UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

PhD thesis

Karl Hans Mikael Nyman

Bifurcations of relaxation oscillations in models of the middle Pleistocene transition of glacial cycles

Supervisor: Peter Ditlevsen

Handed in: April 15, 2019

Thesis submitted to the PhD School of The Faculty of Science, University of Copenhagen in partial fulfillment of the requirements of the PhD program

Department: Physics of Ice, Climate and Earth, Niels Bohr Institute

Author: Karl Hans Mikael Nyman

Supervisor: Peter D. Ditlevsen, University of Copenhagen
Co-supervisor: Peter Ashwin, University of Exeter

Submission date: April 15, 2019

Title: Bifurcations of relaxation oscillations in models of the middle Pleistocene transition of glacial cycles

Abstract

Since 2.7 million years ago Earth has witnessed waxing and waning of global ice volume called glacial cycles. Initially, these cycles had a period of about 40 thousand years, but around one million years ago the period increased to be closer to 100 thousand years long. This puzzling transition, called the middle Pleistocene transition (MPT) is the focus of this thesis. Studying the MPT is important, since only by understanding past climate well can we be confident in projections of climate into the future.

We study possible *dynamical* mechanisms behind the transition, meaning that we focus on transitions in equations describing conceptual models of the glacial cycles, rather than physical processes.

Results are presented in four parts. In the first part we clearly define a mechanism for the MPT which we term *ramping with frequency locking* (RFL). One novel contribution is to clarify that the abruptness of the MPT under RFL depends on a set of factors, some requiring stronger assumptions than others. Another contribution is to point out that the durations between major glacial terminations increased from ∼40 thousand years to ∼80 thousand years approximately 1200 thousand years ago, after which they increased to ∼120 thousand years over the past two glacial cycles. Such a jump and subsequent slow increase in durations is consistent with RFL, unlike mechanisms in some other published models of glacial cycles. We also demonstrate that RFL is relevant for multi-frequency forcing and outline how to evaluate the potential for observing or including RFL in complex models.

In the second part, we classify all typical bifurcations in fast-slow systems with one fast variable that is “infinitely” faster than another, slow, variable. This classification is relevant for conceptual models of glacial cycles, since it lists all possible causes of the MPT given a clearly defined set of model assumptions. One such bifurcation was the cause of the MPT in [Ashwin and Ditlevsen, 2015].

In the third part, we investigate how certain self-sustained glacial cycle models respond to sinusoidal forcing whose amplitude and period varies over time (e.g. astronomical forcing). We invent a score which demonstrates that the observed sequence of durations between glacial terminations in a model run can, in some cases, be understood from the distribution of durations of the *frozen* model with fixed parameters. We are not aware of any other work which addresses the local-in-time influence of astronomical forcing on durations between terminations in glacial cycle models.

In the fourth part, we clarify the dynamical mechanism behind the MPT in the Paillard 1998 model. It is a particular form of RFL, in which glacial cycles after the MPT are frequency locked to the envelope of eccentricity. We construct a transparent model capturing this mechanism, which has previously only been expressed in words. We also show that the weakening of eccentricity from -1 million years to -600 thousand years plays a major role in reproducing an apparent abrupt increase in glacial cycle period, amplified by a set of specific model assumptions. This casts doubt on the generality of the particular form of RFL used to reproduce the MPT in the Paillard 1998 model.

Resumé

Siden for $2,7$ millioner år siden har jorden været vidne til cyklusser af vækst og aftagen af den globale isvolumen kaldet istidscyklusser. I starten havde disse cyklusser en periode på omkring 40 tusind år, men for omkring en million år siden steg perioden til at være tættere på 100 tusind år lang. Denne forunderlige overgang, kaldet den midt-Pleistocæne overgang (MPT), er denne afhandlings fokus. At studere MPT er vigtigt, da vi kun igennem en god forståelse af klimaet som det var engang, kan sætte vores lid til forudsigelser om dets fremtid.

Vi studerer mulige *dynamiske* mekanismer bag overgangen, hvilket betyder, at vi fokuserer på overgange i ligninger, der beskriver konceptuelle modeller af istidscyklusser, snarere end på fysiske processer.

Resultaterne er præsenteret i fire dele. I første del definerer vi en mekanisme for MPT som vi betegner *rampning med frekvenslåsning* (RFL). Et nyt bidrag er, at klarlægge at MPT's abrupthed under RFL afhænger af et sæt faktorer, hvoraf nogle kræver stærkere antagelser end andre. Et andet bidrag er at påpege, at periodernes varighed mellem de store istidsafslutninger steg fra ∼40 tusind år til ∼80 tusind år for ca. 1200 tusind år siden, hvorefter de igen steg til ∼120 tusind år over de sidste to istidscyklusser. Et sådant spring og efterfølgende langsom stigning i varighed er i overensstemmelse med RFL, i modsætning til mekanismer i andre offentliggjorte modeller af istidscyklusser. Vi demonstrerer også, at RFL er relevant for multifrekvens-forcering og beskriver hvordan man vurderer potentialet for at observere eller inkludere RFL i komplekse modeller.

