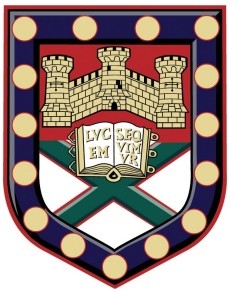
University of Exeter

Thesis in Mathematics

Delayed effects and critical transitions in climate models



Author:
Courtney Quinn

Supervisors:
Jan Sieber
Timothy S. Lenton
Peter Ashwin

Submitted by Courtney Quinn, to the University of Exeter as a thesis for the degree of
Doctor of Philosophy in Mathematics,
October 2018.

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

. . . . . .

"Differences of habit and language are nothing at all if our aims are identical and our hearts are open."

- J. K. Rowling, University of Exeter alumnus

UNIVERSITY OF EXETER

Abstract

Thesis in Mathematics

College of Engineering, Mathematics, and Physical Sciences
PhD in Mathematics

Delayed effects and critical transitions in climate models

by Courtney Quinn

There is a continuous demand for new and improved methods of understanding our climate system. The work in this thesis focuses on the study of delayed feedback and critical transitions. There is much room to develop upon these concepts in their application to the climate system. We explore the two concepts independently, but also note that the two are not mutually exclusive.

The thesis begins with a review of delay differential equation (DDE) theory and the use of delay models in climate, followed by a review of the literature on critical transitions and examples of critical transitions in climate. We introduce various methods of deriving delay models from more complex systems. Our main results center around the Saltzman and Maasch (1988) model for the Pleistocene climate ('Carbon cycle instability as a cause of the late Pleistocene ice age oscillations: modelling the asymmetric response.' *Global biogeochemical cycles*, 2(2):177-185, 1988). We observe that the model contains a chain of first-order reactions. Feedback chains of this type limits to a discrete delay for long chains. We can then approximate the chain by a delay, resulting in scalar DDE for ice mass. Through bifurcation analysis under varying the delay, we discover a previously unexplored bistable region and consider solutions in this parameter region when subjected to periodic and astronomical forcing. The astronomical forcing is highly quasiperiodic, containing many overlapping frequencies from variations in the Earth's orbit. We find that under the astronomical forcing, the model exhibits a transition in time that resembles what is seen in paleoclimate records, known as the Mid-Pleistocene Transition. This transition is a distinct feature of the quasiperiodic forcing, as confirmed by the change in sign of the leading finite-time Lyapunov exponent. Additional results involve a box model of the Atlantic meridional overturning circulation under a future climate scenario and time-dependent freshwater forcing. We find that the model exhibits multiple types of critical transitions, as well as recovery from potential critical transitions. We conclude with an outlook on how the work presented in this thesis can be utilised for further studies of the climate system and beyond.

Acknowledgements

I would like to thank first and foremost my incredible supervisors from the University of Exeter. I am eternally grateful to Prof. Jan Sieber for his constant guidance, endless support, and the many hours of work he dedicated to this project. Also to Prof. Timothy Lenton and Prof. Peter Ashwin for their support throughout the course of my PhD and providing me with opportunities to expand my research.

A special thank you goes to everyone who supported me throughout my secondment at Utrecht University. In particular, to Dr. Anna von der Heydt and Prof. Henk Dijkstra who served as secondary supervisors, and to Swinda Falkena who continued and beautifully executed the work I began.

To my colleagues at the Met Office, Dr. Richard Wood and Dr. Laura Jackson, for their time and inspiration, and to my CRITICS mentor Peter Ditlevsen for showing me that mathematics is more than just proofs and calculations.

Thank you to my CRITICS family, especially Daniele, Kalle, Karl, Michael, and Iacopo, for all the days and nights of long discussions that have widened my views of mathematics and the world beyond.

I would not have made it through these three years without the much needed coffee and biscuit breaks, especially towards the end. Thank you to Chris, Vasilis, and everyone in the Maths Postgraduate Journal Club for your time and support.

I cannot forget my kind officemates who have humoured my questions, complaints, and never-ending smells of coffee in the office. Damian, Hassan, and Paul, these years would not have flown by so quickly and smoothly without your company and friendship. In the end, I'm glad none of you got replaced by a couch.

Finally, to my rocks- the people who have provided me with support not only with my work but with all the pressures and troubles that life can bring. Amanda, Travis, AVP, Kait, ReAnna, Nuria, Andy, Gertjan, and Memo, thank you for it all. The late night discussions, the phone calls across time zones, the Netflix marathons, the spontaneous parties, the shoulders to cry on, and of course the laughter. Most of all, thank you for always believing in me and proving that no distance or time apart can break the bonds you create with people.

This project would not have been possible without the funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 643073.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

The work presented in Chapter 5 was previously published in part in *Dynamics and Statistics of the Climate System* as "The Mid-Pleistocene Transition induced by delayed feedback and bistability" by Courtney Quinn (PhD candidate), Jan Sieber (supervisor), Anna S. von der Heydt, and Timothy M. Lenton (supervisor). This study was conceived by all of the authors. I carried out the majority of manuscript authorship, all numerical results, and creation of figures. The remainder of the work presented has been submitted for peer review as manuscript "Effects of periodic forcing on a paleoclimate delay model" by Courtney Quinn (PhD candidate), Jan Sieber (supervisor), and Anna S. von der Heydt. This study was conceived by all of the authors. I carried out the majority of manuscript authorship, creation of figures, and all numerical results except the calculation for the unstable manifold of a periodic orbit.

The work presented in Chapter 6 is included in the manuscript "Basin bifurcations, oscillatory instability and rate-induced thresholds for AMOC in a global oceanic box model" by Hassan Alkhayuon, Peter Ashwin (supervisor), Laura Jackson, Courtney Quinn (PhD candidate) and Richard Wood. This study was conceived by all of the authors. I carried out the model scaling and model reduction, all numerical bifurcation analyses, and the studies of additional freshwater forcing with instantaneous change. All figures were created by me except where noted otherwise.

This work is dedicated to the memory of Professor Clarence "Larry" Howe who saw potential in me and didn't allow me to waste it.

Chapter 1

Introduction

1.1 Context and motivation

The study of climate is a complex and challenging science. The climate system is comprised of very complex spatio-temporal variability over multiple subsystems such as the atmosphere, the hydrosphere, the cryosphere, the biosphere, and the lithosphere (Dijkstra, 2013). In addition, there is no clear or simple timescale separation between the multiple subsystems (see Table 1.1 for comparison).

Additionally the climate system is constantly changing in the sense that the assumption of stationarity is violated. On short time scales this can be attributed to human impact, and on longer time scales these would be due to, for example, changes of the Earth's orbit or continental shifts. This type of non-stationarity is referred to as climate variability (Salinger, 2005). Nevertheless, it is imperative that society attempts to understand how the system works and when drastic changes can occur. The ability to foresee and respond to these changes will determine our survival as a species on this planet.

One important but overlooked aspect of climate is delayed effects. Delays in climate can be observed through the transport of mass or energy via propagating waves (Dijkstra, 2008) or chains of reactions that form feedback loops (Dijkstra, 2013). Understanding the existence and extent of any delayed

Subsystem	lower timescale	upper timescale
Atmosphere	seconds	days
Hydrosphere	months	years x10^3
Cryosphere	years x10^2	years x10^5
Biosphere	months	years x10^7
Lithosphere	years x10^5	years x10^9

Table 1.1: Characteristic timescales of climate subsystems, compiled from Dijkstra (2013) and Goosse et al. (2010).

effects is useful for long-term predictions of how the climate may be influenced. In this thesis we will analyse delayed effects in different aspects of the climate system through dynamical systems theory.

Another fairly recent research area of interest is the study of *critical transitions*. This is a scientific term that describes the sudden, significant changes to the state of a system. The common scientific assumption is that systems remain at stable equilibrium in the linear response regime, in which it reverts back to equilibrium exponentially after small temporary disturbances and adjusts proportionally to small permanent changes of inputs. For example the notion of climate sensitivity refers to the amount of change in global temperature as a response to a doubling of atmospheric CO_2, considered a small perturbation (Roe and Baker, 2007). A *tipping point* is said to have been crossed when this assumption is violated (Lenton et al., 2008). At the level of conceptual models, most observations of this type of behaviour have been explained by three mathematical mechanisms: slow passage through a bifurcation, exit from a basin of attraction due to noise, and failure to track a quasi-static attractor (Ashwin et al., 2012).

There have been many studies over the past decade which focus on particular critical transitions either observed or suspected to occur within the climate system. One influential study was conducted by Lenton et al. (2008), in which the authors identify 'tipping elements' in the climate, or rather, large-scale components of the Earth system that can potentially be switched into a qualitatively different state by small perturbations. In this thesis we aim to address some of the tipping elements outlined in that study, particularly the melting of the Greenland Ice Sheet (GIS) and changes in North Atlantic Deep Water (NADW) formation. We consider possible critical transitions caused by these elements during past and future climate scenarios. The melting of the GIS can be a trigger for the shutdown (or severe slowing) of the Atlantic meridional overturning circulation (AMOC) (Lenton et al., 2008), which is a primary contributor to the regulation of the earth's climate (Cheng et al., 2013). The AMOC is also highly connected to NADW formation. We investigate changes in NADW in relation to the Mid-Pleistocene Transition (MPT). The MPT is a transition that occurred between 1200 and 700 kyr BP, where proxy records for global temperature show a change from a dominant 41 kyr period to a 100 kyr period (Dijkstra, 2013). While the 41-kyr oscillations have been attributed to external forcing (Paillard, 2001), the 100-kyr cycles have been proposed to be a result of nonlinear responses of the climate system itself (Gildor and Tziperman, 2001, Imbrie et al., 1993, Yiou et al., 1994). We compare the climate record and the external orbital forcing in Figure 1.1.

Our main goal is to explore the ideas of delayed effects and critical transitions within climate in parallel, and combine these ideas where possible (i.e. delayed effects which can lead to a critical transition). The study of either of these topics is useful on its own, however the connection between the two cannot be overlooked. We aim to pave the way for further understanding of such phenomena in the climate system.

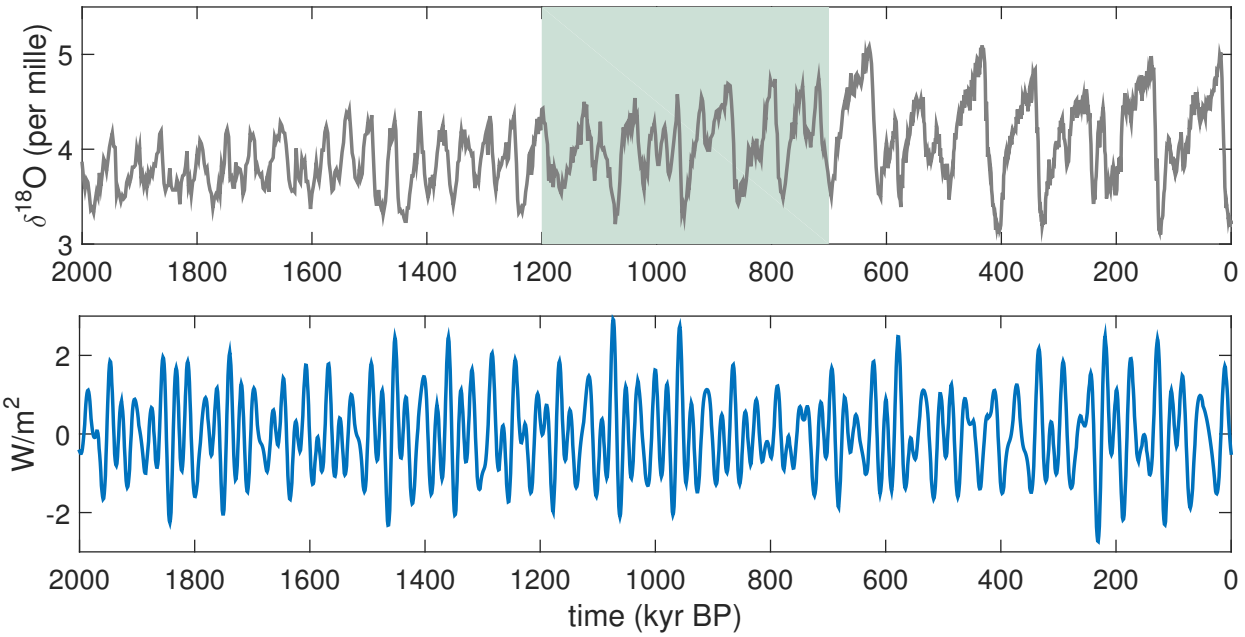


Figure 1.1: Top: Reconstruction of benthic δ^{18}O from 57 globally distributed sediment cores, adapted from Lisiecki and Raymo (2005) data. The green shaded area represents the time frame of the Mid-Pleistocene Transition. Bottom: Normalised integrated July insolation at 65°N, adapted from Huybers and Eisenman (2006).

1.2 Objectives and outline

In this thesis, we will explore potential delay effects and critical transitions in the climate system through considering simplified mathematical models of specific climate phenomena. The simplified models are obtained either through conceptual inception or the reduction of higher complexity models. This approach allows us to then use tools from dynamical systems theory, such as systematic parameter studies, studies of basins of attractions, and finite-time Lyapunov exponents, to explore the possible behaviours one can observe in the system. The thesis is structured as follows.

Chapter 2 reviews some literature around models with delay and critical transitions, specifically in the context of climate science. We first introduce some general properties of delay differential equations (DDEs). We then present examples in which DDEs have been used to model climate processes, namely energy balance models (EBMs) and conceptual models of the El Niño Southern Oscillation (ENSO). Following the discussion of models with delay, we introduce the idea of critical transitions. We review the current understanding around what constitutes a critical transition, and we explain three different types. Finally, we discuss two specific examples of a critical transition in the climate system: the MPT and the shutdown of the AMOC.

In Chapter 3 we shift the focus to the methods of obtaining models with delay from more complex models. The aim is to illustrate different ways of determining whether or not systems have delayed effects which would allow for the reduction of dependent variables. We introduce three different

methods: the linear chain approximation, the Mori-Zwanzig projection framework, and reflection of wave equations. We first show the theoretical derivation of two different types of linear chain systems, one with all identical timescales and one with all distinct timescales. We then derive the Mori-Zwanzig projection for a generic system of ODEs. We illustrate the effectiveness of the projection by applying it to two examples of a linear chain system. We conclude with two case studies of phenomena modelled by wave equations, ENSO and the Atlantic Multidecadal Oscillation (AMO). We show the procedure for deriving delay equations through the reflection conditions of the wave equations. We conclude with a numerical study of the AMO model with delay, where the model is initialised with an idealised history followed by a data-based history.

Chapter 4 starts from a conceptual paleoclimate model with delay by Ghil et al. (1987). The model is comprised of Boolean delay equations (BDEs), in which the variables take only values of 0 and 1 and the equations use logical operations. This model uses three variables (global ice volume, global temperature, and ocean circulation) as well as three delays (ice sheet accumulation time, ice sheet viscoplastic expansion time, and overturning time of the ocean). We transform the BDE model into a discrete map and then regularise it to a DDE with continuous variables. We identify two regular attractors: an equilibrium and a large amplitude periodic orbit. We perform a systematic bifurcation analysis which classifies the dependence of the large amplitude periodic attractor on the ratios between the delays.

The Ghil et al. (1987) BDE model contained little in terms of physical mechanisms, only relying on timescales and logical relations of processes. In Chapter 5 we then turn to a conceptual paleoclimate model with a more realistic representation of underlying physical processes, the Saltzman and Maasch (1988) model. This model included the effects of carbon uptake and transport in the ocean, deep water formation, and the asymmetrical response of the atmospheric CO_2 fluctuations, and thus shows oscillations that match features of the palaeoclimate proxy records. It was developed to explore the possible cause of the MPT. The model is a system of three ODEs representing global ice mass, global atmospheric CO_2, and NADW. These three variables can be related to the variables of the Ghil et al. (1987) model discussed in Chapter 4. We analyse the Saltzman and Maasch (1988) model and find similar feedback effects as in the Ghil et al. (1987) model. We are able to reduce the model to a scalar DDE. We then analyse this DDE under varying delay and external forcing.

The variability of delay in the DDE model of Chapter 5 comes from the variability in the overturning time of the ocean. This is primarily controlled by the AMOC (Hawkins et al., 2011). In Chapter 6 we consider the AMOC in more detail, particularly the possibility of its shutdown (explained in Section 2.3.3). We analyse a box model of the global ocean developed by Rodriguez et al. (2017) through determining the dominant dynamics of the Atmosphere Ocean General Circulation Model FAMOUS. We show that the model can be further reduced in dimensionality without losing any dynamical behaviour. We consider the effects of an altered background state to represent a future

climate scenario. We also explore the model's behaviour when subjected to a increase of fresh water forcing at different rates.

Finally, Chapter 7 summarises our results. We draw connections between previous studies and the work presented in this thesis. In addition, we consider follow-up studies and potential future projects which could benefit from our results.

Chapter 2

Delays and critical transitions in the climate system

In this chapter we review some of the literature around delay differential equations (DDEs) and critical transitions. We take a particular focus on the connection of DDEs and critical transitions with studies of the climate system. Section 2.1 presents some common results regarding solutions and stability analysis of DDEs. Section 2.2 reviews the history of delays used in climate models with specific examples. Section 2.3 discusses critical transitions, with the types of transitions outlined in Section 2.3.1 and two specific examples of possible critical transitions in climate in Section 2.3.2 and Section 2.3.3.

2.1 General theory regarding delay differential equations

We introduce the general form of a DDE:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}_t, p), \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ are state variables, $p \in \mathbb{R}^q$ are parameters, and $f : C([-\tau, 0]; \mathbb{R}^n) \times \mathbb{R}^q \to \mathbb{R}^n$ ($C([-\tau, 0]; \mathbb{R}^n)$ is the space of continuous functions on $[-\tau, 0] \mapsto \mathbb{R}^n$). We will assume f is sufficiently smooth. We define for $t \in [0, T]$, $C([-\tau, T]; \mathbb{R}^n) \ni x(t) \mapsto \mathbf{x}_t(s) = \mathbf{x}(t + s) \in C([-\tau, 0]; \mathbb{R}^n)$ for $s \in [-\tau, 0]$, where $\tau > 0$ is finite. Note that an unbounded delay $\tau = \infty$ would require a more general phase space than the one chosen here. We will consider initial conditions $\mathbf{x}_0 \in C([-\tau, 0]; \mathbb{R}^n)$ and define the norm as $||\mathbf{x}_0|| = \max\{|\mathbf{x}(s)| : s \in [-\tau, 0]\}$.

Existence of solutions for DDEs with finite discrete delay can be shown through the *method of steps*. The method for a linear system is described in Hale and Verduyn-Lunel (1993). We explain the method

here for a general RHS. Consider (2.1) with a single finite delay τ:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}(t), \mathbf{x}(t-\tau), p). \tag{2.2}$$

On the interval $t \in [0, \tau]$, the function $\mathbf{x}(t)$ is the solution of the ordinary differential equation initial value problem (ODE IVP) with forcing:

$$\frac{d\mathbf{x}_1}{dt} = f(\mathbf{x}_1(t), \mathbf{x}_0(t-\tau), p), \qquad \mathbf{x}_1(0) = \mathbf{x}_0(0), \tag{2.3}$$

where $\mathbf{x}_1(t) = \mathbf{x}(t)$ for $t \in [0, \tau]$. Once the solution is known on the interval $t \in [0, \tau]$, $\mathbf{x}(t)$ on the interval $t \in [\tau, 2\tau]$ is the solution of

$$\frac{d\mathbf{x}_2}{dt} = f(\mathbf{x}_2(t), \mathbf{x}_1(t), p), \qquad \mathbf{x}_2(0) = \mathbf{x}_1(\tau), \tag{2.4}$$

where $\mathbf{x}_2(t) = \mathbf{x}(t+\tau)$ for $t \in [0, \tau]$. If we continue this process we end up with a system of ODE IVPs:

$$\frac{d\mathbf{x}_1}{dt} = f(\mathbf{x}_1(t), \tilde{\mathbf{x}}_0(t), p), \qquad \mathbf{x}_1(0) = \tilde{\mathbf{x}}_0(\tau), \tag{2.5a}$$

$$\frac{d\mathbf{x}_2}{dt} = f(\mathbf{x}_2(t), \mathbf{x}_1(t), p), \qquad \mathbf{x}_2(0) = \mathbf{x}_1(\tau), \tag{2.5b}$$

$$\dots \tag{2.5c}$$

$$\frac{d\mathbf{x}_N}{dt} = f(\mathbf{x}_N(t), \mathbf{x}_{N-1}(t), p), \qquad \mathbf{x}_N(0) = \mathbf{x}_{N-1}(\tau), . \tag{2.5d}$$

Note that $\tilde{\mathbf{x}}_0(t) = \mathbf{x}_0(t-\tau)$ is just a time shift of the initial history. If we change the variables back to the original solution through $\mathbf{x}_i(t) = \mathbf{x}(t+(i-1)\tau)$ for $i = 1, 2, \dots, N$, we then have a unique solution for (2.2) on the interval $[-\tau, N\tau]$. If we let $N \to \infty$, we then can recover a unique solution for (2.2) on $[-\tau, \infty)$.

To investigate the local stability of the solutions we analyse the linearisation of (2.2). In DDE systems with a finite delay τ, the linearisation with linearised variable $\hat{\mathbf{x}}$ has the form

$$\frac{d\hat{\mathbf{x}}}{dt} = J_0\hat{\mathbf{x}}(t) + J_\tau\hat{\mathbf{x}}(t-\tau), \tag{2.6}$$

where $J_0 \in \mathbb{R}^{n\times n}$ and $J_\tau \in \mathbb{R}^{n\times n}$ are the Jacobians of the right hand side of (2.2) with respect to $x(t)$ and $x(t-\tau)$, respectively. The linear DDE (2.6) has non-trivial solution of the form $\mathrm{e}^{\lambda t}c$ (for $c \in \mathbb{C}$) if and only if $\delta(\lambda) = 0$, where $\delta(\lambda) : \mathbb{C} \to \mathbb{C}$ defines the characteristic function and is given as follows:

$$\delta(\lambda) = \det(\lambda I - J_0 - J_\tau \mathrm{e}^{-\lambda\tau}). \tag{2.7}$$

Here, I is the $n \times n$ identity matrix. A more general form of the characteristic function for DDEs with unbounded delays can be found in Stépán (1989).

In general, the characteristic function of a DDE has an infinite number of zeros. For finite delays, however, we can make some statements about the behaviour of the zeros. First, we know that for a finite delay, $J_0 + J_\tau e^{-\lambda\tau}$ is finite. This then leads to Theorem 1.7 of Stépán (1989), which states that for every $M \in \mathbb{R}$, there are at most finitely many λ with $\delta(\lambda) = 0$ and $\Re(\lambda) \geq M$. We can make statements about stability of the DDE (2.6) by studying the zeros of the characteristic function (2.7). Stability of the linearised DDE (2.6) then provides information about the local stability of DDE (2.2).

Bifurcation theory, or the study of qualitative change in solution behaviour, of DDEs starts with the study of the behaviour near equilibria. Suppose there exists an equilibrium of (2.2), $\mathbf{x}_{\text{eq}}$, with zeros of $\delta(\lambda) = \det(\lambda I - J_0(\mathbf{x}_{\text{eq}}) - J_\tau(\mathbf{x}_{\text{eq}})e^{-\lambda\tau})$ denoted as $\lambda_{\text{eq},k}$. This equilibrium is called *hyperbolic* if $\Re(\lambda_{\text{eq},k}) \neq 0$ for all $k = 1, 2, \ldots, \infty$. If there exists at least one $k \in \{1, 2, \ldots, \infty\}$ such that $\Re(\lambda_{\text{eq},k}) = 0$, then the equilibrium is called *non-hyperbolic*. We denote j the number of $\lambda_{\text{eq},k}$ which satisfy $\Re(\lambda_{\text{eq},k}) = 0$. Since $j < \infty$, we can construct a center manifold of dimension j in a neighbourhood of the equilibrium $\mathbf{x}_{\text{eq}}$ (Diekmann et al., 1995). This then allows in this neighbourhood for the restriction of the infinite dimensional DDE (2.2) onto a system of ODEs of finite dimension. Through this dimension reduction, ODE bifurcation theory can then be applied in the neighbourhood of the equilibrium (see Section 2.3.1.1 for an introduction to ODE bifurcation theory).

The existence of periodic solutions can be shown through the construction of a Poincaré map as with ODEs. A solution $\mathbf{x}(t)$ of system (2.2) is called periodic if $\mathbf{x}(t)$ is non-constant and there exists $P > 0$ such that $\mathbf{x}(t) = \mathbf{x}(t + P)$ for all $t \in [-\tau, \infty)$ (Diekmann et al., 1995). A Poincaré map can then be constructed, and any fixed points represent periodic orbits (Hale and Verduyn-Lunel, 1993). For every $\rho > 0$ a fixed point will have finitely many eigenvalues μ with $|\mu| > \rho$, such that one can construct finite-dimensional center manifolds. If there exists a non-hyperbolic fixed point of the Poincaré map, then the center manifold theory and bifurcation theory of ODEs can also be applied.

2.2 The use of delays in climate models

Delay differential equations (DDEs) have been used in many different fields of study including population biology, physiology, economics, neural networks, and control of mechanical systems (Smith, 2010). They are typically used to describe delayed feedback mechanisms, i.e. dynamical variables that depend on past states of the system. The need for an initial value in $C([-\tau, 0]; \mathbb{R}^n)$, representing past history, makes systems of DDEs infinite-dimensional (Hale and Verduyn-Lunel, 1993). This added complexity is sometimes criticised in regards to incorporating DDEs into climate models (which are typically already in the form of PDEs, another infinite-dimensional problem). Nevertheless, they can be useful in reducing the state space of partial differential equation (PDE) systems by modelling

feedback loops as DDEs for only a few variables representing the direct effects and the result of the feedbacks (Krauskopf and Sieber, 2014). Methods of reducing such systems are discussed in detail in Chapter 3.

There are a few examples where delays have been incorporated into climate models, typically with conceptual models. For instance, delays can be used to approximate the transport of mass or energy across large distances (Dijkstra, 2008). An early attempt at this is proposed as an energy-balance model (EBM), which models the radiation balance at the earth's surface. For the incoming solar radiation, a delayed albedo term is incorporated. This is typically attributed to the delay in ice-sheet extent due to viscoplastic expansion (discussed in more detail in Chapter 4). The delayed EBMs produce self-sustained oscillations in temperature that are either periodic or aperiodic, and have comparable spectra to paleoclimate data. We show a few examples of these models in Section 2.2.1.

Another prominent example of delays incorporated into climate models involves the El Niño Southern Oscillation (ENSO). The ENSO phenomenon occurs in the equatorial Pacific Ocean and has a period of roughly 4 to 7 years. During this phenomenon, the sea-surface temperature (SST) in the Eastern Tropical Pacific Ocean increases by a few degrees (El Niño) which then drives variations in the surface winds across the tropical Pacific (the Southern Oscillation) (Dijkstra, 2008). Although the driving forces of ENSO are confined to the equatorial Pacific, the effects have a global extent. The delays arise from the time it takes waves to propagate in opposing directions across the equatorial basin. These waves transport perturbations to temperature and density balances. Section 2.2.2 presents some examples of models with delay that have been used to study ENSO, while in Chapter 3 we explore in detail a method of deriving a model with delay for the phenomenon.

2.2.1 Energy balance model with time delay

The first climate model with delay was an energy-balance model (EBM) proposed by Ghil and Bhattacharya (1979). Models of this type are used to describe changes in the earth's average temperature, with the simple processes of the earth receiving solar radiation and emitting radiation as a blackbody. Ghil and Bhattacharya argued there was a time lag related to the growth and decay of ice sheets as a response to global temperature. The extent of the ice sheets affects the global albedo (or reflection of radiation) which in turn affects the global temperature. The authors found that the model can exhibit self-sustained oscillations, which they suggest could have influenced glacial-interglacial cycles of the Pleistocene (2.6 Myr - 11 kyr BP).

Another delayed EBM was suggested by Andersson and Lundberg (1988) as

$$C\frac{dT}{dt} = Q(1 - \alpha(T(t-\tau))) - \sigma g(T(t))T(t)^4. \tag{2.8}$$

The variable T represents annually-averaged global temperature. The parameters represent globally averaged heat capacity C, mean solar radiative input Q, and the Stefan-Boltzmann constant σ. The functions $\alpha(T)$ and $g(T)$ represent the influence of albedo and emissivity (emission of energy as thermal radiation) respectively. In this model, the delayed albedo function $\alpha(T(t-\tau))$ incorporates increased cloudiness due to changes in hydrological cycles at increased global temperatures as well as the changes in ice sheets. The function $\alpha(T)$ is piecewise and predominantly linear (see Andersson and Lundberg (1988) for precise function). Model (2.8) can exhibit self-sustained oscillations along with chaotic solutions.

Bar-Eli and Field (1998) proposed a model of similar form to (2.8),

$$C\frac{dT}{dt} = \frac{S}{4}(1-\alpha_c T(t-\tau))(1-\alpha_s T(t-\tau)) - \sigma g(T(t-\tau))T(t)^4. \tag{2.9}$$

where S is the constant insolation of sun and $\frac{S}{4} \approx Q$. The main differences to (2.8) is in the albedo and emissivity functions. Model (2.9) uses a continuous nonlinear function for the albedo effects with a different albedo constant for clouds (α_c) and surface (α_s). The main difference in the emissivity function $g(T)$ is that in (2.9) there is also a delayed effect. With this model Bar-Eli and Field (1998) find square-wave oscillations and multiple paths to chaos. One critique of the model, however, is that the delays with respect to the three processes (cloud albedo, surface albedo, and emmisivity) are likely not to be equal. The effect of using different delay times on (2.9) has not been studied.

Other studies of EBMs with delays include Bhattacharya et al. (1982) and Hidalgo López et al. (2014). In Chapter 4 and Chapter 5 we expand on the use of different types of models with delay for the Pleistocene in particular.

2.2.2 Conceptual models with delay for ENSO

One of the first models with delay proposed for ENSO was introduced by Suarez and Schopf (1988). The authors suggest a scalar model in perturbations of sea-surface temperature (SST) in the eastern part of the equatorial ocean basin. The model is as follows,

$$\frac{dT}{dt} = T(t) - T(t)^3 - \alpha T(t-\delta), \tag{2.10}$$

where first term of (2.10) represents the positive feedback related to the heat exchange of the sea surface with the atmosphere. The second term is an unspecified nonlinear mechanism meant to limit the growth of unstable perturbations from equilibrium. While the form of this term is suggested without any particular justification in Suarez and Schopf (1988), it was found that this specific nonlinear term can be derived from a projection of the equations of motion using the methods presented in Section 3.2 (Falkena et al., 2018). The final term in (2.10) is where the delayed feedback arises. This represents

the influence of travelling waves in the ocean. An SST perturbation in the east will effect the wind forcing in the atmosphere. This will then drive a wave in the thermocline to travel westward, reflect at the western boundary of the basin, and travel back eastward where it interacts with the model SST again. The full travel time of that wave is captured by delay δ, and α represents the strength of the delayed effects compared to the local effects. The model of Suarez and Schopf (1988) showed stable oscillations with a period of 3-4 years. This is shorter than the observed ENSO cycles, however it shows that the delayed feedback mechanism can cause stable oscillations on the time scales of interest.

Since Suarez and Schopf (1988) there have been numerous other conceptual models with delay proposed for the phenomenon. A model of the same form but with different parameter coefficients was derived by Battisti and Hirst (1989). Other conceptual models have focused on the dynamics of the thermocline rather than SST. Tziperman et al. (1994) were the first to suggest a continuous model with delay of this kind. The authors focus on displacement of the thermocline at the eastern boundary of the equatorial ocean basin. The model is given as

$$\frac{dh}{dt} = aA(\kappa, h(t-\delta_1)) - bA(\kappa, h(t-\delta_2)) + c\cos(\omega t), \tag{2.11}$$

and is based on a discrete model with delay derived by Cane et al. (1990). Here h is the thermocline displacement, δ_1 and δ_2 are the Kelvin (eastward) and Rossby (westward) wave travel times respectively, κ is the ocean-atmosphere coupling strength, and ω is the frequency for the yearly seasonal forcing. The nonlinear function $A(\kappa, h)$ is used to model the effect of wind stress forcing on the thermocline through SST and ocean stratification. Many follow-up studies on this model have been done, including Keane et al. (2016), Zaliapin and Ghil (2010), in which chaotic behaviour can be observed. Tziperman et al. (1998) made a modification to the original model (2.11),

$$\frac{dh}{dt} = aA(\kappa(t-\delta_1), h(t-\delta_1)) - bA(\kappa(t-\delta_2), h(t-\delta_2)) - ch(t). \tag{2.12}$$

In this version the seasonal variability is captured in the time-varying coupling strength $\kappa(t)$, and the final term represents dissipation effects. A full bifurcation analysis of (2.12) was performed in Krauskopf and Sieber (2014). The authors observed phase-locking to the seasonal cycle as well as quasi-periodic behaviour.

It should be noted that there is a common critique of these ENSO models with delay in the assumption of wave reflection. Complete reflection (or even partial reflection) of the Rossby wave is not entirely physical, as the realistic topography of the edge of the ocean basin is far from a rigid wall.

2.3 Examples of critical transitions in climate

Critical transitions, in some cases known as "tipping points", have become an increasingly relevant area of research in the past decade. Events such as disease outbreaks, stock market crashes, and collapses of ecosystems are all examples of a system undergoing a critical transition (Scheffer, 2009).

The concepts of tipping points and critical transitions were introduced as scientific terms. The term "tipping point" generally refers to the phenomenon when a small perturbation of a system has large, long-term consequences, qualitatively changing its future state (Lenton et al., 2008), whereas Scheffer (2009) reserves the term "critical transition" for "transitions in which a positive feedback pushes a runaway change to a contrasting state once a threshold is passed." In this regard, it does not make sense to attempt a mathematical definition for the terms. However, the phenomenon of tipping in mathematical models is well-studied and is covered by some general mathematical mechanisms. The three most common underlying mathematical mechanisms are

- slow passage through a bifurcation (*bifurcation-induced*),
- exit from a basin of attraction due to noise (*noise-induced*), and
- failure to track a quasi-static attractor (*rate-induced*).

In this section we explain some well-understood examples of critical transitions. First we introduce the mathematics behind the three types of critical transitions listed above. Then we discuss two specific examples of a critical transition in the climate system, the Mid-Pleistocene Transition and the shutdown of the Atlantic meridional overturning circulation.

2.3.1 Mathematical mechanisms for critical transitions

We discuss three types of critical transitions introduced by Ashwin et al. (2012): bifurcation-induced (b-tipping), noise-induced (n-tipping), and rate-induced (r-tipping).

2.3.1.1 Bifurcation-induced transitions

The classical theory of critical transitions relates to dynamical bifurcations. We first look at the simplest types of local bifurcations which can be found in systems of dimension $n \geq 1$: *fold*, *transcritical*, and *pitchfork*. The general theory related to bifurcations can be found in Kuznetsov (1998) and Perko (2013).

Consider a dynamical system

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, p). \tag{2.13}$$

Here we consider an n-dimensional vector of state variables $\mathbf{x}$, and a scalar parameter p. An equilibrium $\mathbf{x}_0$ for parameter p_0 is defined such that $f(\mathbf{x}_0, p_0) = 0$. We notate the Jacobian matrix of (2.13) as $Jf(\mathbf{x}, p)$.

We introduce a non-degeneracy condition,

$$\frac{\partial^2}{\partial \mathbf{x}^2} f(\mathbf{x}_0, p_0) \neq 0, \tag{2.14}$$

and a transversality condition,

$$\frac{\partial}{\partial p} f(\mathbf{x}_0, p_0) \neq 0. \tag{2.15}$$

We will assume these conditions hold for the following three definitions.

Definition 2.1. Suppose there exists an equilibrium $\mathbf{x}_0$ for parameter $p_0 = p_{\text{crit}}$ such that $\det(Jf(\mathbf{x}_0, p_{\text{crit}})) = 0$, i.e. the equilibrium is non-hyperbolic, and $Jf(\mathbf{x}_0, p_{\text{crit}})$ has an eigenvalue $\lambda_{\text{crit}} = 0$. If there exist exactly two hyperbolic equilibria arbitrarily close to $\mathbf{x}_0$ for $p \to p_{\text{crit}}^-$ (or $p \to p_{\text{crit}}^+$), and there exist no hyperbolic equilibria arbitrarily close to $\mathbf{x}_0$ for $p \to p_{\text{crit}}^+$ (or $p \to p_{\text{crit}}^-$), then the system (2.13) has a *fold bifurcation* at $p = p_{\text{crit}}$.

A normal form for a bifurcation is the simplest form of a system which is locally topologically equivalent near $(\mathbf{x}_0, p_0)$ to any system (2.13) with equilibrium $\mathbf{x}_0$ satisfying the same bifurcation conditions at p_0 (Kuznetsov, 1998, Def. 2.16). A normal form for a fold bifurcation is given as

$$\frac{dx}{dt} = p \pm x^2, \tag{2.16}$$

where in this case state variable x is scalar. Figure 2.1a shows the bifurcation diagram for the negative case of (2.16) near its fold bifurcation.

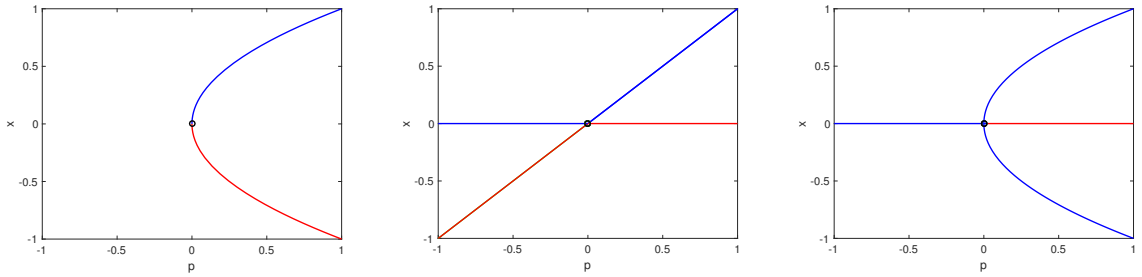
Definition 2.2. Suppose there exists an equilibrium $\mathbf{x}_0$ for parameter $p_0 = p_{\text{crit}}$ such that $\det(Jf(\mathbf{x}_0, p_{\text{crit}})) = 0$, i.e. the equilibrium is non-hyperbolic, and $Jf(\mathbf{x}_0, p_{\text{crit}})$ has an eigenvalue $\lambda_{\text{crit}} = 0$. If there exist exactly two hyperbolic equilibria arbitrarily close to $\mathbf{x}_0$ for $p \to p_{\text{crit}}$, then the system (2.13) has a *transcritical bifurcation* at $p = p_{\text{crit}}$.

A normal form for a transcritical bifurcation (Kuznetsov, 1998) is given as

$$\frac{dx}{dt} = px \pm x^2. \tag{2.17}$$

Figure 2.1b shows the bifurcation diagram for the negative case of (2.17) near its transcritical bifurcation.

Definition 2.3. Suppose there exists an equilibrium $\mathbf{x}_0$ for parameter $p_0 = p_{\text{crit}}$ such that $\text{trace}(Jf(\mathbf{x}_0, p_{\text{crit}})) = 0$, i.e. the equilibrium is non-hyperbolic, and $Jf(\mathbf{x}_0, p_{\text{crit}})$ has an eigenvalue $\lambda_{\text{crit}} = 0$. If there exists



(a) **(b)** **(c)**

Figure 2.1: Bifurcation diagrams for the normal forms of (a) fold, (b) transcritical, and (c) pitchfork bifurcations. The blue curves show stable equilibria, red curves show unstable equilibria, and the black circle denotes the respective bifurcation point.

exactly one hyperbolic equilibrium arbitrarily close to $\mathbf{x}_0$ for $p \to p_{\text{crit}}^-$ (or $p \to p_{\text{crit}}^+$), and there exist exactly three hyperbolic equilibria arbitrarily close to $\mathbf{x}_0$ for $p \to p_{\text{crit}}^+$ (or $p \to p_{\text{crit}}^-$), then the system (2.13) has a *pitchfork bifurcation* at $p = p_{\text{crit}}$.

A normal form for a pitchfork bifurcation (Kuznetsov, 1998) is given as

$$\frac{dx}{dt} = px \pm x^3. \tag{2.18}$$

Figure 2.1c shows the bifurcation diagram for the negative case of (2.18) near its pitchfork bifurcation.

A bifurcation that can occur for systems of dimension $n \geq 2$ is the *Hopf bifurcation.*

Theorem 2.4. *Suppose there exists an equilibrium* $\mathbf{x}_0$ *for parameter* $p_0 = p_{\text{crit}}$ *such that* $Jf(\mathbf{x}_0, p_{\text{crit}})$ *has exactly two purely imaginary eigenvalues* $\lambda_\pm = \pm i\omega(p_{crit})$*, i.e. the equilibrium is non-hyperbolic. If there exists exactly one hyperbolic equilibrium arbitrarily close to* $\mathbf{x}_0$ *for* $p \to p_{\text{crit}}$*, and there also exists a limit cycle which approaches* $\mathbf{x}_0$ *for* $p \to p_{\text{crit}}^-$ *or* $p \to p_{\text{crit}}^+$*, then the system* (2.13) *has a* Hopf bifurcation *at* $p = p_{\text{crit}}$*. Note that the single equilibrium must change stability as* p *crosses* p_{crit}*.*

We assume the following non-degeneracy and transversality conditions (respectively) hold for the above theorem:

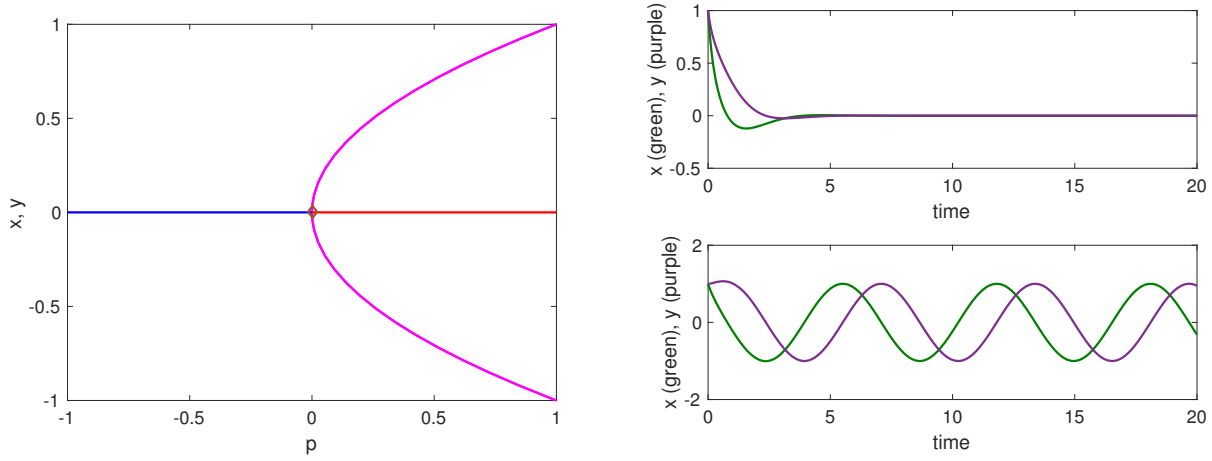
$$l_1(p_{\text{crit}}) \neq 0, \qquad \frac{d}{dp}\mu(p_{\text{crit}}) \neq 0, \tag{2.19}$$

where $l_1(p)$ it the first Lyapunov coefficient (see precise formula for this in Kuznetsov (1998)) and $\mu(p_{\text{crit}}) = 0$ is the real part of the set of purely imaginary eigenvalues of the system for $p_0 = p_{\text{crit}}$.

A normal form for a Hopf bifurcation (Kuznetsov, 1998) is given as

$$\frac{dx}{dt} = -y + x(p \pm (x^2 + y^2)), \tag{2.20a}$$

$$\frac{dy}{dt} = x + y(p \pm (x^2 + y^2)). \tag{2.20b}$$



(a) **(b)**

Figure 2.2: (a) Bifurcation diagram for the normal form of a Hopf bifurcation. The blue curves show stable equilibria, red curves show unstable equilibria, pink curves show maximum and minimum of a periodic orbit, and the brown diamond denotes the Hopf bifurcation point. (b) Example trajectories for (2.20a) when $p = -1$ (top) and $p = 1$ (bottom). Notice the oscillatory behaviour of solution after the Hopf bifurcation.

Figure 2.2a shows the bifurcation diagram for the negative case of (2.20a) near its Hopf bifurcation.

Depending on the physical system in consideration, any of the above bifurcations can induce what could be called a critical transition (see Figure 2.1a-2.2a). Kuehn (2011) gives precise mathematical arguments as to under which conditions the above bifurcations are considered critical transitions. Note that we have only discussed codimension-1 local bifurcations, the codimension being the number of independent conditions determining the bifurcation (Kuznetsov, 1998, Def. 2.13). There exist also global bifurcations, such as homoclinic connections, and codimension-2 bifurcations, such as cusps and Bogdanov-Takens points. We don't go into detail about these bifurcations here, but we will discuss them briefly in Chapter 5 and Chapter 6.

2.3.1.2 Noise-induced transitions

Not all transitions are caused by bifurcations. Stability properties of a system can also be affected by noise, causing a steady state to switch from stable to unstable, or vice versa (Horsthemke, 1984). Ashwin et al. (2012) defined noise-induced transitions (or n-tipping) as when "noisy fluctuations result in the system departing from a neighbourhood of a quasi-static attractor." In other words, random external input to a system can potentially knock it far enough away from a stable state where its internal dynamics no longer allow for the return to that state.

We demonstrate the occurrence of a noise-induced transition using the Stommel box model. Stommel (1961) introduced a simple dynamical model for the thermohaline circulation. The thermohaline circulation is the movement of heat and salt throughout the world's oceans and it has a major impact

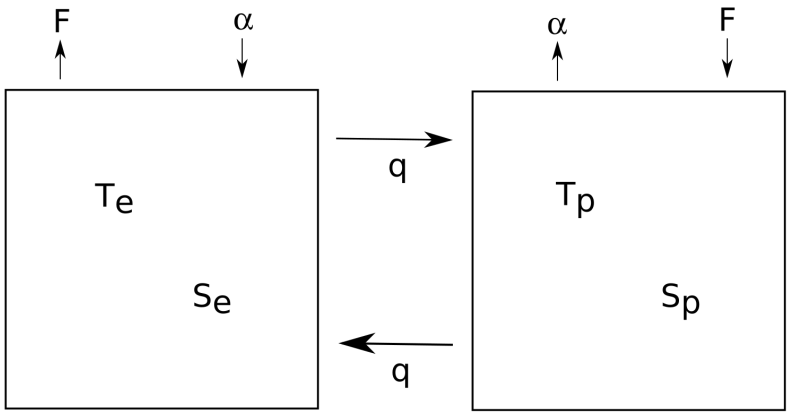


Figure 2.3: Diagram of the Stommel (1961) box model. There are two boxes, equatorial (e) and polar (p), which allow for flow between them. In each box temperature T and salinity S are measured. There is external freshwater forcing (F) in the polar box, and evaporation ($-F$) in the equatorial box. The equatorial box gains heat (α) while the polar box loses heat ($-\alpha$). The flow between boxes is represented by q.

on the earth's climate (Wunsch, 2002). Stommel attempted to understand the mechanism of circulation through an exchange between two connected reservoirs, or 'boxes'. A diagram of his model can be seen in Figure 2.3.

The thermohaline circulation is driven by spatial differences in temperature and salinity. We can then consider the equator to polar differences in these quantities as variables, namely $x \sim \Delta T = T_e - T_p$ and $y \sim \Delta S = S_e - S_p$. With some scaling and nondimensionalisation one can arrive at the following system (for full derivation see Dijkstra (2013)):

$$\frac{dx}{dt} = -\alpha(x-1) - q(x,y)x, \tag{2.21a}$$

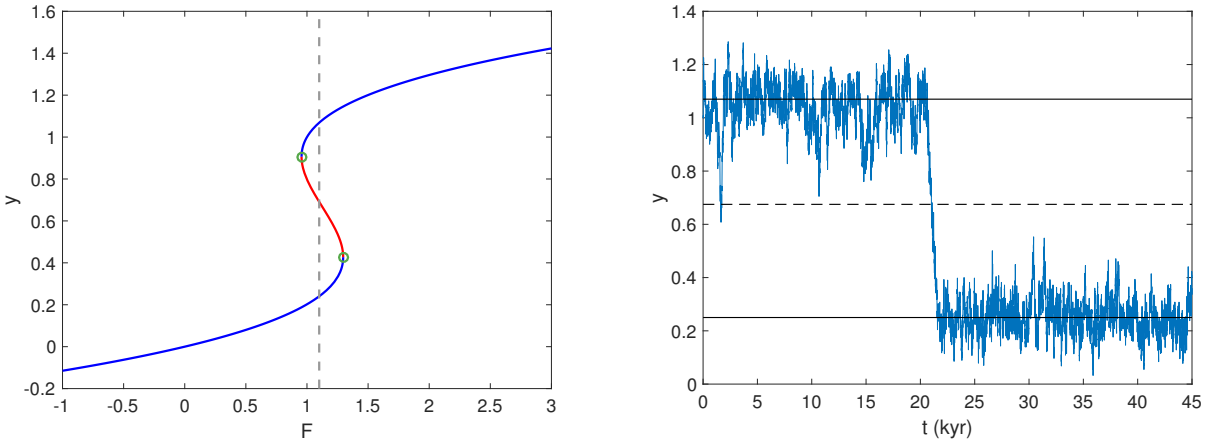
$$\frac{dy}{dt} = F - q(x,y)y, \tag{2.21b}$$

with transport function

$$q(x,y) = 1 + \mu^2(x-y)^2. \tag{2.22}$$

Since α is essentially a ratio of diffusion time to temperature relaxation time, we know $\alpha \approx 2600$ (Dijkstra, 2013). For $\alpha \gg 1$ the approximation $x = 1 + \mathcal{O}(1/\alpha)$ holds. Therefore system (2.21) simplifies to

$$\frac{dy}{dt} = F - y(1 + \mu^2(1-y)^2). \tag{2.23}$$



(a) **(b)**

Figure 2.4: (a) Bifurcation diagram for the reduced Stommel box model (2.23). The blue curves show stable equilibria, red curves show unstable equilibria, and the green circles denote the fold bifurcations. The grey dotted line is the value of F used in noise simulation. (b) Example trajectory for (2.24) when $\sigma = 0.15$. Black lines show stable (solid) and unstable (dotted) deterministic equilibria locations. The noise causes the solution to make a transition just after 20 kyr.

The parameter μ can be determined via known physical quantities which gives $\mu^2 = 6.2$ (Dijkstra, 2013). The parameter F is of much interest, as it represents the freshwater flux into the system. In the physical world this can represent melting of large amounts of glacial ice or changes in the hydrological cycle. As demonstrated in Section 2.3.1.1, we construct a bifurcation diagram for (2.23) when varying F, shown in Figure 2.4a. The system experiences two fold bifurcations when increasing F, leading to a region of bistability for $F \in [0.9556, 1.2962]$.

