modelling $\Delta F(t)$ as stochastic is the fact that it cannot be predicted from past values of $F(t)$ only, as evidenced by the impact on carbon dioxide (CO_2) emissions of the recent SARS-CoV-2 pandemic (Tollefson, 2021). Equation (5.7) is quite general since, beyond the assumption of stationarity, no specific parametric model is assumed for $\Delta F(t)$. Note that $\Delta F(t)$ need not have zero mean: for example, in the case of forcing due to exponentially increasing atmospheric CO_2 concentration, $\Delta F(t)$ would on average (and indeed almost always) take a positive value. Using the backshift operator,

$$F(t) = \frac{1}{1-\mathrm{B}}\Delta F(t), \tag{5.8}$$

whence the term "unit-root" non-stationarity originates, as the polynomial in the transfer function's denominator contains a unit root. To guarantee the existence of a finite climate sensitivity, it is assumed that all roots of the AR polynomial $\boldsymbol{\phi}(\mathrm{B})$ lie strictly outside the unit circle in the complex plane. Noting that the backshift operator B reduces to the identity when the system is in equilibrium, the familiar equilibrium climate sensitivity (ECS) for variable y is then

$$y_{\mathrm{ECS}} = \boldsymbol{\Phi}(1) F_{2\times \mathrm{CO}_2}, \tag{5.9}$$

where $F_{2\times \mathrm{CO}_2}$ denotes the increase in radiative forcing associated with a doubling of atmospheric CO_2 concentration.

5.4.2 Optimal fingerprinting experimental design

Consider an optimal fingerprinting study where observed changes in y are to be attributed to a set of p candidate forcings $F_1, \ldots, F_p$. Using the ARMA model in eqn (5.4) and the backshift operator notation, the study's experimental design may be written

$$y(t) = \mu + \mathbf{\Phi}(\mathrm{B})F(t) + \varepsilon(t), \tag{5.10}$$

$$x_1(t) = \mu' + \mathbf{\Phi}'(\mathrm{B})\pi_1 F_1(t) + \varepsilon_1'(t), \tag{5.11}$$

$$\vdots$$

$$x_p(t) = \mu' + \mathbf{\Phi}'(\mathrm{B})\pi_p F_p(t) + \varepsilon_p'(t). \tag{5.12}$$

The equation for observations $y(t)$ is unchanged. New variables x_i denote output from climate model runs (or ensembles thereof) where forcings F_i have been applied individually. If no important forcing factors are missing from the candidate set, and if there are no interactions between forcing factors, it may be assumed that the total forcing F driving the observed trend in y has the decomposition $F = F_1 + \cdots + F_p$. This is the "additivity assumption" of optimal fingerprinting. Another fundamental assumption of optimal fingerprinting is that radiative forcings driving model runs x_i have the correct temporal structure, i.e. are identical to their real-world counterparts up to multiplicative constants π_i. In practice this assumption may be relaxed using errors-in-variables (EIV) methods, but at the cost of introducing further assumptions such as model exchangeability (Huntingford et al., 2006). Note that, in eqns (5.11) to (5.12), the climate model is not assumed to perfectly reproduce the properties of the true climate. In general, the climate model may have a different mean $\mu' \neq \mu$, a different noise process $\varepsilon' \neq \varepsilon$, and an impulse response differing from the truth in shape and scale $\mathbf{\Phi}' \neq \mathbf{\Phi}$.

5.4.3 Consistency of the least-squares estimator

If at least one of the candidate forcings is I(1) non-stationary then a multiple regression of y on $x_1, \ldots, x_p$ is integrated on both sides of eqn (5.1). An integrated regression of this type is known to be consistent if and only if the I(1) variables are cointegrated (Engle and Granger, 1987). To establish consistency of optimal fingerprinting, as described by eqns (5.10) to (5.12), it is therefore necessary and sufficient to prove that y and $x_1, \ldots, x_p$ cointegrate. From the definition of cointegration, this may be achieved by proving the existence of a linear combination of the x_i which, when subtracted from y, yields a stationary process.

Lemma. *The rational transfer function* $\boldsymbol{\Phi}(\mathrm{B}) = \boldsymbol{\theta}(\mathrm{B})/\boldsymbol{\phi}(\mathrm{B})$ *permits the following decomposition:*

$$\boldsymbol{\Phi}(\mathrm{B}) = \boldsymbol{\Phi}(1) - (1 - \mathrm{B})\frac{\boldsymbol{\psi}(\mathrm{B})}{\boldsymbol{\omega}(\mathrm{B})}, \tag{5.13}$$

where $\boldsymbol{\psi}(\mathrm{B})$ *and* $\boldsymbol{\omega}(\mathrm{B})$ *are polynomial operators, with* $\boldsymbol{\omega}(\mathrm{B})$ *containing no unit root.*

Proof. Define the abstract rational function

$$\boldsymbol{\Omega}(z) = \boldsymbol{\Phi}(1) - \boldsymbol{\Phi}(z), \tag{5.14}$$

where $\boldsymbol{\Phi}(z) = \boldsymbol{\theta}(z)/\boldsymbol{\phi}(z)$ as before. Since $\boldsymbol{\phi}(z)$ has no unit root, $\boldsymbol{\Phi}(1)$ is a finite constant, and it follows that $\boldsymbol{\Omega}(z)$ has no pole at $z = 1$. From eqn (5.14) it may be seen that $\boldsymbol{\Omega}(1) = 0$. Thus $\boldsymbol{\Omega}(z)$ may be factorized

$$\boldsymbol{\Omega}(z) = (1 - z)\frac{\boldsymbol{\psi}(z)}{\boldsymbol{\omega}(z)}, \tag{5.15}$$

where $\boldsymbol{\omega}(z)$ contains no unit root. □

Theorem. *There exists a p-vector of coefficients* $(\beta_1, \ldots, \beta_p)'$ *such that the linear combination*

$$r(t) = y(t) - \beta_1 x_1(t) - \cdots - \beta_p x_p(t) \tag{5.16}$$

is a stationary time series. Specifically, $r(t)$ is stationary when

$$\beta_i = \frac{1}{\pi_i} \frac{\boldsymbol{\Phi}(1)}{\boldsymbol{\Phi}'(1)} \tag{5.17}$$

for all i in $1, \ldots, p$.

Proof. Applying the lemma to eqn (5.10) yields

$$y(t) = \mu + \boldsymbol{\Phi}(1)F(t) - (1 - \mathrm{B})\frac{\boldsymbol{\psi}(\mathrm{B})}{\boldsymbol{\omega}(\mathrm{B})}F(t) + \varepsilon(t). \tag{5.18}$$

Substituting eqn (5.8) into (5.18) gives

$$y(t) = \mu + \boldsymbol{\Phi}(1)F(t) - (1 - \mathrm{B})\frac{\boldsymbol{\psi}(\mathrm{B})}{\boldsymbol{\omega}(\mathrm{B})}\frac{1}{1 - \mathrm{B}}\Delta F(t) + \varepsilon(t) \tag{5.19}$$

$$= \mu + \boldsymbol{\Phi}(1)F(t) - \frac{\boldsymbol{\psi}(\mathrm{B})}{\boldsymbol{\omega}(\mathrm{B})}\Delta F(t) + \varepsilon(t). \tag{5.20}$$

Observe that all terms on the right-hand side (RHS) are stationary except for $\boldsymbol{\Phi}(1)F(t)$, so the non-stationary component of the forced response is simply a scaled version of the forcing series. This holds similarly for series of model output x_i and their respective forcings. It therefore follows that the non-stationarity in y due to forcing F_i may be eliminated by subtracting an appropriately scaled version of the corresponding model output series x_i. Expressions for the scaling factors β_i in eqn (5.17) are readily obtained by considering the relative magnitudes of the non-stationary components of the forced responses. □

Thus it has been established that the optimal fingerprinting regression described in this section is cointegrated, given standard model assumptions, and may be consistently estimated using OLS. The presence of cointegration also renders the OLS estimator superconsistent (Engle and Granger, 1987). In practice, the experimental design of a D&A study can be more complicated: GCM simulations are often run with linearly independent combinations of forcing factors, rather than each forcing being applied separately, in order to reduce collinearity of the forced responses (Jones et al., 2016; Jones and Kennedy, 2017). Due to the additivity assumption of

optimal fingerprinting, the reasoning applied in this section holds similarly in the case of linear combinations of forcings.

5.4.4 Energy-balance model parameterization

The result presented above holds for a general LTI impulse-response model of the form given in eqn (5.4). By choosing a suitable parameterization for the impulse response, this result can be given some physical interpretability. For example, when variable y denotes GMST, the impulse response may be parameterized as a k-box EBM, which is known to have a discrete-time representation as an ARMA(k, $k-1$) filter (Cummins et al., 2020a). In the simplest case, when $k = 1$, the EBM reduces to a single ordinary differential equation,

$$C\dot{T}(t) = F(t) - \lambda(T(t) - T_0), \tag{5.21}$$

where T (K) denotes GMST, T_0 (K) is the pre-industrial baseline temperature, C (W yr m^{-2} K^{-1}) is a heat capacity, and λ (W m^{-2} K^{-1}) is the climate feedback parameter. When $k > 1$ the GMST "box" is coupled to a system of additional boxes representing the heat capacity of the deep ocean. Recent studies have identified $k = 3$ as the optimal EBM complexity for reproducing the thermal characteristics of recent-generation GCMs (Caldeira and Myhrvold, 2013; Tsutsui, 2017; Fredriksen and Rypdal, 2017; Cummins et al., 2020b). Unfortunately, analytical solutions to k-box models quickly become quite complicated, even for $k = 2$ (Geoffroy et al., 2013a), so in this illustrative example equations are shown for the one-box model only.

If radiative forcing is assumed constant between timesteps, i.e. $F(t) = F(s)$ for $s \in (t-1, t]$, eqn (5.21) can be discretized and written in the form of eqn (5.10):

$$T(t) = T_0 + \frac{\lambda^{-1}(1 - \mathrm{e}^{-\lambda/C})}{1 - \mathrm{e}^{-\lambda/C}\mathrm{B}} F(t), \tag{5.22}$$

which may be decomposed into

$$T(t) = T_0 + \left[\frac{1}{\lambda} - (1 - \mathrm{B})\frac{\lambda^{-1}\mathrm{e}^{-\lambda/C}}{1 - \mathrm{e}^{-\lambda/C}\mathrm{B}}\right] F(t) \tag{5.23}$$

$$= T_0 + \frac{F(t)}{\lambda} - \frac{\lambda^{-1}\mathrm{e}^{-\lambda/C}}{1 - \mathrm{e}^{-\lambda/C}\mathrm{B}}\Delta F(t). \tag{5.24}$$

The non-stationary component of $T(t)$ is simply the input forcing $F(t)$ scaled by the climate sensitivity λ^{-1}, while the stationary component is an AR(1)-filtered version of the forcing increment series $\Delta F(t)$. The AR(1) filter is stationary because λ and C are both strictly positive. For EBMs with $k > 1$ the ARMA filter applied to $\Delta F(t)$ in eqn (5.24) will have higher-order polynomials in the numerator and denominator, however the non-stationary component will be unchanged as this term depends only on the climate feedback parameter λ.

Returning to the question of D&A, let $y(t)$ and $x(t)$ denote time series of observed and GCM-simulated historical GMST, driven by forcing series $F(t)$ and $\pi F(t)$ respectively. If temperature series $y(t)$ and $x(t)$ are adequately described by k-box EBMs (not necessarily of the same order) with respective climate feedback parameters λ and λ', then it follows from the theorem that

$$r(t) = y(t) - \frac{\lambda'}{\pi\lambda}x(t) \tag{5.25}$$

is a stationary time series. It also follows that the estimator $\hat{\beta}_{\mathrm{OLS}}$ obtained by an OLS regression of $y(t)$ on $x(t)$ is a superconsistent estimator of $\beta = \lambda'/(\pi\lambda)$ (Engle and Granger, 1987).

5.5 Empirical evidence

The previous section established that, under certain conditions, the variables in optimal fingerprinting regression are provably cointegrated, implying consistency of least-squares parameter estimation. Although cointegration is predicted by the theory, whether it arises in reality will depend on the validity of the model assumptions.

