Relating forced climate change to natural variability and emergent dynamics of the climate-economy system

Submitted by

Owen Kellie-Smith

to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, March 2010.

I certify that all material in this thesis which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.

.....

Owen Kellie-Smith

Abstract

This thesis is in two parts. The first part considers a theoretical relationship between the natural variability of a stochastic model and its response to a small change in forcing. Over a large enough scale, both the real climate and a climate model are characterised as stochastic dynamical systems. The dynamics of the systems are encoded in the probabilities that the systems move from one state into another. When the systems' states are discretised and listed, then transition matrices of all these transition probabilities may be formed. The responses of the systems to a small change in forcing are expanded in terms of the eigenfunctions and eigenvalues of the Fokker-Planck equations governing the systems' transition densities, which may be estimated from the eigenvalues and eigenvectors of the transition matrices. Smoothing the data with a Gaussian kernel improves the estimate of the eigenfunctions, but not the eigenvalues. The significance of differences in two systems' eigenvalues and eigenfunctions is considered. Three time series from HadCM3 are compared with corresponding series from ERA-40 and the eigenvalues derived from the three pairs of series differ significantly.

The second part analyses a model of the coupled climate-economic system, which suggests that the pace of economic growth needs to be reduced and the resilience to climate change needs to be increased in order to avoid a collapse of the human economy. The model condenses the climate-economic system into just three variables: a measure of human wealth, the associated accumulation of greenhouse gases, and the consequent level of global warming. Global warming is assumed to dictate the pace of economic growth. Depending on the sensitivity of economic growth to global warming, the model climate-economy system either reaches an equilibrium or oscillates in century-scale booms and busts.

Acknowledgements

This work is the result of a lot of people's thought and care and encouragement and I am very grateful to you all.

The first part was conceived by Prof John Thuburn, and was developed with Prof Thuburn and Prof Jonathan Gregory and Prof Stuart Townley. The European Research Course on Atmospheres, NCAS Summer Schools, Dynamics Days Europe and courses at the University of Exeter introduced the science. Dr Daan Crommellin suggested the kernel density estimation method. The British Atmospheric Data Centre helped identify relevant datasets. The second part spun off from the training for the first part and was a collaboration with Prof Peter Cox. Participants at the internal seminars in the School of Engineering, Maths and Computing at the University of Exeter made many constructive suggestions. Dr Nigel Byott explained the algebra. Prof Thuburn and anonymous reviewers of an earlier draft clarified the second part. Prof Cox and Dr Tim Osborn placed the whole work in context and drew out its conclusions. Many thanks to my supervisors for your skill, patience, advice and encouragement.

The project was funded by the Natural Environment Resources Council and the Met Office. Thank you for this great opportunity.

A huge community kept me on the road. Special thanks to my family, especially Ewa, Nadia, Maya, and Mum & Dad for your love and interest and encouragement. Thanks to Prof Townley and Dr Sebastian Wieczorek for steering me through the PhD doldrums. Thanks to friends at Exeter and ERCA and NCAS for sharing laughs and rants and entertainment. Finally thanks to my inspiring school Maths teachers: Colin Harding and the late lamented Jonathan Bull.

Contents

List of Figures

Part I

Relating forced climate change to natural variability

Chapter 1

Introduction: testing climate models

1.1 Background

Climate policy will have a major global impact whether climate change is a distraction from other more pressing problems [29] or the most urgent issue of the current time [42]. Since climate policy is heavily based on the projections of climate models, it is natural to ask whether climate models are credible. The Inter Governmental Panel on Climate Change (IPCC) is charged with making policy-relevant assessment of the scientific literature so it discusses the credibility of climate forecasts in detail [41]. To start with, it challenges what is meant by credible. There are different standards of proof or credibility depending on what can realistically be tested.

1.1.1 Standards of proof vary by science

The IPCC's gold standard of credibility is for a model or forecast of an event that can be entirely controlled [41]. An example would be a physical transformation where all the environmental conditions (for example temperature, pressure, atmospheric composition) can be repeated. Such a transformation can be carried out with a high degree of regularity. Given enough repeated demonstrations, the

'model' outcome is treated as a matter of fact.

Less control is possible for, say, the trials of a new medical drug. The boundary conditions are uncontrollable. Patients are all different, so some of the boundary conditions are left to chance. So, no matter how well the other conditions are controlled, all that can be obtained is a probabilistic result. Despite these uncertainties, this is a workable level of reliability for the medical industry.

For weather forecast models, no control is possible. There are no spare Earths to treat as a control group and in any case there are no control mechanisms. Weather forecasters cannot set up a real-world experiment, let alone repeat it [28]. The most reliable forecasters become the most trusted, but this requires that their forecasts are over a short-enough timescale that they can be repeatedly evaluated [23].

Weather forecast models start with an initial state of the Earth-atmosphere system (based on observations and previous forecasts) and project that initial state forward over a brief sequence of short timescales and with high spatial resolution. Climate models project over longer timescales with lower spatial resolution, so that climate model outputs are general features of the model system [22](e.g. global mean surface temperature) and not heavily dependent on initial conditions [11].

For climate models, control is not possible and, furthermore, repeated evaluations of forecasts take too long to be of any use. All that is available to climate scientists are actual observations which may be explained according to some theory or model. But if a climate model encodes universal laws of nature, its explanations can apply to any state of nature. So climate models are challenged to compare the recent and distant past (to the extent that these are actually known).

In short, different standards of credibility apply depending on what test can conceivably be applied to a theory. The more complex the system, the harder it is to test a theory, and the less precise is its credibility. Some of the criticism of climate models comes from applying an inappropriate or indeed impossible standard of credibility to them (for example, by requiring 100% high-resolution accuracy from weather forecasts). But implicit in the question of whether climate projections are credible is the reasonable assumption that observations constrain what the models

can credibly project. So, another way of asking whether models are credible is to ask how observations constrain models.

1.1.2 Observational constraints placed on climate models

As far as possible, climate models are based on well-known laws of dynamics and thermodynamics, laws which are themselves based on observations and controllable experiments. However there are some climate processes (e.g. cloud microphysics) which are not understood to this extent. These processes must be parameterised so that the statistics of their modelled behaviour matches the observed statistics of the real processes' behaviour.

Models are (necessarily) only approximations. The spatial and temporal resolution affordable in climate models is limited by the size and speed of the available computers. Processes like convection and those in the boundary layer, which occur over smaller scales than the affordable resolution must be parameterised.

This parameterisation looks circular: if the parameterisations are chosen to fit the observed data then it may be no surprise if model output fits the observed data. But model output incorporates not just the parameterised processes, but the interaction of those processes with all the other processes. In particular, model feedbacks, where the effect of one process is amplified or dampened by another, are a model 'output' rather than an 'input'. Thus observations place a set of necessary, but not sufficient conditions on the feedbacks of climate models.

Several key feedbacks of climate models have been constrained by the observed data. Observed data, for example of short-term variability due to seasonal variation, has been used to constrain long-term feedbacks and therefore long-term responses to a general change in forcing [18, 20, 16, 14, 31].

Recent observations are the most detailed, so an informal constraint for climate models is to require them to match the recent past. It has been suggested that this informal constraint has resulted in a survivor-bias by which only models that match the twentieth century get developed [5]. Nevertheless, models do now match more closely the mean fields of twentieth century temperature, sea-level-pressure,

meridional overturning circulation, precipitation and cyclones. Furthermore extremes are better matched, as are modes of variability, both spatial patterns like ENSO and the polar-tropical contrast in warming, as well as temporal patterns like the seasonal pattern, and the night-day contrast in warming [41].

Models have also matched aspects of past climates, for example the mid-Holocene, the Last Glacial Maximum, as well as successfully making short-term forecasts of the cooling caused by the eruption of Mount Pinatubo.

Patterns have been noted in the projections of the models which best fit the mean data of the recent past. For example, the smallest errors in matching the twentieth century sea-surface temperatures are from models with the largest projected twenty-first century temperature change [39]. A strong seasonal cycle has been linked to large climate sensitivity [25]. For these associations to carry weight there must be reasons to believe that matching the past means that a model will match the future[41, section 8.1.2.2]. But Kohfeld and Harrison [26] point out that being able to simulate the current climate, while an important necessary condition, is not sufficient to guarantee that a model simulates changes correctly. Reifen and Toumi [37] show that for any 10 year period in the twentieth century, the model which most closely matches observations is never the same for two consecutive decades. A trajectory may be matched by chance, that is, by particular coincidences between the system and the model over the period of evaluation.[1]

Brohan [5] has argued that it is not even necessary to match the twentieth century, since a complex model run twice with identical boundary and near-identical initial conditions will almost certainly produce radically different trajectories.

There may be processes which happen in the future which do not happen in the recent past (or which happened at different rates). If that is the case, then matching the recent past will be insufficient to match the future. For example, adding in a carbon cycle had a big impact on future climate projections [12] though would have had a much smaller impact on reconstructions of the twentieth century.

However, under certain assumptions, the sensitivity of a system's response to a

[1] This has also been observed in financial models, to the extent that one established 'contrarian' strategy is to sell winners and buy losers [7].

small change in forcing is proportional to its natural variability. This is illustrated by Bell [3], who observes that a mechanical spring's strength (and thus response to extra weight) can be inferred from the speed of its oscillations. The *fluctuation dissipation theorem* of Leith [28] proves that under certain conditions, the average response of a system to a small change in forcing is proportionate to the change in forcing, where the proportionality is dictated by a well-defined function of the lag-covariances of the unperturbed system's state variables. North *et al.* [34] compare the response predicted by the theorem with the actual response obtained by perturbing a GCM and find the theorem applies reasonably well, but note that the real climate does not satisfy its assumptions. Gritsun and Branstator [19] extends the fluctuation-dissipation theorem, proving that it applies under more general assumptions than required for [28].

The fluctuation-dissipation theorem [28, 19] states that for a sufficiently small change in forcing (even for non-linear systems), the linear response of the system's expected *state* may be inferred from the natural variability of the unperturbed system. As an alternative to the fluctuation-dissipation theorem, this chapter formalises a linear response of the system's *probability density function* (pdf) to a sufficiently small change in forcing. This is made possible, because, under the same conditions as assumed by [47, 19], even though the dynamics of the system may be non-linear, the dynamics of the system's pdf (encoded in its Fokker-Planck equation) are linear.

1.1.3 Transition densities

One way of formalising the comment in section 1.1.1 that standards of proof vary by science is to consider what experiments can actually show. This can be put into a common currency by considering a system *state space*, that is the (discrete or continous) set of all possible states that a system may be in, and a *transition density*, that is a function that quantifies the likelihood that the system comes out of one state (or region of state space) and goes into another. The probability that the system moves from one state into another is the integral of the transition

density over the pair of states.

For example, an entirely controllable and predictable experiment could be described by two states (the states before and after the transformation) with a 100% rate of transition from one state to the other. The transition density would be a (two dimensional) delta function, centred on the pre- and post-transformation states.

In a test of, say a new medical treatment, then even if there were only two states (say healthy and sick), there will usually be a transition of less than 100% (following treatment) between the two states.

The weather may be described by a continous state space of many dimensions (possibly infinitely many e.g. temperature, pressure, humidity at innumerable places on the globe). For each region of state space there is, over a certain time interval, a transition density to all the other regions of the state space. Given a long-enough time interval, we may assume that the transition density is independent of the starting state. For example, whether it is raining a month from now does not depend (significantly) on whether it is raining right now.

For this long-term transition density, which can be formalised as the limiting density as the time interval increases, it makes more sense to talk of a *probability density*, because the density becomes progressively independent of its starting state.

This leads to a useful definition of 'climate', and climate change. The climate may be defined as the limit (as time increases) of the transition density of the state space of the earth-atmosphere system. Climate change, then, represents a change in the likelihood of various states of the system [47].

With this idea of climate as a density i.e. a probability distribution, it is worth revisiting what has actually been matched between the real climate and climate models (outlined in section 1.1.2). The great progress in increasing agreement between features of the twentieth century climate record and the corresponding output from climate models boils down to agreement in aspects of the observed climate, i.e. in the long term probability distribution. However, even if a (forced) model matches a real system's entire probability distribution (i.e. with the same

relative frequency of all events, the same distribution of extremes etc), then this still does not guarantee that the model will respond to a change in forcing (no matter how small) in the same way as the real system (shown in section 1.1.4).

1.1.4 Differences in forced response even if model matches real mean density exactly

Two models may have the same steady probability density function but still have quite different transition probabilities. For instance, suppose the weather in a particular place alternates so that a sunny day always follows a cloudy day and vice-versa. Compare this to another place which is cloudy every day from October to March, and then sunny every day from April to September. Both places are sunny for half of the year and cloudy for the other half. The climates for both places (according to our definition, which are the steady probability density functions) are the same, but the transition densities are different.

Similarly, two models may have the same steady probability density function (not just the same mean and variance, but the same entire distribution), but still have different steady state responses to a change in forcing, consistent with [26]. This is demonstrated in figure 1.1 on page 40. Therefore, even if a model matches the actual mean climate, that is not enough to conclude that it will forecast the correct equilibrium forced climate, let alone the transient forced climate.

If a model has the same transition density as the real system, then statistically it is identical to the system [22]. It therefore makes sense to suppose that the model will respond in the same way as the real system, at least to a small change in forcing. This is formalised in section 1.2, which also suggests that it is not necessary for a model to match the entire transition density of the real system, but just its dominant modes. The transition density describes the system's 'internal variability', and so gives a relation between internal variability and forced climate change.

1.1.5 Comparing transition densities

The practical problem is to compare the transition densities of a climate model with the densities of the observed climate. This is done in chapter 2 by taking time series of climate model output and reanalysis output, discretising the state space to a finite set of states, sampling transition matrices that count the relative frequencies of transition from one state to another over a constant time-step, and comparing the eigenvectors and eigenvalues of the sampled transition matrices.

The transition matrices are random variables, so any difference between their sampled values may not necessarily indicate an actual underlying difference in transition densities. All that can be said is whether it is likely that the difference arises by chance.

The size of the available data limits the number of states that can credibly be sampled. For instance, with 45 years of daily ERA-40 data, there are 16,425 data points, so at most 16,424 transitions can be observed, and the rate of transition can be credibly estimated only for a much lower number of states. If 1000 states are used, made up by discretizing particular climate variables (e.g. an El Niño index) into, say, 10 discrete intervals, then the number of individual climate variables that can be compared together is no more than 3.

Clearly, there are several choices implicit in this method of comparison. How many and which states should be used? What length of dataset is required? Should the data or densities be smoothed and if so how? These are addressed in chapters 2-3.

Behind those methodological questions is the motivating question of whether a climate model matches the available data sufficiently for all its forced climate change forecasts to be accurate. Just three series from HadCM3 and ERA-40 are tested in in chapter 4. The eigenvalues of the transition densities based on those three series differ by more than can be attributed to natural variability, but the theory needs to be developed further in order to say whether this actually indicates a significant difference between the actual and model responses.

1.2 Theory

The steady pdf response of a particular kind of stochastic system is expressed in this section as a function of the transition density of the system under constant forcing. Thus, two such systems with the same transition densities have the same steady pdf under any small change in forcing. Section 1.2 is based on [46].

1.2.1 Climate defined as a probability density function (pdf)

The state of the earth-atmosphere system may be reduced to a finite dimensional vector $\vec{x}$ [22]. Let $f(\vec{x},t)$ be a time-dependent probability density function, so the relative likelihood that the system is in state $\vec{x}$ at time t is $f(\vec{x},t)$.

1.2.2 Earth-atmosphere dynamics assumed to be a diffusion process

It is assumed, as in [50, 47, 19], that the state vector $\vec{X}$ evolves according to a stochastic differential equation

$$\mathrm{d}\vec{X}_t = \mathbf{M}(\vec{X})\ \mathrm{d}t + \varepsilon\ \mathrm{d}\vec{W}_t, \tag{1.1}$$

The state variable $\vec{X}$ is a random variable. Over a small timestep $\mathrm{d}t$, changes in the state vector $\mathrm{d}\vec{X}_t$ come from a combination of (deterministic) advection in state space, in the form of $\mathbf{M}(\vec{X})\ \mathrm{d}t$, and (random) diffusion in state space, in the form of $\varepsilon\ \mathrm{d}\vec{W}_t$. $\mathbf{M}(\vec{x})$ is called the drift [24], and it represents the large scale resolved processes in a climate model. For a climate to exist (in the sense of section 1.1.3 i.e. a limiting distribution for the transition density), $\mathbf{M}$ must be a restoring force, otherwise the variability introduced by $\mathrm{d}\vec{W}$ would grow indefinitely [22]. $\vec{W}_t$ is a vector of independent Wiener processes, i.e. random walks. Fluctuations in the Wiener processes of size $\varepsilon\ \mathrm{d}\vec{W}_t$ diffuse the probability density independently from the drift $\mathbf{M}$. These random contributions to changes in $\vec{X}_t$ represent the small scale processes which are not resolved by the model, but which are part of the system and whose effects are integrated by the larger scale, resolved processes [22, 36, 19].

ε is a constant matrix, with any finite number of columns, and d rows, where d is the length of $\vec{X}$, i.e. the number of dimensions in the model.

Fokker-Planck equation (FPE)

Because (1.1) is a diffusion process [24, page 34], the probability density function $f(\vec{x},t)$ of the model state satisfies the Fokker-Planck equation [24, page 37]:

$$\frac{\partial f}{\partial t} + \nabla.(\mathbf{M}f) = \partial_i \; \partial_j \; (K_{ij} \; f)\,, \tag{1.2}$$

$$\text{where } \mathbf{K} = \frac{\varepsilon \; \varepsilon^T}{2} \tag{1.3}$$

and the summation convention is used for the diffusion term, so that

$$\partial_i \; \partial_j \; (K_{ij} \; f) = \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(\mathbf{K}_{ij} f(\vec{x})\right). \tag{1.4}$$

As f is a pdf, it must integrate to one and be non-negative everywhere. So, f must fall to zero in the extreme regions of state space. The boundary condition on (1.2) is that, for any polynomial $p(\vec{x})$

$$p(\vec{x}) f(\vec{x}) \to 0 \text{ as } |\vec{x}| \to \infty. \tag{1.5}$$

Equation (1.2) is an advection-diffusion equation for probability. The probability density of a state tends to increase where the advected probability converges and tends to decrease where the advected probability diverges. Probability diffuses into a region if the concentration of probability in that region is convex. This is illustrated in figure 1.2 on page 41.

1.2.3 Eigenfunctions of Fokker Plank equation are used to form a basis for the pdf

The eigenfunctions of equation (1.2) are the separable solutions to equation (1.2) of the form $f_k(\vec{x})e^{-\mu_k t}$, where $f_k(\vec{x})$ are functions over state space and μ_k are constants. Substituting the eigenfunctions into equation (1.2) and dividing by $e^{-\mu_k t}$,

$$-\mu_k f_k + \nabla.(\mathbf{M}f_k) = \partial_i \; \partial_j \; (K_{ij} \; f_k)\,. \tag{1.6}$$

We hypothesise that $\{f_k\}_{k=0}^{k=\infty}$ forms a complete set, that is that any function over the state space can be described in terms of combinations of f_k.[2] So, under this hypothesis, the initial pdf of the system, say $h(\vec{x})$, can be expanded in terms of the eigenfunctions $f_k(\vec{x})$. That is, there are constants b_k for which

$$h(\vec{x}) = f(\vec{x}, 0) = \sum_{k=0}^{\infty} b_k f_k(\vec{x}). \tag{1.7}$$

These constants, b_k, define the pdf $f(\vec{x},t)$ for all t. For, let

$$\hat{f}(\vec{x}, t) = \sum_{k=0}^{\infty} b_k f_k(\vec{x}) e^{-\mu_k t}. \tag{1.8}$$

Then $\hat{f}(\vec{x}, t)$ satisfies the linear equation (1.6) by the definition of f_k, and also $\hat{f}(\vec{x}, t)$ satisfies the initial condition that $\hat{f}(\vec{x}, 0) = h(\vec{x})$. So, $\hat{f}(\vec{x}, t)$ is $f(\vec{x},t)$, the pdf of the system.

Eigenfunctions of adjoint are orthonormal with eigenfunctions of FPE

The eigenfunctions of equation (1.6) are made orthonormal to the eigenfunctions of the adjoint of equation (1.6), in order to expand the steady pdf under a change in forcing, in terms of the eigenfunctions.

Let equation (1.6) be written

$$-\mu_k f_k = \mathcal{L}(f_k) \tag{1.9}$$

where $\mathcal{L}(f)$ is the linear operation $-\nabla.(\mathbf{M}f) + \partial_i\ \partial_j\ (K_{ij}\ f)$.

The adjoint [27] of $\mathcal{L}$ is $\mathcal{L}^*$, defined by

$$\forall f, g \int \mathcal{L}(f)\ g\ \mathrm{d}\vec{x} = \int f\ \mathcal{L}^*(g)\ \mathrm{d}\vec{x}. \tag{1.10}$$

The boundary condition on g is that

$$g(\vec{x}) \text{ and } \nabla g(\vec{x}) \text{ grow no faster than a polynomial as } |\vec{x}| \to \infty. \tag{1.11}$$

[2]For example, $\{f_k\}_{k=0}^{k=\infty}$ forms a complete set for the functions over the 1-dimensional state space for the linear Langevin equation illustrated in chapter 3. For the 1-dimensional linear Langevin equation, the f_k are polynomials of order k times a Gaussian density.

Integrating equation (1.10) by parts, and assuming by boundary conditions (1.5) and (1.11) that $|\mathbf{M}fg|$, $g\nabla.f$ and $f\nabla.g$ tend to zero towards the boundary of state space

$$\int \mathcal{L}(f)\ g\ \mathrm{d}\vec{x} = \int \left(-\nabla.(\mathbf{M}f) + \partial_i\ \partial_j\ \left(K_{ij}\ f\right)\right)\ g\ \mathrm{d}\vec{x} \tag{1.12}$$

$$= \int f\left(\mathbf{M}.\nabla g + K_{ij}\ \partial_i\ \partial_j\ g\right)\ \mathrm{d}\vec{x}. \tag{1.13}$$

So, using the summation convention, the adjoint operator is

$$\mathcal{L}^*(g) = \mathbf{M}.\nabla g + K_{ij}\ \partial_i\ \partial_j\ g. \tag{1.14}$$

Let g_k be eigenfunctions which satisfy the adjoint equation

$$-\nu_k g_k = \mathcal{L}^*(g_k) = \mathbf{M}.\nabla g_k + K_{ij}\ \partial_i\ \partial_j\ g_k. \tag{1.15}$$

The set of eigenvalues $\{\mu_k\}_{k=0}^{\infty}$ is the same as the set of eigenvalues $\{\nu_k\}_{k=0}^{\infty}$. To see this, multiply equation (1.6) by g_l from equation (1.15) and integrate over state space:

$$-\mu_k \int f_k g_l\ \mathrm{d}\vec{x} = \int \left(-\nabla.(\mathbf{M}f_k) + \partial_i\ \partial_j\ \left(K_{ij}\ f_k\right)\right) g_l\ \mathrm{d}\vec{x}. \tag{1.16}$$

Integrate by parts, and make the same assumptions about boundary conditions as for the adjoint equation (1.12). Necessarily, since the adjoint was obtained, equation (1.16) involves the adjoint:

$$-\mu_k \int f_k g_l\ \mathrm{d}\vec{x} = \int f_k\left(\mathbf{M}.\nabla g_l + K_{ij}\ \partial_i\ \partial_j\ g_l\right)\ \mathrm{d}\vec{x} \tag{1.17}$$

$$= -\nu_l \int f_k g_l\ \mathrm{d}\vec{x} \text{ from equation (1.15)} \tag{1.18}$$

$$\Rightarrow (\nu_l - \mu_k) \int f_k g_l\ \mathrm{d}\vec{x} = 0. \tag{1.19}$$

Equation (1.19) implies that either $\nu_l = \mu_k$ or $\int f_k g_l\ \mathrm{d}\vec{x} = 0$. Assuming that for every f_k there is some g_l for which $\int f_k g_l\ \mathrm{d}\vec{x} \neq 0$, there is a ν_l equal to each μ_k. Order and relabel the ν_l so that for all k, $\nu_k = \mu_k$. Then, by equation (1.19)

$$\int f_k g_l\ \mathrm{d}\vec{x} = 0 \text{ for } k \neq l. \tag{1.20}$$

Both f_k and g_k can be multiplied by any constant and still satisfy the (linear) equation (1.6) and equation (1.15). So normalise f_k and g_k so that

$$\int f_k g_k \, \mathrm{d}\vec{x} = 1 \text{ and } \int f_0 \, \mathrm{d}\vec{x} = 1. \tag{1.21}$$

Equation (1.21) does not uniquely define the scale of f_k or g_k, but this does not matter, as only the products of f_k and g_k will be used.

Projection of initial pdf onto eigenfunctions

The constants b_k in equation (1.7) are obtained via (1.21). Multiplying (1.7) by g_l and integrating over state space,

$$\int h(\vec{x}) g_l(\vec{x}) \, \mathrm{d}\vec{x} = \sum_{k=0}^{\infty} b_k \int f_k(\vec{x}) g_l(\vec{x}) \, \mathrm{d}\vec{x} \tag{1.22}$$

$$= b_l \text{ by (1.20) and (1.21).} \tag{1.23}$$

If the state at time t=0 is known with certainty, so that $\vec{X}_t = \vec{x}_t$, then the pdf $h(\vec{x})$ at time t is a delta function centred on $\vec{x}_t$, so that (1.23) implies that

$$b_l = \int \delta(\vec{x}_t) g_l(\vec{x}) \, \mathrm{d}\vec{x} = g_l(\vec{x}_t). \tag{1.24}$$

1.2.4 Expansion of forced steady pdf in terms of eigenfunctions of unforced FPE and adjoint

This section defines the steady pdf[3] as the eigenfunction with eigenvalue zero. The Fokker-Planck equation is used to approximate the effect on the steady pdf of a small change in the drift vector $\mathbf{M}$. The approximate small change in pdf is projected onto the eigenfunctions $\{f_k\}$, and a formula for each eigenfunction coefficient is obtained. The change in $\mathbf{M}$ could, in practice, come from a change in forcing or feedback, or both.

First, observe that the steady pdf of the Fokker-Planck equation has a zero eigenvalue. If, for any k, $\mu_k < 0$ then $e^{-\mu_k t} \to \infty$ as $t \to \infty$ and so for all $\vec{x}$,

[3]Jargon: the steady pdf is not a state of the system e.g. hot or cold, but a probability distribution. It is the equilibrium climate according to the definition in section 1.1.3

$|f(\vec{x},t)| \to \infty$. This would violate the boundary condition that f integrates to one and is always non-negative. So $\mu_k \geq 0$. If $\forall k, \mu_k > 0$ then $e^{-\mu_k t} \to 0$ and so for all $\vec{x}$, $|f(\vec{x},t)| \to 0$ as $t \to \infty$. This would violate the boundary condition that f integrates to one. So there is a k for which $\mu_k = 0$. Let this μ_k be denoted μ_0 so that $f_0(\vec{x})$ is the system's steady pdf. Thus, by (1.15) g_0 is a constant and so by (1.21), $g_0 = 1$.

Second, the effect on f_0 of a small change in $\mathbf{M}$ is estimated. Substituting $\mu_0 = 0$ and f_0 into equation (1.6)

$$\nabla.(\mathbf{M}f_0) = \partial_i \; \partial_j \; (K_{ij} \; f_0) \, . \tag{1.25}$$

If the forcing is changed slightly so the model becomes

$$\mathrm{d}\vec{X} = (\mathbf{M} + \Delta\mathbf{M})\mathrm{d}t + \varepsilon \; \mathrm{d}\vec{W} \tag{1.26}$$

with new steady pdf $(f_0 + \Delta f)$, then the steady pdf equation (1.25) becomes

$$\nabla. \left\{ (\mathbf{M} + \Delta\mathbf{M}) \, (f_0 + \Delta f) \right\} = \partial_i \; \partial_j \; \left(K_{ij} \; \left(f_0 + \Delta f \right) \right) . \tag{1.27}$$

Neglecting the product of small terms $\nabla. \, (\Delta\mathbf{M}\Delta f)$ and subtracting (1.25) from (1.27),

$$\nabla.(\mathbf{M} \; \Delta f) - \partial_i \; \partial_j \; (K_{ij} \; \Delta f) = -\nabla.(\Delta\mathbf{M} \; f_0). \tag{1.28}$$

Third, Δf is expanded in terms of eigenfunctions. Assuming as above that the eigenfunctions are complete, there are constants $\{a_k\}$ for which

$$\Delta f = \sum_{k=0}^{\infty} a_k f_k(\vec{x}). \tag{1.29}$$

$a_0 = 0$ because both f_0 and $f_0 + \Delta f$ are pdfs, so

$$\int f_0 \; \mathrm{d}\vec{x} = \int (f_0 + \Delta f) \; \mathrm{d}\vec{x} = 1 \tag{1.30}$$

and so

$$\int \Delta f \; \mathrm{d}\vec{x} = 0. \tag{1.31}$$

By the orthogonality of f_k and g_0,

$$\int f_k \, \mathrm{d}\vec{x} = 0 \text{ for } k > 0. \tag{1.32}$$

Substituting for Δf from (1.29), and using (1.32)

$$0 = \int \Delta f \, \mathrm{d}\vec{x} = a_0 \int f_0 \, \mathrm{d}\vec{x} = a_0. \tag{1.33}$$

Hence,

$$\Delta f = \sum_{k=1}^{\infty} a_k f_k(\vec{x}). \tag{1.34}$$

Finally, the sum of eigenfunctions is substituted for Δf in equation (1.28). The result is that

$$\sum_{k=1}^{\infty} a_k \left(\nabla.(\mathbf{M} \; f_k) - \partial_i \; \partial_j \; (K_{ij} \; f_k)\right) = -\nabla.(\Delta\mathbf{M} \; f_0) \tag{1.35}$$

$$\Rightarrow \sum_{k=1}^{\infty} a_k \mu_k f_k = -\nabla.(\Delta\mathbf{M} f_0) \text{ by (1.6)}. \tag{1.36}$$

Multiplying equation (1.36) by g_l and integrating over state space, using the orthonormal conditions (1.20) and (1.21),

$$a_l = -\frac{1}{\mu_l} \int g_l \nabla.(\Delta\mathbf{M} f_0) \, \mathrm{d}\vec{x}. \tag{1.37}$$

Integrating equation (1.37) by parts with the assumption (1.5) about far-field boundary conditions, and relabelling l to k, the result is obtained that

$$a_k = \frac{1}{\mu_k} \int f_0 \; \Delta\mathbf{M}.\nabla g_k \, \mathrm{d}\vec{x}, k = 1, 2, ... \tag{1.38}$$

Equation (1.38) is similar to [47, equation(16)]. Since (1.38) is used in (1.34), only products of f_k and g_k are used, so it does not matter that the normalisation of f_k and g_k only defines the scale of their product.

1.2.5 Meaning of eigenvalues and eigenfunctions

Putting (1.38) in words: the projection onto each eigenfunction of the equilibrium response to a change in drift is (to first order) equal to the average change in

convergence of the drift weighted by the corresponding eigenfunction of the adjoint or 'backwards' equation[24] and the timescale of the eigenfunctions (which is the reciprocal of the eigenvalue).

As the time horizon goes forwards, the pdf, conditional on the initial state, tends to the steady state. In other words, differences from the steady state decay to nil as the time horizon increases. The differences are expressible as a linear sum of the eigenfunctions, and the contribution from each eigenfunction decays exponentially over time at the rate of the eigenvalue. (Eigenfunctions with high eigenvalues decay quickly.)

As the time horizon goes backwards, the eigenfunctions of the backwards equation are components of the distribution from which the current state might have come. The current state gets equally likely (as you go back in time) to have come from any previous state. That is, the the backwards density tends to a uniform distribution which is g_0. The rate at which the backwards density tends to the uniform density is set by the eigenvalues.

1.2.6 Significance to climate models

In principle, f_k, g_k and μ_k in equation (1.38) are observable (to some level of confidence) for the real climate, since they are determined by the natural variability of the real climate system. Hence equation (1.38) and (1.34) relates the natural variability of the real climate to the real climate's equilibrium response to a change in forcing. If the eigenvalues and eigenfunctions differ for two models then, by (1.38) there are *some* forcings for which the response differs.

If the change in forcing projects onto the gradient of the adjoint eigenfunctions equally, then, given the $1/\mu_k$ term in equation (1.38), the eigenfunctions with the lowest eigenvalues, and which correspond to the slowest modes of variability are the most significant. If $\Delta\mathbf{M}$ were calculable for the limited dimension state-space of a subset of observed real climate variables, and if the earth-atmosphere system is a diffusion process, then, to first order, and to the level of confidence in the observed estimates of f_k, g_k and μ_k, the real climate's equilibrium response to a change in

forcing could be estimated independently of a climate model.

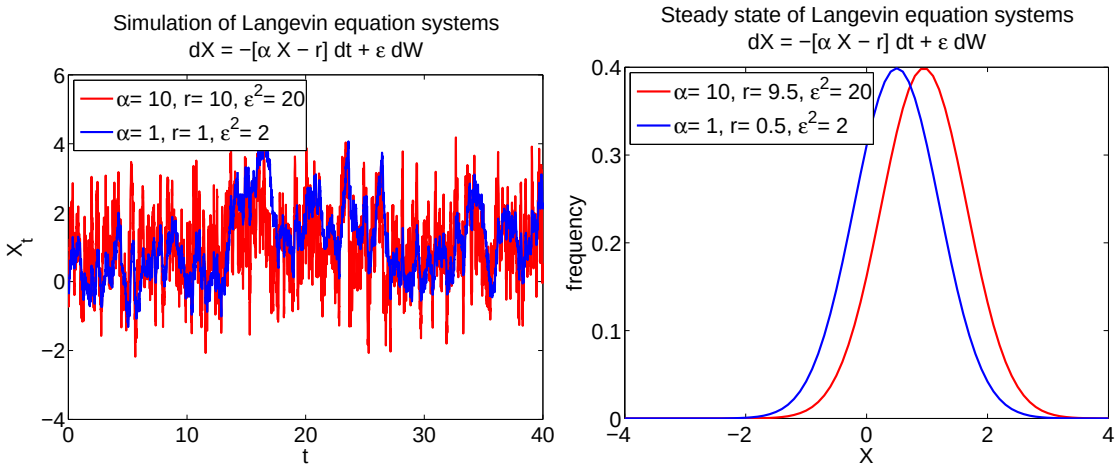
However, it is not possible, in general, to calculate $\Delta\mathbf{M}$ independently of a climate model, so the relation in (1.38) cannot be used, say, to estimate climate sensitivity. $\Delta\mathbf{M}$ is a function of the change in forcing, and not the change in forcing itself. There is necessarily a reduction in dimensions from the infinite-dimensional climate to a finite dimensional model, and the translation of the changing in forcing to a finite dimensional $\Delta\mathbf{M}$ is model-specific. This is illustrated in 3.6.1.

Formally, in (1.38), if the change in forcing is small in relation to the natural variability, then it takes the system into a region of state space that has already been observed, and where the processes that dictate the response to a small change in forcing are largely the same processes that cause natural variability [25]. But if the scale of the change in forcing is much larger than the scale of the natural variability, then the new forcing represents regions of state-space which have not been observed (and of which there is little information), so the natural variability does not give relevant information about the response to a change in forcing. Figure 1.3 on page 42 and figure 1.4 ilustrates the need for the change in forcing to be small in relation to the natural variability for a system whose exact response to forcing is known.