We explore what happens in this bistable region when subjected to noise perturbations. We transform (2.23) into a stochastic differential equation (SDE) by including additive white noise, represented by σdW_t, where σ is the noise strength:

$$dy = (F - y(1 + \mu^2(1 - y)^2))dt + \sigma dW_t. \tag{2.24}$$

We choose a value of freshwater forcing within the bistable region, $F = 1.1$. For an initial condition of $y_0 = 1$ and sufficiently small noise strength σ, we expect the solution of (2.24) to remain near the upper equilibrium of the bifurcation diagram for some time T, before $y(t)$ will eventually transition to the lower equilibrium (see Figure 2.4b for an example trajectory). In this noise realisation, the solution transitions just after 20 kyr.

The expectation of this (random) residence time T can be approximated using *Kramers' law* in the limit of $\sigma \to 0$ for SDEs where the deterministic part is the gradient of a potential V(y) ($f(y) = -\nabla V(y)$). Stable equilibria can be viewed as potential wells defined by the minima of the potential $V(y)$. The wells are connected through a local maximum of $V(y)$, which occurs at the unstable equilibrium. A

simple one-dimensional SDE has the form

$$dy = -V'(y)dt + \sigma dW_t. \tag{2.25}$$

For (2.24) the potential is given as

$$V(y) = -Fy + \frac{1+\mu^2}{2}y^2 - \frac{2\mu^2}{3}y^3 + \frac{\mu^2}{4}y^4. \tag{2.26}$$

When initialising the system in a potential well (near an equilibrium), we can analyse the average escape time from the well through Kramers' law. Kramers' law defines the average escape time from a potential well as follows (see Berglund and Gentz (2006) for further discussion). Consider initial conditions in the potential well of stable equilibrium y_{seq}. The average escape time τ from the corresponding well is given as

$$\tau = \frac{2\pi}{\sqrt{|V''(y_{\text{ueq}})|V''(y_{\text{seq}})}} \mathrm{e}^{\frac{2(V(y_{\text{ueq}})-V(y_{\text{seq}}))}{\sigma^2}} [1 + O(\frac{\sigma^2}{2})], \tag{2.27}$$

where y_{ueq} is the unstable equilibrium and

$$V''(y) = -1 + \mu^2 - 4\mu^2 y + 3\mu^2 y^2. \tag{2.28}$$

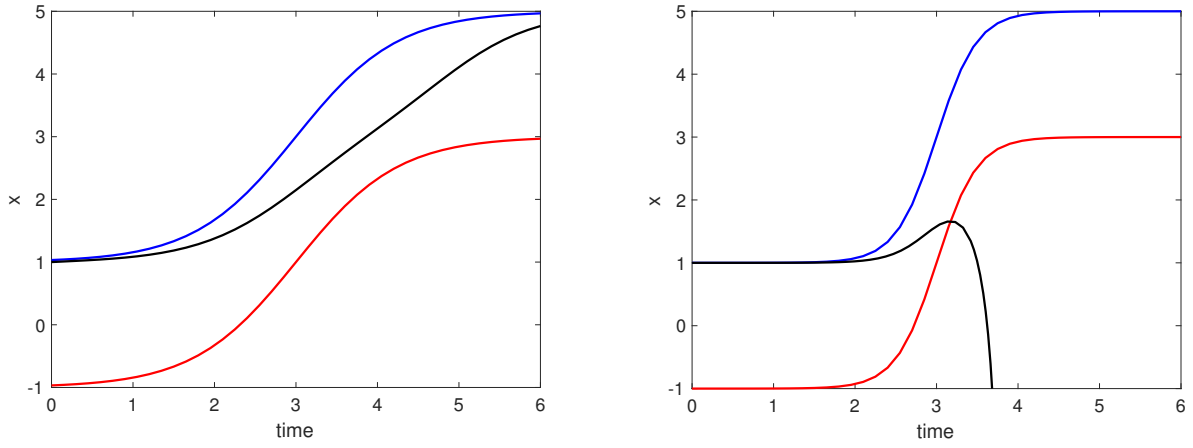
We apply (2.27) to our system. For the upper stable equilibrium $y_{\text{seq}+} = 1.07$ the average escape time is then $\tau_+ \approx 85$ kyr, while for the lower stable equilibrium $y_{\text{seq}-} = 0.25$ the average escape time is $\tau_- \approx 542$ kyr. We therefore conjecture that the lower equilibrium is a more attracting state than the upper equilibrium.

2.3.1.3 Rate-induced transitions

There are also transitions not related to bifurcations or noise, but rather to the speed of nonautonomous changes of parameters in the system. These are known as rate-induced transitions. Ashwin et al. (2012) give the definition as when a system "fails to track a continuously changing quasi-static attractor". A quasi-static attractor is the stable solution for a system given the parameters are considered constant. When a parameter is shifted in time, the stable solution will also change in time. If the shift happens at a critical rate (system dependent), then the solution will fail to continue to track the stable solution and will diverge to another stable solution or to infinity. A formal definition can be found in Ashwin et al. (2017), however in this section we just show an example to illustrate a rate-induced transition.

We use an example system similar to the one used in Ashwin et al. (2012), the fold normal form with drift λ:

$$\frac{dx}{dt} = 1 - (x - \lambda(r,t))^2. \tag{2.29}$$



(a) **(b)**

Figure 2.5: Illustration of a rate-induced transition for system (2.29). Blue curves show the stable quasi-static equilibrium, red curves show the unstable quasi-static equilibrium, and black curve shows solution trajectory. (a) $r = 0.4$, solution tracks quasi-static equilibria. (b) $r = 1$, solution diverges.

We will take $\lambda(r, t)$ of the form

$$\lambda(rt) = 2(\tanh(2rt - 6) + 1). \tag{2.30}$$

This induces a shift of λ from $\lambda_- = 0$ to $\lambda_+ = 4$ with the steepness of the shift determined by r and centered at time $t = 3$. In Figure 2.5 we show the effect of two different values of r. For smaller values of r, the system diverges slightly from the stable quasi-static equilibrium but then recovers (see Figure 2.5a). After a critical rate of $r_{\text{crit}} = 0.5$, the system no longer can recover back to the stable quasi-static equilibrium and diverges. In this particular example, this can be understood by the solution crossing the unstable quasi-static equilibrium (see Figure 2.5b).

As in Section 2.3.1.2, we see again that a solution can transition without passing through a bifurcation (notice there are no stability changes to the quasi-static equilibria in either case of Figure 2.5). All three types of transitions introduced should be considered when analysing an observed transition in data. In the next sections we will explore evidence of specific transition events in the climate system.

2.3.2 The Mid-Pleistocene Transition

The Pleistocene, which began around 2.6 million years ago, is characterised by its successive glacial-interglacial cycles whose documented periodicities have baffled scientists for years. It is generally accepted that orbital forcing plays a large role in the onset and termination of glacial periods (Ashwin and Ditlevsen, 2015, Ganopolski and Calov, 2011, Paillard, 1998). This however does not explain all of the variations, as there are unexplained phenomena such as the prominent 100,000 year timescale of glacial episodes which is only a small signal in the orbital forcing. These different phenomena will be explored in this section, as well as the developments in the related research over the past century.

While current climate conditions can be studied through local observations and satellite data, a different approach must be used to understand the Earth's past states. The paleoclimate record is comprised of many different proxy records. A proxy is a measurable quantity found in different natural materials and organisms that can be related in some way to a climate variable (Dijkstra, 2013). These proxies are found everywhere from ice cores to marine sediments. One commonly used proxy is δ^{18}O. The quantity δ^{18}O measures the ratio of oxygen-18 to oxygen-16 stable isotopes in a sample. Benthic foraminifera record δ^{18}O, which is related to the δ^{18}O in seawater, and can be used to infer global ice volume and ocean salinity as far back as 5.3 Myr BP (Lisiecki and Raymo, 2005). A higher ratio in benthic foraminifera corresponds to greater volumes of ice cover on the planet. A reconstruction of the past climate according to the Lisiecki and Raymo (2005) analysis is shown in Figure 2.6. These $\delta^{18}O$ isotopes can also be measured through ice cores from Greenland and Antarctica.

Another common proxy is δ^{13}C, which measures the ratio of stable isotopes carbon-13 to carbon-12. This proxy is also recorded by the carbonate shells of benthic foraminifera. It measures the attribution of carbon to different carbon reservoirs on land and sea. Organic matter is more rich in carbon-12 than inorganic reservoirs of carbon due to the fact that CO^2 containing carbon-12 is more easily processed during photosynthesis. Since carbon-13 is heavier than carbon-12 it forms slightly stronger chemical bonds, and due to the differences in molecular mass, CO^2 containing carbon-13 diffuses slower (Farquhar et al., 1989). Therefore, increased vegetation on land would imply higher δ^{13}C levels in the ocean. Similarly, an increase of the export of organic matter from the ocean's surface layer to the deep sea would lead to a decrease of ocean δ^{13}C (Indermühle et al., 1999). The proxy is sensitive to orbital precession, as well as overall glacial-interglacial cycles.

Some more recent studies have shown that δ^{15}N can actually be a good paleothermometer. On shorter timescales, such as the Younger Dryas period (12,900 - 11,700 BP), the changes in nitrogen isotopes can be more accurate in identifying warming and cooling events than the classic δ^{18}O records (Severinghaus et al., 1998). δ^{15}N measures the ratio of nitrogen-15 to nitrogen-14 stable isotopes. Severinghuas et al. argue nitrogen isotopes better estimate the magnitude of warming or cooling events, where δ^{18}O has been shown to underestimate such temperature changes (Cuffey et al., 1995).

One actively researched aspect of the Pleistocene is the abrupt change in frequency of major glaciations known as the Mid-Pleistocene Transition (MPT) (Ashwin and Ditlevsen, 2015, Clark et al., 2006, Crucifix, 2012, Maasch, 1988, Maasch and Saltzman, 1990, Paillard, 1998, Pisias and Moore, 1981). Figure 2.6 indicates the time frame for the occurrence of the MPT. Spectral analysis has been performed on this time series, and it has been observed that the signal of the 100-kyr cycle began to rise 1250 kyr BP and was established as the dominant cycle by 700 kyr BP (Clark et al., 2006, Dijkstra, 2013, Lisiecki and Raymo, 2007). Many researchers have tried to determine not only the cause of this switch, but also the driving force behind the 100 kyr cycles themselves (Ashwin and Ditlevsen, 2015, Ganopolski and Calov, 2011, Gildor and Tziperman, 2001, Maasch and Saltzman, 1990, Paillard, 1998,

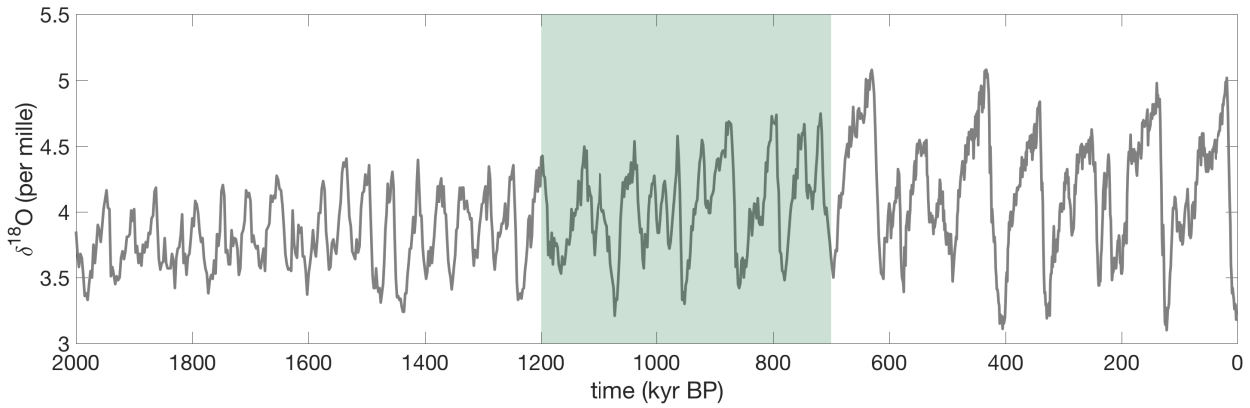


Figure 2.6: Reconstruction of benthic δ^{18}O from 57 globally distributed sediment cores, adapted from Lisiecki and Raymo (2005) data. The green shaded area represents the time frame of the Mid-Pleistocene Transition.

Paillard and Parrenin, 2004, Saltzman and Maasch, 1991). While the 41-kyr oscillations have been attributed to external forcing (Paillard, 2001), the 100-kyr cycles have been proposed to be a result of nonlinear responses of the climate system itself (Gildor and Tziperman, 2001, Imbrie et al., 1993, Yiou et al., 1994).

The major glacial-interglacial cycles were originally attributed to changes in insolation due to orbital variations (i.e. the Milankovitch cycles). Milankovitch argued that summer insolation at high northern latitudes determined the main pacing of glaciations, as these changes control the length of seasons and amount of solar energy being received in high latitudes where the ice resides. The most prominent frequencies of these changes are 19 and 23 kyr due to precession and 41 kyr due to obliquity (Paillard, 2001). Precession is the orientation of the rotational axis of the earth (axial) and the rotation of the orbital axis (apsidal), and obliquity is the angle between the rotational axis and the orbital axis. Milankovitch (1941) argued that the main driving forces of climate fluctuations are obliquity and precession, which agrees with the records prior to the mid-Pleistocene transition. There is a smaller signal of orbital forcing that correlates with the 100-kyr oscillations (Imbrie et al., 1993). Eccentricity, the amount the earth's orbital ellipse deviates from a circle, shifts with a main period of 413 kyr, but there are components that vary with periods between 95 kyr and 125 kyr. The amplitude of these forcing signal bands, however, are an order of magnitude smaller than the signals of the 23- and 41-kyr bands (Hays et al., 1976).

Since Milankovitch's theory, there have been many other attempts to identify the cause of the MPT. Some of the most popular examples are the Saltzman models of the late twentieth century. They were a collection of attempts to model the transition as a bifurcation in low-order dynamical systems due to climate feedbacks. One example of these models, the Saltzman and Maasch (1988) model, is discussed in more detail in Chapter 5. Shortly after the Saltzman models, a non-smooth model consisting of three distinct climate states was introduced by Paillard (1998). In his model, the climate shifts from

interglacial (**i**) to mild glacial (**g**), **g** to full glacial (**G**), and **G** back to **i** based on prescribed insolation and ice volume levels. Recent work has represented this model as a relaxation oscillator with smooth transitions between states (Ashwin and Ditlevsen, 2015). It is shown from this model that the 100-kyr cycles emerge as a response to a transcritical bifurcation along a slow manifold. Another popular mechanism amongst climate modellers is a sea ice switch proposed by Gildor and Tziperman (2001). Through use of a box model of intermediate complexity, sea ice is used to switch the climate system from a growing land ice mode to a retreating land ice mode. This allows for the saw-tooth structure of the oscillations like that seen in proxy data, as well as the 100-kyr periodicity obtained without external forcing. An even more recent study by Ganopolski and Calov (2011) suggests the 100-kyr cycles are related to levels of regolith (terrestrial sediments covering bedrock). The presence of regolith enhances the velocity of ice sheet sliding and increases glaciogenic dust lowering the surface albedo (MacAyeal, 1992). It is hypothesised in that the MPT was a result of gradual removal of regolith, allowing more time for the glaciers to build in volume, and the terminations of glacial periods are phase-locked with orbital forcing.

While most of these theories are good attempts at solving the "100-kyr problem", none of them perfectly encapsulate all of the necessary dynamics and qualities of the system. A more in-depth review of the theories can be found in Crucifix (2012). There is still much room for progress in researching this particular event in climate history.

2.3.3 The potential shutdown of the Atlantic meridional overturning circulation

The Atlantic meridional overturning circulation (AMOC) is responsible for a large amount of transport of heat and salt from the tropics to the northern seas, where then this heat is released to the atmosphere, cooling the water and making it more dense, leading to sinking and deep-water flow southwards (Hawkins et al., 2011). It is also the main transfer of heat from the Southern Hemisphere to the Northern Hemisphere across the equator (Dijkstra, 2008). Because of these processes it has a vital role in regulating the earth's climate (Cheng et al., 2013). Impacts due to changes in the AMOC have been seen in present day climate studies, such as North Atlantic storm tracks (Woollings et al., 2012) and North American and European summer climate (Sutton and Hodson, 2005), but also in paleoclimate studies relating to Dansgaard-Oeschger events (Clement and Peterson, 2008, Ganopolski and Rahmstorf, 2001, Shaffer et al., 2004).

These studies allude to the existence of bistability in the AMOC, namely an 'on' and 'off' state. This was first observed in the simplest of ocean circulation models: the Stommel (1961) box model (discussed in Section 2.3.1.2). In Stommel's model, the bistability is caused by a salinity advection feedback mechanism which can be disturbed by additional freshwater forcing, particularly in the North Atlantic Ocean. This mechanism is particularly of interest in the present day, as the Greenland Ice Sheet is already shrinking at an accelerating rate (Lenton et al., 2008). Stommel's box model has

since been further developed and analysed to varying levels of complexity (Lucarini and Stone, 2005, Tziperman and Gildor, 2002), and the bistability has been observed in general circulation models (GCMs) (Dijkstra, 2007, Hawkins et al., 2011, Marotzke and Willebrand, 1991, Rahmstorf, 1996, Stocker and Wright, 1991, Weaver and Hughes, 1994).

Studies with GCMs have also begun to look at the widespread effects of AMOC shutdown. In addition to the impacts listed above, a recent study has shown that the AMOC 'off' state affects ENSO, shifting the spatial pattern eastward and increasing the oscillation period (Williamson et al., 2018). Precipitation patterns have also been shown to be affected across Europe (Jackson et al., 2015, Jacob et al., 2005), as well as in South America and Africa (Vellinga and Wood, 2002).

Because of the implications of an AMOC shutdown, it is important to be able to predict when it could occur in order to prepare for climate changes. To identify the approach towards a 'tipping point', one can turn to early warning signal theory for models with stochastic perturbations. Scheffer et al. (2009) gives an introduction for the theory of early warning signals for tipping points, particularly related to time series analysis. Some of the tools suggested by the authors include slower recovery from perturbations, increased autocorrelation, increased variance, increased skewness, and increased cross-correlation (spatial coherence). A few studies have looked into implications of these signals in one specific AMOC model: the fully coupled atmosphere-ocean general circulation model FAMOUS (Smith et al., 2008). Boulton et al. (2014) find increased autocorrelation and variance which is latitude dependent. Feng et al. (2014) use an indicator related to complex network theory: kurtosis of the degree distribution. The authors construct a Pearson Correlation Climate Network (Feng and Dijkstra, 2014) for the AMOC. In this set-up, a high degree in the network represents high spatial correlation. An increase in kurtosis is found when increasing the freshwater input to the system. In Chapter 6 we also study tipping points of the FAMOUS model, but through a reduction of it's dynamics to a low-dimensional box model.

Many studies have been presented in this section which show the importance of the AMOC in past and present climate studies. As many elements in the earth's climate system are currently experiencing rapid changes (Lenton et al., 2008), it is of growing interest that the dynamics and implications of an AMOC shutdown are fully understood.

Chapter 3

Methods of deriving models with delay

In this chapter we discuss general methodologies behind deriving models with delay from a variety of dynamical systems. We focus on three particular methods: the linear chain approximation, the Mori-Zwanzig formalism, and integration along wave characteristics. The chapter is organised as follows. In Section 3.1 we discuss the connection between coupling chains of ordinary differential equations (ODEs) and delay differential equations (DDEs). We demonstrate the reduction of two types of ODE systems to a scalar DDE. Section 3.2 discusses general literature on the Mori-Zwanzig projection formalism and how this can be applied to systems of ODEs. The projection introduces a memory term in the form of an integral over the past of the projected variable with a memory kernel, which is in the form of a distributed delay. The particular form of the memory kernel can then in some cases be used as an argument for approximation by a discrete delay. In Section 3.3 we focus on hyperbolic PDEs (e.g. wave equations) in which a delay equation can be derived by integration along wave characteristics. We consider two case studies, the El Niño Southern Oscillation (ENSO) and the Atlantic Multidecal Oscillation (AMO). The ENSO study follows the work of Jin (1997). The AMO study is a novel derivation of a model with delay for the phenomenon where similar methods to those used in the ENSO study are applied.

We note here that in Section 3.2 and Section 3.3 we will be using partial differential equations (PDEs) for some of our analysis. In general, the PDEs will be of the form

$$\frac{\partial T(x,t)}{\partial t} = AT(x,t) + \theta(T(x,t)), \tag{3.1}$$

where A is a linear differential operator. A semigroup is defined as a family of linear operators $T(t)$ on a Banach space X that solve the equation

$$\frac{dT(t)}{dt} = AT(t), \tag{3.2}$$

with infinitesimal generator A (Pazy, 2012). We will notate the semigroup as $\mathrm{e}^{tA} := T(t)$ where $\mathrm{e}^{tA} = \lim_{N\to\infty}(1 - \frac{tA}{N})^{-N}$. Semigroups have the composition property

$$[e^{tA}e^{sA}](\cdot) = e^{(t+s)A}(\cdot), \quad \text{for} \quad t, s \geq 0, \tag{3.3}$$

and

$$[e^{tA}A](\cdot) = [Ae^{tA}](\cdot), \quad \text{for} \quad t \geq 0, \tag{3.4}$$

as well as commutativity with arbitrary functions p:

$$[e^{tA}p](\cdot) = [pe^{tA}](\cdot), \tag{3.5}$$

(Chorin et al., 2000).

3.1 The Linear Chain Approximation

This section describes how the linear chain approximation of DDEs can link DDEs to systems of ODEs with particular forms. The system is in the form of variables in a chain of ODEs:

$$\frac{dy_i}{dt} = a_i(y_{i-1}(t) - y_i(t)), \quad i \in 1, ..., N. \tag{3.6}$$

This is a system of N linear equations that links y_i with the previous variable y_{i-1}. Each equation has a scalar factor and a decay rate of order N.

Literature on the linear chain approximation derivation can be found in Smith (2010). Here we don't focus on this derivation; rather, we start with a system of ODEs and use the fundamental matrix solution and a variation of constants approach to obtain a DDE in our variable of interest. We then show how this equation can be approximated as a DDE with discrete delay. The two cases considered are the degenerate case (all timescales in chain equal) and the non-degenerate case (all timescales in chain distinct). The results presented here can be extended to systems with varying combinations of equal and distinct timescales.

3.1.1 Degenerate case

We first discuss the case where all timescales of the chained variables are equal, which we refer to as the degenerate case. In reference to (3.6), we take $a_i = Na$ for all $i = 1, ..., N$. We consider systems

of the form

$$\frac{dy_0}{dt} = F(y_0(t), y_N(t)), \tag{3.7a}$$

$$\frac{dy_i}{dt} = Na(y_{i-1}(t) - y_i(t)), \quad i \in 1, ..., N, \tag{3.7b}$$

where $a > 0$. Here we show the feedback in (3.7a) as only one component $y_N(t)$, however the theory applies for multiple feedbacks (i.e. different components $y_i(t)$). System (3.7b) is the linear chain part of system (3.7), which can be expressed in matrix-vector form as

$$\frac{d\vec{y}}{dt} = A\vec{y}(t) + \vec{b}(t), \tag{3.8}$$

where A is an $N \times N$ matrix of the form

$$A = \begin{bmatrix} -Na & 0 & \dots & \dots & 0 \\ Na & -Na & \ddots & & \vdots \\ 0 & Na & -Na & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & Na & -Na \end{bmatrix}, \tag{3.9}$$

and $\vec{y}(t), \vec{b}(t)$ are $N \times 1$ vectors of the form

$$\vec{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix}, \qquad \vec{b}(t) = \begin{bmatrix} Nay_0(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.10}$$

Since A is non-diagonalisable, we write the solution matrix, $\Phi(t)$, of $\frac{d\vec{y}}{dt} = A\vec{y}(t)$ ($\vec{b}(t) = \mathbf{0}$) as

$$\Phi(t) = \mathrm{e}^{-Nat}\mathrm{e}^{(A+NaI)t}, \tag{3.11}$$

See Perko (2013) for a detailed derivation of (3.11). Computing the last exponential term in (3.11) gives

$$\Phi(t) = \mathrm{e}^{-Nat}\mathrm{e}^{(A+NaI)t} = \mathrm{e}^{-Nat} \begin{bmatrix} 1 & 0 & \dots & \dots & & 0 \\ Nat & 1 & \ddots & & & \vdots \\ \frac{(Na)^2t^2}{2!} & Nat & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \frac{(Na)^{N-2}t^{N-2}}{(N-2)!} & \frac{(Na)^{N-3}t^{N-3}}{(N-3)!} & \dots & Nat & 1 & 0 \\ \frac{(Na)^{N-1}t^{N-1}}{(N-1)!} & \frac{(Na)^{N-2}t^{N-2}}{(N-2)!} & \dots & \frac{(Na)^2t^2}{2!} & Nat & 1 \end{bmatrix}. \tag{3.12}$$

From Perko (2013) we know the fundamental solution of (3.8) can be expressed as

$$\vec{y}(t) = \Phi(t)\vec{y}(0) + \int_0^t \Phi(t-s)\vec{b}(s)ds. \tag{3.13}$$

In our system this simplifies to

$$\vec{y}(t) = \Phi(t)\vec{y}(0) + \int_0^t \mathrm{e}^{-Na(t-s)} y_0(s)\vec{\Psi}(t-s)ds. \tag{3.14}$$

where

$$\vec{\Psi}(t-s) = \begin{bmatrix} Na \\ (Na)^2(t-s) \\ \frac{(Na)^3(t-s)^2}{2!} \\ \vdots \\ \frac{(Na)^{N-1}(t-s)^{N-2}}{(N-2)!} \\ \frac{(Na)^N(t-s)^{N-1}}{(N-1)!} \end{bmatrix}. \tag{3.15}$$

We are interested in the asymptotic behaviour of (3.7). Hence we assume the system initialised arbitrarily far in the past, which allows us to extend the lower limit of the integral in (3.14) to $-\infty$ and neglect dependence on the initial condition. System (3.14) then becomes

$$\vec{y}(t) = \int_{-\infty}^t \mathrm{e}^{-Na(t-s)} y_0(s)\vec{\Psi}(t-s)ds. \tag{3.16}$$

Looking at the N-th component of (3.16) and using the change of variables $\tau = t - s$,

$$y_N(t) = \int_0^\infty y_0(t-\tau)\frac{(Na)^N\tau^{N-1}}{(N-1)!}\mathrm{e}^{-Na\tau}d\tau. \tag{3.17}$$

The kernel of (3.17) can be extracted as

$$K_N(\tau) = \frac{(Na)^N\tau^{N-1}}{(N-1)!}\mathrm{e}^{-Na\tau}. \tag{3.18}$$

This will be useful to compare with the Mori-Zwanzig projection procedure in Section 3.2. The integral of the kernel (3.18) equals 1, which means that $K_N(\tau)$ is a density of some probability distribution (in this case, the Gamma distribution):

$$\int_0^\infty \frac{(Na)^N\tau^{N-1}}{(N-1)!}\mathrm{e}^{-Na\tau}d\tau = 1. \tag{3.19}$$

As $N \to \infty$, (3.18) approaches a Dirac distribution centered at $\mu = 1/a$ (since (3.18) is in the form of a probability distribution, we can treat τ as a random variable to calculate the mean, $\mathbb{E}[\tau] = N/Na$, and variance, $\mathrm{Var}(\tau) = 1/Na^2$) Smith (2010). This behaviour is visualised in Figure 3.1 for $a = 1$ and increasing values of N. As expected, there is a peak at $\tau = 1$ and the variance approaches 0 as N

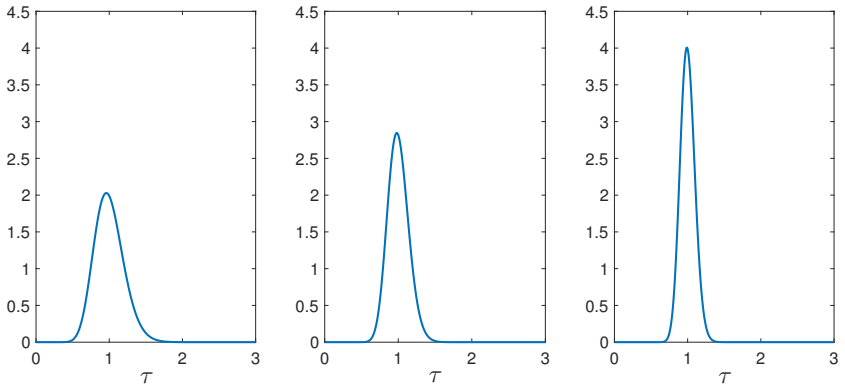


Figure 3.1: Visualisation of the kernel (3.18) for increasing length of chain N and $a = 1$. Left: $N = 25$, $\mathrm{Var}(\tau) = 0.04$. Middle: $N = 50$, $\mathrm{Var}(\tau) = 0.02$. Right: $N = 100$, $\mathrm{Var}(\tau) = 0.01$.

increases. We are interested in bounded, continuous $y_0(t)$ such that histories of $y_0(t)$ are bounded as $t \to -\infty$. Therefore, for all $y_0 \in BC((-\infty, 0]; \mathbb{R}^N)$,

$$\int_0^\infty y_0(t-\tau)K_N(\tau)d\tau \to y_0\Big(t - \frac{1}{a}\Big) \quad \text{as} \quad N \to \infty. \tag{3.20}$$

This allows us to write (3.7a) as

$$\frac{dy_0}{dt} = F\Big(y_0(t), y_0\Big(t - \frac{1}{a}\Big)\Big). \tag{3.21}$$

Equation (3.21) is exact for an infinite chain, and is an approximation for finite N.

3.1.2 Non-degenerate case

The case where the coefficients of system (3.7b) are all mutually distinct we refer to as the non-degenerate case. This can be written as

$$\frac{dy_0}{dt} = F(y_0(t), y_N(t)), \tag{3.22a}$$

$$\frac{dy_i}{dt} = a_i(y_{i-1}(t) - y_i(t)) \quad i \in 1, ..., N. \tag{3.22b}$$

Again extracting system (3.22b) we have

$$\frac{d\vec{y}}{dt} = A\vec{y}(t) + \vec{b}(t), \tag{3.23}$$

where now

$$A = \begin{bmatrix} -a_1 & 0 & \dots & \dots & 0 \\ a_2 & -a_2 & \ddots & & \vdots \\ 0 & a_3 & -a_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_N & -a_N \end{bmatrix}, \qquad \vec{b}(t) = \begin{bmatrix} a_1 y_0(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.24}$$

We will assume $a_i \neq a_j$ for $i \neq j$ such that A is diagonalisable. The the solution matrix, $\Phi(t)$, of $\frac{d\vec{y}}{dt} = A\vec{y}(t)$ ($\vec{b}(t) = \mathbf{0}$) is then given by

$$\Phi(t) = P\mathrm{e}^{\Lambda t}P^{-1}, \tag{3.25}$$

where P is the matrix of eigenvectors associated with the diagonal matrix of eigenvalues Λ,

$$\Lambda = \begin{bmatrix} -a_1 & 0 & \dots & 0 \\ 0 & -a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -a_N \end{bmatrix}, \tag{3.26}$$

The entries of P and P^{-1} have the following forms:

$$P_{i,j} = \begin{cases} 0 & \text{for} \quad i < j \\ \prod_{k=j+1}^{N} \frac{a_i - a_k}{a_k} & \text{for} \quad j \leq i < n \\ 1 & \text{for} \quad i = N \end{cases}, \tag{3.27}$$

$$P^{-1}_{i,j} = \begin{cases} 0 & \text{for} \quad i < j < N \\ \frac{\prod_{k=j+1}^{N} a_k}{\prod_{\substack{k=j \\ k\neq i}}^{N} a_k - a_i} & \text{for} \quad j \leq i \\ 1 & \text{for} \quad i, j = N \end{cases}. \tag{3.28}$$

Again we have the fundamental solution of (3.23) with $\vec{b}(t) \neq \mathbf{0}$,

$$\vec{y}(t) = \Phi(t)\vec{y}(0) + \int_0^t \Phi(t)\Phi^{-1}(s)\vec{b}(s)ds. \tag{3.29}$$

Performing the same manipulations as in the degenerate case, we arise with an equation for y_N,

$$y_N(t) = \int_0^\infty a_1 y_0(t-\tau) \sum_{i=1}^{N} \frac{\prod_{k=2}^{N} a_k}{\prod_{\substack{k=1 \\ k\neq i}}^{N} a_k - a_i} \mathrm{e}^{-a_i\tau} d\tau. \tag{3.30}$$

We can extract the kernel of (3.30) as

$$K_N(\tau) = \sum_{i=1}^{N} a_i \mathrm{e}^{-a_i \tau} \prod_{\substack{k=1 \\ k \neq i}}^{N} \frac{a_k}{a_k - a_i}. \tag{3.31}$$

Note that (3.31) has a limit for $a_i \to a_k$ for any i, k. This kernel has the form of the generalised Erlang distribution. The expected value of this distribution is given as

$$\mathbb{E}[\tau] = \sum_{i=1}^{N} \frac{1}{a_i}. \tag{3.32}$$

We can then write (3.22a) as a DDE with discrete delay

$$\frac{dy_0}{dt} = F\Big(y_0(t), y_0\Big(t - \sum_{i=1}^{N} \frac{1}{a_i}\Big)\Big). \tag{3.33}$$

As in the degenerate case, (3.33) is exact for an infinite chain, and is an approximation for finite N.

3.2 The Mori-Zwanzig Projection framework

The Mori-Zwanzig procedure was developed by Mori (1965) and Zwanzig (1973) as a way to project Hamiltonian systems into a reduced phase space while still accurately capturing the effect of the unresolved variables. The novelty of the procedure is that the dynamic equations of the resolved space are only dependent upon the resolved variables. This is useful when one does not have information about the time evolution of the unresolved variables. In this section we use the Mori-Zwanzig formalism to obtain delay differential equations with distributed delay. Our application starts from non-Hamiltonian dissipative systems resulting in some simplifications.

3.2.1 Derivation of the Mori-Zwanzig equation

We demonstrate the procedure for ODEs, but the method can also be used for PDEs of the form (3.1) when discretised appropriately. Suppose we have a system

$$\frac{d\vec{x}}{dt} = f(\vec{x}(t), \vec{y}(t)), \tag{3.34a}$$

$$\frac{d\vec{y}}{dt} = g(\vec{x}(t), \vec{y}(t)). \tag{3.34b}$$

Let the full dimensionality of the system be $N = m + n$. We assume $f \in C^1(\mathbb{R}^N, \mathbb{R}^m)$ and $g \in C^1(\mathbb{R}^N, \mathbb{R}^n)$ where $C^1(U, V)$ is the space of continuously differentiable functions from space U to

V. We want to project this system onto functions of only $\vec{x}(t) \in \mathbb{R}^m$, which we will now refer to as the resolved variables ($y(t) \in \mathbb{R}^n$ being the unresolved variables). We define a projection $\mathcal{P} : C^1(\mathbb{R}^N, \mathbb{R}^k) \to C^1(\mathbb{R}^N, \mathbb{R}^k)$ for initially arbitrary k as follows:

$$[\mathcal{P}f](\vec{x}, \vec{y}) = f(\vec{x}, 0). \tag{3.35}$$

Here k is the dimension of the output of the function to which the projection is applied. This projection is known as the finite-rank projection (Chorin et al., 2002). Solutions of system (3.34) define a flow map $\phi^t : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, where $\phi^t(\vec{x}, \vec{y})$ maps initial conditions $\vec{x}, \vec{y}$ to their evolution via (3.34) at time t (Givon et al., 2005). The flow map then satisfies the ODE system (3.34), which we write compactly as

$$\frac{d\phi^t(\vec{u})}{dt} = R(\phi^t(\vec{u})), \qquad \phi^0(\vec{u}) = \vec{u}. \tag{3.36}$$

where

$$R(\phi^t(\vec{u})) = \begin{bmatrix} f(\phi_x^t(\vec{u}), \phi_y^t(\vec{u})) \\ g(\phi_x^t(\vec{u}), \phi_y^t(\vec{u})) \end{bmatrix}. \tag{3.37}$$

For simplicity we have combined variables $\vec{x}$ and $\vec{y}$ into one vector, $\vec{u} = [\vec{x}, \vec{y}]$. In the above notation, $\phi_x^t(\vec{u})$ and $\phi_y^t(\vec{u})$ have the same dimensions as $\vec{x}$ and $\vec{y}$ respectively and are the time evolutions of those variables. Applying projection (3.35) to (3.37) then leads to the following:

$$[\mathcal{P}R](\phi^t(\vec{u})) = \begin{bmatrix} f(\phi_x^t(\vec{u}), 0) \\ g(\phi_x^t(\vec{u}), 0) \end{bmatrix} = R([\phi_x^t(\vec{u}), 0]). \tag{3.38}$$

In shorthand notation we will write $[\mathcal{P}R](\phi^t(\vec{u}))$ as simply $R(\phi_x^t(\vec{u}))$.

Given an arbitrary function $h : \mathbb{R}^N \to \mathbb{R}^k$, we define the phase distribution function $\rho : \mathbb{R} \times \mathbb{R}^N \ni (t, \vec{u}) \mapsto h(\phi^t(\vec{u})) \in \mathbb{R}^k$. Then ρ satisfies

$$\frac{\partial}{\partial t}\rho(\vec{u}, t) = \mathcal{L}\rho(\vec{u}, t), \qquad \rho(\vec{u}, 0) = h(\vec{u}). \tag{3.39}$$

Equation (3.39) is the linear PDE known as the Liouville equation (Morriss and Evans, 2013), and $\mathcal{L}$ is the associated Liouville operator which governs the time evolution of the phase distribution ρ:

$$\mathcal{L} := \sum_{j=1}^{N} R_j(\vec{u}) \frac{\partial}{\partial u_j}. \tag{3.40}$$

The PDE (3.39) follows from an application of the chain rule to the ODE system (3.34) governing $\phi^t(\vec{u})$. To illustrate this, consider the following property of ρ:

$$\rho(\vec{x}(0), \vec{y}(0), t) = \rho(\phi_x^t([\vec{x}, \vec{y}]), \phi_y^t([\vec{x}, \vec{y}]), 0). \tag{3.41}$$

This then implies

$$\begin{aligned}\frac{\partial}{\partial t}\rho(\vec{x},\vec{y},t) =& \frac{\partial}{\partial \vec{x}}\rho(\vec{x},\vec{y},t)\frac{d\vec{x}}{dt} + \frac{\partial}{\partial \vec{y}}\rho(\vec{x},\vec{y},t)\frac{d\vec{y}}{dt} \\ =& \sum_{j=1}^{m} f_j(\vec{x},\vec{y})\frac{\partial}{\partial x_j}\rho(\vec{x},\vec{y},t) + \sum_{j=1}^{n} g_j(\vec{x},\vec{y})\frac{\partial}{\partial y_j}\rho(\vec{x},\vec{y},t) \\ =& \mathcal{L}\rho(\vec{u},t). \end{aligned} \tag{3.42}$$

We make the same assumption as in Morriss and Evans (2013) in that there is no explicit time dependence in the original system (3.34). We can then denote the solution to (3.39) as

$$\rho(\vec{u},t) = [\mathrm{e}^{t\mathcal{L}}h](\vec{u}) = h(\phi^t(\vec{u})), \tag{3.43}$$

where $\mathrm{e}^{t\mathcal{L}}$ is the semigroup generated by $\mathcal{L}$. We use the composition property (3.4) to obtain

$$[\mathcal{L}\mathrm{e}^{t\mathcal{L}}h](\vec{u}) = [\mathrm{e}^{t\mathcal{L}}\mathcal{L}h](\vec{u}), \qquad \forall h \in C^1(\mathbb{R}^N,\mathbb{R}). \tag{3.44}$$

Equations (3.39), (3.43), and (3.44) then imply

$$\frac{\partial}{\partial t}\mathrm{e}^{t\mathcal{L}}h(\vec{u}) = [\mathcal{L}\mathrm{e}^{t\mathcal{L}}h](\vec{u}) = [\mathrm{e}^{t\mathcal{L}}\mathcal{L}h](\vec{u}). \tag{3.45}$$

If we consider the decomposition of the identity operator $\mathcal{I} = \mathcal{P} + \mathcal{Q}$ such that $\mathcal{Q} : C^1(\mathbb{R}^N,\mathbb{R}^k) \to C^1(\mathbb{R}^N,\mathbb{R}^k)$, the above equation can be decomposed as

$$\frac{\partial}{\partial t}\mathrm{e}^{t\mathcal{L}}h(\vec{u}) = [\mathrm{e}^{t\mathcal{L}}\mathcal{P}\mathcal{L}h](\vec{u}) + [\mathrm{e}^{t\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}). \tag{3.46}$$

The Dyson decomposition of propagators (Morriss and Evans, 2013) states

$$\mathrm{e}^{(A+B)t} = \mathrm{e}^{At} + \int_0^t \mathrm{e}^{As}B\mathrm{e}^{(A+B)(t-s)}ds. \tag{3.47}$$

Applying this to the last term of the right-hand side of (3.46) with $A = \mathcal{Q}\mathcal{L}$ and $B = \mathcal{P}\mathcal{L}$ gives

$$[\mathrm{e}^{t\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}) = [\mathrm{e}^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}) + \int_0^t [\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}\mathrm{e}^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u})ds, \tag{3.48}$$

where $\mathrm{e}^{s\mathcal{Q}\mathcal{L}}$ is the semigroup operator generated by the orthogonal dynamics equation (Givon et al., 2005):

$$\frac{\partial}{\partial t}\rho_Q(\vec{u},t) = \mathcal{Q}\mathcal{L}\rho_Q(\vec{u}), \qquad \rho_Q(\vec{u},0) = h(\vec{u}). \tag{3.49}$$

The full PDE (3.46) is then given as

$$\frac{\partial}{\partial t}[\mathrm{e}^{t\mathcal{L}}h](\vec{u}) = [\mathrm{e}^{t\mathcal{L}}\mathcal{P}\mathcal{L}h](\vec{u}) + \int_0^t [\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}\mathrm{e}^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}]h(\vec{u})ds + [\mathrm{e}^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}). \tag{3.50}$$

We will consider the case where $h(\vec{u}) = u_i$, i.e. the function h returns the i-th component of $\vec{u}$ onto which one wishes to project. Using this definition of h and applying the semigroup commutativity property (3.5) to the term $[e^{t\mathcal{L}}\mathcal{P}\mathcal{L}h](\vec{u})$, equation (3.50) simplifies to

$$\frac{\partial}{\partial t}\phi_i^t(\vec{u}) = R_i(\phi_x^t(\vec{u})) + \int_0^t [e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}e^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u})ds + [e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}). \tag{3.51}$$

Equation (3.51) is known as the Mori-Zwanzig equation for variable ϕ_i^t. Note that (3.51) is dependent on the resolved variables $(\phi_x^t(\vec{u}))$, but only the intial values $\phi^0(\vec{u}) = \vec{u}$ of the unresolved variables are needed.

In the next section we discuss an approximation to the orthogonal dynamics and the relation to delay equations.

3.2.2 The orthogonal dynamics

The Mori-Zwanzig equation (3.51) has three distinct terms: the Markovian term, the memory term, and the noise term. The first term is denoted the Markovian term, as it depends only on the system in its present state. The integral term, known as the memory term, is dependent upon the past state of the system. The final term is called the noise term since it only depends on the initial values of the unresolved variables, and so the influence of these values on the system is commonly represented as noise. Both the integral term and the final term rely on solving the orthogonal dynamics. The projection should be chosen such that the orthogonal dynamics are stable. With this choice, we can neglect the noise term in (3.51) by extending the upper bound of the integral term to infinity:

$$\frac{\partial}{\partial t}\phi_i^t(\vec{u}) = R_i(\phi_x^t(\vec{u})) + \int_0^\infty [e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}e^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u})ds. \tag{3.52}$$

Equation (3.52) still contains dependence on the past state of the system in the memory term. For the exact form of the integrand, one needs to solve the orthogonal dynamics equation (3.49). In general, (3.49) is not in the form of a Liouville equation of an ODE system. In order to obtain the solution, one would need to solve a ODE in N-dimensional space which is impractical. The following section presents an ODE system that has (3.49) as its Liouville equation for linear functions R. This system can serve as an approximation for nonlinear R in some cases.

3.2.2.1 Approximation of the orthogonal dynamics

The orthogonal dynamics equation (3.49) can be analytically solved for linear systems, however it is unknown for nonlinear systems. In Gouasmi et al. (2017) the authors attempt to approximate the orthogonal dynamics. We summarise their approximation.

We first consider just the integrand in (3.52), $[\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}\mathrm{e}^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u})$. For $h(\vec{u}) = u_i$, we know from the definition of the Liouville operator (3.40) that

$$[\mathcal{Q}\mathcal{L}h](\vec{u}) = \left[\mathcal{Q}\sum_{j=1}^{N} R_j(\cdot)\frac{\partial h}{\partial u_j}\right](\vec{u}) = [\mathcal{Q}R_i](\vec{u}). \tag{3.53}$$

We can then use the definition $\mathcal{Q} = \mathcal{I} - \mathcal{P}$ to obtain

$$[\mathcal{Q}R_i](\vec{u}) = R_i(\vec{u}) - [\mathcal{P}R_i](\vec{u}) = R_i(\vec{u}) - R_i(\vec{u}_x). \tag{3.54}$$

For simplicity we will define $R_Q(\cdot) := R(\cdot) - [\mathcal{P}R](\cdot)$. We here assume that (3.49) is the Liouville equation for a corresponding system of ODEs which describe the flow map of the orthogonal dynamics, $\phi_Q^t(\vec{u}) : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$. The requirement for this would be

$$\frac{\partial}{\partial u_j}[\mathcal{Q}\mathcal{L}h](\vec{u}) = \frac{\partial}{\partial u_j}[\mathcal{Q}\mathcal{L}h](\vec{u}_x) \qquad \forall j \in [1, ..., N] \tag{3.55}$$

For our choice of h, we know from (3.53) and (3.54) that (3.55) is equivalent to

$$\left[\frac{\partial}{\partial u_j}R_{Qi}\right](\vec{u}) = \left[\frac{\partial}{\partial u_j}R_{Qi}\right](\vec{u}_x) \qquad \forall j \in [1, ..., N] \tag{3.56}$$

We then denote the solution $\rho_Q(\vec{u}, t) = [\mathrm{e}^{t\mathcal{Q}\mathcal{L}}h](\vec{u})$ as the solution to (3.49). Using this solution form we can rewrite (3.49) as

$$\frac{\partial}{\partial t}[\mathrm{e}^{t\mathcal{Q}\mathcal{L}}h](\vec{u}) = [\mathcal{Q}\mathcal{L}\mathrm{e}^{t\mathcal{Q}\mathcal{L}}h](\vec{u}). \tag{3.57}$$

If condition (3.56) is met, $\mathrm{e}^{t\mathcal{Q}\mathcal{L}}$ is granted the commutativity property with h:

$$[\mathcal{Q}\mathcal{L}\mathrm{e}^{t\mathcal{Q}\mathcal{L}}h](\vec{u}) = .[\mathcal{Q}\mathcal{L}h]\phi_Q^t(\vec{u}). \tag{3.58}$$

Combining (3.58) with (3.57) then gives

$$\frac{\partial}{\partial t}h(\phi_Q^t(\vec{u})) = [\mathcal{Q}\mathcal{L}h]\phi_Q^t(\vec{u}). \tag{3.59}$$

Using $h(\vec{u}) = u_i$ and applying equivalences (3.53) and (3.54) we obtain an ODE for each component of the flow map $\phi_{Qi}^t(\vec{u})$:

$$\frac{d}{dt}\phi_{Qi}^t(\vec{u}) = R_i(\phi_Q^t(\vec{u})) - R_i(\phi_{Qx}^t(\vec{u})), \qquad \phi_{Qi}^0(\vec{u}) = u_i \tag{3.60}$$

We can then construct the ODE system corresponding to (3.49) as

$$\frac{d\phi_Q^t(\vec{u})}{dt} = R_Q(\phi_Q^t(\vec{u})), \qquad \phi_Q^0(\vec{u}) = \vec{u}. \tag{3.61}$$

This then allows us to write for $h(\vec{u}) = u_i$,

$$[\mathrm{e}^{s\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}h](\vec{u}) \approx R_{Qi}(\phi_Q^s(\vec{u})), \tag{3.62}$$

where $\phi_Q^s(\vec{u})$ is the solution of (3.61) at time $t = s$. The Mori-Zwanzig equation (3.52) can now be approximated by

$$\frac{\partial}{\partial t}\phi_i^t(\vec{u}) \approx R_i(\phi_x^t(\vec{u})) + \int_0^\infty [\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}R_{Qi}](\phi_Q^s(\vec{u}))ds, \tag{3.63}$$

with error $\delta \leq \left| \left[\frac{\partial}{\partial \vec{u}} R_{Qi}\right](\vec{u}) - \left[\frac{\partial}{\partial \vec{u}} R_{Qi}\right](\vec{u}_x) \right|$.

3.2.2.2 Extracting the memory kernel

If we assume the approximation in Section 3.2.2 to be valid, we now have a practical way to compute the approximate memory term of the system (3.52). We consider here just the memory term of (3.52),

$$M_i(\vec{u}, t) = \int_0^\infty [\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}R_{Qi}](\phi_Q^s(\vec{u}))ds. \tag{3.64}$$

By definition we know

$$[\mathcal{L}R_{Qi}](\phi_Q^s(\vec{u})) = \left[\sum_{j=1}^N R_j(\vec{u})\frac{\partial}{\partial u_j}R_{Qi}\right](\phi_Q^s(\vec{u})). \tag{3.65}$$

Application of the Liouville operator to a function is equivalent to taking the directional derivative of the function in the direction $R(\vec{u})$,

$$\nabla_{R(\vec{u})} = \sum_{j=1}^N R_j(\vec{u})\frac{\partial}{\partial u_j}. \tag{3.66}$$

We then have

$$\mathcal{L}R_{Qi}(\phi_Q^s(\vec{u})) = \nabla_{R(\vec{u})}R_{Qi}(\phi_Q^s(\vec{u})). \tag{3.67}$$

The RHS can be written as a limit,

$$\nabla_{R(\vec{u})}R_{Qi}(\phi_Q^s(\vec{u})) = \lim_{\epsilon\to 0}\frac{R_{Qi}(\phi_Q^s(\vec{u}+\epsilon R(\vec{u}))) - R_{Qi}(\phi_Q^s(\vec{u}))}{\epsilon}. \tag{3.68}$$

Combining (3.67) and (3.68) we obtain

$$\mathcal{L}R_{Qi}(\phi_Q^s(\vec{u})) \approx \frac{1}{\epsilon}\Big(R_{Qi}\big(\phi_Q^s(\vec{u}+\epsilon R(\vec{u}))\big) - R_{Qi}\big(\phi_Q^s(\vec{u})\big)\Big). \tag{3.69}$$

Applying the semigroup operator $\mathrm{e}^{(t-s)\mathcal{L}}$ (which commutes due to the Liouville property), we have

$$[\mathrm{e}^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}R_{Qi}](\phi_Q^s(\vec{u})) \approx \frac{1}{\epsilon}\mathcal{P}\Big(R_{Qi}\big(\phi_Q^s(\phi^{t-s}(\vec{u}) + \epsilon R(\phi^{t-s}(\vec{u})))\big) - R_{Qi}\big(\phi_Q^s(\phi^{t-s}(\vec{u}))\big)\Big). \tag{3.70}$$

Applying $\mathcal{P}$ the above gives

$$e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}R_{Qi}(\phi_Q^s(\vec{u})) \approx \frac{1}{\epsilon}R_{Qi}\Big(\phi_Q^s\big(\phi_x^{t-s}(\vec{u}) + \epsilon R(\phi_x^{t-s}(\vec{u}))\big)\Big). \tag{3.71}$$

The last term of (3.69) drops because ϕ_Q^s is the solution to the orthogonal dynamics, so we know by definition of the orthogonal dynamics $\phi_Q^s(\phi_x^{t-s}(\vec{u})) = 0$. Substituting this form of the memory term into (3.63), we now have an ODE formulation of the Mori-Zwanzig equation,

$$\begin{aligned} &\frac{d}{dt}\phi_i^t(\vec{u}) \approx R_i(\phi_x^t(\vec{u})) + \int_0^\infty \frac{1}{\epsilon}R_{Qi}\Big(\phi_Q^s\big(\phi_x^{t-s}(\vec{u}) + \epsilon R(\phi_x^{t-s}(\vec{u}))\big)\Big)ds, \\ &\phi_i^0(\vec{u}) = u_i. \end{aligned} \tag{3.72}$$

Equation (3.72) gives individual differential equations with distributed delay for each projected variable.