Two of the main assumptions used to obtain results in Section 5.4 are standard in optimal fingerprinting:

1. additivity, that the combined effect of multiple forcing factors is the sum of their effects had they been applied separately;

2. correct forcing specification, that radiative forcings in GCMs have correct temporal structure up to a multiplicative constant.

Identifying these assumptions as necessary prerequisites for cointegration of GCM output and historical observations allows them to be assessed using numerical cointegration tests. If there is strong numerical evidence of cointegration, then this gives no reason to doubt the validity of assumptions 1 and 2, and by extension the consistency of optimal fingerprinting. On the other hand, should significant cointegration fail to be detected, then the possibility of violated assumptions, spurious regression and meaningless results cannot be discounted without further investigation.

Two further assumptions were made in Section 5.4 which are non-standard in optimal fingerprinting. Firstly, it was assumed that non-stationary forcings are I(1), for reasons set out in Section 5.2. Since the true underlying forcing series are not directly observable, it is infeasible to assess this assumption directly. However, as with the standard optimal fingerprinting assumptions above, the idealized concept of I(1) forcings may be shown "not inconsistent" with observation in the event that significant cointegration is detected. The second non-standard assumption is that of LTI impulse responses, which may be seen as a strengthening of the standard additivity assumption. To limit the influence of this strengthening, numerical results in this section have been calculated using the GMST climate variable, whose response to a radiative forcing perturbation is known to be well-modelled using LTI impulse responses (Li and Jarvis, 2009; Good et al., 2011; Geoffroy et al., 2013a).

5.5.1 Data

The numerical analyses in this section were performed using observed and GCM-simulated time series of GMST, averaged annually (Jan-Dec) for the period 1880-2014.

Observational datasets were, in alphabetical order: Berkeley Earth (Rohde and Hausfather, 2020), Cowtan and Way 2.0 (Cowtan and Way, 2014), GISTEMP v4 (Lenssen et al., 2019; GISTEMP Team, 2021), HadCRUT5 (Morice et al., 2021) and NOAAGlobalTemp V5 (Smith et al., 2008; Huai-Min Zhang et al., 2019). The choice of observational dataset was found not to affect hypothesis test results. The results presented here were calculated using HadCRUT5, however the whole analysis may be re-run for the other observational datasets by changing a single line of code (see Appendix C for details).

Predicted climate change signals were calculated using simulation output from 13 GCMs of the CMIP6 generation (Eyring et al., 2016). Chosen models are from modelling centres who have contributed runs as part of the DAMIP project (Gillett et al., 2016). For each GCM, ensemble-mean annual-GMST time series were calculated for the *historical* and *hist-GHG* experiments. These two forcing scenarios were chosen because GHG-attributable warming is of primary interest. Jones et al. (2016) recommend a two-way attribution of this form on the grounds of robustness. Table 5.1 gives the respective sizes of the *historical* and *hist-GHG* ensembles for each GCM, as well as the corresponding model citations.

5.5.2 Cointegration tests

Let y denote observed historical GMST and let x_1, x_2 denote GCM-predicted signals corresponding to the *historical* and *hist-GHG* experiments respectively. The theory in Section 5.4 predicts that the time series $y(t), x_1(t), x_2(t)$ are cointegrated. A simple test for cointegration consists of fitting the linear regression model

$$y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \varepsilon(t) \tag{5.26}$$

Model	# *historical*	# *hist-GHG*	Citation
ACCESS-ESM1-5	20	3	Ziehn et al. (2020)
BCC-CSM2-MR	3	3	Wu et al. (2019)
CanESM5	65	50	Swart et al. (2019)
CESM2	11	3	Danabasoglu et al. (2020)
CNRM-CM6-1	30	10	Voldoire et al. (2019)
FGOALS-g3	6	3	Li et al. (2020)
GFDL-ESM4	3	1	Dunne et al. (2020)
GISS-E2-1-G	46	10	Kelley et al. (2020)
HadGEM3-GC31-LL	4	4	Williams et al. (2018)
IPSL-CM6A-LR	32	10	Boucher et al. (2020)
MIROC6	50	3	Tatebe et al. (2019)
MRI-ESM2-0	7	5	Yukimoto et al. (2019)
NorESM2-LM	3	3	Seland et al. (2020)

Table 5.1: CMIP6 climate model ensemble sizes and citations.

using OLS and then testing the series of residuals $\hat{\varepsilon}(t)$ for stationarity (Engle and Granger, 1987). Residual stationarity may be tested using the procedure of Dickey and Fuller (1979), whereby the autoregression

$$\Delta\hat{\varepsilon}(t) = \delta_0 + \delta_1\hat{\varepsilon}(t-1) + u(t) \tag{5.27}$$

is estimated using OLS and the t-statistic corresponding to $\hat{\delta}_1$, denoted $\hat{\tau}_c$, compared with a relevant quantile of the reference "Dickey-Fuller" distribution. The null hypothesis H_0 is that no cointegrating relationship exists between $y(t), x_1(t), x_2(t)$ and the residuals exhibit unit-root non-stationarity. A significant negative estimate $\hat{\delta}_1$ provides evidence of mean-reverting residuals and leads to a rejection of H_0 in favour of the alternative hypothesis: that the residuals are stationary and series $y(t), x_1(t), x_2(t)$ cointegrate. The appropriate reference distribution depends on the number $N = 3$ and length $n = 135$ of potentially I(1) time series being regressed in eqn (5.26). Using the third-order approximation formula in MacKinnon (2010), the critical value for a cointegration test at the one-percent level is $\tau_c = -4.40$. It should be noted that the I(1) assumption refers to the degree of differencing required to achieve stationarity and is best regarded as an upper bound on the level of trendiness anywhere in the time series. The above cointegration test is therefore also valid for

I(0) data: the results would just be more conservative.

Tests of this form were performed for combinations of the HadCRUT5 historical observations with output from each of the 13 CMIP6 GCMs considered in this chapter. Figure 5.1 shows time series of residuals $\hat{\varepsilon}(t)$ from the fitted regression in eqn (5.26), with test statistics $\hat{\tau}_c$ and associated p-values. It can be seen from the residual plots that all 13 time series exhibit strong mean-reverting behaviour. This is confirmed by the results of the cointegration tests: the null hypothesis of "no cointegration" was rejected at the one-percent level for all 13 GCMs.

The residual time series in Figure 5.1 share common features, such as an apparent bump around the year 1940. This is as expected. From the theorem in Section 5.4 it follows that the regression residuals in eqn (5.26) include contributions from the stationary components of the forced trends in $y(t), x_1(t), x_2(t)$ (i.e. those terms involving $\Delta F(t)$), as well as from internal climate variability in those series. Since the GCMs have similar impulse responses, and since the realization of internal variability in HadCRUT5 is common to all 13 regressions, the only truly independent contribution to each residual series comes from that GCM's realizations of internal variability.

5.5.3 Attribution of surface temperature warming

Having detected significant cointegration of $y(t), x_1(t), x_2(t)$, it follows that the coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ in eqn (5.26) may be consistently estimated using OLS. However, the regression residuals (see Figure 5.1) are serially correlated, meaning that the usual formulas for calculating standard errors and confidence regions are invalid. Optimal fingerprinting studies commonly address this problem by estimating the covariance structure of internal climate variability from a GCM's piControl simulation (Allen and Tett, 1999). This approach ignores the stationary forced component of the residuals, which arises due to differences between the impulse responses of GCMs and the true climate. Using piControl also relies on GCMs accurately simulating the pre-industrial climate, which cannot be verified through observation. The use of

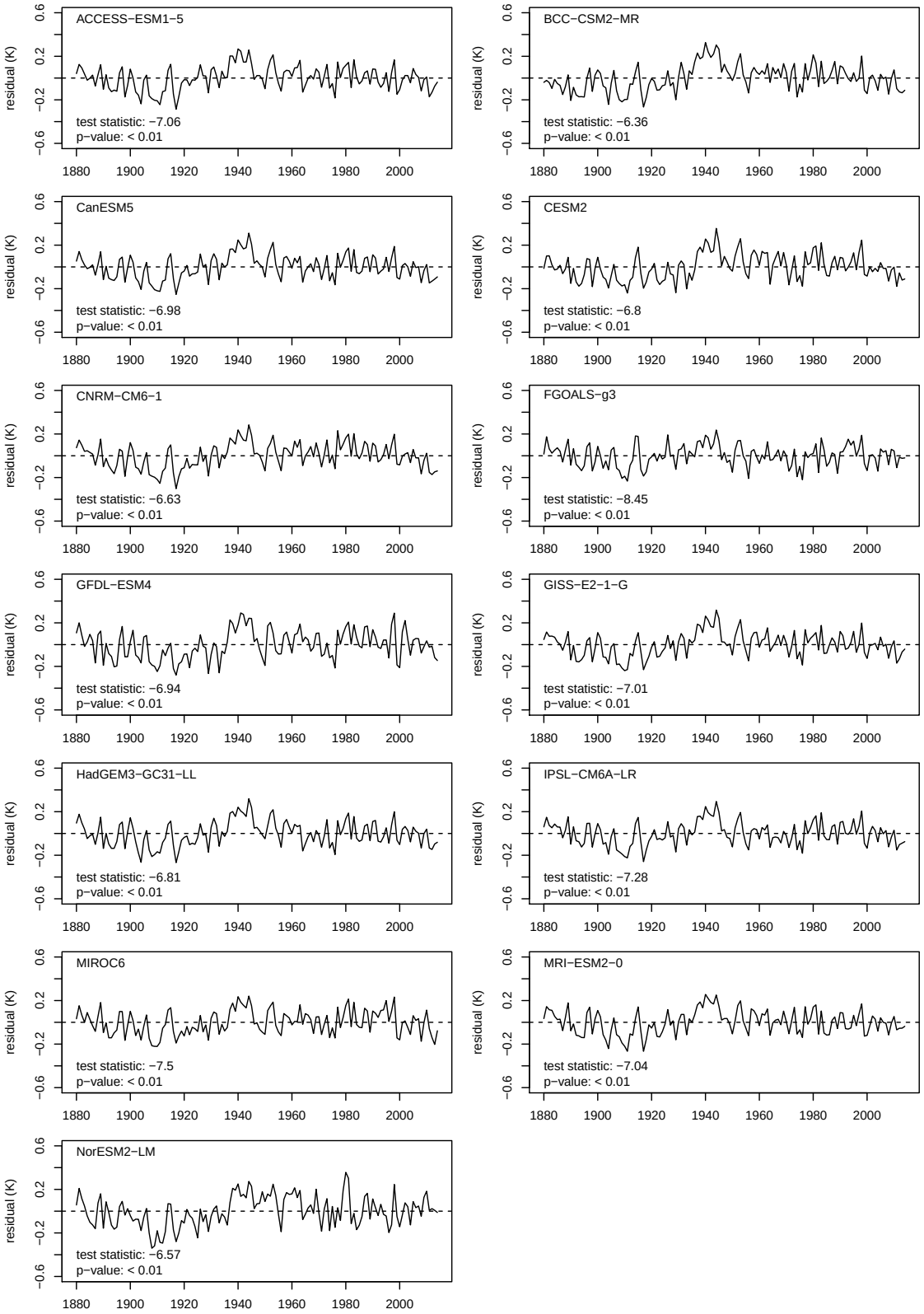

Figure 5.1: Cointegration test results. Time series are residuals from two-way OLS regressions of HadCRUT5 GMST observations on GCM output from *historical* and *hist-GHG* experiments. Test statistics < -4.40 are significant at the one-percent level, indicating residual stationarity and cointegration.

piControl for hypothesis testing in optimal fingerprinting has recently been criticized (McKitrick, 2021).