I anden del klassificerer vi alle typiske bifurkationer i hurtig-langsom-systemer med en hurtig variabel, der er "uendeligt" hurtigere end en anden, langsom variabel. Denne klassifikation er relevant for konceptuelle modeller af istidscyklusser, da den angiver alle mulige årsager til MPT, forudsat et klart defineret sæt af modelantagelser. En sådan bifurkation var årsagen til MPT i [Ashwin and Ditlevsen, 2015].

I tredje del undersøger vi, hvordan visse selvbærende istidscyklusmodeller reagerer på sinusformet forcering, hvis amplitude og periode varierer over tid (fx astronomisk forcering). Vi opfinder en "score", som viser, at den observerede sekvens af varigheder mellem istidsafslutninger i en modelkørsel, i nogle tilfælde kan forstås på baggrund af fordelingen af varigheden af den *frosne* model med faste parametre. Vi har ikke kendskab til nogen anden forskning, som behandler lokal-i-tid indflydelsen af astronomisk forcering på varigheden mellem afslutninger i istidscyklusmodeller.

I fjerde del afklarer vi den dynamiske mekanisme bag MPT i Paillard 1998-modellen. Det er en særlig form for RFL, hvor istidscyklusser efter MPT'en er frekvenslåst til indhyldningskurven af excentriciteten. Vi konstruerer en transparent model, der viser denne mekanisme, som tidligere kun har været beskrevet med ord. Vi viser også, at svækkelsen af excentricitet fra -1 million år til -600 tusind år spiller en vigtig rolle i reproduktionen af en tilsyneladende abrupt stigning i istidscyklusperioden, forstærket af et sæt specifikke modelantagelser. Dette sår tvivl om genereliteten af den specifikke form for RFL, der bruges til at reproducere MPT i Paillard 1998-modellen.

To Charlie

Acknowledgements

I thank Peter Ditlevsen for inspiration, optimism, criticism, and complete academic freedom. I thank Pete Ashwin for persistence, nudging, and new ways of thinking. I am thankful to the taxpayers of the European Union for enabling this research through the ITN CRITICS under the Marie Skłodowska-Curie grant agreement No. 643073, part of the European Union's Horizon 2020 innovation and research programme. I hope that my research will benefit you, if only indirectly. I thank Jeroen Lamb, Martin Rasmussen, all other grant applicants and Chris Richley, for making CRITICS happen. I especially thank Martin Rasmussen for unwavering faith and support. I thank Christian, Chun, Courtney, Damian, Daniele, Flavia, Iacopo, Johannes, Kalle Timperi, Ke, Michael, Moussa, Pablo, Sajjad and Usman of the CRITICS cohort for many good moments academically and otherwise. I thank Abdullah, Hassan, Lamees, Paul and Saad of the Marine Dynamics 301 office, as well as the University of Exeter, for hospitality. I thank everyone at CIC (now PICE) for an open and welcoming environment. I thank especially Sune Rasmussen for sane advice, Tina Bang-Christensen for patience with CWT, Ellen Chrillesen for making sure things are done right, and the MacGyver of Science for shaking myself. I thank Topaz, Air, Kostya and Amethyst of the Eggacy for lulz, and especially Sebastian Jano Nielsen for translating the abstract of this thesis. I also wish to thank my family and friends for patience and support, and last but certainly not the least Carina.

Contents

List of acronyms

kyr Thousand years

Myr Million years

MPT Middle Pleistocene transition

RFL Ramping with frequency locking

Chapter 1

Introduction

Climate change due to increasing concentrations of atmospheric CO_2 is one of the most urgent issues facing mankind in the year 2019. The global warming and increasing sea levels that are likely to follow pose an ecological, economical and societal threat as the livelihood of millions of people may be compromised [IPCC, 2018].

Although state of the art climate models offer climate predictions up to a hundred years in the future [IPCC, 2018], there is a fear that these models may fail to account for rapid transitions, called tipping points and critical transitions, which are currently poorly understood. Such rapid and irreversible shifts are thought to have occurred in Earth's past, such as the End-Permian extinction event 252 million years (Myr) ago and the (hypothesised) Snowball Earth event around 700 Myr ago [Scheffer, 2009]. Despite the obvious complexity of ecological and climatological processes on Earth, some of the fundamental mechanisms behind these rapid transitions have been understood through the use of simple and transparent mathematical models [Scheffer, 2009, Budyko, 1969, Sellers, 1969].