Part I of this thesis examines the practical problem of estimating the eigenfunctions and eigenvalues of the unforced real system. Since equation (1.38) requires an estimate of the transition density of a multi-dimensional system, whose slowest modes of variability are the most significant, there are three practical problems. First, the slowest modes require the longest periods of observation. Second, the more dimensions there are, the higher is the dimension of the transition density, and so the harder it is to estimate, requiring a yet longer time series. The available time series may not be long enough. The ERA-40 reanalysis [49] provides 45 years of a best-estimate of climate data, though reanalysis sets for the last 100 to 150 years are being developed [10]. Third, the forcing on the real climate (for which time series are available) is not constant, as assumed by (1.1). The real forcing forms a daily and seasonal cycle, and there are trends in greenhouse gas and solar

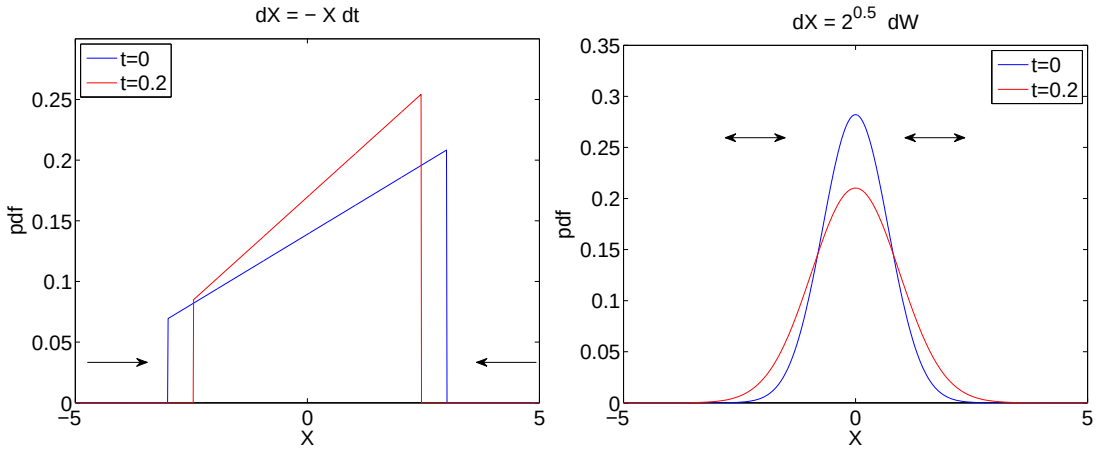
forcings, together with sporadic cooling from volcanoes.

Approaches to reducing the number of dimensions should make optimal use of limited data. Chapter 2 suggests a method for estimating the transition density and chapter 3 applies the method to two test cases. Chapter 4 applies the method to a comparison of HadCM3 and ERA-40.



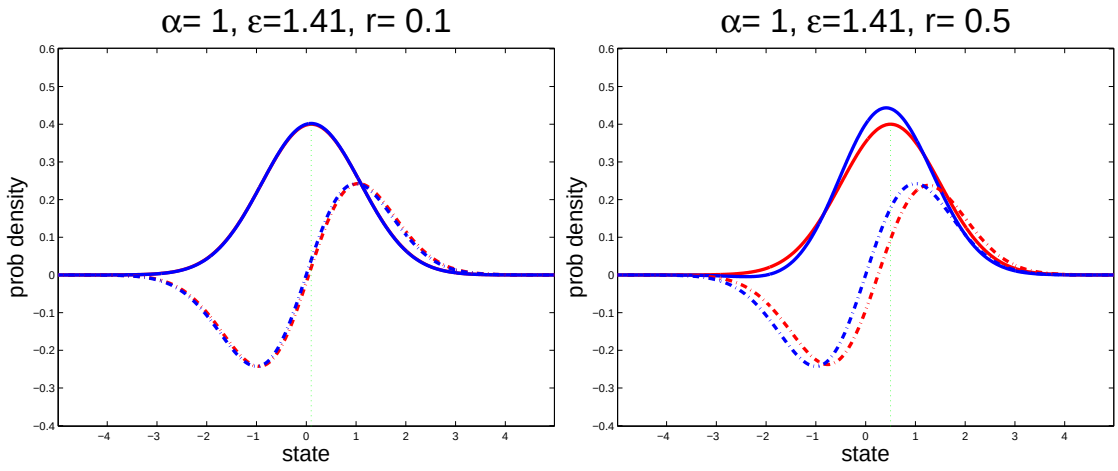
(a) (b)

Figure 1.1: This figure shows that two systems can have the same steady probability density function (pdf), but different steady-state responses to a change in forcing. The left graph shows time series for two stochastic dynamical systems of the form $dX = -(\alpha X - r)\,dt + \sqrt{2\alpha}\;dW$. *$X$ is the state variable and W is the state of a Wiener process, so that dW is gaussian white noise with variance dt over a timescale dt. Both systems have the same steady pdf (a gaussian with mean 1, variance 1) but have different characteristic timescales [43] equal to $1/\alpha$. Because of the different timescales, the two systems have different equilibrium responses to a change in forcing. This is shown in the right plot, where r has been reduced by 0.5 for both systems. The new blue equilibrium for the slow system is a gaussian with mean 0.5 and variance 1. But the red system is ten times as fast as the slow blue system. So, the fast red system's equilibrium mean is reduced by one tenth of the reduction to the slow blue system's equilibrium mean. If the slow system were a model and the fast system were reality, then no matter how small a change in forcing, the slow system would overestimate the real system's response to the change in forcing by a factor of 10.*



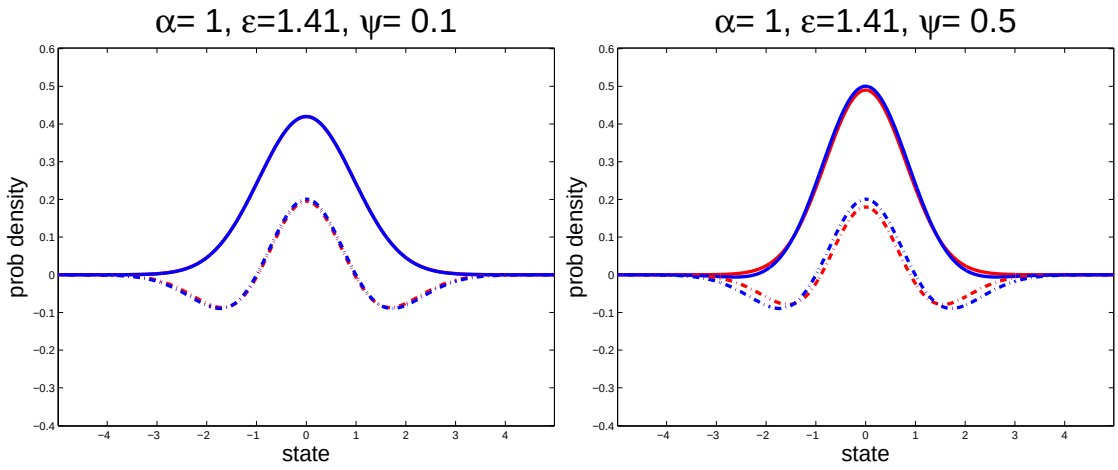
(a) $f(x,t) = e^t f(x\ e^t, 0)$ (b) $f(x,t) = 1/\sqrt{4\pi(1+4t)}e^{-x^2/(1+4t)}$

Figure 1.2: This figure shows the evolution of 2 pdfs under advection and diffusion. The left plot shows advection – there is only drift towards the origin, so the pdf is compressed towards the origin. The right plot shows only diffusion, which evens out the highs and lows i.e. the convexity in the pdf.



(a) (b)

Figure 1.3: Each graph shows, in solid red, the exact steady pdf when a constant forcing of $r\ dt$ is applied to the 1-dimensional Langevin equation $dX = -\alpha\ X\ dt + \epsilon\ dW$. The forced system evolves according $dX = -\alpha\ X\ dt + r\ dt + \epsilon\ dW$. The unforced steady pdf is a Gaussian density with zero mean and variance $\epsilon^2/2\alpha$. The forced steady pdf is also a Gaussian density with variance $\epsilon^2/2\alpha$ but with mean r/α. The best possible estimate via eigenfunctions of the forced steady pdf is shown in solid blue. The exact difference between the unforced and forced steady pdf is in the red broken line. The blue broken line shows the difference calculated using eigenfunctions. The differences are divided by the ratio of the change in mean to the unforced standard deviation, so that the accuracy of the eigenfunction method can be compared for different levels of forcing. As the theory predicts, the eigenfunction method is more accurate for small levels of forcing. In this case, 'small' means in relation to the size of the standard deviation of the unforced system.



(a) (b)

Figure 1.4: Each graph shows, in solid red, the exact steady pdf when a forcing of $-\psi\alpha\ x\ dt$ is applied to the 1-dimensional Langevin equation $dX = -\alpha\ X\ dt + \epsilon\ dW$. The forced system evolves according to $dX = -\alpha\ (1+\psi)\, X\ dt + \epsilon\ dW$. The unforced steady pdf is a Gaussian density with zero mean and variance $\epsilon^2/2\alpha$. The forced steady pdf is also a Gaussian density with zero mean and variance $\epsilon^2/2\alpha(1+\psi)$. The best possible estimate via eigenfunctions of the forced steady pdf is shown in solid blue. The exact difference between the unforced and forced steady pdf is in the red broken line. The blue broken line shows the difference calculated using eigenfunctions. The differences are divided by ψ, so that the accuracy of the eigenfunction method can be compared for different levels of forcing. As the theory predicts, the eigenfunction method is more accurate for small levels of forcing (the blue and red broken lines are closest for small values of ψ).

Chapter 2

Method of analysing time series

Chapter 1 established, to first order, a theoretical equivalence between the transition density of a diffusion process under constant forcing and the steady state response of that process to a small change in forcing. The transition density was expanded in terms of the eigenfunctions and eigenvalues of the Fokker-Planck equation of the process. This chapter approaches the practical problem of estimating the eigenfunctions and eigenvalues for reanalysis data and a climate model.

2.1 Estimating eigenfunctions and eigenvalues from a transition matrix

The eigenfunctions and eigenvalues of equation (1.2) may be estimated by discretising the state space into bins, and then calculating the rates of transition from each bin to each other bin, forming a transition matrix. The eigenvectors of the transition matrix approximate the integral of the eigenfunctions of the Fokker-Planck equation over the bins, and the log of the eigenvalues of the transition matrix approximate the eigenvalues of the Fokker-Planck equation. This is because the transition probabilities are a discretisation of the transition density, which is proven in this section.

2.1.1 Transition matrix is a function of eigenvalues and eigenfunctions

If the system's state space is divided into mutually exclusive and collectively exhaustive bins, then, over any fixed time period, there is a probability of transition from one bin to another. The system has been assumed to be a diffusion process, i.e. a homogeneous Markov process [24, page 34], so the transition probabilities are constant in time. If the (continuous) eigenfunctions of the Fokker-Planck equation are known, then the transition probabilities for the discretised system are known (in principle). This is shown in equation (2.6).

Let T_{ij} be the transition probability, i.e. the probability that the system moves from bin j to bin i over timestep Δt, given that the system starts off in bin j. If the system is in bin j, then, on average, its density within bin j is the steady pdf, restricted to bin j, i.e. if $h_j(\vec{x})$ is the average density within bin j then

$$h_j(\vec{x}) = \begin{cases} 0 & \vec{x} \notin \text{bin } j \\ \dfrac{f_0(\vec{x})}{\int_{\text{bin } j} f_0(\vec{x})\ \mathrm{d}\vec{x}} & x \in \text{bin } j. \end{cases} \tag{2.1}$$

$$\text{Then } T_{ij} = \text{Probability}((X_{t+\Delta t} \in \text{bin } i | \text{ pdf of } X_t \text{ is } h_j(\vec{x})) \tag{2.2}$$

$$= \int_{\text{bin } i} f(\vec{x}, t+\Delta t)\ \mathrm{d}\vec{x} \tag{2.3}$$

$$= \int_{\text{bin } i} \left(\sum_{k=0}^{\infty} b_k f_k(\vec{x}) e^{-\mu_k \Delta t} \right) \mathrm{d}\vec{x} \qquad \text{by (1.8)} \tag{2.4}$$

$$= \sum_{k=0}^{\infty} \underbrace{\int h_j(\vec{x}) g_k(\vec{x})\ \mathrm{d}\vec{x}}_{b_k} e^{-\mu_k \Delta t} \int_{\text{bin } i} f_k(\vec{x})\ \mathrm{d}\vec{x} \qquad \text{by (1.23)} \tag{2.5}$$

$$= \sum_{k=0}^{\infty} \frac{\int_{\text{bin } j} f_0(\vec{x}) g_k(\vec{x})\ \mathrm{d}\vec{x}}{\int_{\text{bin } j} f_0(\vec{x})\ \mathrm{d}\vec{x}} e^{-\mu_k \Delta t} \int_{\text{bin } i} f_k(\vec{x})\ \mathrm{d}\vec{x} \qquad \text{by (2.1).} \tag{2.6}$$

Note that $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$ by (2.3).

2.1.2 Eigenvalues and eigenfunctions are functions of transition matrix

The matrix of transition probabilities defines (up to discretization error) the eigenfunctions and eigenvalues of the Fokker-Planck equation. If the state space is divided into mutually exclusive and collectively exhaustive bins $1, 2, \ldots, n$, then, as the resolution of the bins gets finer, the eigenvectors and eigenvalues of the transition matrix become increasingly accurate estimates of the eigenfunctions and eigenvalues of the Fokker-Planck equation.

To see this, let $\vec{F}^k$ be a vector of integrals over each bin of the eigenfunction f_k and let $\vec{G}^l$ be a vector of the averages over each bin of the eigenfunctions of the adjoint g_l. That is,

$$\vec{F}^k_i = \int_{\text{bin } i} f_k(\vec{x})\ \mathrm{d}\vec{x}, \text{ and } \vec{G}^l_i = \frac{\int_{\text{bin } i} g_l(\vec{x}) f_0(\vec{x})\ \mathrm{d}\vec{x}}{\int_{\text{bin } i} f_0(\vec{x})\ \mathrm{d}\vec{x}}. \tag{2.7}$$

Combining (2.6) and (2.7),

$$T_{ij} = \sum_{k=0}^{\infty} \vec{G}^k_j\ \vec{F}^k_i\ e^{-\mu_k \Delta t}. \tag{2.8}$$

Since the state space is unbounded, some of the bins must be unbounded. As the volume of state space included in bounded bins increases, and the number of bounded bins $\to \infty$,

$$\vec{F}^k \text{ approaches a right eigenvector of the transition matrix } \mathbf{T}, \text{ and} \tag{2.9}$$

$$\vec{G}^l \text{ approaches a left eigenvector of the transition matrix } \mathbf{T}. \tag{2.10}$$

In fact, $\mathbf{T}$ is a discretization of equation (1.2), so that (2.9) simply says that

$$f_k(\vec{x}_i, t + \Delta t) = e^{-\mu_k \Delta t} f_k(\vec{x}_i, t), \text{ which is true by definition.} \tag{2.11}$$

To prove (2.9), partition state space into n+1 mutually exclusive and collectively exhaustive bins so that bins 1, 2, ..., n are bounded, and only bin $n + 1$ is unbounded. (Bin n+1 is probably not a connected set). Let ϵ_1 be an arbitrarily small positive number. By the boundary condition equation (1.5), for every k and every

l there is a partition such that the largest volume of a bounded bin is less than ϵ_1, and the integral of fg over the unbounded bin is less than ϵ_1, i.e.

$$\forall k, l, \ \left| \int_{\text{bin } n+1} f_k(\vec{x}) g_l(\vec{x}) \ \mathrm{d}\vec{x} \right| < \epsilon_1. \tag{2.12}$$

Then,

$$\lim_{\epsilon_1 \to 0} \frac{\left[\mathbf{T}\vec{F}^l\right]_i}{\vec{F}_i^l} = e^{-\mu_l \Delta t}. \tag{2.13}$$

Since ϵ_1 can be arbitrarily small, (2.13) proves (2.9).

Proof of (2.13)**:**

$$\left[\mathbf{T}\vec{F}^l\right]_i / \vec{F}_i^l = \sum_{q=1}^{n+1} T_{iq} \vec{F}_q^l / \vec{F}_i^l \tag{2.14}$$

$$= \sum_{q=1}^{n+1} \left(\sum_{k=0}^{\infty} \vec{G}_q^k \ \vec{F}_i^k \ e^{-\mu_k \Delta t} \right) \ \vec{F}_q^l / \vec{F}_i^l \text{ by (2.8)} \tag{2.15}$$

$$= \sum_{k=0}^{\infty} \left(\sum_{q=1}^{n+1} \vec{G}_q^k \vec{F}_q^l \right) \ e^{-\mu_k \Delta t} \vec{F}_i^k / \vec{F}_i^l. \tag{2.16}$$

$$\lim_{\epsilon_1 \to 0} \sum_{q=1}^{n+1} \vec{G}_q^k \vec{F}_q^l = \int f_k(\vec{x}) g_l(\vec{x}) \ \mathrm{d}\vec{x} = \delta_{k,l} \tag{2.17}$$

$$\Rightarrow \lim_{\epsilon_1 \to 0} \left[\mathbf{T}\vec{F}^l\right]_i / \vec{F}_i^l = \sum_{k=0}^{\infty} \delta_{k,l} \ e^{-\mu_k \Delta t} \vec{F}_i^k / \vec{F}_i^l = e^{-\mu_l \Delta t}. \tag{2.18}$$

Equation (2.13) shows that

- the right eigenvectors of the transition matrix tend to integrals of the eigenfunctions of the Fokker-Planck equation, where the integrals are taken over the bins that discretize state space;
- the log of the eigenvalues of the transition matrix, divided by the time step of the transition, tend to the eigenvalues of the Fokker-Planck equation.

By a similar argument, the left eigenvectors of the transition matrix tend to the eigenfunctions of the adjoint of the Fokker-Planck equation.

2.1.3 Sampling error weakens results

Intuitively, the more that sampled data is dominated by measurement error, the less is truly known about the system, and the less confident are any model forecasts. Consistent with this, the more measurement error there is in the data, the more equation (1.38) underestimates the response to a change in forcing.

The underestimate is due to the fact that as the measurement error increases, the probability of landing in any bin becomes increasingly independent of the bin the system started in. The columns of the transition matrix tend to become identical (and to be in proportion to the equilibrium density). Thus the eigenvalues of the transition matrix tend to zero (for all eigenvectors apart from that of the steady pdf), so the estimated eigenvalues μ_k in (1.38) tend to infinity, and the estimated Δf tends to zero.

Proof For let $\mathbf{T} = \left(\vec{t}\,\vec{t}\,\ldots\,\vec{t}\right)$ be an n x n transition matrix with identical columns $\vec{t} = (t_1, t_2, \ldots, t_n) : t_1 + t_2 + \ldots + t_n = 1$.

Then

$$\mathbf{T}\vec{t} = \begin{pmatrix} t_1\,(t_1 + \ldots + t_n) \\ t_2\,(t_1 + \ldots + t_n) \\ \vdots \\ t_n\,(t_1 + \ldots + t_n) \end{pmatrix} = \begin{pmatrix} t_1.1 \\ t_2.1 \\ \vdots \\ t_n.1 \end{pmatrix} = \vec{t}. \tag{2.19}$$

Let $\vec{e}_i$ be a basis vector $(\delta_{i,1}, \delta_{i,2}, \ldots, \delta_{i,n})$.

Then

$$\mathbf{T}\vec{e}_i = \vec{t} \Rightarrow \mathbf{T}(\vec{e}_i - \vec{e}_j) = \vec{0}. \tag{2.20}$$

$\{(\vec{e}_1 - \vec{e}_j), j = 2, 3, \ldots, n\}$ is a linearly independent set of $n - 1$ vectors, since for any non-zero scalars $\omega_2, \ldots, \omega_n$,

$$\sum_{j=2}^{n} \omega_j(\vec{e}_1 - \vec{e}_j) = \begin{pmatrix} \sum_{j=2}^{n} \omega_j \\ -\omega_2 \\ \vdots \\ -\omega_n \end{pmatrix} \neq \vec{0}. \tag{2.21}$$

Also $\{\vec{t}, (\vec{e}_1 - \vec{e}_j), j = 2, 3, \ldots, n\}$ is linearly independent, since if $\vec{t}$ were a sum of scalar multiples of $(\vec{e}_1 - \vec{e}_j)$, then by (2.20) $\mathbf{T}\vec{t}$ would be $\vec{0}$. But $\mathbf{T}\vec{t} = \vec{t} \neq \vec{0}$.

Thus $\{\vec{t}, (\vec{e}_1 - \vec{e}_j), j = 2, 3, \ldots, n\}$ is a spanning set of n linearly independent eigenvectors, and the eigenvalues of $\mathbf{T}$ are 1 (with multiplicity 1) and 0 (with multiplicity $n - 1$).

2.2 Getting the transition matrix

Two methods of calculating the transition matrices are shown below: bin counting, and kernel density estimation. The methods are applied to test cases in chapter 3.

2.2.1 Bin-counting: getting the transition matrix by counting transitions

A simple way to estimate the transition matrix is to partition the state space into bins, sample the time series at regular intervals, and count the proportion of times that the system changes from each bin to each other [15]. This approach is valid, assuming that the diffusion process in (1.1) is ergodic [24, page 38]. For, by the definition (2.2) of the elements of the transition matrix:

$$T_{ij} = \text{Probability}\left(\vec{X}_{t+\Delta t} \in \text{bin } i | \vec{X}_t \text{ is distributed as per } h_j(\vec{x})\right). \quad (2.22)$$

Let $\mathbf{1}\,()$ be the indicator function for a subset $\mathcal{S}$ of the state-space:

$$\mathbf{1}_{\mathcal{S}}(\vec{x}) = \begin{cases} 1 & \vec{x} \in \mathcal{S} \\ 0 & \vec{x} \notin \mathcal{S}, \text{and} \end{cases} \quad (2.23)$$

$$\mathbf{1}_{\mathcal{S}_1,\mathcal{S}_2}(\vec{x}, \vec{y}) = \mathbf{1}_{\mathcal{S}_1}(\vec{x})\,\mathbf{1}_{\mathcal{S}_2}(\vec{y}). \quad (2.24)$$

Hence, by the definition of conditional probability and (2.22),

$$T_{ij} = \frac{\text{Probability}\left(\vec{X}_{t+\Delta t} \in \text{bin } i \text{ and } \vec{X}_t \text{ is distributed as per } h_j(\vec{x})\right)}{\text{Probability}\left(\vec{X}_t \text{ is distributed as per } h_j(\vec{x})\right)} \tag{2.25}$$

$$= \frac{\mathbb{E}\left(\mathbf{1}_{\text{bin } i}\left(\vec{X}_{t+\Delta t}\right)\ \mathbf{1}_{\text{bin } j}\left(\vec{X}_t\right)\right)}{\mathbb{E}\left(\mathbf{1}_{\text{bin } j}\left(\vec{X}_t\right)\right)}. \tag{2.26}$$

The ergodic assumption [24, page 38] is that a function of many samples from a long time series tends to the expectation of that function over the steady pdf. The discrete time ergodic assumption is that, where u is any function and N samples are taken, each Δt apart in time, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{p=1}^{N} u\left(\vec{X}_{p\Delta t}\right) = \mathbb{E}\left(u\left(\vec{X}\right)\right) \tag{2.27}$$

So, assuming that the diffusion process is ergodic,

$$T_{ij} = \lim_{N\to\infty} \frac{\cancel{1/N}\ \sum_{p=1}^{N-1} \mathbf{1}_{\text{bin } i}\left(\vec{X}_{(p+1)\Delta t}\right)\ \mathbf{1}_{\text{bin } j}\left(\vec{X}_{p\Delta t}\right)}{\cancel{1/N}\ \sum_{p=1}^{N-1} \mathbf{1}_{\text{bin } j}\left(\vec{X}_{p\Delta t}\right)}, \tag{2.28}$$

which is the relative frequency of transitions from bin j to bin i.

Equally likely bins may minimise the variance of the estimated eigenvalues

Ideally, the estimated eigenvalues will be insensitive to 'noise' or sampling error, but sensitive to 'signal' i.e. to real differences in two underlying processes. Equivalently, the ideal approach will minimise the variance of estimated eigenvalues, given several series from the same underlying process, and minimise the bias, i.e. the average difference of the estimated eigenvalues from the true eigenvalues.

The variance of the estimated eigenvalues may be reduced if we choose to use bins which are equally likely, i.e. if the state is sampled in each bin equally frequently. With equally likely bins, the variance of the eigenvalues of the transition matrix appears to be approximately inversely proportional to the number of samples (based on empirical evidence shown in 3.4.3).

Bias in estimating the eigenvalues is reduced by increasing the number of bins. This is shown in section 3.4. If there are too few bins, then eigenfunctions with a high eigenvalue are conflated with eigenfunctions with a low eigenvalue, so the sampled eigenvalue is an overestimate.

Choices of bins for transition matrix obtained by counting transitions

The more bins that are used, the worse each bin is sampled. This introduces noise or sampling error into the estimated eigenfunctions. A trade-off is required between noise and resolution bias. This is illustrated in figure 2.1.

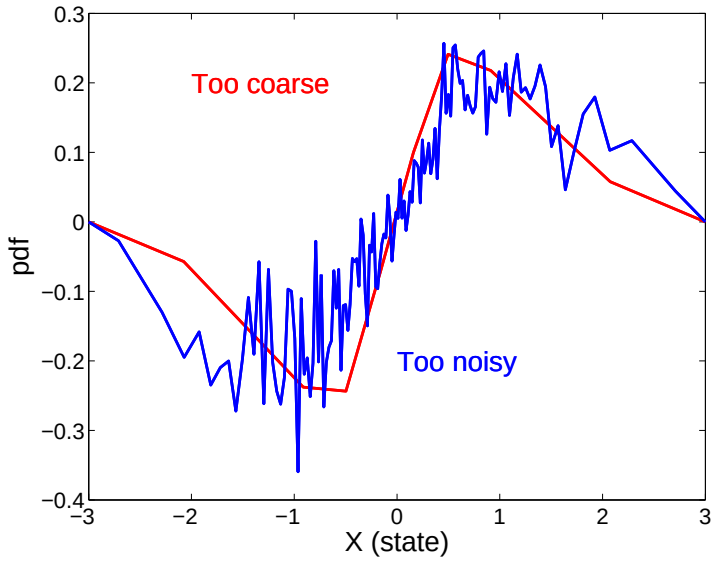


Figure 2.1: Estimate of second eigenfunction for the system $dX = -X\,dt + \sqrt{2}dW$ 100 seconds, sampled at 10 Hz split into 130 (blue line) and 10 (red line) equally probable bins. The exact eigenfunction is the function x weighted by a Gaussian density.

2.2.2 Kernel density estimation of transition density

A generalisation of the bin-counting method is the kernel density estimation method [40]. This method estimates the transition density directly as a continuous function. The 'kernel' is a weighting function $w()$ applied directly to each sample. It estimates

the joint density $f_{X,Y}$ as

$$f_{X,Y}(\vec{x},\vec{y}) = \sum_{p=1}^{N-1} \hat{w}_p(\vec{x},\vec{y}) \tag{2.29}$$

and the steady pdf $f_0(\vec{x})$ as

$$f_0(\vec{x}) = \sum_{p=1}^{N-1} w_p(\vec{x}). \tag{2.30}$$

The weighting function $w_p(\vec{x})$ is defined on the state space and takes maximum value at $\vec{x}_p$, where $\vec{x}_p$ is the p^{th} sample. The joint weighting function $\hat{w}_p(\vec{x},\vec{y})$ takes maximum value at $(\vec{x}_p,\vec{x}_{p+1})$, that is at the point in transition-space where the state $\vec{x}_p$ changes to the state $\vec{x}_{p+1}$. If bins are then chosen, the estimated probability of transition from bin j to bin i is

$$\widehat{T_{ij}} = \frac{\int_{\text{bin } i}\int_{\text{bin } j} f_{X,Y}(\vec{x},\vec{y})\ \mathrm{d}\vec{x}\ \mathrm{d}\vec{y}}{\int_{\text{bin } j} f_0(\vec{x})\ \mathrm{d}\vec{x}} \tag{2.31}$$

$$= \frac{\sum_{p=1}^{N-1}\int_{\text{bin } i}\int_{\text{bin } j} \hat{w}_p(\vec{x},\vec{y})\ \mathrm{d}\vec{x}\ \mathrm{d}\vec{y}}{\sum_{p=1}^{N-1}\int_{\text{bin } j} w_p(\vec{x})\ \mathrm{d}\vec{x}}. \tag{2.32}$$

Bin-counting is equivalent to using weighting functions which are delta functions, centred on the samples. Such delta functions give the same result as indicator functions, and so (2.32) gives the same result for bin-counting as (2.28).

If the weighting function is non-negative everywhere and integrates to one, then the kernel is equivalent to treating each sampled value as a true value plus a random sampling error, where the sampling error is independent and identically distributed for each sample. From the argument in section 2.1.3, any smoothing tends to overestimate the eigenvalues. This is is confirmed in figure 3.9 on page 73.

Optimal bandwidth

The bandwidth of a kernel is a measure of the amount of smoothing it performs. For example, a Gaussian kernel with bandwidth η is a weighting function

$$w_p(x) \propto \frac{1}{\eta\sqrt{2\pi}} e^{-1/2(x/\eta)^2}. \tag{2.33}$$

Silverman [40] defines an integrated square error [40, §3.1.2] to measure the total difference between an estimated and exact eigenfunction, and states that the bandwidth which minimises this error for a 1-dimensional density based on N independent, normally distributed samples with variance σ^2 is

$$\eta = \sigma(4/3N)^{1/5} \approx 1.06\,\sigma\,N^{-1/5}. \tag{2.34}$$

If the samples are not independent, then N is overestimated in equation (2.34) and so the optimal bandwidth will be higher.

Silverman [40] also states [40, Table 4.1 on p.87] that if a d-dimensional multi-variate normal kernel is used (for multi-variate normally distributed samples with covariance matrix $\sigma^2\mathbf{I}$) then the optimal bandwidth is

$$\eta = \sigma\left(\frac{4}{N\,(d+2)}\right)^{1/(d+4)}. \tag{2.35}$$

2.3 Detecting differences in transition matrices

The purpose of estimating the transition matrices is to detect significant differences in their eigenvalues and eigenvectors. This requires a measure of the differences, and a measure of significance, since differences may emerge by chance, because the transition matrices are random variables.

2.3.1 Measure of difference in eigenvectors

The measure used is the integrated square error over state space. Given bins $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n$, and eigenvectors $\vec{V}$ and $\vec{W}$, which are computed using the same bins over state space, the eigenvectors are normalised to have the same absolute sum over state space, and then the sum of their squared differences is taken.

$$\sum_{i=1}^{n} \left\| \frac{\vec{V}_i}{\sum_{i=1}^{n} \left|\vec{V}_i\right|} - \frac{\vec{W}_i}{\sum_{i=1}^{n} \left|\vec{W}_i\right|} \right\|^2 \frac{1}{\mathrm{Vol}\,(\mathcal{I}_i)} \tag{2.36}$$

In order to avoid recording a difference if all that separates two eigenvectors is a factor of -1, the difference measure $\mathrm{d}(\vec{V}, \vec{W})$ is

$$\min\left(\sum_{i=1}^{n}\frac{\left\|\frac{\vec{V}_i}{\sum_{i=1}^{n}|\vec{V}_i|}-\frac{\vec{W}_i}{\sum_{i=1}^{n}|\vec{W}_i|}\right\|^2}{\mathrm{Vol}\left(\mathcal{I}_i\right)},\quad\sum_{i=1}^{n}\frac{\left\|\frac{\vec{V}_i}{\sum_{i=1}^{n}|\vec{V}_i|}+\frac{\vec{W}_i}{\sum_{i=1}^{n}|\vec{W}_i|}\right\|^2}{\mathrm{Vol}\left(\mathcal{I}_i\right)}\right). \tag{2.37}$$

2.3.2 Significance of difference in eigenvectors

The distribution of the differences in sampled eigenvectors or eigenvalues is not known, but is estimated by taking several simulations of identical underlying processes, calculating their mean eigenvector (or eigenvalue) and observing the spread of sampled differences from this mean. A significant difference is one which is outside the range of the sampled differences for identical processes.

2.4 Summary

This chapter described a method of comparing the eigenfunctions and eigenvalues of the Fokker-Planck equation for the transition densities of the real climate and a climate model. The method compares the eigenvectors and eigenvalues of transition matrices based on time series for the model and reanalysis data. The transition matrices could be constructed by first defining bins and then counting the transitions between bins, or by first estimating the density according to some smoothing function and then defining bins over which to integrate the density. An optimal level of smoothing was suggested. This smoothing level is tested in chapter 3 by applying it to a known system, and measuring the integrated square error (2.36) between the known eigenfunctions and the sampled eigenfunctions.

Chapter 3

Test Cases: linear Langevin equations

In this chapter, the theory in chapter 1 is verified for a diffusion process (of Brownian motion in one dimension) whose transition density is known analytically. The methods in chapter 2 are applied to numerical simulations of Brownian motion, and to a multi-dimensional linear diffusion process, in order to investigate:

- what bandwidth should be used to estimate the dominant eigenfunctions;
- what bandwidth should be used to estimate the dominant eigenvalues;
- how the variance of the eigenvalues scales with the length of the series;
- what structure of bins should be used to reduce the variance of the sampled eigenvalues;
- what the effect of adding dimensions is on the detectability of differences between models.

3.1 Test model: Brownian motion

A Brownian motion model (whose solution is known analytically) is used to compare estimated and exact eigenfunctions and eigenvalues. The model may be considered to be a simple energy-balance model, as in [50], where one dimensional

global mean temperature change (X) responds linearly to random white-noise radiative forcing $\epsilon \mathrm{d}W$ at a rate α determined by the thermal inertia of the oceans. So, where α, ϵ are scalar constants,

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \epsilon \mathrm{d}W_t. \tag{3.1}$$

Equation (3.1) is the mixed-layer ocean model of equation (1) of [50] (i.e. with the ΔF term in that equation set to zero). Equation (3.1) is also used to model the velocity X of a dust particle on the surface of water: subject to a braking force αX proportional to its speed, and subject to random perturbations $\epsilon \mathrm{d}W$. The same equation models these two systems because, as pointed out by [22], the climate integrates the many small scale, short-term effects of weather similarly to how the dust particle integrates the random effect of the many smaller particles that make up the braking force.

3.1.1 Exact eigenfunctions and eigenvalues

1. The eigenvalues μ_k of the Fokker-Planck equation for the pdf of the state of the process in (3.1) are integer multiples of α;

2. the eigenfunctions $f_k(x)$ of the Fokker-Planck equation are constant multiples of Hermite polynomials weighted by a Gaussian distribution centred on 0, with variance $\epsilon^2/(2\alpha)$;

3. the eigenfunctions of the adjoint $g_k(x)$ are proportional to the Hermite polynomials.

In other words, let

$$\xi = x\sqrt{\frac{2\alpha}{\epsilon^2}}, \tag{3.2}$$

$$\text{and } \phi_k(\xi) = e^{\xi^2/2}\frac{\partial^k}{\partial \xi^k}\left(e^{-\xi^2/2}\right). \tag{3.3}$$

$$\text{So, for example, } \phi_0(\xi) = 1, \tag{3.4}$$

$$\phi_1(\xi) = -\xi, \tag{3.5}$$

$$\phi_2(\xi) = \xi^2 - 1. \tag{3.6}$$

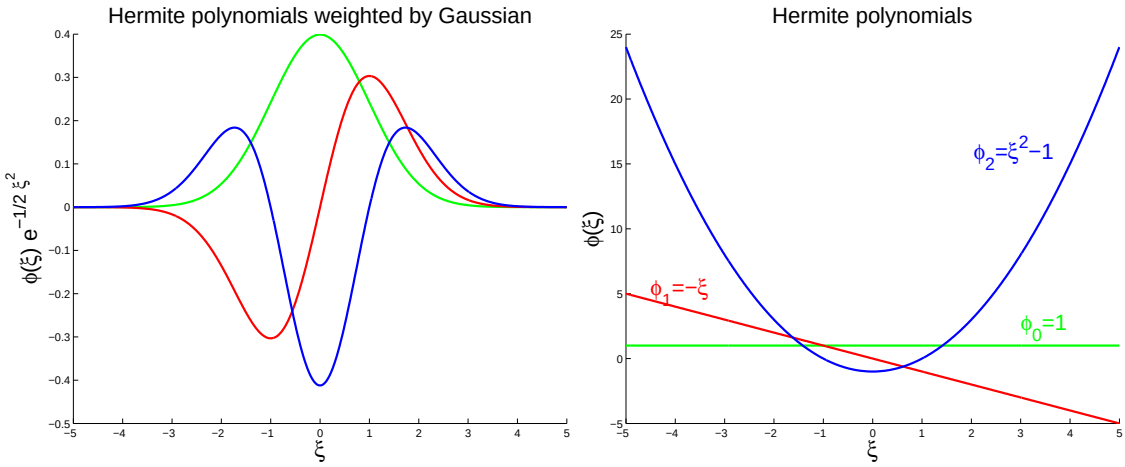
Then, using the notation of (1.38) for the Fokker-Planck equation corresponding to (3.1)

$$\mu_k = k\alpha, \ k = 0, 1, 2, \ldots \tag{3.7}$$

$$f_k(x) \propto \ \phi_k(\xi) e^{-\xi^2/2}, \tag{3.8}$$

$$g_k(x) \propto \ \phi_k. \tag{3.9}$$

The leading eigenfunctions are pictured in figure 3.1.