3.2.3 Examples of computed approximate memory

For nonlinear systems it is not possible to know the exact equations for the orthogonal dynamics, however the existence of solutions to the orthogonal dynamics can be proven for certain classes of projections. Givon et al. (2005) proved that for the finite-rank projection (3.35) there always exists solutions to the orthogonal dynamics (3.49) on a bounded time interval. In this section we will analyse the memory term of two systems: one linear and one nonlinear. We will only consider the finite-rank projection, and we will use the approximation for the orthogonal dynamics derived in Section 3.2.2.

3.2.3.1 Linear chain of ODEs - linear memory term

We will first investigate a linear chain of ODEs which has a linear memory term. This system has been shown to approximate a scalar delay differential equation when the length of chain (N) increases to infinity (see Section 3.1). Consider the system

$$\frac{dy_1}{dt} = b_1 y_1(t) - b_2 y_3(t), \tag{3.73a}$$
$$\frac{dy_2}{dt} = a(y_1(t) - y_2(t)), \tag{3.73b}$$
$$\frac{dy_3}{dt} = a(y_2(t) - y_3(t)). \tag{3.73c}$$

Here the number of variables in the chain is two: y_2 and y_3. We define the projection $\mathcal{P}$ as

$$(\mathcal{P}f)(\vec{y}) = f(\hat{y}), \tag{3.74}$$

where $\vec{y} = [y_1, y_2, y_3]^T$ and $\hat{y} = [y_1, 0, 0]^T$.

We will consider the direct Mori-Zwanzig projection of system (3.73). We define the flow map of (3.73) as simply $\phi^t := \vec{y}(t)$ and R is the righthand side of (3.73). We then use equation (3.72) to define the projection equation:

$$\begin{aligned} &\frac{dy_1}{dt} \approx R_1(\hat{y}(t)) + \frac{1}{\epsilon}\int_0^\infty R_{Q1}\Big(\phi_Q^s\big(\hat{y}(t-s) + \epsilon R(\hat{y}(t-s))\big)\Big)ds, \\ &y_1(0) = y_{10}. \end{aligned} \tag{3.75}$$

The first term of (3.75) simply becomes $b_1 y_1(t)$. For the second term (the memory term) we will need the orthogonal dynamics system (3.61) with the right-hand side $R_Q(\vec{y}^Q) = R(\vec{y}^Q) - R(\vec{y}_x^Q)$. For this system (3.61) has the form

$$\frac{dy_1^Q}{dt} = -b_2 y_3^Q(t), \tag{3.76a}$$

$$\frac{dy_2^Q}{dt} = -a y_2^Q(t), \tag{3.76b}$$

$$\frac{dy_3^Q}{dt} = a(y_2^Q(t) - y_3^Q(t)). \tag{3.76c}$$

For solving (3.76) the initial conditions can be taken as $\hat{y}(s) + \epsilon R(\hat{y}(s))$ where values of the original solution $y_1(s)$ are known. Here, R_Q is the RHS of (3.76), so $R_{Q1}\Big(\phi_Q^s\big(\hat{y}(t-s) + \epsilon R(\hat{y}(t-s))\big)\Big) = -b_2 y_3^Q(s)$. The trajectory $y^Q(s)$ can be obtained by solving (3.76) up to time s with initial condition $\vec{y}^Q(0) = \hat{y}(t-s) + \epsilon R\big(\hat{y}(t-s)\big)$.

Taking $a = 2$, $b_1 = 0.5$, and $b_2 = 1$, we can compare the numerical approximation of the memory term with the exact memory term. To do this we solve the full system (3.73) with initial condition $\vec{y}_0 = [1, 0, 0]^T$ up to time $t = 15$. We then create a modified trajectory $\vec{y}_0^Q(s) = [y_1(s), 0, 0]$ for $s \in [0, 15]$, where $y_1(s)$ is the solution of (3.73) at time s, and use this as the initial condition for solving the approximate orthogonal dynamics (3.76) for every $s \in [0, 15]$. The memory term can then be approximated with

$$\tilde{M}(t) = -\int_0^t \frac{1}{\epsilon} b_2 y_3^Q(s) ds. \tag{3.77}$$

The comparison between the exact memory term, $M(t) = -b_2 y_3(t)$, and the numerically computed memory term is shown in Figure 3.2. It can be clearly seen that for this linear system the memory term is exact.

We also can compare the memory kernels for this system. A memory kernel $K(s)$ is defined for a given memory term as $M(t) = \int_0^t K(s) u_i(t-s) ds$ where $u_i(t)$ is the resolved variable. The kernel of (3.73) is known explicitly from Section 3.1.1 to be

$$K(t) = a^2 t e^{-at}. \tag{3.78}$$

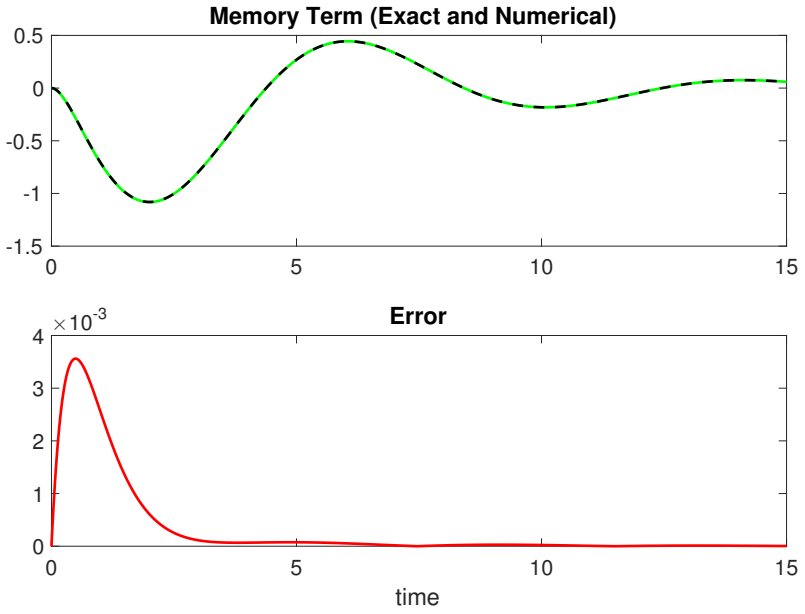


Figure 3.2: Comparison of memory term for linear chain system (3.73). Top: Exact memory $M(t) = -b_2 y_3(t)$ (dotted black) and numerical memory $\tilde{M}(t) = -\int_0^t \frac{1}{\epsilon} b_2 y_3^Q(s) ds$ (green), where $\epsilon = 1$. Bottom: $|M(t) - \tilde{M}(t)|$

For the numerical kernel we take

$$\tilde{K}(t) = y_3^Q(t). \tag{3.79}$$

Figure 3.3 shows the results for $t = 15$. It can also be clearly seen that the memory kernels are identical. In this case, there is a peak that will get sharper if we increase the length of the chain. When the number of variables in the chain approaches infinity, the kernels in Figure 3.3 converge to a delta function centered at the mean (indicated by the grey bar).

3.2.3.2 Linear chain of ODEs - nonlinear memory term

Now we will consider a linear chain of ODEs with a nonlinear memory term,

$$\frac{dy_1}{dt} = b_1 y_1(t)^2 - b_2 y_2(t) y_3(t), \tag{3.80a}$$
$$\frac{dy_2}{dt} = a(y_1(t) - y_2(t)), \tag{3.80b}$$
$$\frac{dy_3}{dt} = a(y_2(t) - y_3(t)). \tag{3.80c}$$

We define the same projection $\mathcal{P}$ as in the previous example,

$$(\mathcal{P} f)(\vec{y}) = f(\hat{y}), \tag{3.81}$$

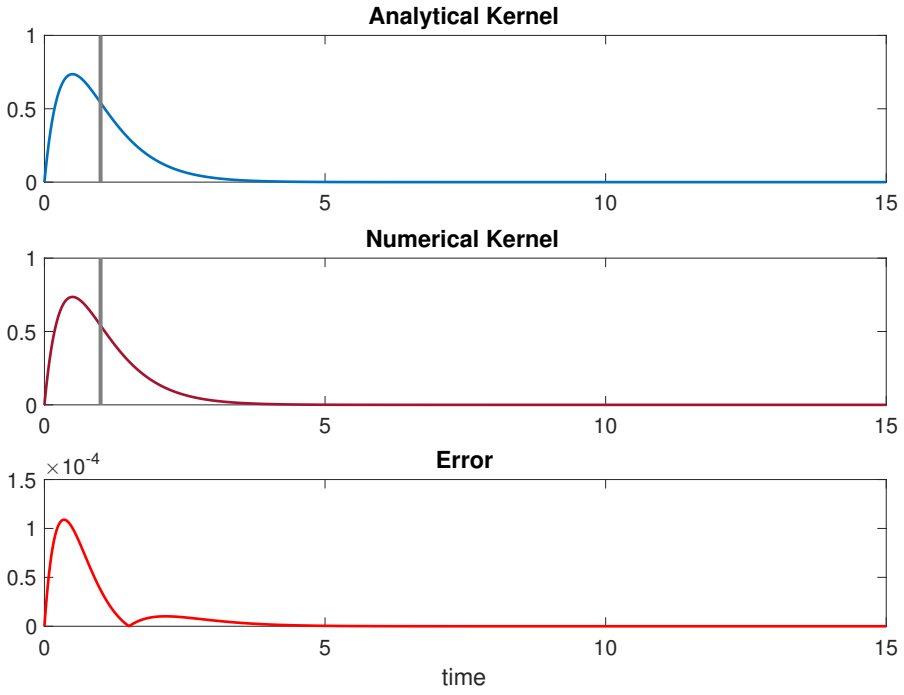


Figure 3.3: Comparison of memory kernels for linear chain system (3.73). Grey bar indicates location of mean. Top: $K(t) = a^2 t e^{-at}$. Middle: $\tilde{K}(t) = y_3^Q(t)$. Bottom: $|K(t) - \tilde{K}(t)|$.

where $\vec{y} = [y_1, y_2, y_3]^T$ and $\hat{y} = [y_1, 0, 0]^T$. The Mori-Zwanzig projection of the system is given again as in (3.75),

$$\begin{aligned} &\frac{dy_1}{dt} \approx b_1 y_1(t)^2 + \frac{1}{\epsilon}\int_0^\infty R_{Q1}\Big(\phi_Q^s\big(\hat{y}(t-s) + \epsilon R(\hat{y}(t-s))\big)\Big)ds, \\ &y_1(0) = y_{10}. \end{aligned} \tag{3.82}$$

The orthogonal dynamics system R_Q is the right-hand side of the following,

$$\frac{dy_1^Q}{dt} = -b_2 y_2^Q(t) y_3^Q(t), \tag{3.83a}$$

$$\frac{dy_2^Q}{dt} = -a y_2^Q(t), \tag{3.83b}$$

$$\frac{dy_3^Q}{dt} = a(y_2^Q(t) - y_3^Q(t)). \tag{3.83c}$$

The initial conditions of (3.83) are again $\hat{y}(s) + \epsilon R(\hat{y}(s))$ where values of the original solution $y_1(s)$ are given as history. We have $R_{Q1}\Big(\phi_Q^s\big(\hat{y}(t-s) + \epsilon R(\hat{y}(t-s))\big)\Big) = -b_2 y_2^Q(s) y_3^Q(s)$. We solve (3.83) up to time s with initial condition $\vec{y}^Q(0) = \hat{y}(t-s) + \epsilon R\big(\hat{y}(t-s)\big)$ to obtain the trajectory $y^Q(s)$.

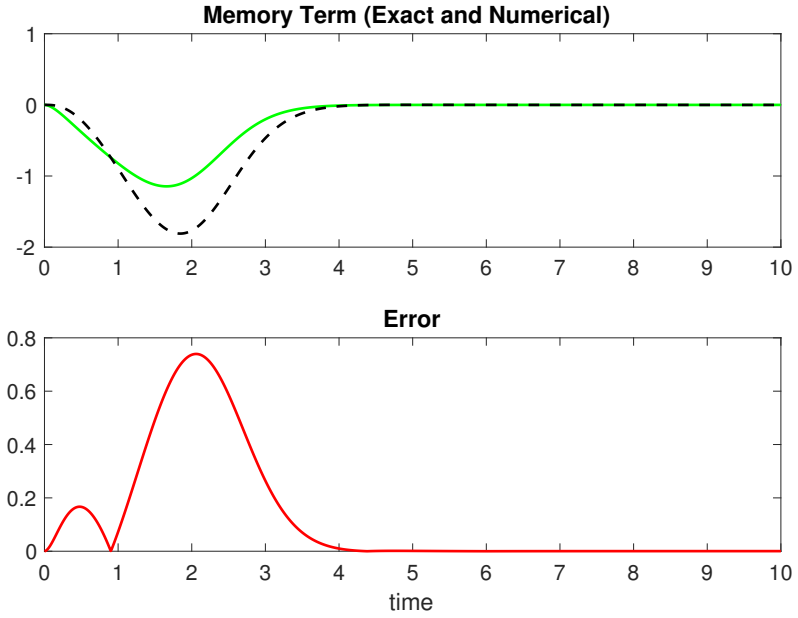


Figure 3.4: Comparison of memory for linear chain system with nonlinear memory (3.80). Top: Exact memory $M(t) = -b_2 y_2(t) y_3(t)$ (dotted black) and numerical memory $\tilde{M}(t) = -\int_0^t \frac{1}{\epsilon} b_2 y_2^Q(s) y_3^Q(s) ds$ (green), where $\epsilon = 1$. Bottom: $|M(t) - \tilde{M}(t)|$

We use the same method for computing the memory term as in the previous section. The memory term is again approximated with

$$\tilde{M}(t) = -\int_0^t \frac{1}{\epsilon} b_2 y_2^Q(s) y_3^Q(s) ds, \tag{3.84}$$

The numerical kernel can be computed with the orthogonal dynamics trajectory,

$$\tilde{K}(t) = y_2^Q(t) y_3^Q(t). \tag{3.85}$$

The exact memory term and kernel for this system are

$$M(t) = -b_2 y_2(t) y_3(t), \qquad K(t) = a^3 t \mathrm{e}^{-2at}. \tag{3.86}$$

We take the same parameter values as in the previous example: $a = 2$, $b_1 = 0.5$, and $b_2 = 1$. The resulting memory term and kernels are compared in Figure 3.4 and Figure 3.5. In this system the kernel is exact (which is expected as the subsystem for the unresolved variables y_2, y_3 is linear). However, since the memory term of the system is nonlinear, there is some error in using (3.61) to compute it. This example suggests that the orthogonal dynamics approximation is an appropriate method for obtaining the memory kernel of a system where the subsystem containing the unresolved variables is linear. Determining the shape of the kernel is a first step in understanding how a system can be reduced to a simplified delay equation.

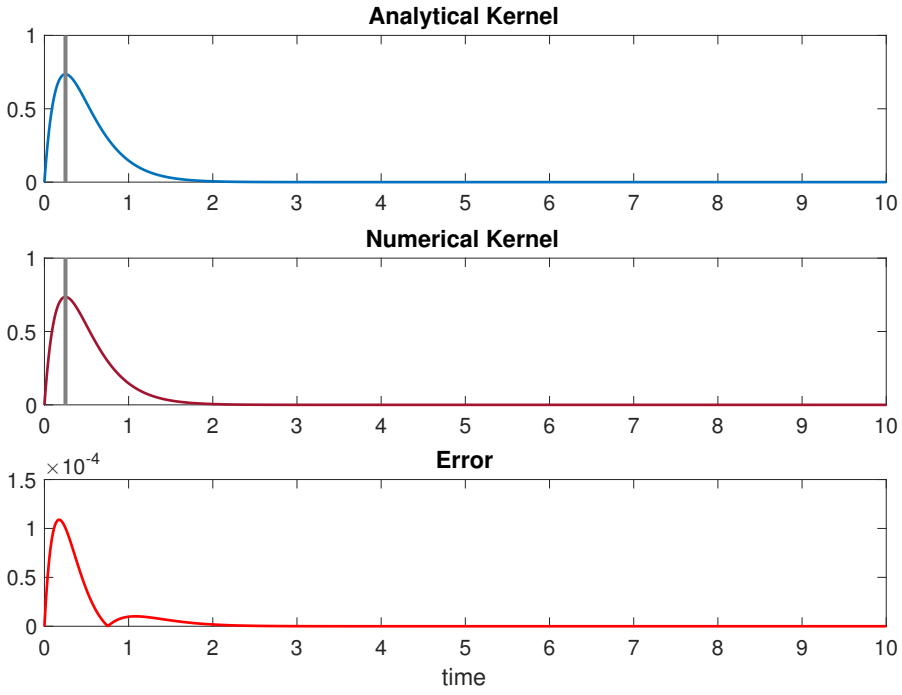


Figure 3.5: Comparison of memory kernels for linear chain system with nonlinear memory (3.80). Grey bar indicates location of mean. Top: $K(t) = a^3 t e^{-2at}$. Middle: $\tilde{K}(t) = y_2^Q(t) y_3^Q(t)$. Bottom: $|K(t) - \tilde{K}(t)|$.

3.3 Derivation of models with delay for wave equations

Waves are known to transport quantities such as heat, energy, material, and more. When waves return to the source of the disturbance, they can cause a feedback effect (Dijkstra, 2008). This can be represented mathematically by a delay. In this section we demonstrate how such a delay is derived from hyperbolic PDE systems. We focus on two phenomena: the El Niño Southern Oscillation and the Atlantic Multidecal Oscillation. We show how the couplings in the different systems lead to a delay differential equation and delay difference equations respectively.

3.3.1 El Niño Southern Oscillation (ENSO)

The El Niño Southern Oscillation (ENSO) is a phenomenon that occurs in the equatorial Pacific Ocean. Its primary effects are interannual sea surface temperature (SST) variations between warm anomalies (El Niño) and cold anomalies (La Niña) in the central and eastern Pacific. Secondary effects include weakening or strengthening of the trade winds, known as the Southern Oscillation (Dijkstra, 2005). Together the effects are known as ENSO and the whole phenomenon has a periodicity on interannual timescales, typically of 4-7 years (Dijkstra, 2008).

Value	Unit	Value	Unit	Value	Unit
$L = 1.7240\text{x}10^{7}$	m	$\beta_0 = 2.28\text{x}10^{-11}$	$\text{m}^{-1}\text{s}^{-1}$	$c_0 = \sqrt{g'H} \approx 2.9$	ms^{-1}
$H = 1.50\text{x}10^{2}$	m	$g' = 5.61\text{x}10^{-2}$	ms^{-2}	$\lambda_0 = \sqrt{\frac{c_0}{\beta_0}} \approx 3.5664\text{x}10^{5}$	m

[Scaling for ENSO PDE model]

Table 3.1: Typical scaling values for parameters in (3.88) taken from (Zebiak and Cane, 1987).

In this section we revisit the works of Dijkstra (2008) and Jin (1997) to derive a model with delay for ENSO. We start with the Navier-Stokes equations (equations of motion of a fluid) and use appropriate approximations to reduce the system to one equation for SST anomalies. The resulting model can be compared against the conceptual models proposed by Suarez and Schopf (1988) and Battisti and Hirst (1989).

3.3.1.1 The dynamic equations of ENSO

When considering the spatial scales of ENSO, the horizontal scales ($\mathcal{O}(10^7)$ m) dominate the vertical scale ($\mathcal{O}(10^2)$ m). In this regime we may start from the shallow water beta-plane approximation to the Navier-Stokes equations on a rotating body (Dijkstra, 2008):

$$\frac{Du_*}{dt_*} - \beta_0 y_* v_* + g' \frac{\partial h_*}{\partial x_*} = 0, \tag{3.87a}$$

$$\frac{Dv_*}{dt_*} + \beta_0 y_* u_* + g' \frac{\partial h_*}{\partial y_*} = 0, \tag{3.87b}$$

$$\frac{\partial h_*}{\partial t_*} + \frac{\partial (u_* h_*)}{\partial x_*} + \frac{\partial (v_* h_*)}{\partial y_*} = 0. \tag{3.87c}$$

Equations (3.87a) and (3.87b) are conservation of momentum and equation (3.87c) is conservation of mass. In the above equations, $\frac{D}{dt} = \frac{\partial}{\partial t} + u_* \frac{\partial}{\partial x} + v_* \frac{\partial}{\partial y}$ is the material derivative; u_* and v_* are velocities in the x_* and y_* directions; β_0 is the variation of the Coriolis force with latitude y_*; h_* is the thickness of the shallow layer (in the ocean this corresponds to the thermocline); g' is the reduced acceleration due to gravity in the shallow water layer at the equator. We will use the following scaling,

$$t_* = \frac{L}{c_0} t \quad ; \quad h_* = Hh \tag{3.88a}$$

$$x_* = Lx \quad ; \quad u_* = c_0 u \tag{3.88b}$$

$$y_* = \lambda_0 y \quad ; \quad v_* = \frac{\lambda_0}{L} c_0 v \tag{3.88c}$$

The typical scaling values for these parameters in an equatorial basin are taken from (Zebiak and Cane, 1987) and listed in Table 3.1

Linearising (3.87) around a motionless reference state $(\bar{u}_*, \bar{v}_*, \bar{h}_*) = (0, 0, H)$ and using the above scaling to nondimensionalise gives

$$\frac{\partial u}{\partial t} - yv + \frac{\partial h}{\partial x} = 0, \tag{3.89a}$$

$$\zeta^2 \frac{\partial v}{\partial t} + yu + \frac{\partial h}{\partial y} = 0, \tag{3.89b}$$

$$\frac{\partial h}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \tag{3.89c}$$

ζ is a ratio of the equatorial Rossby radius of deformation to the basin length ($\zeta = \lambda_0/L$).

3.3.1.2 Reduction of shallow water equations

We will begin with a simplification of the shallow water response to zonal wind forcing. We add a zonal wind stress τ to (3.89a) and damping with constant ϵ to all three equations (3.89). We will consider the dimensionless long-wave approximation to the shallow water equations (3.89) in which $\zeta \to 0$ (note this also allows us to drop damping term in (3.89b) due to scaling):

$$\frac{\partial u}{\partial t} - yv + \frac{\partial h}{\partial x} + \epsilon u = \tau, \tag{3.90a}$$

$$yu + \frac{\partial h}{\partial y} = 0, \tag{3.90b}$$

$$\frac{\partial h}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \epsilon h = 0. \tag{3.90c}$$

As before, x is zonal (east-west) direction, y is meridional (north-south) direction, and h is thermocline depth. Velocities are u and v in the x and y direction, respectively.

We will reduce (3.90) to a one-dimensional PDE for h. From (3.90b) we can see that u is a function of $\frac{\partial h}{\partial y}$, therefore $\frac{\partial u}{\partial t}$ can be expressed a function of h as well. Through this, equation (3.90a) can be used to express v as a function of h. With expressions for u and v as functions of h, both can be inserted into (3.90c) to recover a scalar PDE. We will now demonstrate this process explicitly.

Differentiating (3.90a) by y and multiplying through by y gives

$$yu_{yt} - yv - y^2 v_y + yh_{xy} + y\epsilon u_y = \tau_y y, \tag{3.91}$$

where the subscripts denote partial differentiation. Subtracting (3.90a) from (3.91) yields

$$(yu_y - u)_t - y^2 v_y + yh_{xy} + \epsilon y u_y - h_x - \epsilon u = y\tau_y - \tau. \tag{3.92}$$

Next we differentiate (3.90b) by both x and y. This results in the following PDEs,

$$y^2 u_x + y h_{xy} = 0, \tag{3.93a}$$

$$u + y u_y + h_{yy} = 0. \tag{3.93b}$$

Note that (3.93a) has also been multiplied by y. This allows for the relations of the following terms:

$$y h_{xy} = -y^2 u_x, \tag{3.94a}$$

$$y u_y = -u - h_{yy}. \tag{3.94b}$$

By substitution of the terms in (3.94a) and (3.94b) into (3.92) we have

$$(-2u - h_{yy})_t - y^2 v_y - y^2 u_x + \epsilon y u_y - h_x - \epsilon u = y\tau_y - \tau. \tag{3.95}$$

From (3.90b) and (3.90c) we also have the relations

$$u = -\frac{h_y}{y}, \tag{3.96a}$$

$$u_x + v_y = -h_t - \epsilon h. \tag{3.96b}$$

Using (3.96) in (3.95) we arrive at the scalar PDE

$$\left(y^2 + \frac{2}{y}\frac{\partial}{\partial y} - \frac{\partial^2}{\partial y^2}\right)\left(\frac{\partial h}{\partial t} + \epsilon h\right) - \frac{\partial h}{\partial x} = y\frac{\partial \tau}{\partial y} - \tau. \tag{3.97}$$

We have returned to the original notation of partial derivatives for grouping purposes.

The boundary conditions must be determined for (3.97). We take $x = 0, 1$ as rigid boundaries, meaning the reflected wave is inverted. The boundary conditions are given by Cane et al. (1981), and can be summarised as no zonal velocity at the eastern boundary ($x = 1$) and zero integrated zonal mass flux at the western boundary ($x = 0$). This can be written in terms of u as

$$u = 0 \quad \text{at} \quad x = 1, \tag{3.98a}$$

$$\int_{-\lambda_0/2}^{\lambda_0/2} u dy = 0 \quad \text{at} \quad x = 0. \tag{3.98b}$$

Through (3.96a) these can be written in terms of h:

$$\frac{\partial h}{\partial y} = 0 \quad \text{at} \quad x = 1, \tag{3.99a}$$

$$\int_{-\lambda_0/2}^{\lambda_0/2} \frac{1}{y}\frac{\partial h}{\partial y} dy = 0 \quad \text{at} \quad x = 0. \tag{3.99b}$$

3.3.1.3 A two-strip model

We can further reduce (3.97) to just one spatial dimension through appropriate approximations.

Equatorial strip Near the equator the thermocline depth can be approximated as

$$h(x,y,t) = h_e(x,t) + \frac{y^2}{2}(h_n(x,t) - h_e(x,t)). \tag{3.100}$$

where h_e is the depth of the thermocline at the equator and h_n represents the thermocline depth at an off-equatorial strip (note $h_n > h_e$). The zonal velocity is then obtained through (3.96a):

$$u = h_e(x,t) - h_n(x,t). \tag{3.101}$$

The definition of h in (3.100) can be substituted into (3.97) and results in the equation

$$\begin{aligned} y^2\Big(\frac{\partial h_e}{\partial t} + \epsilon h_e + \frac{y^2}{2}\Big(\frac{\partial h_n}{\partial t} - \frac{\partial h_e}{\partial t} + \epsilon(h_n - h_e)\Big)\Big) + \frac{\partial h_n}{\partial t} - \frac{\partial h_e}{\partial t} \\ +\epsilon(h_n - h_e) - \frac{\partial h_e}{\partial x} - \frac{y^2}{2}\Big(\frac{\partial h_n}{\partial x} - \frac{\partial h_e}{\partial x}\Big) = y\frac{\partial \tau}{\partial y} - \tau. \end{aligned} \tag{3.102}$$

The above is an equatorial approximation to (3.97). Applying (3.102) at $y = 0$, we arrive at the PDE with only one spatial dimension,

$$\Big(\frac{\partial}{\partial t} + \epsilon\Big)(h_n - h_e) - \frac{\partial h_e}{\partial x} = -\tau_{|y=0}. \tag{3.103}$$

Off-equatorial strip On the other hand, if we consider y large (i.e. $y \geq 2$), then the $\frac{\partial}{\partial y}$ and $\frac{\partial^2}{\partial y^2}$ terms in (3.97) become negligible. This allows for an off-equatorial approximation to (3.97) where h can be replaced with h_n giving

$$\Big(\frac{\partial}{\partial t} + \epsilon\Big)h_n - \frac{1}{y_n^2}\frac{\partial h_n}{\partial x} = \frac{\partial}{\partial y}\Big(\frac{\tau}{y}\Big)_{|y=y_n}. \tag{3.104}$$

Equation (3.103) represents the equatorial Kelvin waves including off-equatorial Rossby waves influence, while equation (3.104) represents the off-equatorial long Rossy waves. These two equations together form the two-strip model that we will analyse for internal delay effects.

The boundary conditions for (3.103) and (3.104) are given as

$$h_n(x_E,t) = r_E h_e(x_E,t) \quad \text{where} \quad x_E = 1, \tag{3.105a}$$

$$h_e(x_W,t) = r_W h_n(x_W,t) \quad \text{where} \quad x_W = 0, \tag{3.105b}$$

(Jin, 1997). The coefficients $r_{E,W}$ are reflection parameters; $r_{E,W} = 1$ signifies complete reflection of the waves at either boundary, while $r_{E,W} < 1$ represents imperfect reflection with mass exchange through boundary layer currents. Note that $r_E = r_W = 1$ make (3.105) equivalent to (3.99).

3.3.1.4 Incorporating SST variability

Near the equator, τ is a symmetric zonal wind stress and can be expressed as a function of SST $T_e(x,t)$ and west-east location x:

$$\tau(x,y,t) = \mu A(T_e(x,t),x)\mathrm{e}^{-\frac{1}{2}\left(\frac{yL_0}{L_a}\right)^2}. \tag{3.106}$$

Through an empirical relation between wind stress and SST (Deser and Wallace, 1990), (3.106) can be approximated as

$$\tau(x,y,t) = \mu A_0 T_{eE}(t) f(x)\mathrm{e}^{-\frac{1}{2}\left(\frac{yL_0}{L_a}\right)^2}, \tag{3.107}$$

where the west-east profile of wind stress $f(x)$ can be prescribed and doesn't depend on time. Here we assume the time dependence comes only from $T_{eE}(t)$, which is the SST at the eastern boundary of the basin ($x = 1$). Equation (3.107) then implies that the forcing terms of (3.103) and (3.104) take the values

$$-\tau_{|y=0} = -\mu A_0 T_{eE}(t) f(x), \tag{3.108a}$$

$$\frac{\partial}{\partial y}\left(\frac{\tau}{y}\right)_{|y=y_n} = -\theta\frac{\mu A_0}{y_n^2} T_{eE}(t) f(x), \tag{3.108b}$$

where $\theta = [\left(\frac{y_n L_0}{L_a}\right)^2 + 1]\mathrm{e}^{-\frac{1}{2}\left(\frac{y_n L_0}{L_a}\right)^2}$ is an $O(1)$ constant that varies between 1.00 and 1.25 depending on y_n.

In order to fully couple SST to the thermohaline, Jin and Neelin (1993) introduce an equatorial box approximation to the thermodynamic SST equation. Jin (1997) further simplifies this approximation through linearisation around an upwelling (vertically rising) background state to get

$$\frac{\partial T_e}{\partial t} = -c(x)T_e(x,t) + \gamma(x)h_e(x,t). \tag{3.109}$$

The function $c(x)$ captures local damping and $\gamma(x)$ is the zonal dependence on thermocline variations. In comparing the eastern and western contributions to SST variations, the dominant effects are where the thermocline is shallow in the Eastern Pacific. We can write (3.109) as just variations at the eastern boundary ($x = 1$),

$$\frac{dT_{eE}}{dt} = -c_E T_{eE}(t) + \gamma_E h_{eE}(t), \tag{3.110}$$

where the coefficients c_E and γ_E are now just constants. Equation (3.110) couples the SST anomaly to the equations of motion through the thermocline height at the eastern boundary.

As the two-strip model keeps y_n fixed as a parameter, (3.108) is a function of just $T_{eE}(t)$ and x, resulting in a closed system of equations for $h_e(x,t), h_n(x,t), T_{eE}(t)$.

3.3.1.5 Exploring wave-like solutions of two-strip model

We now have a simplified model of one spatial dimension for the equatorial Pacific given as

$$\Big(\frac{\partial}{\partial t}+\epsilon\Big)(h_n-h_e)-\frac{\partial h_e}{\partial x}=-\mu A_0 T_{eE}(t)f(x), \tag{3.111a}$$

$$\Big(\frac{\partial}{\partial t}+\epsilon\Big)h_n-\frac{1}{y_n^2}\frac{\partial h_n}{\partial x}=-\theta\frac{\mu A_0}{y_n^2}T_{eE}(t)f(x), \tag{3.111b}$$

$$\frac{dT_{eE}}{dt}=-c_E T_{eE}(t)+\gamma_E h_{eE}(t). \tag{3.111c}$$

We consider the undamped ($\epsilon=0$) system (3.111a-b):

$$\frac{\partial h_e}{\partial t}+\frac{\partial h_e}{\partial x}=\mu A_0 T_{eE}(t)f(x)+\frac{\partial h_n}{\partial t}, \tag{3.112a}$$

$$\frac{\partial h_n}{\partial t}-\frac{1}{y_n^2}\frac{\partial h_n}{\partial x}=-\theta\frac{\mu A_0}{y_n^2}T_{eE}(t)f(x). \tag{3.112b}$$

Rossby Wave Rossby waves travel westward across the ocean basin just above the equator. (3.112b) represents this wave and has the following characteristics,

$$\frac{dx}{ds}=-\frac{1}{y_n^2} \qquad x(s)=-\frac{s}{y_n^2}+x_0, \tag{3.113a}$$

$$\frac{dt}{ds}=1 \qquad t(s)=s+t_0, \tag{3.113b}$$

where $x_0=x(0)$ and $t_0=t(0)$. If we combine (3.113a) and (3.113b) then we get the characteristic

$$h_n(x,t)=h_n(x_0+\frac{t_0-t}{y_n^2},t). \tag{3.114}$$

Taking x as a function of t along constant characteristics of (3.114), (3.112b) is equivalent to

$$\frac{dh_n}{dt}=-\frac{\mu A_0\theta}{y_n^2}T_{eE}(t)f(x(t)). \tag{3.115}$$

If we consider a wave travelling westward from $x=1$ to $x=0$ (where $x_0=0$) and integrate (3.115) along characteristics then we have

$$\int_{t_0-y_n^2}^{t_0}\frac{dh_n}{dt}dt=-\frac{\mu A_0\theta}{y_n^2}\int_{t_0-y_n^2}^{t_0}T_{eE}(t)f(x(t))dt. \tag{3.116}$$

We can assume the wind stress profile $f(x)$ only acts in a region $x_1 < x < x_2$ and choose $f(x)$ as piecewise constant (Dijkstra, 2008):

$$f(x) = \begin{cases} \frac{1}{x_2 - x_1} & \text{for} \quad x_1 < x < x_2., \\ 0 & \text{otherwise.} \end{cases} \tag{3.117}$$

This implies that we have the relation

$$h_n(0,t) = h_n(1, t - y_n^2) - \frac{\mu A_0 \theta}{y_n^2 (x_2 - x_1)} \int_{t - y_n^2 x_1}^{t - y_n^2 x_2} T_{eE}(\tau) d\tau. \tag{3.118}$$

Here t_0 has been interchanged with the variable t.

Kelvin Wave Kelvin Waves travel eastward along the equator. Equation (3.112a) describes this wave (along with some coupling to the off-equatorial Rossby wave). It has the characteristics

$$\frac{dx}{ds} = 1 \qquad x(s) = s + x_0, \tag{3.119a}$$

$$\frac{dt}{ds} = 1 \qquad t(s) = s + t_0, \tag{3.119b}$$

Combining (3.119a) and (3.119b) gives the characteristic

$$h_e(x,t) = h_e(x_0 + t - t_0, t). \tag{3.120}$$

Again in (3.120), x can be considered as a function of t and (3.112a) becomes

$$\frac{dh_e}{dt} = \mu A_0 T_{eE}(t) f(x(t)) + \frac{\partial h_n}{\partial t}. \tag{3.121}$$

Now we consider a wave travelling eastward from $x = 0$ to $x = 1$ (where $x_0 = 0$) and integrate (3.121) along characteristics to obtain

$$\int_{t_0}^{t_0+1} \frac{dh_e}{dt} dt = \int_{t_0}^{t_0+1} \mu A_0 T_{eE}(t) f(x(t)) dt + \int_{t_0}^{t_0+1} \frac{\partial h_n}{\partial t} dt. \tag{3.122}$$

The function $f(x)$ is again taken as (3.117) which gives

$$h_e(1, t+1) = h_e(0,t) + \frac{\mu A_0}{(x_2 - x_1)} \int_{t+x_1}^{t+x_2} T_{eE}(\tau) d\tau + O\Big(\frac{1}{1 + y_n^2}\Big) \tag{3.123}$$

where t_0 is again replaced by t. To further simplify (3.118) and (3.123), we use the mean-value approximation:

$$\int_a^b g(t) dt = g(c)(b - a) \quad \text{where} \quad g(c) = \overline{g(t)}, t \in [a, b]. \tag{3.124}$$

This assumes we know the mean value of $T_{eE}(\tau)$ for $\tau \in [t+x_1, t+x_2]$. Also ignoring terms of $O\left(\frac{1}{1+y_n^2}\right)$, we get

$$h_e(1,t) = h_e(0,t-1) - \mu A_0 T_{eE}(t+x_p-1), \tag{3.125a}$$

$$h_n(0,t) = h_n(1,t-y_n^2) - \mu A_0 \theta T_{eE}(t-y_n^2). \tag{3.125b}$$

where $x_p \in [x_1, x_2]$ is the east-west position at which $T_{eE}(\tau)$ achieves its mean value. Since all the x values in (3.125) are at the boundaries, we can apply boundary conditions (3.105). Through substitution and use of boundary conditions (3.105), we arrive at the delay-difference system

$$\begin{aligned} h_{eW}(t) =& r_W r_E h_{eW}(t-1-y_n^2) \\ &+ \mu A_0 r_W (\theta T_{eE}(t-y_n^2 x_p) - r_E T_{eE}(t+x_p-1)), \end{aligned} \tag{3.126a}$$

$$\begin{aligned} h_{eE}(t) =& r_W r_E h_{eE}(t-y_n^2-1) \\ &+ \mu A_0 (T_{eE}(t+x_p-1) - r_W \theta T_{eE}(t-y_n^2 x_p-1)), \end{aligned} \tag{3.126b}$$

$$\frac{dT_{eE}}{dt} = -c_E T_{eE}(t) + \gamma_E h_{eE}(t) \tag{3.126c}$$

where $h_{eW}(t) = h_e(0,t)$, $h_{eE}(t) = h_e(1,t)$, and . The delay oscillator model is valid in the limiting case $r_E = 0$:

$$\frac{dT_{eE}}{dt} = -c_E T_{eE}(t) + \gamma_E \mu A_0 (T_{eE}(t-1+x_p) - \theta r_W T_{eE}(t-1-y_n^2 x_p)). \tag{3.127}$$

Introducing the coefficients $\delta_1 = 1 - x_p$, $\delta_2 = 1 + y_n^2 x_p$, $\alpha = c_E$, $\beta = \gamma_E \mu A_0$, and $b = \gamma_E \mu A_0 \theta r_W$ we can rewrite (3.127 - with $T = T_{eE}$) as

$$\frac{dT}{dt} = -\alpha T(t) + \beta T(t-\delta_1) - bT(t-\delta_2). \tag{3.128}$$

Considering long timescales and therefore setting $\delta_1 = 0$, $\delta_2 = \tau$, and $\beta - \alpha = a$, (3.128) becomes the linear part of

$$\frac{dT}{dt} = aT(t) - bT(t-\tau) - cT^3(t), \tag{3.129}$$

which is the Battisti and Hirst (1989) delayed oscillator model. A rescaling of $a = c = 1$ gives the original model proposed by Suarez and Schopf (1988). The nonlinear part in both models is added as an approximation of all the nonlinear effects known to limit the growth of unstable modes in the system, including advective processes in the ocean and interaction with atmospheric convection (Suarez and Schopf, 1988). This model has been studied extensively through bifurcation analysis and extensions (discussed in Chapter 2, Section 2.2.2).

3.3.2 Atlantic Multidecadal Oscillation (AMO)

Another phenomenon affecting sea-surface temperature (SST) is the Atlantic Multidecadal Oscillation (AMO). This is an oscillation on decadal timescales which affects surface and subsurface temperatures in the North Atlantic Ocean. Despite the relatively short climate record that exists for this variability, the period is believed to be on the order of 50 to 70 years (Dijkstra, 2005). The full physical mechanism is explained in Te Raa and Dijkstra (2002), but the most significant effect is westward propagation of temperature anomalies in the northern part of the Atlantic Ocean basin.

In this section we analyse a previously proposed model of the AMO introduced by Sévellec and Huck (2015). This model is a PDE system which describes the wave-like dynamics of surface, subsurface, and deep ocean temperatures in the North Atlantic. Using methods similar to those introduced in the previous section for ENSO, we recover a delay difference equation that captures the essential dynamics of the AMO.

3.3.2.1 The model

The AMO model defined in Sévellec and Huck (2015) is as follows,

$$\frac{\partial T_1}{\partial t} = -a_1\frac{\partial T_1}{\partial x} - b_1\frac{\partial T_2}{\partial x} - c_1\frac{\partial T_3}{\partial x} + \kappa\frac{\partial^2 T_1}{\partial x^2}, \tag{3.130a}$$

$$\frac{\partial T_2}{\partial t} = -a_2\frac{\partial T_2}{\partial x} - b_2\frac{\partial T_2}{\partial x} - c_2\frac{\partial T_3}{\partial x} + \kappa\frac{\partial^2 T_2}{\partial x^2}, \tag{3.130b}$$

$$\frac{\partial T_3}{\partial t} = \kappa\frac{\partial^2 T_3}{\partial x^2}. \tag{3.130c}$$

These equations describe temperature anomalies from a background state for three layers of the ocean. Layers 1,2, and 3 are the surface, subsurface, and deep ocean, respectively. It is important to note for physical purposes that the coefficients $a_1, b_1, c_1, a_2, b_2, c_2$ are negative. The constant κ is a diffusion coefficient that is the same for all three layers.

We will consider a basin of $x \in [0, 1]$ where 0 is west and 1 is east. The boundary conditions are given as

$$T_i(1, t) = -T_i(0, t), \tag{3.131a}$$

$$\frac{\partial T_i}{\partial x}(1, t) = -\frac{\partial T_i}{\partial x}(0, t). \tag{3.131b}$$

These boundary conditions imply an instantaneous temperature adjustment across the ocean. The physical interpretation of this is that the timescale of the eastward propagating Kelvin wave travelling back across the basin (from 0 to 1) is much faster than the Rossby wave described by the model, so we can consider it as an instantaneous adjustment at the boundaries. This is supported by the studies

of Johnson and Marshall (2002) who showed the timescale of Kelvin wave adjustment is 2-3 months, while the westward wave propagation of interest is on decadal timescales.

3.3.2.2 Free-wave solutions

First we derive explicit expressions for the free-wave solutions of (3.130), meaning we neglect the diffusion terms ($\kappa = 0$). System (3.130) becomes

$$\frac{\partial T_1}{\partial t} = -a_1\frac{\partial T_1}{\partial x} - b_1\frac{\partial T_2}{\partial x} - c_1\frac{\partial T_3}{\partial x}, \tag{3.132a}$$

$$\frac{\partial T_2}{\partial t} = -a_2\frac{\partial T_2}{\partial x} - b_2\frac{\partial T_2}{\partial x} - c_2\frac{\partial T_3}{\partial x}, \tag{3.132b}$$

$$\frac{\partial T_3}{\partial t} = 0. \tag{3.132c}$$

Equation (3.132c) shows that T_3 is constant in time. Since T_1, T_2 are not coupled into (3.132c), we know that as $t \to \infty$, T_3 is constant in x. This simplifies (3.132) further to

$$\frac{\partial T_1}{\partial t} = -a_1\frac{\partial T_1}{\partial x} - b_1\frac{\partial T_2}{\partial x}, \tag{3.133a}$$

$$\frac{\partial T_2}{\partial t} = -a_2\frac{\partial T_2}{\partial x} - b_2\frac{\partial T_2}{\partial x}. \tag{3.133b}$$

This is a system of two coupled wave equations. To solve along characteristics we write the system in matrix form,

$$\frac{\partial \vec{T}}{\partial t} + B\frac{\partial \vec{T}}{\partial x} = \vec{0}, \tag{3.134}$$

where

$$\vec{T} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, \qquad B = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}. \tag{3.135}$$

Using coordinates in $\mathbb{R}^2$ which diagonalise B into the form

$$B = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \tag{3.136}$$

where

$$\lambda_{1,2} = \frac{1}{2}\Big(a_1 + b_2 \pm \sqrt{(a_1+b_2)^2 - 4a_1b_2 + 4a_2b_1}\Big). \tag{3.137}$$

we create a new system of uncoupled variables $\tilde{T}$ that are governed by the equations

$$\frac{\partial \tilde{T}_1}{\partial t} + \lambda_1\frac{\partial \tilde{T}_1}{\partial x} = 0, \tag{3.138a}$$

$$\frac{\partial \tilde{T}_2}{\partial t} + \lambda_2\frac{\partial \tilde{T}_2}{\partial x} = 0.. \tag{3.138b}$$

The characteristic speeds of the waves are also given by (3.137). The new variables are related back to our original variables through

$$\vec{T} = P\tilde{T}, \qquad P = [\vec{e}_1, \vec{e}_2], \tag{3.139}$$

where

$$\vec{e}_1 = \begin{bmatrix} 1 \\ \frac{\lambda_1 - a_1}{b_1} \end{bmatrix}, \qquad \vec{e}_2 = \begin{bmatrix} \frac{b1}{\lambda_2 - a_1} \\ 1 \end{bmatrix}. \tag{3.140}$$

The general solution of (3.138) is of the form

$$\tilde{T}_i(x,t) = \tilde{T}_i(\lambda_i t + 1, t), \qquad i = 1, 2. \tag{3.141}$$

This implies the following relations at the boundaries:

$$\tilde{T}_1(t,1) = \tilde{T}_1(t - \frac{1}{\lambda_1}, 0), \tag{3.142a}$$

$$\tilde{T}_2(t,1) = \tilde{T}_2(t - \frac{1}{\lambda_2}, 0). \tag{3.142b}$$

Through (3.131) and (3.139) the boundary conditions for the new system become

$$\tilde{T}_{1W}(t) = -\tilde{T}_{1E}(t), \tag{3.143a}$$

$$\tilde{T}_{2W}(t) = -\tilde{T}_{2E}(t). \tag{3.143b}$$

We use the shorthand notation $\tilde{T}_{iE}(t) = \tilde{T}_i(t,1)$ and $\tilde{T}_{iW}(t) = \tilde{T}_i(t,0)$. The general solution (3.141) allows us to express the value of $\tilde{T}_i$ at any x in terms of $\tilde{T}_i$ at a chosen location (here we choose the eastern boundary of the basin, $x = 1$). Applying (3.143) to (3.142) we obtain

$$\tilde{T}_{1E}(t) = -\tilde{T}_{1E}(t - \frac{1}{\lambda_1}), \tag{3.144a}$$

$$\tilde{T}_{2E}(t) = -\tilde{T}_{2E}(t - \frac{1}{\lambda_2}). \tag{3.144b}$$

Now we can change back to our original variables through

$$\tilde{T} = P^{-1}\vec{T}, \qquad P^{-1} = \begin{bmatrix} \frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} & \frac{-b_1}{\lambda_2 - \lambda_1} \\ \frac{-(\lambda_2 - a_1)(\lambda_1 - a_1)}{(\lambda_2 - \lambda_1)b_1} & \frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} \end{bmatrix}. \tag{3.145}$$

This gives the system

$$\begin{aligned} T_{1E}(t) =& -\frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} T_{1E}\Big(t - \frac{1}{\lambda_1}\Big) + \frac{b_1}{\lambda_2 - \lambda_1} T_{2E}\Big(t - \frac{1}{\lambda_1}\Big) \\ &+ \frac{\lambda_1 - a_1}{\lambda_2 - \lambda_1} T_{1E}\Big(t - \frac{1}{\lambda_2}\Big) - \frac{b_1}{\lambda_2 - \lambda_1} T_{2E}\Big(t - \frac{1}{\lambda_2}\Big), \end{aligned} \tag{3.146a}$$

$$\begin{aligned} T_{2E}(t) =& -\frac{(\lambda_2 - a_1)(\lambda_1 - a_1)}{b_1(\lambda_2 - \lambda_1)} T_{1E}\Big(t - \frac{1}{\lambda_1}\Big) + \frac{\lambda_1 - a_1}{\lambda_2 - \lambda_1} T_{2E}\Big(t - \frac{1}{\lambda_1}\Big) \\ &+ \frac{(\lambda_2 - a_1)(\lambda_1 - a_1)}{b_1(\lambda_2 - \lambda_1)} T_{1E}\Big(t - \frac{1}{\lambda_2}\Big) - \frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} T_{2E}\Big(t - \frac{1}{\lambda_2}\Big). \end{aligned} \tag{3.146b}$$

System (3.146) is a delay-difference system. We can write the system in matrix form to analyse the combined effect of both delays,

$$\vec{T}_E(t) = C_1 \vec{T}_E(t - \tau_1) + C_2 \vec{T}_E(t - \tau_2), \tag{3.147}$$

where

$$C_1 = \begin{bmatrix} -\frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} & \frac{b_1}{\lambda_2 - \lambda_1} \\ -\frac{(\lambda_2 - a_1)(\lambda_1 - a_1)}{b_1(\lambda_2 - \lambda_1)} & \frac{\lambda_1 - a_1}{\lambda_2 - a_1} \end{bmatrix}, \qquad C_2 = \begin{bmatrix} \frac{\lambda_1 - a_1}{\lambda_2 - \lambda_1} & -\frac{b_1}{\lambda_2 - \lambda_1} \\ \frac{(\lambda_2 - a_1)(\lambda_1 - a_1)}{b_1(\lambda_2 - \lambda_1)} & -\frac{\lambda_2 - a_1}{\lambda_2 - \lambda_1} \end{bmatrix},$$

and

$$\tau_1 = -\frac{1}{\lambda_1}, \qquad \tau_2 = -\frac{1}{\lambda_2}.$$

Note that we have made a time shift in (3.147) since $\lambda_{1,2}$ will always be negative.

3.3.2.3 Numerical exploration of AMO model with delay

In order to explore solutions of delay-difference system, we convert (3.147) to a system of DDEs. The method for doing this is explained in Chapter 4. Here we just present the resulting system:

$$\epsilon \frac{d\vec{T}_E(t)}{dt} = -\vec{T}_E(t) + C_1 \vec{T}_E(t - \tau_1) + C_2 \vec{T}_E(t - \tau_2), \tag{3.148}$$

where $\epsilon \ll 1$. The choice of ϵ can be related to the discretisation of the original PDE system (3.133) through $\epsilon = 1/N$, where N is the number of discretisation steps using an 'upwind' scheme (discretising in the direction of the wave).