In this thesis we study the middle Pleistocene transition (MPT), arguably a critical transition, around 1 Myr ago when the period of glacial cycles increased from about 40 thousand years (kyr) to about 100 kyr. We model the glacial cycles as a dynamical system, a climate state evolving in time, and analyse the transition using bifurcation theory, a theory that describes abrupt change in state due to small changes to system parameters.

Although the glacial cycles occur on time scales longer than the existence of modern civilisation, we see three reasons why they are important to study. The first is that we want to have confidence in our models of future climate change. Lacking the ability to model past critical transitions, how can we trust that these models are able to predict future ones? The second is understanding for the sake of understanding. Many great discoveries in science have no immediate applications, but were motivated by a quest to expand the body of human knowledge. Lastly, the study of glacial cycles raises mathematical questions that are of more general interest; in what ways can interacting systems evolving on different time scales bifurcate, how do dynamical systems respond to time-varying and irregular input, and how can we detect whether a dynamical mechanism generates a set of data?

This thesis is divided into five parts. In the first part (Chapters 2 to 3) we introduce the problem of the ice ages and the MPT, as well as theory common to the remaining four parts. In the second part (Chapter 7) we present a theory for the MPT based on a slow change in the internal climate state and entrainment to variations in the incoming insolation at Northern latitudes. In the third part (Chapter 8) we set astronomical forcing aside and propose a classification of bifurcation in singular fast-slow systems with one fast variable. One of these bifurcations may be the cause of the MPT. In the fourth part (Chapter 9) we expand on Chapter 7 and present a framework for understanding how a class of self-sustained oscillators respond to modulated sinusoidal forcing, of which insolation variations are a particular kind. In the fifth part (Chapter 10), we study how the MPT is realised in the Paillard 1998 model of glacial cycles. At the time of writing, the second part is a summary of a paper published online in Climate Dynamics, and the third part is a summary of and extension to a paper submitted to Nonlinearity. The fourth and fifth parts are planned to be written up as articles in the future.

Original contributions are predominantly collected in chapters 7 to 10. We will refer to the paper in A as "Paper A" and the paper in A as "Paper B".

If the reader is familiar with paleoclimate proxies and the phenomenology of glacial cycles they can skip past chapters 2 and 3 without loss of continuity. Similarly, if the reader is familiar with dynamical systems, frequency locking, and bifurcation theory they can skip chapters 4 and 5. Likewise, Chapter 6 can be skipped if the reader is well versed in conceptual glacial cycle models.

The only assumed prerequisites are basic calculus, algebra, statistics and spectral analysis. We have attempted to introduce all other concepts sufficiently well that no prior knowledge of them is required.

For the remainder of the thesis, the pronoun "we" will be used. This can mean either "we" as in the author of the thesis and the reader, or "we" as in the author of the thesis and collaborators. In both cases I, the author of the thesis, take full responsibility for any statement or claim written as if originating from "we".

Chapter 2

The glacial cycles

Written account of the idea that parts of the world that are now free from permanent ice cover were once covered with large ice sheets stems from the 18:th century [1][Mathews et al., 1898, Remy and Testut, 2006]. It was first in the 19:th century when it was suggested that parts of Europe had been covered by ice not once, but multiple times, that the concept of glacial cycles was born [Imbrie, 1982].

In the following we give a brief introduction to the theory of glacial cycles. This review is incomplete, but we aim to provide an overview and mention important breakthroughs. The part of the introduction covering the 19:th and early 20:th century is largely adapted from [Imbrie, 1982] (where also historical references can be found).

2.1 Early history and link to insolation variations

Louis Agassiz in 1837 was not the first to propose that Earth had undergone a state of widespread glaciation, but he was the first to gain widespread credit for the hypothesis [Imbrie, 1982]. Evidence for this theory is the large boulders mysteriously scattered in the Alps. Agassiz suggested that they had been brought there by glaciers which had since then melted. Furthermore, he proposed that there had been multiple such glacial cycles, and idea which might have inspired Joséph Adhémar to propose in 1842 that variations in insolation onto Earth could be the cause. This idea was refined by James Croll in 1864, who aided by estimates of variations in eccentricity, the ovalness of Earth's orbit around the sun, proposed that an ice age would occur when Earth's orbit is oval, since then winters are particularly cold. This was later shown to be wrong, however. It was only after 1920 when Milutin Milanković calculated insolation onto Earth with decent accuracy that it became possible to link variation in insolation at different latitudes to the glacial cycles. Over the coming decades, Milanković postulated that the summer insolation in the Northern hemisphere was the key driver of ice ages. However, the insolation curves that he found suggested major glaciations every 40 thousand years or so, as opposed to every 100 thousand years over the past million years, which is closer to the truth.