(a) Eigenfunctions of FPE

(b) Eigenfunctions of adjoint

Figure 3.1: $dX = -Xdt + \sqrt{2}dW$; first three eigenfunctions of Fokker-Planck equation for Langevin model in (3.1).

Proof of (3.7) **-** (3.9)

By substituting (3.1) into (1.2), the Fokker-Planck equation for the evolution of the pdf of the system is

$$\frac{\partial f}{\partial t} = \frac{\partial(\alpha x f)}{\partial x} + \frac{\partial^2 (Kf)}{\partial x^2} \tag{3.10}$$

$$\text{where } K = \frac{\epsilon^2}{2}. \tag{3.11}$$

Suppose that equation (3.10) is satisfied by solutions of the form

$$e^{-\mu t} f_k(x) = e^{-\mu t} \phi(\xi) e^{-\frac{\xi^2}{2}}, \tag{3.12}$$

By substituting $e^{-\mu t} f_k(x)$ for f in (3.10) and dividing by $e^{-\mu t - \frac{\xi^2}{2}}$

$$\phi'' - \xi \phi' + \frac{\mu}{\alpha} \phi = 0. \tag{3.13}$$

Equation (3.13) is the probabilistic form of Hermite's equation, so ϕ is a Hermite polynomial and μ/α is an integer [1, pp 773–785].

Similarly by substituting (3.1) into (1.15), the adjoint of the Fokker-Planck equation is

$$-\nu g = -\alpha x \frac{\partial g}{\partial x} + K \, \frac{\partial^2 g}{\partial x^2}. \tag{3.14}$$

By substituting $\phi(\xi)$ for g in (3.14)

$$\phi'' - \xi \phi' + \frac{\nu}{\alpha} \phi = 0, \text{ which is (3.13) with } \nu = \mu. \tag{3.15}$$

Normalisation of eigenfunctions

$f_k(x)$ and $g_k(x)$ may be normalised according to (1.21) so that

$$f_k(x) = \frac{1}{\sigma\sqrt{2\pi}} \, \phi_k(\xi) e^{-\xi^2/2}, \text{ and} \tag{3.16}$$

$$g_k(x) = \gamma_k \, \phi_k(\xi), \text{ where} \tag{3.17}$$

$$1/\gamma_k = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi_k^2 \, e^{-\xi^2/2} \, \mathrm{d}\xi, \text{ and} \tag{3.18}$$

$$\sigma^2 = \epsilon^2 / 2\alpha. \tag{3.19}$$

3.2 Verification of theory

By substituting $e^{-\frac{\xi^2}{2}}$ for f in (3.10), the steady pdf of the process in (3.1) is a Gaussian, with zero mean and variance σ^2. Under a constant change in forcing, or a change in α, the steady pdf is still a Gaussian, though with a shifted mean or a change in variance. The change in the steady pdf is directly calculated in this section, and verifies (1.38).

3.2.1 Direct calculation of response of 1-d Langevin equation to constant forcing

If a constant r is added to the drift then (3.1) becomes

$$\mathrm{d}Y = (-\alpha Y + r)\ \mathrm{d}t + \epsilon \mathrm{d}W. \tag{3.20}$$

By substituting $X = Y - r/\alpha$ into (3.20), the steady distribution of Y is a Gaussian, with mean r/α and variance σ^2, illustrated in figure 1.1(a). So the steady pdf of Y is

$$\begin{aligned}
f_0 + \Delta f &= \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\ \left(\frac{x-r/\alpha}{\sigma}\right)^2} && (3.21)\\
&= \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\ \left(\frac{x}{\sigma}\right)^2} e^{\left(\frac{xr/\alpha}{\sigma^2} + O\left(r^2\right)\right)} && (3.22)\\
&= \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\xi^2} \left(1 + \frac{r/\alpha}{\sigma}\xi + O\left(r^2\right)\right) && (3.23)\\
&= f_0(x) - \frac{r/\alpha}{\sigma} f_1(x) + O\left(r^2\right). && (3.24)
\end{aligned}$$

3.2.2 Direct calculation of response of 1-d Langevin equation to forcing proportional to state variable

If a constant $\psi\alpha$ is added to the drift parameter α then the steady distribution of (3.1) changes from $N(0, \sigma^2)$ to $N(0, \frac{\sigma^2}{1+\psi})$, illustrated in figure 1.1(b). So, as above, the altered steady pdf is

$$\begin{aligned}
f_0 + \Delta f &= \frac{\sqrt{1+\psi}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2(1+\psi)}, && (3.25)\\
&= \left(1 + \frac{\psi}{2} + O(\psi^2)\right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2} \left(1 - \frac{\psi\xi^2}{2} + O(\psi^2)\right) && (3.26)\\
&= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2} \left(1 + \frac{\psi}{2} - \frac{\psi\xi^2}{2}\right) + O(\psi^2) && (3.27)\\
&= f_0 + \frac{-\psi}{2}\left(\xi^2 - 1\right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2} + O(\psi^2) && (3.28)\\
&= f_0 + \frac{-\psi}{2} f_2(x) + O(\psi^2). && (3.29)
\end{aligned}$$

3.2.3 Equilibrium response to forcing by eigenfunction is next eigenfunction

If the drift in (3.1) is changed in proportion to a hermite polynomial, i.e. if $\Delta\mathbf{M} = r\ \phi_l(\xi)$ where r is a small scalar constant, then (1.38) implies that, to first order, the steady state response is proportional to the $(l+1)^{th}$ eigenfunction, i.e. the $(l+1)^{th}$ Hermite polynomial weighted by the Gaussian density.

For (1.38) gives the change in steady pdf as

$$\Delta f = \sum_{k=1}^{\infty} a_k f_k(x) \tag{3.30}$$

$$\text{where } a_k = -\frac{1}{k\alpha}\int g_k(x)\frac{\partial\left(f_0(x)\ \Delta\mathbf{M}\right)}{\partial x}\ \mathrm{d}x \tag{3.31}$$

$$= -\frac{1}{k\alpha}\int \gamma_k\phi_k(\xi)\frac{\partial}{\partial\xi}\left(\frac{e^{-\xi^2/2}}{\sigma\sqrt{2\pi}}\ r\phi_l(\xi)\right)\ \mathrm{d}\xi \tag{3.32}$$

$$= -\frac{r\gamma_k}{k\alpha\sigma}\int \frac{\phi_k}{\sqrt{2\pi}}\frac{\partial}{\partial\xi}\left(\frac{\partial^l}{\partial\xi^l}\left(e^{-\xi^2/2}\right)\right)\ \mathrm{d}\xi \tag{3.33}$$

$$= -\frac{r\gamma_k}{k\alpha\sigma}\int \frac{\phi_k\ \phi_{l+1}}{\sqrt{2\pi}}\ e^{-\xi^2/2}\ \mathrm{d}\xi \tag{3.34}$$

$$= -\frac{r}{(l+1)\,\alpha\sigma}\ \delta_{k,l+1} \text{ by orthonormality of } \phi. \tag{3.35}$$

$$\therefore \Delta f = -\frac{r}{(l+1)\,\alpha\sigma}f_{l+1}(x), \text{ agreeing with (3.24) and (3.29).} \tag{3.36}$$

3.3 Estimating eigenfunctions via transition matrices

To test the method in chapter 2 for estimating eigenfunctions and eigenvalues, time series of the process in (3.1) are generated, and then the eigenfunctions deduced from their transition matrices are compared with the known eigenfunctions in (3.8).

3.3.1 Generated time series

Figure 3.2 on the next page shows a series generated by discretising (3.1) in time with a predictor-corrector method, though similar results are obtained with a simple Euler method. On the right of figure 3.2, the sample auto-correlation function

for the series fits the theoretical auto-correlation function, which is $e^{-\alpha t}$ [24, page 44]. This confirms that the series is generated sensibly.

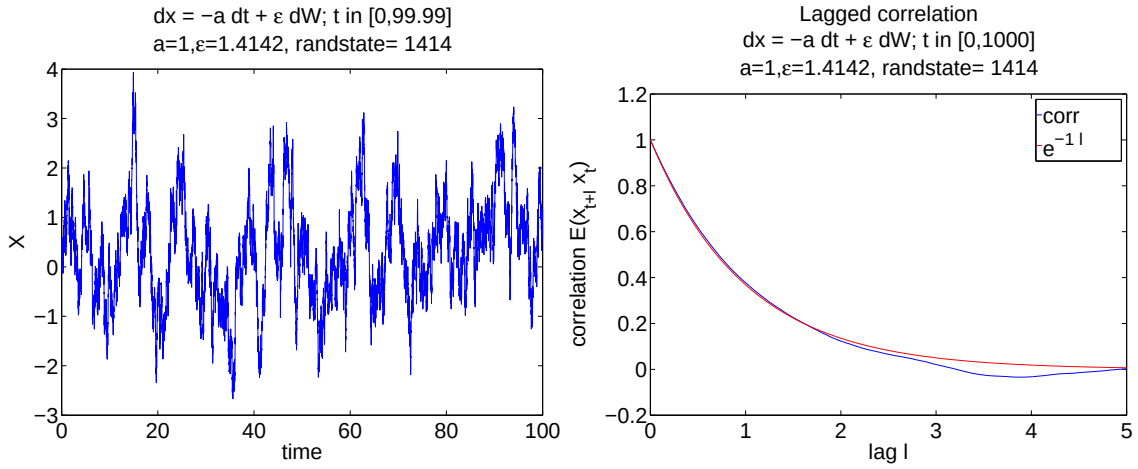
The series is generated, given initial value x_0 and a timestep Δt, for each time N, by letting

$$r_N = \epsilon \ \rho_N \ \sqrt{\Delta t}, \tag{3.37}$$

where ρ_N is a pseudo-random sample from a standard normal distribution.

$$\text{predictor} \ = x_{N-1} - \alpha x_{N-1} \Delta t + r_N; \tag{3.38}$$

$$x_N = x_{N-1} - \alpha \frac{x_{N-1} + \text{predictor}}{2} \Delta t + r_N. \tag{3.39}$$



(a) Series

(b) autocorrelation

Figure 3.2: dX = - X dt + $\sqrt{2}$ dW. Series and sample autocorrelation function (blue line). The sample autocorrelation fits the theoretical autocorrelation function (red line) which is e^{-l}. The series was generated with a time step of 0.0001 time units but was sampled only every 0.01 time units i.e. at 100 Hz.

3.3.2 Estimated eigenfunctions

Figures 3.3–3.5 show the integrated square error when using a Gaussian kernel density to estimate the first 3 eigenfunctions for the system in (3.1) with $\alpha = 1$, $\epsilon =$

$\sqrt{2}$. The estimates are based on a single series of 100 time units sampled at 10 Hz, generated using (3.39). (To put this in the context of the available climate data, ERA-40 contains 45 years of at least daily reanalysis output. For processes like El Niño this represents about 30 multiples of the process's characteristic timescale.) The estimated transition density is integrated over various numbers of bins to form a transition matrix. The bins are equally sized, mutually exclusive intervals whose union is [-5,5] with $(\infty, 5]$ and $[5, \infty)$ as semi-infinite bins. The eigenvectors of the transition matrices are linearly interpolated to estimate the eigenfunctions of the Fokker-Planck equation for (3.1). The estimated eigenfunctions are compared with the known eigenfunctions from (3.8) in order to see what level of smoothing gives the best estimate of the true eigenfunctions.

Figure 3.3 shows that where there is too little smoothing, then after a point, adding more bins increases the error (as noise dominates the signal). Figure 3.5 shows that where there is too much smoothing, adding more bins reduces the error (by an insignificant amount, and only because the smoothing dominates both the noise and the signal). Figure 3.4 shows a better level of smoothing, with lower errors for most eigenfunctions. Since these observations are only relevant to one time series, the mean error is shown in figure 3.6 for 2000 independent time series at various levels of smoothing. The optimal bandwidth is higher for higher eigenfunctions (which are progressively worse estimated). The sample mean integrated square error is minimised for the steady state pdf when the bandwidth is approximately 0.4.

Figure 3.7 shows the sample mean and variance of 2000 series, each of samples 10 times as long as those in figure 3.6. For the longer series, the optimal bandwidth was approximately 0.2, compared with the bandwidth recommended by (2.34) of $1.06 * 10000^{-1/5} = 0.17$. The sample mean integrated square error (ISE) and variance appeared (based on only two sample lengths) to scale inversely with sample length, as the sample mean ISE and variance for the longer series was approximately one tenth of the sample mean ISE and variance of the shorter series.

The optimal bandwidths are higher than the estimates provided by (2.34) be-

cause (2.34) is for independent samples, but the samples in the time series are not independent. Each sample depends explicitly on the previous sample. However, the optimal bandwidths are close enough to the estimate in (2.34) that (2.34) can be used as a preferred bandwidth for smoothing unknown eigenfunctions.

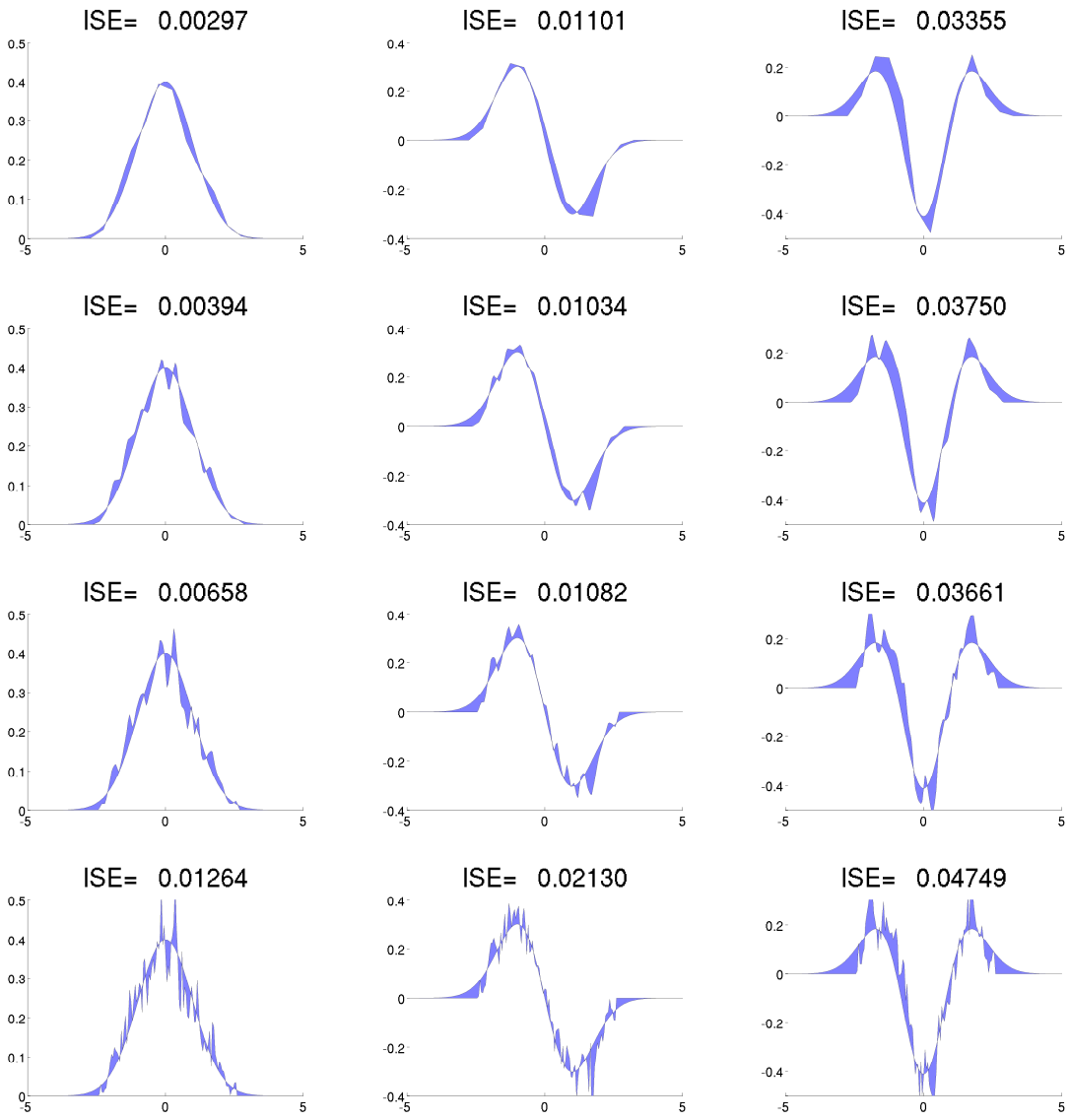


Figure 3.3: Unsmoothed eigenfunctions; first three eigenfunctions (going across the page) of the Fokker-Planck equation for a 1-dimensional Langevin equation. The time series is a simulation of $dX = -X\,dt + \sqrt{2}dW$*, sampled at 10Hz for a series 100 s long. The smoothing bandwidth is* 10^{-5}*, which is next to no smoothing. The state space is split into 20, 40, 80, 160 (increasing down the page) equally spaced bins on [-5,5]. The absolute error between the exact and sampled eigenfunction is shaded. Both the exact and the sampled eigenfunctions are normalised to have absolute area of one.*

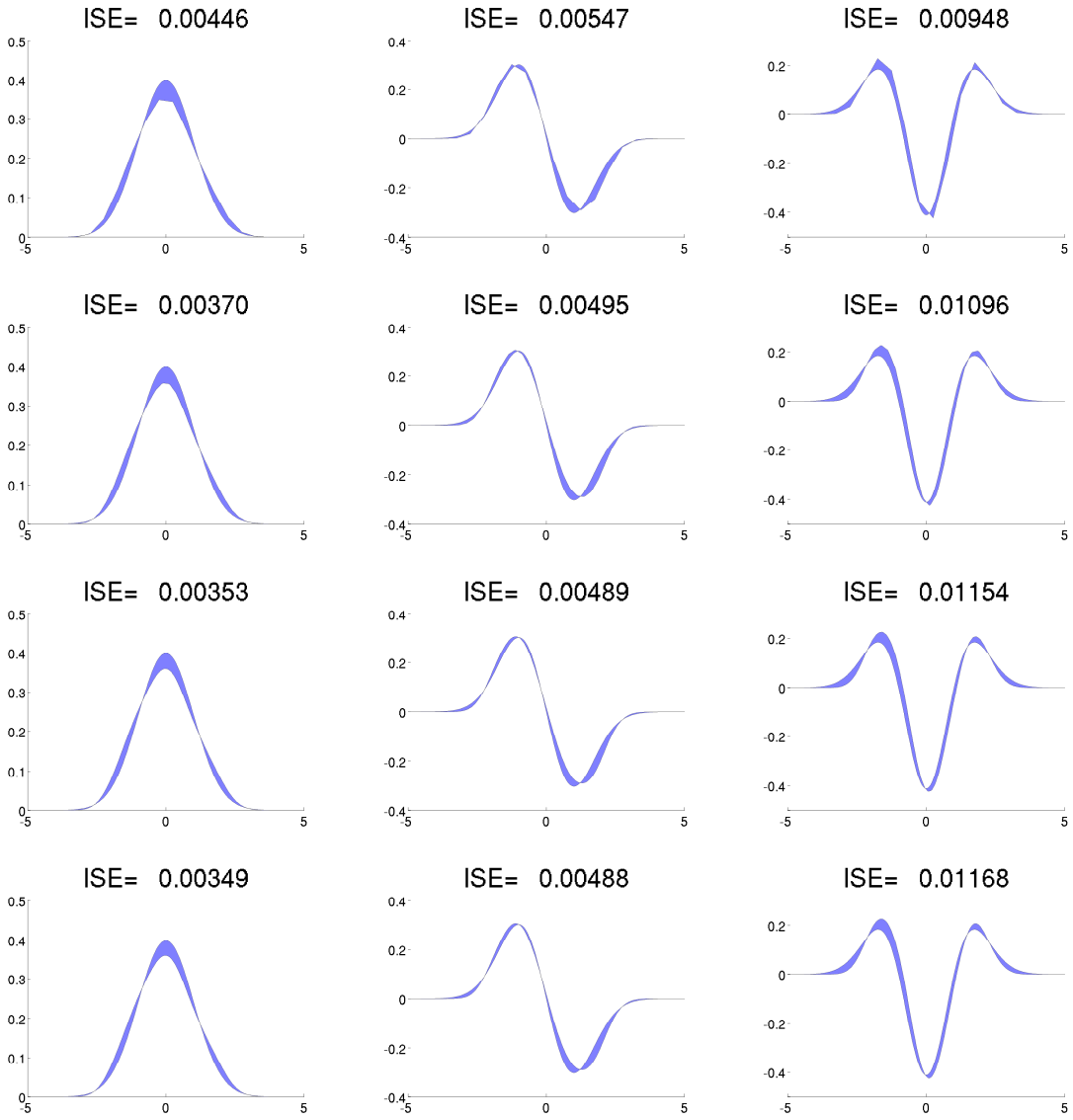


Figure 3.4: Optimally smoothed eigenfunctions; first three eigenfunctions (going across the page) of the Fokker-Planck equation for a 1-dimensional Langevin equation. The time series is a simulation of $dX = -X\,dt + \sqrt{2}dW$*, sampled at 10Hz for a series 100 s long. The smoothing bandwidth is 0.4, which is optimal on average for estimating the steady pdf.*

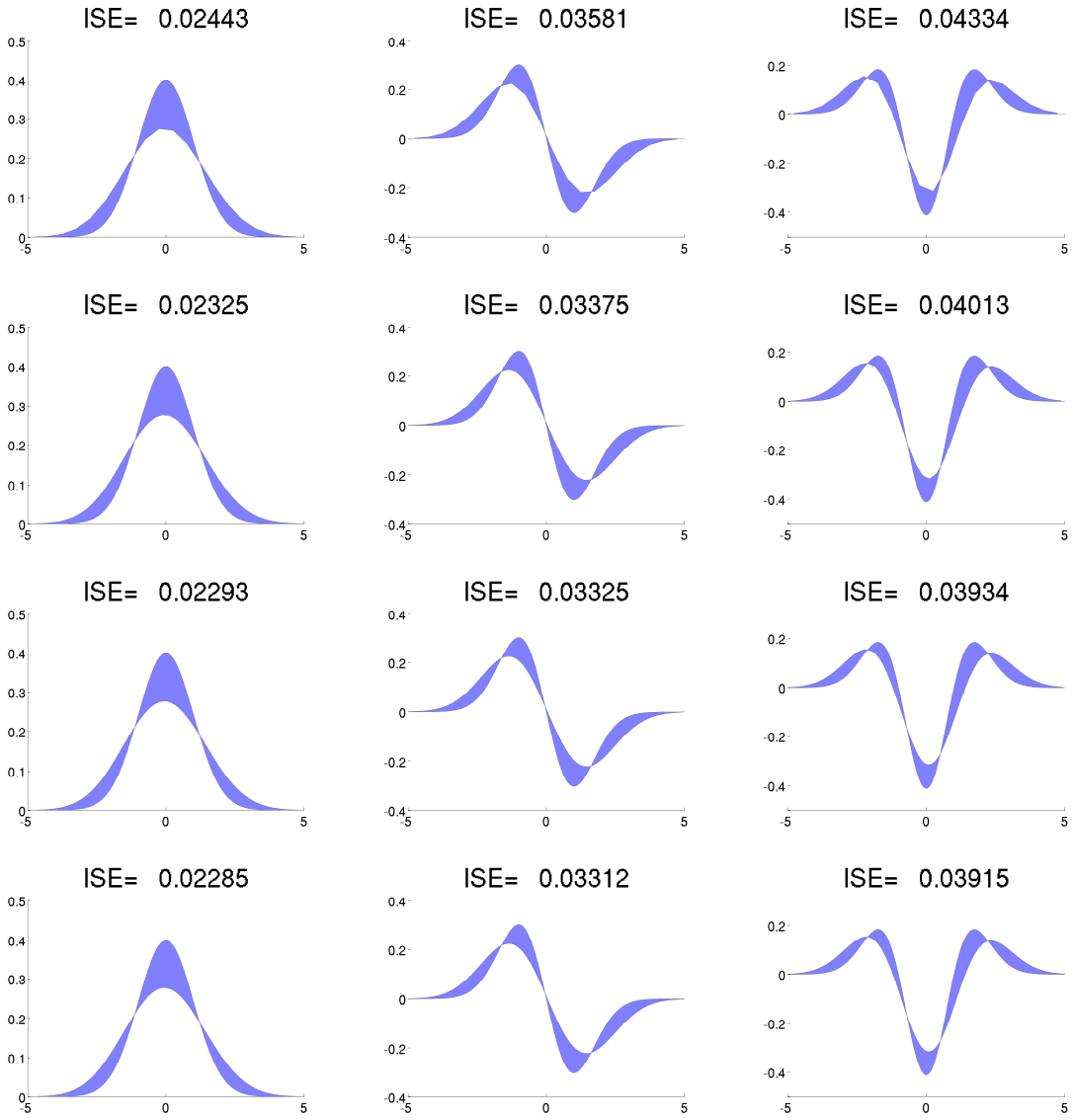
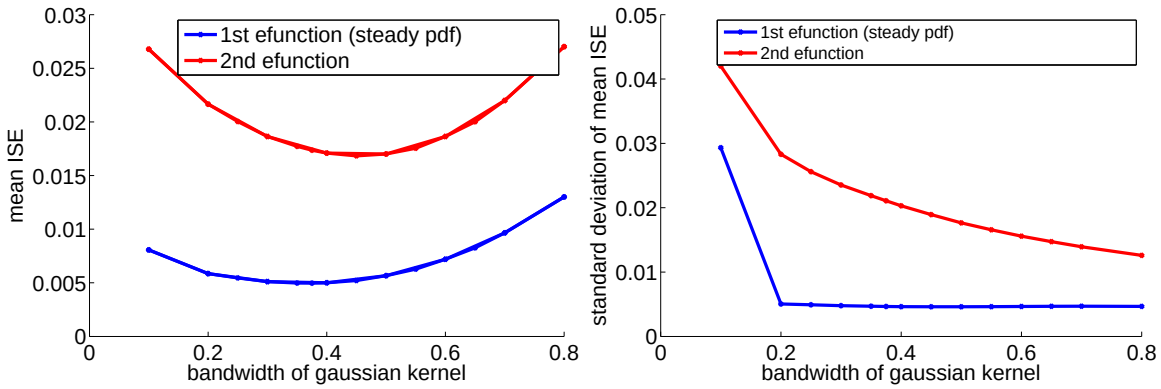


Figure 3.5: Oversmoothed eigenfunctions; first three eigenfunctions (going across the page) of the Fokker-Planck equation for a 1-dimensional Langevin equation. The time series is a simulation of $dX = -X\,dt + \sqrt{2}dW$*, sampled at 10Hz for a series 100 s long. The smoothing bandwidth is 1, which is about four times the optimal level suggested by* (2.34)*.*



(a) (b)

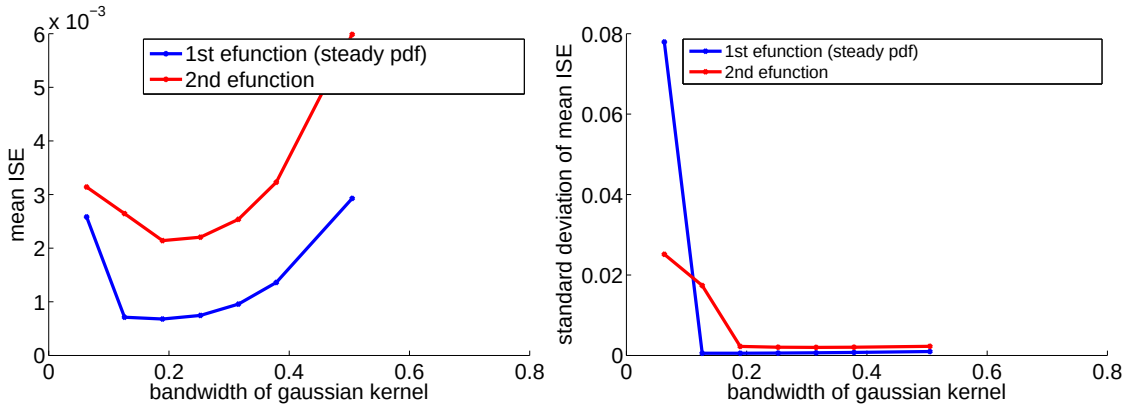
Figure 3.6: sample mean ISE from 2000 runs each of 1001 samples of $dX = -X\,dt + \sqrt{2}dW$, *sampled at* 10 *Hz. Eigenfunctions normalised to have absolute area of one. 80 bins on [-5,5]. Transition matrix obtained by Gaussian kernel density with various bandwidths. The best bandwidth for estimating the dominant eigenfunctions is approximately 0.4.*

3.4 Estimating eigenvalues via transition matrices

This section shows that the best way to estimate the eigenvalues appears to be to have a fine grid of equally likely bins, and to simply count the transitions from one bin to another.

3.4.1 Too few bins causes an overestimate of eigenvalues

The eigenvalues of the Fokker-Planck equation are overestimated if too few bins are used. Figure 3.8 on page 72 shows that the eigenvalues for the problem (3.1) are overestimated if there are less than 100 bins, even if a long time series is available. This overestimate in the eigenvalues is worsened by taking more frequent time samples. In the extreme of using only two semi-infinite bins, then even if



(a) (b)

Figure 3.7: sample mean ISE from 2000 runs each of 10001 samples of $dX = -X\,dt + \sqrt{2}dW$*, sampled at* 10 *Hz. Eigenfunctions normalised to have absolute area of one. 80 bins on [-5,5]. The best bandwidth for estimating the dominant eigenfunctions is approximately 0.2.*

the transition matrix can be perfectly estimated, the overestimate of the non-zero eigenvalue for $\mathrm{d}X = -X\mathrm{d}t + \sqrt{2}\mathrm{d}W$ is approximately

$$\frac{2\sqrt{2}}{\pi\sqrt{\Delta t}} - 1 \to \infty \text{ as } \Delta t \to 0. \tag{3.40}$$

Proof of (3.40)

With only 2 symmetric bins, the true probability transition matrix

$$\mathbf{T} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \tag{3.41}$$

$$\text{where } p = \text{Probability}(X_{t+\Delta t} \leq 0 | X_t \leq 0)\,, \tag{3.42}$$

$$\text{so that as } \Delta t \to 0,\ p \to 1. \tag{3.43}$$

If the process is $\mathrm{d}X = -X\mathrm{d}t + \sqrt{2}\mathrm{d}W$ then by (3.7)-(3.8) the first non-zero eigenvalue is 1, and the steady distribution $f_0(x) \sim N(0,1)$ and for any Δt,

$X_{t+\Delta t} \sim N(e^{-\Delta t}x_t, 1-e^{-2\Delta t})$ [24, p. 35, equation (7.4)], so that

$$\text{Probability}(X_{t+\Delta t} < x_0) = \Phi\left(\frac{x_0 - \ x_t e^{-\Delta t}}{\sqrt{1-e^{-2\Delta t}}}\right), \tag{3.44}$$

$$\text{where } \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{\infty}^{y} e^{-\frac{t^2}{2}}\ \mathrm{d}t. \tag{3.45}$$

$$\therefore p = \frac{\text{Probability}(X_{t+\Delta t} \le 0\ \&\ X_t \le 0)}{\text{Probability}(X_{t+\Delta t} \le 0)} \ \text{ by (3.42)} \tag{3.46}$$

$$= \frac{\text{Probability}(X_{t+\Delta t} \le 0\ \&\ X_t \le 0)}{1/2} \tag{3.47}$$

$$= 2\int_{-\infty}^{0} \text{Probability}(X_{t+\Delta t} \le 0\ |X_t = x)\, f_0(x)\ \mathrm{d}x \tag{3.48}$$

$$= 2\int_{-\infty}^{0} \Phi\left(\frac{-\ xe^{-\Delta t}}{\sqrt{1-e^{-2\Delta t}}}\right) f_0(x)\ \mathrm{d}x \text{ by (3.44).} \tag{3.49}$$

By expanding the exponential and square root in (3.49) for small Δt,

$$p \approx 2\int_{-\infty}^{0} \Phi\left(\frac{-x}{\sqrt{2\Delta t}}\right) f_0(x)\ \mathrm{d}x \tag{3.50}$$

$$= 2\int_{-\infty}^{0} \left\{1 - \Phi\left(\frac{x}{\sqrt{2\Delta t}}\right)\right\} f_0(x)\ \mathrm{d}x, \text{ since } \Phi(-y) = 1-\Phi(y), \tag{3.51}$$

$$\Rightarrow p \approx= 1 - 2\int_{-\infty}^{0} \Phi\left(\frac{x}{\sqrt{2\Delta t}}\right) f_0(x)\ \mathrm{d}x. \tag{3.52}$$

When $\sqrt{2\Delta t}$ is small, then, as $x \to -\infty$, $\Phi\left(\frac{x}{\sqrt{2\Delta t}}\right) \to \Phi(-\infty) = 0$ much faster than $f_0(x)$ changes. So (3.52) may be approximated by treating $f_0(x)$ within the integral as constant i.e.

$$p \approx 1 - 2\ f_0(0)\int_{-\infty}^{0} \Phi\left(\frac{x}{\sqrt{2\Delta t}}\right)\ \mathrm{d}x \tag{3.53}$$

$$= 1 - \frac{2}{\sqrt{2\pi}}\sqrt{2\Delta t}\int_{-\infty}^{0} \Phi(y)\ \mathrm{d}y, \tag{3.54}$$

by changing the dummy variable in the integral in (3.53) to $y = x/\sqrt{2\Delta t}$.

$$\text{Now } \int_{-\infty}^{0} \Phi(y)\ \mathrm{d}y = \int_{-\infty}^{0}\int_{-\infty}^{y} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\ \mathrm{d}x\ \mathrm{d}y \tag{3.55}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0}\int_{x}^{0} e^{-x^2/2}\ \mathrm{d}y\ \mathrm{d}x \tag{3.56}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} -x\ e^{-x^2/2}\ \ \mathrm{d}x \tag{3.57}$$

$$= \frac{1}{\sqrt{2\pi}}. \tag{3.58}$$

$$\text{So, by (3.54) and (3.58), } p \approx 1 - \frac{\sqrt{2\Delta t}}{\pi}, \text{ satisfying (3.43).} \tag{3.59}$$

The eigenvalues of $\mathbf{T}$ are 1 and $(2p-1)$, so the non-zero eigenvalue of the Fokker-Planck equation is estimated as $-\log(2p-1)/\Delta t$

$$= -\log(1 + (2p-2))/\Delta t \tag{3.60}$$

$$\approx (2-2p)/\Delta t \text{ by expanding log} \tag{3.61}$$

$$\approx \frac{2\sqrt{2}}{\pi\sqrt{\Delta t}} \text{ by (3.59).} \tag{3.62}$$

Equation (3.62) has also been checked empirically, by generating long series, sampling frequently, counting the transitions over $(-\infty, 0]$ and $[0, \infty)$ and calculating the smaller eigenvalue of the resulting 2x2 matrix.

3.4.2 Kernel density smoothing overestimates eigenvalues

The artificial noise assumed by kernel smoothing has the same effect as actual noise, noted in section 2.1.3. That is, Gaussian kernel smoothing makes the overestimates of the eigenvalues worse, as it adds an upwards bias to the estimate. Also, kernel smoothing does not provide any compensating reduction in the variance of the estimated eigenvalues, so it appears to add no net benefit. This is shown in figure 3.9, which shows estimates of the first non-zero eigenvalue for 100 different time series, each of which is a simulation of $\mathrm{d}X = -1X\mathrm{d}t + \sqrt{2}\mathrm{d}W$ (so the exact eigenvalue is 1). This suggests that kernel density estimation should not be used to estimate the eigenvalues.