An alternative way to avoid the problem of serially correlated residuals, without introducing dependence on piControl simulations, is to fit a dynamic regression model which includes lagged versions of time series $y(t), x_1(t), x_2(t)$ (Hendry and Juselius, 2000). Fitting dynamic regressions of the form

$$y(t) = \beta_0' + \beta_1' x_1(t) + \beta_2' x_2(t) + \beta_3' y(t-1) + \beta_4' x_1(t-1) + \beta_5' x_2(t-1) + \varepsilon'(t) \quad (5.28)$$

using OLS yields serially uncorrelated residual series for each of the 13 GCMs. The usual normal distribution theory may then be assumed to hold asymptotically for estimates of the coefficients $\boldsymbol{\beta}' = (\beta_1', \ldots, \beta_5')'$ in eqn (5.28).

Dynamic regression coefficients $\boldsymbol{\beta}'$ in eqn (5.28) can be related back to coefficients $\boldsymbol{\beta}$ in eqn (5.26) via the Granger representation theorem, which requires that systems of cointegrated series have equivalent representations as error-correction models (ECMs) (Engle and Granger, 1987). Observe that eqn (5.28) may be written

$$\Delta y(t) = \beta_1' \Delta x_1(t) + \beta_2' \Delta x_2(t) - \alpha \left[y(t-1) - \beta_0 - \beta_1 x_1(t-1) - \beta_2 x_2(t-1) \right] + \varepsilon'(t), \quad (5.29)$$

where $\alpha = 1 - \beta_3'$, $\beta_0 = \beta_0'/\alpha$, $\beta_1 = (\beta_1' + \beta_4')/\alpha$, and $\beta_2 = (\beta_2' + \beta_5')/\alpha$. The expression inside the square brackets, called the error-correction term, is a stationary linear combination of $y(t), x_1(t), x_2(t)$. By estimating the dynamic regression model in eqn (5.28) using OLS, and then reparameterizing to obtain the ECM in eqn (5.29), it is possible to recover estimates of coefficients $\boldsymbol{\beta}$ in eqn (5.26).

Because the *historical* CMIP6 experiment includes GHG forcing, parameters β_1 and β_2 must be transformed to obtain the scaling factors of primary interest. Following the notation of Jones et al. (2016), let $\beta_{\mathrm{G}} = \beta_1 + \beta_2$ and $\beta_{\mathrm{OAN}} = \beta_1$ denote the scaling factors to be applied to GCM-predicted signals forced by GHG emissions and "other anthropogenic and natural" factors respectively. Although

$\boldsymbol{\beta}^* = (\beta_{\mathrm{G}}, \beta_{\mathrm{OAN}})'$ is linearly related to coefficients $\boldsymbol{\beta}$, the function relating $\boldsymbol{\beta}$ to $\boldsymbol{\beta}'$ is non-linear. The partially linear function $f : \boldsymbol{\beta}' \mapsto \boldsymbol{\beta}^*$ may be written

$$f(\boldsymbol{\beta}') = g(\boldsymbol{\beta}')h(\boldsymbol{\beta}'), \tag{5.30}$$

where $g(\boldsymbol{\beta}') = (1 - \beta_3')^{-1}$ is the non-linear part of f, and $h(\boldsymbol{\beta}') = M\boldsymbol{\beta}'$ is the linear part where

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{5.31}$$

Let

$$J(\boldsymbol{\beta}') = h(\boldsymbol{\beta}')\nabla g(\boldsymbol{\beta}')' + \frac{\partial h(\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}'} g(\boldsymbol{\beta}') \tag{5.32}$$

denote the Jacobian of f. If coefficient estimates $\hat{\boldsymbol{\beta}}'$, obtained by fitting eqn (5.28) using OLS, have estimated covariance $\hat{\Sigma}'$, then a linearized estimate of the covariance of $\hat{\boldsymbol{\beta}}^* = f(\hat{\boldsymbol{\beta}}')$ is given by $\hat{\Sigma}^* = J(\hat{\boldsymbol{\beta}}')\hat{\Sigma}' J(\hat{\boldsymbol{\beta}}')'$. An approximate 90 % confidence ellipse for $\boldsymbol{\beta}^*$ satisfies

$$(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*)' \left(\hat{\Sigma}^*\right)^{-1} (\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*) < 2F_{2,128}(0.90), \tag{5.33}$$

where $F_{2,128}(0.90)$ denotes the 90th quantile of the F distribution with degrees of freedom two and 128. Residual degrees of freedom are 135 (years of observations from 1880-2014), minus one (due to differencing), minus six (parameters estimated to fit eqn (5.28)), giving 128.

Point estimates and confidence ellipses have been calculated for scaling factors β_{G} and β_{OAN} using the methodology described above (see Figure 5.2 and Table 5.2). Collinearity between time series $y(t), x_1(t), x_2(t)$ and their lagged counterparts means that, taken individually, coefficient estimates $\hat{\boldsymbol{\beta}}'$ are subject to greater uncertainty than the classical OLS estimates $\hat{\boldsymbol{\beta}}$. However, the estimates $\hat{\beta}_{\mathrm{G}}$ and $\hat{\beta}_{\mathrm{OAN}}$ obtained by back-transforming $\hat{\boldsymbol{\beta}}'$ are well-constrained and closely resemble the classical estimates (proportional change has mean 0.008 and standard devia-

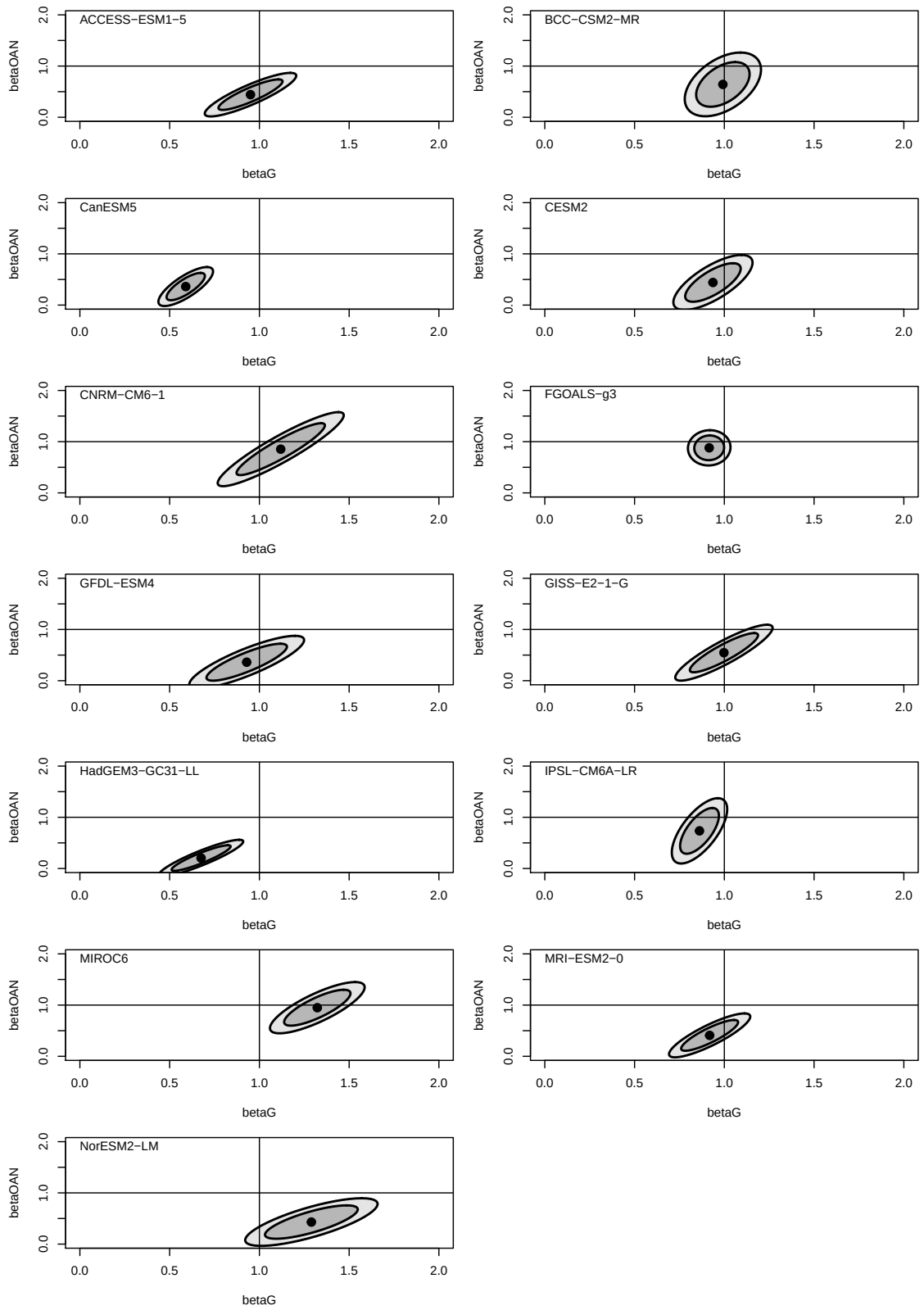

Figure 5.2: Scaling factor confidence ellipses. Black dots are point estimates of scaling factors $\hat{\beta}_{\mathrm{G}}, \hat{\beta}_{\mathrm{OAN}}$ for each GCM, obtained using dynamic OLS regression. Smaller and larger shaded ellipses are approximate 90 % and 99 % confidence regions respectively.

Model	$\hat{\beta}_{\mathrm{G}}$	(s.e.)	$\hat{\beta}_{\mathrm{OAN}}$	(s.e.)	$\Delta T_{\mathrm{total}}$ (K)	ΔT_{GHG} (K)
ACCESS-ESM1-5	0.95	(0.083)	0.44	(0.14)	1.02	1.22
BCC-CSM2-MR	0.99	(0.069)	0.64	(0.20)	0.99	1.04
CanESM5	0.59	(0.049)	0.36	(0.12)	1.01	1.18
CESM2	0.94	(0.071)	0.44	(0.17)	1.05	1.14
CNRM-CM6-1	1.12	(0.114)	0.85	(0.23)	1.08	1.47
FGOALS-g3	0.92	(0.038)	0.88	(0.11)	1.02	0.93
GFDL-ESM4	0.93	(0.104)	0.36	(0.17)	1.05	1.20
GISS-E2-1-G	1.00	(0.088)	0.55	(0.18)	1.04	1.28
HadGEM3-GC31-LL	0.68	(0.076)	0.20	(0.12)	1.05	1.22
IPSL-CM6A-LR	0.86	(0.050)	0.73	(0.21)	1.07	1.18
MIROC6	1.32	(0.085)	0.95	(0.16)	1.05	1.29
MRI-ESM2-0	0.92	(0.073)	0.41	(0.14)	1.04	1.25
NorESM2-LM	1.29	(0.119)	0.43	(0.15)	1.02	1.20

Table 5.2: Estimated scaling factors and attributable warming. Point estimates of $\hat{\beta}_{\mathrm{G}}$ and $\hat{\beta}_{\mathrm{OAN}}$ are obtained from OLS fits of the dynamic regression model in eqn (5.28) to HadCRUT5 GMST observations. Reported standard errors are calculated by linearizing $f : \boldsymbol{\beta}' \mapsto \boldsymbol{\beta}^*$ about $\boldsymbol{\beta}' = \hat{\boldsymbol{\beta}}'$. Columns $\Delta T_{\mathrm{total}}$ and ΔT_{GHG} are corresponding estimates of the total and GHG-attributable increases in GMST between the reference periods 1880-1899 and 2005-2014, obtained by appropriate scaling of GCM-predicted signals.

tion 0.07). The inflation of parameter uncertainty which results from including lagged variables is a necessary consequence of accounting for residual autocorrelation. Scaling factor standard errors vary across different GCMs. This is partly a consequence of variation in the size of available ensembles, with increased ensemble size leading to smaller standard errors. Standard errors are also affected by the GCMs' ability to reproduce the forced patterns in the observations. Table 5.2 also includes estimates of attributable warming between the reference periods 1880-1899 and 2005-2014. Attributable warming was calculated by first computing the mean difference in GMST between the reference periods for each of the *historical* and *hist-GHG* GCM experiments, and then taking appropriate linear combinations of the temperature differences with coefficients determined by the relevant scaling factors. These results are visualized in Figure 5.3, where total historical and GHG-attributable warming are plotted side-by-side. From Figure 5.3 it may be seen that for 12 out of 13 CMIP6 GCMs GHG-attributable warming exceeds total historical warming.