2.2 Sediment δ^{18}O as a proxy of ice temperature

To ascertain not only that multiple glacial cycles occurred, but when they occurred, required improved sources of data. In 1955 Cesare Emiliani analysed the ratios of the oxygen isotopes O^{16} and O^{18} in the shells of small creatures called foraminifera obtained from twelve

---

[1] "It is to be observed, that the glacier is not level, and all the ice has a motion from the higher parts towards the lower; that is to say, that it slides continually towards the outlets into the valley, which has been remarked by many circumstances. First, by great stones, which have been carried quite into the valley of Chamouny; they showed us one of a very large size, which several old people assured us, that they had been upon the ice." From [Mathews et al., 1898] p. 349

sediment cores, cylindrical samples of the ocean floor [Emiliani, 1955]. The ratio between the elements O^{18}/O^{16} in calcium carbonate CaC_3O reflects the temperature at which the shells are formed, since at colder temperatures O^{18} is more prone to form CaC_3O than O^{16}, while at higher temperatures the difference between the isotopes is smaller. Thus, a higher ratio reflects colder waters and a lower ratio warmer waters. In practice, one compares the isotope ratio in a measured sample with a standard, to produce the isotope ratio anomaly $\delta^{18}O$, which typically is on the order of magnitude 10‰.

2.2.1 Contribution of ice volume to $\delta^{18}O$

However, the oxygen isotope ratio $\delta^{18}O$ also reflects the composition of O^{16} and O^{18} in the water where the shells were formed. Since lighter isotopes are preferentially removed through evaporation and then deposited on ice sheets, more light isotopes stored in ice sheets mean that heavier isotopes are left in the ocean, increasing $\delta^{18}O$. Emiliani down-played this effect, but Nicolas Shackleton argued in 1967 that it forms a significant contribution to $\delta^{18}O$, and thus that $\delta^{18}O$ reflects changes in global ice volume [Shackleton, 1967]. Since then, the relative contribution of temperature and ice volume has been disputed. Attempts at separating the two signals through modelling [Bintanja et al., 2005, Mix and Ruddiman, 1984] and measurement of the ratio between calcium and magnesium Ca/Mg in foraminiferal shells [Katz, 1973, Lear et al., 2000, Sosdian and Rosenthal, 2009], indicate that the relative contribution of temperature and ice volume to $\delta^{18}O$ is not constant over time. However, by-and-large the contributions covary, so variations in $\delta^{18}O$ do seem to represent variations in ice volume.

2.2.2 Relating core depth and age

While on the topic of ice volume proxies, we spend a few words on how age is inferred from depth in a sediment core. A classic method is to infer age at a certain depth from the decay of radioactive isotopes of carbon, thorium, uranium and protactinium [Emiliani, 1955, Sakanoue et al., 1967]. The age is estimated at easily identifiable markers such as magnetic reversals, when the polarity of Earth's inner magnet, and material deposited in the sediment core, changed. Once the age of these marker events has been estimated, one has to assume a (possibly time-varying) sedimentation rate, which when integrated gives a relation between depth and time. Recently, multiple cores have been aggregated in a method called stacking in order to improve age models [Lisiecki and Raymo, 2005, Huybers, 2007, Huybers and Wunsch, 2004]. Lastly, we mention that age models are sometimes tuned by assuming a phase relationship between glacial cycles and insolation variations (e.g. [Lisiecki and Raymo, 2005]). This, however, reduces the ability to infer relationships between the ice volume record and the insolation variations used for tuning [Huybers, 2007].

2.3 The 1970s until present

As dating techniques improved in the 1970s, the confidence in the age models increased and the sequence of ∼100 thousand years (kyr) long glaciations over the past million years became well established [Shackleton and Opdyke, 1973, Broecker and van Donk, 1970]. Along with this, interest in the Milanković's theory rose. In 1976 Hays et al. [1976] performed spectral analysis on a 450 kyr long $\delta^{18}O$ record and concluded that the orbital periods $\sim 21, 41$ kyr and ∼100 kyr present in insolation variations at Northern latitudes were also present in the $\delta^{18}O$ record. This seemed to agree with Milanković's theory. However, the analysis revealed a puzzle that lingers to this day; the dominant frequency in the $\delta^{18}O$ record is ∼100 kyr while the changes in solar output due to the 100 kyr period of eccentricity are negligible. This hinted that non-linear mechanism is responsible for the saw-tooth shaped glacial cycles with slow growth of ice volume and rapid deglaciation. The physical and