3.4.3 Eigenvalue variance scales with reciprocal of length of time series

Figure 3.10 on page 74 shows that when equally probable bins are used with bin-counting, the variance of the eigenvalues of the transition matrix is approximately proportional to $1/N$, where N is the length of the series. This is consistent with the argument of [19] which uses the Central Limit Theorem to scale errors in the estimated response to a change in forcing proportionately to $1/\sqrt{N}$. When logs are taken and the eigenvalues of the Fokker-Planck equation are estimated, their variance is also approximately proportional to $1/N$. The variance is larger if equally sized (rather than equally probable) bins are used. The difference between equally sized and equally likely bins is amplified if the number of bins is small.

The relevant length of the time series is the multiple of the system's characteristic timescale $1/\alpha$. Figure 3.11 on page 75 shows that if the system evolves slowly (with $\alpha = \alpha_0$, say) then a sample at 10Hz over a period of 1000 time units will give eigenvalue estimates with a similar variance to those for a system with α=10 α_0 also taken at 10Hz over 100 time units.

Figure 3.12 shows the sample standard deviation and variance of the estimate of the eigenvalues (in proportion to the actual size of the eigenvalues) for increasing lengths of time series. This demonstrates a limitation of the theory in chapter 1. By (1.38) the slowest i.e. smallest eigenvalues are most significant, but data limitations make them the hardest to estimate with confidence. Assuming an approximately Gaussian distribution of the sampled eigenvalues, then two systems will be detected as having significantly different timescales only if the proportional difference in their eigenvalues is more than about twice the eigenvalues' standard deviation. So if the available time series are shorter than about 80 multiples of the characteristic timescale $1/\alpha$, then two systems will be detected as significantly different only if their eigenvalues differ by more than 40%.

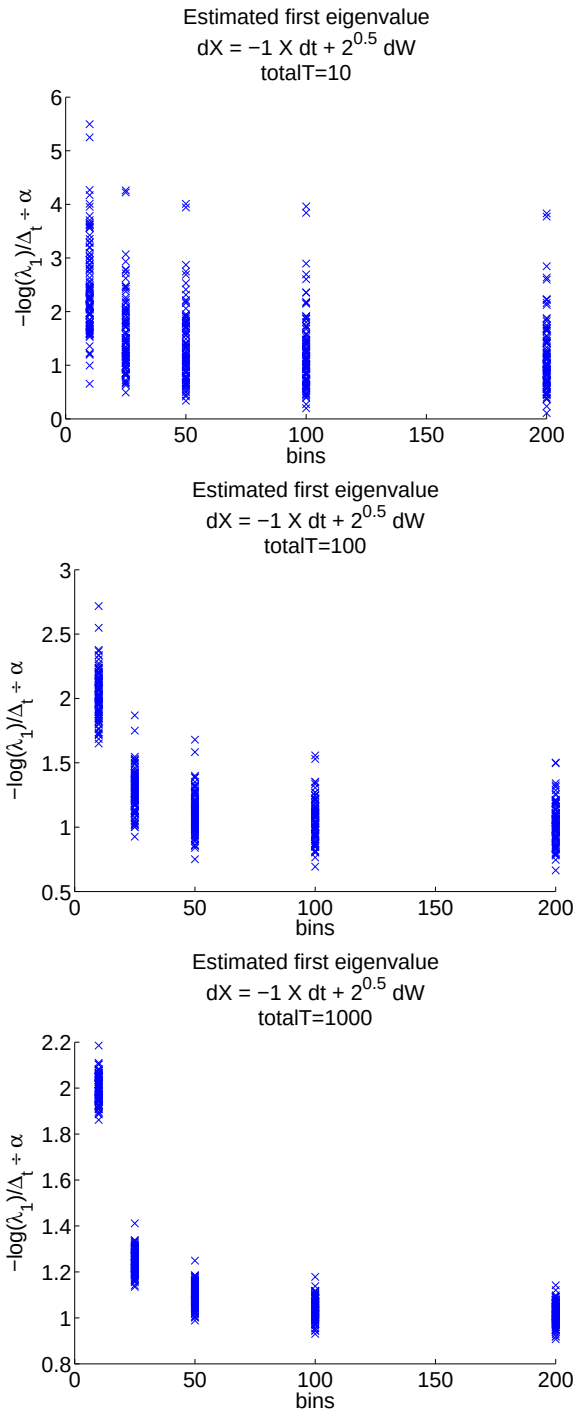


Figure 3.8: Eigenvalues for the problem (3.1) *are overestimated if there are less than 100 equally likely bins. The graph shows the result of estimating the first non-zero eigenvalue for* $dX = -1X\,dt + \sqrt{2}dW$. *100 simulations of the system over 10, 100, 1000 model seconds were made and sampled at 10 Hz. Transition matrices were calculated using 10, 25, 50, 100, 200 equally likely bins, and using bin-counting, with no kernel density smoothing. The theoretical eigenvalue is 1. The sampled eigenvalues have ranges of about 1 which reduce only a little if more bins are used. The range also reduces slowly with larger series. In addition, there is an upwards bias in the sampled values if fewer than 100 bins are used.*

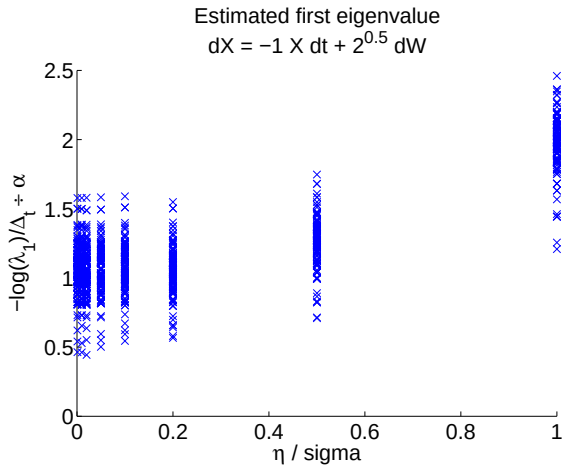
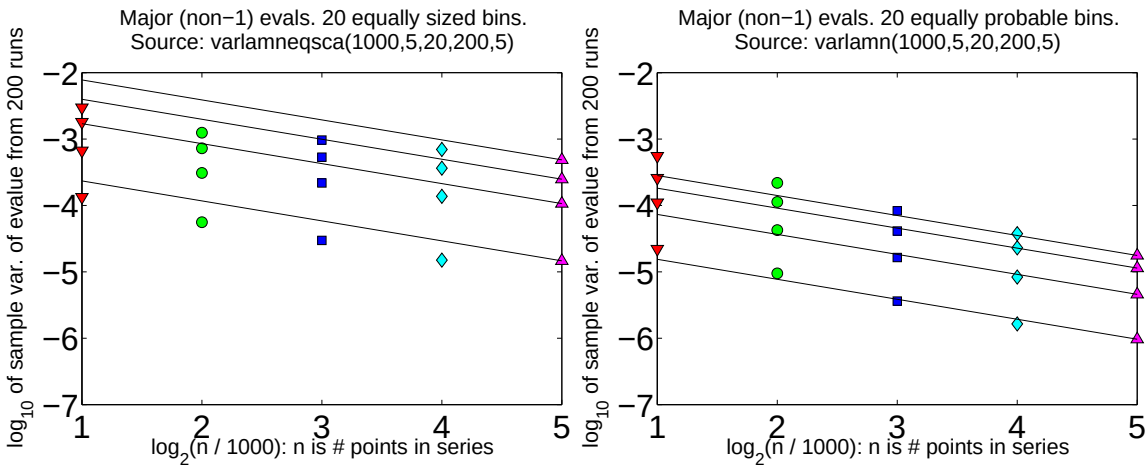
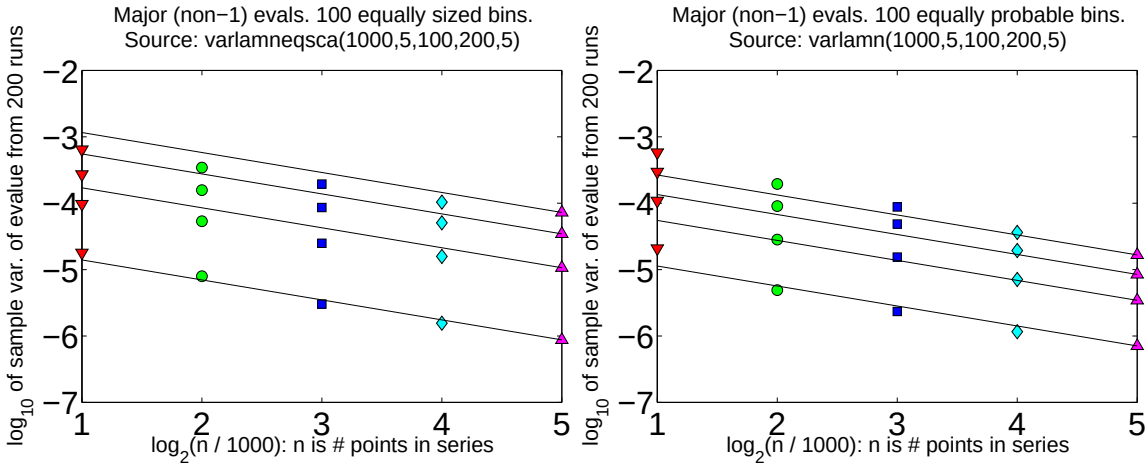


Figure 3.9: Estimate of first non-zero eigenvalue is biased upwards with Gaussian kernel smoothing. The graph shows the result of estimating the first non-zero eigenvalue for $dX = -1X\,dt + \sqrt{2}\,dW$*. 100 simulations of the system over 100 time units were made and sampled at 10 Hz. Transition densities were estimated via gaussian kernel estimation with bandwidths which were various multiples of the sample standard deviation of the series. The densities were integrated over 100 bins, equally spaced between the lowest sample (less 3 times the bandwidth) to the highest sample (plus 3 times the bandwidth).*



(a) Equally spaced bins

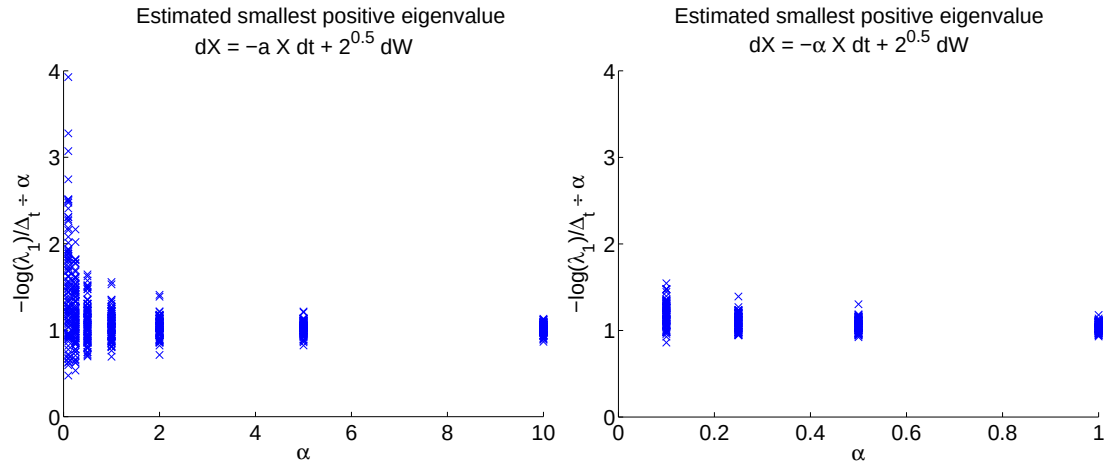
(b) Equally likely bins



(c) Equally spaced bins

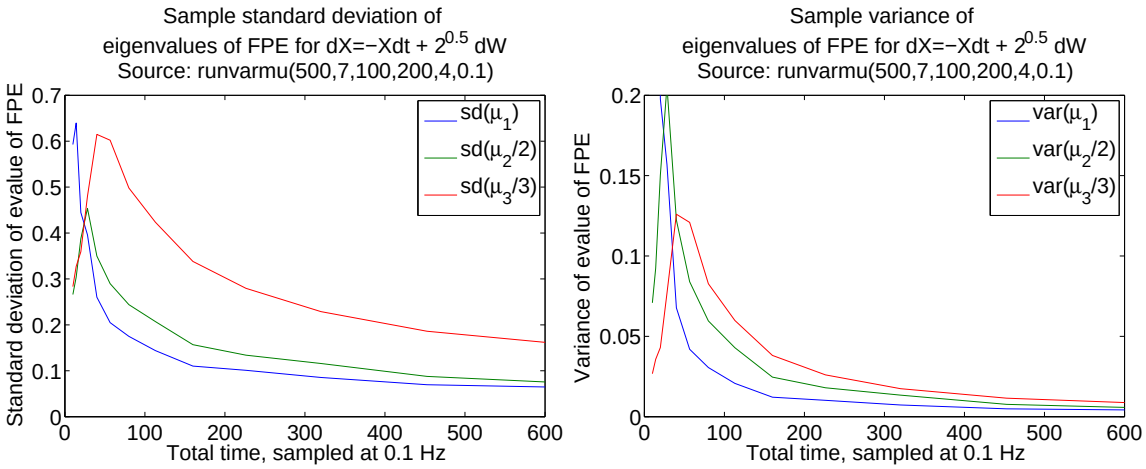
(d) Equally likely bins

Figure 3.10: Each graph shows the sample variance of the first 4 non-zero eigenvalues of transition matrices based on simulations of $dX = -1X\,dt + \sqrt{2}dW$. *The vertical scale is determined by the timestep* Δt. *200 simulations of the system over N=20, 40, 80, 160, 320 model seconds were made and sampled at 100 Hz. The transition matrices were calculated using 20 (top) or 100 (bottom) equally spaced or equally likely bins, using bin-counting, with no kernel density smoothing. The straight lines are each a constant times* $1/N$. *The sample variance of each sampled eigenvalue lies roughly on one of the straight lines, indicating that the sample variance scales approximately with* $1/N$. *Where the bins are equally sized, and especially for the higher eigenvalues, the variance scales more slowly than the reduction in* $1/N$.



(a) 100 time units sampled at 10Hz (b) 1000 time units sampled at 10Hz

Figure 3.11: Variance of estimate of smallest positive eigenvalue increases as timescale $1/\alpha$ increases. A slower system, with lower α, requires a longer time-series to be sampled well and obtain reliable estimates of eigenvalues. 100 equally likely bins. Transition matrix obtained by bin-counting with no smoothing.



(a) Standard deviation (b) Variance

Figure 3.12: Variance of estimate of eigenvalues falls as length of time series falls.

3.5 Test Case: Box model

The test case in 3.1 is idealised, being both linear and one-dimensional, whereas the climate system is non-linear and effectively infinite-dimensional. To analyse the multi-dimensional real (or model) climate requires some strategy for reducing the number of dimensions, since finite computer power limits the number of bins that can be used, and finite observations limit confidence in the sampled transition densities. Adding the complexity of multi-dimensionality, this section defines a box model of one hemisphere and analyses it using the bin-counting method in section 3.4.

Figure 3.13 represents the model, which is of the atmosphere and ocean temperatures in tropical and polar boxes of one hemisphere. The hemisphere is heated constantly with no diurnal or annual cycle. Noise comes from fluctuations to the tropical sea-surface temperature. The tropical sea-surface temperature anomalies form an independent stochastic process, which evolves according to a 1-dimensional linear Langevin equation with a similar timescale to El Niño. For convenience, the polar and tropical boxes cover equal surface area.

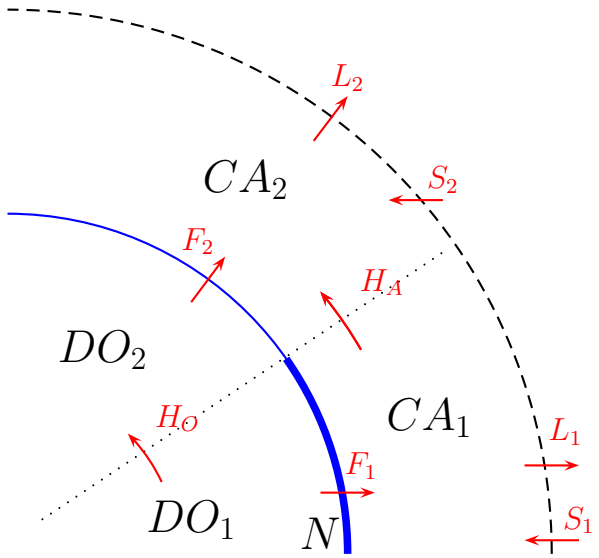


Figure 3.13: Stocks and flows of fourbox model.

The form of the model is

$$d\vec{X} = \left(\mathbf{M}\vec{X} + \vec{R}\right) dt + \vec{\epsilon}\, dW \tag{3.63}$$

where $\vec{X}$ is the state variable, $\mathbf{M}$ is a constant matrix, $\vec{R}$, $\vec{\epsilon}$ are constant vectors, and W is a single Wiener process. The components of the state variable $\vec{X} = (A_1, A_2, O_1, O_2, N)'$ are:

A_1, A_2, O_1, O_2 Tropical, polar, near-surface air and ocean temperature

N Surface temperature anomaly (ENSO index)

In the model, heat is introduced by solar radiation of constant heat fluxes S_1 and S_2 (different due to the different average angles of incidence of tropical and polar solar radiation). Heat transfers at rates proportionate to the temperature difference between boxes, and at different rates depending on the processes of heat transfer (which depend on whether the transfer is ocean-ocean, ocean-atmosphere etc). The assumed heat fluxes are:

$$F_1 = \Lambda(O_1 + \alpha N - A_1), \tag{3.64}$$

$$F_2 = \Lambda(O_2 - A_2), \tag{3.65}$$

$$H_A = K_A(A_1 - A_2), \tag{3.66}$$

$$H_O = K_O(O_1 - O_2), \tag{3.67}$$

$$L_i = p_i + BA_i,\ i = 1, 2. \tag{3.68}$$

where $\Lambda, \alpha, K_A, K_O, p_i, B$ are the scalar constants in table 3.1 on page 79. The long-wave radiation L_i is a linearisation of the rate of grey-body radiation. If, say, an increase in greenhouses gases further insulated outgoing radiation, then the body would get greyer and both p_i and B would fall.

Multiplying the box temperatures by assumed constant box heat capacities C and D gives the total heat fluxes into each box as:

$$C\dot{A}_1 = S_1 + F_1 - L_1 - H_A, \tag{3.69}$$

$$C\dot{A}_2 = S_2 + F_2 - L_2 + H_A, \tag{3.70}$$

$$D\dot{O}_1 = -F_1 - H_O, \tag{3.71}$$

$$D\dot{O}_2 = -F_2 + H_O. \tag{3.72}$$

The sea-surface temperature anomaly is assumed to evolve independently according to a Langevin equation:

$$\mathrm{d}N_t = -aN_t\mathrm{d}t + \epsilon\mathrm{d}W_t, \tag{3.73}$$

where a and ϵ are constants and W_t is a Wiener process. Hence, in (3.63)

$$\vec{R} = \begin{pmatrix} (S_1 - p_1)/C \\ (S_2 - p_2)/C \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \epsilon \end{pmatrix}, \tag{3.74}$$

$$M = \begin{pmatrix} -(B+\Lambda+K_A)/C & K_A/C & \Lambda/C & 0 & \alpha\Lambda/C \\ K_A/C & -(B+\Lambda+K_A)/C & 0 & \Lambda/C & 0 \\ \Lambda/D & 0 & -(\Lambda+K_O)/D & K_O/D & -\alpha\Lambda/D \\ 0 & \Lambda/D & K_O/D & -(\Lambda+K_O)/D & 0 \\ 0 & 0 & 0 & 0 & -a \end{pmatrix}. \tag{3.75}$$

3.5.1 Model timescales

The box model in (3.63) decomposes into 5 linear one-dimensional Langevin equations, each with a timescale inversely proportional to an eigenvalue of the matrix $\mathbf{M}$. For $\mathbf{M}$ is invertible and diagonalizable. So, there is a matrix $P : P^{-1}\mathbf{M}P = D$, where D is diagonal.

Parameter	Value	Units	Based on
B	1.7	W m^{-2} K^{-1}	[30, p.1353]
$S_1 - p_1$	554.4	W m^{-2}	[30, p.1353] Switch from $^\circ C$ to K : $90 + 273.15 \times B = 90 + 273 \times 1.7$.
$S_2 - p_2$	424.4	W m^{-2}	[30, p.1353] Switch from $^\circ C$ to K : $-40 + 273.15 \times B = -40 + 273 \times 1.7$.
Ocean depth	1000	m	[13, p.2808]
C	10^7	J m^{-2} K^{-1}	Air mass 10^4 kg m$^{-2}\times$ Specific heat capacity 1005 J kg^{-1} K^{-1}
D	4×10^9	J m^{-2} K^{-1}	Density 1000 kg m^{-3} $\times$ 1000 m depth $\times$ specific heat capacity 4000 J kg^{-1} K^{-1}
Λ	15	W m^{-2} K^{-1}	[13, p.2808]
K_A	1.3	W m^{-2} K^{-1}	[30, p.1353 (χ)]
K_O	2.0	W m^{-2} K^{-1}	[30, p.1354, thermohaline heat flux]
a	0.6	K year^{-1}	13 peaks in 120 years (as El Niño)
ϵ	$\sqrt{2a}$	K year^{-1}	Gaussian steady state with unit variance
α	1	1	

Table 3.1: Constants in four-box model: An increase in CO_2 to, say, a doubling of CO_2 would cause B and p_i to fall, and thus the equilibrium temperature to rise.

$$\text{Let } \vec{y} = \vec{x} + \mathbf{M}^{-1}\vec{R} \tag{3.76}$$

$$\Rightarrow \mathrm{d}\vec{y} = \mathrm{d}\vec{x} = \mathbf{M}\vec{y}\,\mathrm{d}t + \vec{\epsilon}\,\mathrm{d}W \qquad \text{by (3.63)} \tag{3.77}$$

$$\text{Let } \vec{b} = P^{-1}\vec{y} \tag{3.78}$$

$$\Rightarrow \mathrm{d}\vec{b} = P^{-1}\left(\mathbf{M}\vec{y}\,\mathrm{d}t + \vec{\epsilon}\,\mathrm{d}W\right) \tag{3.79}$$

$$= D\vec{b}\,\mathrm{d}t + P^{-1}\vec{\epsilon}\,\mathrm{d}W. \tag{3.80}$$

In this (linear) case, the eigenvectors and eigenvalues of the drift matrix $\mathbf{M}$ are

related to the eigenvectors and eigenvalues of the Fokker-Planck equation, because equation (3.63) decomposes into 5 one-dimensional linear Langevin equations in (3.80). From (3.7) or by considering a change in the time unit, the timescale of the one-dimensional linear Langevin equations are determined by the size of $D_{i,i}$ in (3.80), i.e. by the size of the eigenvalues of $\mathbf{M}$. So, from the solutions to the characteristic equation of $\mathbf{M}$, the timescales of each component b_i, in ascending order, are

$$C\left(\frac{1}{\Lambda+B+2K_A}+O\left(\frac{C}{D}\right)\right)\approx\frac{C}{\Lambda}\approx 8 \text{ days}, \tag{3.81}$$

$$C\left(\frac{1}{\Lambda+B}+O\left(\frac{C}{D}\right)\right)\approx\frac{C}{\Lambda}\approx 8 \text{ days}, \tag{3.82}$$

$$\frac{1}{a}\approx 1.7 \text{ years}, \tag{3.83}$$

$$D\left(\frac{1}{2K_O+\frac{B+2K_A}{1+B/\Lambda+2K_A/\Lambda}}+O\left(\frac{C}{D}\right)\right)\approx\frac{D}{2K_O+2K_A+B}\approx 15 \text{ years}, \tag{3.84}$$

$$D\left(\frac{1}{B}+\frac{1}{\Lambda}+O\left(\frac{C}{D}\right)\right)\approx\frac{D}{B}\approx 75 \text{ years}, \tag{3.85}$$

As observed in section 3.4.3, as a timescale lengthens, a longer time series is required to sample the state space of its variable. By (3.8) and (3.80), the true steady state of each component is a Gaussian distribution. But figure 3.18 shows the bell-shape of the Gaussian distribution is only seen when the time series lasts for about 100 multiples of the timescale, consistent with 3.4.3. The eigenvector with the longest timescale, approximated in figure 3.14, is a temperature pattern of general warming. All the other eigenvectors are patterns of temperature contrasts, shown in figures 3.15–3.17. An increase in radiative forcing projects onto the slowest eigenvector, indicating that a very long time series would be needed to use this method to estimate e.g. climate sensitivity. This is consistent with [50], who showed with a stochastic climate model of the form of (1.1), that high-frequency information is of little use in estimating climate sensitivity, as the high-frequency response is approximately independent of climate sensitivity. Still, it should be borne in mind that the results in figure 3.18 are for a system subject to constant forcing (with no daily or annual cycle), so only internal variability is generated. The real system is subject to varying external forcing which provokes a varying response, and so does not require such a very long time series to explore the system's state space.

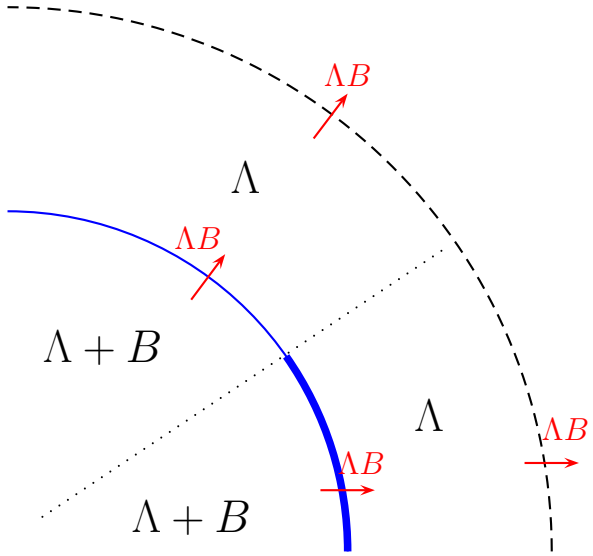


Figure 3.14: Approximate eigenvector of general warming. Ocean heat capacity is D, and ocean heat is lost to space at rate ΛB and time scale is $\frac{D(\Lambda+B)}{\Lambda B}$.

3.6 Detecting a change in eigenvalues of the Fokker-Planck equation

3.6.1 Dimension reduction

In order to analyse real climate data by this method it is necessary to reduce the number of dimensions analysed at once. The multi-dimensional time series must be condensed to lower-dimensional time series. The dimensions or variables of the lower-dimensional time series could be a subset of the original dimensions, or a projection of them onto some other lower-dimensional space.

Possible dimensions for the lower-dimensional series are components of the raw time series, the raw series projected onto the eigenvectors of the drift matrix $\mathbf{M}$, or the principal components of the time series, i.e. the raw series projected onto the eigenvectors of the covariance matrix of the series. In general, the eigenvectors of the drift matrix are unknown or non-existent, so may be of limited use for real climate data.

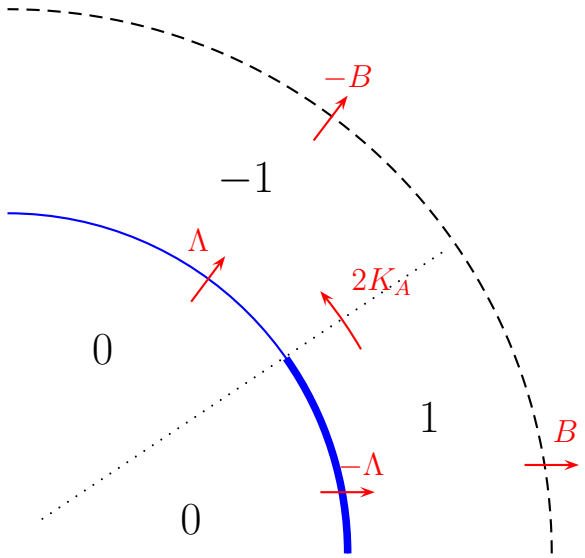


Figure 3.15: Approximate eigenvector of meridional air temperature difference. Air heat capacity is C, and time scale is $\frac{C}{\Lambda + B + 2K_A}$.

Insufficiency of identical eigenvalues and eigenvectors

If the one-dimensional (identical) Niño series from two models with, say different values of B were compared, then clearly there would be no difference in the statistics of the identical series, and this would be insufficient for the two models to have the same response to every forcing. (The response to radative forcing is proportional to $1/B$.)

3.6.2 Method

In figures 3.19 – 3.24, one of the parameters in $\mathbf{M}$ is varied. For each level of the parameter, 20 realisations of the model are generated, starting from the model's equilibrium point and sampling each 5 model days. For each realisation a transition matrix is formed. If just one time-series is analysed, then 100 equally likely bins are used. If two time-series are analysed, then the range of each series is split into 15 equally likely bins, and the bins are combined to make 225 two-dimensional bins over which to analyse the two-dimensional series. The log of the largest non-unity eigenvalue of the transition matrix, divided by the time step in the time series is plotted. A 'signal to noise ratio' can be defined as the sensitivity of the mean sampled eigenvalue to the change in parameter

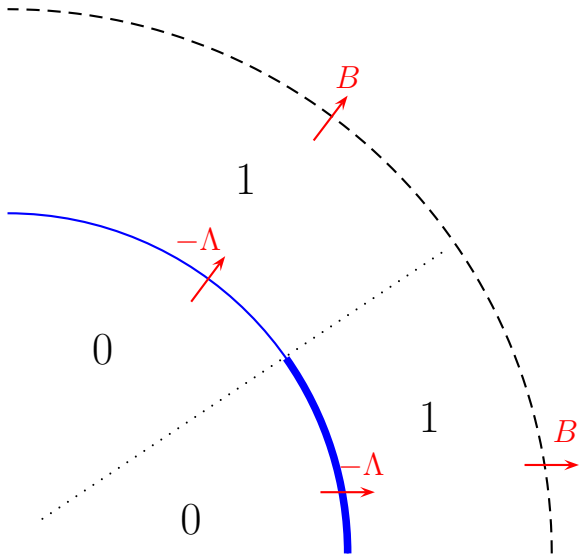


Figure 3.16: Approximate eigenvector of air-ocean temperature difference. Air heat capacity is C, and time scale is $\frac{C}{\Lambda + B}$.

divided by the sample variance of the eigenvalue. The signal to noise ratio is a property of the dimension (i.e. the variable whose timeseries is analysed) as different dimensions have different sensitivities and variances. A low signal to noise ratio is shown by a large spread of eigenvalues with little difference when a parameter is varied. Such a dimension does not easily detect a significant difference in models.

3.6.3 Expected result

Equations (3.81)–(3.85) show the degree to which an eigenvalue of the drift matrix $\mathbf{M}$ is affected by a change in one of the parameters. They also show the timescales of the processes of the components b_i. Thus they show the detectability via that component, of a significant difference between two models. For example, (3.85) shows that the smallest eigenvalue is approximately proportional to the size of B, but also a difference in B between two models will be hard to detect via this component, as it will require a very long time series to measure with confidence. On the other hand (3.81) shows that the largest eigenvalue is approximately proportional to Λ, and is for a fast process, so this component will most easily show differences in Λ between two models. This is confirmed

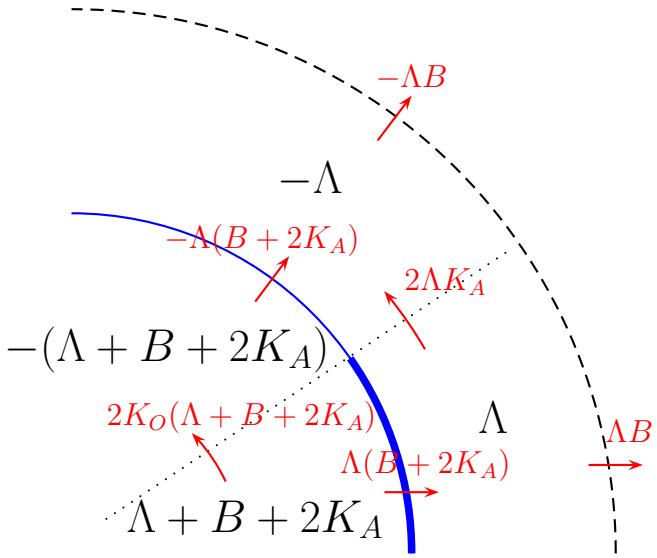
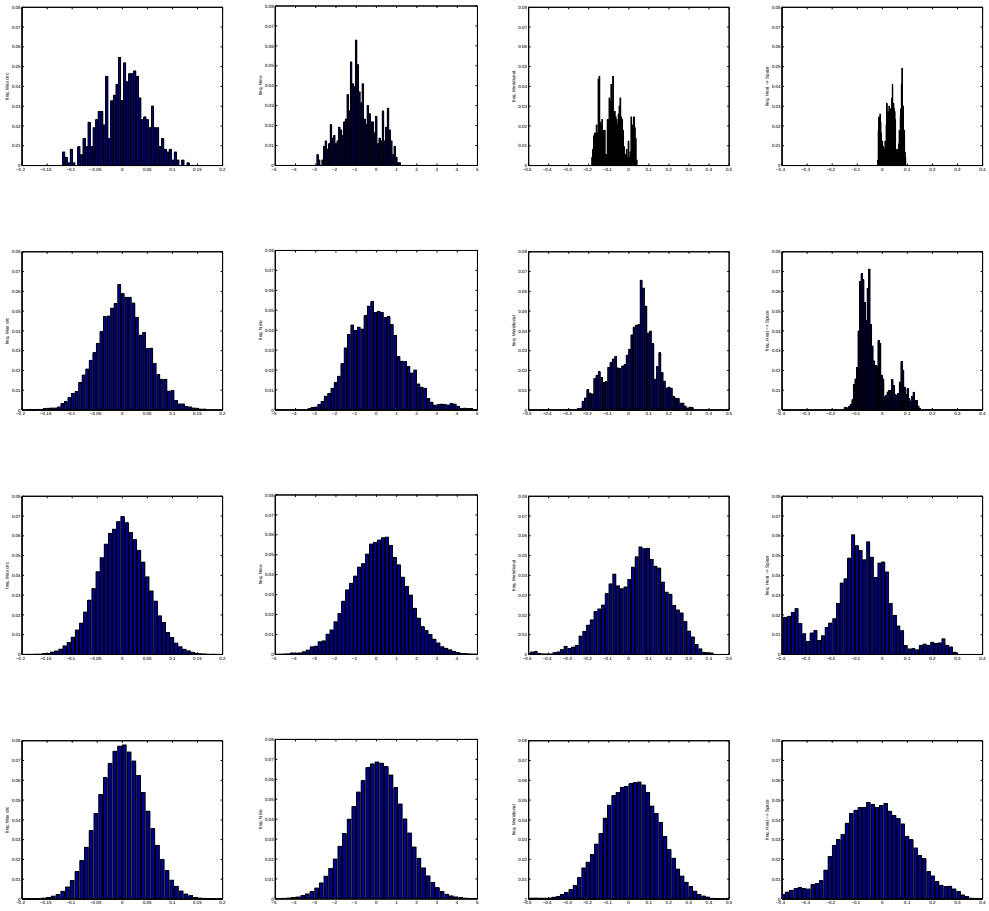


Figure 3.17: Eigenvector of meridional ocean temperature difference. Ocean heat capacity is D, and time scale is $\dfrac{D}{2K_O + \dfrac{\Lambda(B+2K_A)}{\Lambda+B+2K_A}}$.

by figures 3.19–3.24, which show the difference in estimated eigenvalues of the Fokker-Planck equation when just one variable at a time is considered.