We implement the system in MATLAB. Table 3.2 shows the parameter values used, taken as suggested in Sévellec and Huck (2015). The last two columns of the table list the approximate resulting wave speeds and delays.

To compute the history needed for the difference equation, the PDE system (3.133) is solved numerically for an initial profile of the basin using an upwind discretisation scheme for 34.14 years. We take

a_1	-0.1083	$\lambda_{1,\text{app}}$	-0.0293
a_2	-0.0352	$\lambda_{2,\text{app}}$	-0.2482
b_1	-0.3140	$\tau_{1,\text{app}}$	34.14
b_2	-0.1692	$\tau_{2,\text{app}}$	4.03

Table 3.2: Parameters used in numerical computations of (3.148).

a Gaussian initial distribution profile (shown in Figure 3.6a). The DDE system (3.148) is then evolved for a further 200 time-steps. Figure 3.6 show the results.

A spectral analysis is performed on the resulting trajectories to identify the most prominent oscillation periods. A dominant signal of an 8.08 year cycle is obtained, and two smaller signals for 69.40 and 22.91 year cycles. The two most prominent signals correspond to period doublings of the two delay values. The smaller peak corresponds to approximately $\frac{2}{3}\tau_1$. The two smaller signals, corresponding to 69.40 and 22.91 year cycles, align with the literature regarding possible cycle lengths of the AMO (Chylek et al., 2011, Delworth and Mann, 2000).

We can compare this behaviour when we use a more realistic history to initialise DDE system (3.148). We use the Kaplan SST V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at https://www.esrl.noaa.gov/psd/. The historical recordings begin in 1856, so we initialise our model at 1890. This history, along with the rest of the data until 2018, is depicted in Figure 3.7a. Our model results from using the data history is also shown in the same figure for comparison (note we only show the top layer, T_1). At first glance it appears that the model captures the low frequency variations well. This is further confirmed through a spectral analysis shown in Figure 3.7b. Here we see that the dominant low frequency variation of the model and the data are almost exact. There appear to be some additional resonances in the higher frequency variations of the model. It is important to note, however, that with the data history the strongest signal is in fact the low frequency variation of approximately 69 years, rather than the period doubling of the smaller delay ($\approx$ 8 years) seen in the previous example (see Figure 3.6d). This provides evidence that the AMO is a self-sustained oscillation induced by delay effects.

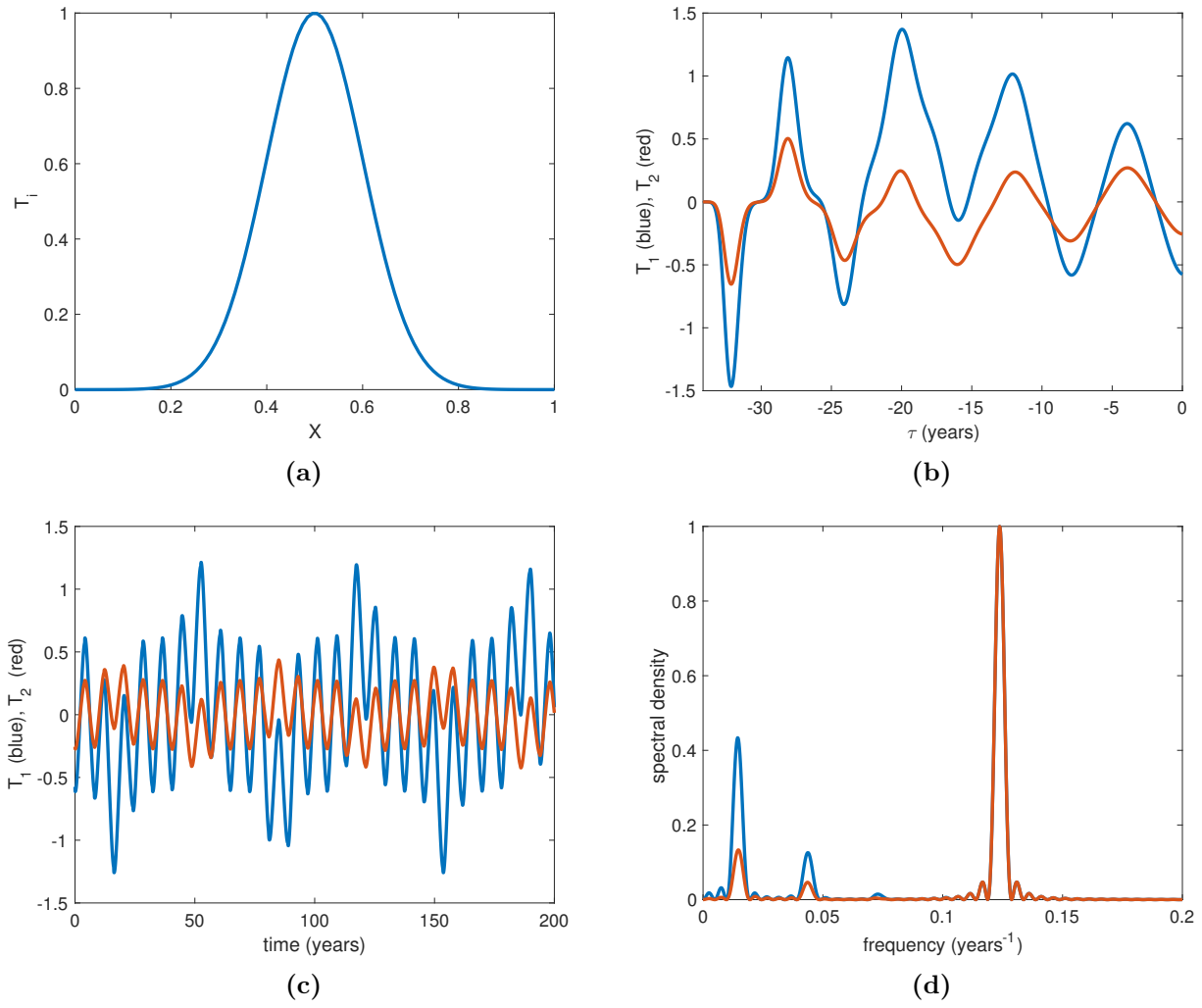
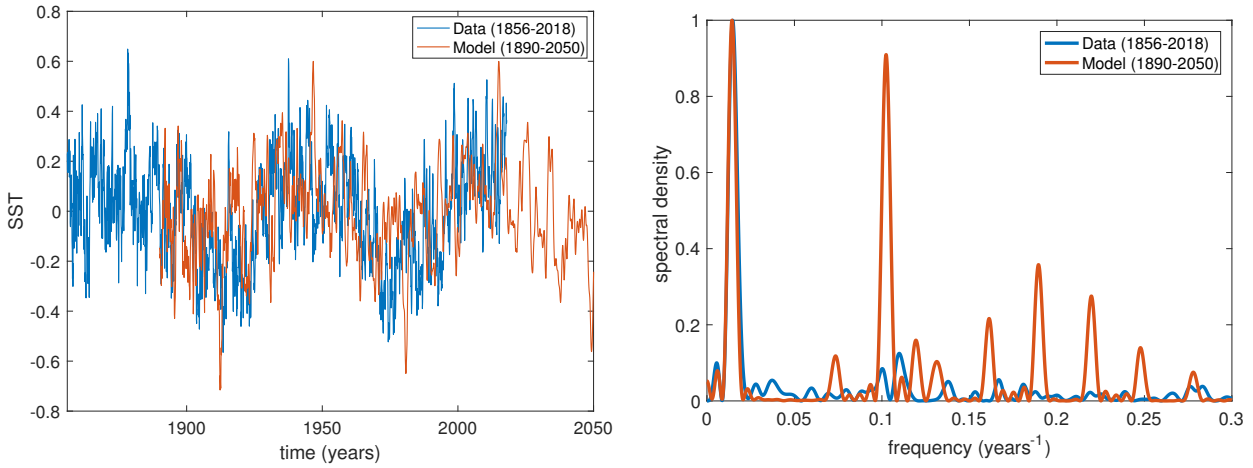


Figure 3.6: Numerical results for (3.148) with parameters from Table 3.2. (a) Initial distribution profile for each T. (b) History function. (c) Trajectory for 200 years. (d) Power spectral density.



(a) **(b)**

Figure 3.7: Numerical results for (3.148) initialised with Kaplan SST V2 data provided by NOAA/OAR/ESRL PSD (https://www.esrl.noaa.gov/psd/) . (a) Time series of Kaplan SST data compared with model output. (b) Power spectral density of Kaplan SST data and model output.

Chapter 4

Analysing a Boolean Delay Model of the Pleistocene

We have introduced models with delay for the climate system in Chapter 2. Here we focus on one particular model for the Pleistocene, a period of climate characterised by a succession of glacial-interglacial cycles (see Section 2.3.2 for full description). This model was proposed by Ghil et al. (1987) and presents the climate system in the Boolean sense: *on* or *off* states with logical relations between variables. The aim of the original study was to conceptually understand the oscillatory nature of the ice ages through internal dynamics of the climate system. In Section 4.1 we describe the model in detail, including its physical rationale. Section 4.2 details a method to embed Boolean Delay Equations (BDEs) into difference equations, while Section 4.3 describes the regularisation of difference equations by singularly perturbed delay differential equations (DDEs). Finally we perform a numerical bifurcation analysis in Section 4.4 and collect its scientifically relevant conclusions.

4.1 The Boolean Delay Model

Ghil, Mullhaupt, and Pestiaux introduced Boolean delay equations (BDEs) to theoretical climate dynamics studies. In 1987 they developed a model to describe the interaction between ice volume, global temperature, and the deep-ocean circulation during the Pleistocene (Ghil et al., 1987). They wanted to investigate the effects that deep-ocean circulation changes could have on the climate fluctuations in that time period. Previously it had been hypothesised that the ocean circulation could have been slower in the past, causing a longer transport time of cold, dense deep water from formation sites to the rest of the ocean (Worthington, 1968, as cited in Ghil et al. (1987)). Ghil et al. considered the possibility of delays relating to the ocean circulation (as well as other physical processes) playing a role in the observed oscillatory behaviour of paleoclimate records. Ghil et al.'s interest in BDEs arose

from the simplicity of the equation states. The state could be described by a simple "high" or "low" value, represented as 1 and 0 respectively. This requires only the understanding of the logical nature of mechanisms and the delays after which effects take place. With this approach one can avoid the use of largely unknown physical parameters, but still attain an understanding of how delays can impact the long-term behaviour of a system (Ghil et al., 2008).

The model the authors constructed has of three state variables: northern hemisphere ice volume (V), global air temperature (T), and intensity of deep-ocean circulation (C). They chose to focus solely on northern hemisphere ice cover because of its dominating effect on deep water formation (Ghil et al., 1987). The two main types of deep water are Antarctic Bottom Water (AABW) from the southern hemisphere and North Atlantic Deep Water (NADW) from the northern hemisphere. These cold, dense waters form due to the cooling of dense, warm water from the tropics as it is transported to the poles. Brine rejection also contributes to the process, which is the expulsion of salt during sea-ice formation, making the surrounding surface water much denser and prompting convective mixing. The present-day locations of all deep water formation and sinking are the North Atlantic Subpolar Sea (north of the Greenland-Iceland-Scotland ridge), the Weddell Sea, and the Ross Sea (Stommel, 1958; Weyl, 1968; Worthington, 1968; as cited in Ghil et al., 1987). The latter two are located off the coast of Antarctica. Ghil et al. use the assumption that the present extent and volume of the Antarctic ice sheet were reached before the Pleistocene and maintained throughout that period. This then assumes that there are no changes in AABW production, leaving NADW the main driver of circulation fluctuations. In the Boolean notation used here, the variables taking a value of 1 mean "high" for V and T, and "strong" for C. A value of 0 denotes "low" for V and T, and "weak" for C. There are also three delays that are incorporated into the system: τ_{acc} is the time it takes for one layer of an ice sheet to accumulate, τ_{if} is the time it takes for the surface area of an ice sheet to increase due to viscoplastic expansion, and τ_{dw} is the overturning time of the ocean.

The original Boolean model is given as

$$V(t) = T(t - \tau_{\text{acc}}), \tag{4.1a}$$

$$T(t) = \overline{V}(t - \tau_{\text{if}}) \wedge \overline{C}(t - \tau_{\text{dw}}), \tag{4.1b}$$

$$C(t) = \overline{V(t - \tau_{\text{dw}}) \wedge \overline{V}(t)} = \overline{V}(t - \tau_{\text{dw}}) \vee V(t). \tag{4.1c}$$

The Boolean operators are as follows: $\wedge$ is conjuction (*and*), $\vee$ is alternative (*or*), and the overbar indicates negation (*not*).

The physical rationale for the model is explained as follows. Equation (4.1a) is derived from the precipitation-temperature effect. The end result is that there exists a volume increase with a temperature increase having occurred τ_{acc} in the past, which sparked a more active hydrological cycle. Equation (4.1b) results from the temperature-albedo feedback. The first term is the more obvious relation, saying if there was more ice cover in the past, the overall surface albedo will increase after

τ_{if} time for the ice to expand and the global temperature will decrease. Along with this effect comes the ice volume effect on circulation, which is influenced by increased NADW formation during periods of glaciation. This increased NADW formation would strengthen the overall ocean circulation, therefore transporting cold surface waters to the rest of the ocean which also supports decreasing global temperature. Equation (4.1c) explains the ocean circulation's dependence on total ice cover. It is better understood by the first equivalence $C(t) = \overline{V(t-\tau_{\text{dw}}) \wedge \overline{V}(t)}$. Namely, if there is much ice melt during time interval τ_{dw}, fresh water runoff will stabilise the North Atlantic water column, producing less NADW (i.e. less convective overturning), and the overall circulation rate will decrease. $\overline{V}(t-\tau_{\text{dw}}) \vee V(t)$ is the logical equivalence to this statement and used in this particular analysis for its algebraic simplicity (discussed in the following section).

4.2 Boolean Delay Equations to Difference Equations

In order to enable a systematic study of the behaviour in dependence of the parameters, the basic Boolean equations must first be converted to continuous maps on the interval [0,1]. The two fundamental Boolean operators are the *identity* and *not* functions. Since the maps will be further converted to a continuous function, it is not necessary that the maps are strictly defined on the interval [0,1] as long as the orbits of any real number close to 0 or 1 have the same behaviour as the closest integer.

The map to function as the *identity* logical operator would require that there are two stable fixed points at 0 and 1. We use

$$x_{n+1} = x_n^2(x_n - 2)^2. \tag{4.2}$$

As can be seen in Figure 4.1, values close to either fixed point will converge to that fixed point. The map corresponding to the *not* logical operator would require a stable periodic orbit of period 2, with the periodic orbit visiting the points 0 and 1. This means that a past value of 1 will be sent to 0 and vice versa, in agreement with the Boolean logical operator. This applies to any values close to the periodic points. Figure 4.2 illustrates this behaviour. The map is defined as

$$x_{n+1} = (x_n + 1)(x_n - 1)^2. \tag{4.3}$$

The logical operators $\vee$ and $\wedge$ can be expressed directly using algebraic operators. This is done by using the integer values 0,1 to represent false and true (respectively). Given two logical values p and q, the truth tables must match the final algebraic quantity. The appropriate algebraic expressions are then (using $*$ to indicate multiplication)

$$p \wedge q = p * q \tag{4.4}$$

$$p \vee q = p + q - p * q \tag{4.5}$$

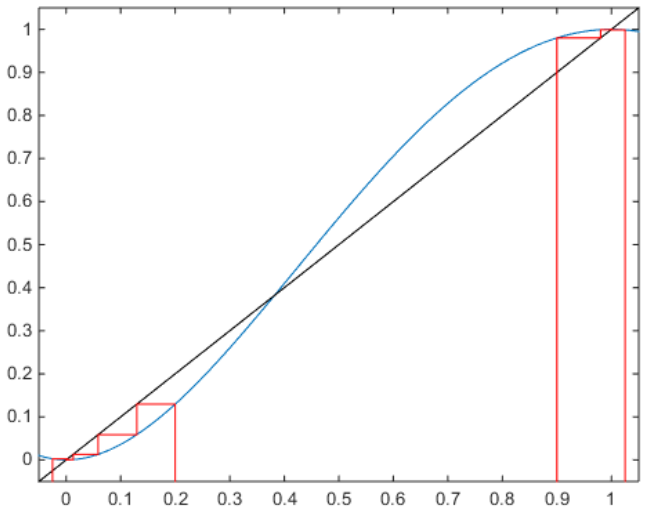
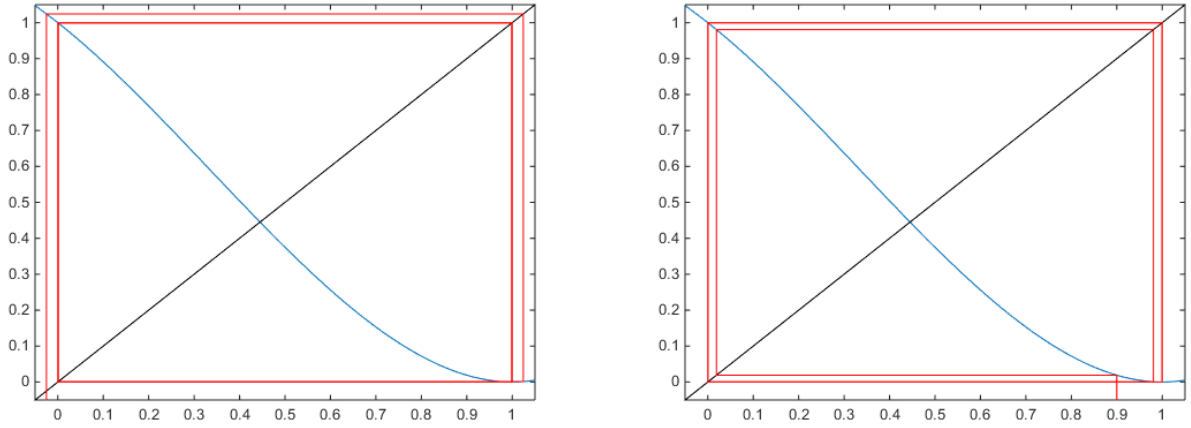


Figure 4.1: Orbits of initial points on the *identity* map (4.2) after 100 iterations. Initial values $x_0 = \{-0.025, 0.2, 0.9, 1.025\}$



(a) **(b)**

Figure 4.2: Orbits of initial points on the *not* map (4.3) after 100 iterations with initial values (a) $x_0 = -0.025$ and (b) $x_0 = 0.9$.

Note the truth tables of the Boolean operations, with added columns for the algebraic expressions, in Table 4.1. The calculations are trivial. Using the *identity* map, *not* map, and algebraic operators, the BDE (4.1) is embedded in the following system of delay difference equations:

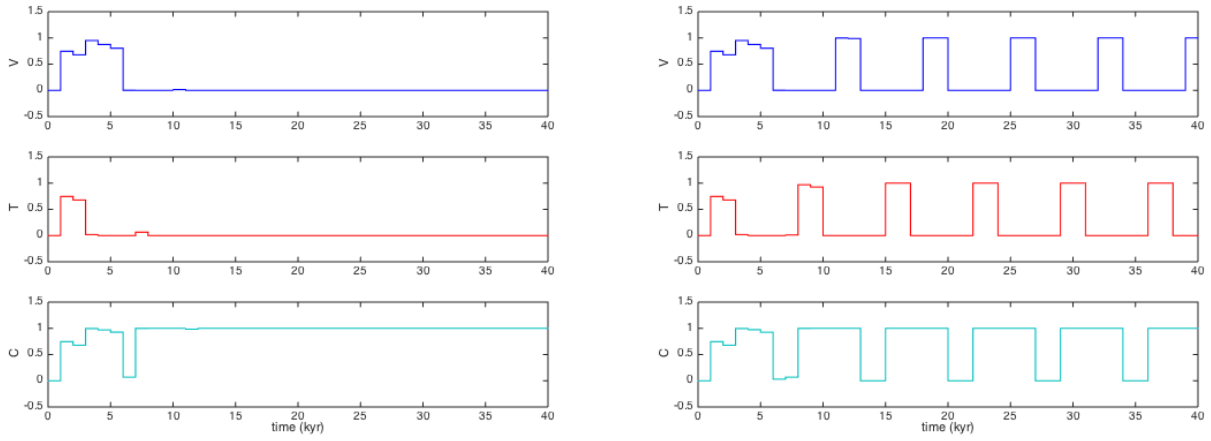
$$V(t) = T(t-\tau_{\text{acc}})^2(T(t-\tau_{\text{acc}})-2)^2, \tag{4.6a}$$

$$T(t) = (V(t-\tau_{\text{if}})+1)(V(t-\tau_{\text{if}})-1)^2(C(t-\tau_{\text{dw}})+1)(C(t-\tau_{\text{dw}})-1)^2, \tag{4.6b}$$

$$\begin{aligned} C(t) = {} & (V(t-\tau_{\text{dw}})+1)(V(t-\tau_{\text{dw}})-1)^2 + V(t)^2(V(t)-2)^2 \\ & - (V(t-\tau_{\text{dw}})+1)(V(t-\tau_{\text{dw}})-1)^2V(t)^2(V(t)-2)^2. \end{aligned} \tag{4.6c}$$

Table 4.1: Truth tables for $\wedge$ and $\vee$ logical operators, with algebraic equivalents.

p	q	$p \wedge q$	$p \vee q$	$p * q$	$p + q - p * q$
0	0	0	0	0	0
0	1	0	1	0	1
1	0	0	1	0	1
1	1	1	1	1	1



(a) **(b)**

Figure 4.3: Solution of (4.6) with (a) $\tau_{\text{acc}} = 3$, $\tau_{\text{if}} = 2$, $\tau_{\text{dw}} = 1$ and (b) $\tau_{\text{acc}} = 3$, $\tau_{\text{if}} = 1$, $\tau_{\text{dw}} = 2$.

We find that the asymptotic behaviour of the trajectory starting from $V(0) = C(0) = T(0) = 0$ changes depending on the relations between the delay values. Take for instance the simplest case of the delays being consecutive discrete values (namely 1,2,3). For $\tau_{\text{acc}} > \tau_{\text{if}} > \tau_{\text{dw}}$ the solution converges to an equilibrium, while for $\tau_{\text{acc}} > \tau_{\text{dw}} > \tau_{\text{if}}$ the solution oscillates. These solutions are shown in Figure 4.3a and Figure 4.3b respectively. Because this model has no additional parameters, we will investigate how the attractors (such as equilibria and periodic orbits) change as we vary the delays of the system.

4.3 Regularisation of Difference Equations to Delay Differential Equations

It is worthwhile to analyse the model on a continuous time spectrum to determine the exact delay relationships where periodic orbits appear. For this reason, we regularise the difference equation into a delay differential equation. A detailed explanation is shown only for one variable, but the other two are done identically. Consider (4.6a),

$$V(t) = T(t - \tau_{\text{acc}})^2 (T(t - \tau_{\text{acc}}) - 2)^2. \tag{4.7}$$

We rewrite (4.7) as

$$0 = -V(t) + T(t-\tau_{\text{acc}})^2(T(t-\tau_{\text{acc}})-2)^2. \tag{4.8}$$

Introducing an additional small time step $\epsilon = \Delta t$, we then regularise by replacing 0 with $\epsilon\frac{dV}{dt}$:

$$\epsilon\frac{dV}{dt} = -V(t) + T(t-\tau_{\text{acc}})^2(T(t-\tau_{\text{acc}})-2)^2. \tag{4.9}$$

The DDE (4.9) is then equivalent to difference equation (4.7) in the limit $\epsilon \to 0$. For all three variables, the adopted DDE system is then

$$\epsilon\frac{dV}{dt} = -V(t) + T(t-\tau_{\text{acc}})^2(T(t-\tau_{\text{acc}})-2)^2 \tag{4.10a}$$

$$\epsilon\frac{dT}{dt} = -T(t) + (V(t-\tau_{\text{if}})+1)(V(t-\tau_{\text{if}})-1)^2(C(t-\tau_{\text{dw}})+1)(C(t-\tau_{\text{dw}})-1)^2 \tag{4.10b}$$

$$\begin{aligned}\epsilon\frac{dC}{dt} = &-C(t) + (V(t-\tau_{\text{dw}})+1)(V(t-\tau_{\text{dw}})-1)^2 + V(t)^2(V(t)-2)^2 \\ &- (V(t-\tau_{\text{dw}})+1)(V(t-\tau_{\text{dw}})-1)^2V(t)^2(V(t)-2)^2\end{aligned} \tag{4.10c}$$

where ϵ is a time-scaling parameter. When $\epsilon = 0$ system (4.10) is identical to system (4.6). Larger ϵ values slows down transitions between 0 and 1.

4.3.1 Boundedness of solutions

We show that solutions starting with a history in $[0,1]$ are bounded within the interval $[0,1]$. Assume that the initial history for (4.10) satisfies $(V(s),T(s),C(s)) \in [0,1]$ for all $s \in [-\max(\tau_{\text{acc}},\tau_{\text{if}},\tau_{\text{dw}}),0]$. Then $(V(t),T(t),C(t)) \in [0,1]^3$ for all $t \geq 0$.

Without loss of generality, the boundedness of each individual variable will be explored for $\epsilon = 1$. To assess whether or not a variable will leave the given interval, only the boundaries of the interval need be considered.

Boundedness of V(t) Suppose $V(t)$ is the first variable to leave the interval $[0,1]$. First we consider the case where $V(t)$ leaves $[0,1]$ crossing 1. Thus, there exists a t_* such that $V(t_*) = 1$ and $\frac{dV}{dt}(t_*) > 0$, where $0 \leq V(t) \leq 1$ for all $t \leq t_*$. From (4.10a),

$$\frac{dV}{dt}(t_*) = -1 + T(t_*-\tau_{\text{acc}})^2(T(t_*-\tau_{\text{acc}})-2)^2. \tag{4.11}$$

Since $0 \leq T(t_*-\tau_{\text{acc}}) \leq 1$ by assumption,

$$T(t_*-\tau_{\text{acc}})^2(T(t_*-\tau_{\text{acc}})-2)^2 \leq (1)^2(1-2)^2 = 1. \tag{4.12}$$

Thus,

$$\frac{dV}{dt}(t_*) \leq -1 + 1 = 0. \tag{4.13}$$

This is a contradiction. Therefore, $V(t)$ cannot be the first variable to leave $[0, 1]$ through 1, so $V(t) \leq 1$.

The other possible case is $V(t)$ leaves $[0, 1]$ through 0. In this case, there exists a t_* such that $V(t_*) = 0$ and $\frac{dV}{dt}(t_*) < 0$, where $0 \leq V(t) \leq 1$ for all $t \leq t_*$. From (4.10a),

$$\frac{dV}{dt}(t_*) = T(t_* - \tau_{\text{acc}})^2(T(t_* - \tau_{\text{acc}}) - 2)^2 \tag{4.14}$$

Again, since $0 \leq T(t_* - \tau_{\text{acc}}) \leq 1$ by assumption,

$$T(t_* - \tau_{\text{acc}})^2(T(t_* - \tau_{\text{acc}}) - 2)^2 \geq 0^2(-2)^2 = 0 \tag{4.15}$$

Hence,

$$\frac{dV}{dt}(t_*) \geq 0. \tag{4.16}$$

This is again a contradiction. Hence $V(t)$ cannot be the first variable to leave $[0, 1]$ through 0. Since $V(t)$ cannot be the first variable to cross 0 or 1, it cannot be the first to leave $[0, 1]$ overall. The boundedness of solutions of the other two variables are proved in a similar fashion.

Boundedness of T(t) Suppose $T(t)$ is the first variable to leave the interval $[0, 1]$ through 1. Thus, there exists a t_* such that $T(t_*) = 1$ and $\frac{dT}{dt}(t_*) > 0$, where $0 \leq T(t) \leq 1$ for all $t \leq t_*$. From (4.10b),

$$\frac{dT}{dt}(t_*) = -1 + (V(t_* - \tau_{\text{if}}) + 1)(V(t_* - \tau_{\text{if}}) - 1)^2(C(t_* - \tau_{\text{dw}}) + 1)(C(t_* - \tau_{\text{dw}}) - 1)^2. \tag{4.17}$$

Applying the assumed inequality $\frac{dT}{dt}(t_*) > 0$ and rearranging we obtain

$$(V(t_* - \tau_{\text{if}}) + 1)(V(t_* - \tau_{\text{if}}) - 1)^2(C(t_* - \tau_{\text{dw}}) + 1)(C(t_* - \tau_{\text{dw}}) - 1)^2 > 1. \tag{4.18}$$

Using the difference of two squares equality, the left-hand side can be written equivalently as

$$(V(t_* - \tau_{\text{if}}) - 1)(V(t_* - \tau_{\text{if}})^2 - 1)(C(t_* - \tau_{\text{dw}}) - 1)(C(t_* - \tau_{\text{dw}})^2 - 1) > 1. \tag{4.19}$$

It is assumed that $0 \leq V(t_* - \tau_{\text{if}}) \leq 1$, which means that $(V(t_* - \tau_{\text{if}}) - 1)(V(t_* - \tau_{\text{if}})^2 - 1)$ is positive and less than (or equal to) one. The same argument is made for $(C(t_* - \tau_{\text{dw}}) - 1)(C(t_* - \tau_{\text{dw}})^2 - 1)$. Their multiplication then implies the left-hand side is indeed less than or equal to one, and a contradiction arises. Therefore, $T(t) \leq 1$ for all $t > 0$.

Suppose $T(t)$ is the first variable to leave the interval $[0,1]$ through 0. Thus, there exists a t_* such that $T(t_*) = 0$ and $\frac{dT}{dt}(t_*) < 0$, where $0 \leq T(t) \leq 1$ for all $t \leq t_*$. From (4.10b),

$$\frac{dT}{dt}(t_*) = (V(t_* - \tau_{\text{if}}) + 1)(V(t_* - \tau_{\text{if}}) - 1)^2(C(t_* - \tau_{\text{dw}}) + 1)(C(t_* - \tau_{\text{dw}}) - 1)^2. \tag{4.20}$$

From the assumption that $T(t)$ is the first variable to leave $[0,1]$, all terms on the right-hand side are greater than or equal to 0, meaning $\frac{dT}{dt}(t_*) \geq 0$. This is a direct contradiction. Thus, $T(t)$ cannot be the first to leave $[0,1]$.

Boundedness of C(t) Suppose $C(t)$ is the first variable to leave the interval $[0,1]$ through 1. Thus, there exists a t_* such that $C(t_*) = 1$ and $\frac{dC}{dt}(t_*) > 0$, where $0 \leq C(t) \leq 1$ for all $t \leq t_*$. From (4.10c),

$$\begin{aligned}\frac{dC}{dt}(t_*) = &- 1 + (V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2 + V(t_*)^2(V(t_*) - 2)^2 \\ &- (V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2V(t_*)^2(V(t_*) - 2)^2.\end{aligned} \tag{4.21}$$

Applying the assumed inequality $\frac{dC}{dt}(t_*) > 0$ and rearranging gives

$$(V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2[1 - V(t_*)^2(V(t_*) - 2)^2] > 1 - V(t_*)^2(V(t_*) - 2)^2. \tag{4.22}$$

Dividing the inequality by the right-hand side,

$$(V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2 > 1. \tag{4.23}$$

Rewriting the left-hand side in a different form,

$$(V(t_* - \tau_{\text{dw}}) - 1)(V(t_* - \tau_{\text{dw}})^2 - 1) > 1. \tag{4.24}$$

From the assumption that $C(t)$ is the first variable to leave $[0,1]$, the modulus of the two terms on the left-hand side are both less than or equal to one, therefore their product is also less than or equal to one. This contradiction implies $C(t) \leq 1$ for all $t > 0$.

Suppose $C(t)$ is the first variable to leave the interval $[0,1]$ through 0. Thus, there exists a t_* such that $C(t_*) = 0$ and $\frac{dC}{dt}(t_*) < 0$, where $0 \leq C(t) \leq 1$ for all $t \leq t_*$. From (4.10c),

$$\begin{aligned}\frac{dC}{dt}(t_*) = &(V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2 + V(t_*)^2(V(t_*) - 2)^2 \\ &- (V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2V(t_*)^2(V(t_*) - 2)^2.\end{aligned} \tag{4.25}$$

Substituting into $\frac{dC}{dt}(t_*) < 0$ and rearranging, the inequality becomes

$$(V(t_* - \tau_{\text{dw}}) + 1)(V(t_* - \tau_{\text{dw}}) - 1)^2[1 - V(t_*)^2(V(t_*) - 2)^2] \leq -V(t_*)^2(V(t_*) - 2)^2. \tag{4.26}$$

From the boundedness already proven for $V(t)$, the quantity $V(t_*)^2(V(t_*)-2)^2$ is positive and less than one. This implies that the above left-hand side is strictly positive, while the right-hand side is strictly negative, which leads to a contradiction in the inequality. Hence, $C(t) \geq 0$, and $C(t)$ cannot leave $[0,1]$.

In summary, it has been proven for all three variables, V, T, and C, that none can be the first to leave the interval $[0,1]$. From this it follows that the solution $(V(t), C(t), T(t))^T$ must stay in $[0,1]^3$ for all times $t > 0$ if the initial history is in $[0,1]$.

4.3.2 Stability of equilibrium

Let $\vec{x}(t) = (V(t), T(t), C(t))^T$. One equilibrium of (4.10) is $\vec{x}_0 = (0,0,1)^T$. This is the only equilibrium with physical values that is in $[0,1]$. To determine stability of the equilibrium, the Jacobian of the right-hand side of (4.10) must be calculated. For the delayed system, the linearisation has the form

$$\frac{d\vec{x}}{dt} \approx J_0\vec{x}(t) + J_{\tau_{\text{acc}}}\vec{x}(t-\tau_{\text{acc}}) + J_{\tau_{if}}\vec{x}(t-\tau_{\text{if}}) + J_{\tau_{\text{dw}}}\vec{x}(t-\tau_{\text{dw}}), \tag{4.27}$$

where the explicit Jacobian matrices are

$$J_0 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ J_{0,31} & 0 & -1 \end{bmatrix}, \qquad J_{\tau_{\text{acc}}} = \begin{bmatrix} 0 & J_{\tau_{\text{acc}},12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$J_{\tau_{\text{if}}} = \begin{bmatrix} 0 & 0 & 0 \\ J_{\tau_{\text{if}},21} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad J_{\tau_{\text{dw}}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & J_{\tau_{\text{dw}},23} \\ J_{\tau_{\text{dw}},31} & 0 & 0 \end{bmatrix},$$

$$\begin{aligned}
J_{0,31} &= [2V_0(V_0-2)^2 + 2V_0^2(V_0-2)][1 - V_{\tau_{\text{dw}}} + 1)(V_{\tau_{\text{dw}}} - 1)^2], \\
J_{\tau_{\text{acc}},12} &= 2T_{\tau_{\text{acc}}}(T_{\tau_{\text{acc}}} - 2)^2 + 2T_{\tau_{\text{acc}}}^2(T_{\tau_{\text{acc}}} - 2), \\
J_{\tau_{\text{if}},31} &= [(V_{\tau_{\text{if}}} - 1)^2 + 2(V_{\tau_{\text{if}}} + 1)(V_{\tau_{\text{if}}} - 1)](C_{\tau_{\text{dw}}} + 1)(C_{\tau_{\text{dw}}} - 1)^2, \\
J_{\tau_{\text{dw}},23} &= [(C_{\tau_{\text{dw}}} - 1)^2 + 2(C_{\tau_{\text{dw}}} + 1)(C_{\tau_{\text{dw}}} - 1)](V(t-\tau_{\text{if}} + 1)(V(t-\tau_{\text{if}}) - 1)^2, \\
J_{\tau_{\text{dw}},31} &= [(V_{\tau_{\text{dw}}} - 1)^2 + 2(V_{\tau_{\text{dw}}} + 1)(V_{\tau_{\text{dw}}} - 1)][1 + V_0^2(V_0-2)^2].
\end{aligned}$$

For ease of notation, the variable subscripts denote the value of the delay at which the respective variables are evaluated. For instance, $V_{\tau_{\text{acc}}} = V(t-\tau_{\text{acc}})$ and $V_0 = V(t)$. At $\vec{x}_0$ the Jacobian matrices are evaluated for $\vec{x}_0 = \vec{x}_{\tau_{\text{acc}}} = \vec{x}_{\tau_{\text{if}}} = \vec{x}_{\tau_{\text{dw}}}$. This gives

$$J_0 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad J_{\tau_{\text{acc}}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$J_{\tau_{\text{if}}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad J_{\tau_{\text{dw}}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

This implies (4.27) can be reduced to

$$\dot{\vec{x}}(t) \approx J_0\vec{x}(t) + J_{\tau_{\text{dw}}}\vec{x}(t - \tau_{\text{dw}}). \tag{4.28}$$

Let $\vec{x}(t) = \vec{c}\mathrm{e}^{\lambda t}$, where $\vec{c} = (c_1, c_2, c_3)^T$. Then $\vec{x}(t - \tau_{\text{dw}}) = \vec{c}\mathrm{e}^{\lambda(t-\tau_{\text{dw}})}$. The exponential stability of equilibria is determined by the linearisation. Substituting the ansatz into (4.10), the linearisation at $(0, 0, 1)^T$ becomes

$$\epsilon\lambda\vec{c}\mathrm{e}^{\lambda t} = (J_0\mathrm{e}^{\lambda t} + J_{\tau_{\text{dw}}}\mathrm{e}^{\lambda(t-\tau_{\text{dw}})})\vec{c}.$$

Rearranging and dividing through by $\mathrm{e}^{\lambda t}$ gives

$$(\epsilon\lambda I - J_0 - J_{\tau_{\text{dw}}}\mathrm{e}^{-\lambda\tau_{\text{dw}}})\vec{c} = 0.$$

Since $\vec{c} = \vec{0}$ would be a trivial solution, the following will be considered:

$$\det\left(\begin{bmatrix} \epsilon\lambda + 1 & 0 & 0 \\ 0 & \epsilon\lambda + 1 & 0 \\ \mathrm{e}^{-\lambda\tau_{\text{dw}}} & 0 & \epsilon\lambda + 1 \end{bmatrix}\right) = 0.$$

This simplifies to

$$(\epsilon\lambda + 1)^3 = 0,$$

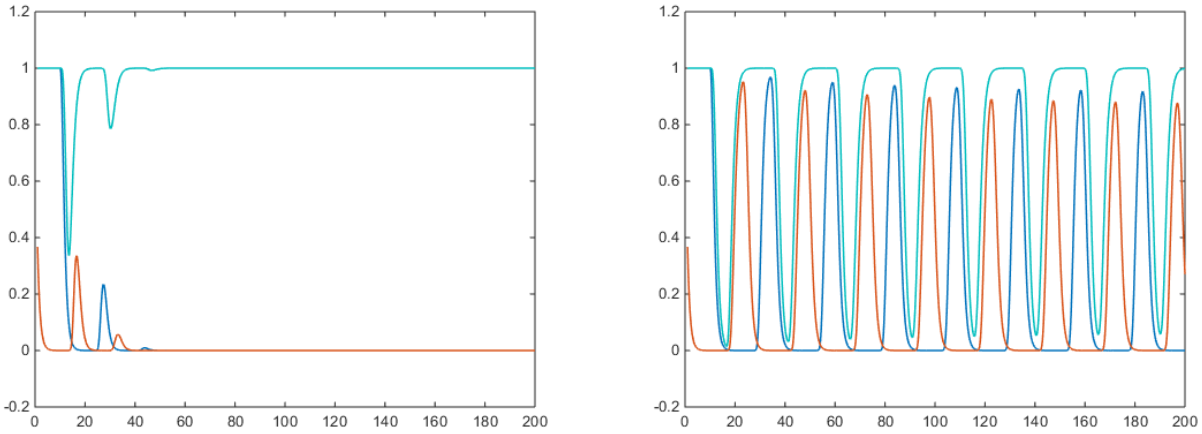
which gives eigenvalues

$$\lambda_{1,2,3} = \frac{-1}{\epsilon}.$$

Therefore, the equilibrium $(0, 0, 1)^T$ is always stable independent of the values of the delays.

4.4 Numerical Analysis of DDE System

As the above equilibrium does not pass through any bifurcations, we turn to numerical methods to find any oscillatory behaviour. In the following analyses, ϵ is taken to equal 1. The aim is to identify the precise onset of periodic orbits based on the relationships between the delays. Since the most uncertain delay value is the overturning time of the ocean in the past, τ_{dw} is varied while $\tau_{\text{if}} = 3$



(a) **(b)**

Figure 4.4: Solution of (4.10) for (a) $\tau_{\text{acc}} = 10$ kyr, $\tau_{\text{if}} = 3$ kyr, $\tau_{\text{dw}} = 2$ kyr and (b) $\tau_{\text{acc}} = 10$ kyr, $\tau_{\text{if}} = 3$ kyr, $\tau_{\text{dw}} = 6$ kyr. The blue curve is V, red is T, and cyan is C. Time (kyr) is along the horizontal axis.

and $\tau_{\text{acc}} = 10$ are kept constant. The values chosen for the latter two were taken from Ghil et al. (1987). We try two different values of τ_{dw} with the initial conditions $(V(s), T(s), C(s)) = (1, 1, 1)$ for $s \in [-\tau_{\text{acc}}, 0]$. For $\tau_{\text{dw}} = 2$ (4.10) settles to the equilibrium $(0, 0, 1)^T$ (Figure 4.4a). As τ_{dw} is increased, however, the long term behaviour becomes oscillatory (see Figure 4.4b for $\tau_{\text{dw}} = 6$).

The periodic orbit solution shown in Figure 4.4b can be tracked through different values of τ_{dw} using the numerical bifurcation software DDE-BIFTOOL (Engelborghs et al., 2001, 2002, Sieber et al., 2014). Through this a fold of periodic orbits is evident (Figure 4.5). The two-parameter bifurcation diagrams show the fold bifurcation's dependence on the other delays (Figure 4.6). For small values of τ_{if}, the fold is nearly independent of this delay value. Once it approaches 5 kyr, though, there is a transition to a nearly linear relationship between τ_{if} and τ_{dw} for larger values of τ_{if}. This is depicted in Figure 4.6a. The opposite is seen for τ_{acc}. The fold is essentially independent of this delay for large values, however for small values it becomes independent of τ_{dw} (see Figure 4.6b).

Because time can be rescaled using the parameter ϵ, larger delay values can be used without them being unrealistic. In this case, only the relationship between τ_{if} and τ_{dw} is important to the onset of periodic orbits. There is then two ways the bifurcation could occur. As proposed in Ghil et al. (1987), the overturning time of the ocean could have been different in the past than present-day likely due to a slow-down of the thermohaline circulation, which would cause a change in τ_{dw}. If a shift in τ_{if} causes the bifurcation, then this implies there was a change in the characteristic time of ice-sheet expansion. This is not a new idea, as a physical explanation for such a change has been previously proposed. The effect of regolith on ice sheet sliding (i.e. basal sliding) was discussed in Ganopolski and Calov (2011), where the authors hypothesised that basal sliding accelerated ice-sheet expansion (lower τ_{if}), but the gradual removal of regolith could increase the time of ice-sheet expansion (higher

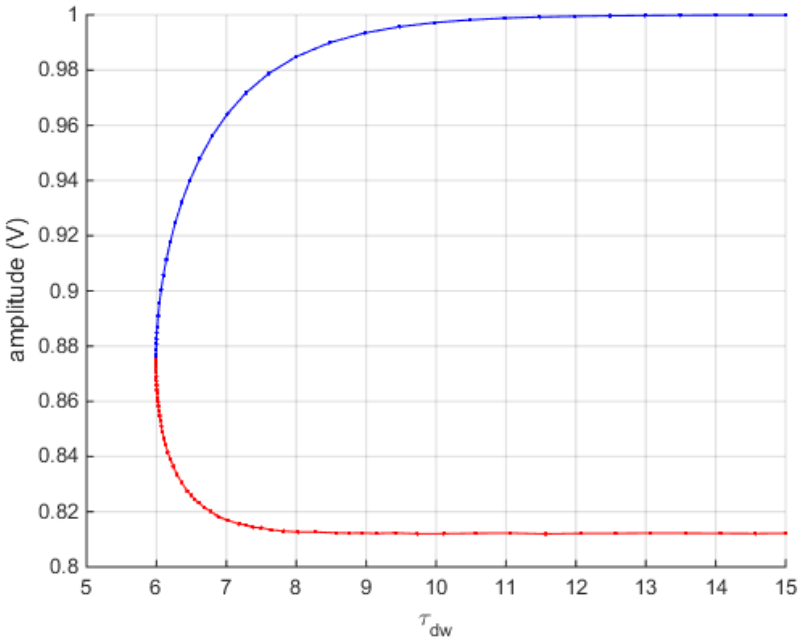
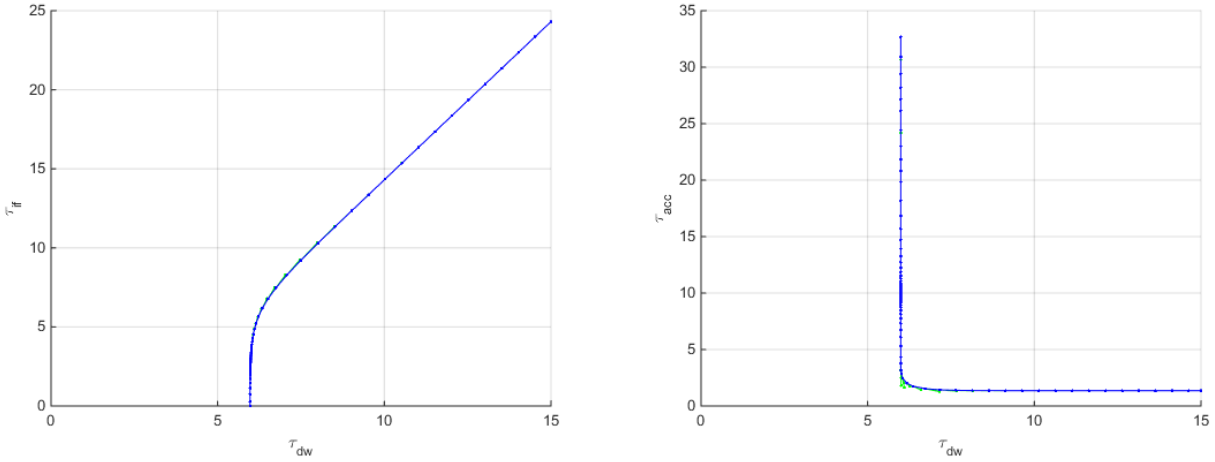


Figure 4.5: Amplitude of periodic orbits as a function of τ_{dw} (kyr). The blue branch represents stable periodic orbits, while the red branch represents unstable.



(a) **(b)**

Figure 4.6: Two-parameter bifurcation diagrams showing fold of periodic orbits for (a) τ_{dw} vs. τ_{if} and (b) τ_{dw} vs. τ_{acc}. All axes are in kyr.

τ_{if}). This could be one physical mechanism to explain a shift in that delay. It has also been suggested through geochemical studies of deep-sea cores that the lag between temperature changes and ice sheet extent changes is anywhere from 1 to 6 kyr (Ruddiman and McIntyre, 1981).

While this model is extremely idealistic, it is a first step towards the inclusion of delays in continuous paleoclimate models. It provides a hypothesis of which delays might have the biggest impact on the dynamical behaviour of a system and thus should be included in models that include similar mechanisms but are closer to the realistic physical mechanisms. This is described in Chapter 5, where a model with delay is derived from the Saltzman and Maasch (1988) model. The more realistic model then allows us to include the effect of forcing on the system, and gives insight on how external forcing (e.g. orbital forcing) could influence oscillations caused by internal feedbacks.

Chapter 5

A conceptual model with delay for the Mid-Pleistocene Transition

In the previous chapter we discussed delay effects which contributed to the glacial and interglacial cycles during the Pleistocene. We introduced a first attempt at developing a model with delay for the climate system during this time. In this chapter we extend upon the ideas of the Ghil et al. (1987) Boolean delay equation model, and apply the linear chain approximation from Chapter 3, Section 3.1 to a previously developed ordinary differential equation (ODE) model for the Pleistocene, the Saltzman and Maasch (1988) model (SM88). This model was originally used to study the Mid-Pleistocene Transition (MPT, see Chapter 2, Section 2.3.2). We first show that the dynamics of SM88 without forcing is qualitatively the same to a scalar delay differential equation (DDE). By exploring realistic values for the delay time, we discover a bistable parameter region of the model which was not previously explored. When subjecting this region to two types of external forcing, periodic and quasiperiodic (astronomical), we observe a distinct transition behaviour. In particular, the quasiperiodic forcing induces a transition similar to the MPT without any change in parameter. We also explore the robustness of this transition when noise is included into the model.

The chapter is organised as follows. Section 5.1 introduces the original SM88 model and details the derivation of our DDE model. Section 5.2 compares the bifurcation diagrams of both models without forcing. Section 5.3 and Section 5.4 study the model subjected to periodic forcing and quasiperiodic forcing, respectively. Finally, Section 5.5 considers the combined effect of quasiperiodic forcing and noise. Sections 5.1, 5.2, 5.4, and 5.5 have been published in:

> C. Quinn, J. Sieber, A. S. von der Heydt, and T. M. Lenton. The mid-pleistocene transition induced by delayed feedback and bistability. *Dynamics and Statistics of the Climate System*, 3 (1):1–18, 2018b. doi: 10.1093/climsys/dzy005

. Section 5.3 is currently under review as the manuscript:

> C. Quinn, J. Sieber, and A. S. von der Heydt. Effects of periodic forcing on a paleoclimate delay model. *preprint arXiv:1808.02310*, 2018a

Any coauthor contributions will be noted where applicable. All numerical results have been obtained by the candidate, apart from the calculation for the unstable manifold of a periodic orbit which was performed by Jan Sieber.

5.1 Introduction of the model with delay

In this section we explain the origin of the delay differential equation model used in our study. Section 5.1.1 recounts the study of Saltzman and Maasch (1988), where a low-dimensional ODE model for the Mid-Pleistocene Transition was developed. Section 5.1.2 explains how a scalar delay differential equation can be derived from that original model.