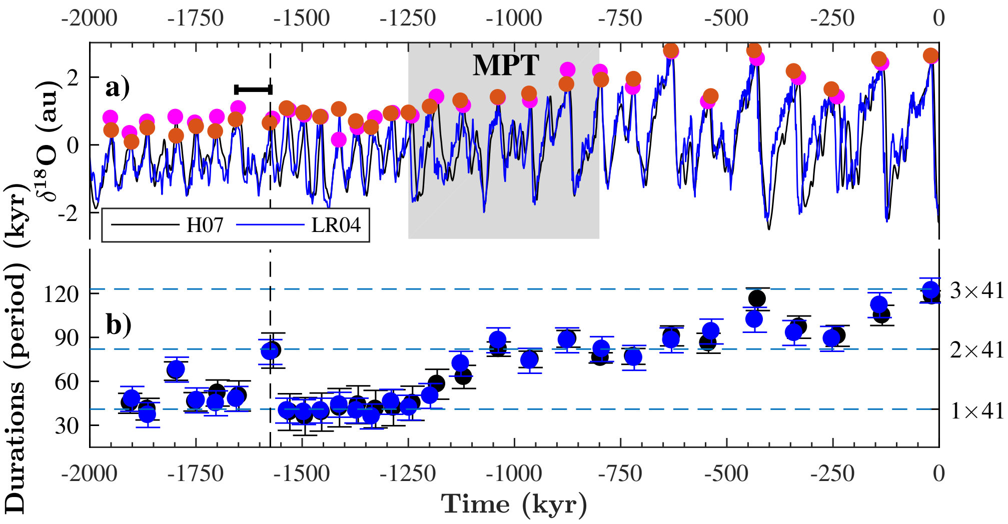


Figure 2.1: Global ice volume and deep ocean temperature proxies. a) LR04 is a stack of δ^{18}O from benthic (bottom dwelling) foraminifera, and H07 is a stack of benthic and planktic (top ocean dwelling) foraminifera [Lisiecki and Raymo, 2005, Huybers, 2007]. Magenta and red dots mark major glacial terminations in the respective stacks. b) Durations between glacial terminations in the two stacks. Error bars show one standard deviation of durations between major glacial terminations (taken to be $\sqrt{2} \cdot 6$ kyr for LR04 and $\sqrt{2}\sigma_{H07}(t)$), where $\sigma_{H07}(t)$ is a time-varying age model error taken from [Huybers, 2007].

dynamical mechanisms reponsible for these glacial cycles occupy researchers and modellers to this day. We delay the discussion of models until Chapter 6, however.

2.3.1 The middle Pleistocene transition

The record in Shackleton and Opdyke [1973] stretched back to 1600 kyr ago, far enough to reveal that the major ∼100 kyr glacial cycles at some point around 1000 kyr ago started to give way to ∼40 kyr cycles, a transition that would later be known as the middle Pleistocene transition (or variations on the name), see Figure 2.1. This point was made once more in Shackleton and Opdyke [1976], Pisias and Moore [1981]. Since then, new observational data has not triggered new ideas on the origins of the middle Pleistocene transition. Therefore most attention has gone into modelling the transition and analysing it statistically, as we do in this thesis. As mentioned before, we treat the modelling of the glacial cycles separately in Chapter 6.

2.3.2 New sources of data

The most prominent sources of new data relating to glacial cycles comes from Antarctic ice cores. The longest set of Vostok ice cores from 1998 stretches back to 420 kyr[Petit et al., 1999] and the Dome C core from 2004 goes back to about 800 kyr [EPICA community members, 2004, Parrenin et al., 2007, Lüthi et al., 2008]. Ice cores, just like sediment cores, contain water molecules whose isotope ratios carry information about climate conditions. However, unlike in sediment cores variations in the isotopic ratios are mainly believed to reflect the temperatures where pricipitation is formed and where it is deposited. Additionally, gas bubbles are trapped in the ice, which enables atmospheric CO_2 and other chemical elements to be measured.

Interestingly, the records revealed that CO_2 oscillates cyclically along with local temperature variations. This implies that ice volume, (local) temperature, and CO_2 are intimately linked and has to be accounted for in a complete model of the glacial cycles [Dijkstra, 2013]. There is currently hope that a new ice core extending over 1 million years back in time can reveal the role of CO_2 across the MPT.