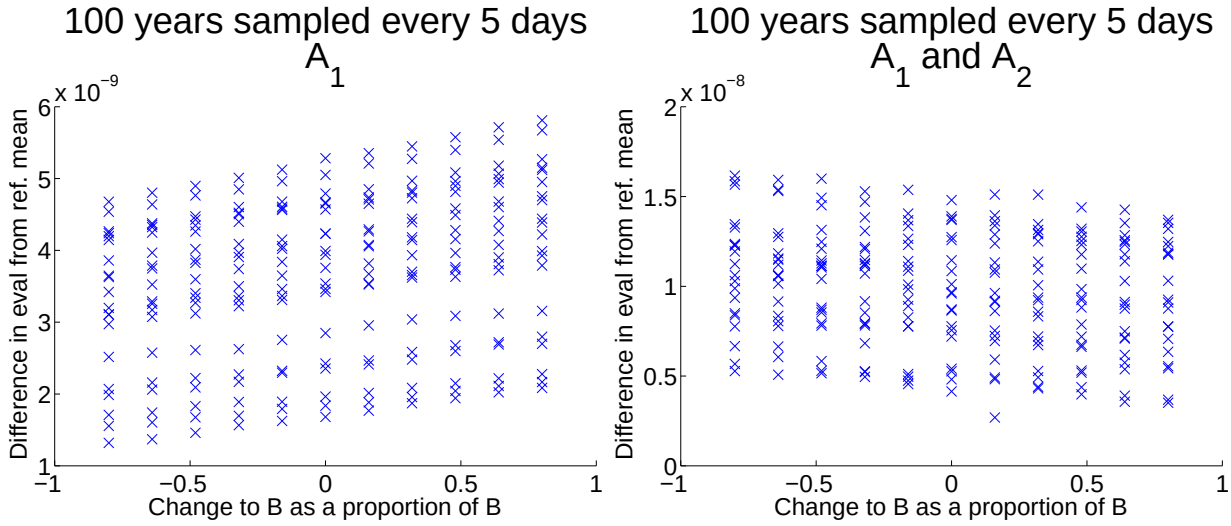
3.6.4 Actual result

Figures 3.19 – 3.21 show the spread of estimated eigenvalues based on timeseries over different dimensions where B is varied. All that differs between 3.19 – 3.21 is the (reduced) dimension-set chosen to analyse the time series over. For each dimension set (raw data, data projected onto eigenvectors of drift matrix, data projected onto EOFs), the graph of the best detector is shown, i.e. the dimension and pair of dimensions with the greatest signal to noise ratio. The signal to noise ratios have been estimated by eye from the graphs of the various dimension sets, so that the set with the highest signal to noise ratio is the steepest and least spread out. Figures 3.22 – 3.24 show a similar set of estimated eigenvalues where heat flux parameter Λ is varied.



10 yrs				
100 yrs				
1000 yrs				
10000 yrs				
Pattern	Meridional air difference	El Nino	Meridional ocean difference	General warming
Approx timescale (yrs)	$\frac{8}{365}$	$1\frac{2}{3}$	15	75

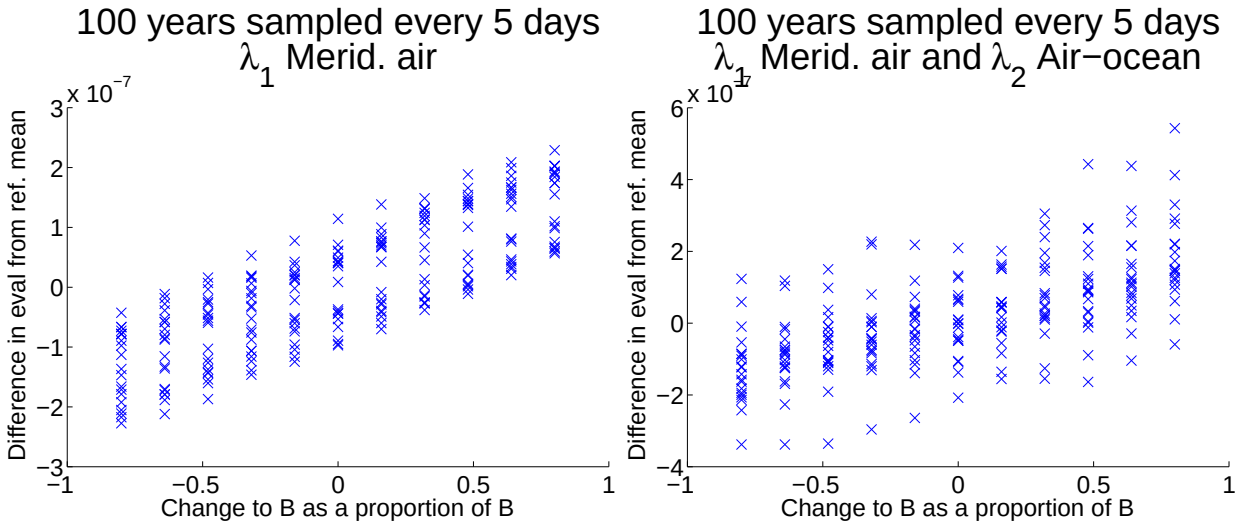
Figure 3.18: Histograms of samples of the four-box state variables projected onto the eigenvectors of the drift matrix $\mathbf{M}$*. As the relaxation timescale of the eigenvector increases, the required length of series to recognise a Gaussian distribution also increases. The model system has no varying forcing, so the variation observed is only from internal variability. If varying forcing were included (as for the real world) then a much shorter observation period would be required to observe the system's variability.*



(a) Single series

(b) Pair of series

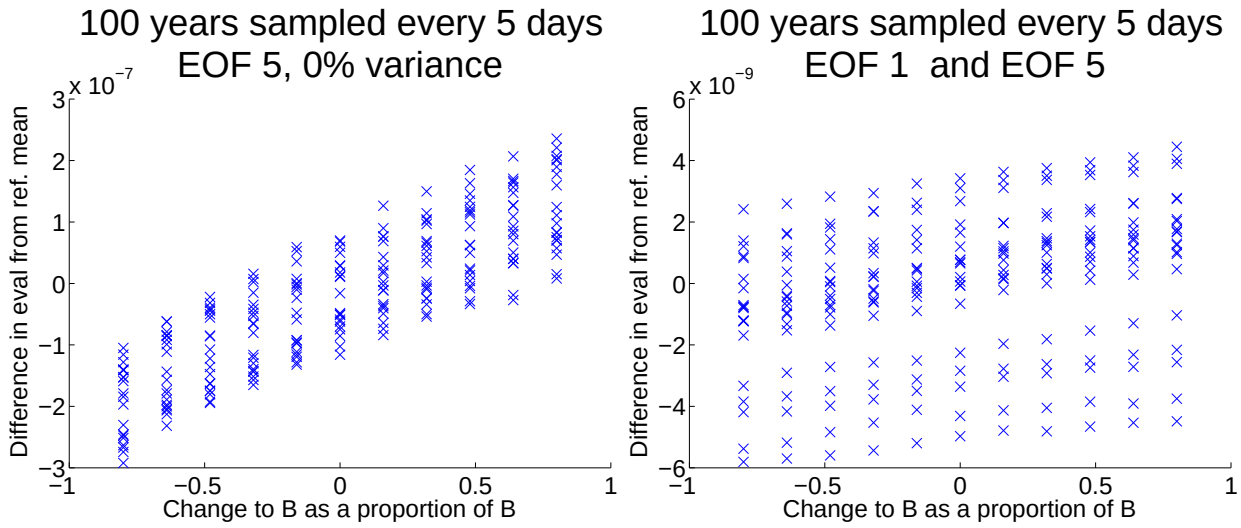
Figure 3.19: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, when model parameter B is varied. The better a series is at detecting a change in model parameter, the greater is the ratio of the gradient of the mean estimate to the variance of the estimates. Tropical atmospheric temperature variable A_1 shows (from all the single raw variables) the greatest detectability of the change in model climate sensitivity parameter B. Variables (A_1, A_2) show almost no detectability of the change in B, though other pairs of raw variables are even worse at detecting a change in B.



(a) Single series

(b) Pair of series

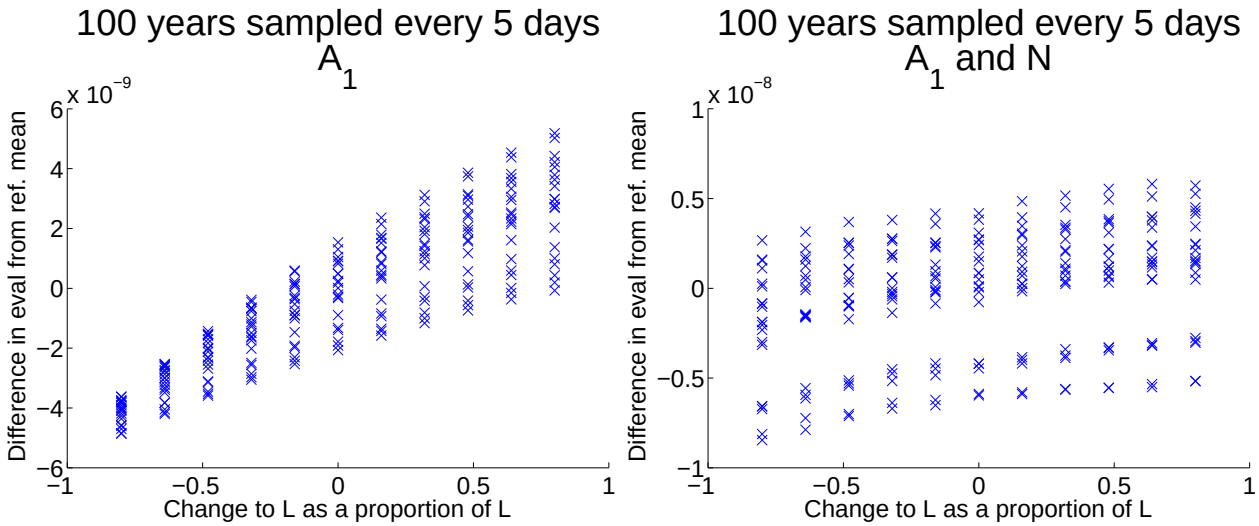
Figure 3.20: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, projected onto eigenvectors of the drift matrix. Meridional air temperature contrast variable λ_1 detects most easily (from all the projections onto single eigenvectors of the drift matrix) the change in model climate sensitivity parameter B. Variables (λ_1, λ_2) detect the change in B most easily for a pair of projected variables, but less easily than just λ_1 alone.



(a) Single series

(b) Pair of series

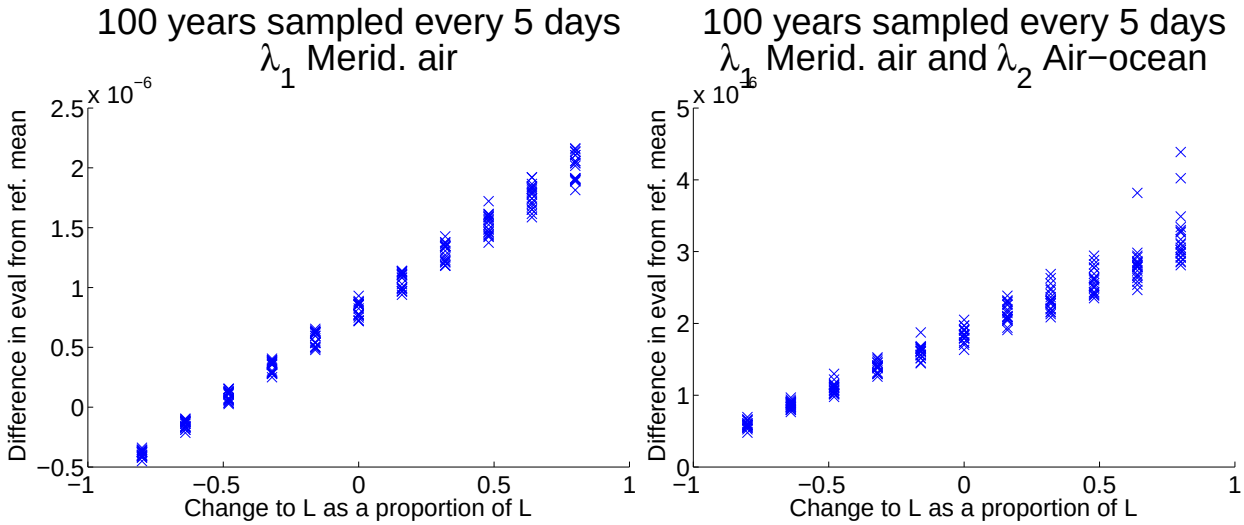
Figure 3.21: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, projected onto the empirical orthogonal functions of the sampled series. The principal component that explains the least *variance detects most easily (from all the principal components) the change in model climate sensitivity parameter B. The pair of principal components that explain respectively the least and most variance detect the change in B most easily for a pair of principal components, but less easily than the best principal component alone.*



(a) Single series

(b) Pair of series

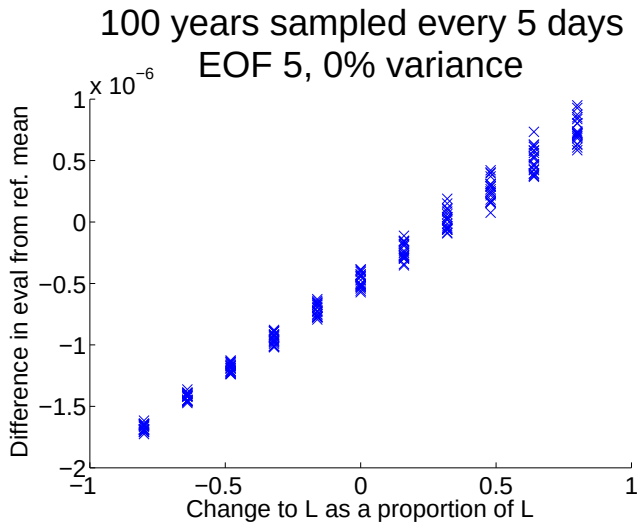
Figure 3.22: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, when model heat flux parameter Λ is varied. The tropical atmospheric temperature variable A_1 shows (from all the single raw variables) the greatest detectability of the change in Λ. As Λ increases, the noise from the Niño variable is amplified throughout the system and so the variance in the sampled eigenvalue increases. Variables (A_1, N) where N is the Niño variable itself, detect the change in Λ most easily for a pair of raw variables, but detect the change less easily than does just variable A_1 alone.



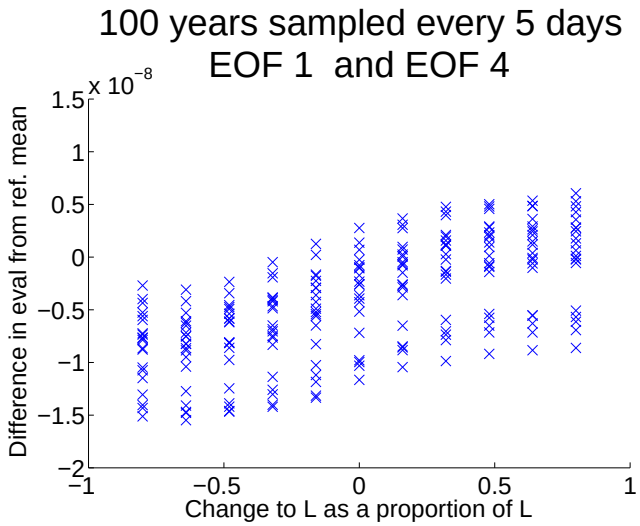
(a) Single series

(b) Pair of series

Figure 3.23: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, projected onto eigenvectors of the drift matrix. Meridional air temperature contrast variable λ_1 detects most easily (from all single variables) the change in model heat flux parameter Λ. Equation (3.81) *shows that the eigenvalue itself is proportional to Λ, and so it is most sensitive to a change in Λ. Also, as the characteristic timescale of the series is small, the available series is effectively long, so there is less sampling noise and the eigenvalues are less spread out. Variables (λ_1, λ_2) detect the change in Λ most easily for a pair of projected variables, but less easily than just λ_1 alone.*



(a) Single series



(b) Pair of series

Figure 3.24: Estimates of the smallest positive eigenvalue of Fokker-Planck equation for the pdf of the reduced-dimension box model, projected onto the empirical orthogonal functions of the sampled series. The principal component that explains the least *variance detects most easily (from all the principal components) the change in model heat flux parameter* Λ*. The pair of principal components that explain respectively the most and fourth most variance detect the change in* Λ *most easily for a pair of principal components, but less easily than the best principal component alone.*

3.6.5 Conclusions

The response to a change in drift $\Delta\mathbf{M}$ is not (as suggested by the first part of (1.38)) simply proportional to the reciprocal of the eigenvalues. As shown in 3.2.3, the projection of the change in drift onto the eigenfunctions of the adjoint (which is the other part of (1.38)) is significant.

The best way to estimate eigenvalues appears to be to have a fine grid of equally likely bins, and to simply count the transitions from one bin to another without smoothing.

Bins

- If the finite set of n calculated eigenvalues is treated as estimates of the first n eigenvalues in the infinite set of actual eigenvalues, then the fewer bins that are used, the more the estimates of the eigenvalues are biased upwards.

- when equally probable bins are used with bin-counting, the variance of eigenvalues is approximately inversely proportional to the length of the series. If the available time series are shorter than about 80 multiples of the system's characteristic timescale, then the method will detect two systems as significantly different only if their eigenvalues differ by more than 40%,

Smoothing

- the best bandwidth from a Gaussian kernel to estimate the dominant eigenfunctions is approximately 0.4 times the system's standard deviation, when the series length is 100 multiples of the system's characteristic timescale. The best bandwidth is approximately 0.2 times the system's standard deviation when the series length is 1,000 multiples of the system's characteristic timescale;

- the best estimate of eigenvalues for the simplest test problem is made with zero bandwidth, i.e. by the bin-counting technique;

Dimensionality

- For a multi-dimensional linear model, differences between models are most detectable when dimensions are treated separately. The eigenvalues based on 2-

dimensional series are more spread out than those based on a single series, as shown in each of figures 3.19 – 3.21.

- The best series to use for detecting differences in models are the quickly-varying components of the projection onto the eigenvectors of the drift matrix $\mathbf{M}$. This may simply be due to the limited length of the time series, consistent with 3.4.3.

Chapter 4

Comparison of climate model and reanalysis data

Chapter 3 compared the eigenvalues of the Fokker-Planck equations of linear, low-dimensional models. This chapter compares the corresponding eigenvalues of a general circulation climate model, HadCM3, with those of the ERA-40 reanalysis of the real earth-atmosphere system [48, 49]. The HadCM3 control run fixes levels of greenhouse gases and aerosols at levels representative of the pre-industrial era [9]. Ideally, we would compare its output with a long time series from the pre-industrial era, as this would compare the natural model variability with the relevant actual natural variability. Long datasets are available, e.g. those based on ice cores. Chapter 3 showed that the longer the dataset, the more reliable the result. But in order to use more than one dataset with the method of chapter 3, it is necessary that each sample from each dataset is taken simultaneously. So we have decided to use the high frequency, multi-dimensional dataset from ERA-40, even though this is not a pre-industrial time series, but an industrial era time series. A similar compromise was made in [51].

Both the climate model and the real system carry the added complexity of being non-linear, and having a very high number of dimensions. The non-linearity means that neither system decomposes into one-dimensional components as in 3.5.1. But the two systems must be condensed to no more than about 3 dimensions or time series at a time in order for us to be able to construct their transition matrices.

The time series chosen to analyse are Niño 3.4, an Arctic Oscillation index, which

evolves over many timescales, and global mean surface temperature, which is a direct measure of global warming. All three indices are large scale summaries of the earth-atmosphere system. The surface temperature index is used because it is well-known and is the basis for descriptions of climate sensitivity [41]. The Niño index is used because its characteristic timescale is longer than that of the Arctic Oscillation or surface temperature indices, so the eigenvalues based on it will be small, so equation (1.38) suggests that differences in the Niño index's behaviour indicate more significant differences in overall dynamics and response. The Arctic Oscillation index is an atmospheric index unlike the other two temperature indices, whose response to climate change is of wide interest [17, 35]. A more systematic selection of dimensions is made by [19], albeit using many more dimensions.

Section 4.3 describes how the time series were constructed and section 4.4 uses the time series to compare the model with the reanalysis data.

4.1 Data source

The ERA-40 and HadCM3 control run data is provided by the British Atmospheric Data Centre [49, 48]. The ERA-40 data has 4 samples per day, on a 2.5 x 2.5 degree grid for 1-Sep-1957 to 31-Aug-2002. The HadCM3 data has one sample per model day, on a 2.5 x 3.75 degree grid of the globe, over 310 model years from a notional 1849 to 2159.

4.2 Eigenvalues and eigenfunctions of Fokker-Planck equation

For each time series, estimated eigenvalues of the corresponding one-dimensional Fokker-Planck equation are obtained by bin-counting (with no smoothing). The first three eigenfunctions of the Fokker-Planck equation are estimated using gaussian kernel density estimation with bandwidth equal to $(\frac{4n}{3})^{-\frac{1}{6}}$ times the sample standard deviation of the series, where n is the number of samples in the series. The graphs of the eigenfunctions have been normalised so that the integral of their absolute area is equal to one.

4.3 Construction and validation of time series

4.3.1 ERA-40 Nino 3.4 index

The Niño 3.4 index is the mean sea-surface temperature in the Equatorial Pacific. The mean is taken over latitudes 5S to 5N and longitudes 170W to 120W [4], and area-weighted by the cosine of the latitude (the area-weights are all close to one). The data samples are treated as average temperatures for a grid-box centred on the sampled latitude and longitude, so only half of the most northerly and southerly boxes are treated as being relevant, and just half of the most easterly and westerly boxes are treated as relevant (illustrated in figure 4.1). So the data points on the boundaries of the box are weighted by a half, and the datapoints on the corner are weighted by a quarter.

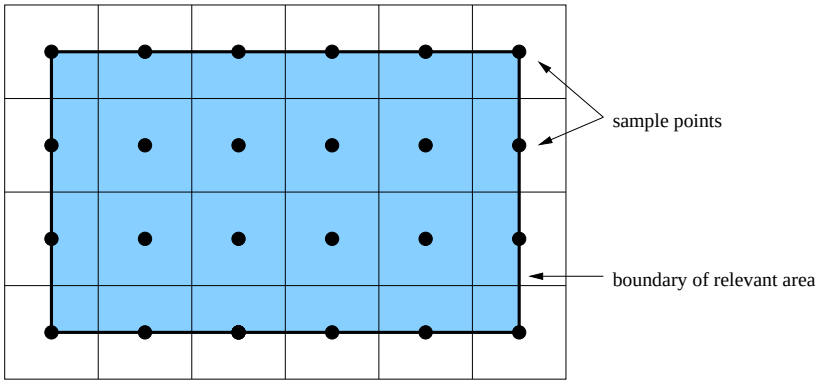


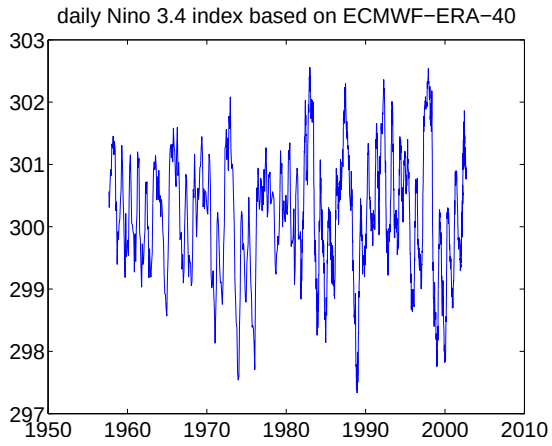
Figure 4.1: Illustration of area weighting of datapoints, used to calculate Niño3.4 index. Points on the boundary of the relevant area are given half the weighting of points in its interior. Points on the corner are given a quarter of the weighting of points in the interior.

The seasonal cycle for the index is the mean (over the 45 years of reanalysis) for each day of the year. The seasonal mean is calculated only for a mean year of 365 days, so, in calculating the mean, the last day in leap years is ignored, for the sake of simplicity. The first day of the series is 1 September (1957), so the last day of the year which is excluded from the mean in leap years is 31 August. The anomaly index is the difference between the index for a day and the mean for that day of the year, counting days since 1 September. In a leap year, the anomaly for 31 August is taken as the difference between the index for 31 August and the mean index for 1 September. The seasonal cycle and

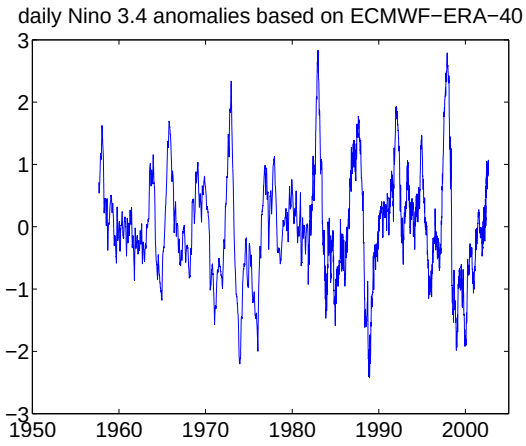
anomaly index are shown in figure 4.2, which also shows the estimated autocorrelation function for the anomaly index.

Figure 4.3 shows an adequate visual match between the calculated daily anomaly, averaged over a month, and the same index calculated by National Oceanic and Atmospheric Administration (NOAA).

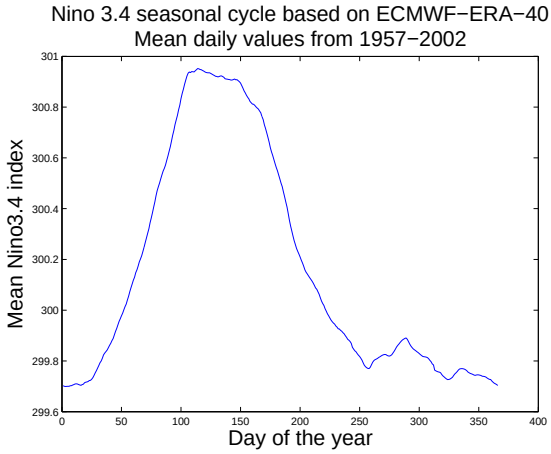
Chang *et al.* [8] and Burgers [6] (reproduced in figure 4.4) report similar Niño autocorrelation functions to the function calculated in figure 4.2 for ERA-40.



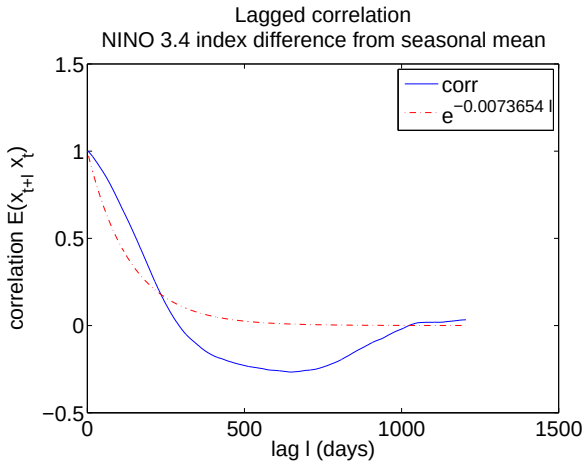
(a) raw series



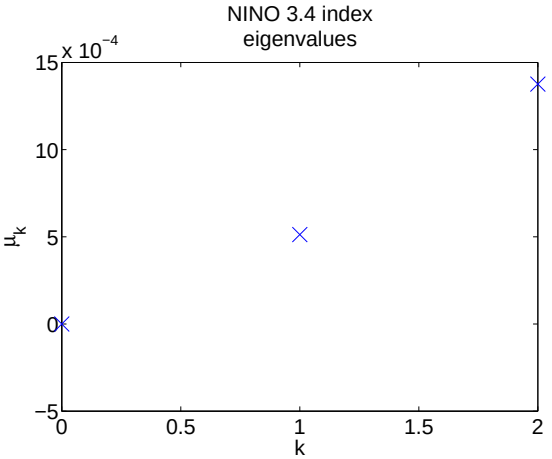
(b) anomalies



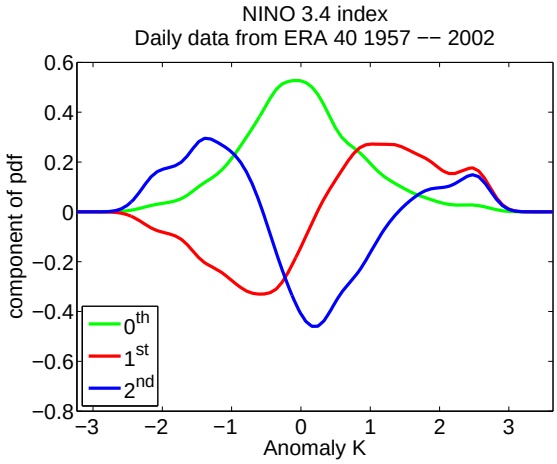
(c) cycle



(d) auto-correlation



(e) evalues



(f) efunctions

Figure 4.2: ERA-40 Nino3.4 index, calculated from approx 45 years of reanalysis. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

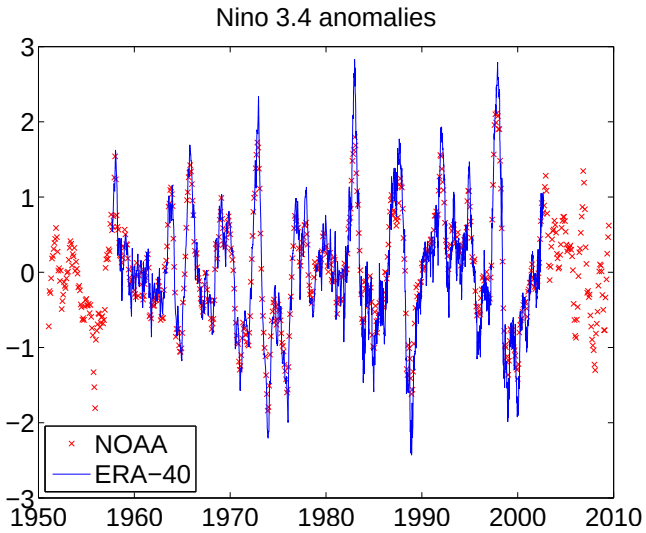


Figure 4.3: Niño 3.4 anomalies based on ERA-40 match monthly means as calculated by NOAA. NOAA figures are from [32].

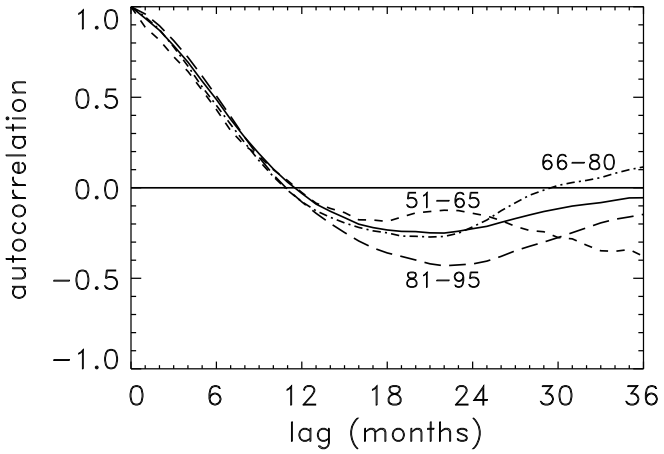


Figure 4.4: The autocorrelation of the observed NINO3.4 index from 1951–1995 (solid line), which shows a similar shape to the autocorrelation calculated in figure 4.2. From [6] based on NCEP data. Reproduced with kind permission of G.Burgers.

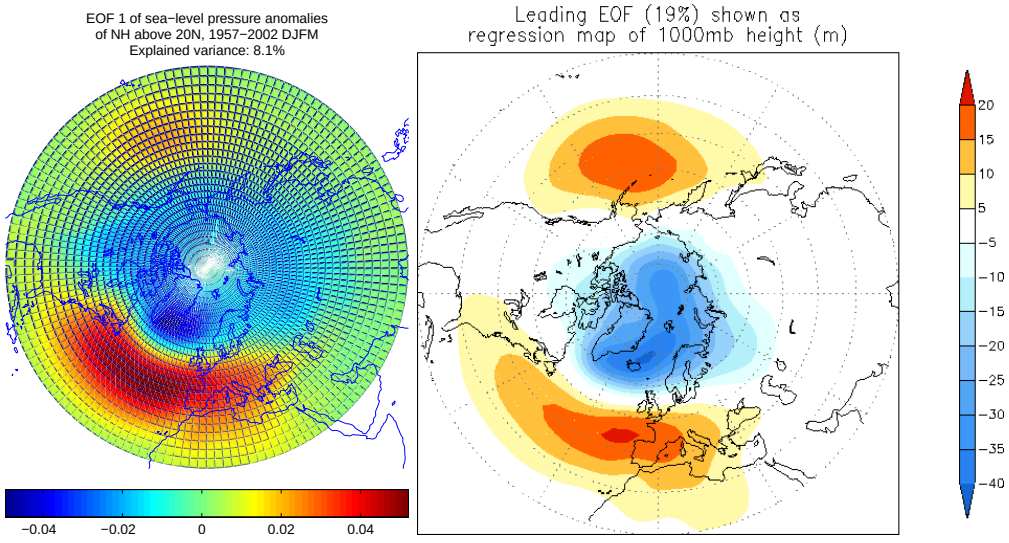
4.3.2 ERA-40 Arctic Oscillation

The Arctic Oscillation index is the principal component of the area-weighted surface pressure anomalies for the northern hemisphere above and including 20 degrees north[2, 45]. The anomalies are the difference between the surface pressure at any time, and the mean surface pressure for that time of day on that day of the year. The mean is taken over the 45 years of reanalysis. The area weighting (which is applied *before* the covariance matrix is calculated) is the average of the cosines of the latitude ± 1.25 degrees. In fact, the *square root* of the cosine of latitude (as for the NOAA index [33]) is the correct weighting, because it causes the *variance* of the anomalies to be weighted by the cosine of latitude, that is by the area-weighting. However, since the same error has been made in calculating the HadCM3 Niño index, the two series may be compared.

The dominant empirical orthogonal function (EOF), shown in figure 4.5, is the dominant eigenvector of the covariance matrix of the winter anomalies (those from December 1 to March 31). The Arctic Oscillation index for any day is the dot product of the area-weighted surface pressure anomaly for that day and the dominant empirical orthogonal vector.

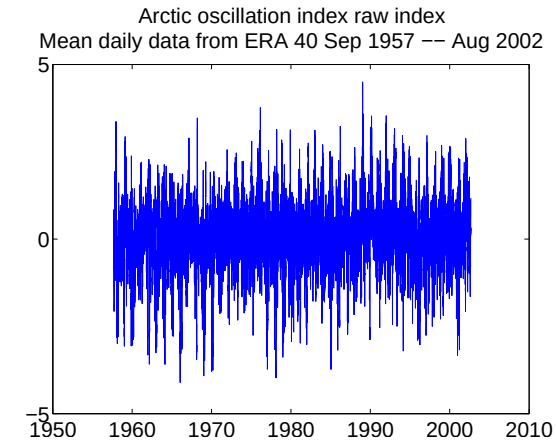
By construction, the mean over all years of every anomaly is zero, so there is no seasonal cycle. This is verified in figure 4.6.

Figure 4.5 on the next page shows that the dominant EOF is similar to that calculated by NOAA, though the NOAA low is more centred over the pole. The monthly mean of the AO index (the first principal component) is correlated with the NOAA AO index with correlation coefficient 0.92. Known differences between the calculations are that the NOAA calculation is based on monthly means for the whole year, a different period (1979-2000), and NOAA weights pressure anomalies (correctly) by the square root of the cosine of latitude, giving more weight to anomalies at the pole [33]. The NOAA data was not used directly because it provided only monthly mean indices, and our method required high temporal resolution.