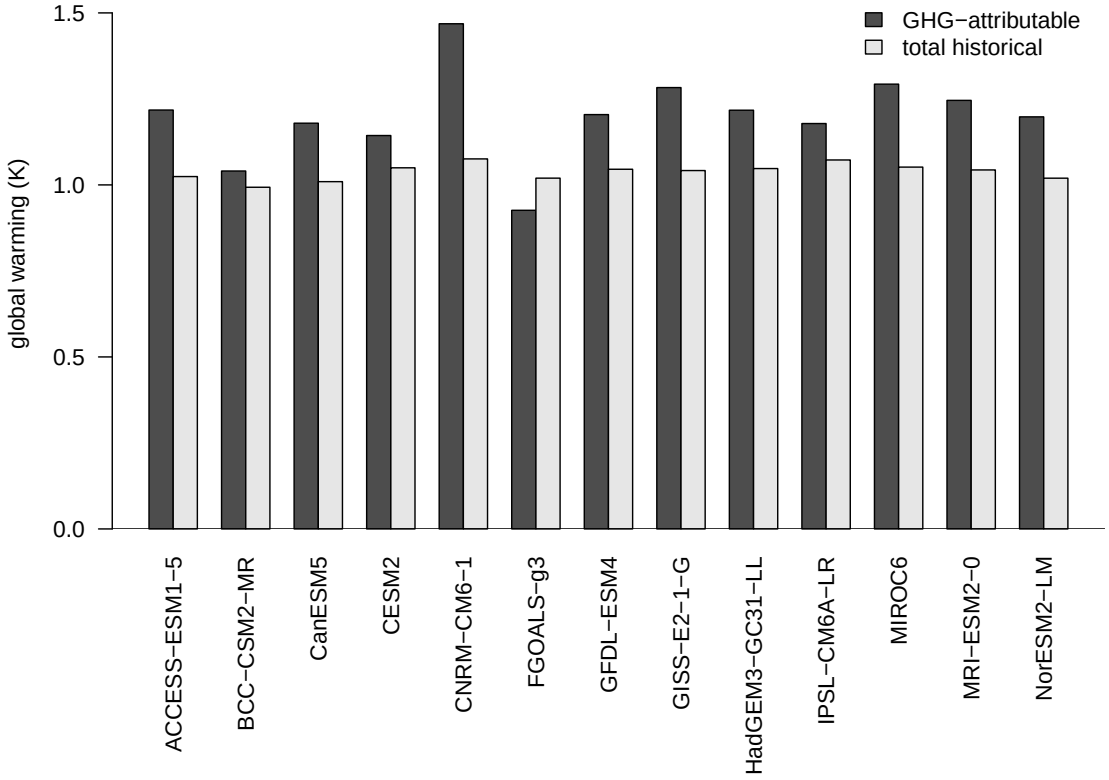

Figure 5.3: Estimated total and GHG-attributable increases in GMST between the reference periods 1880-1899 and 2005-2014, obtained by appropriate scaling of GCM-predicted signals.

5.6 Summary

Optimal fingerprinting, the statistical methodology commonly used for D&A of climate change trends, is typically performed by linearly regressing non-stationary climate variables. Non-stationary time series regressions are, in general, statistically inconsistent and liable to produce spurious results. This chapter has shown, by modelling radiative forcing as an integrated stochastic process within an idealized linear-response-model framework, that the optimal fingerprinting estimator is consistent under standard D&A assumptions. Hypothesis tests, combining observations of historical GMST with simulation output from 13 CMIP6-generation GCMs, produce no evidence that standard assumptions have been violated. It is therefore concluded that, at least in the case of GMST, detection and attribution of climate change trends is very likely not spurious regression. Furthermore, detection of significant cointegration between observations and GCM output indicates that the OLS estimator is superconsistent, with better convergence properties than might previously have been assumed. Finally, a new method has been developed for quantifying D&A uncertainty, which exploits the notion of cointegration to eliminate the need to rely on piControl GCM simulations and the corresponding strong assumptions.

Chapter 6

Conclusions and future directions

6.1 Thesis summary

The primary contribution of this thesis has been to establish a foundational basis for the validity of "optimal fingerprinting", a statistical methodology whose variants are commonly used in the detection and attribution (D&A) of climate change trends. This has been achieved through the use of idealized linear-impulse-response models as tractable analytical reasoning tools. The use of linear impulse responses was originally motivated by energy-balance considerations and has led to the formulation and fulfillment of multiple secondary research aims.

In Chapter 3, a procedure was developed for estimating parameters of a k-box stochastic energy-balance model (EBM), a simple but flexible representation of global climate in terms of surface temperature and top-of-atmosphere (TOA) net radiative flux. The method is based on the principle of maximum likelihood and estimates parameters of EBMs with any $k > 0$ number of boxes from CO2-quadrupling general circulation model (GCM) experiments. A software implementation of the method as an R package has been released as free software for public use[1]. It has been determined that a three-box model is optimal for emulating the global mean surface temperature (GMST) impulse responses of GCMs in the Coupled Model Intercomparison Project Phase 5 (CMIP5). The research presented in Chapter 3 led

[1] `https://github.com/donaldcummins/EBM`

naturally onto the research of Chapter 4.

In Chapter 4, a link was established between EBMs and the general class of linear time series models known as autoregressive moving-average (ARMA). It has been shown that the response of any k-box EBM can be represented as an ARMA(k, $k-1$) filter. Such "energy-balance" ARMA filters have direct applicability to the D&A problem. For example, it has been shown how, by inverting the ARMA filter, time series of surface temperature may be converted into radiative forcing, thus providing a means of estimating historical effective radiative forcing (ERF) from time series of GMST. This ARMA filter estimator has been compared with an established method ("ERF_trans"), using historical simulations from HadGEM3-GC31-LL, and it has been found that the ARMA method gives an ERF time series that closely matches published results (correlation of 0.83). Unlike ERF_trans, the ARMA method does not require historical observations of TOA net radiative flux. An application of the ARMA method to historical temperature observations, in combination with HadGEM3, produced evidence of a significant increase in ERF over the historical period with an estimated forcing in 2018 of 1.45 ± 0.504 W m^{-2}.

The primary research aims of the thesis were addressed in Chapter 5. It was decided that the energy-balance constraints used in Chapters 3 and 4 should be relaxed to allow modelling of climate model responses as general causal ARMA filters (a superset of EBMs). The results of Chapter 5 thus have considerable generality, while still permitting the EBM parameterization as a motivating and physically interpretable example. To investigate susceptibility of fingerprinting methods to spurious regression in the case of trending climate variables, radiative forcing was represented as an integrated stochastic process which accumulates over time in a manner analogous to atmospheric carbon dioxide concentration. It has been proved that if standard assumptions hold then the optimal fingerprinting estimator is consistent, and hence robust against spurious regression. Hypothesis tests, conducted using historical GMST observations and simulation output from 13 GCMs of the CMIP6 generation, were used to check for evidence of these assumptions being

violated in practice. No such evidence has been found. Therefore, it has been provisionally concluded that the historical trends in GMST which are detected and attributed using these GCMs are very likely not spurious.

It was noted in Chapter 5 that, under the assumption of integrated forcing, consistency of the fingerprinting estimator depends on "cointegration" between historical observations and GCM output. That such a cointegration was detected in Chapter 5 for the GMST variable indicates that the least-squares estimator is "superconsistent", with better convergence properties than might previously have been assumed. Finally, a new method has been developed for quantifying D&A uncertainty, which exploits the connection between cointegration and error-correction time series models to eliminate the need for pre-industrial control simulations. This new method works by using the Granger representation theorem (Engle and Granger, 1987) to represent an optimal fingerprinting study as an "error-correction" time series model, whose residuals are serially uncorrelated.

6.2 Future directions

The research presented in Chapters 3 to 5 of this thesis may be readily extended in future studies. Follow-up topics of possible interest include:

- **EBMs with non-integral number of boxes k:** The number of parameters in the k-box EBMs estimated in Chapter 3 increases by two for each ocean box added, which then feeds into the AIC score for choosing the number of boxes (see subsection 3.5.7). Of the GCMs in Table 3.4, some models (e.g. MIROC5) reported only a modest improvement in AIC score moving from two to three boxes. Rather than using a full three-box model, which may be overparameterized for such GCMs, an alternative compromise would be to use a 2.5-box model of the form used in Grieser and Schönwiese (2001). The "half box" is obtained by setting $C_1 = 0$ so that the top-layer box has zero heat capacity. The temperature T_1 of this half box is then a weighted

average of the forcing F and the temperature T_2. A 2.5-box model shares an important advantage of the three-box model, which is that it can emulate a GCM's impulse response on one-year and 15-year timescales (roughly) at the same time, while retaining the simpler two-dimensional temperature state vector of the two-box model. A disadvantage is that the zero-heat-capacity assumption is unrealistic, implying an instantaneous adjustment of surface temperature to an imposed forcing. A counterargument would then be that an EBM is not typically intended for use on such short timescales, although including the one-year timescale as an instant adjustment during estimation prevents biasing estimates of the intermediate 15-year timescale.

- **Rigorous uncertainty quantification for ARMA-filter estimates of historical ERF:** Possible approaches were discussed in subsection 4.6.3. The natural next step will be to apply the ARMA filter validation procedure in Section 4.5 to the full multi-model ensemble of GCMs in CMIP6. In this way, it will be seen whether the overall good performance of the ARMA filter, in the case of the two GCMs from the UK Met Office Hadley Centre, generalizes to GCMs from other modelling centres. It will also be informative to see whether observed deficiencies of the method, e.g. the underestimation of volcanic aerosol forcing, are intrinsic to the ARMA / EBM approach, or specific to the Hadley Centre GCMs.

- **Generalization of theoretical results to the full spatio-temporal D&A problem:** The theoretical results in Chapter 5 have been proved in a simplified setting where observed and simulated climate variables $y(t), x(t), F(t)$ are indexed by time only. In practice, optimal fingerprinting studies are commonly performed in the spatio-temporal domain (see Chapter 2), which brings its own set of problems (e.g. dimension reduction, ill-conditioning, curse of dimensionality etc.). However, the structure of the problem remains fundamentally unchanged. Variables $y(t), x(t), F(t)$, treated as time series in this thesis, could be replaced by vector equivalents $\boldsymbol{y}(t), \boldsymbol{x}(t), \boldsymbol{F}(t)$ describing relevant spatial

fields at each timestep. The ARMA impulse-response model in Chapter 5 could then be replaced by its vector equivalent: the vector-ARMA or VARMA model (Lütkepohl, 1991). There would no doubt exist possible physical parameterizations for such a model, e.g. as a spatially resolved EBM. The scientific value of generalizing Chapter 5 in this way would of course depend on whether GCMs' spatio-temporal dynamics (as opposed to just their globally averaged temporal dynamics) could be realistically captured by a linear-multivariate model. Low-dimensional linear approximation of fluid dynamical systems is an established practice in applied mathematics so this requirement may not be prohibitive (Wynn et al., 2013).

Appendix A

Stochastic energy-balance model results

A.1 Proof that eigenvalues of matrix A are real and non-positive when $\varepsilon = 1$

Consider the $k \times k$ matrices $A^\dagger$ and D where

$$A^\dagger_{i,j} = \begin{cases} -(\kappa_i + \kappa_{i+1})/C_i & \text{if } i = j, \\ \kappa_j/\sqrt{C_i C_j} & \text{if } j = i+1, \\ \kappa_i/\sqrt{C_j C_i} & \text{if } j = i-1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

and D is a diagonal matrix with leading diagonal $(1, \sqrt{C_2/C_1}, \sqrt{C_2 C_3/C_1 C_2}, \dots)'$. If $\varepsilon = 1$ the matrix A is related to $A^\dagger$ by the similarity transform $A = D^{-1} A^\dagger D$. Since matrix $A^\dagger$ is real and symmetric it must have all real eigenvalues. By similarity, all eigenvalues of A are therefore real. Applying the Geršgorin circle theorem (Gershgorin, 1931) to A it follows that the eigenvalues of A are also non-positive.