Two stacks of sediment cores (the LR04 and H07 stacks) extending 5 and 2 million years back in time respectively were published in the 2000s [Lisiecki and Raymo, 2005, Huybers, 2007] (Figure 2.1). These stacks are superpositions of individual cores, and therefore dating can be improved. Furthermore, they are thought to represent global ice volume and average deep ocean temperatures which is relevant for modelling the glacial cycles with simple models. Because of their advantages, these cores are commonly used as benchmarks for comparison with models.

2.3.3 A note on terminology

The term *ice age* has been used to denote both the glacial cycles of length ∼40 to ∼120 kyr, which this thesis deal with, in addition to other phenomena such as Dansgaard-Oeschger events, Snowball Earth events, and the “Little Ice Age”. In this thesis, we use exclusively the term glacial cycles.

2.4 Summary

Glacial cydes have been known to exist since the 19:th century. Throughout the 20:th and early 21:st century improved proxies, data from which past climate can be inferred, and methods revealed that glacial cycles had a dominant period of about 40 thousand years (kyr) prior to 1 Million years ago, which then changed to a period of about 100 kyr. This shift in dominant period is called the middle Pleistocene transition. The period 41 thousand years was predicted by Milutin Milanković in the 1940s based on calculations of insolation variation due to variations to Earth’s orbit and tilt. However, the origin of the $\sim 100kyr$ period is still debated. In this thesis we shall encounter different theories for this period, and we shall question the appropriateness of characterising the glacial cycles over the past Million years by the average period $\sim 100kyr$.

Chapter 3

Insolation variations

Adhémar and Croll were first to link variations in insolation to glacial cycles [Imbrie, 1982]. But the usefulness of the insolation fields depend on the quality of the calculation of the underlying equations. This improved in stages, with Milanković (1920), Vernekar (1972), Berger (1976) and Laskar (2004) providing milestones to the present time in 2019, when it is possible to determine insolation fields with decent accuracy up to 50 million years back in time.

In this section we give an introduction to the insolation variations on Earth and the astronomical elements responsible for these variations. First I sketch how the astronomical elements relate to the insolation field, and then we review a discussion about which latitudes and times of the year are the most relevant for the glacial cycles.

3.1 Insolation from orbital parameters

The insolation at any day of the year at a given latitude is given by four parameters; the solar output power S_0, eccentricity (ovalness) of Earth's orbit around the sun e, the tilt of Earths axis of rotation relative to the sun (obliquity) ϵ, and the longitude of perihelion ω (Figure 3.1) [Berger, 1978].

The solar output power S_0 we consider constant on time scales relevant for the glacial cycles.

Obliquity ϵ is the tilt of Earth's axis of rotation relative to the ecliptic plane, the plane in which Earth rotates around the Sun (Figure 3.1 d)). Obliquity changes between ca 22.1° and 24.5° degrees with a period of about 41 kyr.

Eccentricity $e = \sqrt{a^2 - b^2}/a$ relates the lengths of the longer and shorter semi-axes of the orbit of Earth around the sun and is a measure of ovalness (Figure 3.1 a)). Since a is approximately constant on glacial time scales [Berger, 1978], changes to e mostly occur through b. Eccentricity variations are not quite sinusoidal but have a range of periods around 100 and 400 kyr.

The longitude of perihelion ω is the angle between the spring equinox (when days and nights are equally long when progressing towards summer) and perihelion (the point on Earth's orbit closest to the sun) (Figure 3.1 c)). This variable influences how close Earth is to the sun during different seasons. For example, if the North pole points away from the Sun when Earth is the farthest from the Sun, then Northern latitude winters are colder compared to if the North Pole points away from the Sun when closest to it. The longitude of perihelion changes over time due to two precessional effects. Apsidal precession occurs because Earth's orbit around the sun rotates relative to the fixed stars (Figure 3.1 b)). As perihelion and aphelion (where Earth is the farthest from the sun) change position, the relative position of summer of winter on the orbit changes as well. Axial precession occurs because Earth's axis of rotation is not fixed relative to the sun, but rotates (Figure 3.1 e)). For this reason, even with no apsidal precession the seasons would occur at varying