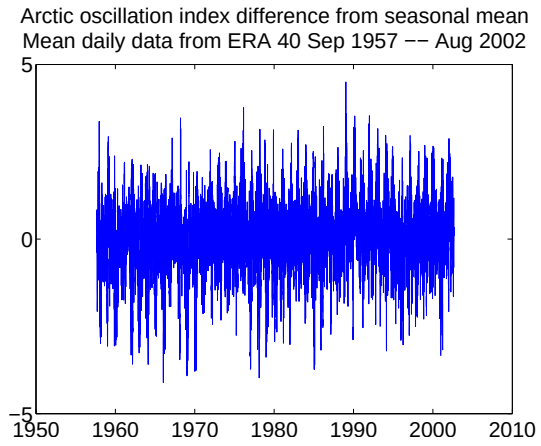


(a) ERA-40 (b) NOAA

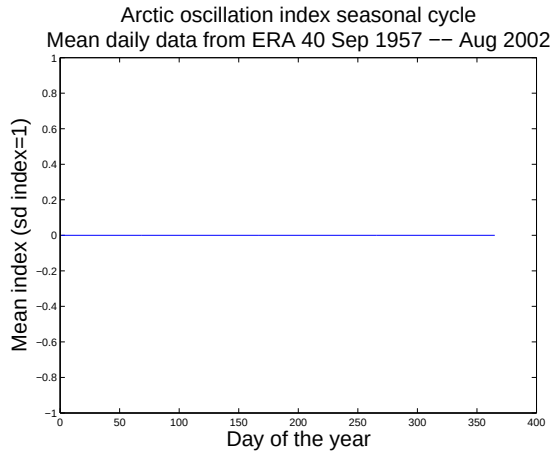
Figure 4.5: First empirical orthogonal function of mean sea level pressure anomalies. The EOF based on ERA-40 data, has the same main features as those of the first EOF calculated by NOAA.



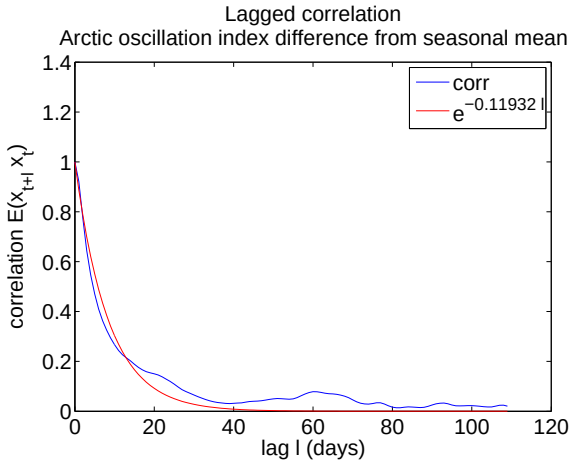
(a) raw series



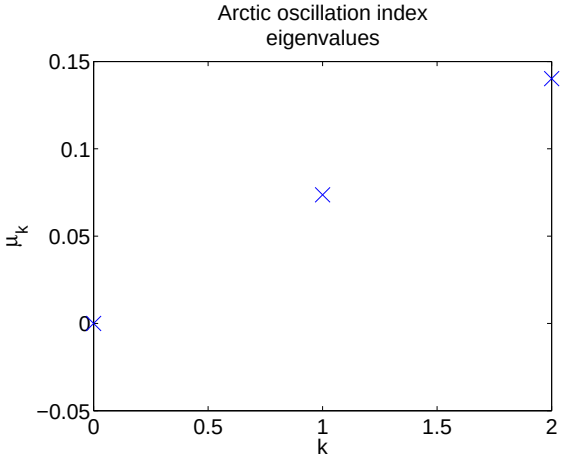
(b) anomalies



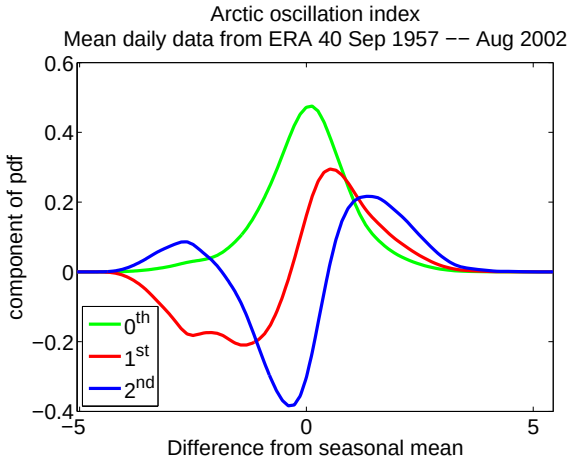
(c) cycle



(d) auto-correlation



(e) evalues



(f) efunctions

Figure 4.6: ERA-40 Arctic Oscillation index, calculated from approx 45 years of reanalysis. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

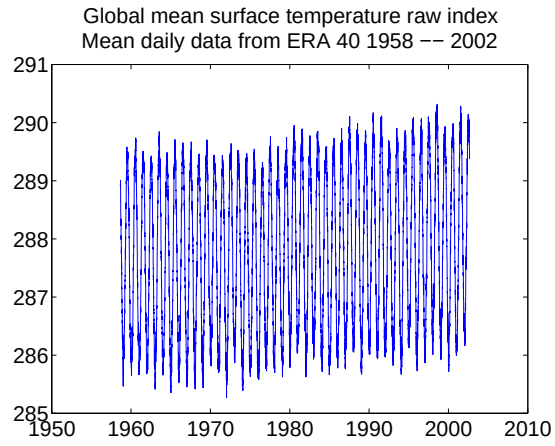
4.3.3 ERA-40 surface air temperature

The Surface Air Temperature index is the area-weighted global mean near-surface air temperature. The area weighting is the average of the cosines of the latitude ± 1.25 degrees. As for the Niño 3.4 index, the seasonal cycle for the surface air temperature is obtained by taking the mean (over the 45 years of reanalysis) for each day of the year. Days are indexed from 1 Sep (the first date in the series) so, in a leap year, 31 August is ignored. The anomaly index is the difference between the index for a date and the mean for that day of the year. In a leap year, the anomaly for 31 August is the difference between the index for 31 August and the mean index for 1 September.

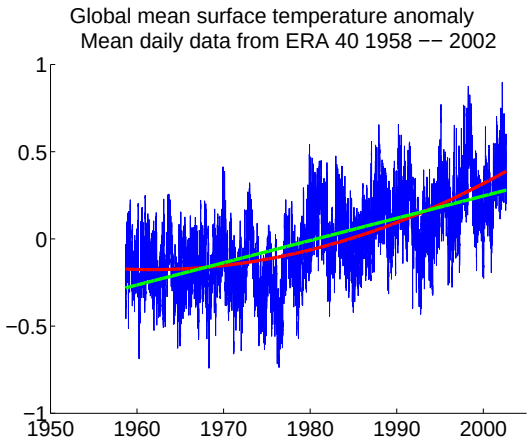
The surface air temperature index is calculated over land and sea, but temperature varies over land more than over sea due to the greater effective thermal capacity of the sea [44]. The northern hemisphere has a smaller sea surface area than the southern hemisphere, and so has a greater temperature variability. Thus the variability of the global index is dominated by the northern hemisphere, and the global seasonal cycle peaks in the northern hemisphere summer, as shown in figure 4.7.

Figure 4.7 shows an increasing trend in the temperature anomaly, corresponding to the well-known global warming over the twentieth century, consistent with, for example [38] in figure 4.8. This trend is very unlikely to be a feature of the system's natural variability, and is 'very likely' to be a feature of the system's drift [41], the result of a change in the radiative forcing on the system. There is 'very high confidence' that the radiative forcing increased over the twentieth century [41], and so the assumption in equation (1.1) of a constant drift does not hold. But the system's response to the change in forcing over the twentieth century gives information about its sensitivity, just as its response to random changes gives information about its sensitivity. So an area of useful future research would be to extend the theory behind (1.38) to allow for a time-dependant change in drift e.g. a time-dependant change in forcing.

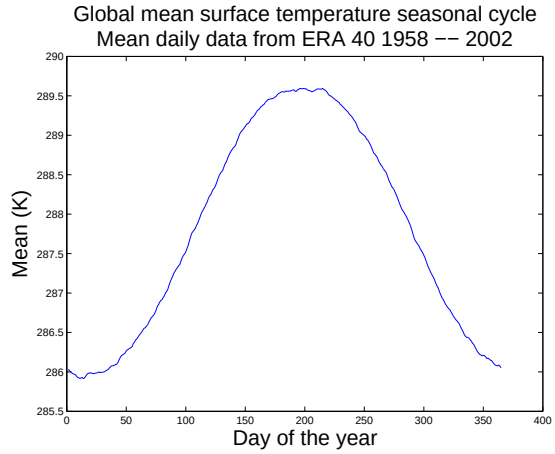
More immediately, a quadratic trend (red line in figure 4.7b) has been obtained by the method of least squares, and been removed from the series to form a de-trended series, shown in figure 4.9.



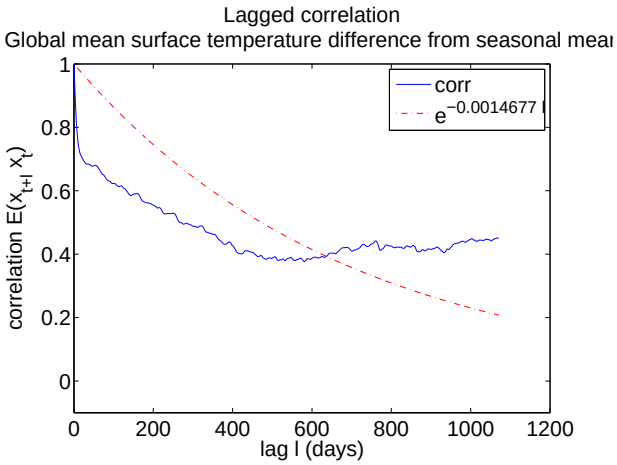
(a) raw series



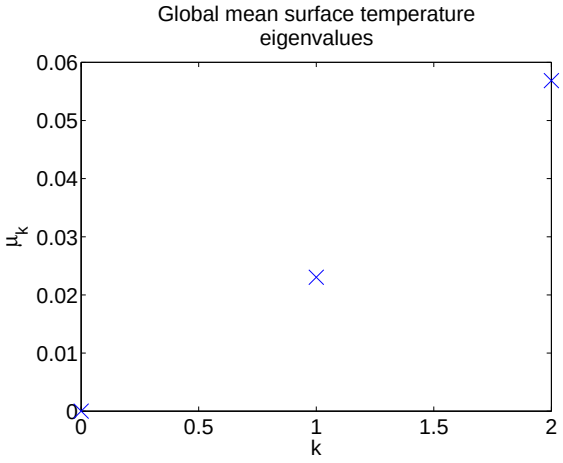
(b) anomalies



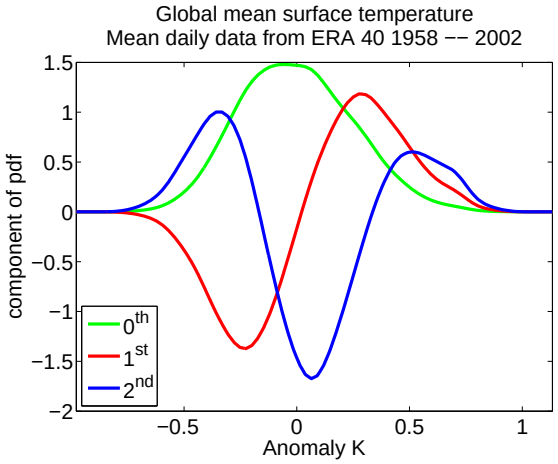
(c) cycle



(d) auto-correlation

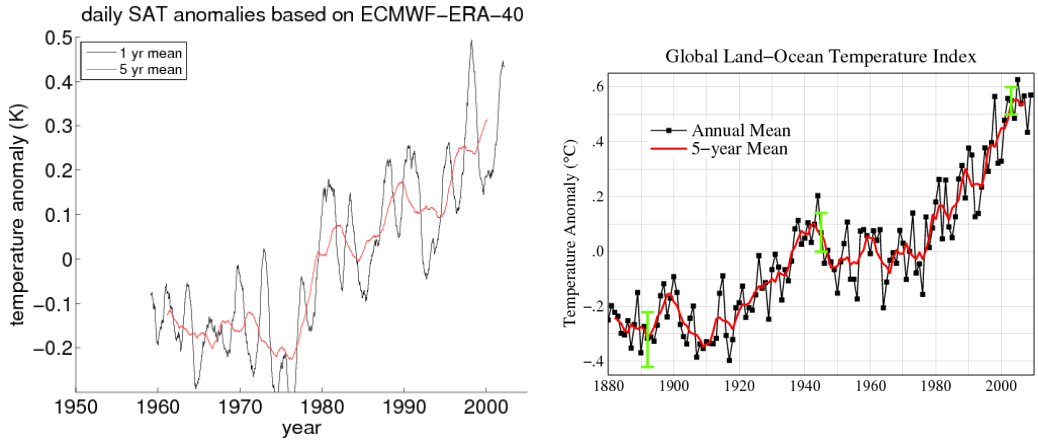


(e) evalues



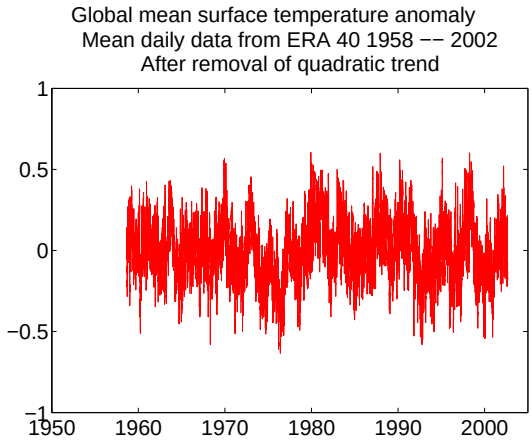
(f) efunctions

Figure 4.7: ERA-40 Surface Air temperature index, calculated from 45 years of reanalysis. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

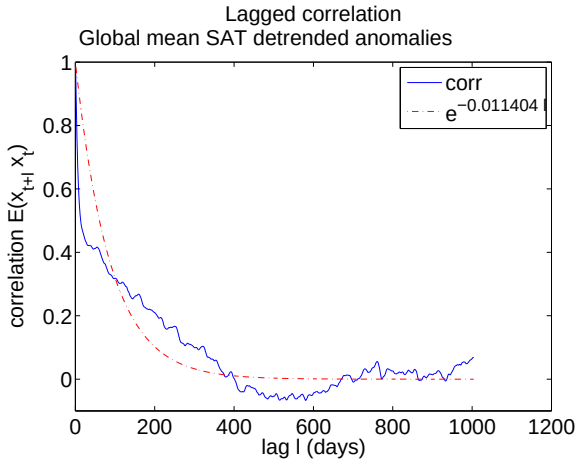


(a) ERA-40 (b) NASA

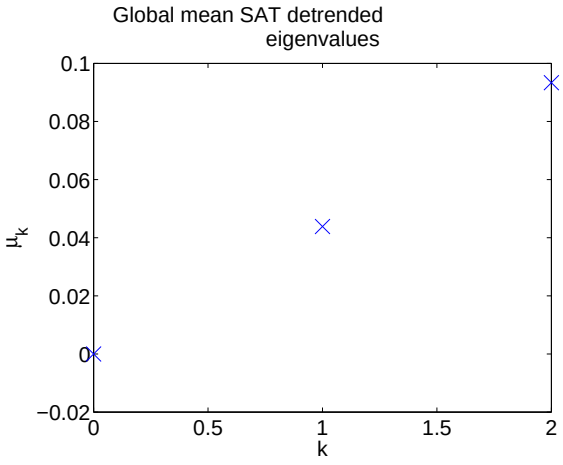
Figure 4.8: Annual and Five-year means of global surface temperature anomaly, ERA-40 vs NASA. Note that the NASA graph covers twice the period of the ERA-40 graph and has a different reference mean. The NASA graph is from [38] which is an update of [21].



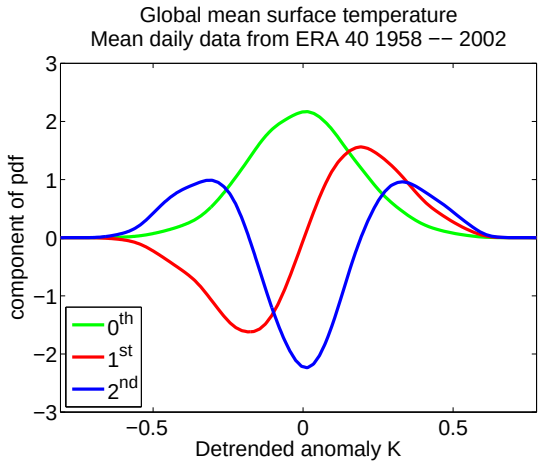
(a) detrended anomalies



(b) auto-correlation of detrended anomalies



(c) evalues of detrended anomalies

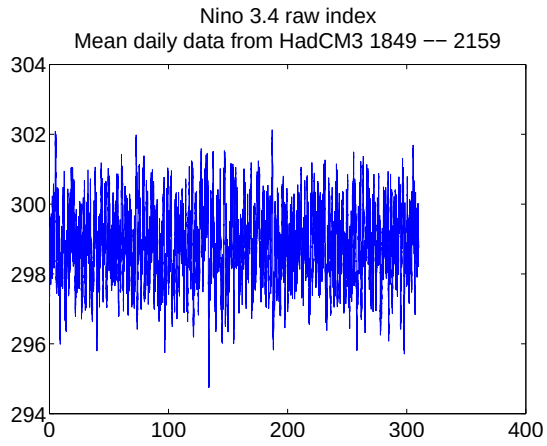


(d) efunctions of detrended anomalies

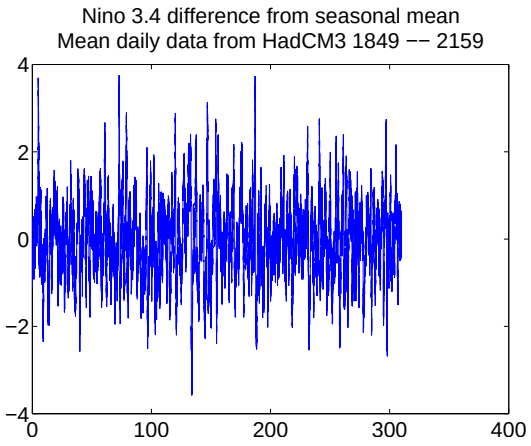
Figure 4.9: ERA-40 Global mean surface air temperature index. (a) detrended anomaly, which is excess over seasonal mean minus the quadratic trend. (b) autocorrelation function of the detrended anomaly. (c) estimated eigenvalues of the Fokker-Planck equation corresponding to the detrended anomaly. (d) estimated eigenfunctions of the Fokker-Planck equation corresponding to the detrended anomaly. Once the trend is removed the autocorrelation function becomes more similar to the autocorrelation function of the HadCM3 surface air temperature anomaly in figure 4.13.

4.3.4 HadCM3 Nino 3.4 index

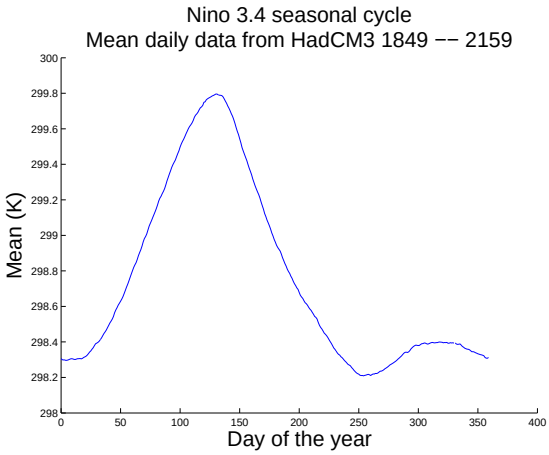
The HadCM3 Niño 3.4 index, shown in figure 4.10, is calculated similarly to the ERA-40 Niño 3.4 index in 4.3.1. The HadCM3 data has a coarser longitudinal resolution, so the longitudes over which the Sea Surface temperatures are averaged are approximately 168.75W to 120W.



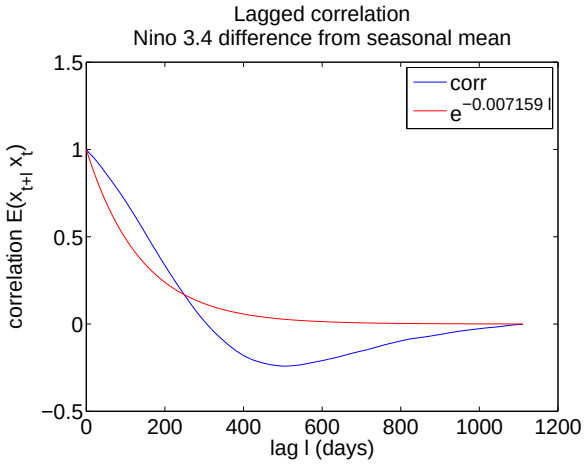
(a) raw series



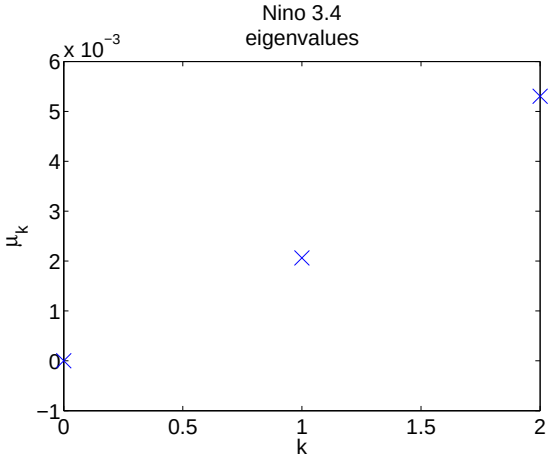
(b) anomalies



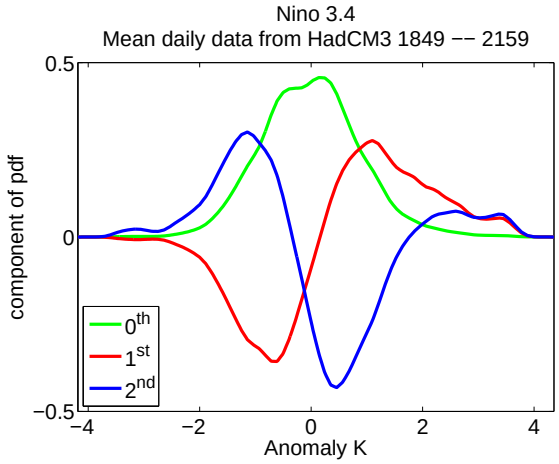
(c) cycle



(d) auto-correlation



(e) evalues

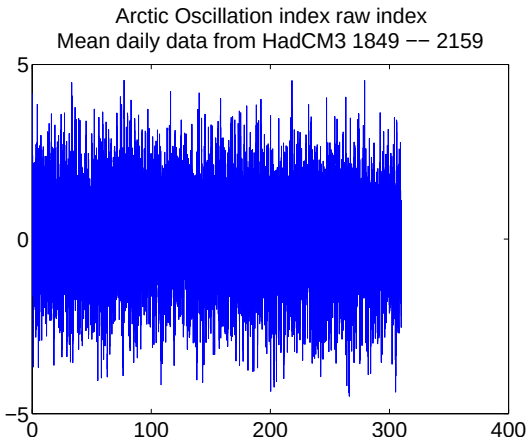


(f) efunctions

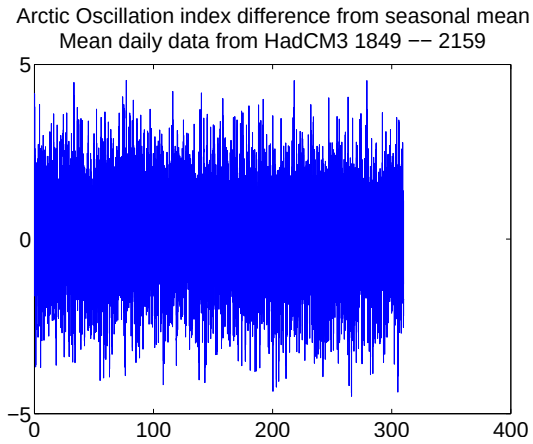
Figure 4.10: HadCM3 Nino3.4 index, calculated from approx 310 model years. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

4.3.5 HadCM3 Arctic Oscillation

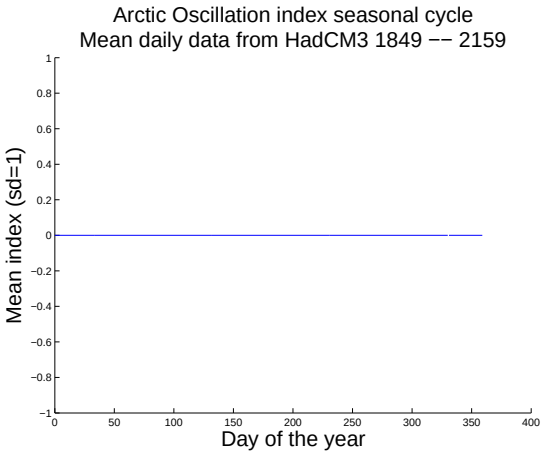
The HadCM3 Arctic Oscillation (AO) index, shown in figure 4.11 is calculated similarly to the ERA-40 AO index in 4.3.2. The only difference between the HadCM3 and ERA-40 calculation is that the HadCM3 series is longer, and has lower longitudinal resolution. The 'loading pattern' for the HadCM3 AO index [33] is the EOF for the HadCM3 winter surface pressure anomalies (shown in figure 4.12 to be similar to the loading pattern for the ERA-40 AO index). The HadCM3 AO index is the dot-product of this loading pattern with the daily area-weighted surface pressure anomaly.



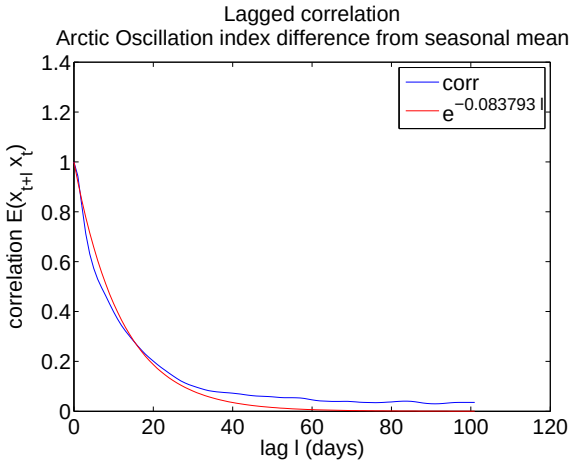
(a) raw series



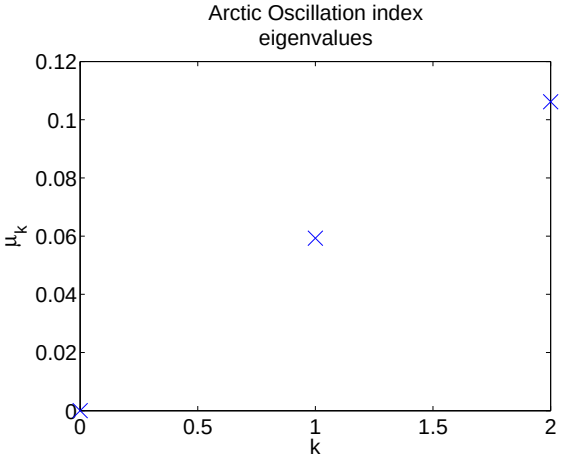
(b) anomalies



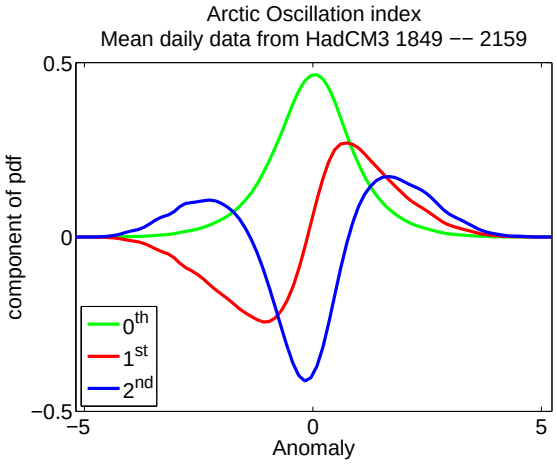
(c) cycle



(d) auto-correlation

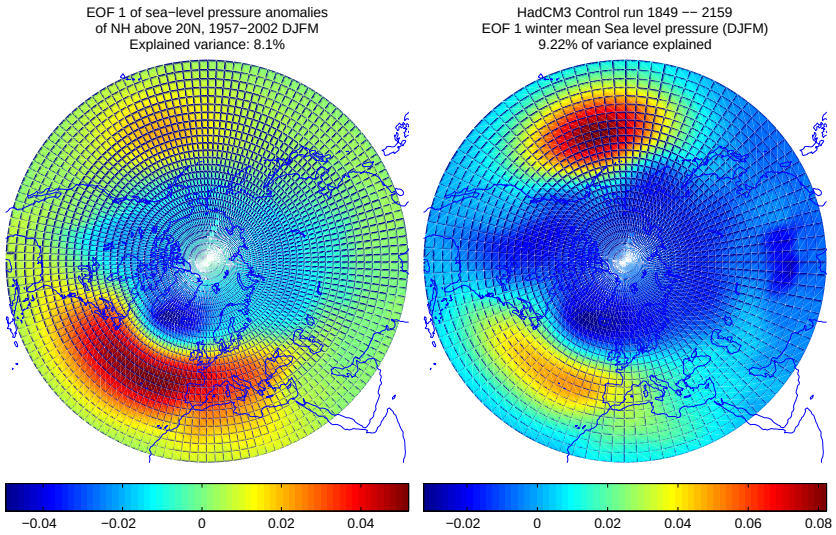


(e) evalues



(f) efunctions

Figure 4.11: HadCM3 Arctic Oscillation index, calculated from 310 model years. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

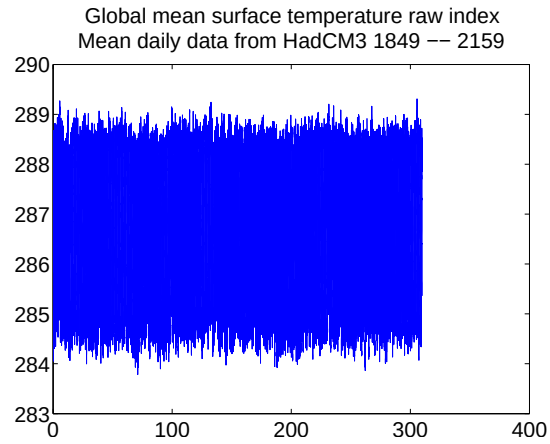


(a) ERA-40 (b) HadCM3

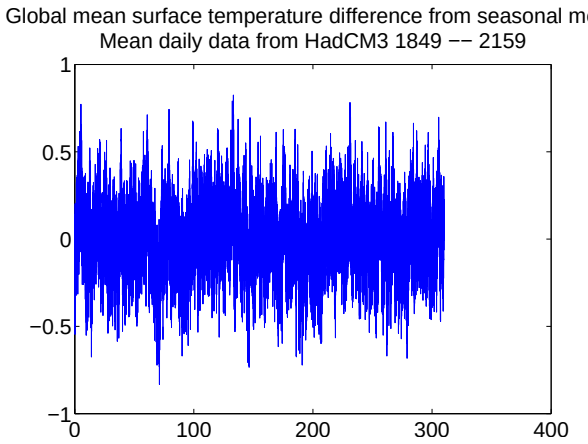
Figure 4.12: Dominant empirical orthogonal function (EOF1) of Mean sea level pressure anomalies, based on ERA-40, is similar to the corresponding EOF1 based on HadCM3.

4.3.6 HadCM3 surface air temperature

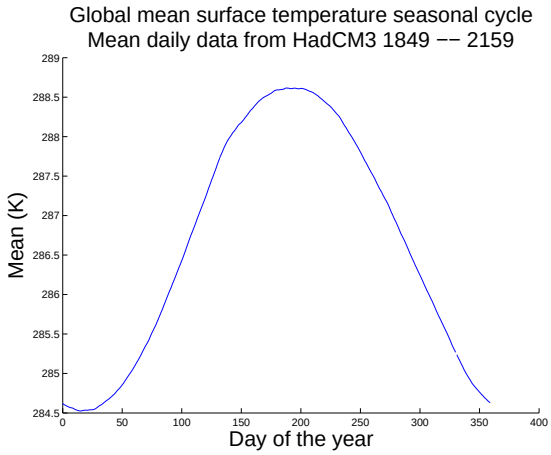
The HadCM3 surface air temperature (SAT) index, shown in figure 4.13, is calculated with the same approach as was used for the ERA-40 SAT index in 4.3.3. The HadCM3 index does not exhibit any trend as it uses a pre-industrial level of radiative forcing throughout.



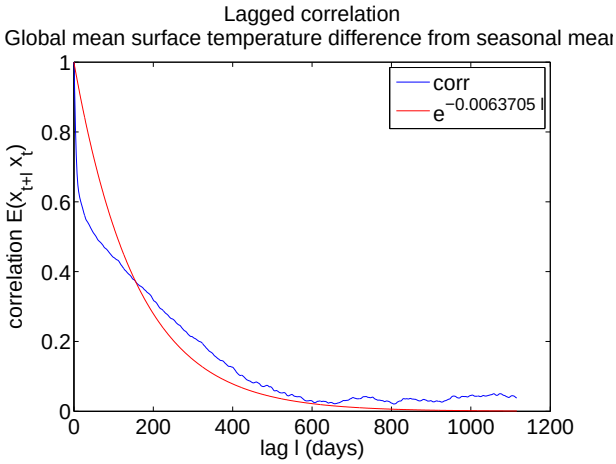
(a) raw series



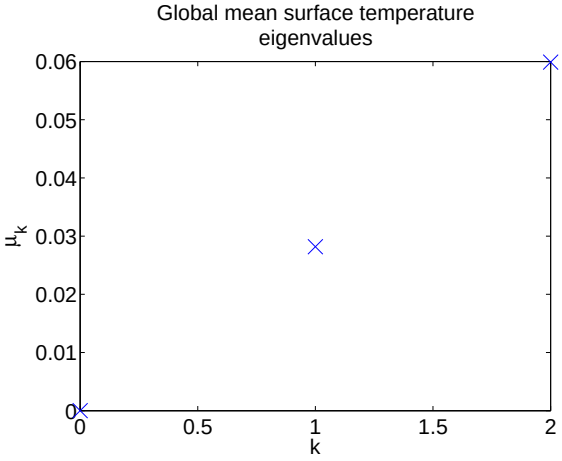
(b) anomalies



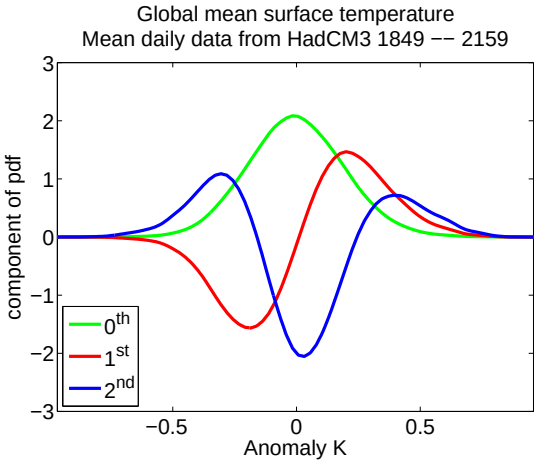
(c) cycle



(d) auto-correlation



(e) evalues



(f) efunctions

Figure 4.13: HadCM3 Surface Air temperature index, calculated from approx 310 model years. (a) actual index. (b) anomaly index, that is excess of index over seasonal mean. (c) seasonal mean (mean daily value). (d) autocorrelation of anomaly index, fitted to exponential. (e) estimated smallest eigenvalues of Fokker-Planck equation. (f) estimated dominant eigenfunctions of Fokker-Planck equation.

4.4 Comparison of series

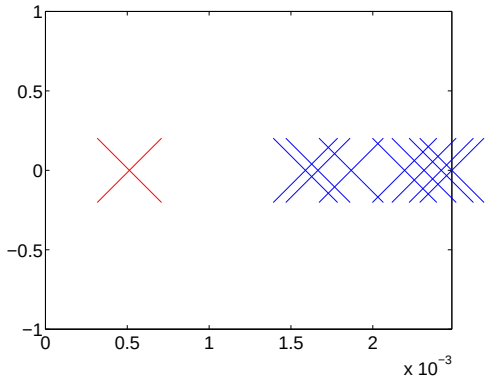
The anomaly time series were compared (for one dimension only) by splitting their state spaces into 80 equally probable bins, and then by counting bins without smoothing. This approach was taken in order to get the best estimate of differences in eigenvalues, as it was observed in section 3.4.2 that smoothing worsens the estimate of eigenvalues.