5.1.1 The Saltzman and Maasch 1988 model

We begin by considering the original model of Saltzman and Maasch (1988), which we will refer to as SM88,

$$\frac{dX}{dt} = -X(t) - Y(t), \tag{5.1a}$$

$$\frac{dY}{dt} = -pZ(t) + rY(t) + sZ(t)^2 - Z(t)^2Y(t), \tag{5.1b}$$

$$\frac{dZ}{dt} = q(-X(t) - Z(t)). \tag{5.1c}$$

Here the variables are non-dimensionalised versions of global land ice mass (X), atmospheric CO_2 (Y), and North Atlantic Deep Water (NADW, Z). More precisely, the variables represent deviations from a background state (anomalies). The first term, $-X(t)$ in (5.1a) combines the feedback mechanisms affecting global ice mass: damping and negative feedback from ablation (melting, evaporation, and loss of ice through icebergs). The second term, $-Y(t)$ in (5.1a) is the direct effect of higher temperatures (caused by higher CO_2 levels) leading to loss of ice mass. In this equation the time scale of t is 10 kyr, which is the characteristic time scale of ablation. In equation (5.1c) the term $-X(t)$ represents the reduction of NADW production with increasing of ice mass - this comes from the fact that ice will begin to form in the shallow ocean shelf near glaciers which is the primary NADW production zone. This equation has a time constant q which is the ratio of ice sheet time constant (10 kyr) to the response time of the deep ocean (between 1 and 6 kyr). The equation for atmospheric CO_2

(5.1b) has a negative term $-pZ$ for the negative effect due to NADW- more NADW leads to more ventilation of the deep ocean (stronger mixing between surface and deep waters) and an increased downdraw of atmospheric CO_2 into the deep ocean. The term rY models the positive feedbacks of atmospheric CO_2 resulting from changes in sea surface temperatures, sea ice extent, and sea level, which outweigh negative feedbacks (such as linear damping). The nonlinearity $(s-Y)Z^2$ represents locally enhanced instabilities in the Southern Ocean due to increased NADW production. The NADW meets colder, denser Antarctic Bottom Water and induces vertical mixing. This brings CO_2-rich water to the surface, reducing downdraw of atmospheric CO_2 into the ocean and potentially releasing CO_2 into the atmosphere depending on the current level of atmospheric CO_2 (Y). Each of these effects (except damping of the nonlinear terms, $-YZ^2$) have an associated parameter which control their relative strengths. All unit factors come from the nondimensionalisation of the system. These were the feedbacks between global ice mass, atmospheric CO_2, and NADW deemed most important by Saltzman and Maasch (1988). Present day studies continue to stress the interaction between these three variables as being essential for the MPT, including a study by Chalk et al. (2017) which uses a geochemical box model to confirm that carbon cycle feedbacks related to ocean circulation are necessary to sustain the long glacial cycles in the late Pleistocene (see Chapter 2, Section 2.3.2 for discussion of the Pleistocene climate record).

5.1.2 Derivation of DDE

In Chapter 3 we explained in detail how a linear chain of first-order ODEs approximates a delay (see Section 3.1). Here we adapt the derivation to our particular system.

First, we employ a change of variables to system (5.1) to clarify the presence of a linear chain. We set $V(t) = -X(t)$, where V is the negative of the global ice mass perturbations, replacing the two linear equations (5.1a) for X and (5.1c) for Z by a chain of linear first-order filters

$$\frac{dV}{dt} = Y(t) - V(t), \qquad \frac{dZ}{dt} = q[V(t) - Z(t)]. \tag{5.2}$$

The 3-dimensional SM88 model (5.1), using (5.2), has a 2-dimensional linear chain and is of the following form:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = F(Y(t), Z(t)), \tag{5.3a}$$
$$\frac{\mathrm{d}V}{\mathrm{d}t} = (Y(t) - V(t)), \tag{5.3b}$$
$$\frac{\mathrm{d}Z}{\mathrm{d}t} = q(V(t) - Z(t)). \tag{5.3c}$$

Extracting the linear chain system we have

$$\frac{\mathrm{d}\vec{y}}{dt} = A\vec{y}(t) + \vec{b}(t), \tag{5.4}$$

where $\vec{y} = [V, Z]^T$. We have A and $\vec{b}(t)$ as follows:

$$A = \begin{bmatrix} -1 & 0 \\ q & -q \end{bmatrix}, \qquad \vec{b}(t) = \begin{bmatrix} y_1(t) \\ 0 \end{bmatrix}. \tag{5.5}$$

Since A is diagonalisable, the solution matrix of (5.4) for the homogeneous system ($\vec{b} = \mathbf{0}$) is then given by

$$\Phi(t) = P\mathrm{e}^{\Lambda t}P^{-1} \tag{5.6}$$

where P is the matrix of eigenvectors associated with eigenmatrix Λ,

$$\Lambda = \begin{bmatrix} -1 & 0 \\ 0 & -q \end{bmatrix}, \tag{5.7}$$

We have the following basis of A and it's inverse,

$$P = \begin{bmatrix} \frac{q-1}{q} & 0 \\ 1 & 1 \end{bmatrix}, \qquad P^{-1} = \begin{bmatrix} \frac{q}{q-1} & 0 \\ \frac{-q}{q-1} & 1 \end{bmatrix}. \tag{5.8}$$

The fundamental solution of (5.4) can be expressed as

$$\vec{y}(t) = \Phi(t)\vec{y}(0) + \int_0^t \Phi(t)\Phi^{-1}(s)\vec{b}(s)\mathrm{d}s. \tag{5.9}$$

Assuming the solution has existed arbitrarily far in the past we can extract an equation for $Z(t)$,

$$Z(t) = \int_{-\infty}^{t} Y(s)K(t-s)ds, \tag{5.10}$$

where the kernel, $K(\tau)$ in (5.10) is

$$K(\tau) = \frac{q}{q-1}\left[\mathrm{e}^{-\tau} - \mathrm{e}^{-q\tau}\right]. \tag{5.11}$$

Note that this is a hypoexponential distribution with mean,

$$\mathbf{E}[\tau] = \int_0^\infty \frac{q}{q-1}\left[\mathrm{e}^{-\tau} - \mathrm{e}^{-q\tau}\right]\tau\mathrm{d}\tau = \frac{q+1}{q}, \tag{5.12}$$

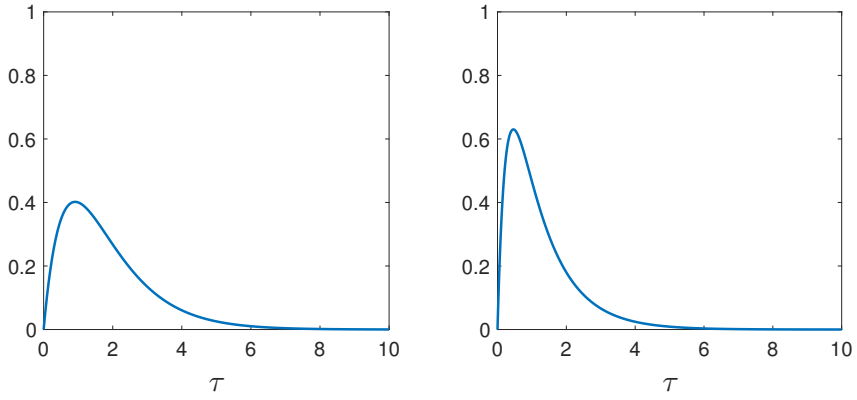


Figure 5.1: Visualisation of the kernel (5.11) for different values of q. Left: $q = 1.2$, $\mathrm{Var}(\tau) \approx 1.69$. Right: $q = 4$, $\mathrm{Var}(\tau) \approx 1.06$.

and variance

$$\mathrm{Var}(\tau) = \int_0^\infty \left(\tau - \frac{q+1}{q}\right) \frac{q}{q-1} \left[\mathrm{e}^{-\tau} - \mathrm{e}^{-q\tau}\right] \mathrm{d}\tau = \frac{q^2+1}{q^2}. \tag{5.13}$$

This allows us to write (5.3a) as a DDE with a distributed delay,

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = F(Y(t), \int_{-\infty}^{t} Y(s)K(t-s)ds). \tag{5.14}$$

Using a change of variables $\tau = t - s$ and approximating the true distributed delay kernel K in (5.10) with a delta distribution at its expected value (5.12) we approximate (5.3) by a DDE with a discrete delay

$$\frac{dY}{dt} = F\Big(Y(t), Y\Big(t - \frac{q+1}{q}\Big)\Big). \tag{5.15}$$

To facilitate the comparison to paleoclimate records, we shift time and exploit that in the formulation as a DDE, the ice mass anomaly $V(t) = -X(t)$ corresponds to $Y(t-1)$ and $Z(t)$ corresponds to $Y(t-\tau)$. Thus, we may express the DDE model in terms of ice mass anomaly X:

$$\frac{dX}{dt} = rX(t) - pX(t-\tau) - X(t-\tau)^2[s + X(t)]. \tag{5.16}$$

In DDE (5.16) one of the parameters r, p, s, τ is redundant and could be removed by a rescaling of X and time. Thus, (5.16) is simpler than the original SM88 model (5.1) in the sense that it has fewer free parameters. For the purpose of model comparison, the DDE system will be left as above.

5.2 Analysis of the DDE model

In this section we show the qualitative similarity between the original SM88 model (5.1) and our scalar DDE (5.16) by comparing their bifurcation diagrams. This involves comparing the locations of equilibria and their respective stability changes. We will then focus on a region in parameter space with bistability that was not previously explored.

5.2.1 Equilibria and Stability

The first property of the scalar DDE model that we can study analytically is the location of equilibria. Since delays don't affect location of equilibria, i.e. $X(t) = X(t-\tau)$, we can look at the solutions of

$$0 = -pX_{\text{eq}} + rX_{\text{eq}} - sX_{\text{eq}}^2 - X_{\text{eq}}^3. \tag{5.17}$$

Here the subscript eq signifies the equilibrium values. There is always the trivial equilibrium $X_{\text{eq},1} = 0$. Dividing (5.17) by X_{eq} for non-zero equilibria, we get a quadratic equation for the other equilibria which we can solve:

$$X_{\text{eq}_{2,3}} = \frac{-s \pm \sqrt{s^2 - 4(p-r)}}{2}. \tag{5.18}$$

These equilibria exist if

$$s^2 > 4(p-r). \tag{5.19}$$

This implies that a fold bifurcation occurs at

$$s^2 = 4(p-r). \tag{5.20}$$

In addition, as the trivial equilibrium always exists, there will be a transcritical bifurcation if

$$0 = \frac{-s \pm \sqrt{s^2 - 4(p-r)}}{2}, \tag{5.21}$$

or simplified,

$$p = r. \tag{5.22}$$

While the location of equilibria does not depend on delay, their stability can. For this we analyse the linearisation of (5.16). In DDE systems with a single delay τ, the linearisation with linearised variable $\hat{X}$ has the form

$$\frac{d\hat{X}}{dt} = J_0\hat{X}(t) + J_\tau\hat{X}(t-\tau) \tag{5.23}$$

where J_0 is the Jacobian with respect to $X(t)$ and J_τ is the Jacobian with respect to $X(t-\tau)$. In our system they are

$$J_0 = r - X_{\text{eq}}^2, \tag{5.24a}$$
$$J_\tau = -p - 2sX_{\text{eq}} - 2X_{\text{eq}}^2. \tag{5.24b}$$

If we consider solutions of the form $X(t) = X_0 \mathrm{e}^{\lambda t}$, from (5.23) we have

$$\lambda X_0 \mathrm{e}^{\lambda t} = [J_0 + J_\tau \mathrm{e}^{-\lambda\tau}] X_0 \mathrm{e}^{\lambda t}. \tag{5.25}$$

Since (5.25) should hold true for nontrivial solutions, i.e. the case where $X_0 \neq 0$, we conclude that a nontrivial solution $\mathrm{e}^{\lambda t}$ exists if λ is a root of the characteristic equation of (5.25):

$$\delta_{\text{DDE}}(\lambda) = \lambda - r + X_{\text{eq}}^2 + (p + 2sX_{\text{eq}} + 2X_{\text{eq}}^2)\mathrm{e}^{-\lambda\tau}. \tag{5.26}$$

Where (5.26) is equal to zero, λ is called an eigenvalue of equilibrium X_{eq}. These eigenvalues cannot be found explicitly, but it can be seen that they in fact depend upon the delay τ. We can then expect the value of τ to affect stability of equilibria. A change in stability will occur when an eigenvalue crosses the imaginary axis under variation of τ. One case of this is a fold bifurcation, when the crossing occurs in $\lambda = 0$. We already identified the parameters for such a bifurcation to occur through equation (5.20). When two eigenvalues cross the imaginary axis simultaneously, a Hopf bifurcation occurs. If a fold and Hopf bifurcation coincide, a codimension 2 bifurcation occurs known as a Bogdanov-Takens (BT) point. At a BT point the derivative of the characteristic equation with respect to λ should also be zero. The derivative of (5.26) is

$$\delta'_{\text{DDE}}(\lambda) = 1 - \tau(p + 2sX_{\text{eq}} + 2X_{\text{eq}}^2)\mathrm{e}^{-\lambda\tau}. \tag{5.27}$$

Considering solutions where the above equation is equal to zero with $\lambda = 0$ (as this is a constraint for a fold bifurcation), we get an equation for τ depending on the other parameters of the system. Combining this with equilibrium equation (5.17) and fold bifurcation equation (5.20) gives the following set of equations to obtain a BT point,

$$\begin{aligned} 0 &= -pX_{\text{eq}} + rX_{\text{eq}} - sX_{\text{eq}}^2 - X_{\text{eq}}^3, \\ \tau &= \frac{1}{p + 2sX_{\text{eq}} + 2X_{\text{eq}}^2}, \\ p &= \frac{s^2}{4} + r. \end{aligned}$$

Here our second parameter of interest is p, but (5.20) can easily be rearranged for any of the other parameters.

5.2.1.1 Comparison to SM88

Let us now check how the local bifurcations of equilibria of this DDE model compare to those of the original SM88. To claim that our model captures the same behaviour that was in the original model, some main features need to agree. In order to have consistency in the meaning of the parameters between the models, we will use $q = \frac{1}{\tau - 1}$.

First we consider the equilibria. The equilibria of SM88 satisfy the following:

$$0 = -X_{\text{eq}} - Y_{\text{eq}}, \tag{5.29a}$$

$$0 = -pZ_{\text{eq}} + rY_{\text{eq}} + sZ_{\text{eq}}^2 - Z_{\text{eq}}^2 Y_{\text{eq}}, \tag{5.29b}$$

$$0 = \frac{1}{\tau - 1}(-X_{\text{eq}} - Z_{\text{eq}}). \tag{5.29c}$$

From (5.29a) and (5.29c) we get the relation

$$Z_{\text{eq}} = -X_{\text{eq}} = Y_{\text{eq}}. \tag{5.30}$$

This allows (5.29b) to be written as

$$0 = -pX_{\text{eq}} + rX_{\text{eq}} - sX_{\text{eq}}^2 - X_{\text{eq}}^3. \tag{5.31}$$

It is clear that (5.31) is identical to (5.17), confirming that the equilibria of X are identical between the two models.

Since the equilibria are equivalent, the fold bifurcations and transcritical bifurcations will occur at the same parameter values. Hopf bifurcations, however, can occur in different places. As before, the stability is determined by the characteristic equation. For SM88 this is

$$\begin{aligned}\delta_{\text{ODE}}(\lambda) = &-\lambda^3 + \left(r - X_{\text{eq}}^2 - 1 - \frac{1}{\tau - 1}\right)\lambda^2 + \left(\left(1 + \frac{1}{\tau - 1}\right)\left(r - X_{\text{eq}}^2\right) - \frac{1}{\tau - 1}\right)\lambda + \\ &+ \frac{1}{\tau - 1}\left(r - X_{\text{eq}}^2 + p + 2sX_{\text{eq}} + 2X_{\text{eq}}^2\right).\end{aligned} \tag{5.32}$$

As (5.32) is different from (5.26), we expect that the Hopf bifurcations will occur at different parameter values.

5.2.2 Numerical bifurcation analysis

Figure 5.2 shows the bifurcation diagrams for varying τ in DDE (5.16) and SM88 model (5.1), respectively, using the system parameter values p, r, and s from SM88. Figure 5.2a was computed using numerical continuation software DDE-BIFTOOL (Engelborghs et al., 2001, 2002, Sieber et al., 2014)

and Figure 5.2b was computed using numerical continuation software COCO (Dankowicz and Schilder, 2013). Although the locations of bifurcations are slightly shifted, qualitatively the two figures agree nicely. In particular, even though the space of possible initial conditions for DDE (5.16) is infinite-dimensional (every possible history on $[-\tau, 0]$ gives a different trajectory), the long-time behaviour of trajectories in the $(X(t), X(t-\tau))$-plane follows a two-dimensional ODE in the range we explored; evidence of this is seen in Figure 5.3- trajectories do not cross inside the limit cycle. The diagrams in Figure 5.2 are partitioned into five main regions of different global behaviour separated by grey vertical lines can be seen in both figures. From left to right the attractors in each region are:

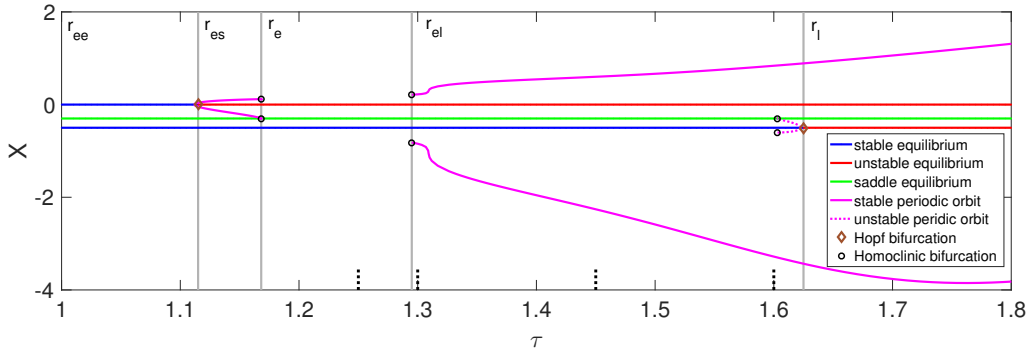
- $[r_{\text{ee}}]$ two stable equilibria,
- $[r_{\text{es}}]$ one stable equilibrium and one stable small-amplitude period solution,
- $[r_{\text{e}}]$ one stable equilibrium,
- $[r_{\text{el}}]$ one stable equilibrium and one stable large-amplitude periodic solution, and
- $[r_{\text{l}}]$ one stable large-amplitude periodic solution.

The region of interest in this study is the bistable region with large-amplitude periodic orbits $[r_{\text{el}}]$. Note that $[r_{\text{el}}]$ can be split further into two sections: one containing unstable periodic orbits and the other not. The effect of the unstable periodic orbits is a drastically reduced basin of attraction for the stable equilibrium. However, since there is no change in the attractors, we will consider these as one region. Region $[r_{\text{el}}]$ was not discussed in a parameter study of the original SM88 model by Maasch and Saltzman (1990). For the remainder of this paper we will discuss only the DDE model, but similar results can be seen for the ODE model.

5.2.3 The bistable region

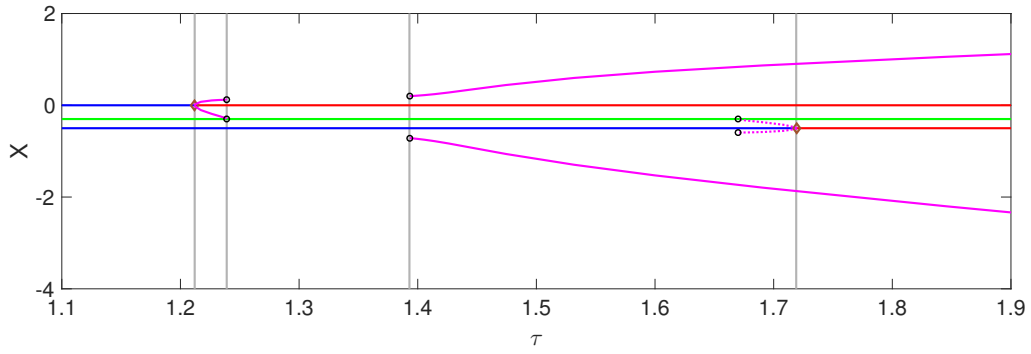
We focus on the bistability seen for $\tau \in [1.295, 1.625]$ in the DDE system. In this region there are two possible stable solutions (shown in Figure 5.3): a stable equilibrium and stable large amplitude periodic orbits. The time profile of the periodic orbits has the asymmetrical shape of the ice age cycles in the late Pleistocene (seen in climate record - Chapter 2, Figure 5.13). The asymmetry is attributed to a slow accumulation of ice mass followed by rapid melting. In addition, the period remains between 109 and 120 kyr throughout the bistable region (with an exception of τ very close to the $[r_{\text{e}}$-$r_{\text{el}}]$ boundary where the period approaches infinity - see Figure 5.4). This cycle length agrees with what is seen in the data (even more so when one adds the external forcing; see Section 2.3.2). We show the persistence of this bistable region in a two-parameter plane in Figure 5.5.

In previous studies of SM88 model (5.1), a transition between two stable states was enforced by a parameter shift through a Hopf bifurcation (Maasch and Saltzman, 1990, Saltzman and Maasch, 1988). We will demonstrate in Section 5.4 that the bistability in region $[r_{\text{el}}]$ makes transitions between



(a) DDE model (5.16)

The dotted black lines indicate values of τ used in Section 2.3.2: $\tau_{\text{ref}} = 1.25$, $\tau_1 = 1.3$, $\tau_2 = 1.45$, $\tau_3 = 1.6$.



(b) SM88 ($\tau = \frac{1}{q} + 1$)

Figure 5.2: Bifurcation diagrams of both models for delay parameter τ. Other parameters: $p = 0.95$, $r = 0.8$, $s = 0.8$.

the two states possible without any change of parameter when the model is subjected to external forcing. We focus on two types of external forcing: periodic and quasiperiodic (astronomical).

5.3 Periodic Forcing

We are interested in the behaviour of this model when a sinusoidal forcing with period 41 kyr is included. This corresponds to the most prominent frequency found in orbital forcing - the obliquity variations, *i.e.* the changes in the angle between the rotational and orbital axes. Thus, we choose

$$F_P(t) = \sin(2\pi t/T). \tag{5.33}$$

The forcing period is $T = 4.1$, corresponding to 41 kyr. Equation (5.33) is then included as a linear forcing with amplitude u,

$$\frac{dX}{dt} = -pX(t-\tau) + rX(t) - sX(t-\tau)^2 - X(t-\tau)^2X(t) - uF_P(t). \tag{5.34}$$

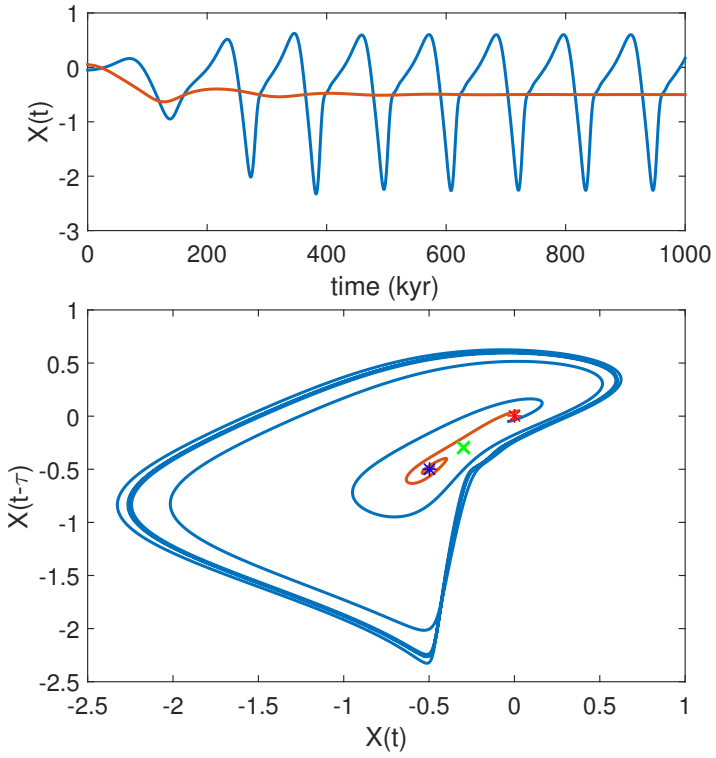


Figure 5.3: Example trajectories in the bistable region $[r_{\mathrm{el}}]$ for $\tau = 1.45$; $X(t) = -0.05$ (blue) and $X(t) = 0.05$ (red) for $t < 0$. Top- time profiles, bottom - phase portrait. Locations of saddle (green cross), stable (blue star) and unstable (red star) equilibria indicated in phase portrait.

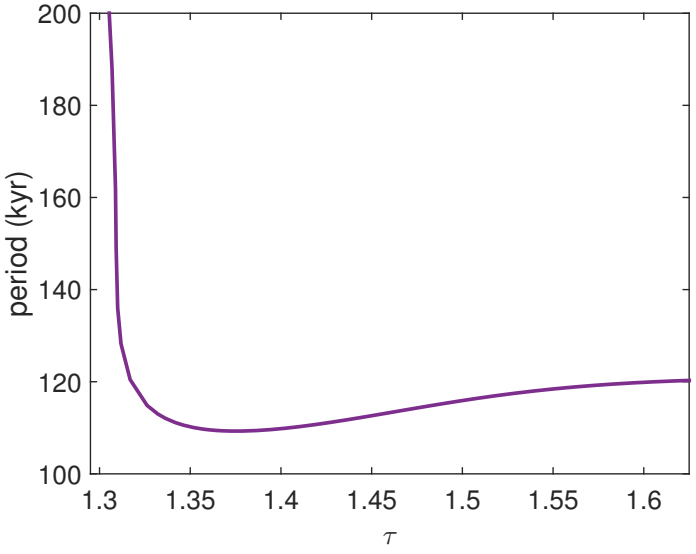


Figure 5.4: Period of the large-amplitude periodic orbit in the bistable region $[r_{\mathrm{el}}]$.

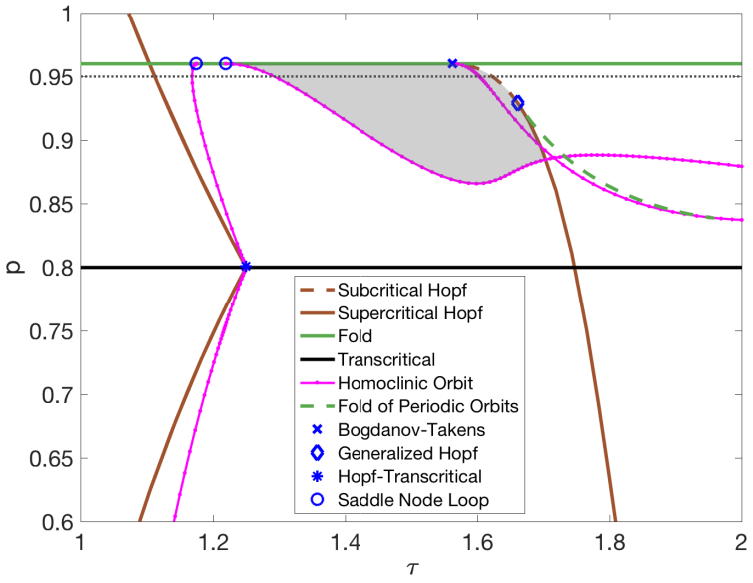
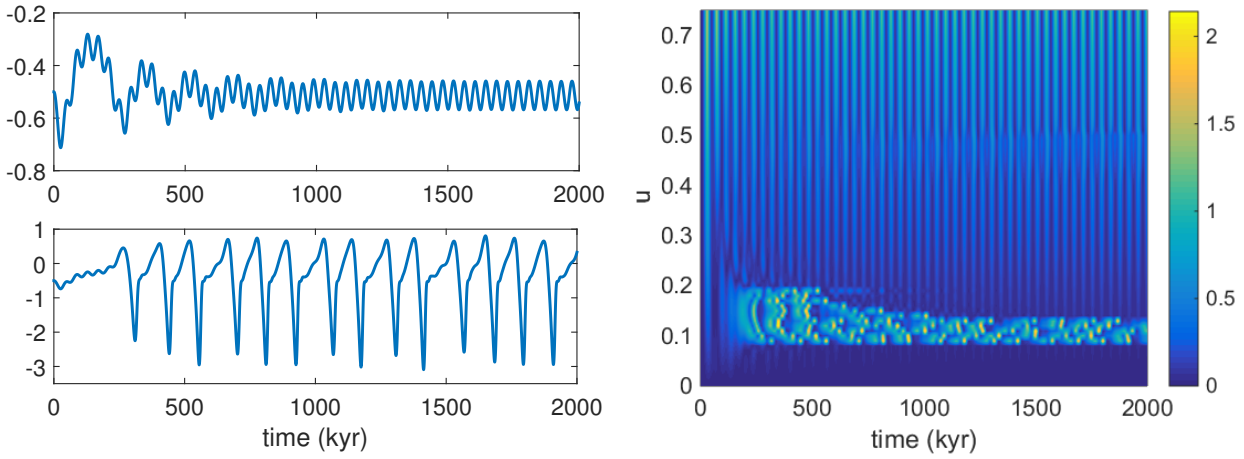


Figure 5.5: Two parameter bifurcation diagram of (5.16) for τ and p. Bistable region is depicted in grey. The black dotted line shows the slice taken for the one-parameter bifurcation diagram.

DDE (5.34) is a dynamical system with the phase space $U = C([-\tau, 0]; \mathbb{R})$, where $C([-\tau, 0]; \mathbb{R})$ is the space of continuous functions on the interval $[-\tau, 0]$ with the maximum norm $\|X\|_0 = \max\{|X(t)| : t \in [-\tau, 0]\}$. At any given time $t \geq 0$, the state is $X_t : [-\tau, 0] \ni s \mapsto X(t + s) \in \mathbb{R}$. For sufficiently small values of u and $\tau \in r_{\text{el}}$, there exist

- a stable small amplitude periodic orbit (with period T), which is a perturbation of the stable equilibrium at $u = 0$, and
- a stable quasiperiodic large amplitude solution, which is a perturbation from the large amplitude periodic orbit at $u = 0$.

These two attractors will persist for a range of u and we will refer to them as the *small-amplitude response* and the *large-amplitude response*, as we did in the case of astronomical forcing. Both types of stable long-time regimes are shown in Figure 5.6a including a transient. We observed in simulations that the large-amplitude response changes from quasiperiodic to chaotic as u increases. Large-amplitude chaotic responses have been observed previously in conceptual ice age models subject to periodic forcing in the literature. Ashwin et al. (2018) find significant regions of chaotic responses for the van der Pol-Duffing oscillator, the Saltzman and Maasch (1991) model, and the Paillard and Parrenin (2004) model. The chaos exists both for simple periodic forcing defined by (5.33) and more complex quasiperiodic forcing. In contrast to our scenario, in Ashwin et al. (2018) all of the models were considered in parameter regions where the unforced dynamics has a single large-amplitude stable periodic orbit. Our simulations suggest that large-amplitude chaotic solutions are also present in a periodically forced bistable regime.



(a) Top: $u = 0.08$, Bottom: $u = 0.09$ **(b)** Distance from unforced stable equilibrium

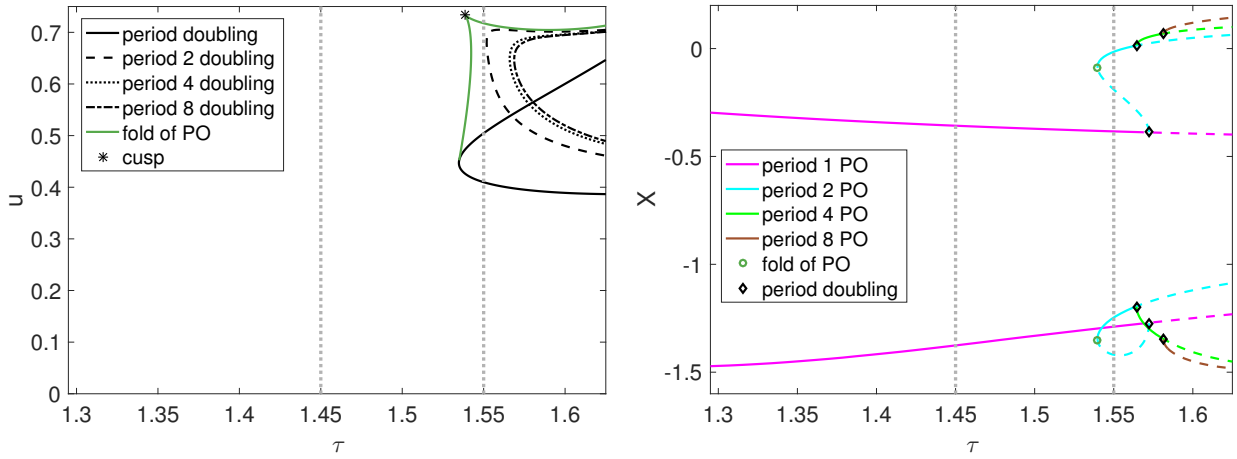
Figure 5.6: (a) Example trajectories of (5.34) for two qualitatively different long-time regimes. Top: small-amplitude response (a small-amplitude periodic orbit), bottom: large-amplitude response (longer time series suggest that it is chaotic). (b) Distance of solution X from $X : s \mapsto -0.5$ (a stable equilibrium of the unforced system) for varying forcing amplitudes u. Other parameters: $\tau = 1.55$, $T = 4.1$, $p = 0.95$, $r = s = 0.8$, $\phi = 0$; initial condition $X(t) = -0.5$ for $t \leq 0$.

The heat map in Figure 5.6b shows the model response over a larger range of forcing amplitudes u. For Figure 5.6b we keep the delay constant at $\tau = 1.55$ and increase u from 0 to 0.75. All trajectories start from the constant initial history $X_0 : [-\tau, 0] \ni s \mapsto -0.5 \in \mathbb{R}$ corresponding to the stable equilibrium of the unforced system. We then compute the distance of $X_t : [-\tau, 0] \ni s \mapsto X(t+s) \in \mathbb{R}$ to X_0, using the mean absolute error (MAE), $\mathrm{MAE}(X_t, X_0) = \frac{1}{\tau} \int_{-\tau}^{0} |X_t(s) - X_0(s)| ds$. Bright colours in Figure 5.6b indicate large distances, corresponding to large amplitude responses. We notice an obvious shift in behaviour between $u = 0.08$ and $u = 0.09$ where the model goes from exhibiting the small-amplitude periodic orbit to following the large-amplitude solution. This lower threshold is similar to the observations when applying non-periodic insolation forcing (compare to Figure 5.12b).

5.3.1 Bifurcation analysis

In order to examine the cause of the shift in behaviour observed in Figure 5.6b, we start a numerical bifurcation analysis of DDE (5.34). The forcing period is kept constant at $T = 4.1$ (corresponding to 41 kyr). We consider forcing amplitudes $u \in [0, 0.75]$ and delays τ in the bistable region $\tau \in r_{\mathrm{el}} \approx [1.295, 1.625]$ of the unforced system ($u = 0$).

Figure 5.7a shows the bifurcations of the small-amplitude periodic orbit. Bifurcations only occur for $\tau > 1.53$ and $u > 0.38$. For a range of $u > 0.4$ there exists a cascade of period doubling bifurcations for increasing τ, evidence of which is also visible in Figure 5.6b. Figure 5.7b shows a cross section of the two-parameter bifurcation diagram Figure 5.7a along the horizontal line $u = 0.55$ displaying the maximum and minimum of the periodic orbits on the y-axis. We observe that the small-amplitude



(a) Bifurcation diagram for varying τ and u.

(b) Cross section for $u = 0.55$.

Figure 5.7: Bifurcations of small-amplitude periodic orbit. Dotted vertical lines indicate values of τ used in Section 5.3.2, Section 5.3.3, and Section 5.3.4; other parameters: $T = 4.1$, $\tau = 1.55$, $p = 0.95$, $r = s = 0.8$.

motion does *not* experience any bifurcation for $u < 0.3$. Moreover, because the bifurcations are restricted large values of τ, they cannot be used to explain any transitions in time. Therefore periodic forcing, even with a time-dependent modulated amplitude, is not sufficient to induce an MPT-like transition.

The large amplitude solution also goes through some bifurcations. We do not show a detailed bifurcation analysis, but evidence of the collapse of the large amplitude solution can be seen in Figure 5.6b. These large responses are stable in a range of forcing amplitudes $u \in [0.09, 0.15]$ in Figure 5.6b. For $u \in [0.15, 0.2]$ the trajectories make transient large-amplitude excursions before converging to a small-amplitude periodic orbit, which suggests a collapse of the (then chaotic) large-amplitude attractor.

As Figure 5.7a establishes, the observed transition in Figure 5.6b from small- to large-amplitude oscillations at $u = 0.09$ must have been caused by some phenomenon other than a bifurcation. Since the unforced system is bistable for $\tau \in r_{\text{el}}$, we expect this bistability to persist for small forcing amplitudes u. Thus, the initial condition may cross from the basin of attraction of the small-amplitude periodic orbit to the basin of attraction of the large-amplitude response. Both example trajectories in Figure 5.6a started from the same initial condition but were computed with slightly different forcing amplitude ($u = 0.08$ and $u = 0.09$). Figure 5.6b suggests that the constant initial condition $X_0 : s \mapsto -0.5$ leaves the basin of attraction of the small-amplitude periodic orbit at the lower threshold $u \approx 0.09$.

5.3.2 Dynamics on a two-dimensional slow manifold

Since DDEs are infinite-dimensional, it is not feasible to determine the basin of attraction in all dimensions. However, previous studies have proven results of Poincaré-Bendixson type (there exists a plane in $\mathbb{R}^2$ such that trajectories cannot cross each other) for scalar DDEs with monotone feedback (Mallet-Paret and Sell, 1996, Mallet-Paret and Smith, 1990). These are DDEs of the form

$$\frac{dx}{dt} = f(x(t), x(t-\tau)), \quad \text{where} \quad f(0,0) = 0 \text{ and } \delta y\, f(0,y) > 0 \text{ for all } y \neq 0,\ \delta \in \{-1, 1\}, \tag{5.35}$$

($\delta = 1$ corresponds to positive delayed feedback, $\delta = -1$ corresponds to negative delayed feedback). The right-hand side $f(X(t), X(t-\tau))$ in (5.16) does not satisfy the feedback conditions in (5.35) since, for our right-hand side f, $f(0,y) = -py - sy^2$ changes sign also at $y = -p/s = -1.1875$, which is reached in the unforced large-amplitude periodic orbit (see Figure 5.2a). However, the phase portrait in Figure 5.3 suggests that the unforced DDE (5.34) (with $u = 0$) has an attracting two-dimensional slow manifold. We expect this manifold to persist for small forcing amplitudes u. Within this persistent slow manifold the time-T map of the forced DDE (5.34) is a locally invertible two-dimensional map. For two-dimensional maps the basin of attraction for a periodic orbit is often bounded by the stable manifold of a saddle periodic orbit. We can investigate the basin of attraction directly by using an implicit computational dimension reduction introduced by Kevrekidis *et al* as *equation-free methods* (see reviews Kevrekidis and Samaey (2009, 2010)). In their notation one needs to define the following operators. Recall that $U = C([-\tau, 0]; \mathbb{R})$ is the phase space of the DDE (5.34).

$$\textit{lifting} \qquad L : \mathbb{R}^2 \ni (x_1, x_2) \mapsto (y_0, \tilde{y}) \in \mathbb{R} \times \mathbb{L}^\infty([-\tau, 0]; \mathbb{R}), \tag{5.36}$$
$$\text{where } y_0 = x_1 \text{ and } \tilde{y}(t) = x_2 \text{ for } t \in [-\tau, 0],$$
$$\textit{evolution} \text{ map} \qquad M : [0,\infty) \times [0,\infty) \times \operatorname{rg} L \ni (t_0 + t, t_0, (y_0, \tilde{y})) = X_{t_0} \mapsto X_{t+t_0} \in \operatorname{rg} L, \tag{5.37}$$
$$\textit{restriction} \qquad R : U \ni X \mapsto (X(0), X(-\tau))^T \in \mathbb{R}^2. \tag{5.38}$$

The space $\mathbb{R} \times \mathbb{L}^\infty([-\tau, 0]; \mathbb{R})$ (called $U^{\odot,*}$ in the terminology of Diekmann et al. (1995)) is a natural extension of the phase space $U = C([-\tau, 0]; \mathbb{R})$ of the DDE (5.34). The trajectories of the DDE (5.34) starting from initial values in $U^{\odot,*}$ admit discontinuous essentially bounded initial history segments $\tilde{y} : [-\tau, 0] \mapsto \mathbb{R}$ and have y_0 as the right-side limit for $t \searrow 0$. For an element $(y_0, \tilde{y})$ of $U^{\odot,*}$, y_0 is usually called the *head point*, while $\tilde{y}$ is the *history segment*. The notation $\mathbb{L}^\infty([-\tau, 0]; \mathbb{R})$ refers to the space of essentially bounded functions on $[-\tau, 0]$ with essential maximum norm $\|\tilde{y}\|_0 = \inf\{m \geq 0 : \text{Leb}\{t \in [-\tau, 0] : |\tilde{y}(t)| \geq m\} = 0\}$ (Leb A is the Lebesgue measure of a set $A \subset \mathbb{R}$). As explained in the textbook Diekmann et al. (1995) trajectories starting from an element in the larger space $U^{\odot,*}$ return to the smaller phase space U after time τ and the dependence $U^{\odot,*} \ni X_0 \mapsto X_t \in U$ of the solution on its initial condition is as regular as the right-hand side of the

DDE (5.34). Thus,

$$U^{\odot,*} \ni X \mapsto M(t, t_0; X) \in U$$

is smooth for all $t \geq \tau$ and $t_0 \in \mathbb{R}$.

In our computations we approximate elements $(y_0, \tilde{y}) \in U^{\odot,*}$ by vectors $Y \in \mathbb{R}^N$, where Y_k is an approximation of $\tilde{y}(-\tau(N-k)/(N-1))$ for $1 \leq k < N$ and the head point is $Y_N = y_0$. We use the discretised map M based on the Euler-Heun approximation with $h = 0.01$, where a single step has the form

$$\begin{aligned} M(h,t,\cdot) : \mathbb{R}^N \ni Y \mapsto (Y_2, \ldots, Y_N, Y_N + h(f_0 + f_E)/2)^T \in \mathbb{R}^N, & \qquad \text{where} \\ f_0 = f(t, Y_N, Y_1, u), \quad Y_N^0 = Y_N + h f_0, \quad f_E = f(t+h, Y_N^0, Y_2, u), & \qquad \text{and} \\ f(t, x_1, x_2, u) = -p x_2 + r x_1 - s x_2^2 - x_2^2 x_1 - u F_P(t) & \end{aligned}$$

is the right-hand side of the DDE (5.34). For larger time spans we apply the composition rule $M(t+s, r, \cdot) = M(t, s+r, \cdot) \circ M(s, r, \cdot)$ for $s, t \geq 0$, such that the discretisation using $N - 1 = s/h$ steps converges to the continuous map $M(t+s, s, \cdot)$ uniformly for bounded $t \geq \tau$ and bounded initial values in $U^{\odot,*}$. We restrict ourselves to stroboscopic maps $M(t+s, s, \cdot)$, where t is a multiple of the period: $t = kT$ with $k \in \mathbb{Z}$ and $T = 4.1$, such that we may write

$$M^k Y = M(kT, 0, Y)$$

for integers $k \geq 0$. The map $M^k : \operatorname{rg} L \to \operatorname{dom} R$ ($\operatorname{rg} L$ is the range of L, $\operatorname{dom} R$ is the domain of definition of R) is autonomous and smooth, since $M((k+j)T, jT, \cdot) = M(kT, 0, \cdot)$ for all integers $k \geq 0$ and j, and periodic forcing with period T. With this notation, M^{k+j} equals $M^k M^j$. Compatible with the discretisation of M, the discretisations of lifting and restriction are

$$\begin{aligned} L &: \mathbb{R}^2 \ni (x_1, x_2)^T \mapsto (x_2, \ldots, x_2, x_1)^T \in \mathbb{R}^N, \\ R &: \mathbb{R}^N \ni Y \mapsto (Y_N, Y_1)^T \in \mathbb{R}^2. \end{aligned}$$

Figure 5.8a shows that the linearisation of the map M^1 has a spectral gap after the first two eigenvalues such that ∂M^1 is a small perturbation of a rank 2 matrix for all y in a neighbourhood of $L([-0.75, 0.15] \times [-0.75, 0.15])$. This is numerical evidence for the suspected time scale separation leading to a two-dimensional slow manifold. We do not need to construct the slow manifold explicitly, but rather may construct an approximate two-dimensional map $\mathcal{M}_\ell$ from the slow manifold back to itself implicitly, using the coordinates in $\operatorname{dom} L$:

$$\mathcal{M}_\ell : \operatorname{dom} L \ni x \mapsto y \in \operatorname{dom} L, \quad \text{where } y \text{ is the solution of } \quad R M^{\ell+1} L x = R M^\ell L y. \tag{5.39}$$

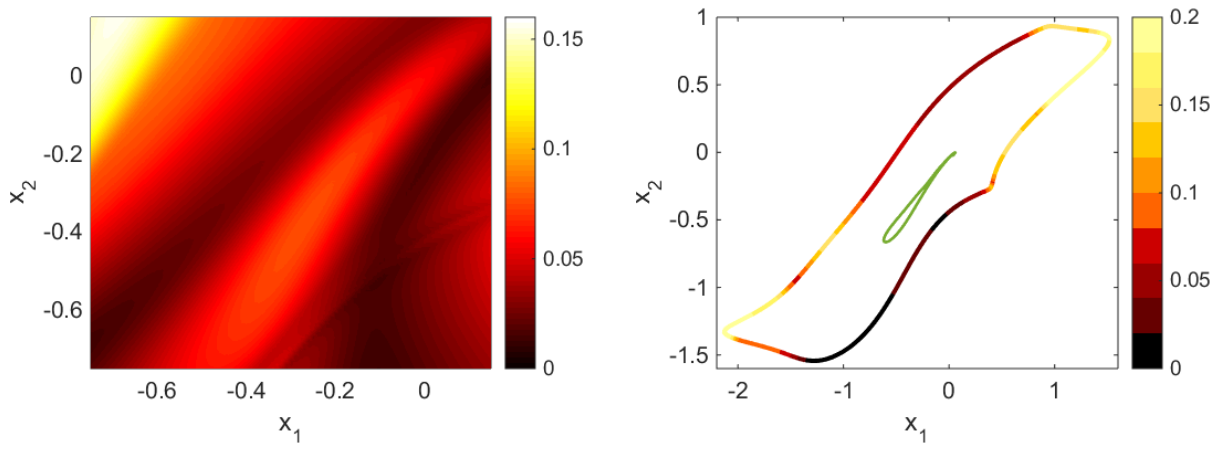


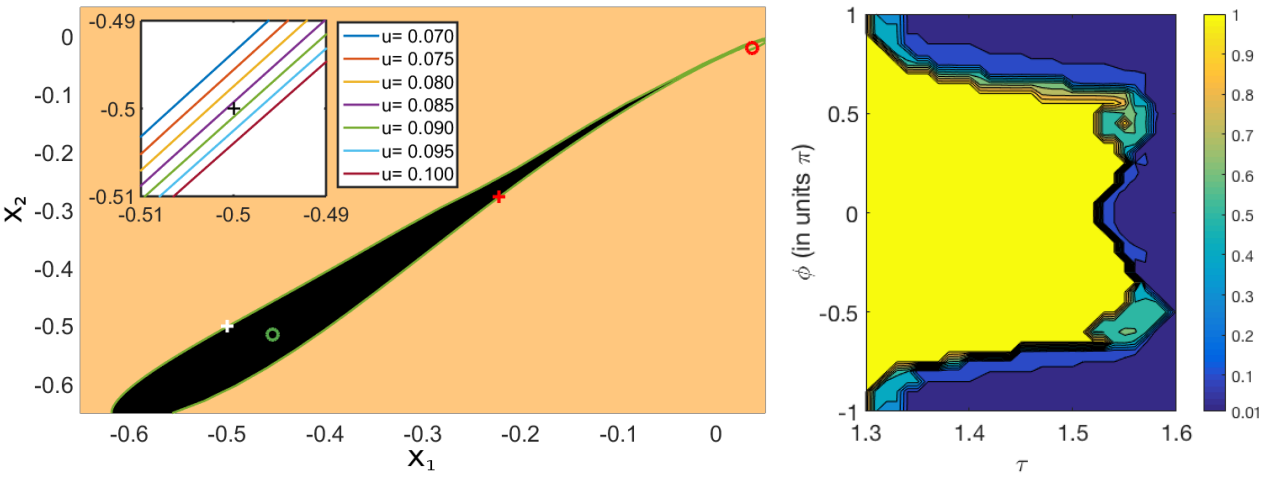
Figure 5.8: Spectral gap of linearisation of M^1 (a) on the domain $L([-0.75, 0.15] \times [-0.75, 0.15])$ and (b) on the boundary for which the Jacobian of RM^1L becomes singular. Colour indicates the ratio between the third and second largest singular values of ∂M^1. The green boundary in figure (b) depicts the basin discussed in Section 5.3.3 and Section 5.3.4. Parameters: $u = 0.09$, $T = 4.1$, $\tau = 1.55$, $p = 0.95$, $r = s = 0.8$, $\phi = 0$.

The integer ℓ is the *healing time* in the notation of Kevrekidis and Samaey (2009, 2010). The map $\mathcal{M}_\ell$ approximates the true stroboscopic map generated by the DDE (5.34) on the slow manifold (Sieber et al., 2017). The approximation improves for increasing healing time ℓ if lifting L and restriction R satisfy some genericity conditions (implying that the map R is a diffeomorphism between the slow manifold and $\mathbb{R}^2$, and that $RM^\ell L : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism). The convergence result in Sieber et al. (2017) does not require a large separation of time scale, only a sufficiently large healing time. In our case $\ell = 1$ (a healing time of one period $T = 4.1$) is sufficient: the results only change by less than 10^{-2} when increasing ℓ to 2 (a large ℓ increases the condition number of $\partial[RM^\ell L]$).

Figure 5.8a justifies using a planar rectangle in $\operatorname{dom} L$ to visualise the basins of attraction in the slow manifold. Furthermore, since we can evaluate the stroboscopic map on the two-dimensional slow manifold by using $\mathcal{M}_\ell$, we can employ algorithms designed for the computation of stable manifolds of fixed points in planar maps. In particular, we continue all three fixed points present for zero forcing ($u = 0$; see Figure 5.2a) in the parameter u, using the defining equation

$$RM^\ell L x_{\text{fix}} = RM^{\ell+1} L x_{\text{fix}},$$

which is a system of two equations for the two-dimensional variable x_{fix} and the parameter u (results will be shown for healing time $\ell = 1$). One of the fixed points is of saddle type. We compute the stable manifold of the saddle fixed point using the algorithm for maps that are not globally invertible, proposed by England et al. (2004) and originally implemented for two-dimensional maps in `DsTool`. Since the map $\mathcal{M}_\ell$ is implicitly defined, the algorithm as originally implemented would require the solution of the nonlinear system (5.39) every time the map gets evaluated. This turns out not to be



(a) $\tau = 1.55$ **(b)**

Figure 5.9: (a) Basin of attraction for the small amplitude stable periodic orbit for $\tau = 1.55$. Initial conditions in the black regions are attracted to the small amplitude stable periodic orbit (green circle) intersected with $\operatorname{dom} L$. The white cross indicates the initial condition $(x_1, x_2) = (-0.5, -0.5)$ used to create Figure 5.6b. The red circle is the unstable small amplitude periodic orbit, while the red cross represents the saddle periodic orbit. Figure zoom in top left shows stable manifold for different values of u close to the initial condition $(x_1, x_2) = (-0.5, -0.5)$ (black cross). (b) Threshold values for u at which a transition to the large-amplitude response is observed as a function of delay τ and phase shift ϕ. Values for which no transitions were observed are shown with $u = 1$. Other parameters for both figures: $T = 4.1$, $p = 0.95$, $r = s = 0.8$; initial condition $X_0 = -0.5$.

necessary: with a small modification the stable manifold algorithm does not require any solution of a nonlinear system (see Appendix A for a brief explanation). The stable manifold of the saddle fixed point will determine the basin of attraction for the stable fixed point (the other fixed point is a source for all u) on the slow manifold.