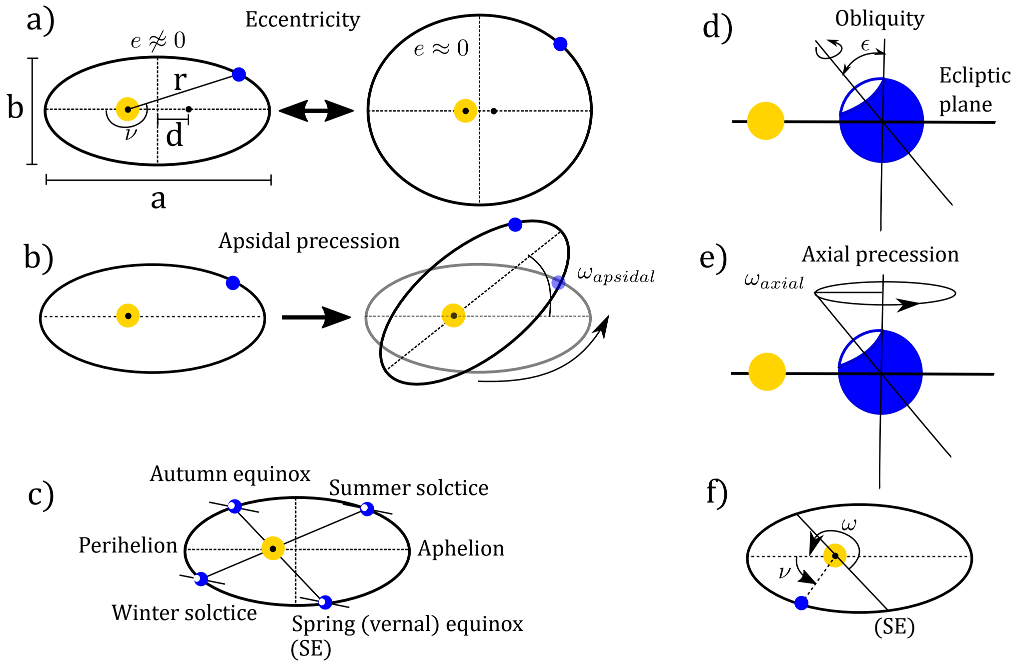


Figure 3.1: Earth's (blue disk) astronomical configuration relative to the sun (yellow disk) (heavily exaggerated relative to reality). a) Earth orbits on an ellipse with the Sun at one of the focal points. The distance between Earth and the Sun is denoted r, a and b are the major and minor semi-axes, d is the distance between the Sun and the centre of the ellipse, and ν is the true anomaly, the counter-clockwise angle between perihelion (the point on the orbit closest to the Sun) and Earth. Eccentricity $e = \sqrt{a^2 - b^2}/a = d/a$ is larger for more oblong ellipses. b) Apsidal precession is the rotation of the apsides (perihelion and aphelion) around the sun, which causes the intensity of the seasons to vary. c) The intensity of the seasons on different latitudes on Earth depends on how close is to the sun at each season. d) Obliquity is the tilt of Earth relative to the ecliptic plane, the plane in which Earth rotates around the Sun. e) Axial precession, the precession of Earth's axis of rotation in the ecliptic plane. f) The longitude of perihelion (the angle between the spring equinox (SE)) and the true anomaly ν, the angle between the Spring equinox and perihelion, whose sum is the true longitude $\lambda = \nu + \omega$.

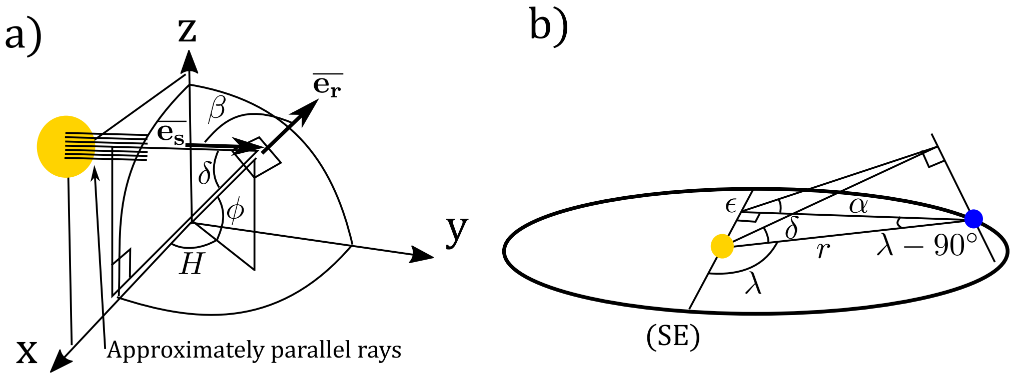


Figure 3.2: Angles in the formula for insolation on a given latitude and hour. a) The azimuth angle β is the angle between the Sun to Earth unit vector $\overline{e_s}$ lying in the (x, z) plane and the outward unit normal vector $\overline{e_r}$ from the surface of Earth, approximated as a sphere. b) A relation between δ, the obliquity ϵ and the true latitude λ can be derived through trigonometry.

distances from the sun. These two effects combined produce insolation variations with an average period of about 23 kyr.