When two-dimensional series are compared, the eigenvalues are calculated by splitting the state space for each single series into 15 equally probable bins, and then forming 225 two-dimensional bins from the pairs of one-dimensional bins. The two-dimensional bins are not equally probable unless the one-dimensional series are uncorrelated. Bin counting was used with no smoothing to estimate the transition matrix for the 225 discrete bins.

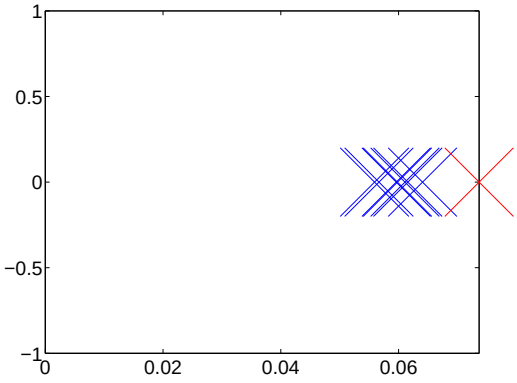
4.4.1 Eigenvalues for single series

Figures 4.2–4.13 show that the eigenvalues based on the transition matrices for the single indices of Nino 3.4, Arctic Oscillation index, and global surface air temperature, are significantly different between ERA40 and HadCM3. The eigenvalues that most differ are for the Niño index, which has the smallest eigenvalue and is therefore subject to the greatest sampling error or variance, consistent with section 3.4.3 and [19].

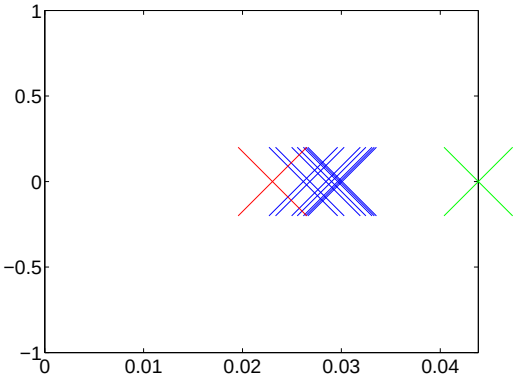
Figure 4.14 shows the 7 eigenvalues from 7 different sections of the HadCM3 time series, obtained by splitting the HadCM3 series into series of the same length as ERA-40. These are compared with the corresponding eigenvalue from ERA-40. The 7 eigenvalues give a range of variation, such that any sampled eigenvalue which falls within that range could clearly be from a statistically identical system. If a sampled eigenvalue falls outside that range, then it is still possible that it came from a statistically identical system, but this is less likely, the further it gets outside the range. In order to quantify this likelihood it would be necessary to state an explicit model for the distribution of eigenvalues. We have not modelled the distribution of eigenvalues explicitly, but observe that the eigenvalues based on each ERA40 series lie outside the range of the eigenvalues based on HadCM3, suggesting a significant difference between the actual eigenvalues of the two systems.



(a) Nino 3.4



(b) Arctic Oscillation

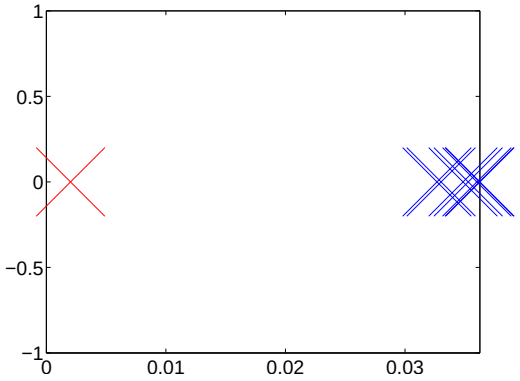


(c) Surface Air Temperature

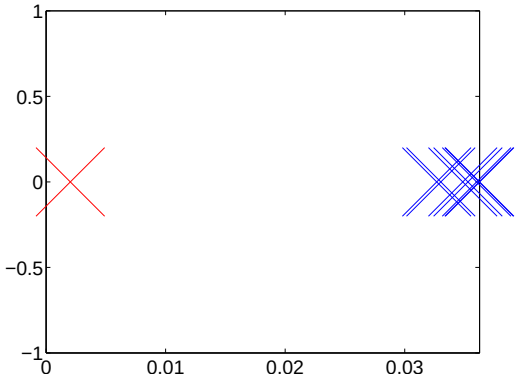
Figure 4.14: Smallest non-zero eigenvalues for Fokker-Planck equations based on single series. Blue crosses are for HadCM3 control runs. Red cross is for ERA 40. Green cross in graph (c) is for detrended ERA 40. (Vertical axes are meaningless).

4.4.2 Eigenvalues for pairs of series

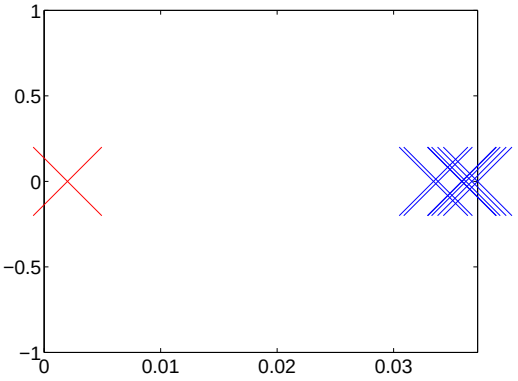
Figure 4.15 shows the corresponding eigenvalues if two series are taken at a time. That is, figure 4.15 shows the eigenvalues for the two-dimensional Fokker-Planck equations from the same 7 different sections of the HadCM3 time series. The eigenvalues based on ERA-40 are outside the range of sampled eigenvalues based on HadCM3, and the scale of the differences is much greater than the differences based on the one-dimensional series.



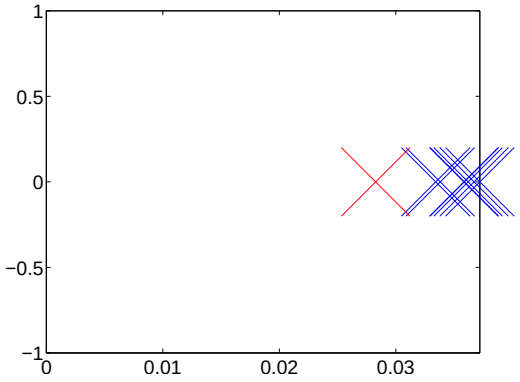
(a) Nino and Surface Air Temperature



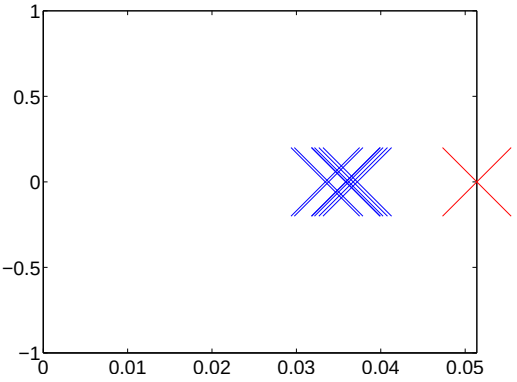
(b) Nino and detrended Surface Air Temperature



(c) Nino and Arctic Oscillation



(d) Arctic Oscillation and Surface Air Temperature



(e) Arctic Oscillation and detrended Surface Air Temperature

Figure 4.15: Smallest non-zero eigenvalues for Fokker-Planck equations based on two series. Blue crosses are for HadCM3 control runs. Red cross is for ERA 40. (Vertical axis is meaningless). Left column is compares ERA-40 with HadCM3 control run. Right column compares ERA-40 (with the quadratic trend removed from the surface air temperature anomaly) with HadCM3 control run. The difference between eigenvalues involving the Niño series persists after the trend in surface air temperature is removed.

4.5 Conclusion

There is an adequate match between the main features of the calculated indices and those calculated by others using different datasets, so differences in the statistics are features of the method of calculating them (or of the system itself) but are not artefacts of the method of calculating the indices.

Based on the autocorrelation functions (and consistent with [6]), the Niño index cannot be modelled as a linear Langevin system since neither the model nor the actual autocorrelation functions decay simply to zero as they would for a linear system. Of the three indices examined, the Arctic Oscillation index shows, via its autocorrelation function, the most similarity to a linear 1-dimensional Langevin system.

The autocorrelation function of the Surface Air Temperature series shows a persistent memory for the ERA-40 series (and not for the HadCM3 series), though this is due to the global warming trend, as the feature disappears for the detrended series. The (detrended) ERA-40 SAT series and HadCM3 control run SAT series show two timescales: the autocorrelation falls very quickly initially (showing a short-term memory of the order of days) but a longer-term autocorrelation persists.

Certain differences between the series are only brought out by the eigenvalue analysis. For example, the Niño series have similar autocorrelation functions, but have eigenvalues that differ by a factor of 4. Significant differences exist between the eigenvalues of the model and actual systems. Unlike in the linear case in section 3.5, these differences are accentuated by comparing two series at a time.

However, the theory needs to be developed further in order to clarify how differences in the eigenvalues represent a different response to forcing, and in particular, to the radiative forcings that are of genereal interest. A lower eigenvalue may represent a system which is harder to vary, but which, when varied, responds more stongly. However this is not enough to conclude that the real and model systems respond differently to a particular forcing.

Future work could usefully develop the theory in chapter 1 and more firmly establish

- how it could be verified to what extent the real system (or even a model) is of the form in (1.1);
- how the theory could be extended to systems with different kinds of noise e.g.

multiplicative rather than additive;

- the adaptation of the theory to systems with varying forcing;
- the circumstances, if any, under which $\Delta\mathbf{M}$ in (1.38) can be calculated independently of a model, and hence
- the significance (in terms of responses to changes in forcings) of differences in the eigenvalues and eigenvectors of different models' transition matrices.

References

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications, New York, ninth Dover printing, tenth GPO printing edition. 1046 pages.

[2] Ambaum, M. H. P., Hoskins, B. J., and Stephenson, D. B. (2001). Arctic Oscillation or North Atlantic Oscillation? *Journal of Climate*, **14**(16), 3495–3507.

[3] Bell, T. L. (1980). Climate Sensitivity from Fluctuation Dissipation: Some Simple Model Tests. *Journal of the Atmospheric Sciences*, **37**(8), 1700–1707.

[4] British Atmospheric Data Centre (2009). El Nino Southern Oscillation (ENSO). `http://badc.nerc.ac.uk/data/higem/xbyld/nino3.html` accessed 2009.

[5] Brohan, P. (2008). Are we too good? Met Office Seminar.

[6] Burgers, G. (1999). The El Nino stochastic oscillator. *Climate Dynamics*, **15**(7), 521–531.

[7] Chan, K. C. (1988). On the Contrarian Investment Strategy. *Journal of Business*, **61**(2), 147.

[8] Chang, P., Zhang, L., Saravanan, R., Vimont, D., Chiang, J., Ji, L., Seidel, H., and Tippett, M. (2007). Pacific meridional mode and El Nino-southern oscillation. *Geophysical Research Letters*, **34**(16), 16608.

[9] Collins, M., Tett, S. F. B., and Cooper, C. (2001). The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **17**(1), 61–81.

[10] Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D. (2006). Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, **87**(2), 175–190.

[11] Cox, P. M. and Stephenson, D. B. (2007). Climate Change: a Changing Climate for Prediction. *Science*, **317**(5835), 207.

[12] Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**(6809), 184–187.

[13] de Verdiere, A. C. and Huck, T. (2000). A 2 degree of freedom dynamical system for interdecadal oscillations of the ocean-atmosphere. *Journal of Climate*, **13**(15), 2801–2817.

[14] Dessler, A. E. and Sherwood, S. C. (2009). A matter of humidity. *Science*, **323**, 1020–1021.

[15] Egger, J. (2001). Master equations for climatic parameter sets. *Climate Dynamics*, **18**(1/2), 169–177.

[16] Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R., and Webster, M. D. (2002). Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**(5552), 113–117.

[17] Gillett, N. P., Allen, M. R., McDonald, R. E., Senior, C. A., Shindell, D. T., and Schmidt, G. A. (2002). How linear is the Arctic Oscillation response to greenhouse gases. *J. Geophys. Res*, **107**(D3), 4022–4028.

[18] Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A., and Rayner, N. A. (2002). An Observationally Based Estimate of the Climate Sensitivity. *Journal of Climate*, **15**(22), 3117–3121.

[19] Gritsun, A. and Branstator, G. (2007). Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *Journal of the Atmospheric Sciences*, **64**(7), 2558–2575.

[20] Hall, A. and Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett*, **33**, L03502.

[21] Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., and Medina-Elizade, M. (2006). Global temperature change. *Proceedings of the National Academy of Sciences*, **103**(39), 14288–14293.

[22] Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, **28**(6), 473–485.

[23] Jolliffe, I. T. and Stephenson, D. B. (2003). *Forecast verification: a practitioner's guide in atmospheric science*. Wiley. 254 pages.

[24] Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer. 636 pages.

[25] Knutti, R., Meehl, G. A., Allen, M. R., and Stainforth, D. A. (2006). Constraining climate sensitivity from the seasonal cycle in surface temperature. *Journal of Climate*, **19**(17), 4224–4233.

[26] Kohfeld, K. E. and Harrison, S. P. (2000). How well can we simulate past climates? Evaluating the models using global palaeoenvironmental datasets. *Quaternary Science Reviews*, **19**(1-5), 321–346.

[27] Lanczos, C. (1961). *Linear Differential Operators*. Society for Industrial Mathematics. 564 pages.

[28] Leith, C. E. (1975). Climate Response and Fluctuation Dissipation. *Journal of the Atmospheric Sciences*, **32**(10), 2022–2026.

[29] Lomborg, B. (2007). *Cool it: the skeptical environmentalist's guide to global warming*. Vintage Books. 234 pages.

[30] Marotzke, J. and Stone, P. H. (1995). Atmospheric Transports, the Thermohaline Circulation, and Flux Adjustments in a Simple Coupled Model. *Journal of Physical Oceanography*, **25**(6), 1350–1364.

[31] Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**(7001), 768–72.

[32] National Oceanic and Atmospheric Administration (2009). Monthly nino 3.4 indices. `http://www.esrl.noaa.gov/psd/forecasts/sstlim/global/indices_global` accessed March 2010.

[33] National Weather Service Climate Prediction Center (2009). Arctic oscillation index. http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index. accessed December 2009.

[34] North, G. R., Bell, R. E., and Hardin, J. W. (1993). Fluctuation dissipation in a general circulation model. *Climate Dynamics*, **8**(6), 259–264.

[35] Osborn, T. J. (2004). Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing. *Climate Dynamics*, **22**(6), 605–623.

[36] PALMER, T. N. (2001). A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, **127**(572), 279–304.

[37] Reifen, C. and Toumi, R. (2009). Climate projections: Past performance no guarantee of future skill? *Geophysical Research Letters*, **36**(13), L13704.

[38] Sato, M. (2009). Global annual mean surface air temperature change. "`http://data.giss.nasa.gov/gistemp/graphs/`". accessed 2009.

[39] Shukla, J., DelSole, T., Fennessy, M., Kinter, J., and Paolino, D. (2006). Climate model fidelity and projections of climate change. *Geophys. Res. Lett*, **33**.

[40] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC. 178 pages.

[41] Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (2007). *IPCC, 2007: 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. 996 pages.

[42] Stern, N. H. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge University Press. 692 pages.

[43] Strogatz, S. H. and Herbert, D. E. (1994). *Nonlinear dynamics and chaos*. Addison-Wesley Reading, MA. 498 pages.

[44] Sutton, R. T., Dong, B., and Gregory, J. M. (2007). Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations. *Geophysical Research Letters*, **34**(2), 2701.

[45] 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**, 1297–1300.

[46] Thuburn, J. (2005a). personal communication.

[47] Thuburn, J. (2005b). Climate sensitivities via a Fokker-Planck adjoint approach. *Quarterly Journal of the Royal Meteorological Society*, **131**(605), 73–92.

[48] UK Meteorological Office, Hadley Centre (2006–2010). Hadcm3 control run model data. British Atmospheric Data Centre. Available from http://badc.nerc.ac.uk/data/hadcm3-control.

[49] Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., *et al.* (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**(612), 2961–3012. Data available from http://badc.nerc.ac.uk/data/ecmwf-e40/.

[50] Wigley, T. M. L. and Raper, S. C. B. (1990). Natural variability of the climate system and detection of the greenhouse effect. *Nature*, **344**, 324–327.

[51] Woollings, T., Hannachi, A., Hoskins, B., and Turner, A. (2010). A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing. *Journal of Climate*, **23**, 1291–1307.

Part II

Emergent dynamics of the climate-economy system

Chapter 5

Emergent dynamics of the climate-economy system

5.1 Introduction

It is widely accepted that climate change will have major impacts on humankind. Depending on the magnitude of 21st century climate change, negative impacts are expected on water, food, human health [20, 11] and ultimately economic growth [19, 28]. Global CO_2 emissions, which are the largest contributor to anthropogenic climate change [29], have, to date, been highly correlated with economic output [2]. As a result there is a negative feedback between climate change and economic growth that is mediated by CO_2 emissions: an increase in human wealth causes an increase in emissions and global warming, but the warming damages human wealth, slowing its rise or even making it fall.

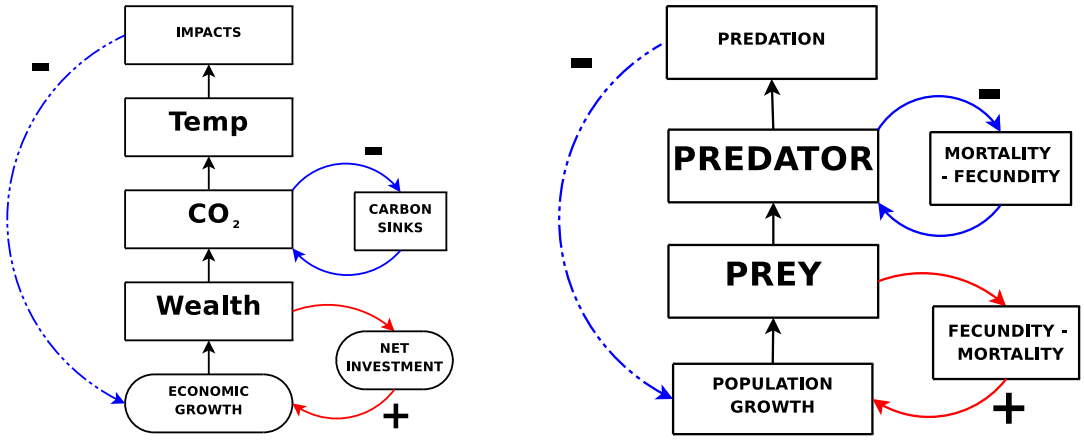
This climate-economy feedback is typically neglected in a standard climate change assessment [27], which is largely a serial process going from socioeconomic scenarios to emissions to climate change to impacts [4]. Integrated assessment models, e.g. [19] do include the feedback but only weakly. A feasible sensitivity of the economy to the climate results in important emergent properties, which are the subject of this chapter. Dangerous rates of change [25] can be defined as those rates that cause instabilities or long-term boom-bust oscillations, thereby preventing a 'soft-landing' of the climate-economy system. A soft-landing would be a process leading smoothly to an equilibrium.

Using a simple model of the coupled climate-economy system this chapter derives formulae for the critical rates of growth of global CO_2 emissions which define the edges of stable and unstable regimes. On the basis of this model, historical rates of economic growth and decarbonisation, which together have led to the historical rates of growth of CO_2 emissions, put the climate-economy system in a dangerous boom-bust regime.

The model is defined in section 5.2 and the stability of its equilibria are analysed in section 5.3. The model is fitted to the 20^{th} century experience in section 5.4 and projected in section 5.5. The model is compared with the well known DICE model [19] in section 5.6.

5.2 Model definition

The model presented here describes the global human-environment system with just 3 state variables: atmospheric CO_2 concentration ($\hat{C}$), global warming ($\hat{T}$), and global wealth ($\hat{W}$), interdependent as in figure 5.1a.



(a) Climate-economy model (b) Predator-prey model

Figure 5.1: Schematic of climate-economy model, with predator-prey model for comparison. Red lines indicate positive feedbacks and blue lines indicate negative feedbacks. On the left, the blue dot-dash line is the climate change impact on the economy, which is the main subject of this chapter.

Schematically (see figure 5.1b) the model has a similar form to a predator-prey model [32]. Global wealth has the role of the prey. It supplies the 'predator' of pollution and is

reduced by the pollution's impacts.

5.2.1 Model dynamics

Global warming is assumed to increase with the atmospheric CO_2 concentration according to the standard logarithmic dependence on CO_2 [10, 7], moderated by the extra outgoing radiation from the higher temperature on earth. Equilibrium climate sensitivity [8] is ΔT_{2*CO2}(for a doubling of CO_2 from an assumed pre-industrial level of C_{PI}), approached on a characteristic climate timescale of τ_T, which is set by the thermal capacity of the oceans.

$$\frac{\mathrm{d}\hat{T}}{\mathrm{d}\hat{t}} = \frac{1}{\tau_T}\left\{\frac{\Delta T_{2*CO2}}{\ln 2}\ln(\hat{C}/C_{PI}) - \hat{T}\right\}. \tag{5.1}$$

Atmospheric carbon dioxide ($\hat{C}$) increases in proportion to global CO_2 emissions ($\hat{E}$), but the excess $\hat{C}$ above the preindustrial level is reduced by the combined effect of land and ocean carbon sinks with an assumed characteristic timescale of τ_C years. In reality CO_2 is removed from the atmosphere on a large range of timescales [1], but a single effective timescale of τ_C=50 is consistent with the historical airborne fraction [23].

$$\frac{\mathrm{d}\hat{C}}{\mathrm{d}\hat{t}} = \hat{E} - \frac{1}{\tau_C}\left(\hat{C} - C_{PI}\right). \tag{5.2}$$

CO_2 emissions $\hat{E}$ increase with global wealth $\hat{W}$, which is human and material capital. Initially, $\hat{E}$= $\chi\hat{W}$, where χ is a constant carbon intensity, that is the amount of CO_2 required to service each unit of wealth. Consistent with historical records of emissions in [15, 17, 3], the carbon intensity is assumed to fall exponentially over time by a constant, positive decarbonisation rate of μ per year, so that after Y years, one unit of wealth can be serviced by CO_2 emissions of $e^{-\mu Y}$ times the amount currently required. If $\hat{W}$ grows faster than at μ per year, then emissions will increase, that is

$$\hat{E} = \chi e^{-\mu\hat{t}}\hat{W}. \tag{5.3}$$

Global wealth ($\hat{W}$) grows through net investment in social capital, technology, productivity [9, 14], and is shown as the positive feedback loop on the right-hand-side of Figure 5.1a. An increase in $\hat{W}$ comes from world product exceeding world consumption and depreciation. Within the model, $\hat{W}$ is theoretically infinite, only constrained by the condition of natural resources, i.e. by $\hat{T}$. The model simplifications imply that there are

sources (of resources) and sinks (for waste) with infinite capacity. In reality, both are limited [16]. In particular, a finite fossil fuel reserve could be modelled, with the cost of fossil fuel increasing as the reserve fell. An increasing real cost of fossil fuel would require a higher proportion of resources to be allocated to production of energy, and a lower proportion to be allocated to production of goods and services, so that the rate of increase in global wealth would fall as the fossil fuel reserve fell (which would be as the integral of emissions over time rose). However, for simplicity, effectively infinite fossil fuel reserves with a constant price of extraction have been assumed, so that the rate of increase in wealth does not depend on the size of wealth, nor on its integral over time.

The Kaya identity [22] separates the components of emissions into ratios so that emissions are a product of population, GDP per person and carbon intensity of GDP. Population is not modelled explicitly in this model. An increase in global wealth could be the result of greater wealth per person, or of a rising population with static or falling wealth per person.

The climate-economy feedback loop is closed by assuming that global warming suppresses economic growth. An extensive literature (surveyed in e.g. [28, 21]) estimates the total impacts of climate change from the 'bottom-up' – aggregating the different impacts (water stress, sea-level rise, weather extremes etc) on different regions (low-lying, developing, industrial etc) and sectors (food supply, human health, security etc). Parry *et al.* [21] concludes that the degree to which economic growth is affected by climate change depends on the projected development pathway. This implies damage that depends somehow on the time-integral of the level of wealth. Also, Parry *et al.* [21] suggests that the convexity of the damage function depends on inequalities of wealth within the population. However, for the sake of simplicity we have neglected inequalities by using a single variable for global wealth, and assumed that the net rate of growth in wealth, $\hat{\mathrm{x}}(\hat{T})$, depends only on the level of global warming, and falls as global warming rises. Similarly, we have assumed the rate of growth in wealth is independent of the pace of global warming, so that $\hat{\mathrm{x}}(\hat{T})$ does not depend on $\frac{\mathrm{d}\hat{T}}{\mathrm{d}t}$. Also, the proportional rate of growth $\frac{1}{\hat{W}}\frac{\mathrm{d}\hat{W}}{\mathrm{d}t}$ is assumed to be independent even of the size of wealth, so the model does not allow for climate change to have less proportionate impact if the level of wealth is high, and unlike e.g. [16] the model does not assume a carrying capacity.

The actual ranges of temperature allowing economic growth or forcing economic con-

traction are unknown, so $\hat{\mathrm{x}}(\hat{T})$ is only specified as far as assuming that at one (positive) level of global warming the rate of economic growth will fall to the decarbonisation rate μ, at which point, by (5.3), the growth of CO_2 emissions will be zero, though may not necessarily be stable.

$$\frac{\mathrm{d}\hat{W}}{\mathrm{d}\hat{t}} = \hat{\mathrm{x}}(\hat{T})\hat{W}; \tag{5.4}$$

$$\exists \hat{T}_e > 0 : \hat{\mathrm{x}}(\hat{T}_e) = \mu \tag{5.5}$$

$$\frac{\mathrm{d}\hat{\mathrm{x}}}{\mathrm{d}\hat{T}} < 0. \tag{5.6}$$

5.2.2 Non-dimensional form of model

In non-dimensional variables, the model is autonomous:

$$\dot{C} := \frac{\mathrm{d}C}{\mathrm{d}t} = E - \phi C, \tag{5.7}$$

$$\dot{T} = \log(1 + C) - T, \tag{5.8}$$

$$\dot{E} = \mathrm{x}(T)E, \tag{5.9}$$

where

$$t = \frac{\hat{t}}{\tau_T}, \tag{5.10}$$

$$C = \frac{\hat{C}}{C_{PI}} - 1, \tag{5.11}$$

$$T = \frac{\hat{T}\log 2}{\Delta T_{2*CO2}}, \tag{5.12}$$

$$E = \frac{\tau_T \chi e^{-\mu\hat{t}}\hat{W}}{C_{PI}}, \tag{5.13}$$

$$\phi = \frac{\tau_T}{\tau_C}, \tag{5.14}$$

$$\mathrm{x}(T) = \tau_T \left\{ \hat{\mathrm{x}}(\hat{T}) - \mu \right\}. \tag{5.15}$$

5.3 Model equilibria

Without making any approximations, the model has two equilibrium points, i.e. combinations of C, T, E which are in balance, and so can (in theory) be permanent. These

points are obtained by setting (5.7)–(5.9) to zero and solving for C, T, E. The equilibrium points are (0,0,0) and, using the assumptions in equations (5.5–5.6), the positive equilibrium

$$\mathbf{s}_{\mathrm{e}} \equiv \begin{pmatrix} C_e \\ T_e \\ E_e \end{pmatrix} = \begin{pmatrix} e^{T_e} - 1 \\ \mathrm{x}^{-1}(0) \\ \phi\left(e^{T_e} - 1\right) \end{pmatrix}. \tag{5.16}$$

At the equilibrium level of emissions, CO_2 concentration and global warming are constant. From (5.13) $\hat{W} \propto e^{\mu \hat{t}} E$ so that an equilibrium level of E means $\hat{W}$ increases exponentially at rate μ.

5.3.1 Zero equilibrium is unstable

The zero equilibrium (no emissions, a preindustrial level of CO_2 and no warming) is unstable. A small level of emissions grows exponentially at rate x(0) without (initially) any significant impact on $\hat{T}$, because the accumulated emissions are initially small, so the radiative forcing is small, and the increase in temperature only emerges over the timescale τ_T.

5.3.2 Stability of positive equilibrium

The positive equilibrium may be stable or unstable. If it is stable, differences from the equilibrium get smaller over time, so that configurations of the variables C, T, E which are only slightly different from the equilibrium configuration, tend over time, towards the equilibrium. If it is unstable, differences from the equilibrium get larger over time, so that the equilibrium is practically unattainable.

Whether the equilibrium is stable or unstable depends on the relative timescales for economic growth or decay, global warming and the carbon cycle. In dimensionless variables, the stability depends on the relationship between x and ϕ.

$$\text{For, let } \sigma = \sigma(\mathrm{x}) = -\frac{\mathrm{dx}(T_e)}{\mathrm{d}T}\left\{1 - e^{-T_e}\right\} \tag{5.17}$$

$$= -\frac{\mathrm{dx}}{\mathrm{d}T}\Big|_{\mathrm{x}^{-1}(0)} \left\{1 - e^{-\mathrm{x}^{-1}(0)}\right\}, \text{ by (5.16).} \tag{5.18}$$

By (5.5), (5.6) and (5.15)

$$\sigma > 0. \tag{5.19}$$

Then (proven in section 5.3.3) the equilibrium is stable if and only if

$$\sigma < 1 + \phi. \tag{5.20}$$

Equations (5.18) and (5.20) show that the stability of the equilibrium depends entirely on the damage function x and the relative timescales of the warming and carbon cycles, ϕ. The equilibrium gets less stable the higher the level of warming at which emissions can continue to grow, and the more severe the change in the damage function near the equilibrium. In other words, the slacker the control, but the more suddenly it is applied, the less stable is the equilibrium. This confirms the idea of instability being a function of delays in responses to an oncoming limit [16]. There are many physical analogies, e.g. when braking smoothly and early, or suddenly and at the last moment.

Even if the equilibrium is stable, the system oscillates on its way to achieving the equilibrium unless (proven in section 5.3.3)

$$\sigma < \frac{1}{\phi} pq^2, \tag{5.21}$$

$$\text{where } pq^2 = \frac{1}{27}\left(-2 + 3\phi + 3\phi^2 - 2\phi^3 + 2\left(1 - \phi + \phi^2\right)^{3/2}\right). \tag{5.22}$$

It follows (estimated in (5.61)) that the system does not oscillate near the equilibrium if

$$\sigma < \frac{1}{8}\min\left(1, \phi\right) \tag{5.23}$$

and does oscillate near the equilibrium if

$$\sigma > \frac{1}{4}\min\left(1, \phi\right). \tag{5.24}$$

5.3.3 Proof of stability conditions

Jacobian of linearized system

Standard linear stability theory [12] proves equations (5.20) – (5.21). By the Hartman-Grobman theorem [30], the qualitative behaviour of small displacements from the equilibrium is the same as the behaviour of small displacements in the linearized system. Let $\mathbf{s}$(t) be the state of the system at time t and $\mathbf{s}_\mathrm{e}$ be the equilibrium point.

Let $\mathbf{s}_{\mathrm{d}}(\mathrm{t}) = \mathbf{s}(\mathrm{t}) - \mathbf{s}_{\mathrm{e}}$. Let $F = \dot{C}, G = \dot{T}$ and $H = \dot{E}$. Then let $\mathbf{J}$ be the Jacobian matrix evaluated at $\mathbf{s}_{\mathrm{e}}$:

$$\mathbf{J}(\mathbf{s}_{\mathrm{e}}) = \begin{pmatrix} \partial F/\partial C & \partial F/\partial T & \partial F/\partial E \\ \partial G/\partial C & \partial G/\partial T & \partial G/\partial E \\ \partial H/\partial C & \partial H/\partial T & \partial H/\partial E \end{pmatrix} \tag{5.25}$$

$$= \begin{pmatrix} -\phi & 0 & 1 \\ \frac{1}{1+C_e} & -1 & 0 \\ 0 & \frac{\mathrm{dx}(T_e)}{\mathrm{d}T} E_e & \mathrm{x}(T_e) \end{pmatrix} \tag{5.26}$$

$$= \begin{pmatrix} -\phi & 0 & 1 \\ e^{-T_e} & -1 & 0 \\ 0 & \frac{\mathrm{dx}(T_e)}{\mathrm{d}T} \phi \left\{ e^{T_e} - 1 \right\} & 0 \end{pmatrix}, \text{ by (5.16).} \tag{5.27}$$

If, as $t \to \infty$, $|\mathbf{J}^t\ \mathbf{s}_{\mathrm{d}}| \to 0$ so that the linearized system tends to the equilibrium point then $\mathbf{s}(\mathrm{t}) \to \mathbf{s}_{\mathrm{e}}$ for the non-linear system. If $|\mathbf{J}^t\ \mathbf{s}_{\mathrm{d}}| \to \infty$, so that the fixed point is unstable for the linearized system, then it is also unstable for the non-linear system.

$|\mathbf{J}^t\ \mathbf{s}_{\mathrm{d}}| \to 0$ or ∞ according to whether the eigenvalues of $\mathbf{J}$ have negative or positive real part. $\mathbf{s}_{\mathrm{d}}$ spirals towards 0 or away from it if any of the eigenvalues of $\mathbf{J}$ have non-zero imaginary parts.

Characteristic equation of Jacobian

This section shows that the eigenvalues of $\mathbf{J}$ depend on the size of

$$h = \phi\sigma > 0, \tag{5.28}$$

because the characteristic equation of the Jacobian is

$$\Rightarrow \lambda\left(\lambda + \phi\right)\left(\lambda + 1\right) + h = 0. \tag{5.29}$$

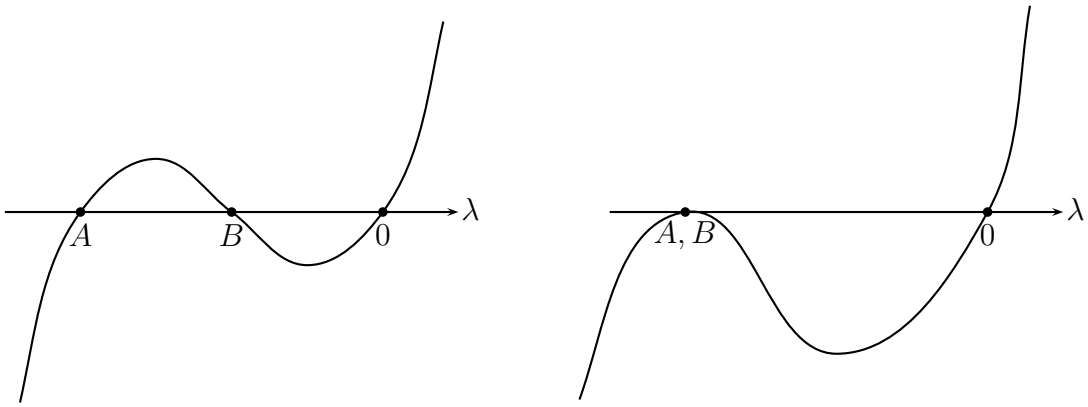
The eigenvalues λ of $\mathbf{J}(\mathbf{s}_{\mathrm{e}})$ are solutions to its characteristic equation i.e.

$$| \lambda\mathbf{I} - \mathbf{J} | = 0 \tag{5.30}$$

$$\Rightarrow \lambda\left(\lambda + \phi\right)\left(\lambda + 1\right) - \frac{\mathrm{dx}(T_e)}{\mathrm{d}T}\phi\left(1 - e^{-T_e}\right) = 0 \qquad \text{by (5.27)}, \tag{5.31}$$

$$\Rightarrow \lambda\left(\lambda + \phi\right)\left(\lambda + 1\right) + h = 0 \qquad \text{by (5.17) , (5.28)}. \tag{5.32}$$

Consider the polynomial in (5.32) when h= 0. When h=0, the roots of the polynomial are 0, -ϕ and -1. Since $\phi > 0$ (by (5.14)), and the coefficient of λ^3 is positive, the graph



(a) $\phi \neq 1$ (b) $\phi = 1$

Figure 5.2: $\lambda(\lambda+\phi)(\lambda+1)$ *given that* $\phi > 0$. $A = -\max(\phi, 1)$. $B = -\min(\phi, 1)$.

of the polynomial (when h= 0) is one of the two in figure 5.2. h is a positive constant so it can be considered a height that shifts the graph up the vertical axis, as in figure 5.3. As h increases, the root at zero becomes negative, so (5.32) never has a non-negative real root. Also, as h increases, the largest negative root gets larger i.e. tends towards $-\infty$, so that (5.32) always has at least one negative root. Since (5.32) always has a negative, real root, the cubic (5.32) factorises into a linear part and a quadratic part. The nature of the other roots depends on the solution of the quadratic part, which depends on h. So, the height h determines whether the polynomial has, in addition to the real negative root: two real negative roots, or one repeated negative root, or two complex conjugate roots.