The large amplitude response (attracting initial conditions on the other side of the stable manifold) is partially outside of the domain of validity of the coordinates introduced by lifting L and restriction R. Figure 5.8b shows the curve in the $\operatorname{dom} L$ plane along which the Jacobian of RM^1L becomes singular, which violates one of the assumptions made in the implicit definition (5.39) of $\mathcal{M}_1$. Outside of this curve (where the large amplitude solution lies), our chosen plane is no longer valid. However, the spectral gap values indicate that the dynamics may still be confined to a two-dimensional slow manifold.

5.3.3 Basins of attraction and stable manifold in the plane

Figure Figure 5.9a shows the basins in the rectangle $[-0.65, 0.05] \times [-0.65, 0.05]$ for $u = 0.09$.

The constant history Lx with $x_1 = x_2 = -0.5$, corresponding to an initial head point y_0 and initial history $\tilde{y}$ both equal to -0.5, was used in the parameter scan for increasing forcing amplitude u,

depicted in Figure 5.6b. This point is indicated by a white cross in Figure 5.9a. Black regions in Figure 5.9a are initial values in $\operatorname{dom} L$ that converge to the stable small-amplitude periodic orbit in the center of the black region. The beige region contains initial conditions that escape to the large amplitude response. The saddle fixed point is located near $(-0.2, -0.3)$. Its stable manifold (in green) is the boundary between the two basins of attraction. The inset in Figure 5.9a shows how the stable manifold of the saddle fixed point changes as the forcing amplitude u increases. In particular, we observe how the initial condition $(-0.5, -0.5)^T$ is crossed by the stable manifold, which shifts downward as u increases.

5.3.4 Dependence on forcing phase

As the basin of attraction in Figure 5.9a shows, the critical forcing amplitude $u = 0.09$ for the transition depends strongly on the initial condition, which we chose as $X_0 = -0.5$ for the heat map in Figure 5.6b. A simple alternative way to illustrate the importance of the initial condition is to vary the phase of the periodic forcing. We adjust the forcing equation accordingly,

$$F_P(t) = \sin((2\pi/T)\, t - \phi), \quad \phi \in [-\pi, \pi]. \tag{5.40}$$

The variable ϕ represents the phase shift of the forcing. Note that the bifurcation diagram Figure 5.7a is independent of the forcing phase ϕ. However, Figure 5.9b shows that the phase affects the threshold value for the forcing amplitude u at which a transition to the large-amplitude response occurs for initial value $X_0 = -0.5$. Figure 5.9b shows contours of the smallest value of u for which we observe a transition to large-amplitude response in simulations for different forcing phases ϕ and delays τ in the bistable region r_{el} of Figure 5.2a. For all points in Figure 5.9b we chose the initial value $X_0 = -0.5$. For some parameter combinations the response is always small amplitude. In these points we set the contour level to its maximum ($u = 1$). A distinct boundary can be seen between parameter combinations that exhibit transitions at low values of u and those that do not. For a forcing phase $\phi = \pi$, a transition can always occur within the bistable region.

In Appendix B we show the effect of a phase shift on the basin of attraction for the small-amplitude periodic orbit in the plane $\operatorname{dom} L$. This implies that, for some phases ϕ, trajectories from the initial history $X_0 = -0.5$ will converge to the small amplitude periodic orbit, while for other phases trajectories starting from $X_0 = -0.5$ will converge to the large amplitude response.

5.4 Astronomical Forcing

In this section we show that when adding astronomical forcing to model (5.16), a transition typically occurs at the same time as the MPT is seen in recorded data without further parameter tuning. In

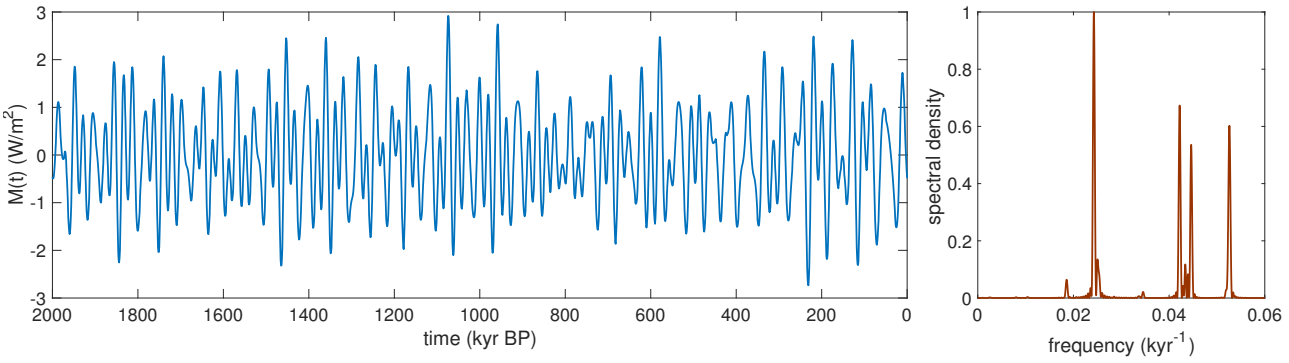


Figure 5.10: Normalised integrated July insolation at 65°N adapted from Huybers and Eisenman (2006) (left), and the corresponding spectrum (right).

Section 5.4.1 we describe the astronomical forcing considered and how we include it in the model. We then examine the responses for different delays and different forcing strengths in Section 5.4.2 and Section 5.4.3.

Hays et al. (1976) have provided evidence that the glacial cycles during the Pleistocene are driven primarily by variations in the earth's orbital cycle. This includes changes in precession (orientation of the rotational axis), obliquity (angle between the rotational axis and orbital axis), and eccentricity (orbital ellipse's deviation from a circle). These modes vary approximately periodically, with cycle lengths of 19/23 kyr, 41 kyr, and 100 kyr respectively (Berger, 1978, Huybers and Eisenman, 2006, Milankovitch, 1941). Figure 5.10 shows a time series of average daily summer insolation at 65°N computed by Huybers and Eisenman (2006) based on the model introduced by Huybers (2006).

5.4.1 Extraction of integrated summer insolation forcing $M(t)$ from data

To investigate the effect of this forcing on DDE (5.16), we include the astronomical forcing in the same way as in Maasch and Saltzman (1990). We add forcing signal $M(t)$ shown in Figure 5.10 with negative amplitude u,

$$\frac{dX}{dt} = rX(t) - pX(t-\tau) - X(t-\tau)^2[s + X(t)] - uM(t). \tag{5.41}$$

The precise procedure for extracting $M(t)$ from the publicly available data source is explained as follows.

We use the Integrated Summer Insolation Calculations data set provided by Huybers and Eisenman (2006) (found at `https://www1.ncdc.noaa.gov/pub/data/paleo/climate_forcing/orbital_variations/huybers2006insolation/j_65north.txt`). The data is provided as daily average summer energy in GJ/m^2 calculated from number of days the insolation was above a given threshold (W/m^2). We used the data provided for the latitude of 65° North. Rather than selecting a particular

threshold, we took the average of all the thresholds. This has the effect of giving a linear weighting to the threshold reached in a day: days that reached higher insolation levels are given a proportionally larger weight. The precise expression is as follows,

$$I_{\text{agg}}(t) = \sum_{i=1}^{25} I_i(t). \tag{5.42}$$

Here $I_i(t)$ is the average daily summer insolation calculated by Huybers and Eisenman (2006) for the threshold $25(i-1)$ in year t (column number $i+1$ in the data file). We then normalise this data through

$$M(t) = \frac{I_{\text{agg}}(t) - \text{mean}_t\, I_{\text{agg}}}{\text{std}_t\, I_{\text{agg}}}, \tag{5.43}$$

where $\text{mean}_t\, I_{\text{agg}}$ and $\text{std}_t\, I_{\text{agg}}$ are the mean and standard deviation of $I_{\text{agg}}(t)$ over all times t. The timeseries of (5.43) is depicted in Figure 5.10. Its spectrum is shown in Figure 5.10. The dominant signal of 0.0242 kyr^{-1} (period $\approx$ 41 kyr) comes from the obliquity variations. The other dominant signals of 0.0422, 0.0446, and 0.0526 kyr^{-1} (periods around 23 and 19 kyr) are a result of changes in precession. The primary eccentricity variations (frequencies around 0.01 kyr^{-1}) are negligible in this forcing. Dueto its composition from several frequencies the forcing is called quasiperiodic in the literature (Crucifix, 2012).

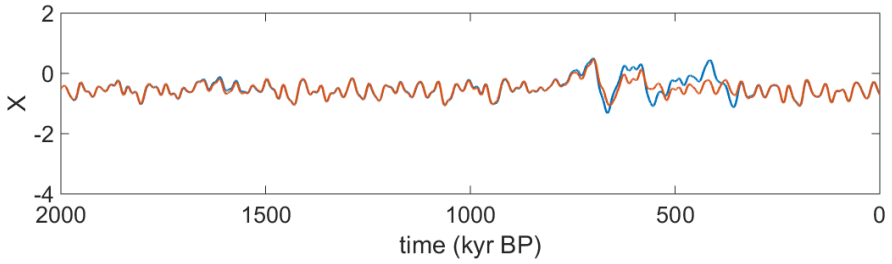
We are interested in how the bistable region responds to this external forcing in dependence of two parameters: the delay τ and the forcing amplitude u.

For small u the system is expected to exhibit two types of responses to forcing in the bistable region. Each of them is a perturbation of an attractor of the unforced system, namely the equilibrium and the large-amplitude periodic orbit, which persist for small u. We will refer to these responses as the *small-amplitude* and the *large-amplitude* response (compare red and blue time profiles in Figure 5.11c). For increasing u we expect to observe increasingly frequent transitions between these responses. For large forcing amplitudes u the internal dynamics of the model will be dominated by the forcing.

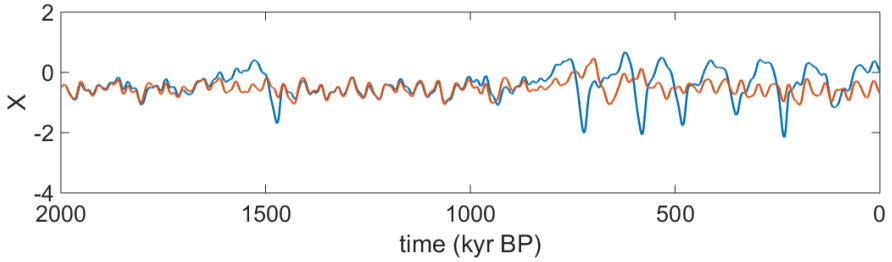
5.4.2 Responses for different delays in the bistable region

We choose a moderate value of the forcing amplitude $u = 0.25$ and investigate the response at four values of τ, labelled in Figure 5.2a by τ_{ref}, τ_1, τ_2, and τ_3. The first value $\tau_{\text{ref}} = 1.25$ is outside of the region of bistability and is used as a reference trajectory to which the solution trajectories for the other delays τ_1, τ_2 and τ_3 are compared. The results are shown in Figure 5.11 (response for τ_{ref} shown in red, for other delays in blue)

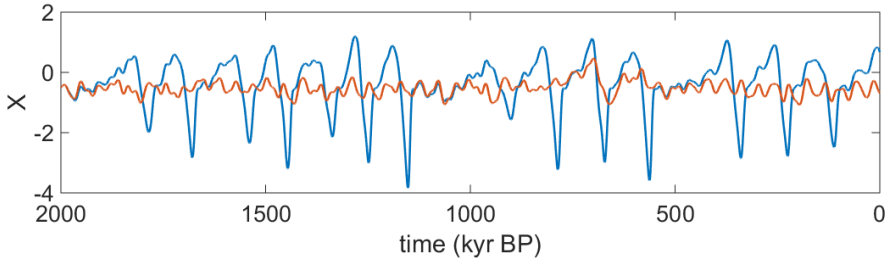
The first comparison is made close to the boundary $[r_{\text{e}}]$–$[r_{\text{el}}]$ at $\tau_1 = 1.3$. We see in Figure 5.11a that both trajectories change in synchrony, exhibiting only the small-amplitude response. In the middle of



(a) $\tau = \tau_1$ (blue), $\tau = \tau_{\text{ref}}$ (red)



(b) $\tau = \tau_2$ (blue), $\tau = \tau_{\text{ref}}$ (red)



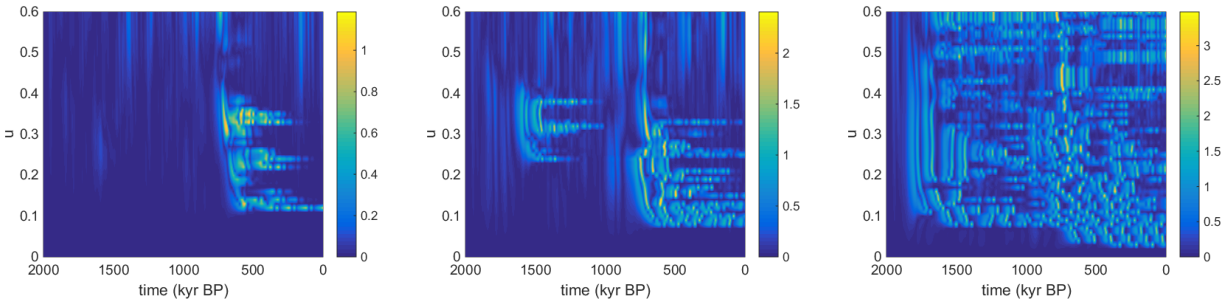
(c) $\tau = \tau_3$ (blue), $\tau = \tau_{\text{ref}}$ (red)

Figure 5.11: Trajectories in bistable region (blue) compared to a reference trajectory at $\tau_{\text{ref}} = 1.25$ (red) for delays (a) $\tau_1 = 1.30$, close to homoclinic connection, (b) $\tau_2 = 1.45$, middle of bistable region, and (c) $\tau_3 = 1.60$, close to Hopf bifurcation. Values of τ are indicated on bifurcation diagram in Figure 5.2a. Other parameters: $p = 0.95$, $r = 0.8$, $s = 0.8$, and $u = 0.25$. Initial condition $X(t) = -0.5$ for $t \in [2 + \tau, 2]$Myr BP in all cases.

the bistable region ($\tau_2 = 1.45$, shown in Figure 5.11b), the solution shows a small-amplitude response in most of the first half of the time window and a large-amplitude response in the second half. Close to the Hopf bifurcation of the autonomous stable equilibrium ($\tau_3 = 1.6$, shown in Figure 5.11c) the solution exhibits primarily a large-amplitude response. This is expected due to the weakening attraction of the autonomous stable equilibrium and its shrinking basin of attraction. We will demonstrate in the following section that these transitions generate dynamics with time profiles that are qualitatively similar to the records of the MPT.

5.4.3 Variable forcing strength

For the exploration of the effect of different forcing strengths, we take the same values for τ as in Figure 5.11, covering the range of the bistable region. For each τ we compute trajectories for different u, ranging from $u = 0$ to $u = 0.6$, with the initial history $X(t) = -0.5$ for $t \in [2 + \tau, 2]$Myr BP.



(a) **(b)** **(c)**

Figure 5.12: Distance from reference solution at $\tau_{\text{ref}} = 1.25$, for different forcing strengths u. Averages taken over window length of size τ. (a) $\tau = 1.30$, close to homoclinic connection. (b) $\tau = 1.45$, middle of bistable region. (c) $\tau = 1.60$, close to Hopf bifurcation. Initial condition $X(t) = -0.5$ for $t \in [2 + \tau, 2]$Myr BP in all cases.

Figure 5.12 shows the difference of these trajectories to a respective reference trajectory for $\tau_{\text{ref}} = 1.25$ and the same initial condition and forcing strength u, averaged over a window of length τ. We note that differences to a reference trajectory at $u = 0$ and identical delay τ (using same initial condition) give qualitatively similar results.

Remarkably, Figure 5.12a and Figure 5.12b show that there is a distinct period around 700-800 kyr BP where the solutions diverge from the reference trajectory in a large range of forcing strengths u. This suggests that some aspect of the forcing around this time kicks the trajectories into the basin of attraction of the large-amplitude response. An additional area like this is seen in Figure 5.12b near 1600 kyr BP.

The second interesting feature in all three figures is a threshold behaviour in the forcing strength u: below a certain value of u specific to each τ, transitions do not occur such that the solutions just track the reference trajectory. Above this value trajectories suddenly can make this transition. For some values of delay τ we observe an additional larger threshold value for the transition at 1600 kyr BP.

Thresholds in u were also seen when periodic forcing is added to the model instead of the insolation time series. When creating Figure 5.12b with periodic forcing, we saw the same threshold transition at $u = 0.08$ (see Figure 5.6b). We showed that this transition is due to moving basins of attraction for the two stable solutions. The other sharp transition (in time) we see in Figure 5.12 is not possible with periodic forcing. Either the solution would start to transition immediately, or it would not transition at all (see Section 5.3). This leads us to believe that the quasiperiodicity is necessary for a transition of this type.

5.5 Model sensitivity to noise: desynchronisation and increased robustness of transition

As in most physical systems, the sensitivity of our model to small perturbations should be considered. While the astronomical forcing is the main driver of the glacial cycles, there are other processes (typically on faster timescales) that also can influence accumulation and loss of ice mass. In this section we will represent such processes as noise, and examine the effect on our results.

5.5.1 Finite-time Lyapunov Exponents

To analyse how trajectories depend on their history at specific instances of the forcing, we compute finite-time Lyapunov exponents (FTLEs) using a QR decomposition method. This method was previous used in De Saedeleer et al. (2013) to illustrate the desynchronisation of nearby trajectories in a van der Pol-type oscillator model. We will apply the ideas presented in that study to our forced model.

Our algorithm for computing Lyapunov exponents for a delay differential equation (DDE) follows from Farmer (1982).

Linearisation We consider a trajectory $X(t)$ of (5.41) and linearise (5.41) along this trajectory:

$$\dot{x}(t) = J_0(t)x(t) + J_\tau(t)x(t-\tau), \tag{5.44}$$

where J_0 an J_τ are the derivatives of the right-hand side of (5.41) with respect to $X(t)$ and $X(t-\tau)$:

$$J_0(t) = r - X(t-\tau)^2, \tag{5.45}$$

$$J_\tau(t) = -p - 2sX(t-\tau) - 2X(t)X(t-\tau). \tag{5.46}$$

Discretisation We consider an approximate solution of (5.44) $x_i = x(t_i)$ for $t_i = t_0 + ih$, where h is a small step size, obtained using a second order trapezoidal numerical solver for DDEs. We then discretise (5.44) into $m = \tau/h$ steps,

$$\vec{y}_{i+1} = M_i\vec{y}_i; \quad \vec{y}_i = [x_i, ..., x_{i-m}]^T, \tag{5.47}$$

where M_i is a square $(m+1)\times(m+1)$ matrix of the form

$$M_i = \begin{bmatrix} M_{i,1} & 0 & \dots & M_{i,m} & M_{i,m+1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

with entries

$$\begin{aligned} M_{i,1} =& 1 + \frac{h}{2}(J_0(t_{i+1}) + J_0(t_i)) + \frac{h^2}{2} J_0(t_{i+1}) J_0(t_i), \\ M_{i,m} =& \frac{h}{2} J_\tau(t_{i+1}), \\ M_{i,m+1} =& \frac{h}{2} J_\tau(t_i) + \frac{h^2}{2} J_0(t_{i+1}) J_\tau(t_i). \end{aligned}$$

QR method We use a continuous QR algorithm for computing the Lyapunov exponents (see Dieci et al. (1997)). The QR algorithm is based on the numerical linear algebra factorisation of a matrix M into an orthogonal matrix Q and an upper triangular matrix R. Although DDEs have an infinite number of Lyapunov exponents, this method allows us to compute up to $m+1$ through the discretisation. We can choose to compute only the largest l Lyapunov exponents by using a non-square Q. We initialise an arbitrary orthogonal Q_0 as an $m \times l$ matrix of form

$$Q_0 = \begin{bmatrix} I_l \\ 0_{(m-l)\times l} \end{bmatrix}.$$

Note that I_l is the $l \times l$ identity matrix and $0_{(m-l)\times l}$ is an $(m-l)\times l$ zero matrix. We define iteratively Q_i and R_i by the QR decomposition of $M_i Q_{i-1}$ (Matlab command $\mathrm{qr}(M_i Q_{i-1}, 0)$),

$$Q_i R_i = M_i Q_{i-1}, \tag{5.48}$$

which produces a square $l \times l$ upper triangular matrix R_i with eigenvalues $R_{i,jj} > 0$ $(j = 1, ..., l)$. We store R_i for each time-step. After N time-steps we have the relation

$$Q_N R_N ... R_1 = M_N ... M_1 Q_0. \tag{5.49}$$

The infinite Lyapunov exponents can then be approximated by

$$\lambda_j = \frac{1}{N} \sum_{i=0}^{N} \ln R_{i,jj}. \tag{5.50}$$

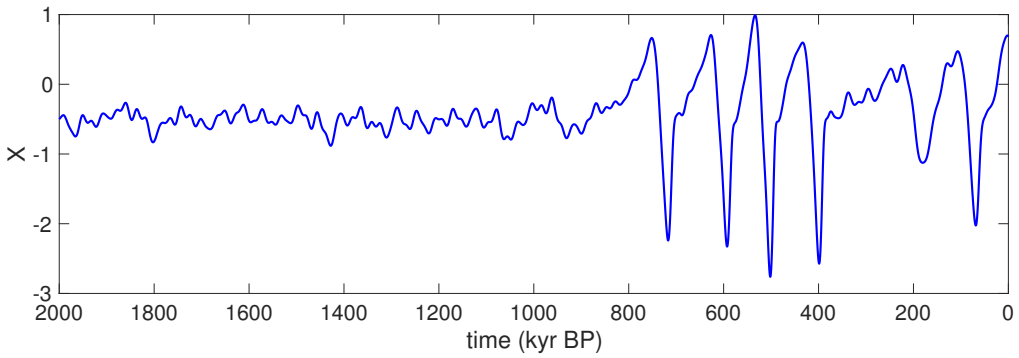


Figure 5.13: Example trajectory that exhibits an MPT-like transition with $\tau = 1.45$ and $u = 0.15$.

For FTLEs we truncate the above sum after W time-steps, and view how this truncation changes in time, i.e.

$$\lambda_{j,n} = \frac{1}{hW} \sum_{i=n-W}^{n} \ln R_{i,jj}. \tag{5.51}$$

Here $W = w/h$, where w is the desired time window length. Note that n initialises at time W.

The difference between FTLEs and classical Lyapunov exponents is that FTLEs are recorded for a family of time windows of finite length w rather than over the entire long time run. Thus, FTLEs are time-dependent functions instead of real numbers. A positive FTLE along a given trajectory $X(t)$ at time t_0 indicates that some nearby trajectories diverge exponentially from $X(t)$ in the time window $[t_0 - w, t_0]$. Therefore, $X(t)$ is sensitive to small perturbations in the time window $[t_0 - w, t_0]$. While a trajectory could be asymptotically stable, a positive FTLE at a time t_0 indicates temporary amplification of perturbations from the attracting trajectory (observed as temporary desynchronisation, see (De Saedeleer et al., 2013)).

5.5.2 FTLE implications on MPT and timing of major deglaciations

We analyse one trajectory, showing a forcing-induced MPT ($\tau = 1.45$, $u = 0.15$, shown in Figure 5.13) . We compute the FTLE over a sliding window of our example MPT trajectory with a window length of $w = 250$ kyr. This window length is chosen in order to filter out the dominant frequencies of the forcing. Similar results are seen with any window lengths w from 150-500 kyr.

Figure 5.14 shows the time profile of the largest FTLE along the trajectory. Before the transition around 800 kyr BP, the FTLE generally remains negative apart from a few short excursions above zero. At 1000 kyr BP the FTLE approaches zero and remains there for some time. Just before 800 kyr BP it goes positive and then on average stays positive for the remainder of the trajectory.

The negative FTLEs leading up to the transition confirm that the trajectory forgets its initial history and the effect of past disturbances. In particular, this implies that the infinite-dimensional nature

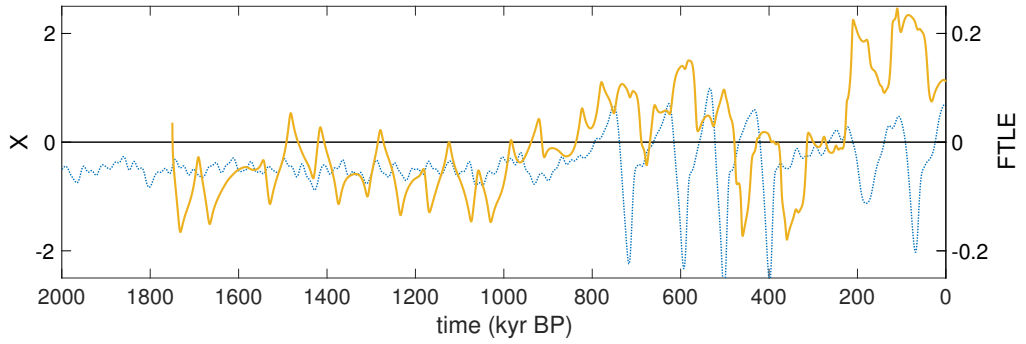


Figure 5.14: Finite-time Lyapunov exponents for window length $w = 250$ kyr computed from model run with $\tau = 1.45$ and $u = 0.15$: $X(t)$ (dotted blue), FTLE($[t - w, t]$) (gold).

of the DDE's possible initial conditions does not play a role for the MPT. Whenever the system is exhibiting the small-amplitude response we observed negative FTLEs.

The positive FTLEs indicate a sensitivity of the trajectory during the large-amplitude response. This sensitivity affects the precise timing of the deglaciations, as we now demonstrate by noise-induced desynchronisation of nearby trajectories.

To explore further the desynchronisation phenomenon outlined in De Saedeleer et al. (2013), we add noise to our system. The stochastic delay differential equation (SDDE) is then given as

$$\begin{aligned} dX(t) = \quad & [-pX(t-\tau) + rX(t) - sX(t-\tau)^2 \\ & -X(t-\tau)^2 X(t) - uM(t)]dt + \sigma dW(t), \end{aligned} \tag{5.52}$$

Here, $W(t)$ is standard white noise and σ is the noise amplitude. We will always consider σ as a fraction of the forcing amplitude u, i.e. $\sigma = \frac{u}{30}$.

We set the deterministic forcing strength $u = 0.15$ and the noise amplitude $\sigma = 0.005$. We ran 500 realisations of the model, all with the same initial history. The results of 10 randomly selected realisations can be seen in Figure 5.15. Up until 1000 kyr BP all trajectories generally track the same solution with only minor short desynchronisations. Shortly after 1000 kyr BP the trajectories begin to diverge, making the transition to the large amplitude oscillation state at different times. The trajectories then stay desynchronised, which corresponds to being at a different phase along the large amplitude oscillation. We illustrate this phenomenon in Figure 5.16, where we show the distribution of trajectories in the $(X(t), X(t-\tau))$-plane along the unforced periodic orbit, and in Figure 5.17.

Figure 5.17 shows the transition from synchronisation to desynchronisation in a density plot. At approximately 800 kyr BP (indicated by the grey line) the probability density makes a sharp transition from a small-variance to a large-variance-low-maximum density. The timing of this transition agrees with the first large positive excursion of the FTLE in the deterministic case and with the MPT. The standard deviation shows this transition as well, but with a lag of about 50 kyr. Figure 5.16

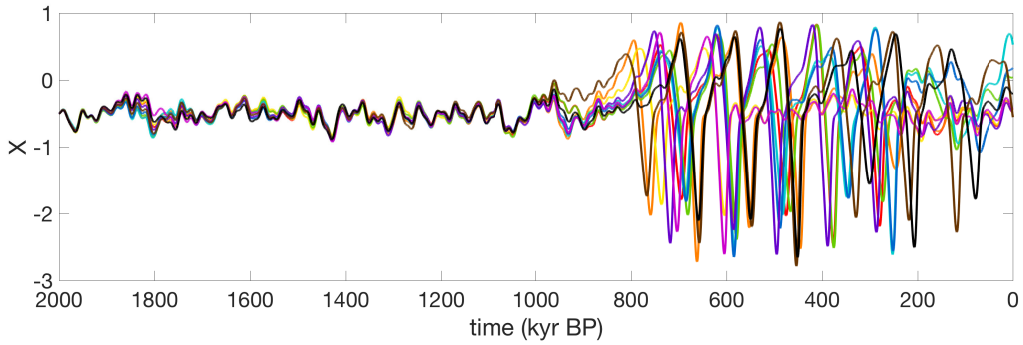


Figure 5.15: Trajectories of DDE model with noise (5.52): $u = 0.15$, $\sigma = 0.005$, $\tau = 1.45$.

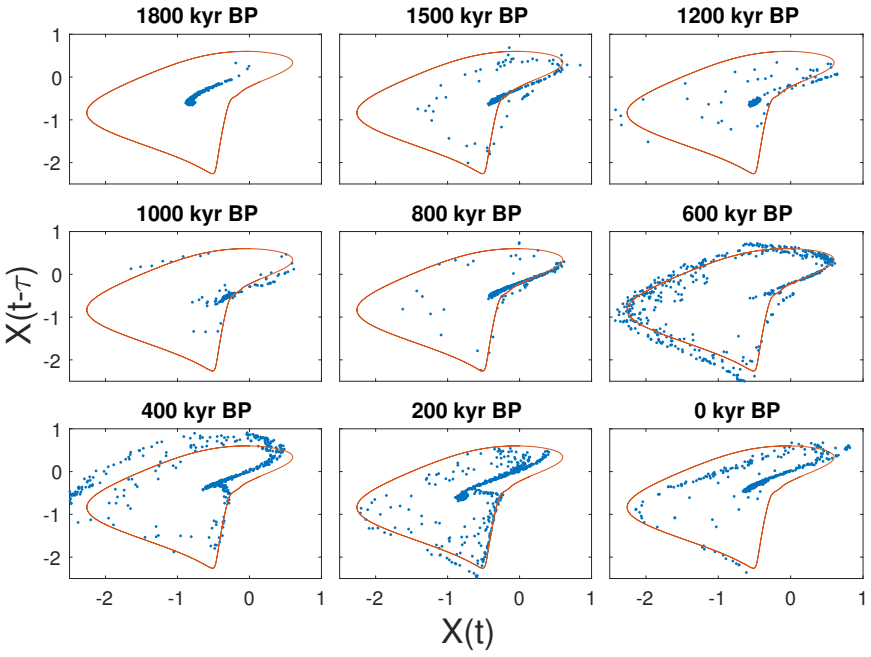


Figure 5.16: Phase portraits at specific time instances of 500 model runs with noise: $u = 0.15$, $\sigma = 0.005$, $\tau = 1.45$. The red curve represents the unforced periodic orbit.

and Figure 5.17 also show that the distribution by desynchronisation is far from uniform. There are still distinct deglaciation times favoured for many realisations. For this reason the desynchronisation observed in Figure 5.16 and Figure 5.17 does not contradict studies that have argued for phase-locking to different components of the orbital forcing. One of the original hypotheses of phase-locking in the late Pleistocene was suggested to be related to eccentricity, specifically associated to events of low eccentricity (Hays et al., 1976, Paillard, 2015). More recent studies have looked at the possibility of locking to precession or obliquity. In Ridgwell et al. (1999) the authors argue locking to every 4th or 5th precession cycle, while Huybers and Wunsch (2005) argue for locking to every 2nd or 3rd obliquity cycle. Later, Huybers (2011) attributes locking to a combination of precession and obliquity, and states strongly that the pacing cannot be attributed solely to one of these components. Robust evidence for locking requires testing on a range of initial conditions or noise realisations (or long time series),

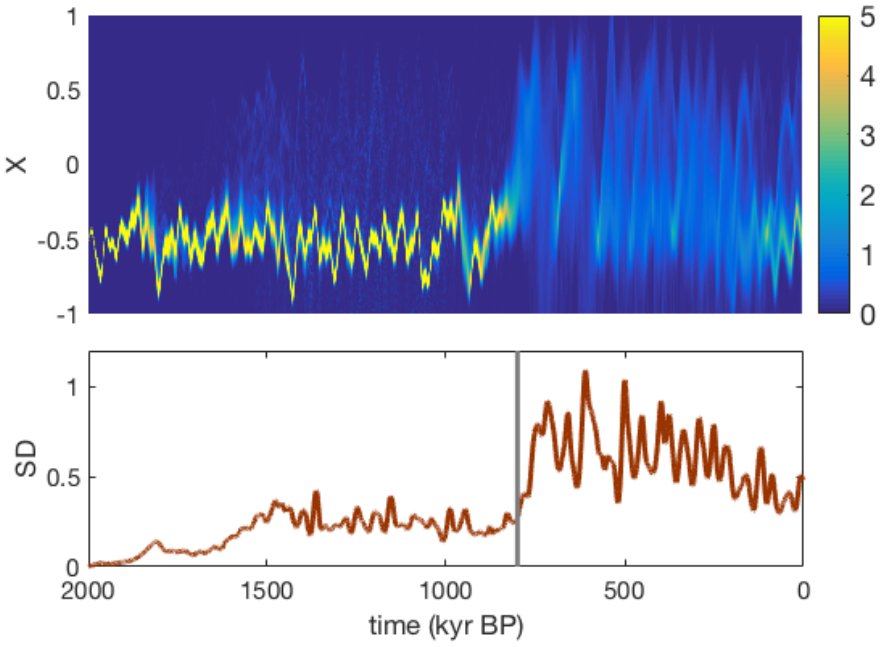


Figure 5.17: Probability density in time (top) and its standard deviation (bottom) of 500 model runs with noise: $u = 0.15$, $\sigma = 0.005$, $\tau = 1.45$. The grey line indicates 800 kyr BP

but the short data record corresponds to only one realisation of noise disturbance and one initial condition. Thus, data available may be insufficient to distinguish the higher-order locking proposed in the literature from the level of desynchronisation we report in Figure 5.16 and Figure 5.17.

5.5.3 Effect of noise with stronger forcing

Figure 5.18 shows that the addition of noise enhances the MPT for strong forcing. When adding noise, we observe transitions for forcing strengths u for which there was no transition in the deterministic case (compare Figure 5.12b and Figure 5.19). In Figure 5.12b the last persistent transition occurs at $u = 0.33$. Fixing the forcing strength at $u = 0.45$, we compute ten realisations. Here, two realisations exhibit transitions that persist until the end of the run. Although the transitions don't appear as commonly as in the weaker forcing case, with noise it is possible to observe an MPT-like transition across a wider range of forcing amplitudes.

Figure 5.19 shows a systematic overview of the transition enhancing effect of noise. We added noise of amplitude $\sigma = \frac{u}{30}$ and compute the distance diagram as in Figure 5.12b. We observe transitions occurring above the maximal value of u for which transitions occurred in the deterministic case.

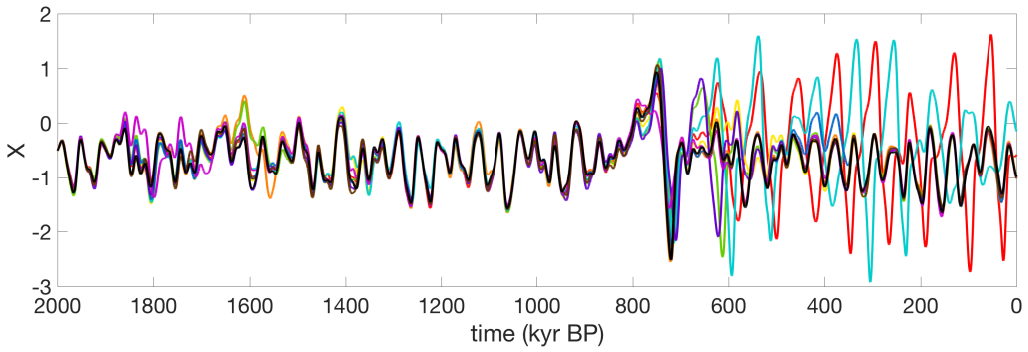


Figure 5.18: Trajectories of DDE model with noise: $u = 0.45$, $\sigma = 0.015$, $\tau = 1.45$.

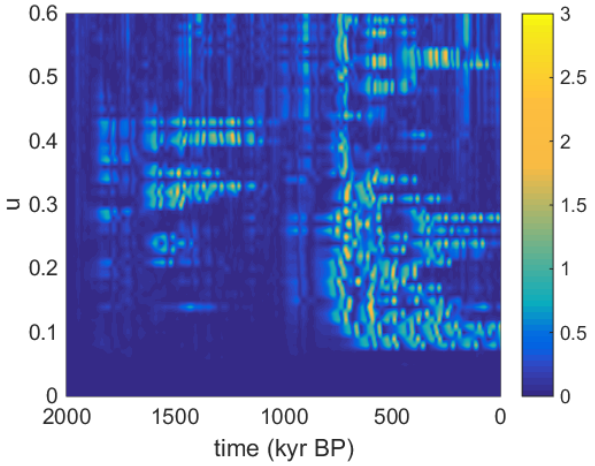


Figure 5.19: Distance from reference solution at $\tau_{\text{ref}} = 1.25$, for different forcing strengths u and noise strength $\sigma = u/30$. Averages taken over window length of size $\tau = 1.45$. Initial condition $X(t) = -0.5$ for $t \in [2 + \tau, 2]$ Myr BP.

Chapter 6

Exploring AMOC shutdown in a global ocean box model

In Chapter 2 we introduced the possibility of the Atlantic meridional overturning circulation (AMOC) shutting down (see Section 2.3.3). In this chapter we explore that possibility in a simplified box model of the global oceans. An empirical five-box model for global ocean circulation that exhibits AMOC switching has been developed by Rodriguez et al. (2017). By examining the unperturbed flows in the Atmosphere-Ocean Global Circulation Model (AOGCM) FAMOUS (Smith, 2012), the authors are able to reduce the global ocean circulation to a system of five boxes. This allows direct quantitative comparisons between states and fluxes for the AOGCM runs and the much simpler five-box model. Jackson et al. (2017) examine timescales of collapse within the model, and these studies form the starting point for the current chapter.

The model used is more complex than the original Stommel (1961) box model (see Chapter 2, Section 2.3.1.2), however it still retains many of the original features including the bistability. There are several notable differences:

- The first loss of stability of the 'on'-state is via a subcritical Hopf bifurcation rather than a fold bifurcation.
- There is an oscillatory relaxation to the 'on'-state - evidence of oscillatory behaviour is seen in the climate record of Dansgaard-Oeschger events (Clement and Peterson, 2008).
- The basin of attraction of the on state goes from infinite volume to finite volume at a homoclinic bifurcation.

The chapter is organised as follows. Section 6.1 introduces the five-box model of Rodriguez et al. (2017) and discusses a three-box reduction through a timescale separation. In Section 6.2 we perform

a comparative bifurcation analysis of these models and find good agreement. Section 6.3 explores the model calibrated to a potential future climate scenario, namely doubled atmospheric CO_2. We investigate the behaviour of the system under time-dependent perturbations to the hosing, in particular identifying examples of bifurcation-induced tipping (B-tipping) and rate-induced tipping (R-tipping). All of the sections in this chapter are included in the manuscript:

> H. Alkhayuon, P. Ashwin, L. Jackson, C. Quinn, and R. Wood. Basin bifurcations, oscillatory instability and rate-induced thresholds for amoc in a global oceanic box model. *In preparation*, 2018

Any coauthor contributions will be noted where applicable. The model scaling and model reduction, all numerical bifurcation analyses, and the studies of additional freshwater forcing with instantaneous change were carried out by the candidate. All figures were created by the candidate except where noted otherwise.

6.1 The model

By examining the geometry and behaviour of ocean currents for the FAMOUS runs (Smith, 2012), Rodriguez et al. (2017) propose a realistic box model with five boxes corresponding to large scale water mass structures in the global ocean. Only the balance of salinity in the boxes are considered, as temperature fluctuations occur on a much faster timescale and therefore can be considered fixed (with the exception of temperature in the North Atlantic box which will be explained later). The five boxes are as follows: North Atlantic (N), Tropical Atlantic (T), Indo-Pacific (IP), Southern (S), and Bottom (B). Figure 6.1 schematically illustrates the salinity fluxes between boxes.

This five-box model is described by the following systems of equations, where the initial values of the variables and parameters used by Rodriguez et al. (2017) are displayed in Table 6.2 and Table 6.3. We define the overturning as

$$q = \frac{\lambda[\alpha(T_S - T_0) + \beta(S_N - S_S)]}{1 + \lambda\alpha\mu} \tag{6.1}$$

For $q \geq 0$,

$$\left.\begin{aligned}
V_N \tfrac{dS_N}{dt} &= q(S_T - S_N) + K_N(S_T - S_N) - F_N S_0, \\
V_T \tfrac{dS_T}{dt} &= q[\gamma S_S + (1-\gamma)S_{IP} - S_T] + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0, \\
V_S \tfrac{dS_S}{dt} &= \gamma q(S_B - S_S) + K_{IP}(S_{IP} - S_S) + K_S(S_T - S_S) + \eta(S_B - S_S) - F_S S_0, \\
V_{IP} \tfrac{dS_{IP}}{dt} &= (1-\gamma)q(S_B - S_{IP}) + K_{IP}(S_S - S_{IP}) - F_{IP} S_0, \\
V_B \tfrac{dS_B}{dt} &= q(S_N - S_B) + \eta(S_S - S_B),
\end{aligned}\right\} \tag{6.2}$$

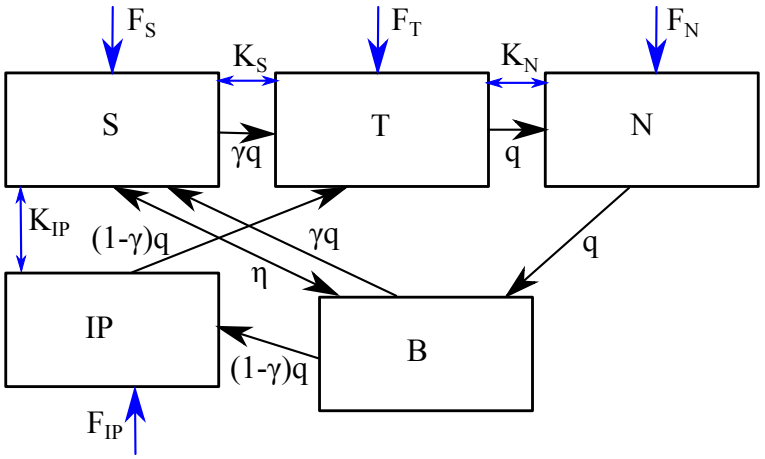


Figure 6.1: The flows between the five boxes in the global ocean box model of Rodriguez et al. (2017). From left to right: Southern Ocean, Tropical Atlantic, North Atlantic, Indo-Pacific and Bottom Ocean. The exchange of salinity is indicated by the black arrows. Freshwater surface fluxes F_X and wind-driven fluxes K_X are shown with blue arrows, while the overturning strength q is driven by a balance of salinity and temperature gradients described by (6.1). There is a mixing parameter η for the S and B boxes. The parameter $\gamma \in [0,1]$ represents the proportion of the overturning q which follows a cold water path from B to S, while the remainder follows a warm water path from B to IP. This figure was created by co-author Peter Ashwin.

while for $q < 0$,

$$\left.\begin{aligned} V_N \tfrac{dS_N}{dt} &= |q|(S_B - S_N) + K_N(S_T - S_N) - F_N S_0, \\ V_T \tfrac{dS_T}{dt} &= |q|(S_N - S_T) + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0, \\ V_S \tfrac{dS_S}{dt} &= \gamma|q|(S_T - S_S) + K_{IP}(S_{IP} - S_S) + K_S(S_T - S_S) + \eta(S_B - S_S) - F_S S_0, \\ V_{IP} \tfrac{dS_{IP}}{dt} &= (1-\gamma)|q|(S_T - S_{IP}) + K_{IP}(S_S - S_{IP}) - F_{IP} S_0, \\ V_B \tfrac{dS_B}{dt} &= |q|[\gamma S_S + (1-\gamma)S_{IP} - S_B] + \eta(S_S - S_B). \end{aligned}\right\} \tag{6.3}$$

If we consider the total salt content C

$$C = V_N S_N + V_T S_T + V_S S_S + V_{IP} S_{IP} + V_B S_B \tag{6.4}$$

then

$$dC/dt = -(F_N + F_T + F_S + F_{IP})S_0.$$

We want to conserve total salt content, so the total surface fresh water fluxes must satisfy

$$F_N + F_T + F_S + F_{IP} = 0, \tag{6.5}$$

and C is constant on trajectories of (6.2,6.3). In order to satisfy (6.5), we consider variation of the fresh water fluxes as follows: we prescribe a "hosing" parameter H (which may depend on time) that affects all of the fresh water fluxes simultaneously in a way that preserves (6.5). We use the same fluxes as in Rodriguez et al. (2017), given by Table 6.1.

F_N	$0.384 + 0.070H$	F_S	$1.078 - 0.257H$
F_T	$-0.723 + 0.752H$	F_{IP}	$-0.738 - 0.565H$

Table 6.1: Freshwater fluxes (all in Sv) depend on a non-dimensional "hosing" variable H. The total flux is zero for all H and corresponds to the baseline values in Table 6.2 for $H = 0$.

	Volume (x10^{17} m^3)	Salinity (psu)	Flux (Sv)
North Atlantic	$V_N = 0.3261$	$S_N = 0.034912$	$F_N = 0.384$
Tropical Atlantic	$V_T = 0.7777$	$S_T = 0.035435$	$F_T = -0.723$
Southern Ocean	$V_S = 0.8897$	$S_S = 0.034427$	$F_S = 1.078$
Indo-Pacific	$V_{IP} = 2.2020$	$S_{IP} = 0.034668$	$F_{IP} = -0.738$
Bottom Ocean	$V_B = 8.6490$	$S_B = 0.034538$	

Table 6.2: The volumes and standard (baseline) values for the AMOC models, based on the FAMOUS$_B$ runs (Smith, 2012) that are used in Rodriguez et al. (2017). The salinity values correspond to an equilibrium of the system. Note that the fluxes are assumed to balance: $F_N + F_T + F_S + F_{IP} = 0$.

Name	Default value	Units	Name	Default value	Units
α	0.12	kg m^{-3} °C^{-1}	K_N	5.456	Sv
β	790.0	kg m^{-3}	K_S	5.447	Sv
S_0	0.035		K_{IP}	96.817	Sv
T_S	4.773	°C	λ	2.79x10^7	m^6 kg^{-1}s^{-1}
T_0	2.650	°C	γ	0.39	
η	74.492	Sv	μ	5.5	°C$^{-1}m^{-3}s$x10^{-8}

Table 6.3: Baseline parameters used in the model, taken from FAMOUS$_B$ runs (Smith, 2012) in Rodriguez et al. (2017). Note that 1 Sv = 10^6m^{3}s^{-1}.

This formulation also allows us to eliminate one of the equations in (6.2) and (6.3). We therefore assume C is constant and solve (6.4) to give S_B in terms of the other variables (note that this choice of variable is arbitrary). Solutions are equivalent to the solutions of the original system as long as (6.5) is satisfied.

6.1.1 Three-box reduction

We consider an empirical three-box reduction as follows. We first perform a time-dependent hosing experiment such that the freshwater flux varies slowly in time. The variable H is defined as follows:

$$H = \begin{cases} 0.0005t & t \leq 2000, \\ 1 - 0.0005(t - 2000) & t > 2000. \end{cases} \tag{6.6}$$

Through (6.6) the freshwater forcing starts with a values of $H(0) = 0$, slowly increases to $H(2000) = 1$, and then slowly decreases at the same rate to $H(5000) = -0.4$. Figure 6.2 shows the behaviour of the salinity values in each box throughout the model run, from which we see that the variations of S_S and S_B are much smaller and slower than those of S_N and S_T. We then can perform a dimension

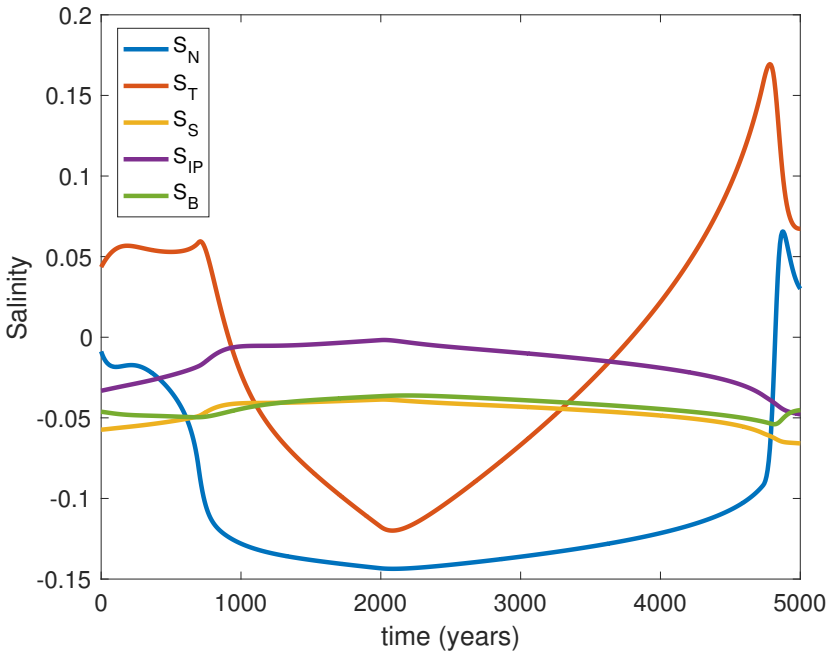


Figure 6.2: Transient behaviour of (6.2,6.3) when subjected to time-dependent hosing (6.6). Note the minimal variation seen for S_S and S_B when compared to S_N and S_T.

reduction by fixing S_S and S_B at their equilibrium values and considering only S_N, S_T, and S_{IP} as dynamic variables. Conservation of salt (6.4) means we can solve for one of the variables (we choose S_{IP}) to give

$$\left.\begin{aligned} V_N \tfrac{dS_N}{dt} &= q(S_T - S_N) + K_N(S_T - S_N) - F_N S_0, \\ V_T \tfrac{dS_T}{dt} &= q[\gamma S_S + (1-\gamma)S_{IP} - S_T] + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0, \\ S_{IP} &= \tfrac{1}{V_{IP}}(C - V_N S_N - V_T S_T - V_S S_S - V_B S_B), \end{aligned}\right\} \tag{6.7}$$

for $q \geq 0$ and

$$\left.\begin{aligned} V_N \tfrac{dS_N}{dt} &= |q|(S_B - S_N) + K_N(S_T - S_N) - F_N S_0, \\ V_T \tfrac{dS_T}{dt} &= |q|(S_N - S_T) + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0, \\ S_{IP} &= \tfrac{1}{V_{IP}}(C - V_N S_N - V_T S_T - V_S S_S - V_B S_B). \end{aligned}\right\} \tag{6.8}$$

for $q < 0$. The parameters are the same as those given in Table 6.2 and Table 6.3.