A.2 Discretization of the full k-box model

Equation (3.13) can be rearranged:

$$\dot{\boldsymbol{x}}(t) - A^{+}\boldsymbol{x}(t) = \boldsymbol{b}u(t) + \boldsymbol{w}(t). \tag{A.2}$$

Multiplying by the integrating factor $\mathrm{e}^{-A^{+}t}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{-A^{+}t}\boldsymbol{x}(t)\right) = \mathrm{e}^{-A^{+}t}\boldsymbol{b}u(t) + \mathrm{e}^{-A^{+}t}\boldsymbol{w}(t). \tag{A.3}$$

Integrating with respect to time,

$$\mathrm{e}^{-A^{+}t}\boldsymbol{x}(t) = \int_{s=-\infty}^{t} \mathrm{e}^{-A^{+}s}\boldsymbol{b}u(s) + \mathrm{e}^{-A^{+}s}\boldsymbol{w}(s)\,\mathrm{d}s, \tag{A.4}$$

so that, multiplying by $\mathrm{e}^{A^{+}t}$, we obtain

$$\boldsymbol{x}(t) = \int_{s=-\infty}^{t} \mathrm{e}^{A^{+}(t-s)}\boldsymbol{b}u(s) + \mathrm{e}^{A^{+}(t-s)}\boldsymbol{w}(s)\,\mathrm{d}s. \tag{A.5}$$

As a linear function of Gaussian random variables $\boldsymbol{x}(t)$ is itself Gaussian and hence fully characterized by its mean and covariance. Since $\mathrm{E}(\boldsymbol{w}(t)) = \boldsymbol{0}$ for all t,

$$\mathrm{E}\left(\boldsymbol{x}(t)|\boldsymbol{x}(t-1)\right) = \mathrm{e}^{A^{+}}\boldsymbol{x}(t-1) + \int_{s=t-1}^{t} \mathrm{e}^{A^{+}(t-s)}\boldsymbol{b}u(s)\,\mathrm{d}s, \tag{A.6}$$

where, assuming $u(s) = u(t-1)$ for $s \in [t-1, t)$,

$$\begin{aligned}
\int_{s=t-1}^{t} \mathrm{e}^{A^{+}(t-s)}\boldsymbol{b}u(s)\,\mathrm{d}s &= \int_{s=t-1}^{t} \mathrm{e}^{A^{+}(t-s)}\boldsymbol{b}u(t-1)\,\mathrm{d}s && \text{(A.7)}\\
&= \left[-(A^{+})^{-1}\mathrm{e}^{A^{+}(t-s)}\boldsymbol{b}u(t-1)\right]_{s=t-1}^{t} && \text{(A.8)}\\
&= -(A^{+})^{-1}\left(I - \mathrm{e}^{A^{+}}\right)\boldsymbol{b}u(t-1) && \text{(A.9)}\\
&= (A^{+})^{-1}\left(\mathrm{e}^{A^{+}} - I\right)\boldsymbol{b}u(t-1). && \text{(A.10)}
\end{aligned}$$

For the covariance we have

$$\operatorname{cov}\left(\boldsymbol{x}(t)|\boldsymbol{x}(t-1)\right) = \operatorname{cov}\left(\int_{s=t-1}^{t} \mathrm{e}^{A^{+}(t-s)}\boldsymbol{w}(s)\,\mathrm{d}s\right) \tag{A.11}$$

$$= \operatorname{cov}\left(\int_{s=0}^{1} \mathrm{e}^{A^{+}(1-s)}\boldsymbol{w}(s)\,\mathrm{d}s\right) \tag{A.12}$$

$$= \operatorname{cov}\left(\int_{s=0}^{1} \mathrm{e}^{A^{+}s}\boldsymbol{w}(s)\,\mathrm{d}s\right), \tag{A.13}$$

where, since $\boldsymbol{w}(t)$ is white noise and hence uncorrelated in time,

$$\operatorname{cov}\left(\int_{s=0}^{1} \mathrm{e}^{A^{+}s}\boldsymbol{w}(s)\,\mathrm{d}s\right) = \int_{s=0}^{1} \operatorname{cov}\left(\mathrm{e}^{A^{+}s}\boldsymbol{w}(s)\right)\,\mathrm{d}s \tag{A.14}$$

$$= \int_{s=0}^{1} \mathrm{e}^{A^{+}s}\operatorname{cov}\left(\boldsymbol{w}(s)\right)\mathrm{e}^{A^{+'}s}\,\mathrm{d}s \tag{A.15}$$

$$= \int_{s=0}^{1} \mathrm{e}^{A^{+}s}Q\mathrm{e}^{A^{+'}s}\,\mathrm{d}s. \tag{A.16}$$

For additional information on this type of discretization scheme see Section 4.3 of Ljung (1987).

A.3 Marginal covariance of the stochastic response

The k-box model is a linear dynamic system. Therefore the response to a linear combination of inputs is equal to the sum of the responses to individual inputs. In this way we can separate the model responses to deterministic and stochastic forcing components. The stochastic component of the response is driven by a purely stochastic input and may be written

$$\boldsymbol{x}(t) = A_d\boldsymbol{x}(t-1) + \boldsymbol{w}_d(t) \tag{A.17}$$

which is a vector autoregressive process of order one (VAR(1)). The matrix-valued auto-cross covariance function $\Gamma(h)$ is defined

$$\Gamma(h) = \Gamma(-h) = \mathrm{E}(\boldsymbol{x}(t)\boldsymbol{x}(t+h)') \tag{A.18}$$

where the lag h is an integer. We seek the marginal auto-cross covariance matrix $\Gamma(0)$ which is the *a priori* covariance of $\boldsymbol{x}_0$ in the Kalman filter. Define the backshift operator B such that

$$\mathrm{B}^i \boldsymbol{x}(t) = \boldsymbol{x}(t-i). \tag{A.19}$$

We can write

$$(I - A_d \mathrm{B})\boldsymbol{x}(t) = \boldsymbol{w}_d(t) \tag{A.20}$$

and

$$\boldsymbol{x}(t) = (I - A_d \mathrm{B})^{-1} \boldsymbol{w}_d(t) \tag{A.21}$$

$$= (I + A_d \mathrm{B} + A_d^2 \mathrm{B}^2 + \dots)\boldsymbol{w}_d(t). \tag{A.22}$$

The geometric series converges when the VAR(1) process is stationary, i.e. all eigenvalues of A_d lie within the unit circle in the complex plane.

$$\Gamma(0) = \mathrm{E}(\boldsymbol{x}(t)\boldsymbol{x}(t)') \tag{A.23}$$

$$= \mathrm{E}\left[(I + A_d \mathrm{B} + \dots)\boldsymbol{w}_d(t)\boldsymbol{w}_d(t)'(I + A_d' \mathrm{B} + \dots)\right]. \tag{A.24}$$

Since

$$\mathrm{E}(\mathrm{B}^i \boldsymbol{w}_d(t)\boldsymbol{w}_d(t)'\mathrm{B}^j) = Q_d \delta_{ij}, \tag{A.25}$$

where δ_{ij} denotes the Kronecker delta, we have

$$\Gamma(0) = Q_d + A_d Q_d A_d' + A_d^2 Q_d A_d^{2'} + \dots \tag{A.26}$$

The infinite series can be computed as follows using the vec operator and the Kronecker product (Lütkepohl, 1991).

$$\begin{aligned}\mathrm{vec}(\Gamma(0)) &= \mathrm{vec}(Q_d) + \mathrm{vec}(A_d Q_d A_d') + \mathrm{vec}(A_d^2 Q_d A_d^{2\prime}) + \dots && \text{(A.27)}\\ &= \mathrm{vec}(Q_d) + (A_d \otimes A_d)\mathrm{vec}(Q_d) + (A_d^2 \otimes A_d^2)\mathrm{vec}(Q_d) + \dots && \text{(A.28)}\\ &= (I - A_d \otimes A_d)^{-1}\mathrm{vec}(Q_d). && \text{(A.29)}\end{aligned}$$

Note $(I - A_d \otimes A_d)$ is invertible because eigenvalues of $A_d \otimes A_d$ are products of eigenvalues of A_d and hence have modulus < 1 when the VAR(1) process is stationary.

A.4 Analytical responses under idealized forcing scenarios

A.4.1 Unit step forcing

The k-box model response under a unit step-forcing scenario

$$F_{\mathrm{step}}(t) = \begin{cases} 1 & \text{if } t \geq 0, \\ 0 & \text{otherwise;} \end{cases} \tag{A.30}$$

can be written

$$\boldsymbol{x}_{\mathrm{step}}(t) = \frac{1}{\kappa_1}\left(\mathbf{1} - \mathrm{e}^{At}\mathbf{1}\right), \tag{A.31}$$

where $\mathbf{1}$ denotes the k-vector of ones $(1, \dots, 1)'$. The unit-forced equilibrium temperature is $1/\kappa_1$, which is obtained by setting equation (3.8) equal to zero and solving for $T_1 = \cdots = T_k$. As the k-box model is linear, transient relaxation to the new equilibrium temperature is exponential.

A.4.2 Unit impulse forcing

Differentiating $\boldsymbol{x}_{\text{step}}(t)$ with respect to time we obtain the response to a unit-impulse forcing

$$F_{\text{imp}}(t) = \delta(t), \tag{A.32}$$

where $\delta(t)$ denotes the Dirac delta function:

$$\boldsymbol{x}_{\text{imp}}(t) = -\frac{1}{\kappa_1} A \mathrm{e}^{At} \mathbf{1}. \tag{A.33}$$

This follows from the fact that an impulse is the time derivative of a step forcing.

A.4.3 Transient climate response

Integrating $\boldsymbol{x}_{\text{step}}(t)$ with respect to time and scaling appropriately we obtain the transient climate response (TCR)

$$\boldsymbol{x}_{\text{TCR}}(t) = \frac{\log 1.01}{\log 4} \int_{s=0}^{t} \frac{F_{4\times\text{CO}_2}}{\kappa_1} \left(\mathbf{1} - \mathrm{e}^{As}\mathbf{1}\right) \mathrm{d}s \tag{A.34}$$

$$= \frac{\log 1.01}{\log 4} \frac{F_{4\times\text{CO}_2}}{\kappa_1} \left[t\mathbf{1} - A^{-1}\left(\mathrm{e}^{At} - I\right)\mathbf{1}\right], \tag{A.35}$$

which is the response to atmospheric CO_2 concentration increasing at a rate of 1% per year starting at time $t = 0$. This follows from the fact that an exponentially increasing CO_2 input is equivalent to a sequence of superimposed $1.01 \times \text{CO}_2$ step-forcing inputs. By linearity, the k-box model response to this superposition of forcing inputs is a superposition of the corresponding temperature outputs.

Appendix B

ARMA filter derivation

The step response of the k-box EBM is a linear combination of saturating exponentials (Geoffroy et al., 2013a; Tsutsui, 2020; Cummins et al., 2020b). The discrete-time impulse response $R(v)$ is a linear function of the step response and is a sum of decaying exponentials:

$$R(v) = \sum_{i=1}^{k} a_i \mathrm{e}^{-v/\tau_i} \tag{B.1}$$

$$= \sum_{i=1}^{k} a_i (r_i)^v, \tag{B.2}$$

where $a_i, \tau_i > 0$ and $r_i = \mathrm{e}^{-1/\tau_i}$. Then

$$\mathbf{\Phi}(\mathrm{B}) = \sum_{v=0}^{\infty} R(v)\mathrm{B}^v \tag{B.3}$$

$$= \sum_{v=0}^{\infty} \sum_{i=1}^{k} a_i (r_i)^v \mathrm{B}^v \tag{B.4}$$

$$= \sum_{i=1}^{k} \sum_{v=0}^{\infty} a_i (r_i)^v \mathrm{B}^v \tag{B.5}$$

$$= \sum_{i=1}^{k} \frac{a_i}{1 - r_i \mathrm{B}} \tag{B.6}$$

$$= \frac{\sum_{i=1}^{k} \left(a_i \prod_{j \neq i} (1 - r_j \mathrm{B}) \right)}{\prod_{i=1}^{k} (1 - r_i \mathrm{B})}. \tag{B.7}$$

Provided the r_i are distinct, it follows that the operator $\boldsymbol{\Phi}(\mathrm{B})$ is a ratio of polynomials $\boldsymbol{\theta}(\mathrm{B})$ and $\boldsymbol{\phi}(\mathrm{B})$ of degree $k-1$ and k respectively. Cases $k = 1, 2, 3$ are shown below.