Condition for stability: no roots with positive real part

This subsection proves equation (5.20). The equilibrium is stable if **J** has no eigenvalues with positive real part. In order to prove (5.20), it is shown that (5.32) has no solutions with positive real part, if and only if

$$h < \phi\,(1+\phi)\,. \tag{5.33}$$

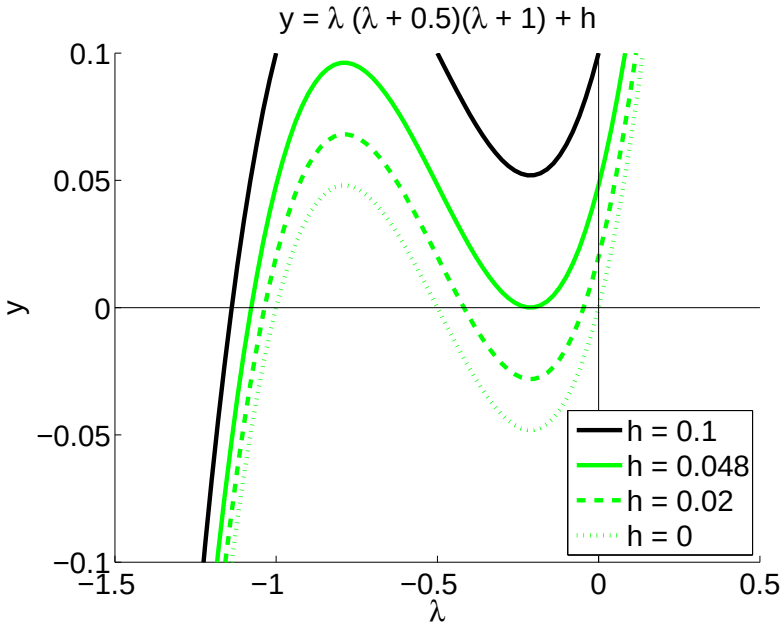


Figure 5.3: $\lambda\left(\lambda+0.5\right)\left(\lambda+1\right)+h, h \geq 0$

To prove (5.33), the polynomial in (5.32) can be factorised as

$$\lambda\left(\lambda+\phi\right)\left(\lambda+1\right)+h=\left(\lambda+p\right)\left\{\lambda^{2}+\left(\phi+1-p\right)\lambda+\phi-\left(\phi+1-p\right)p\right\} \tag{5.34}$$

$$\text{where } h=p\left\{\phi-\left(\phi+1-p\right)p\right\} \tag{5.35}$$

$$\text{and } -p\leq -1 \text{ is the real root of (5.31).} \tag{5.36}$$

By the quadratic formula, $\left\{\lambda^{2}+\left(\phi+1-p\right)\lambda+\phi-\left(\phi+1-p\right)p\right\}$ has roots with negative real part if and only if

$$p<1+\phi \tag{5.37}$$

$$\Rightarrow h<\left(1+\phi\right)\phi \text{ by (5.35), (5.36) since } \left(1+\phi\right)>p>0. \tag{5.38}$$

If $h=\left(1+\phi\right)\phi$ then the roots of (5.32) are $-p=-(1+\phi)$ and $\pm i\sqrt{\phi}$. As discussed above using figure 5.3, as h decreases, p decreases, so if $h<\left(1+\phi\right)\phi$, then $p<1+\phi$. Together with (5.38) this proves (5.33).

Condition for no oscillation: no complex roots

This subsection proves (5.21). Small displacements from the equilibrium $\mathbf{s}_{\mathrm{e}}$ tend smoothly to 0, with no oscillations, so long as h is small enough for all the roots of (5.32) to be real

and negative. Thus h must be smaller than it is in the borderline case, where (5.32) has a negative root $-p$ and two repeated negative roots $-q$. By considering how the sketches in figure 5.2 are shifted upwards by $h > 0$, the repeated negative roots have the value of λ at the turning point between 0 and point B, so that

$$q < \min(1, \phi). \tag{5.39}$$

In the borderline case, (5.32) is of the form $(\lambda + p)(\lambda + q)^2$. So, $h < pq^2$ and it remains to show that the equation for pq^2 in equation (5.22) is correct. This is done by factorising (5.32) into a linear and quadratic part, as before, and then comparing coefficients of powers of λ. In the borderline case,

$$(\lambda + p)(\lambda + q)^2 = \lambda(\lambda + \phi)(\lambda + 1) + h. \tag{5.40}$$

Expanding,

$$\lambda^3 + (p + 2q)\lambda^2 + (2pq + q^2)\lambda + pq^2 = \lambda^3 + (\phi + 1)\lambda^2 + \phi\lambda + h \tag{5.41}$$

Equating coefficients of powers of λ,

$$p + 2q = \phi + 1 \tag{5.42}$$

$$2pq + q^2 = \phi \tag{5.43}$$

$$pq^2 = h. \tag{5.44}$$

pq^2 is obtained from (5.42) and (5.43). Substituting p from (5.42) into (5.43)

$$2(\phi + 1 - 2q)q + q^2 = \phi \tag{5.45}$$

$$\Rightarrow 3q^2 - 2(\phi + 1)q + \phi = 0 \tag{5.46}$$

$$\Rightarrow q = \frac{1}{3}\left\{\phi + 1 \pm \sqrt{(\phi + 1)^2 - 3\phi}\right\}. \tag{5.47}$$

The negative root of (5.47) must be taken, otherwise a contradiction appears.

$$\text{For, suppose } q = \frac{1}{3}\left\{\phi + 1 + \sqrt{(\phi+1)^2 - 3\phi}\right\} \tag{5.48}$$

$$\text{and consider the case where } \phi < 1 \tag{5.49}$$

$$\text{so that } q < \phi, \text{ by (5.39).} \tag{5.50}$$

$$\text{Thus } q = \frac{1}{3}\left\{\phi + 1 + \sqrt{(\phi+1)^2 - 3\phi}\right\} < \phi \tag{5.51}$$

$$\Rightarrow \sqrt{(\phi+1)^2 - 3\phi} < 2\phi - 1 \tag{5.52}$$

$$\Rightarrow (\phi+1)^2 - 3\phi < (2\phi-1)^2 \tag{5.53}$$

$$\Rightarrow 0 < 3\phi\,(\phi - 1) \tag{5.54}$$

$$\Rightarrow 1 < \phi, \text{which contradicts (5.49).} \tag{5.55}$$

A similar contradiction appears for the case where $\phi > 1$. If $\phi = 1$, then the positive root of (5.47) makes $q = 1$, contradicting (5.39). So,

$$q = \frac{1}{3}\left\{\phi + 1 - \sqrt{\phi^2 - \phi + 1}\right\} \tag{5.56}$$

$$\Rightarrow p = \frac{1}{3}\left\{\phi + 1 + 2\sqrt{\phi^2 - \phi + 1}\right\} \text{ by (5.42)} \tag{5.57}$$

$$\Rightarrow pq^2 = \frac{1}{27}\left\{-2 + 3\phi + 3\phi^2 - 2\phi^3 + 2\left(1 - \phi + \phi^2\right)^{3/2}\right\}. \tag{5.58}$$

Approximate values for the stable equilibrium to have no oscillations

This subsection justifies (5.23) and (5.24). Via binomial expansion of $\left(1 - \phi + \phi^2\right)^{3/2}$ in equation (5.58), for small ϕ

$$pq^2 = \frac{\phi^2}{4} + O(\phi^3). \tag{5.59}$$

For large ϕ, via binomial expansion of $\phi^3\left(1/\phi^2 - 1/\phi + 1\right)^{3/2}$,

$$pq^2 = \frac{\phi}{4} - \frac{1}{8} + O\left(\frac{1}{\phi}\right). \tag{5.60}$$

Furthermore, it appears empirically, as shown in figure 5.4, for all ϕ in the likely relevant range for the model, that

$$\min\left(\frac{\phi}{8}, \frac{\phi^2}{8}\right) < pq^2 < \min\left(\frac{\phi}{4}, \frac{\phi^2}{4}\right). \tag{5.61}$$

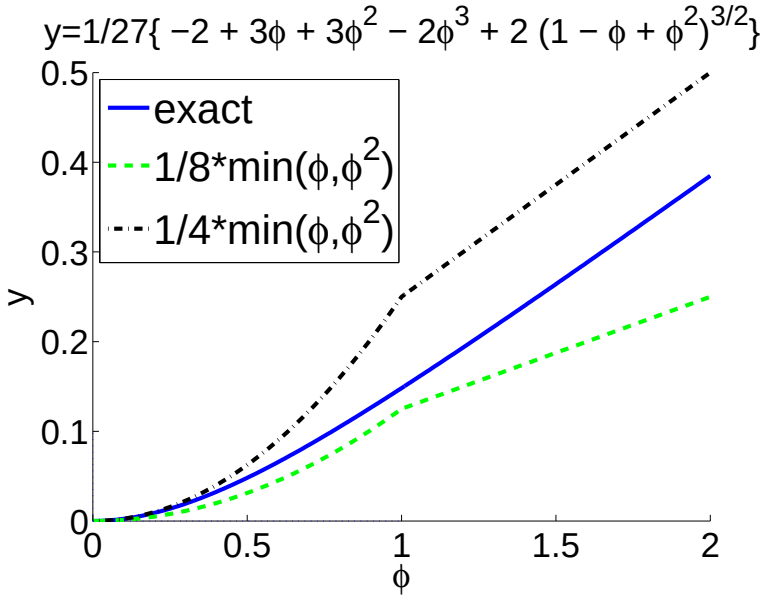


Figure 5.4: The critical value of h for the stable equilibrium to have no oscillations appears to lie between 1/8 and 1/4 of $\min(\phi, \phi^2)$

5.3.4 Period of oscillations

The period of the oscillations toward the stable equilibrium is $\frac{2\pi}{\omega}$ where ω is the imaginary part of the complex eigenvalues [12].

By applying the quadratic formula to the quadratic part of (5.34),

$$\omega = \frac{1}{2}\sqrt{4\{\phi - (\phi + 1 - p)\, p\} - (\phi + 1 - p)^2}, \tag{5.62}$$

$$\text{where } 4\{\phi - (\phi + 1 - p)\, p\} > (\phi + 1 - p)^2. \tag{5.63}$$

$$\text{Hence } \omega < \frac{1}{2}\sqrt{4\{\phi - (\phi + 1 - p)\, p\}} \tag{5.64}$$

$$= \sqrt{\phi - (\phi + 1 - p)\, p} \tag{5.65}$$

$$< \sqrt{\phi} \quad \text{by (5.36) \& (5.37).} \tag{5.66}$$

So the period of oscillations towards the stable equilibrium is

$$> \frac{2\pi}{\sqrt{\phi}}. \tag{5.67}$$

5.3.5 Critical values in dimensional variables

The above conditions for stability and the period of oscillations may be expressed in the original dimensions by applying (5.10)–(5.15) to the non-dimensional results.

The equilibrium point for $\hat{C}$ and $\hat{T}$ is, from (5.16)

$$\begin{pmatrix} \hat{C}_e \\ \hat{T}_e \\ \hat{W}_e \end{pmatrix} = \begin{pmatrix} C_{PI} e^{\hat{\mathrm{x}}^{-1}(\mu)\log 2/\Delta T_{2*CO2}} \\ \hat{\mathrm{x}}^{-1}(\mu) \\ \frac{C_{PI}}{\tau_C}\left\{e^{\hat{\mathrm{x}}^{-1}(\mu)\log 2/\Delta T_{2*CO2}} - 1\right\}\frac{e^{\mu\hat{t}}}{\chi} \end{pmatrix}. \tag{5.68}$$

From (5.20), the equilibrium is stable if and only if

$$-\frac{\Delta T_{2*CO2}}{\log 2}\frac{\mathrm{d}\hat{\mathrm{x}}(\hat{T}_e)}{\mathrm{d}\hat{T}}\left\{1 - e^{-\hat{\mathrm{x}}^{-1}(\mu)\log 2/\Delta T_{2*CO2}}\right\} < \frac{1}{\tau_T} + \frac{1}{\tau_C} \tag{5.69}$$

From (5.23), the system does not oscillate near the equilibrium if

$$-\frac{\Delta T_{2*CO2}}{\log 2}\frac{\mathrm{d}\hat{\mathrm{x}}(\hat{T}_e)}{\mathrm{d}\hat{T}}\left\{1 - e^{-\hat{\mathrm{x}}^{-1}(\mu)\log 2/\Delta T_{2*CO2}}\right\} < \frac{1}{8}\min\left(\frac{1}{\tau_T}, \frac{1}{\tau_C}\right) \tag{5.70}$$

and, from (5.24), the system does oscillate near the equilibrium if

$$-\frac{\Delta T_{2*CO2}}{\log 2}\frac{\mathrm{d}\hat{\mathrm{x}}(\hat{T}_e)}{\mathrm{d}\hat{T}}\left\{1 - e^{-\hat{\mathrm{x}}^{-1}(\mu)\log 2/\Delta T_{2*CO2}}\right\} > \frac{1}{4}\min\left(\frac{1}{\tau_T}, \frac{1}{\tau_C}\right). \tag{5.71}$$

From (5.67), the period of oscillations towards the stable equilibrium is

$$> 2\pi\sqrt{\tau_T \tau_C} \quad \text{years.} \tag{5.72}$$

5.4 Model parameters

Clearly, from (5.69), the parameters chosen for the model determine whether its equilibrium is stable or not. On the left hand side of (5.69), a higher level of tolerable global warming or a decrease in the decarbonisation rate are destabilising, as they allow longer lags before the system has to adjust. A higher level of climate sensitivity and a steeper damage function are also destabilising, as they imply a faster pace of change to which the system must adjust. On the right hand side of (5.69), greater thermal or carbon cycle inertia is destabilising, as it means the system can only adjust slowly. Consistent with this, climate sensitivity and response lags have been found to be sensitive determinants of the social cost of carbon [21]. The initial conditions, including the initial carbon intensity, affect the system's trajectory, but do not affect the stability of the equilibrium.

The numerical simulations in this section use parameters based on the following:

1. An initial carbon intensity χ of 0.025 ppmv / \$ tr, consistent with current CO_2 rises of approx 2 ppmv per year.

2. Initial levels of CO_2 concentation, global warming and global wealth of 380 ppmv, 0.7K, \$160 tr [5]

3. A central estimate for the equilibrium climate sensitivity of ΔT_{2*CO2}=3K, approached on a characteristic climate timescale of $\tau_T = 50$ years.

4. A preindustrial level of CO_2 of 280 ppmv.

5. A characteristic carbon timescale of $\tau_C = 50$ years, consistent with a fixed airborne fraction.

6. A decarbonisation rate $\mu = 1\%$ per year, based on records of economic and CO_2 emissions growth for the late 20th century [15, 17, 3]

7. A linear expansion (or damage) function for wealth of

$$\frac{\mathrm{d}\hat{W}}{\mathrm{d}\hat{t}} = \xi\left(1 - \delta\hat{T}\right)\hat{W}. \tag{5.73}$$

Equation (5.73) assumes the rate of economic growth is reduced by a fraction δ for each kelvin of global warming. The orthodox climate prediction chain essentially assumes that $\delta \approx 0$, such that there is no feedback from climate change to economic growth. The actual fraction δ is unknown, though it is constrained by the 20th century experience. Parry *et al.* [21] estimates global mean losses of GDP (for a 4K warming) as between 1-5% GDP, but considers this likely to be an underestimate. Rearranging the linear damage function in (5.73),

$$\xi = \frac{\hat{\mathrm{x}}(\hat{T})}{1 - \delta\hat{T}} \tag{5.74}$$

$$\Rightarrow \xi - \mu = \frac{\hat{\mathrm{x}}(\hat{T})}{1 - \delta\hat{T}} - \mu. \tag{5.75}$$

Using the observed values of μ, $\hat{T}$, $\hat{\mathrm{x}}(\hat{T})$ for the late twentieth century which are 1%, 0.7K and 3%, and allowing for a 10% error in each measurement, then

$$\Rightarrow \xi - \mu = \frac{3\% \pm 0.3\%}{1 - \delta(0.7 \pm 0.07)} - (1\% \pm 0.1\%) \tag{5.76}$$

which constrains $(\xi - \mu)$ and δ in figure 5.6 to the brown region marked as observed.

Parry *et al.* [21] suggests a net *positive* effect on GDP for a small level of warming, so that 3% pa growth could in fact be *above* the background level of economic growth. The actual growth experienced since the 1950s could be considered atypical, being attributed to the recovery from World War II or advances in technology. While the results of (5.69) – (5.72) are more general, resting only on assumptions of section 5.2.1, figures 5.5–5.7 are less certain, as they rely on the more specific parameters in this section.

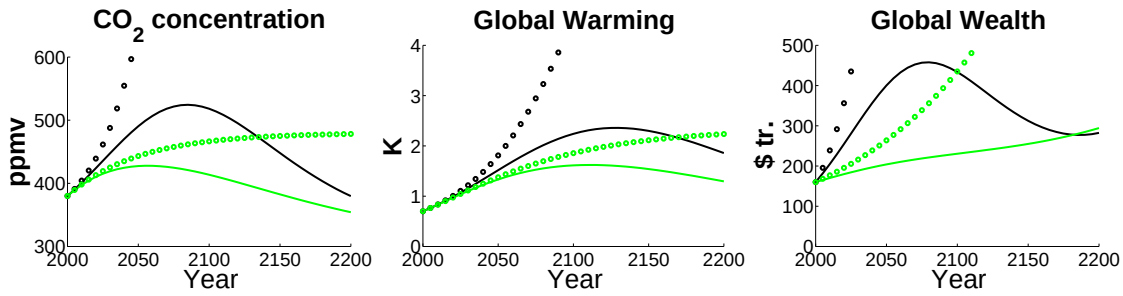
In low-level climate change impacts (5.73) would imply an exponential growth of $\hat{W}$ at a constant background-rate of ξ. In the late 20th century the global economic growth-rate, ξ, averaged about 3% per year [31]. However, the Special Report on Emissions Scenarios for 21st century growth [18] translate into growth-rates of between 1 and 4 % per year, and these extremes for ξ are used in the next section.

5.5 Model results

Figure 5.5 compares the model projections of the 21st and 22nd centuries for the standard no-feedback case (dashed lines) with projections when δ=0.5, a value that would produce an equilibrium global warming of $\hat{T}$=2(1-μ/ξ)K. A low background economic growth-rate of ξ=1% per year is considered (green lines) as is a high background economic growth-rate of ξ=4% per year (black lines). In both cases the closure of the climate-economy feedback loop significantly affects the projections, especially in the 22nd century. However, the emergent dynamics are very different in the low and high growth cases.

In the low growth-rate case the impact of climate change on economic growth leads to a *soft-landing* at the equilibrium in which the negative climate-economy feedback loop counteracts the background economic growth-rate, the CO_2 emission rate stabilises, and the economy grows at the decarbonisation rate of μ per year. In contrast, in the high growth case the negative feedback loop is too slow to balance the background growth-rate. This leads to an overshoot of the climate equilibrium that precedes an economic crash (see Figure 5.5c, red continuous line). The high growth-rate case projects an economic depression for the whole of the 22nd century, although rather ironically the CO_2 concentration and climate recover as a result.

The linear analysis in section 5.3 identified three regimes: the soft-landing associated with low background economic growth-rates, the instability associated with high growth-



(a) (b) (c)

Figure 5.5: Impact of the climate-economy feedback on projections for the 21st and 22nd century. Coupled projections are shown by the continuous lines, and uncoupled simulations are shown by the broken lines. Black lines assume a background economic growth-rate of 4%pa; green lines assume 1% pa.

rates, and oscillations at intermediate growth-rates. Figure 5.6 shows the location of these regimes in the parameter-space defined by the background growth-rate of CO_2 emissions ($\xi - \mu$), and the fractional suppression of economic growth per unit of global warming, δ. The possible parameters are constrained by the historical level of global warming and economic growth. In the absence of any intervention, a soft landing requires the economy to be almost insensitive to global warming, that is for δ to be less than about 0.05. This is equivalent to requiring that the economy can withstand global warming of more than 20K without contracting, which seems very unlikely. It therefore seems likely that the climate-economy system is currently in an oscillatory regime, with the possibility of even more abrupt instabilities if economic growth is faster in the future or if the damage function for wealth is more steep or nonlinear than we have supposed.

Figure 5.7 shows the stability regimes in the parameter-space defined by the background economic growth-rate and the decarbonisation rate, for two values of δ, the economic damage due to climate change. It is clear that decarbonisation raises the threshold under which a soft landing is possible.

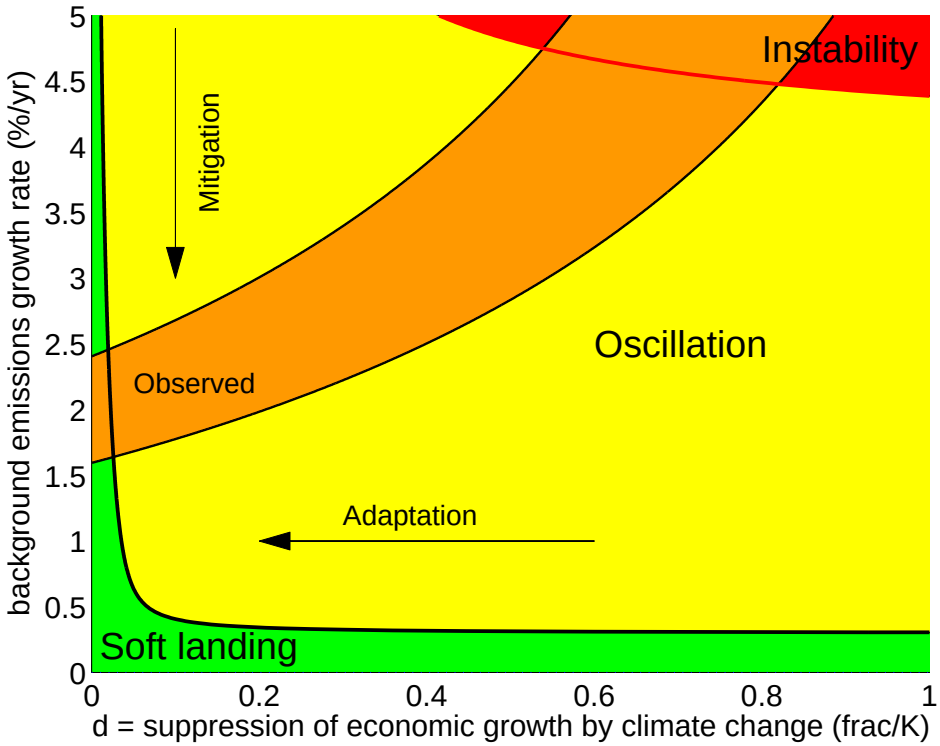
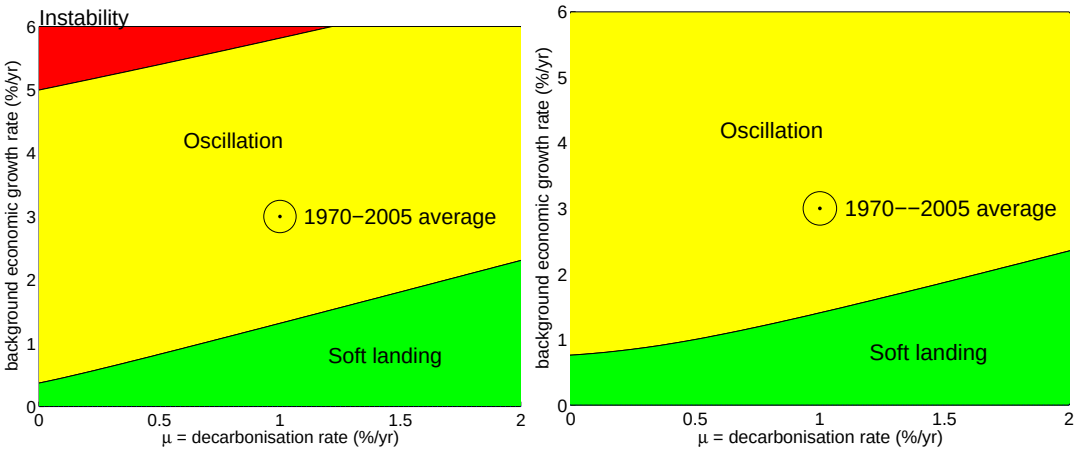


Figure 5.6: Stability regimes of the climate-economy system as a function of the background rate of growth of CO_2 emissions $\xi - \mu$ and the economic damages due to global warming. The brown area is consistent with the observed level of global warming and recent economic growth, according to the data in section 5.4. Climate sensitivity is assumed to be 3K and the characteristic timescales for $\hat{T}$ and CO_2 are both taken as 50 years.

5.6 Relation to more sophisticated Integrated Assessment Models

In the climate-emissions equilibrium state the economic growth rate is equal to the decarbonisation rate. Decarbonisation is typically understood to represent an economic cost [13], but in this model it is only through decarbonisation that the economy can grow sustainably.

A limitation of the model is that the price of decarbonisation and the resources spent on decarbonisation are not explicitly modelled. The parameter μ is externally prescribed and does not depend on wealth. This appears to conflict with the theory



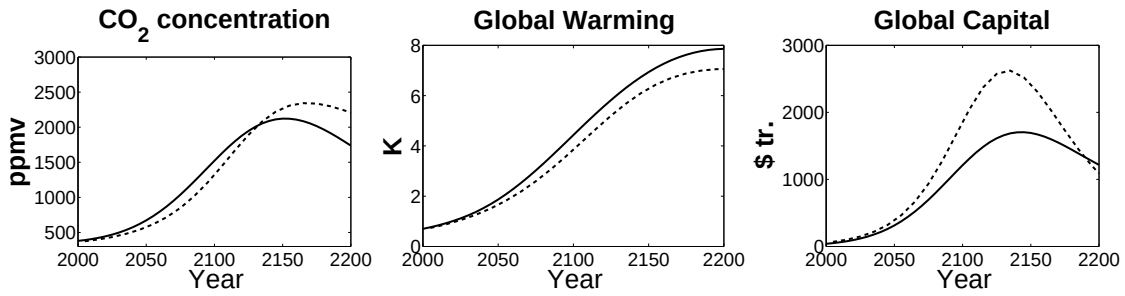
(a) (b)

Figure 5.7: Stability regimes of the climate-economy system as a function of the background economic growth-rate ξ and the rate of decarbonisation of the economy (μ). Left and right panels show different economic damages due to global warming; (a) δ=0.5 per K; (b) δ=0.1 per K.

of the Environmental Kuznets Curve (EKC) which holds that a clean environment is a luxury good [6], so that under the EKC theory, μ might be an increasing function of wealth. However, [6] shows two reasons why the EKC may not apply to global CO_2 emissions. First, there is a clearer case for an EKC in pollutants whose effects are short-term and local, than for pollutants whose effects are long-term and global. Second, and consistent with the first reason, the apparent drop in pollution in richer societies may be achieved to some degree by just exporting that pollution to poorer societies. No such export is possible at a global level.

The assumption of a linear damage function is more optimistic than most commentators' [28] though its size is an order of magnitude more severe. The well known DICE model, if suitably simplified and with a scaled-up damage function, exhibits similar behaviour to the three-variable model presented here. Figure 5.8 compares a simplified version of the global DICE model (dashed lines) with the our model (solid lines). The simplifications made to the DICE model are:

i The DICE capital share is set to one, removing the sensitivity to DICE's exogeneous population growth. This is defensible if population is treated as a function, indeed



(a) (b) (c)

Figure 5.8: Impact of the climate-economy feedback on projections for the 21st and 22nd century. Simplified DICE (dashed lines) is compared with the 3-variable model in this chapter (solid lines).

a component of global wealth parameters. It is consistent with [9] which attributes the bulk of productivity differences to the accumulation of social infrastructure.

ii The DICE exogeneous productivity growth rate is set to zero. This is no more arbitrary than setting the exogeneous decarbonisation rate to a constant.

iii The DICE carbon intensity is set to reduce by 10% per decade, consistent with the historical record.

iv The DICE savings rate is fixed at 23%, approximately the level set by the optimised DICE model.

v The DICE damage function is *multiplied by ten*. This is the most striking change, but part of the increase in the damage-to-production function is due to the fact that the DICE-99 depreciation function is fixed, whereas it is reasonable to suppose that the replacement cost of assets will increase along with the production cost.

The parameters for our model are the same as those in section 5.5 except that:

i the damage coefficient is set to *0.15* (implying a climate induced recession at *7K* of global warming).

ii The background economic growth rate is set to *5%*.

5.7 Conclusion

Even damped oscillations are likely to be damaging to the long-term well-being and security of humanity [26], so how can they be avoided? Figures 5.1a and 5.6 suggest four possible ways to ensure a soft-landing for the climate-economy system. The first is to reduce the background rate of economic growth to rates ($\xi < 0.5\%$ per year) that can be gradually counteracted by the climate-economy feedback loop. The second to increase the decarbonisation rate, though this is unlikely to be sustainable for long, if decarbonisation gets progressively harder to achieve. The third is to reduce the sensitivity of the economy [24] to global warming (i.e. reduce δ) through adaptation measures. The fourth way is to break the link between economic output and climate change, for instance by geo-engineering to offset the radiative forcing of CO_2.

The inclusion of even a relatively weak feedback loop between climate change and economic growth leads to projections for the 21st and 22nd centuries that differ fundamentally from the standard no-feedback case. The climate-economy feedback permits a climate equilibrium state in which the background economic growth-rate (i.e. in the absence of climate change) is counteracted by climate change impacts on the economy. Economic growth in this climate state is equal to the rate of decarbonisation, so mitigation efforts are critical to ensure long-term sustainable growth. However, the assumptions behind figure 5.7 indicate that decarbonisation will not be enough to ensure a soft-landing on this sustainable trajectory. Instead they anticipate over-shoot oscillations or even instabilities under historical rates of economic growth, and feasible levels of economic damage due to climate change. Navigating the climate-economy system to a soft-landing will require massive efforts in both mitigation and adaptation, but also lower but more sustainable rates of global economic growth.

References

[1] Archer, D., Eby, M., Tokos, K., Brovkin, V., Ridgwell, A., Cao, L., Mikolajewicz, U., Caldeira, K., Matsumoto, K., Munhoven, G., *et al.* (2009). Atmospheric Lifetime of Fossil-Fuel Carbon Dioxide. *Annual Review of Earth and Planetary Sciences*, **37**(1).

[2] Barker, T., Ekins, P., and Johnstone, N. (1995). *Global warming and energy demand*. Routledge. 336 pages.

[3] Behrens, A., Giljum, S., Kovanda, J., and Niza, S. (2007). The material basis of the global economy Worldwide patterns of natural resource extraction and their implications for sustainable resource use policies. *Ecological Economics*, **64**(2), 444–453.

[4] Cox, P. M. and Stephenson, D. B. (2007). Climate Change: a Changing Climate for Prediction. *Science*, **317**(5835), 207.

[5] Davies, J. B., Sandstrom, S., Shorrocks, A., and Wolff, E. (2007). Estimating the level and distribution of global household wealth. *UNU-WIDER Research Paper*, **77**, 1–56.

[6] Dinda, S. (2004). Environmental Kuznets Curve hypothesis: A survey. *Ecological Economics*, **49**(4), 431–455.

[7] Eby, M., Zickfeld, K., Montenegro, A., Archer, D., Meissner, K. J., and Weaver, A. J. (2009). Lifetime of anthropogenic climate change: millennial time scales of potential CO2 and surface temperature perturbations. *Journal of Climate*, **22**(10), 2501–2511.

[8] Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R., and Webster, M. D. (2002). Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**(5552), 113–117.

[9] Hall, R. E. and Jones, C. I. (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others?*. *Quarterly Journal of Economics*, **114**(1), 83–116.

[10] Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., and Stone, P. (1985). Climate response times: dependence on climate sensitivity and ocean mixing. *Science*, **229**(4716), 857–859.

[11] Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S., and Samuel, M. D. (2002). Climate warming and disease risks for terrestrial and marine biota. *Science*, **296**(5576), 2158–2162.

[12] José, J. V. and Saletan, E. J. (1998). *Classical dynamics: a contemporary approach*. Cambridge University Press. 676 pages.

[13] Keller, K., Hall, M., Kim, S. R., Bradford, D. F., and Oppenheimer, M. (2005). Avoiding dangerous anthropogenic interference with the climate system. *Climatic Change*, **73**(3), 227–238.

[14] Knack, S. and Keefer, P. (1997). Does Social Capital Have An Economic Payoff? A Cross-Country Investigation*. *Quarterly Journal of Economics*, **112**(4), 1251–1288.

[15] Marland, G., Boden, T. A., Andres, R. J., Brenkert, A. L., and Johnston, C. A. (2006). Global, regional, and national fossil fuel CO_2 emissions. *Trends: A compendium of data on global change*. Carbon Dioxide Information Analysis Center.

[16] Meadows, D., Meadows, D. H., and Randers, J. (2004). *Limits to Growth: The 30 Year Global Update*. Chelsea Green Publishing Company. 338 pages.

[17] Metz, B. and Davidson, O. R. (2007). *Climate Change 2007 Mitigation: Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Published for the Intergovernmental Panel on Climate Change by Cambridge Univ. Press. 851 pages.

[18] Nakicenovic, N. and Swart, R. (2000). Special report on emissions scenarios. *Cambridge University Press.* 599 pages.

[19] Nordhaus, W. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press. 234 pages.

[20] Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., and Hanson, C. E. (2007a). *Climate change 2007: Impacts, adaptation and vulnerability: Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. 976 pages.

[21] Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P., and Hanson, C. E. (2007b). *IPCC, 2007: Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. 976 pages.

[22] Raupach, M., Marland, G., Ciais, P., Le Quéré, C., Canadell, J., Klepper, G., and Field, C. (2007). Global and regional drivers of accelerating CO2 emissions. *Proceedings of the National Academy of Sciences*, **104**(24), 10288.

[23] Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., *et al.* (2004). The oceanic sink for anthropogenic CO2. *Science*, **305**(5682), 367–371.

[24] Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B. (2001). Catastrophic Shifts in Ecosystems. *Nature*, **413**(6856), 591–596.

[25] Schnellnhuber, H. J. and Cramer, W. P. (2006). *Avoiding dangerous climate change*. Cambridge University Press. 392 pages.

[26] Silk, L. (1992). Dangers of slow growth. *Foreign Affairs*, **72**, 167.

[27] Soden, B. J. and Held, I. M. (2006). An assessment of climate feedbacks in coupled ocean–atmosphere models. *Journal of Climate*, **19**(14), 3354–3360.

[28] Stern, N. H. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge University Press. 692 pages.

[29] Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Mitchell, J. F. B., and Jenkins, G. J. (2000). External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**(5499), 2133–2137.

[30] Strogatz, S. H. and Herbert, D. E. (1994). *Nonlinear dynamics and chaos*. Addison-Wesley Reading, MA. 498 pages.

[31] United Nations Statistical Division (2008). National accounts main aggregates database. Economics and Statistics Branch, United Nations Statistics Division.

[32] Volterra, V. (1928). Variations and fluctuations of the number of individuals in animal species living together. *ICES J. Mar. Sci*, **3**, 3–51.