In order to improve computational efficiency (particularly for the continuation software), we rescale both time and state variables of the models. The time unit t of the original model is in seconds; we will consider a time unit of years $\tau = tY^{-1}$ where $Y = 3.15\text{x}10^7$ seconds. In addition we will consider the state variables as scaled perturbations from a background state, namely

$$\tilde{S}_i = 100(S_i - S_0), \qquad i \in [N, T, S, IP, B]. \tag{6.9}$$

This leads to a modified equation for q,

$$q = \frac{\lambda[\alpha(T_S - T_0) + \frac{\beta}{100}(\tilde{S}_N - \tilde{S}_S)]}{1 + \lambda\alpha\mu},$$

and the following ODEs: For $q \geq 0$ we have

$$\left.\begin{aligned} \tfrac{V_N}{Y}\tfrac{d\tilde{S}_N}{d\tau} &= q(\tilde{S}_T - \tilde{S}_N) + K_N(\tilde{S}_T - \tilde{S}_N) - 100F_N S_0, \\ \tfrac{V_T}{Y}\tfrac{d\tilde{S}_T}{d\tau} &= q[\gamma\tilde{S}_S + (1-\gamma)\tilde{S}_{IP} - \tilde{S}_T] + K_S(\tilde{S}_S - \tilde{S}_T) + K_N(\tilde{S}_N - \tilde{S}_T) - 100F_T S_0, \\ \tilde{S}_{IP} &= \tfrac{100}{V_{IP}}(C - \tfrac{V_N S_N + V_T S_T + V_S S_S + V_B S_B}{100} - S_0(V_N + V_T + V_{IP} + V_S + V_B)), \end{aligned}\right\} \quad (6.10)$$

while for $q < 0$ we have

$$\left.\begin{aligned} \tfrac{V_N}{Y}\tfrac{d\tilde{S}_N}{d\tau} &= |q|(\tilde{S}_B - \tilde{S}_N) + K_N(\tilde{S}_T - \tilde{S}_N) - 100F_N S_0, \\ \tfrac{V_T}{Y}\tfrac{d\tilde{S}_T}{d\tau} &= |q|(\tilde{S}_N - \tilde{S}_T) + K_S(\tilde{S}_S - \tilde{S}_T) + K_N(\tilde{S}_N - \tilde{S}_T) - 100F_T S_0, \\ \tilde{S}_{IP} &= \tfrac{100}{V_{IP}}(C - \tfrac{V_N S_N + V_T S_T + V_S S_S + V_B S_B}{100} - S_0(V_N + V_T + V_{IP} + V_S + V_B)). \end{aligned}\right\} \quad (6.11)$$

Note that the total salt content is still conserved with

$$C = \frac{1}{100}(V_N\tilde{S}_N + V_T\tilde{S}_T + V_S\tilde{S}_S + V_{IP}\tilde{S}_{IP} + V_B\tilde{S}_B) + S_0(V_N + V_T + V_S + V_{IP} + V_B). \quad (6.12)$$

We have only shown the scaled version of the three-box model, but the five-box model is scaled equally.

6.2 Bifurcation analysis

As discussed in Chapter 2, the AMOC can potentially be forced into an alternative stable state than the present one with increased freshwater fluxes, particularly in the North Atlantic. To explore the possibility of this in our model, we consider the effects of varying the surface freshwater fluxes F_X. We perform a bifurcation analysis on both the five-box and three-box model and compare the results.

6.2.1 Five-box model

Using COCO (Dankowicz and Schilder, 2013) continuation software we are able to verify that (6.2,6.3) exhibits a hysteresis between two branches of stable equilibria on varying H and imposing C constant (ie. using (6.4) to specify S_B), as shown in Figure 6.3. The upper and lower branches will be referred to as the 'on' and 'off' states respectively. Although both equilibria are destroyed through a fold bifurcation, the 'on' state first loses stability through a subcritical Hopf bifurcation. In addition there is a homoclinic bifurcation that leads to a sudden collapse in the size of the basin of attraction before the Hopf bifurcation occurs (we explain this further in Section 6.3.1).

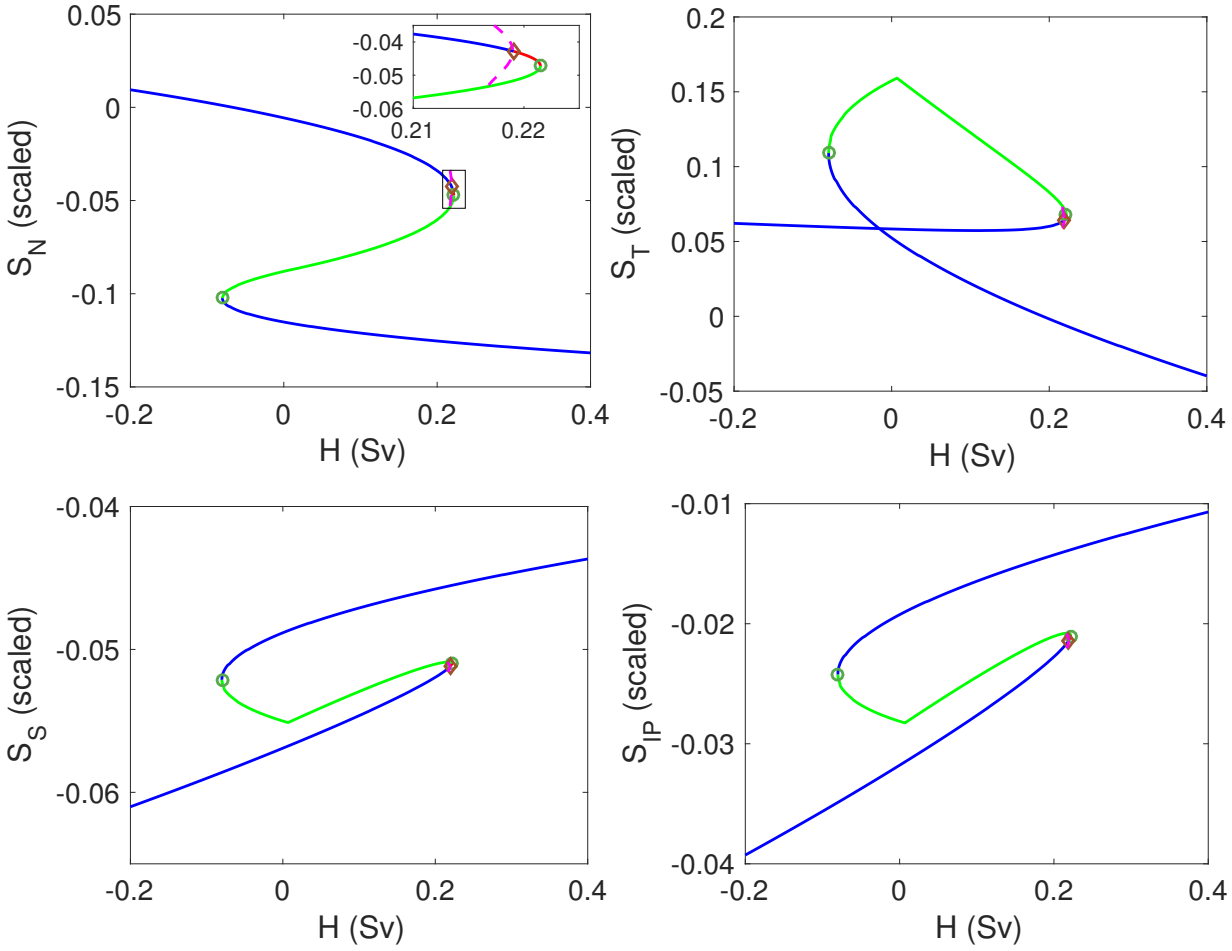


Figure 6.3: The bifurcation diagram of the five box model (6.2,6.3) on varying H. The stable equilibria are shown in blue, unstable in green (saddle) and red (source). The green circles indicate saddle node bifurcations, and the brown diamonds indicate Hopf bifurcations. The dotted pink line shows the minimum and maximum of the unstable periodic orbit. A zoom of the upper bifurcation region is shown for S_N.

6.2.2 Three-box model

We now demonstrate that the bifurcation structures for the five-box model are contained in the dynamics of the three-box reduction (6.7,6.8) when varying H: Figure 6.4 illustrates that all of the bifurcations are present in the same order and roughly the same location. More precisely, Table 6.4 compares the locations of the four bifurcations for the two systems. Using Maple it is possible to directly calculate the local fold and Hopf bifurcation points for the three box model; these parameter values are also included on Table 6.4. The homoclinic bifurcations are approximated by determining a large but fixed period orbit at the end of the branch of periodic orbits that emerges from the Hopf bifurcation.

Comparing the bifurcations of the five- and three-box models on varying H, there is close agreement

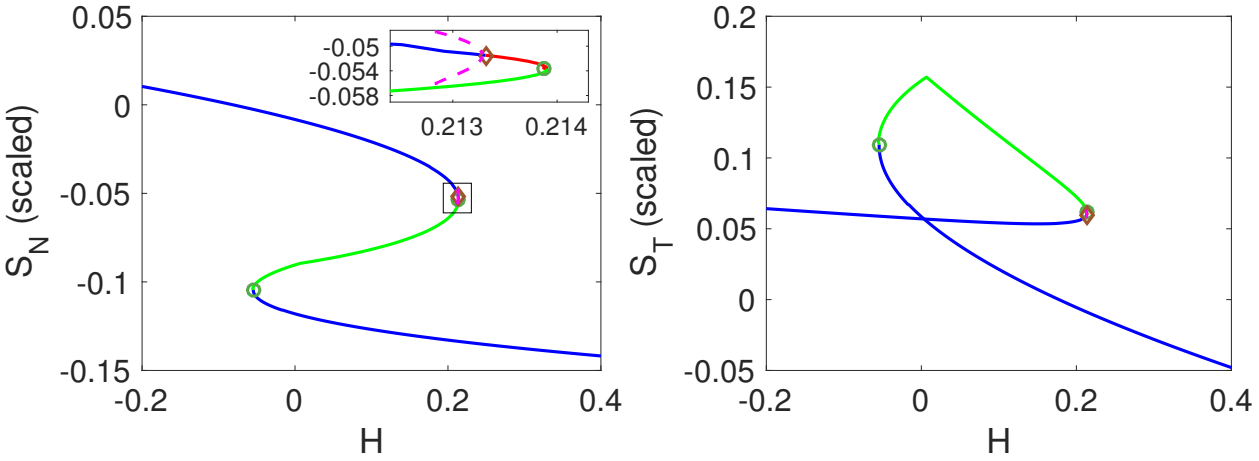


Figure 6.4: The bifurcation diagram of the three box model (6.7,6.8) on varying H. The stable equilibria are shown in blue, unstable in green (saddle) and red (source). The green circles indicate saddle node bifurcations, and the brown diamonds indicate Hopf bifurcations. The dotted pink line shows the minimum and maximum of the unstable periodic orbit. A zoom of the upper bifurcation region in shown for S_N.

		Five box numerical	Three box numerical	Three box Maple
Lower saddle node	H_{lsn}	-0.07996	-0.05446	-0.05445
Upper homoclinic	H_{hom}	0.2165	0.2128	NA
Upper Hopf	H_{Hopf}	0.2191	0.2134	0.2133
Upper saddle node	H_{usn}	0.2214	0.2139	0.2138

Table 6.4: Comparison of the locations of bifurcations for H for the five box (6.2,6.3) and the three box (6.7,6.8) models with standard parameters: the numerical solution by path following is using COCO apart from the homoclinic bifurcations that are estimated using xppaut. The last column gives numerical approximation using Maple (not available for the homoclinic bifurcation).

both qualitatively and quantitatively. The bifurcation scenario is also robust on changing the parameter γ (the proportion of cold-water path). If γ is increased, the structure remains. However, for decreasing γ we find the Hopf curve meets the upper fold curve (see Figure 6.5). The two curves meet at a Bogdanov-Takens (BT) point at $(H, \gamma) = (0.2268, 0.1559)$ that can be determined (with the aid of Maple) by simultaneous solution of an equilibrium of (6.7) with $q > 0$ whose Jacobian has both determinant and trace zero. Note that the BT point represents a convergence not just of Hopf and fold bifurcations, but also we can infer (Kuznetsov, 1998) that the homoclinic bifurcation continues to the BT point.

This is not the first time a bifurcation structure of this nature was seen in an model for the AMOC. A similar bifurcation structure was discovered and explored in the studies of Titz et al. (2002a,b).

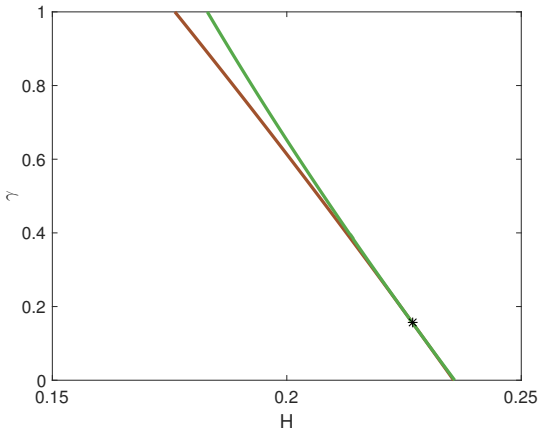


Figure 6.5: Two parameter plot showing the location of H_{Hopf} (brown) and H_{usn} (green) for the three box model on varying γ away from the default value 0.39. The two curves meet at a Bogdanov-Takens point at $(H, \gamma) = (0.2268, 0.1559)$ while the Hopf and upper saddle-node bifurcations move apart for larger γ and smaller H.

6.3 Exploring a future climate scenario

With the discovery of the underlying bifurcation structure, it is of interest to consider how it might change under an altered background climate that corresponds to a potential future climate scenario. For this we take parameter values from the FAMOUS$_B$ model calibrated to a world with doubled atmospheric CO_2 (also explored in Rodriguez et al. (2017)). The adjusted baseline and parameter values are listed in Table 6.5 and Table 6.6 respectively. We explore changes to the bifurcation structure and experiment with time-dependent hosing scenarios. For the latter we use the modified freshwater flux functions in Table 6.7.

	Volume (x10^{17} m^3)	Salinity (psu)	Flux (Sv)
North Atlantic	$V_N = 0.3683$	$S_N = 0.034912$	$F_N = 0.486$
Tropical Atlantic	$V_T = 0.5418$	$S_T = 0.035435$	$F_T = -0.997$
Southern Ocean	$V_S = 0.6097$	$S_S = 0.034427$	$F_S = 1.265$
Indo-Pacific	$V_{IP} = 1.4860$	$S_{IP} = 0.034668$	$F_{IP} = -0.754$
Bottom Ocean	$V_B = 9.9250$	$S_B = 0.034538$	

Table 6.5: The volumes and standard (baseline) values for the AMOC models with doubled atmospheric CO_2, based on the FAMOUS$_B$ runs (Smith, 2012) that are used in Rodriguez et al. (2017). The salinity values correspond to an equilibrium of the system. Note that the fluxes are assumed to balance: $F_N + F_T + F_S + F_{IP} = 0$.

Through conducting a bifurcation analysis as in Section 6.2, we see that the main effect of doubled atmospheric CO_2 is an increased unstable region before the saddle-node bifurcation (see Figure 6.6). From this we can infer that the system has moved further away from the Bogdanov-Takens point. This increased region of instability has implications on how the system would act under additional

Name	Default value	Units	Name	Default value	Units
α	0.12	kg m^{-3} °C^{-1}	K_N	1.762	Sv
β	790.0	kg m^{-3}	K_S	1.872	Sv
S_0	0.035		K_{IP}	99.977	Sv
T_S	7.919	°C	λ	1.62x10^7	m^6 kg^{-1}s^{-1}
T_0	3.870	°C	γ	0.36	
η	33.264	Sv	μ	22	°C$^{-1}m^{-3}s$x10^{-8}

Table 6.6: Baseline parameters used in the model, taken from FAMOUS$_B$ runs (Smith, 2012) with doubled atmospheric CO_2 in Rodriguez et al. (2017). Note that 1 Sv = 10^6m^{3}s^{-1}.

F_N	$0.486 + 0.131H$	F_S	$1.265 - 0.263H$
F_T	$-0.997 + 0.696H$	F_{IP}	$-0.754 - 0.564H$

Table 6.7: Freshwater fluxes (all in Sv) for doubled atmospheric CO_2 scenarios. The total flux is zero for all H and corresponds to the baseline values in Table 6.5 for $H = 0$.

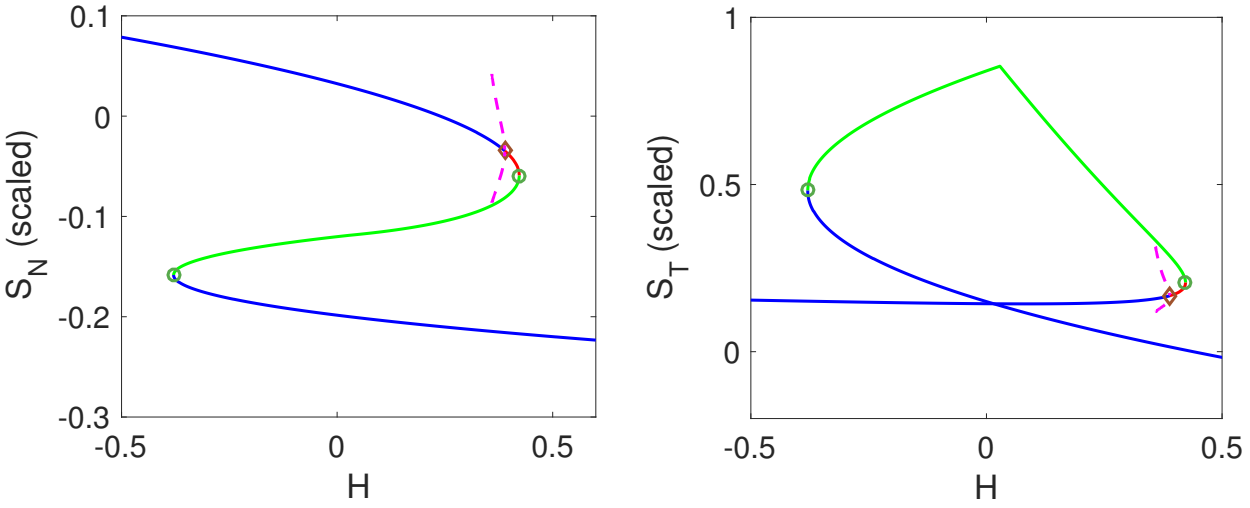


Figure 6.6: The bifurcation diagram of the three box model (6.7,6.8) on varying H with a background climate of doubled atmospheric CO_2. The stable equilibria are shown in blue, unstable in green. The green circles indicate saddle-node bifurcations, and the brown diamonds indicate Hopf bifurcations. The dotted pink line shows the minimum and maximum of the unstable periodic orbit. A zoom of the upper bifurcation region in shown for S_N.

freshwater forcing. We explain the phenomenon and explore some possible scenarios in the following sections.

6.3.1 Basin bifurcations

To understand the behaviour of the system in the bistable region for changing parameter H, we examine the basins of attraction for our equilibria. A schematic diagram of the expected behaviour is shown in Figure 6.7. We will outline the expected behaviour for increasing H, starting from the left

side of the bifurcation diagram. Before reaching the lower saddle-node bifurcation ($H < H_{lsn}$), there is only one equilibrium, hence all initial conditions converge to this point (shown in the bottom left). After crossing the lower saddle-node ($H_{lsn} < H < H_{hom}$), the system is bistable between two stable equilibria and the phase plane is divided by the stable manifold of the saddle equilibrium. Increasing H further reaches a homoclinic bifurcation. This is where the most significant change to the basins of attraction is seen. At this point, the unstable manifold of the saddle equilibrium connects to the stable manifold of the saddle equilibrium. This causes the basin of attraction of the upper equilibrium to go from infinite to finite. For $H_{hom} < H < H_{Hopf}$, the basin of attraction of the upper equilibrium is bounded by an unstable periodic orbit, whereas the rest of the phase plane converges to the lower stable equilibrium. The unstable periodic orbit shrinks as H increases, eventually leading to a Hopf bifurcation. After the Hopf bifurcation ($H_{Hopf} < H$) only the lower stable equilibrium exists, and so all initial conditions in the phase plane are attracted to this.

The expected behaviour is realised in our system, shown through Figure 6.8. The black line indicates the stable manifold of the saddle equilibrium, which also acts as the boundary between the basins of attraction for the upper and lower stable equilibria. The colour contour indicates the years it takes to converge to the respective equilibrium (indicated by the blue circles). In the upper left panel we have $H_{lsn} < H < H_{hom}$, meaning there exist two basins of attraction separated by the stable manifold of the saddle equilibrium. The upper right panel and lower left panel also have $H_{lsn} < H < H_{hom}$, but the shrinking basin of attraction is already visible. In the lower right panel we have $H \approx H_{hom}$ where the upper basin becomes finite.

The shrinking basin of attraction for $H_{hom} < H < H_{Hopf}$ has some counter-intuitive effects for perturbations. As the basin of the upper branch shrinks to a small size, small perturbations in any direction can take the system out of the basin of the upper branch. We demonstrate this type of behaviour in the following section.

6.3.2 Hosing experiments

We now turn to the influence of time-dependent perturbations of (6.7), (6.8) where the freshwater fluxes F_X are varied via $H(t)$ according to Table 6.7.

If $H(t)$ is varied slowly enough, standard arguments of adiabatic reduction (Kuehn, 2015) can be used to predict the behaviour of the nonautonomous perturbed system. However, it is not always clear how slow this needs to be, and in many cases there are scientifically important questions that need exploration of a range of perturbation speeds.

The particular protocol we consider here is inspired by the question of how quickly changes to hosing need to be reversed to avoid tipping from the upper branch onto the lower branch (Ritchie et al., 2017).

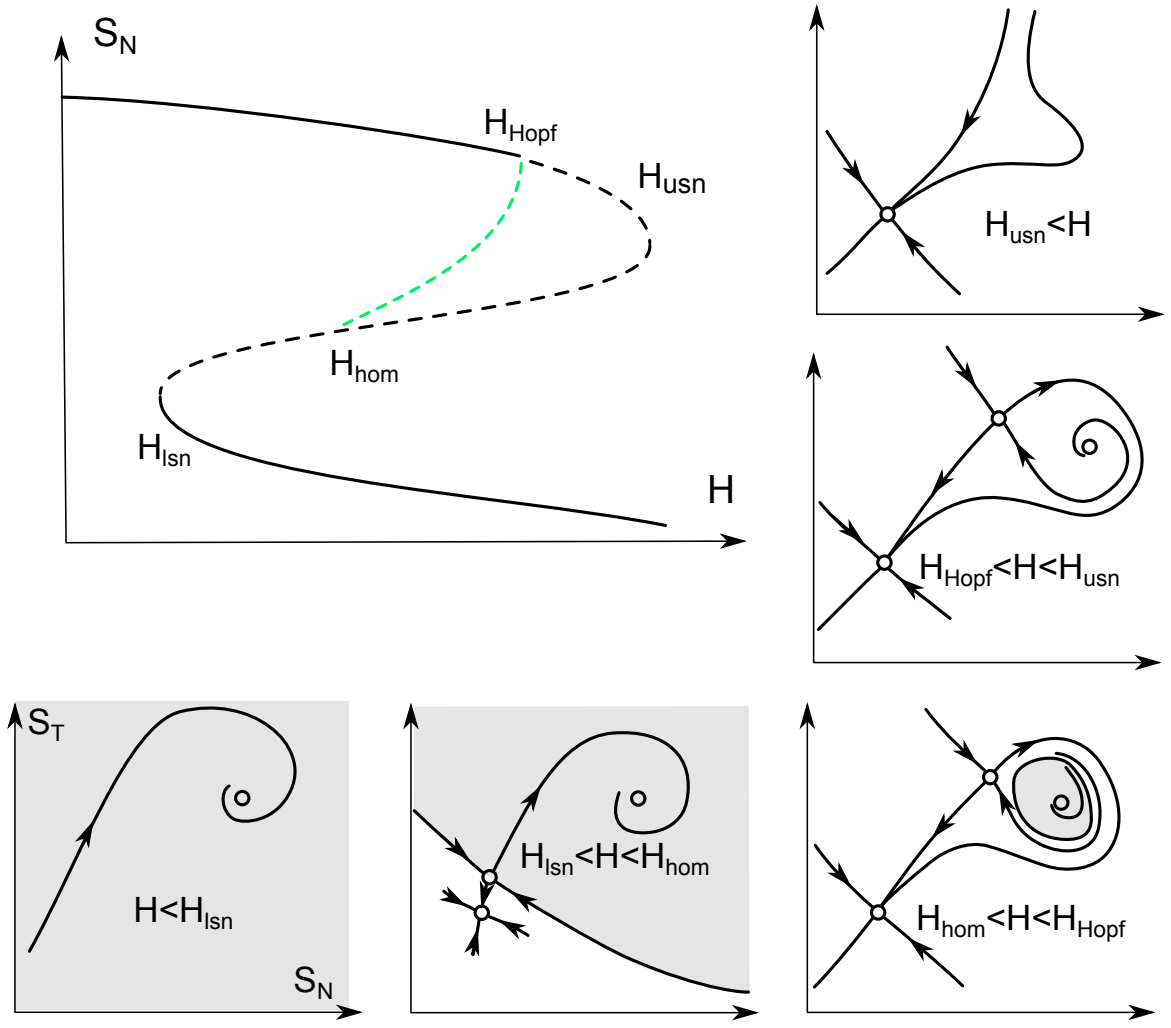


Figure 6.7: Schematic of the bifurcation diagram (top left) and phase portraits on varying H through various points on the bifurcation diagram of (6.7,6.8). The shaded region shows the basin of attraction for the upper branch - this shrinks from being semi-infinite for $H < H_{hom}$ to being quite small for $H_{hom} < H < H_{Hopf}$. Note also that the upper branch loses stability before the saddle node. This figure was created by co-author Peter Ashwin.

To this end, we consider the effect of piecewise linear (PWL) perturbations $H(t) = H_{pwl}(t - t_0)$ as in Figure 6.9 on the system, where

$$H_{pwl}(\tau) = \begin{cases} H_0 & \tau < 0, \\ \alpha(\tau) & \tau \in [0, T_{rise}], \\ H_{pert} & \tau - T_{rise} \in [0, T_{pert}], \\ \beta(\tau) & \tau - T_{rise} - T_{pert} \in [0, T_{fall}], \\ H_0 & \tau \geq T_{rise} + T_{pert} + T_{fall}, \end{cases} \tag{6.13}$$

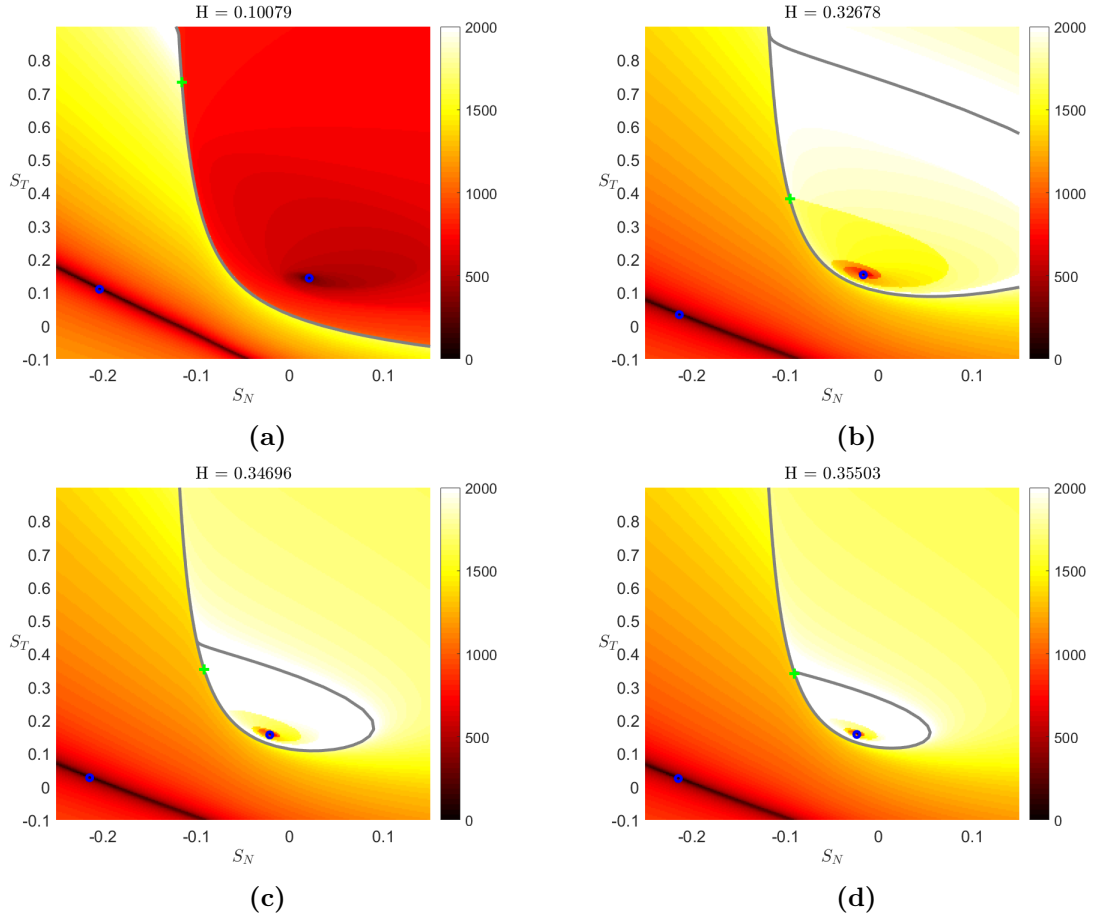


Figure 6.8: The basins of attraction of the two stable equilibria for different values of H, system (6.7) and (6.8). The colour map represents how long (in years) it takes to reach the corresponding equilibria. These figures were created by co-author Hassan Alkhayuon.

where $\alpha(t)$ and $\beta(t)$ are linear functions such that H is continuous and $\tau = t - t_0$. This depends on constants T_{rise}, T_{pert}, T_{fall}, levels H_0 and H_{pert}. If we define the rise and fall rates

$$r_{rise} = \frac{|H_0 - H_{pert}|}{T_{rise}}, \; r_{fall} = \frac{|H_0 - H_{pert}|}{T_{fall}}. \tag{6.14}$$

then we can write $\alpha(\tau) = H_0 + r_{rise}\tau$, $\beta(\tau) = H_{pert} - r_{fall}(\tau - T_{rise} - T_{pert})$.

As an example of the sort of critical transition this can lead to, Figure 6.10 shows the effect of two very similar PWL perturbations that vary only in T_{Rise} by a relatively small amount. As can be seen in the phase plane (a) and time series (b) these lead to quite different outcomes at the end of the perturbation.

A special case of the PWL perturbation (6.13) is when $T_{rise} = T_{fall} = 0$: an instantaneous change from H_0 to the perturbed H_{pert} and then back. This type of perturbation has been studied extensively in ecological settings (Bolt et al., 2018, Ratajczak et al., 2017) where it is called a "press". We explore the behaviour of this case when varying H_{pert} and T_{pert}, fixing $H_0 = 0$: an example is depicted in

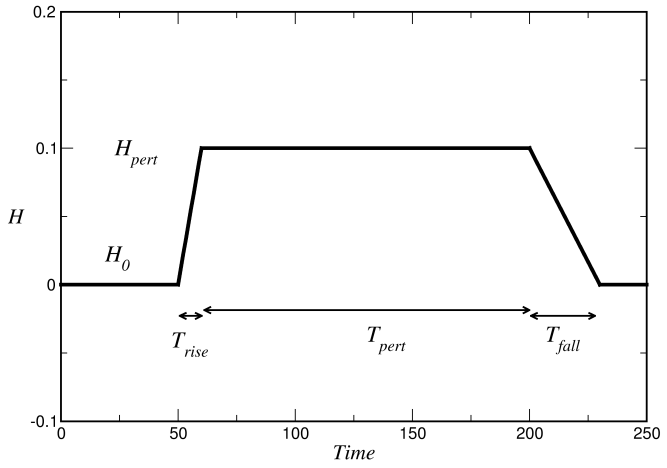
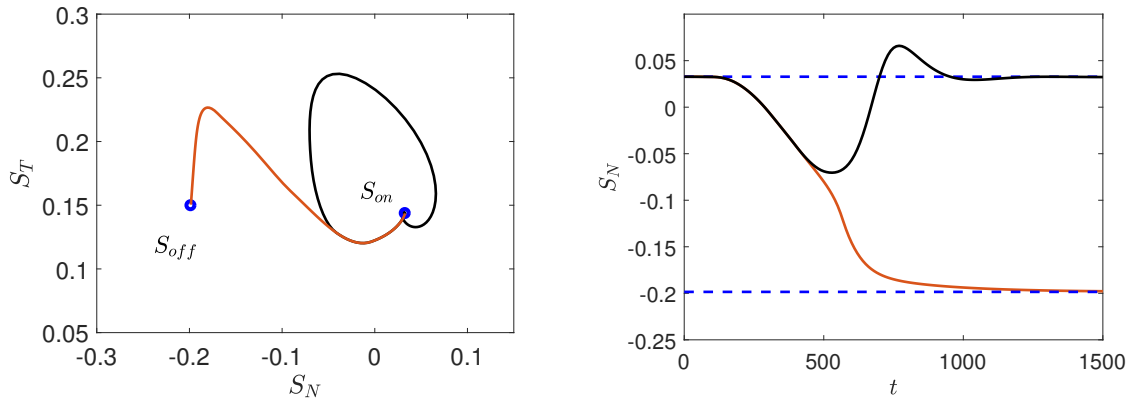


Figure 6.9: Schematic diagram showing a piecewise linear time-dependent perturbation $H(t) = H_{pwl}(t - t_0)$ where $t_0 = 50$, $T_{rise} = 10$, $T_{pert} = 140$, $T_{fall} = 30$, $H_0 = 0$ and $H_{pert} = 0.1$. This figure was created by co-author Peter Ashwin.



(a) **(b)**

Figure 6.10: Trajectories of (6.7,6.8) with piecewise linear time-dependent parameter H defined by (6.13), the fixed parameter values are $T_{pert} = 200$, $H_0 = 0$, $H_{pert} = 0.5$, $T_{rise} = 200$. We use $T_{fall} = 100$ for the black trajectory and $T_{fall} = 200$ for the dark red trajectory. The subfigure (a) shows the phase portrait and (b) the time profile of S_N. The blue dashed lines denote the stable equilibria values for $H_0 = 0$. These figures were created by co-author Hassan Alkhayuon.

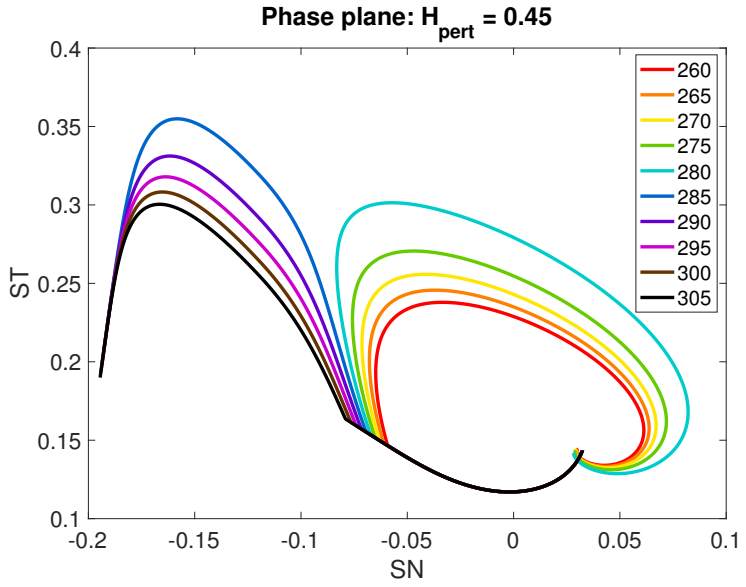


Figure 6.11: Trajectories of (6.7,6.8) with parameter H defined by (6.13) for $H_{pert} = 0.45$ with $T_{rise} = T_{fall} = 0$. The initial hosing value is fixed at $H_0 = 0$, and T_{pert} (in years) is indicated by the line colour. An initial condition is taken close to the upper branch for $H = 0$.

Figure 6.11. The black line shows the initial departure from the equilibrium state for $H = H_0$. All trajectories follow this path in the $H = H_{pert}$ phase plane for the time T_{pert} corresponding to the line colour. When $t > T_{pert} + 100$ trajectories then make an excursion back to one of the two stable equilibria for $H = H_0$. The maximal press duration which allows a return to the upper branch can be estimated between the two diverging trajectories. Comparing Figure 6.11 and Figure 6.8 we note that the divergence corresponds to where the black trajectory intersects with the basin boundary. This means we can find the critical duration of the perturbation that results in switch of branches. We illustrate this critical value for different values of H_{pert} in Figure 6.12.

We now consider some illustrative examples of the effect of more general perturbations of the form (6.13) starting at $H_0 = 0$ and the 'on'-state.

B-tipping and avoided B-tipping In the case that $H_0 < H_{Hopf} < H_{pert}$, the bifurcation diagram Figure 6.4 shows that the lower branch is the unique attractor for $H = H_{pert}$. This means that regardless of T_{rise} and T_{fall} we expect to see tipping onto the lower branch if T_{pert} is large enough. However, if some combination of T_{rise}, T_{fall} and T_{pert} is sufficiently small we may avoid B-tipping. Figure 6.13 illustrates these two scenarios, the black trajectory shows the avoiding tipping behaviour while the dark red trajectory shows B-tipping. Note the overshoot behaviour of the black trajectory when it returns to the upper branch - this is due to the oscillatory nature of the 'on'-state which is also observed in the AOGCM.

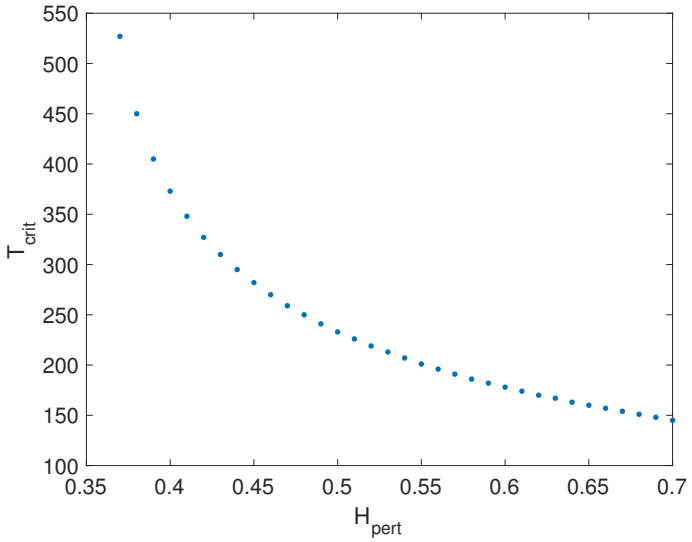


Figure 6.12: Critical values of T_{pert} for different values of H_{pert} ($T_{rise} = T_{fall} = 0$). Values below the curve will not tip, while values above will.

R-tipping and avoided R-tipping The bifurcation diagram Figure 6.4 might seem to suggest that a perturbation from H_0 to $H_{pert} < H_{Hopf}$ will not result in tipping to the lower branch for any choice of T_{rise}, T_{fall} and T_{pert}. Indeed, Figure 6.14 shows the effect of a perturbation that increases and then decreases H slow enough for there to be no tipping from the upper branch. On the other hand, for a more rapid perturbation (the dark red trajectory in Figure 6.14) we can find situations that switch to the lower branch. Again we see the overshoot in the black trajectory on return to the upper branch.

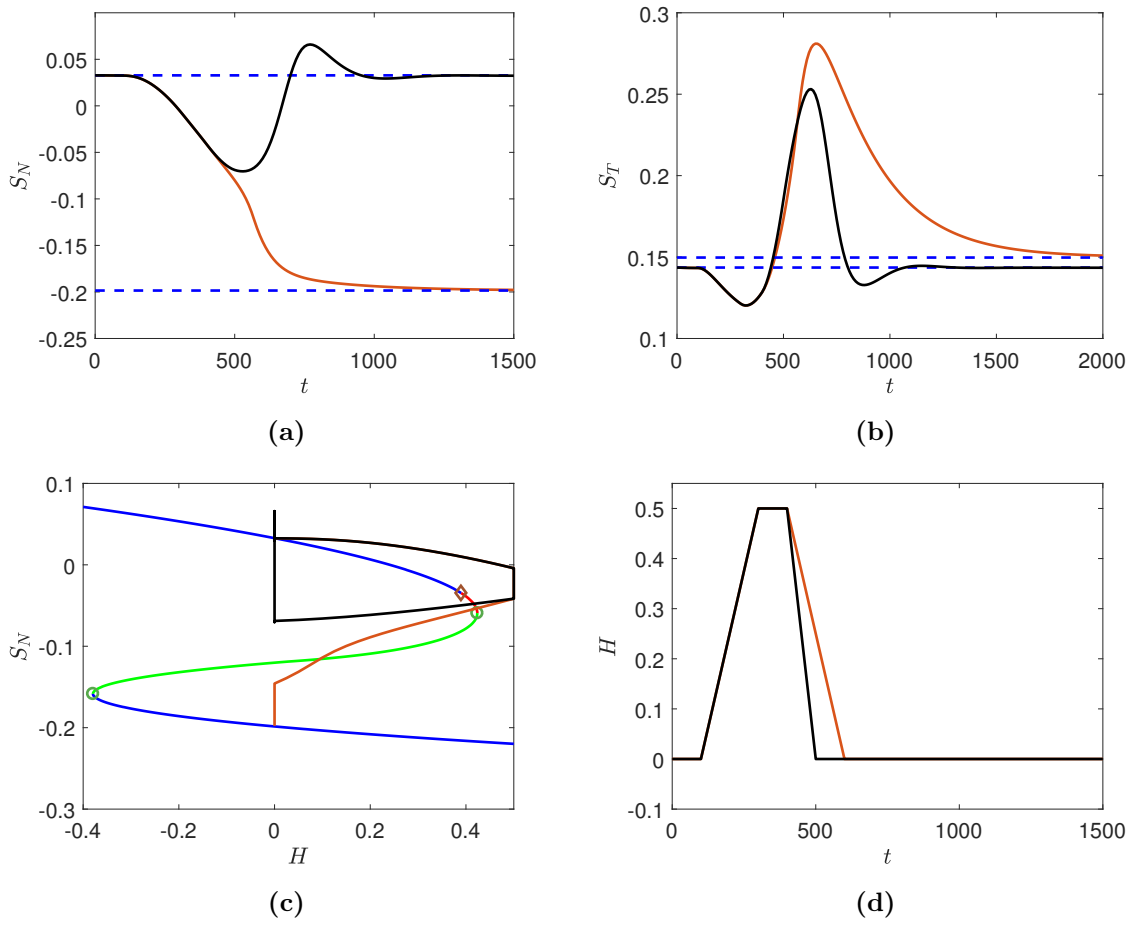


Figure 6.13: Bifurcation-induced tipping system (6.7, 6.8) for two trajectories, one tips (dark red) and the other does not (black). Subfigures (a) and (b) show the time series of S_N and S_T respectively. The dashed blue lines represent the two equilibria for $H_0 = 0$. Subfigure (c) shows the trajectories of S_N plotted over the bifurcation diagram. Subfigure (d) shows the piecewise linear forcing against time. We use $T_{fall} = 200$ for the dark red trajectory and $T_{fall} = 100$ for the black trajectory. The other parameter values are $H_0 = 0$, $H_{pert} = 0.5$, $T_0 = 100$, $T_{pert} = 200$, and $T_{rise} = 200$. These figures were primarily created by co-author Hassan Alkhayuon- bifurcation curve was created by PhD candidate.

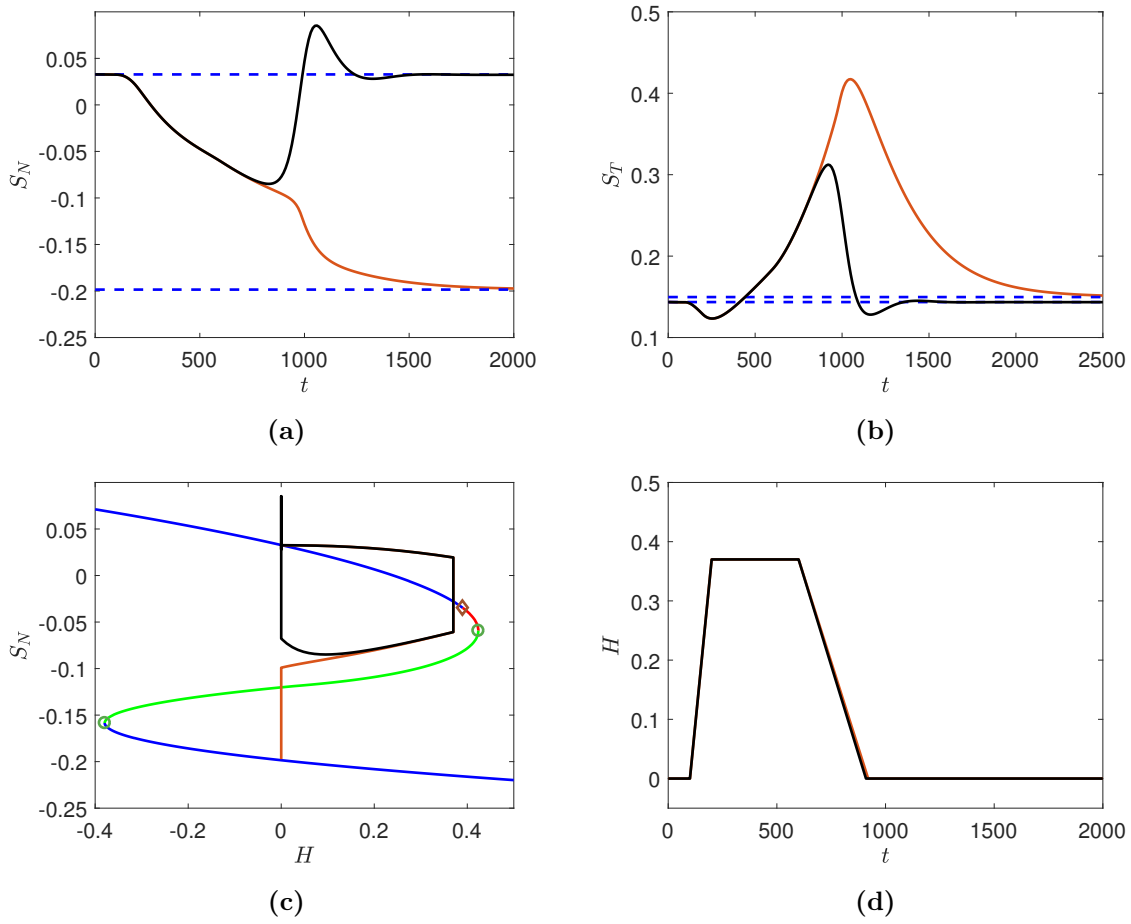


Figure 6.14: Rate-induced tipping for system (6.7, 6.8) for two trajectories, one tips (dark red) and the other does not (black). Subfigures (a) and (b) show the time series of S_N and S_T respectively. The dashed blue lines represent the two equilibria for $H_0 = 0$. Subfigure (c) shows the trajectories of S_N plotted over the bifurcation diagram. Subfigure (d) shows the piecewise linear forcing against time. Here, $H_0 = 0$ and $H_{pert} = 0.37$, which is less than H_{Hopf} where the upper branch loses its stability. The other parameter values are $T_{rise} = 100$, $T_0 = 100$ and $T_{pert} = 400$. We use $T_{fall} = 320$ for the the dark red trajectory and $T_{fall} = 310$ for the black trajectory. These figures were primarily created by co-author Hassan Alkhayuon- bifurcation curve was created by PhD candidate.

Chapter 7

Discussion

7.1 Summary

In this thesis we explored the possibility of delayed feedback and critical transitions in models of the climate system. We do this primarily through conceptual models, but discuss the connection to more complex models.

Methods for deriving models with delay

In Chapter 3 we introduced three different methods for obtaining models with delay from other types of systems. First, we described the linear chain approximation, which allows for the reduction of large ODE system comprised of linear first-order filters into a scalar delay equation. We showed two example systems: (i) all equal and (ii) all distinct timescales of the filters. The method is not restricted to scalar feedback- any large ODE system with subsystems of linear first-order filters can be reduced into delay equations for the variables not in any of the subsystems.

The second method introduced was the Mori-Zwanzig Projection framework. This method also applies to ODE systems, but those of a more general form than in the linear chain case. Given that a system can be separated into resolved variables and unresolved variables such that the system of unresolved variables is stable and decays exponentially, we derived equations for the resolved variables which are comprised of a direct (Markovian) term and a memory term. The behaviour of the memory term is defined by the orthogonal dynamics. To compute the behaviour of the memory term, we introduce a way of approximating the orthogonal dynamics which is exact for linear systems and has a quantifiable error for nonlinear systems. We illustrate the effectiveness of the approximation on two systems where the exact equation for the memory term is known.

The last method discussed was applicable to PDE systems, particularly those comprised of wave equations. We showed that through the boundary conditions we can obtain an equation for the delayed feedback that propagates with the wave. The delay is equal to the time it takes a wave to return to a location, which is calculated through the characteristic wave speed. We showed the use of this method in relation to ocean phenomena. Both the El Niño Southern Oscillation (ENSO) and the Atlantic Multidecadal Oscillation (AMO) were captured through delayed feedback. We explored solutions of the AMO model with delay, as this was an original model for the phenomenon. We found that the model captures the low-frequency oscillations that are characteristic of the AMO. We also found that when initialising the model with historical data, the low-frequency signal was even more prominent.

Exploration of climate models

In the following chapters we turned our attention to studying a few different climate models in detail. Chapter 4 began with the analysis of a previously conceived model with delay for the Pleistocene climate. Ghil et al. (1987) developed a Boolean Delay Equation (BDE) model for global ice volume, global temperature, and global ocean circulation on the timescale of thousands of years. They considered delays due to ice sheet accumulation, ice sheet expansion, and overturning time of the ocean. We converted the logically-based model into a discrete mapping system using algebraic equivalent operations to the logical operators. We analysed this new model in its discrete form and also through a conversion to a system of continuous DDEs. We found that both models exhibited either an equilibrium or a periodic orbit based on the relative sizes of the delays. In particular, for a large enough ocean overturning delay time, the models exhibited self-sustained oscillations.

These findings led us to our Chapter 5 study, in which we revisited another conceptual model for the Pleistocene climate, the Saltzman and Maasch (1988) model. This model has analogous state variables to those in the Ghil et al. (1987) BDE model: global ice mass, atmospheric CO_2, and North Atlantic Deep Water (NADW). We discovered that the Saltzman and Maasch (1988) model can be reduced to a scalar DDE in ice mass perturbations. The delay in this case was a combination of the three delay effects mentioned in the Ghil et al. (1987) model, with the most variable again being ocean overturning time. We analysed the attractors of the DDE model for different delay times and found a bistable region between a stable equilibrium and a stable large-amplitude periodic orbit which wasn't studied in the original model. We considered the behaviour within this bistable region when subjected to periodic and astronomical (quasiperiodic) forcing. We found that the astronomical forcing could cause a transition between the two attractors without any change in the parameters. The timing of the transition agreed with what is understood to be the timing of the Mid-Pleistocene Transition (MPT), seen in paleoclimate records. In the DDE model, the transition is possible over a large range of parameters in the bistable region and is more robust with added noise.