Case $k = 1$:

$$\boldsymbol{\Phi}(\mathrm{B}) = \frac{a}{1 - r\mathrm{B}} \tag{B.8}$$

so $\boldsymbol{\phi}(\mathrm{B})$ and $\boldsymbol{\theta}(\mathrm{B})$ have respective degrees one and zero, i.e. the system is an ARMA(1, 0) or simply AR(1) filter.

Case $k = 2$:

$$\boldsymbol{\Phi}(\mathrm{B}) = \frac{a_1}{1 - r_1\mathrm{B}} + \frac{a_2}{1 - r_2\mathrm{B}} \tag{B.9}$$

$$= \frac{a_1(1 - r_2\mathrm{B}) + a_2(1 - r_1\mathrm{B})}{(1 - r_1\mathrm{B})(1 - r_2\mathrm{B})} \tag{B.10}$$

so $\boldsymbol{\phi}(\mathrm{B})$ and $\boldsymbol{\theta}(\mathrm{B})$ have respective degrees two and one, i.e. the system is an ARMA(2, 1) filter.

Case $k = 3$:

$$\boldsymbol{\Phi}(\mathrm{B}) = \frac{a_1}{1 - r_1\mathrm{B}} + \frac{a_2}{1 - r_2\mathrm{B}} + \frac{a_3}{1 - r_3\mathrm{B}} \tag{B.11}$$

$$= \frac{a_1(1 - r_2\mathrm{B})(1 - r_3\mathrm{B}) + a_2(1 - r_1\mathrm{B})(1 - r_3\mathrm{B}) + a_3(1 - r_1\mathrm{B})(1 - r_2\mathrm{B})}{(1 - r_1\mathrm{B})(1 - r_2\mathrm{B})(1 - r_3\mathrm{B})} \tag{B.12}$$

so $\boldsymbol{\phi}(\mathrm{B})$ and $\boldsymbol{\theta}(\mathrm{B})$ have respective degrees three and two, i.e. the system is an ARMA(3, 2) filter.

Note that, in the statistics literature, the term "ARMA process" generally refers to a an ARMA filter driven by a Gaussian white noise input. For more information about ARMA models and the backshift operator, see Brockwell and Davis (2002).

Appendix C

Code and data availability

C.1 Estimation of energy-balance models

I have developed a package for the R software environment (R Core Team, 2021) for simulation, fitting, filtering and predicting with k-box EBMs. The package estimates parameters of k-box models from time series of GMST and TOA net downward radiative flux by numerically maximizing the likelihood function. The R package, which includes the datasets used in Chapter 3, is available for download at `https://github.com/donaldcummins/EBM`.

C.2 Surface temperature and radiative forcing data

The HadGEM2-ES and HadGEM3-GC3.1-LL surface temperature and TOA radiative flux datasets used in Chapter 4 are available for download from Earth System Grid Federation (ESGF) portals, e.g. `https://esgf-data.dkrz.de/`. The time series of historical ERF for HadGEM3-GC3.1-LL estimated by Andrews et al. (2019) is available at `https://github.com/timothyandrews/HadGEM3-ERF-Timeseries`. The Cowtan and Way 2.0 historical surface temperature series is available at `https://www-users.york.ac.uk/~kdc3/papers/coverage2013/had4_krig_annual_v2_0_0.txt`.

C.3 Cointegration tests and attribution of GHG warming

The datasets and code used to perform the analyses in Chapter 5 are available online at `https://doi.org/10.5281/zenodo.5008827`. Code is written in the statistical programming language R (R Core Team, 2021). The ellipses in Figure 5.2 were plotted using the "ellipse" function from the *car* package by Fox et al. (2011).

Appendix D

Glossary of acronyms

ACF	AutoCorrelation Function
AIC	Akaike Information Criterion
AOGCM	Atmosphere-Ocean General Circulation Model
AR	AutoRegressive
ARMA	AutoRegressive Moving-Average
BOBYQA	Bound Optimization BY Quadratic Approximation
CMIP5	Coupled Model Intercomparison Project Phase 5
CMIP6	Coupled Model Intercomparison Project Phase 6
CW2.0	Cowtan and Way 2.0
D&A	Detection and Attribution
EBM	Energy-Balance Model
ECM	Error-Correction Model
ECS	Equilibrium Climate Sensitivity
EIV	Errors In Variables

EMIC	Earth-Sytem Model of Intermediate Complexity
EOF	Empirical Orthogonal Function
ERF	Effective Radiative Forcing
ESM	Earth System Model
FIR	Finite Impulse Response
GAM	Generalized Additive Model
GCM	General Circulation Model
GMST	Global Mean Surface Temperature
IIR	Infinite Impulse Response
IPCC	Intergovernmental Panel on Climate Change
LTI	Linear-Time-Invariant
MA	Moving-Average
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimate
MMSE	Minimum Mean-Square-Error
MSE	Mean Square Error
ODE	Ordinary Differential Equation
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PDE	Partial Differential Equation
PDF	Probability Density Function

ROF	Regularized Optimal Fingerprinting
SAT	Surface Air Temperature
SST	Sea Surface Temperature
TCR	Transient Climate Response
TLS	Total Least Squares
TOA	Top-Of-Atmosphere
TOANDRF	Top-Of-Atmosphere Net Downward Radiative Flux
VAR	Vector AutoRegressive
VARMA	Vector AutoRegressive Moving-Average

Bibliography

Adler, A. (2020). Pade: Padé Approximant Coefficients. https://CRAN.R-project.org/package=Pade.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Aldrin, M., Holden, M., Guttorp, P., Skeie, R. B., Myhre, G., and Berntsen, T. K. (2012). Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, 23(3):253–271. https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2140.

Allen, M. R. and Stott, P. A. (2003). Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dynamics*, 21(5):477–491. https://doi.org/10.1007/s00382-003-0313-9.

Allen, M. R. and Tett, S. F. B. (1999). Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, 15(6):419–434. https://doi.org/10.1007/s003820050291.

Alvarez, M., Luengo, D., and Lawrence, N. D. (2009). Latent Force Models. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 9–16, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA. PMLR. http://proceedings.mlr.press/v5/alvarez09a.html.

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., and Tang, Y. (2019). Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, 11(12):4377–4394. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001866.

Andrews, T. and Forster, P. M. (2020). Energy budget constraints on historical radiative forcing. *Nature Climate Change*, 10(4):313–316. https://www.nature.com/articles/s41558-020-0696-1.

Annan, J. D., Hargreaves, J. C., Edwards, N. R., and Marsh, R. (2005). Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, 8(1):135–154. https://www.sciencedirect.com/science/article/pii/S1463500303000726.

Beenstock, M., Reingewertz, Y., and Paldor, N. (2012). Polynomial cointegration tests of anthropogenic impact on global warming. *Earth System Dynamics*, 3(2):173–188. https://esd.copernicus.org/articles/3/173/2012/.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013). Chapter 10 - Detection and Attribution of Climate Change: From Global to Regional. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter10_FINAL.pdf.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F.,

Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12(7):e2019MS002010. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002010.

Brockwell, P. J. and Davis, R. A., editors (2002). *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer New York, New York, NY. http://link.springer.com/10.1007/b97391.

Bruns, S. B., Csereklyei, Z., and Stern, D. I. (2020). A multicointegration model of global climate change. *Journal of Econometrics*, 214(1):175–197. https://www.sciencedirect.com/science/article/pii/S0304407619301137.

Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the Earth. *Tellus*, 21(5):611–619. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2153-3490.1969.tb00466.x.

Caldeira, K. and Myhrvold, N. P. (2013). Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. *Environmental Research Letters*, 8(3):034039. https://doi.org/10.1088/1748-9326/8/3/034039.

Chang, M. K., Kwiatkowski, J. W., Nau, R. F., Oliver, R. M., and Pister, K. S. (1982). Arma models for earthquake ground motions. *Earthquake Engineering &*

Structural Dynamics, 10(5):651–662. https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.4290100503.

Chung, E.-S. and Soden, B. J. (2015). An assessment of methods for computing radiative forcing in climate models. *Environmental Research Letters*, 10(7):074004. https://doi.org/10.1088/1748-9326/10/7/074004.

Cohen, J. B. and Wang, C. (2014). Estimating global black carbon emissions using a top-down Kalman Filter approach. *Journal of Geophysical Research: Atmospheres*, 119(1):307–323. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JD019912.

Cowtan, K. and Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1935–1944. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2297.

Cox, P. M., Huntingford, C., and Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, 553(7688):319–322. https://www.nature.com/articles/nature25450.

Crowley, T. J. (2000). Causes of Climate Change Over the Past 1000 Years. *Science*, 289(5477):270–277. https://science.sciencemag.org/content/289/5477/270.

Cummins, D. P., Stephenson, D. B., and Stott, P. A. (2020a). A new energy-balance approach to linear filtering for estimating effective radiative forcing from temperature time series. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(2):91–102. https://ascmo.copernicus.org/articles/6/91/2020/.

Cummins, D. P., Stephenson, D. B., and Stott, P. A. (2020b). Optimal Estimation of Stochastic Energy Balance Model Parameters. *Journal of Climate*,

33(18):7909–7926. https://journals.ametsoc.org/view/journals/clim/33/18/jcliD190589.xml.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001916. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001916.

De Groen, P. and De Moor, B. (1987). The fit of a sum of exponentials to noisy data. *Journal of Computational and Applied Mathematics*, 20:175–187. https://www.sciencedirect.com/science/article/pii/037704278790135X.

de Jong, P. and Penzer, J. (2004). The ARMA model in state space form. *Statistics & Probability Letters*, 70(1):119–125. https://www.sciencedirect.com/science/article/pii/S0167715204002330.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a):427–431. https://doi.org/10.1080/01621459.1979.10482531.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly,

P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., and Zhao, M. (2020). The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics. *Journal of Advances in Modeling Earth Systems*, 12(11):e2019MS002015. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002015.

Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276. https://www.jstor.org/stable/1913236.

Estrada, F. and Perron, P. (2017). Extracting and Analyzing the Warming Trend in Global and Hemispheric Temperatures. *Journal of Time Series Analysis*, 38(5):711–732. https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12246.

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. https://gmd.copernicus.org/articles/9/1937/2016/.

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y. (2021). Chapter 3 - Human Influence on the Climate System. In Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Pean, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekci, O., Yu, R., and Zhou, B., editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press,

Cambridge, UK. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_03.pdf.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). Chapter 9 - Evaluation of Climate Models. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter09_FINAL.pdf.

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., Andrews, T., Pincus, R., and Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, 121(20):12,460–12,475. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025320.

Fox, J., Weisberg, S., and Fox, J. (2011). *An R Companion to Applied Regression*. SAGE Publications, Los Angeles, Calif. ; London, 2nd ed edition.

Fredriksen, H.-B. and Rypdal, M. (2017). Long-Range Persistence in Global Surface Temperatures Explained by Linear Multibox Energy Balance Models. *Journal of Climate*, 30(18):7157–7168. https://journals.ametsoc.org/view/journals/clim/30/18/jcli-d-16-0877.1.xml.

Gay-Garcia, C., Estrada, F., and Sánchez, A. (2009). Global and hemispheric temperatures revisited. *Climatic Change*, 94(3):333–349. https://doi.org/10.1007/s10584-008-9524-8.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., and Tytéca, S. (2013a). Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation

for CMIP5 AOGCMs. *Journal of Climate*, 26(6):1859–1876. https://journals.ametsoc.org/view/journals/clim/26/6/jcli-d-12-00196.1.xml.