We then shifted our focus to the ocean circulation itself and looked into the mechanisms that affect the overturning time. Chapter 6 explores a box model for the global ocean circulation. This model was originally developed by Rodriguez et al. (2017) through analysis of the Atmosphere Ocean Global Circulation Model (AOGCM) FAMOUS. We were able to reduce the five-box model to a three-box model while preserving the dynamical behaviour. We find that when modifying the amount of additional freshwater to the system, the model can exhibit multiple stable states. An interesting feature discovered is that one of the stable states loses stability through a subcritical Hopf bifurcation rather than a fold bifurcation, which was traditionally seen in similar ocean circulation models. This type of bifurcation has implications on the basin of attraction of one of the stable states- there exists a point before the bifurcation where the basin goes from being infinite to finite. We observed a significant shrinking basin of attraction for the current state of the global ocean circulation with increased freshwater forcing. This phenomenon was even more pronounced when the model was calibrated to a future climate scenario of doubled atmospheric CO_2 levels. We explored different examples of time-dependent freshwater forcing and found that the model can exhibit bifurcation-induced tipping and rate-induced tipping, but additionally it can exhibit tipping prevention.

7.2 Outlook

Overall, this thesis presents some novel studies of climate phenomena. We showed the insight that can be gained about physical processes by analysing the contribution of any delayed feedback. Additionally, we demonstrated that the dominant behaviour of some complex processes can be explained by delay equations with only a few state variables. This is invaluable for getting a mechanistic understanding of what drives the climate system.

We expanded the use of delay equation models in paleoclimate. Our DDE model for the Pleistocene not only shone a light on the delayed feedback nature of the glacial cycles, but also introduced a new null hypothesis for the MPT: the transition being induced by a combination of bistability and external forcing. We conjecture that this null hypothesis could be tested for different models of the Pleistocene that exhibit similar bistability. We also found that in the mathematical framework, the transition is a result of interesting dynamical behaviour in the forced DDE model which appears to be attributed to the quasiperiodic nature of the forcing. This can be explored further through follow-up studies.

The dynamical behaviour discovered in the global ocean circulation model is of great significance to the study of the Atlantic Meridional Overturning Circulation (AMOC) shutdown. The mechanism for loss of stability differs from the widely accepted fold bifurcation theory. This has implications for tipping behaviour as well as recovery. As the box model was derived from an AOGCM, we expect the mechanism found in this thesis to be present in more complex models of ocean circulation. This

should be investigated in order to better understand the reaction of the AMOC to an ever-changing background climate.

Appendix A

Detailed description of modified Search Circle Algorithm

This appendix was written by co-author Jan Sieber and describes how one can modify the search circle (SC) algorithm for stable manifolds in England et al. (2004) for maps $\mathcal{M}_\ell$ given implicitly through

$$\mathcal{M}_\ell : \operatorname{dom} L \ni x \mapsto y \in \operatorname{dom} L, \quad \text{where } y \text{ is the solution of } \quad RM^{\ell+1}Lx = RM^{\ell}Ly. \tag{A.1}$$

The original SC algorithm (England et al., 2004) grows the stable curve of a map iteratively. In particular, the algorithm of England et al. (2004) does not rely on root-finding using Newton iterations, but rather on a bisection on a suitable small circle to find intersections between the image of the small circle and the collection of previously computed segments. In principle, the original SC algorithm could be applied directly, if one solves the defining system (A.1), $RM^{\ell+1}Lx = RM^{\ell}Ly$, for y every time the original algorithm applies its map (in our case $\mathcal{M}_\ell$) to a point $x \in \mathbb{R}^2$. However, a small modification of the SC algorithm avoids the need to solve the nonlinear equation (A.1). The paragraphs below describe the modified version of the SC algorithm in England et al. (2004), as implemented in the demo scripts in the supplementary material and used for the computation of the stable manifold shown in the paper.

Evaluation of parametric curves As part of the algorithm we need to manage smooth curves with potentially sharp turns, stored in the form of a sequence $S = (x^0, \ldots, x^k)$ of points in $\mathbb{R}^2$. In general, in order to associate an sequence $S = (x^0, \ldots, x^k)$ to a continuous parametric curve one needs an interpolation scheme. In our implementation we use optionally either a piecewise linear or cubic interpolation. For a sequence $S = (x^0, \ldots, x^k)$ we denote

$$\mathcal{I}_m : [0, k] \ni t \mapsto \mathcal{I}_m^{\mathrm{fl}(t)}(\mathrm{mod}_1(t)) \in \mathbb{R}^2, \tag{A.2}$$

where $\mathrm{fl}(t)$ is the largest integer $j \leq t$, and $\mathrm{mod}_1(t) = t - \mathrm{fl}(t)$ is the fractional part of t. The subscript $m = 0$ or 1 denotes the two types of interpolation implemented (linear and cubic Bezier curve):

$$\mathcal{I}_0^j : [0,1] \ni t \mapsto (1-t)x^j + tx^{j+1} \in \mathbb{R}^2, \qquad \mathcal{I}_1^j : [0,1] \ni t \mapsto \sum_{\ell=0}^{3} \binom{3}{\ell} t^\ell (1-t)^{3-\ell} x^{j,\ell} \in \mathbb{R}^2$$

for $j = 0, \ldots, k-1$, where in $\mathcal{I}_1$ (using $|\cdot|$ for the Euclidean norm)

$$x^{j,0} = x^j, \qquad x^{j,1} = x^j + \nu |x^{j+1} - x^j|\, b^j, \qquad x^{j,2} = x^{j+1} - \nu |x^{j+1} - x^j|\, b^{j+1}, \qquad x^{j,3} = x^{j+1}.$$

The vectors $b^j = n(n(x^{j+1} - x^j) + n(x^j - x^{j-1}))$ are the normalised bisector between $x^{j+1} - x^j$ and $x^j - x^{j-1}$ ($b^0 = b^k = 0$), where we write and $n(v) = v/|v|$ for the normalisation of a vector. The curve $\mathcal{I}_0$ is continuous on $[0, k]$ but has corners with non-zero angles at each integer value of t, at point x^j with $j \in \{1, \ldots, k-1\}$. The curve $\mathcal{I}_1$ is continuously differentiable as a curve since the tangent at point x^j is parallel to b^j in both directions (the parametrisation is singular in the end points $t = 0$ and $t = k$ and continuous but not differentiable in t at $t = 1, \ldots, k-1$). The form of $\mathcal{I}_1$ is the composite cubic Bezier form with $x^{j,\ell}$ as guides. We choose the fraction $\nu = 1/3$ as default in our implementation. Smaller ν brings the piecewise cubic form closer to the piecewise linear form $\mathcal{I}_0$ (in a different parametrisation in t). The particular choice of guides, using the bisector b^j, ensures that each segment $\mathcal{I}_1^j$ is a graph over the abscissa $t \mapsto x^j + t(x^{j+1} - x^j)$. Using the cubic interpolation $\mathcal{I}_1$ may not make the results accurate to a higher interpolation order since the guides are chosen arbitrary. However, the algorithm based on $\mathcal{I}_0$ is more susceptible to failure when angles between successive points in S become too sharp. All following elements of the SC algorithm are independent of the interpolation type m, such that we drop the subscript m from $\mathcal{I}$.

Parametric curves as local graphs We transform the representation of a curve in the parametric form (A.2) to make it easier to determine intersections between the curve and another curve (an image of a small circle under the implicitly defined map). Consider a sequence of points $S = (x^0, \ldots, x^k)$ and its interpolation $\mathcal{I}$, and an arbitrary index $j \in \{0, \ldots, k-1\}$. We represent the curve as a graph locally over the abscissa $t \mapsto x^j + t(x^{j+1} - x^j)$, and then define a signed distance from this graph for root finding in the following way. Define the Householder reflection

$$Q_j = I - 2v^j (v^j)^T \in \mathbb{R}^{2\times 2}, \quad \text{where} \quad v^j := n(x^{j+1} - x^j - |x^{j+1} - x^j| e^1) \in \mathbb{R}^2,$$

using e^ℓ for the ℓth unit vector (Q_j is orthogonal). With the help of Q_j we define the orthogonal projections (and translations)

$$\tau_j : \mathbb{R}^2 \ni x \mapsto (e^1)^T Q_j [x - x^j] \in \mathbb{R}, \qquad \sigma_j : \mathbb{R}^2 \ni x \mapsto (e^2)^T Q_j [x - x^j] \in \mathbb{R},$$

such that $\tau_j(x)$ is the coordinate of x along the abscissa and $\sigma_j(x)$ the coordinate along the ordinate. We pick a small ϵ ($\epsilon = 0.25$ by default in the implementation) and define the indices low_j and $\mathrm{up}_j \in \{0, \ldots, k-1\}$ such that $\mathrm{low}_j \le j \le \mathrm{up}_j$,

$$\begin{aligned} (e^1)^T Q_j [x^{\ell+1} - x^\ell] &\ge \epsilon |x^{\ell+1} - x^\ell| && \text{for all } \ell \in \{\mathrm{low}_j, \ldots, \mathrm{up}_j\}, \text{ and} \\ (e^1)^T Q_j b^\ell &> \epsilon |b^\ell| && \text{for all } \ell \in \{\mathrm{low}_j, \ldots, \mathrm{up}_j + 1\} \end{aligned}$$

(the latter is required only when using the cubic interpolation $\mathcal{I}_1$). By construction of the interpolations $\mathcal{I}$ the range $\{\mathrm{low}_j, \ldots, \mathrm{up}_j\}$ contains at least j. The conditions ensure that $\sigma_j(\mathcal{I}([\mathrm{low}_j, \mathrm{up}_j + 1]))$ is locally a graph over $\tau_j(\mathcal{I}([\mathrm{low}_j, \mathrm{up}_j + 1]))$ for τ between $\tau^j_{\mathrm{low}} := \tau_j(x^{\mathrm{low}_j})$ and $\tau^j_{\mathrm{up}} := \tau_j(x^{\mathrm{up}_j + 1})$. In particular, the set of $\tau_j(x^\ell)$ is strictly increasing for $\ell \in \{\mathrm{low}_j, \ldots, \mathrm{up}_j + 1\}$. For $\hat\tau \in [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}]$, we first pick an index $\ell \in \{\mathrm{low}_j, \ldots, \mathrm{up}_j + 1\}$ such that $\tau_j(x) \in [\tau_j(x^\ell), \tau_j(x^{\ell+1})]$. Then we determine the unique root $\hat t$ in the interval $[0, 1]$ of $t \mapsto \mathcal{I}^\ell(t) - \hat\tau$. (For interpolation order $m = 0$ this is an affine equation, for $m = 1$ this is a cubic polynomial equation with a guaranteed unique root in $[0, 1]$.) Then, with $\hat\sigma = \sigma_j(\mathcal{I}^\ell(\hat t))$, the pair $(\hat\tau, \hat\sigma)$ are the τ_j and σ_j projections of the point $\mathcal{I}^\ell(\hat t)$ on the curve $\mathcal{I}([\mathrm{low}_j, \mathrm{up}_j + 1])$, thus, establishing that the curve $\mathcal{I}([\mathrm{low}_j, \mathrm{up}_j + 1])$ is a graph of the function

$$g_j : [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}] \ni \hat\tau \mapsto \hat\sigma \in \mathbb{R}.$$

We extend the function g_j linearly as $g_{j,\mathrm{ext}}$ to arguments in $\mathbb{R}$ for $\hat\tau \notin [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}]$ in a way such that g_j is differentiable in τ^j_{low} and τ^j_{up}:

$$g_{j,\mathrm{ext}} : \mathbb{R} \ni \hat\tau \mapsto \begin{cases} g_j(\hat\tau) & \text{if } \hat\tau \in [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}], \\ g_j(\tau^j_{\mathrm{low}}) + g_j'(\tau^j_{\mathrm{low}})[\hat\tau - \tau^j_{\mathrm{low}}] & \text{if } \hat\tau < \tau^j_{\mathrm{low}}, \\ g_j(\tau^j_{\mathrm{up}}) + g_j'(\tau^j_{\mathrm{up}})[\hat\tau - \tau^j_{\mathrm{up}}] & \text{if } \hat\tau > \tau^j_{\mathrm{up}}, \end{cases} \in \mathbb{R}$$

$$g_j'(\tau^j_{\mathrm{low}}) = \frac{(e^2)^T Q_j \left[\left(\mathcal{I}_m^{\mathrm{low}_j} \right)' (0) \right]}{(e^1)^T Q_j \left[\left(\mathcal{I}_m^{\mathrm{low}_j} \right)' (0) \right]}, \qquad g_j'(\tau^j_{\mathrm{up}}) = \frac{(e^2)^T Q_j \left[\left(\mathcal{I}_m^{\mathrm{up}_j} \right)' (1) \right]}{(e^1)^T Q_j \left[\left(\mathcal{I}_m^{\mathrm{up}_j} \right)' (1) \right]}.$$

With the above construction we have, for any index $j \le k - 1$ a function $g_{j,\mathrm{ext}} : \mathbb{R} \to \mathbb{R}$ such that a point x is on the parametric curve $\{\mathcal{I}^\ell(t) : \mathrm{low}_j \le \ell \le \mathrm{up}_j, t \in [0, 1]\}$, if and only if $\tau_j(x) \in [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}]$ and $g_{j,\mathrm{ext}}(\tau_j(x)) - \sigma_j(x) = 0$.

This implies that $g_{j,\mathrm{ext}}(\tau_j(x)) - \sigma_j(x)$ is a signed distance from the curve for any $x \in \mathbb{R}^2$ with $\tau_j(x) \in [\tau^j_{\mathrm{low}}, \tau^j_{\mathrm{up}}]$. This is not the shortest Euclidean distance to the curve, but a distance orthogonal to the axis $t \mapsto x^j + t(x^{j+1} - x^j)$.

We note that the constructions of $\mathcal{I}$ and $g_{j,\text{ext}}$ also work for n-dimensional spaces with $n > 2$. Then σ_j maps into $\mathbb{R}^{n-1}$, such that one should replace e^2 by $[e^2, \dots, e^n] \in \mathbb{R}^{n\times(n-1)}$ in all expressions above. Furthermore, the function $g_{j,\text{ext}}$ is independent of the parameterisations of the curve segments, $\mathcal{I}^\ell$. So, any reparameterisation of the curve $\mathcal{I}([0,k])$ would give identical sets of local graph functions $g_{j,\text{ext}}$.

Modified SC algorithm for implicit maps We assume the following situation after step k of the SC algorithm. We have grown the stable curve starting from $x_L^0 \in \operatorname{dom} L$ up to some end point $x^k \in \operatorname{dom} L$ and stored the curve in the form of a sequence of points $S_L^k = (x_L^0, x_L^1, \dots, x_L^k) \subset \operatorname{dom} L \subset \mathbb{R}^2$. We modify the algorithm by also storing the image of the curve S_L^k, $S_R^k = (x_R^0, \dots, x_R^k) \subset \operatorname{rg} R$ where $x_R^j = RM^\ell L x_L^j$ for all $j = 0, \dots, k$. We also have a pointer $j_k \in \{1, \dots, k\}$ after step k. This pointer stores to which segment of S_R^k the point $x^k \in S_L^k$ is mapped by $RM^\ell L$, up to some tolerance `tol`: so there exists a t_k (not stored) such that $RM^{\ell+1} L x_L^k = \mathcal{I}^{j_k}(t_k; S_R^k)$. In practice the equality is only met up to some tolerance, so $|RM^{\ell+1} L x_L^k - \mathcal{I}^{j_k}(t_k; S_R^k)| <$ `tol`. We added the argument S_R^k to $\mathcal{I}^{j_k}$ to clarify from which sequence the interpolated curve is generated. We will not add the generating sequence S_R^k as an argument to $g_{j,\text{ext}}$, τ_j and σ_j below, since all interpolations and transformations all refer to the generating sequence S_R^k.

Using the above notation and depending on the method parameters listed in Chapter A.1, we describe the iterative Search Circle Algorithm (England et al., 2004) and its modification as follows.

parameter	default value	description
$\alpha_{\max}$	0.2	maximal desired angle between successive points of S_L^k
$[\alpha\Delta]_{\max}$	10^{-2}	maximal admissible product of angle between successive points of S_L^k and search radius
$\alpha_{\min}$	0.05	angle admissible for increasing the next search radius Δ
α_{bdmax}	$2\pi/3$	maximal admissible angle between successive points of S_L^k
$[\alpha\Delta]_{\min}$	10^{-4}	product of angle α and search radius Δ admissible for increasing next search radius
Δ_{stop}	10^{-6}	stop if search radius Δ drops below
$\Delta_{\min}$	10^{-4}	permit angles larger than $\alpha_{\max}$ if $\Delta < \Delta_{\min}$
$\Delta_{\max}$	10^{-2}	maximal search radius
`tol`	10^{-6}	tolerance for root bracketing (should not be larger than Δ_{stop})

Table A.1: Algorithm parameters and default values

After step k we have

- a domain sequence $S_L^k = (x_L^0, x^1, \dots, x_L^k) \subset \operatorname{dom} L \subset \mathbb{R}^2$,
- an image sequence $S_R^k = (x_R^0, x_R^1, \dots, x_R^k) \subset \operatorname{rg} R \subset \mathbb{R}^2$,
- a pointer $j_k \in \{0, \dots, k\}$,
- a current radius Δ under consideration.

At step $k+1$ we perform a circle arc search.

1. (*Check if search radius not too small.*) Stop if Δ is smaller than Δ_{stop}.
2. Initialise an empty list of already tested new pointer candidates: $\mathcal{J} = \emptyset$.
3. Set the new provisional pointer into S_R^k as $j = j_k$,
4. (*Check if stuck in a loop.*) If $j \in \mathcal{J}$, go back to step 1 with decreased radius $\Delta \mapsto \Delta/2$. Otherwise, include j into the list $\mathcal{J}$.
5. Define a circle arc ϕ of radius Δ around x_L^k (denoting the normalised previous secant $y_L^k = n(x_L^k - x_L^{k-1})$, recalling that $n(v) = v/|v|$):
$$\phi : (-\alpha_{\text{bd}}, \alpha_{\text{bd}}] \ni \alpha \mapsto x_L^k + \Delta \left[y_L^k \cos\alpha + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} y_L^k \sin\alpha \right] \in \operatorname{dom} L \subset \mathbb{R}^2.$$
6. (*Construct one-dimensional zero problem*) We construct $g_{j,\text{ext}}$, τ_j, σ_j, τ_{low}^j, τ_{up}^j for S_R^k and pointer j, and consider the one-dimensional root-finding problem for
$$\gamma : \alpha \mapsto g_{j,\text{ext}}\left(\tau_j\left(x_R(\alpha)\right)\right) - \sigma_j\left(x_R(\alpha)\right), \text{ where } x_R(\alpha) = RM^{\ell+1}L\phi(\alpha)$$
By construction of $g_{j,\text{ext}}$, a root α of γ with $\tau_j(x_R(\alpha)) \in [\tau_{\text{low}}^j, \tau_{\text{up}}^j]$ corresponds to a point $\phi(\alpha)$ that maps into the curve $\mathcal{I}_m([\text{low}_j, \text{up}_j + 1])$ generated from S_R^k under the implicitly defined map $\mathcal{M}_\ell$. Note the power $\ell + 1$ in the definition of $x_R(\alpha)$.
7. (*Find root*) We gradually increase α_{bd}, starting from $\alpha_{\max}$, up to α_{bdmax}, until $\gamma(-\alpha_{\text{bd}})$ and $\gamma(\alpha_{\text{bd}})$ have the different signs. If this is not the case for any α_{bd} up to α_{bdmax} we go back to step 1, decreasing the radius $\Delta \mapsto \Delta/2$. Otherwise, we find a root $\alpha_* \in [-\alpha_{\text{bd}}, \alpha_{\text{bd}}]$ of γ by bracketing, up to tolerance `tol`.
8. (*Adjust pointer down*) If $\tau_j(x_R(\alpha_*)) < \tau_{\text{low}}^j$ we decrease the pointer j to $j-1$, and go back to step 4.
9. (*Adjust pointer up*) If $\tau_j(x_R(\alpha_*)) > \tau_{\text{up}}^j$ we increase the pointer j to $j+1$, and go back to step 4.
10. (*Acceptance*) If $|\alpha_*| \le \alpha_{\max}$ and $|\alpha_*|\Delta < [\alpha\Delta]_{\max}$, we accept α_* as an admissible angle for the next point on the stable manifold.
11. (*Acceptance despite sharp angle*) If $|\alpha_*| \in (\alpha_{\max}, \alpha_{\text{bdmax}}]$ and $\alpha_*\Delta < [\alpha\Delta]_{\max}$ and $\Delta < \Delta_{\min}$, also accept α_* as the next admissible angle.
12. (*Step rejection*) If α_* is not accepted, go back to step 1, decreasing the radius $\Delta \mapsto \Delta/2$.

13. (*Include new points into sequences*) $x_L^{k+1} = \phi(\alpha_*)$ and $x_R^{k+1} = RM^\ell L\phi(\alpha_*)$. Note that the power of M is only ℓ in the definition of x_R^{k+1}. This extends the sequences defining the manifold: $S_L^{k+1} = (S_L^k, x_L^{k+1})$ and $S_R^{k+1} = (S_L^k, x_R^{k+1})$.

14. (*Increase search radius*) If $|\alpha_*| < \alpha_{\min}$ and $|\alpha_*|\Delta < [\alpha\Delta]_{\min}$, increase $\Delta \mapsto 2\Delta$.

15. (*Update initial pointer*) Set the initial pointer for the next step to $j_{k+1} = j$.

Steps 1, 3 5, 7–12, 14 and 15 are identical to the original SC algorithm by England et al. (2004). Steps 2 and 4 were included to avoid occurrences of infinite loop repetitions. The zero problem in step 6 and the addition of the new point in step 13 have been modified to take into account the implicit nature of the map $\mathcal{M}_\ell$: step 6 uses $RM^{\ell+1}L\phi(\alpha)$, while step 13 uses $RM^\ell L\phi(\alpha_*)$.

Furthermore, if one considers the count of evaluations of the map M as the main computational expense, one may increase efficiency by storing the values $x_R(\alpha)$ of all α tested on a circle arc of radius Δ throughout multiple loops between steps 4 and step 8 and use a super-linearly convergent bracketing scheme in step 7. The didactic code example uses the modified Regula Falsi.

Appendix B

Periodically forced DDE model: dependence on forcing phase

Here we consider the forced scalar DDE

$$\dot{X}(t) = -pX(t-\tau) + rX(t) - sX(t-\tau)^2 - X(t-\tau)^2X(t) - uF_P(t), \tag{B.1}$$

where

$$F_P(t) = \sin((2\pi/T)\,t - \phi), \quad \phi \in [0, 2\pi]. \tag{B.2}$$

Figure B.1 shows the effect of a phase shift on the basin of attraction for the small-amplitude periodic orbit in the plane $L([-0.65, -0.05] \times [-0.65, -0.05])$. We see that the phase shift can cause the initial conditions to cross the boundary of basins of attraction. This is the same mechanism that is seen for keeping the phase constant and increasing the forcing strength u (see Chapter 5, Section 5.3.3).

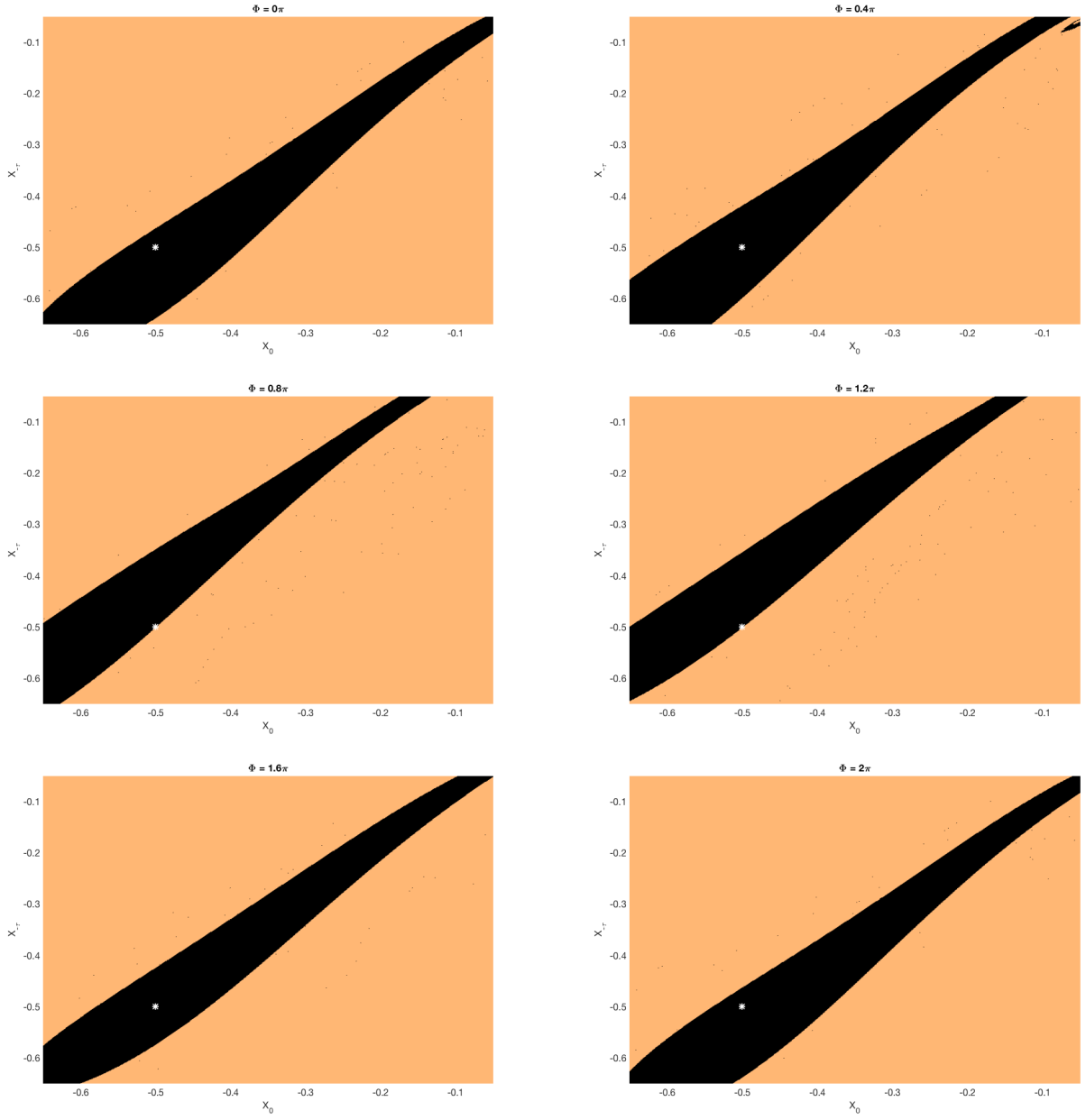


Figure B.1: Basins of attraction for stable solutions of (B.1) under varying ϕ. Black indicates attraction to small-amplitude periodic orbit and beige indicates attraction to large-amplitude quasiperiodic or chaotic response. (a) $\phi = 0$, (b) $\phi = \frac{1\pi}{5}$, (c) $\phi = \frac{2\pi}{5}$, (d) $\phi = \frac{3\pi}{5}$, (e) $\phi = \frac{4\pi}{5}$, (f) $\phi = 2\pi$; Other parameters: $p = 0.95$, $r = s = 0.8$, $u = 0.09$, $T = 4.1$.

Bibliography

H. Alkhayuon, P. Ashwin, L. Jackson, C. Quinn, and R. Wood. Basin bifurcations, oscillatory instability and rate-induced thresholds for amoc in a global oceanic box model. *In preparation*, 2018.

L. S. Andersson and P. A. Lundberg. Delayed albedo effects in a zero-dimensional climate model. *Journal of the atmospheric sciences*, 45(16):2294–2305, 1988.

P. Ashwin and P. Ditlevsen. The middle pleistocene transition as a generic bifurcation on a slow manifold. *Climate Dynamics*, 45(9-10):2683–2695, 2015.

P. Ashwin, S. Wieczorek, R. Vitolo, and P. Cox. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Phil. Trans. R. Soc. A*, 370(1962): 1166–1184, 2012.

P. Ashwin, C. Perryman, and S. Wieczorek. Parameter shifts for nonautonomous systems in low dimension: bifurcation-and rate-induced tipping. *Nonlinearity*, 30(6):2185, 2017.

P. Ashwin, C. D. Camp, and A. S. von der Heydt. Chaotic and non-chaotic response to quasiperiodic forcing: limits to predictability of ice ages paced by milankovitch forcing. *Dynamics and Statistics of the Climate System*, 2018.

K. Bar-Eli and R. J. Field. Earth-average temperature: A time delay approach. *Journal of Geophysical Research: Atmospheres*, 103(D20):25949–25956, 1998.

D. S. Battisti and A. C. Hirst. Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry and nonlinearity. *Journal of the Atmospheric Sciences*, 46(12): 1687–1712, 1989.

A. L. Berger. Long-term variations of daily insolation and quaternary climatic changes. *Journal of the Atmospheric Sciences*, 35(12):2362–2367, 1978.

N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems: a sample-paths approach*. Springer Science & Business Media, 2006.

K. Bhattacharya, M. Ghil, and I. Vulis. Internal variability of an energy-balance model with delayed albedo effects. *Journal of the Atmospheric Sciences*, 39(8):1747–1773, 1982.

B. Bolt, E. H. Nes, S. Bathiany, M. E. Vollebregt, and M. Scheffer. Climate reddening increases the chance of critical transitions. *Nature Climate Change*, 8(6):478, 2018.

C. A. Boulton, L. C. Allison, and T. M. Lenton. Early warning signals of atlantic meridional overturning circulation collapse in a fully coupled climate model. *Nature communications*, 5:5752, 2014.

M. A. Cane, E. Sarachik, D. Moore, and E. Sarachik. The response of a linear baroclinic equatorial ocean to periodic forcing. *Journal of Marine Research*, 39(4):651–693, 1981.

M. A. Cane, M. Münnich, and S. F. Zebiak. A study of self-excited oscillations of the tropical ocean-atmosphere system. part i: Linear analysis. *Journal of the atmospheric sciences*, 47(13):1562–1577, 1990.

T. B. Chalk, M. P. Hain, G. L. Foster, E. J. Rohling, P. F. Sexton, M. P. Badger, S. G. Cherry, A. P. Hasenfratz, G. H. Haug, S. L. Jaccard, et al. Causes of ice age intensification across the mid-pleistocene transition. *Proceedings of the National Academy of Sciences*, 114(50):13114–13119, 2017.

W. Cheng, J. C. Chiang, and D. Zhang. Atlantic meridional overturning circulation (amoc) in cmip5 models: Rcp and historical simulations. *Journal of Climate*, 26(18):7187–7197, 2013.

A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction and the mori–zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences*, 97(7):2968–2973, 2000.

A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D: Nonlinear Phenomena*, 166(3-4):239–257, 2002.

P. Chylek, C. K. Folland, H. A. Dijkstra, G. Lesins, and M. K. Dubey. Ice-core data evidence for a prominent near 20 year time-scale of the atlantic multidecadal oscillation. *Geophysical Research Letters*, 38(13), 2011.

P. U. Clark, D. Archer, D. Pollard, J. D. Blum, J. A. Rial, V. Brovkin, A. C. Mix, N. G. Pisias, and M. Roy. The middle pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pco 2. *Quaternary Science Reviews*, 25(23):3150–3184, 2006.

A. C. Clement and L. C. Peterson. Mechanisms of abrupt climate change of the last glacial period. *Reviews of Geophysics*, 46(4), 2008.

M. Crucifix. Oscillators and relaxation phenomena in pleistocene climate theory. *Phil. Trans. R. Soc. A*, 370(1962):1140–1165, 2012.

K. M. Cuffey, G. D. Clow, R. B. Alley, M. Stuiver, et al. Large arctic temperature change at the wisconsin-holocene glacial transition. *Science*, 270(5235):455, 1995.

H. Dankowicz and F. Schilder. *Recipes for continuation*, volume 11 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013. ISBN 978-1-611972-56-6. doi: 10.1137/1.9781611972573. URL https://doi.org/10.1137/1.9781611972573.

B. De Saedeleer, M. Crucifix, and S. Wieczorek. Is the astronomical forcing a reliable and unique pacemaker for climate? a conceptual model study. *Climate Dynamics*, 40(1-2):273–294, 2013.

T. L. Delworth and M. E. Mann. Observed and simulated multidecadal variability in the northern hemisphere. *Climate Dynamics*, 16(9):661–676, 2000.

C. Deser and J. M. Wallace. Large-scale atmospheric circulation features of warm and cold episodes in the tropical pacific. *Journal of Climate*, 3(11):1254–1281, 1990.

L. Dieci, R. D. Russell, and E. S. Van Vleck. On the compuation of lyapunov exponents for continuous dynamical systems. *SIAM journal on numerical analysis*, 34(1):402–423, 1997.

O. Diekmann, S. van Gils, S. Verduyn Lunel, and H.-O. Walther. *Delay equations*, volume 110 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1995.

H. A. Dijkstra. *Nonlinear physical oceanography: a dynamical systems approach to the large scale ocean circulation and El Nino*, volume 28. Springer Science & Business Media, 2005.

H. A. Dijkstra. Characterization of the multiple equilibria regime in a global ocean model. *Tellus A: Dynamic Meteorology and Oceanography*, 59(5):695–705, 2007.

H. A. Dijkstra. *Dynamical oceanography*. Springer Science & Business Media, 2008.

H. A. Dijkstra. *Nonlinear climate dynamics*. Cambridge University Press, 2013.

K. Engelborghs, T. Luzyanina, and G. Samaey. Dde-biftool v. 2.00: a matlab package for bifurcation analysis of delay differential equations. 2001.

K. Engelborghs, T. Luzyanina, and D. Roose. Numerical bifurcation analysis of delay differential equations using dde-biftool. *ACM Transactions on Mathematical Software (TOMS)*, 28(1):1–21, 2002.

J. P. England, B. Krauskopf, and H. M. Osinga. Computing one-dimensional stable manifolds and stable sets of planar maps without the inverse. *SIAM Journal on Applied Dynamical Systems*, 3(2): 161–190, 2004.

S. K. Falkena, C. Quinn, J. Sieber, J. Frank, and H. A. Dijkstra. Derivation of delay equation climate models using the mori-zwanzig formalism. *In preparation*, 2018.

J. D. Farmer. Chaotic attractors of an infinite-dimensional dynamical system. *Physica D: Nonlinear Phenomena*, 4(3):366–393, 1982.

G. D. Farquhar, J. R. Ehleringer, and K. T. Hubick. Carbon isotope discrimination and photosynthesis. *Annual review of plant biology*, 40(1):503–537, 1989.

Q. Y. Feng and H. Dijkstra. Are north atlantic multidecadal sst anomalies westward propagating? *Geophysical Research Letters*, 41(2):541–546, 2014.

Q. Y. Feng, J. P. Viebahn, and H. A. Dijkstra. Deep ocean early warning signals of an atlantic moc collapse. *Geophysical Research Letters*, 41(16):6009–6015, 2014.

A. Ganopolski and R. Calov. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. *Climate of the Past*, 7(4):1415–1425, 2011.

A. Ganopolski and S. Rahmstorf. Rapid changes of glacial climate simulated in a coupled climate model. *Nature*, 409(6817):153, 2001.

M. Ghil and K. Bhattacharya. An energy-balance model of glaciation cycles. 1979.

M. Ghil, A. Mullhaupt, and P. Pestiaux. Deep water formation and quaternary glaciations. *Climate dynamics*, 2(1):1–10, 1987.

M. Ghil, I. Zaliapin, and B. Coluzzi. Boolean delay equations: A simple way of looking at complex systems. *Physica D: Nonlinear Phenomena*, 237(23):2967–2986, 2008.

H. Gildor and E. Tziperman. A sea ice climate switch mechanism for the 100-kyr glacial cycles. *Journal of Geophysical Research: Oceans*, 106(C5):9117–9133, 2001.

D. Givon, R. Kupferman, and O. H. Hald. Existence proof for orthogonal dynamics and the mori-zwanzig formalism. *Israel Journal of Mathematics*, 145(1):221–241, 2005.

H. Goosse, P. BARRIAT, W. LEFEBVRE, M. LOUTRE, and V. ZUNZ. Brief history of climate: causes and mechanisms. *Introduction to climate dynamics and climate modelling*, pages 109–144, 2010.

A. Gouasmi, E. J. Parish, and K. Duraisamy. A priori estimation of memory effects in reduced-order models of nonlinear systems using the mori–zwanzig formalism. *Proc. R. Soc. A*, 473(2205): 20170385, 2017.

J. K. Hale and S. M. Verduyn-Lunel. *Introduction to functional differential equations*, volume 99. Springer Science & Business Media, 1993.

E. Hawkins, R. S. Smith, L. C. Allison, J. M. Gregory, T. J. Woollings, H. Pohlmann, and B. De Cuevas. Bistability of the atlantic overturning circulation in a global climate model and links to ocean freshwater transport. *Geophysical Research Letters*, 38(10), 2011.

J. D. Hays, J. Imbrie, and N. J. Shackleton. Variations in the earth's orbit: pacemaker of the ice ages. *Science*, 194(4270):1121–1132, 1976.

A. Hidalgo López, M. L. Tello Del Castillo, and J. I. Díaz Díaz. Multiple solutions and numerical analysis to the dynamic and stationary models coupling a delayed energy balance model involving latent heat and discontinuous albedo with a deep ocean. *Proceedings of the Royal Society a-Mathematical Physical And Engineering Sciences*, 470(2170):0–20, 2014.

W. Horsthemke. Noise induced transitions. In *Non-Equilibrium Dynamics in Chemical Systems*, pages 150–160. Springer, 1984.

P. Huybers. Early pleistocene glacial cycles and the integrated summer insolation forcing. *Science*, 313(5786):508–511, 2006.

P. Huybers. Combined obliquity and precession pacing of late pleistocene deglaciations. *Nature*, 480 (7376):229–232, 2011.

P. Huybers and I. Eisenman. Integrated summer insolation calculations. noaa/ncdc paleoclimatology program data contribution# 2006-079, 2006.

P. Huybers and C. Wunsch. Obliquity pacing of the late pleistocene glacial terminations. *Nature*, 434 (7032):491, 2005.

J. Imbrie, A. Berger, E. Boyle, S. Clemens, A. Duffy, W. Howard, G. Kukla, J. Kutzbach, D. Martinson, A. McIntyre, et al. On the structure and origin of major glaciation cycles 2. the 100,000-year cycle. *Paleoceanography*, 8(6):699–735, 1993.

A. Indermühle, T. Stocker, F. Joos, H. Fischer, H. Smith, M. Wahlen, B. Deck, D. Mastroianni, J. Tschumi, T. Blunier, et al. Holocene carbon-cycle dynamics based on co2 trapped in ice at taylor dome, antarctica. *Nature*, 398(6723):121–126, 1999.

L. Jackson, R. Kahana, T. Graham, M. Ringer, T. Woollings, J. Mecking, and R. Wood. Global and european climate impacts of a slowdown of the amoc in a high resolution gcm. *Climate dynamics*, 45(11-12):3299–3316, 2015.

L. Jackson, R. S. Smith, and R. Wood. Ocean and atmosphere feedbacks affecting amoc hysteresis in a gcm. *Climate Dynamics*, 49(1-2):173–191, 2017.

D. Jacob, H. Goettel, J. Jungclaus, M. Muskulus, R. Podzun, and J. Marotzke. Slowdown of the thermohaline circulation causes enhanced maritime climate influence and snow cover over europe. *Geophysical research letters*, 32(21), 2005.

F.-F. Jin. An equatorial ocean recharge paradigm for enso. part ii: A stripped-down coupled model. *Journal of the Atmospheric Sciences*, 54(7):830–847, 1997.

F.-F. Jin and J. D. Neelin. Modes of interannual tropical ocean–atmosphere interaction?a unified view. part i: Numerical results. *Journal of the Atmospheric Sciences*, 50(21):3477–3503, 1993.

H. L. Johnson and D. P. Marshall. A theory for the surface atlantic response to thermohaline variability. *Journal of Physical Oceanography*, 32(4):1121–1132, 2002.

A. Keane, B. Krauskopf, and C. Postlethwaite. Investigating irregular behavior in a model for the el niño southern oscillation with positive and negative delayed feedback. *SIAM Journal on Applied Dynamical Systems*, 15(3):1656–1689, 2016.

Y. Kevrekidis and G. Samaey. Equation-free multiscale computation: algorithms and applications. *Review of Physical Chemistry*, 60:321–344, 2009.

Y. Kevrekidis and G. Samaey. Equation-free modeling. *Scholarpedia*, 5(9):4847, 2010.

B. Krauskopf and J. Sieber. Bifurcation analysis of delay-induced resonances of the el-niño southern oscillation. *Proc. R. Soc. A*, 470(2169):1–18, 2014.

C. Kuehn. A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena*, 240(12):1020–1035, 2011.

C. Kuehn. *Multiple Time Scale Dynamics*. Springer, New York, 2015. ISBN 3319123165. doi: 10.1007/978-3-319-12316-5. URL https://books.google.com/books?id=7fjOBgAAQBAJ{&}pgis=1.

Y. A. Kuznetsov. *Elements of Applied Bifurcation Theory (2nd Ed.)*. Springer-Verlag, Berlin, Heidelberg, 1998. ISBN 0-387-98382-1.

T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber. Tipping elements in the earth's climate system. *Proceedings of the national Academy of Sciences*, 105(6):1786–1793, 2008.

L. E. Lisiecki and M. E. Raymo. A pliocene-pleistocene stack of 57 globally distributed benthic δ18o records. *Paleoceanography*, 20(1), 2005.

L. E. Lisiecki and M. E. Raymo. Plio–pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, 26(1):56–69, 2007.

V. Lucarini and P. H. Stone. Thermohaline circulation stability: A box model study. part i: Uncoupled model. *Journal of climate*, 18(4):501–513, 2005.

K. Maasch. Statistical detection of the mid-pleistocene transition. *Climate dynamics*, 2(3):133–143, 1988.

K. A. Maasch and B. Saltzman. A low-order dynamical model of global climatic variability over the full pleistocene. *Journal of Geophysical Research: Atmospheres*, 95(D2):1955–1963, 1990.

D. R. MacAyeal. The basal stress distribution of ice stream e, antarctica, inferred by control methods. *Journal of Geophysical Research: Solid Earth*, 97(B1):595–603, 1992.

J. Mallet-Paret and G. R. Sell. The poincaré–bendixson theorem for monotone cyclic feedback systems with delay. *Journal of differential equations*, 125(2):441–489, 1996.

J. Mallet-Paret and H. L. Smith. The poincaré-bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations*, 2(4):367–421, 1990.

J. Marotzke and J. Willebrand. Multiple equilibria of the global thermohaline circulation. *Journal of Physical Oceanography*, 21(9):1372–1385, 1991.

M. Milankovitch. History of radiation on the earth and its use for the problem of the ice ages. *K. Serb. Akad. Beogr*, 1941.

H. Mori. Transport, collective motion, and brownian motion. *Progress of theoretical physics*, 33(3): 423–455, 1965.

G. P. Morriss and D. J. Evans. *Statistical Mechanics of Nonequilbrium Liquids*. ANU Press, 2013.

D. Paillard. The timing of pleistocene glaciations from a simple multiple-state climate model. *Nature*, 391(6665):378–381, 1998.

D. Paillard. Glacial cycles: toward a new paradigm. *Reviews of Geophysics*, 39(3):325–346, 2001.

D. Paillard. Quaternary glaciations: from observations to theories. *Quaternary Science Reviews*, 107: 11–24, 2015.

D. Paillard and F. Parrenin. The antarctic ice sheet and the triggering of deglaciations. *Earth and Planetary Science Letters*, 227(3):263–271, 2004.

A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer Science & Business Media, 2012.

L. Perko. *Differential equations and dynamical systems*, volume 7. Springer Science & Business Media, 2013.

N. G. Pisias and T. Moore. The evolution of pleistocene climate: a time series approach. *Earth and Planetary Science Letters*, 52(2):450–458, 1981.

C. Quinn, J. Sieber, and A. S. von der Heydt. Effects of periodic forcing on a paleoclimate delay model. *preprint arXiv:1808.02310*, 2018a.

C. Quinn, J. Sieber, A. S. von der Heydt, and T. M. Lenton. The mid-pleistocene transition induced by delayed feedback and bistability. *Dynamics and Statistics of the Climate System*, 3(1):1–18, 2018b. doi: 10.1093/climsys/dzy005.

S. Rahmstorf. On the freshwater forcing and transport of the atlantic thermohaline circulation. *Climate Dynamics*, 12(12):799–811, 1996.

Z. Ratajczak, P. D'odorico, S. L. Collins, B. T. Bestelmeyer, F. I. Isbell, and J. B. Nippert. The interactive effects of press/pulse intensity and duration on regime shifts at multiple scales. *Ecological Monographs*, 87(2):198–218, 2017.

A. J. Ridgwell, A. J. Watson, and M. E. Raymo. Is the spectral signature of the 100 kyr glacial cycle consistent with a milankovitch origin? *Paleoceanography*, 14(4):437–440, 1999.

P. Ritchie, O. Karabacak, and J. Sieber. Inverse-square law between time and amplitude for crossing tipping thresholds. 2017. URL http://arxiv.org/abs/1709.02645.

J. Rodriguez, R. Wood, R. Smith, and E. Hawkins. Estimating present and future thresholds of the atlantic meridional overturning circulation from observable ocean variables. *preprint*, 2017.

G. H. Roe and M. B. Baker. Why is climate sensitivity so unpredictable? *Science*, 318(5850):629–632, 2007.

W. F. Ruddiman and A. McIntyre. Oceanic mechanisms for amplification of the 23,000-year ice-volume cycle. *Science*, 212(4495):617–627, 1981.

M. J. Salinger. Climate variability and change: past, present and future?an overview. In *Increasing climate variability and change*, pages 9–29. Springer, 2005.

B. Saltzman and K. A. Maasch. Carbon cycle instability as a cause of the late pleistocene ice age oscillations: modeling the asymmetric response. *Global biogeochemical cycles*, 2(2):177–185, 1988.

B. Saltzman and K. A. Maasch. A first-order global model of late cenozoic climatic change ii. further analysis based on a simplification of co 2 dynamics. *Climate Dynamics*, 5(4):201–210, 1991.

M. Scheffer. *Critical transitions in nature and society*, volume 16. Princeton University Press, 2009.

M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara. Early-warning signals for critical transitions. *Nature*, 461 (7260):53, 2009.

F. Sévellec and T. Huck. Theoretical investigation of the atlantic multidecadal oscillation. *Journal of Physical Oceanography*, 45(9):2189–2208, 2015.

J. P. Severinghaus, T. Sowers, E. J. Brook, R. B. Alley, and M. L. Bender. Timing of abrupt climate change at the end of the younger dryas interval from thermally fractionated gases in polar ice. *Nature*, 391(6663):141–146, 1998.

G. Shaffer, S. M. Olsen, and C. J. Bjerrum. Ocean subsurface warming as a mechanism for coupling dansgaard-oeschger climate cycles and ice-rafting events. *Geophysical Research Letters*, 31(24), 2004.

J. Sieber, K. Engelborghs, T. Luzyanina, G. Samaey, and D. Roose. Dde-biftool manual-bifurcation analysis of delay differential equations. *arXiv preprint arXiv:1406.7144*, 2014.

J. Sieber, C. Marschler, and J. Starke. Convergence of equation-free methods in the case of finite time scale separation with application to stochastic systems. *arXiv preprint arXiv:1701.08999*, 2017.

H. Smith. *An introduction to delay differential equations with applications to the life sciences*, volume 57. Springer Science & Business Media, 2010.

R. S. Smith. The famous climate model (versions xfxwb and xfhcc): description update to version xdbua. *Geoscientific Model Development*, 5(1):269–276, 2012. doi: 10.5194/gmd-5-269-2012. URL https://www.geosci-model-dev.net/5/269/2012/.

R. S. Smith, J. M. Gregory, and A. Osprey. A description of the famous (version xdbua) climate model and control run. *Geoscientific Model Development*, 1(1):53–68, 2008.

G. Stépán. *Retarded dynamical systems: stability and characteristic functions*. Longman Scientific & Technical, 1989.

T. F. Stocker and D. G. Wright. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature*, 351(6329):729, 1991.

H. Stommel. Thermohaline convection with two stable regimes of flow. *Tellus*, 13(2):224–230, 1961.

M. J. Suarez and P. S. Schopf. A delayed action oscillator for enso. *Journal of the Atmospheric Sciences*, 45(21):3283–3287, 1988.

R. T. Sutton and D. L. Hodson. Atlantic ocean forcing of north american and european summer climate. *science*, 309(5731):115–118, 2005.

L. A. Te Raa and H. A. Dijkstra. Instability of the thermohaline ocean circulation on interdecadal timescales. *Journal of Physical Oceanography*, 32(1):138–160, 2002.

S. Titz, T. Kuhlbrodt, and U. Feudel. Homoclinic bifurcation in an ocean circulation box model. *International Journal of Bifurcation and Chaos*, 12(04):869–875, 2002a. doi: 10.1142/S0218127402004759. URL https://doi.org/10.1142/S0218127402004759.

S. Titz, T. Kuhlbrodt, S. Rahmstorf, and U. Feudel. On freshwater-dependent bifurcations in box models of the interhemispheric thermohaline circulation. *Tellus A*, 54(1):89–98, 2002b. doi: 10.1034/j.1600-0870.2002.00303.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0870.2002.00303.x.

E. Tziperman and H. Gildor. The stabilization of the thermohaline circulation by the temperature–precipitation feedback. *Journal of Physical Oceanography*, 32(9):2707–2714, 2002.

E. Tziperman, L. Stone, M. A. Cane, and H. Jarosh. El niño chaos: Overlapping of resonances between the seasonal cycle and the pacific ocean-atmosphere oscillator. *Science*, 264(5155):72–74, 1994.

E. Tziperman, M. A. Cane, S. E. Zebiak, Y. Xue, and B. Blumenthal. Locking of el niño's peak time to the end of the calendar year in the delayed oscillator picture of enso. *Journal of climate*, 11(9): 2191–2199, 1998.

M. Vellinga and R. A. Wood. Global climatic impacts of a collapse of the atlantic thermohaline circulation. *Climatic change*, 54(3):251–267, 2002.

A. J. Weaver and T. M. Hughes. Rapid interglacial climate fluctuations driven by north atlantic ocean circulation. *Nature*, 367(6462):447, 1994.

M. S. Williamson, M. Collins, S. S. Drijfhout, R. Kahana, J. V. Mecking, and T. M. Lenton. Effect of amoc collapse on enso in a high resolution general circulation model. *Climate Dynamics*, 50(7-8): 2537–2552, 2018.

T. Woollings, J. M. Gregory, J. G. Pinto, M. Reyers, and D. J. Brayshaw. Response of the north atlantic storm track to climate change shaped by ocean–atmosphere coupling. *Nature Geoscience*, 5(5):313, 2012.

C. Wunsch. What is the thermohaline circulation? *Science*, 298(5596):1179–1181, 2002.

P. Yiou, M. Ghil, J. Jouzel, D. Paillard, and R. Vautard. Nonlinear variability of the climatic system from singular and power spectra of late quaternary records. *Climate Dynamics*, 9(8):371–389, 1994.

I. Zaliapin and M. Ghil. A delay differential model of enso variability, part 2: Phase locking, multiple solutions, and dynamics of extrema. *arXiv preprint arXiv:1003.0028*, 2010.

S. E. Zebiak and M. A. Cane. A model el niñ–southern oscillation. *Monthly Weather Review*, 115(10): 2262–2278, 1987.

R. Zwanzig. Nonlinear generalized langevin equations. *Journal of Statistical Physics*, 9(3):215–220, 1973.