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., and Tytéca, S. (2013b). Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments. *Journal of Climate*, 26(6):1841–1857. https://journals.ametsoc.org/view/journals/clim/26/6/jcli-d-12-00195.1.xml.

Gershgorin, S. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de l'Académie des Sciences de l'URSS. VII. Série*, 1931(6):749–754.

Gilbert, P. and Varadhan, R. (2019). numDeriv: Accurate Numerical Derivatives. https://CRAN.R-project.org/package=numDeriv.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C. (2016). The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geoscientific Model Development*, 9(10):3685–3697. https://gmd.copernicus.org/articles/9/3685/2016/.

GISTEMP Team (2021). GISS Surface Temperature Analysis (GISTEMP), version 4. https://data.giss.nasa.gov/gistemp/.

Good, P., Gregory, J. M., and Lowe, J. A. (2011). A step-response simple climate model to reconstruct and interpret AOGCM projections. *Geophysical Research Letters*, 38(1). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL045208.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., and Stadelmann, M. (2021). Expm: Matrix Exponential, Log, 'etc'. https://CRAN.R-project.org/package=expm.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120. https://www.sciencedirect.com/science/article/pii/0304407674900347.

Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, 16(7):501–515. https://doi.org/10.1007/s003820000059.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31(3). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL018747.

Grieser, J. and Schönwiese, C.-D. (2001). Process, Forcing, and Signal Analysis of Global Mean TemperatureVariations by Means of a Three-Box Energy Balance Model. *Climatic Change*, 48(4):617–646. https://doi.org/10.1023/A:1005629309829.

Hannart, A. (2016). Integrated Optimal Fingerprinting: Method Description and Illustration. *Journal of Climate*, 29(6):1977–1998. https://journals.ametsoc.org/view/journals/clim/29/6/jcli-d-14-00124.1.xml.

Hannart, A. (2019). An improved projection of climate observations for detection and attribution. *Advances in Statistical Climatology, Meteorology and Oceanography*, 5(2):161–171. https://ascmo.copernicus.org/articles/5/161/2019/.

Hannart, A., Ribes, A., and Naveau, P. (2014). Optimal fingerprinting under multiple sources of uncertainty. *Geophysical Research Letters*, 41(4):1261–1268. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058653.

Hansen, J., Sato, M., Kharecha, P., and von Schuckmann, K. (2011). Earth's energy imbalance and implications. *Atmospheric Chemistry and*

Physics, 11(24):13421–13449. https://acp.copernicus.org/articles/11/13421/2011/acp-11-13421-2011.html.

Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, 28(6):473–485. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2153-3490.1976.tb00696.x.

Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, 13(9):601–611. https://doi.org/10.1007/s003820050185.

Hasselmann, K. F. (1979). On the signal-to-noise problem in atmospheric response studies. In *Joint Conference of Royal Meteorological Society, American Meteorological Society, Deutsche Meteorologische Gesellschaft and the Royal Society*, pages 251–259. Royal Meteorological Society. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3030122.

Haustein, K., Allen, M. R., Forster, P. M., Otto, F. E. L., Mitchell, D. M., Matthews, H. D., and Frame, D. J. (2017). A real-time Global Warming Index. *Scientific Reports*, 7(1):15417. https://www.nature.com/articles/s41598-017-14828-5.

Hegerl, G. and Zwiers, F. (2011). Use of models in detection and attribution of climate change. *WIREs Climate Change*, 2(4):570–591. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.121.

Hegerl, G. C., Hasselmann, K., Cubasch, U., Mitchell, J. F. B., Roeckner, E., Voss, R., and Waszkewitz, J. (1997). Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dynamics*, 13(9):613–634. https://doi.org/10.1007/s003820050186.

Hegerl, G. C., von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U., and Jones, P. D. (1996). Detecting Greenhouse-Gas-Induced Climate Change with an Optimal Fingerprint Method. *Journal of Climate*,

9(10):2281–2306. https://journals.ametsoc.org/view/journals/clim/9/10/1520-0442_1996_009_2281_dggicc_2_0_co_2.xml.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N., Luo, Y. M., Marengo Orsini, J., Nicholls, N., Penner, J., and Stott, P. (2007). Chapter 9 - Understanding and Attributing Climate Change. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK. https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter9-1.pdf.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G. K. (2010). Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing. *Journal of Climate*, 23(9):2418–2427. https://journals.ametsoc.org/view/journals/clim/23/9/2009jcli3466.1.xml.

Hendry, D. F. and Juselius, K. (2000). Explaining Cointegration Analysis: Part 1. *The Energy Journal*, 21(1):1–42. https://www.jstor.org/stable/41322853.

Huai-Min Zhang, Jay H. Lawrimore, Boyin Huang, Matthew J. Menne, Xungang Yin, Ahira Sánchez-Lugo, Byron E. Gleason, Russell Vose, Derek Arndt, J. Jared Rennie, and Claude N. Williams (2019). Updated Temperature Data Give a Sharper View of Climate Trends. https://doi.org/10.1029/2019EO128229.

Huntingford, C., Stott, P. A., Allen, M. R., and Lambert, F. H. (2006). Incorporating model uncertainty into attribution of observed temperature change. *Geophysical Research Letters*, 33(5). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL024831.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge,

UK. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf.

Johansson, D. J. A., O'Neill, B. C., Tebaldi, C., and Häggström, O. (2015). Equilibrium climate sensitivity in light of observations over the warming hiatus. *Nature Climate Change*, 5(5):449–453. https://www.nature.com/articles/nclimate2573.

Johnson, S. G. (2020). The NLopt nonlinear-optimization package. http://github.com/stevengj/nlopt.

Jones, G. S. and Kennedy, J. J. (2017). Sensitivity of Attribution of Anthropogenic Near-Surface Warming to Observational Uncertainty. *Journal of Climate*, 30(12):4677–4691. https://journals.ametsoc.org/view/journals/clim/30/12/jcli-d-16-0628.1.xml.

Jones, G. S., Stott, P. A., and Mitchell, J. F. B. (2016). Uncertainties in the attribution of greenhouse gas warming and implications for climate prediction. *Journal of Geophysical Research: Atmospheres*, 121(12):6969–6992. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD024337.

Jonko, A., Urban, N. M., and Nadiga, B. (2018). Towards Bayesian hierarchical inference of equilibrium climate sensitivity from a combination of CMIP5 climate models and observational data. *Climatic Change*, 149(2):247–260. https://doi.org/10.1007/s10584-018-2232-0.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45. https://doi.org/10.1115/1.3662552.

Katzfuss, M., Hammerling, D., and Smith, R. L. (2017). A Bayesian hierarchical model for climate change detection and attribution. *Geophysical Research Letters*, 44(11):5720–5728. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073688.

Kaufmann, B. (2003). Fitting a Sum of Exponentials to Numerical Data. *arXiv:physics/0305019*. http://arxiv.org/abs/physics/0305019.

Kaufmann, R. K., Kauppi, H., Mann, M. L., and Stock, J. H. (2011). Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proceedings of the National Academy of Sciences*, 108(29):11790–11793. https://www.pnas.org/content/108/29/11790.

Kaufmann, R. K., Kauppi, H., Mann, M. L., and Stock, J. H. (2013). Does temperature contain a stochastic trend: Linking statistical results to physical mechanisms. *Climatic Change*, 118(3):729–743. https://doi.org/10.1007/s10584-012-0683-2.

Kaufmann, R. K., Kauppi, H., and Stock, J. H. (2006). Emissions, Concentrations, & Temperature: A Time Series Analysis. *Climatic Change*, 77(3):249–278. https://doi.org/10.1007/s10584-006-9062-1.

Kaufmann, R. K. and Stern, D. I. (1997). Evidence for human influence on climate from hemispheric temperature relations. *Nature*, 388(6637):39–44. https://www.nature.com/articles/40332.

Kaufmann, R. K. and Stern, D. I. (2002). Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research: Atmospheres*, 107(D2):ACL 8–1–ACL 8–10. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD000174.

Kawamura, R. (1994). A Rotated EOF Analysis of Global Sea Surface Temperature Variability with Interannual and Interdecadal Scales. *Journal of Physical Oceanography*, 24(3):707–715. https://journals.ametsoc.org/view/journals/phoc/24/3/1520-0485_1994_024_0707_areaog_2_0_co_2.xml.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S.,

Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Miller, R. L., Murray, L. T., Oinas, V., Orbe, C., García-Pando, C. P., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., and Yao, M.-S. (2020). GISS-E2.1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, 12(8):e2019MS002025. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002025.

Lee, T. C. K., Zwiers, F. W., Hegerl, G. C., Zhang, X., and Tsao, M. (2005). A Bayesian Climate Change Detection and Attribution Assessment. *Journal of Climate*, 18(13):2429–2440. https://journals.ametsoc.org/view/journals/clim/18/13/jcli3402.1.xml.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres*, 124(12):6307–6326. https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029522.

Li, L., Yu, Y., Tang, Y., Lin, P., Xie, J., Song, M., Dong, L., Zhou, T., Liu, L., Wang, L., Pu, Y., Chen, X., Chen, L., Xie, Z., Liu, H., Zhang, L., Huang, X., Feng, T., Zheng, W., Xia, K., Liu, H., Liu, J., Wang, Y., Wang, L., Jia, B., Xie, F., Wang, B., Zhao, S., Yu, Z., Zhao, B., and Wei, J. (2020). The Flexible Global Ocean-Atmosphere-Land System Model Grid-Point Version 3 (FGOALS-g3): Description and Evaluation. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS002012. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002012.

Li, S. and Jarvis, A. (2009). Long run surface temperature dynamics of an A-OGCM: The HadCM3 4×CO2 forcing experiment revisited. *Climate Dynamics*, 33(6):817–825. https://doi.org/10.1007/s00382-009-0581-0.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall Information and System Sciences Series. Prentice-Hall, Englewood Cliffs, N.J. ; London.

Ljungqvist, G. (2015). Decomposing global warming using Bayesian statistics. https://odr.chalmers.se/handle/20.500.12380/219282.

Lourens, E., Reynders, E., De Roeck, G., Degrande, G., and Lombaert, G. (2012). An augmented Kalman filter for force identification in structural dynamics. *Mechanical Systems and Signal Processing*, 27:446–460. https://www.sciencedirect.com/science/article/pii/S0888327011003931.

Lucarini, V., Ragone, F., and Lunkeit, F. (2017). Predicting Climate Change Using Response Theory: Global Averages and Spatial Patterns. *Journal of Statistical Physics*, 166(3):1036–1064. https://doi.org/10.1007/s10955-016-1506-z.

Luethi, D., Erb, P., and Otziger, S. (2020). FKF: Fast Kalman Filter. https://CRAN.R-project.org/package=FKF.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin Heidelberg. https://www.springer.com/gp/book/9783662026915.

MacKinnon, J. G. (2010). Critical values for cointegration tests. Queen's Economics Department Working Paper 1227, Queen's University, Department of Economics, Kingston (Ontario). http://hdl.handle.net/10419/67744.

McKitrick, R. (2021). Checking for model consistency in optimal fingerprinting: A comment. *Climate Dynamics*. https://doi.org/10.1007/s00382-021-05913-7.

Mills, T. C. (2008). How robust is the long-run relationship between temperature and radiative forcing? *Climatic Change*, 94(3):351. https://doi.org/10.1007/s10584-008-9525-7.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W. (2005). Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, 433(7026):613–617. https://www.nature.com/articles/nature03265.

Monahan, J. F. (1983). Fully Bayesian analysis of ARMA time series models. *Journal of Econometrics*, 21(3):307–331. https://www.sciencedirect.com/science/article/pii/0304407683900489.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, 117(D8). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD017187.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, 126(3):e2019JD032361. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032361.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H. (2013). Chapter 8 - Anthropogenic and Natural Radiative Forcing. In Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Doschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 659–740. Cambridge University Press, Cambridge, UK. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf.

Otto, F. E. L., Frame, D. J., Otto, A., and Allen, M. R. (2015). Embracing

uncertainty in climate change policy. *Nature Climate Change*, 5(10):917–920. https://www.nature.com/articles/nclimate2716.

Padilla, L. E., Vallis, G. K., and Rowley, C. W. (2011). Probabilistic Estimates of Transient Climate Sensitivity Subject to Uncertainty in Forcing and Natural Variability. *Journal of Climate*, 24(21):5521–5537. https://journals.ametsoc.org/view/journals/clim/24/21/2011jcli3989.1.xml.

Portmann, R. W., Solomon, S., and Hegerl, G. C. (2009). Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proceedings of the National Academy of Sciences*, 106(18):7324–7329. https://www.pnas.org/content/106/18/7324.

Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. https://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf.

Pretis, F., Mann, M. L., and Kaufmann, R. K. (2015). Testing competing models of the temperature hiatus: Assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change*, 131(4):705–718. https://doi.org/10.1007/s10584-015-1391-5.

Proistosescu, C. and Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, 3(7):e1602821. https://advances.sciencemag.org/content/3/7/e1602821.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Reid, I. (2001). Estimation II. Lecture notes, University of Oxford, Oxford. https://www.robots.ox.ac.uk/~ian/Teaching/Estimation/LectureNotes2.pdf.

Ribes, A., Azaïs, J.-M., and Planton, S. (2009). Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance

matrix estimate. *Climate Dynamics*, 33(5):707–722. https://doi.org/10.1007/s00382-009-0561-4.

Ribes, A., Planton, S., and Terray, L. (2013). Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate Dynamics*, 41(11):2817–2836. https://doi.org/10.1007/s00382-013-1735-7.

Ribes, A., Zwiers, F. W., Azaïs, J.-M., and Naveau, P. (2017). A new statistical approach to climate change detection and attribution. *Climate Dynamics*, 48(1):367–386. https://doi.org/10.1007/s00382-016-3079-6.

Richardson, M., Cowtan, K., Hawkins, E., and Stolpe, M. B. (2016). Reconciled climate response estimates from climate models and the energy budget of Earth. *Nature Climate Change*, 6(10):931–935. https://www.nature.com/articles/nclimate3066.

Rohde, R. A. and Hausfather, Z. (2020). The Berkeley Earth Land/Ocean Temperature Record. *Earth System Science Data*, 12(4):3469–3479. https://essd.copernicus.org/articles/12/3469/2020/.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., Frölicher, T. L., Paynter, D., Danabasoglu, G., Yang, S., Dufresne, J.-L., Cao, L., Schmidt, G. A., Abe-Ouchi, A., Geoffroy, O., and Knutti, R. (2020). Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models. *Geophysical Research Letters*, 47(4):e2019GL083898. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083898.

Rypdal, K. (2015). Attribution in the presence of a long-memory climate response. *Earth System Dynamics*, 6(2):719–730. https://esd.copernicus.org/articles/6/719/2015/.

Rypdal, M., Fredriksen, H.-B., Myrvoll-Nilsen, E., Rypdal, K., and Sørbye, S. H.

(2018). Emergent Scale Invariance and Climate Sensitivity. *Climate*, 6(4):93. https://www.mdpi.com/2225-1154/6/4/93.

Rypdal, M. and Rypdal, K. (2014). Long-Memory Effects in Linear Response Models of Earth's Temperature and Implications for Future Global Warming. *Journal of Climate*, 27(14):5240–5258. https://journals.ametsoc.org/view/journals/clim/27/14/jcli-d-13-00296.1.xml.

Särkkä, S., Alvarez, M. A., and Lawrence, N. D. (2019). Gaussian Process Latent Force Models for Learning and Stochastic Control of Physical Systems. *IEEE Transactions on Automatic Control*, 64(7):2953–2960.

Schervish, M. J. (1995). Large Sample Theory. In Schervish, M. J., editor, *Theory of Statistics*, Springer Series in Statistics, pages 394–475. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4250-5_7.

Schmith, T., Johansen, S., and Thejll, P. (2012). Statistical Analysis of Global Surface Temperature and Sea Level Using Cointegration Methods. *Journal of Climate*, 25(22):7822–7833. https://journals.ametsoc.org/view/journals/clim/25/22/jcli-d-11-00598.1.xml.

Schneider, T. and Held, I. M. (2001). Discriminants of Twentieth-Century Changes in Earth Surface Temperatures. *Journal of Climate*, 14(3):249–254. https://journals.ametsoc.org/view/journals/clim/14/3/1520-0442_2001_014_0249_ldotcc_2.0.co_2.xml.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M. (2020). Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci-*

entific Model Development, 13(12):6165–6200. https://gmd.copernicus.org/articles/13/6165/2020/.

Sellers, W. D. (1969). A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System. *Journal of Applied Meteorology and Climatology*, 8(3):392–400. https://journals.ametsoc.org/view/journals/apme/8/3/1520-0450_1969_008_0392_agcmbo_2_0_co_2.xml.

signal developers (2021). Signal: Signal processing. https://cran.r-project.org/package=signal.

Smith, C. J., Kramer, R. J., Myhre, G., Forster, P. M., Soden, B. J., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Hodnebrog, O., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Mülmenstädt, J., Olivié, D., Richardson, T., Samset, B. H., Shindell, D., Stier, P., Takemura, T., Voulgarakis, A., and Watson-Parris, D. (2018). Understanding Rapid Adjustments to Diverse Forcing Agents. *Geophysical Research Letters*, 45(21):12,023–12,031. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL079826.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J. (2008). Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006). *Journal of Climate*, 21(10):2283–2296. https://journals.ametsoc.org/view/journals/clim/21/10/2007jcli2100.1.xml.

Solomon, A., Goddard, L., Kumar, A., Carton, J., Deser, C., Fukumori, I., Greene, A. M., Hegerl, G., Kirtman, B., Kushnir, Y., Newman, M., Smith, D., Vimont, D., Delworth, T., Meehl, G. A., and Stockdale, T. (2011). Distinguishing the Roles of Natural and Anthropogenically Forced Decadal Climate Variability: Implications for Prediction. *Bulletin of the American Meteorological Society*, 92(2):141–156. https://journals.ametsoc.org/view/journals/bams/92/2/2010bams2962_1.xml.

Spolia, S. K. and Chander, S. (1974). Modelling of surface runoff systems

by an ARMA model. *Journal of Hydrology*, 22(3):317–332. https://www.sciencedirect.com/science/article/pii/0022169474900845.

Stern, D. (2005). A Three-Layer Atmosphere-Ocean Time Series Model of Global Climate Change.

Stern, D. I. (2006). An atmosphere–ocean time series model of global climate change. *Computational Statistics & Data Analysis*, 51(2):1330–1346. https://www.sciencedirect.com/science/article/pii/S0167947305003038.

Stern, D. I. and Kaufmann, R. K. (2000). Detecting a Global Warming Signal in Hemispheric Temperature Series: AStructural Time Series Analysis. *Climatic Change*, 47(4):411–438. https://doi.org/10.1023/A:1005672231474.

Stern, D. I. and Kaufmann, R. K. (2014). Anthropogenic and natural causes of climate change. *Climatic Change*, 122(1):257–269. https://doi.org/10.1007/s10584-013-1007-x.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P. C. B., and Wild, M. (2016). Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity. *Nature Geoscience*, 9(4):286–289. https://www.nature.com/articles/ngeo2670.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W. (2016). Attribution of extreme weather and climate-related events. *WIREs Climate Change*, 7(1):23–41. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.380.

Sutton, R., Suckling, E., and Hawkins, E. (2015). What does global mean temperature tell us about local climate? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2054):20140426. https://royalsocietypublishing.org/doi/10.1098/rsta.2014.0426.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11):4823–4873. https://gmd.copernicus.org/articles/12/4823/2019/.

Tanaka, K., Raddatz, T., O'Neill, B. C., and Reick, C. H. (2009). Insufficient forcing uncertainty underestimates the risk of high climate sensitivity. *Geophysical Research Letters*, 36(16). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL039642.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7):2727–2765. https://gmd.copernicus.org/articles/12/2727/2019/.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498. https://journals.ametsoc.org/view/journals/bams/93/4/bams-d-11-00094.1.xml.

Thompson, D. W. J. and Wallace, J. M. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, 25(9):1297–1300. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL00950.

Thompson, D. W. J., Wallace, J. M., Jones, P. D., and Kennedy, J. J. (2009). Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights. *Journal of Climate*,

22(22):6120–6141. https://journals.ametsoc.org/view/journals/clim/22/22/2009jcli3089.1.xml.

Tollefson, J. (2021). COVID curbed carbon emissions in 2020 — but not by much. *Nature*, 589(7842):343–343. https://www.nature.com/articles/d41586-021-00090-3.

Tsutsui, J. (2017). Quantification of temperature response to CO2 forcing in atmosphere–ocean general circulation models. *Climatic Change*, 140(2):287–305. https://doi.org/10.1007/s10584-016-1832-9.

Tsutsui, J. (2020). Diagnosing Transient Response to CO2 Forcing in Coupled Atmosphere-Ocean Model Experiments Using a Climate Model Emulator. *Geophysical Research Letters*, 47(7):e2019GL085844. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085844.

Turasie, A. A. (2012). Cointegration modelling of climatic time series. page 166.

Tusell, F. (2011). Kalman Filtering in R. *Journal of Statistical Software*, 39(1):1–27. https://www.jstatsoft.org/index.php/jss/article/view/v039i02.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, 36(5):823–841. https://link.aps.org/doi/10.1103/PhysRev.36.823.

Urban, N. M., Holden, P. B., Edwards, N. R., Sriver, R. L., and Keller, K. (2014). Historical and future learning about climate sensitivity. *Geophysical Research Letters*, 41(7):2543–2552. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL059484.

Urban, N. M. and Keller, K. (2010). Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: A Bayesian fusion of century-scale observations with a simple model. *Tellus A: Dynamic Meteorology and Oceanography*, 62(5):737–750. https://doi.org/10.1111/j.1600-0870.2010.00471.x.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 23(3):395–404.

Venzke, S., Allen, M. R., Sutton, R. T., and Rowell, D. P. (1999). The Atmospheric Response over the North Atlantic to Decadal Changes in Sea Surface Temperature. *Journal of Climate*, 12(8):2562–2584. https://journals.ametsoc.org/view/journals/clim/12/8/1520-0442_1999_012_2562_tarotn_2.0.co_2.xml.

Vial, J., Dufresne, J.-L., and Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, 41(11):3339–3362. https://doi.org/10.1007/s00382-013-1725-9.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R. (2019). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, 11(7):2177–2213. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001683.

Wallace, J. M., Zhang, Y., and Bajuk, L. (1996). Interpretation of Interdecadal Trends in Northern Hemisphere Surface Air Temperature. *Journal of Climate*, 9(2):249–259. https://journals.ametsoc.org/view/journals/clim/9/2/1520-0442_1996_009_0249_ioitin_2_0_co_2.xml.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.

(2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, 10(2):357–380. https://onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001115.

Wood, S. (2021). Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. https://cran.r-project.org/package=mgcv.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00749.x.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y., and Liu, X. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, 12(4):1573–1600. https://gmd.copernicus.org/articles/12/1573/2019/.

Wynn, A., Pearson, D. S., Ganapathisubramani, B., and Goulart, P. J. (2013). Optimal mode decomposition for unsteady flows. *Journal of Fluid Mechanics*, 733:473–503. https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/abs/optimal-mode-decomposition-for-unsteady-flows/7F006F5D7F3DA64D8C2658B3B3B55609.

Ypma, J., Johnson, S. G., Borchers, H. W., Eddelbuettel, D., Ripley, B., Hornik, K., Chiquet, J., and Adler, A. (2020). Nloptr: R Interface to NLopt. https://cran.r-project.org/package=nloptr.

Yu, D. and Chakravorty, S. (2015). An autoregressive (AR) model based stochastic unknown input realization and filtering technique. In *2015 American Control Conference (ACC)*, pages 1499–1504.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M. (2019). The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component. *Journal of the Meteorological Society of Japan. Ser. II*, 97(5):931–965.

Yule, G. U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society*, 89(1):1–63. `https://www.jstor.org/stable/2341482`.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., Srbinovsky, J., Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, 70(1):193–214. `https://www.publish.csiro.au/es/ES19035`.