The longitude of perihelion is often combined with eccentricity to form the precessional parameter

$$p(t) = e(t) \sin(\omega(t)),$$

where we note that ω is an angle varying in time and not an angular frequency. This parameter expresses a sine function (often with near constant rate of change of its argument $\omega(t)$) that is modulated by eccentricity. This makes sense, since for a circular orbit $e = 0$, the longitude of perihelion does not influence insolation onto different latitudes of Earth.

The orbital parameters can be combined to form an expression for the insolation at a given latitude and time of the day

$$W(t) = S_a \left(\frac{a}{r}\right)^2 \cos\beta, \tag{3.1}$$

where S_a is the solar constant renormalised to the major semi-axis a of Earth's orbit around the sun, r is the distance from Earth to the Sun, and β is the azimuth angle, the angle between Earth"s outward normal and the incoming rays from the Sun, assumed parallel, (Figure 3.2 a)). The cosine azimuth angel can be computed from dot product of unit vectors [Sproul, 2007], to give

$$\cos\beta = \cos\phi\cos\delta\cos H + \sin\phi\sin\delta,$$

where ϕ is the latitude on earth, δ is the angle of the Sun's rays onto Earth and H is the hourly angle, ranging from 0 to 360°).

The constant S_a is the standard solar constant S_0 renormalised to the major semi-axis a [Berger et al., 2010]

$$S_a = S_0\sqrt{1 - e^2}, \tag{3.2}$$

which includes explicit dependence on the eccentricity.

The radius r can be expressed in the true anomaly ν (the angle between perihelion and Earth), eccentricity e and the length of the major semi-axis a as

$$r = \frac{a(1 - e^2)}{1 - e\cos\nu},$$

by noting that the sun is located at a focal point of the ellipse at a distance $d = ae$ from the centre, and inserting the expression for the position of Earth in (r, ν)-coordinates into the equation for the ellipse (Figure 3.1 a)). The angle between the sun's rays and Earth's equatorial plane δ can be found trigonometrically as

$$\sin\delta = \sin\lambda \sin\epsilon, \tag{3.3}$$

where again ϵ is the obliquity, and where λ is the true longitude, the angle between the spring equinox and Earth (Figure 3.2 b)). The true longitude is related to the true anomaly ν and the longitude of perihelion as

$$\lambda = \nu + \omega,$$

as can be seen from Figure 3.1 d). This allows us to express the instantaneous insolation (3.1) as

$$W(t) = S_a(e)\left(\frac{1 + e\cos(\lambda - \omega)}{1 - e^2}\right)^2 (\cos\phi\cos(\delta(\epsilon,\lambda))\cos H + \sin\phi\sin\epsilon\sin\lambda), \tag{3.4}$$

where we have emphasised the dependence of δ on ϵ and λ (Equation (3.3)) and of S_a on e (Equation (3.2)). This expression can be integrated to derive a number of insolation quantities such as the insolation at summer solstice at 65° North, or the total insolation over one year on the Northern hemisphere. In practice, it is a bit involved, however. The main purpose of deriving (3.4) here is to demonstrate that insolation variations on thousand year time scales are mainly a function of just three parameters: obliquity ϵ, eccentricity e, and precession $e\sin\omega$.

3.2 The most relevant insolation curve

In 1930 Milanković proposed that the most important insolation curve for the glacial cycles is the caloric summer insolation, irradiance integrated over the summer half year where the summer is defined such that any summer day receives more insolation than any day in the winter half [Tzedakis et al., 2017], Figure 3.3b). Ever since, there has been a debate about which scalar insolation curve is the most important for the glacial cycles [Imbrie, 1982]. Milanković argued that the insolation at 65N is the most relevant, since in glacial times the largest and most sensitive ice sheets would lie on the Northern hemisphere (which has more land mass), and their ablation lines (where ice mass has net loss) would lie around 65N.

But since insolation varies on all latitudes and over all times, why is there a focus on identifying a scalar time series as "the" relevant insolation curve? There are two reasons for this. The first is practicality; as the δ^{18}O proxy curve (e.g. Figure 2.1) representing global variations in ice volume is scalar, it is more straightforward to compare it with a scalar insolation curve, rather than a spatially distributed insolation field. Furthermore, if a model does not have spatial extent, then a spatially varying insolation field cannot be part of it. The second is parsimony; one wishes to include only the necessary components in an explanation for the glacial cycles. If indeed integrated summer insolation at 65N can explain the succession of glacial cycles in a satisfactory manner, and the winter insolation at the equator does not add explainatory value, then the equatorial winter insolation should not be included in a model.

An observation that helps the discussion on insolation curves is that almost all insolation curves F averaged over latitudes and times of the year can be well approximated as a linear combination of obliquity ϵ and precession $e\sin(\omega + \phi)$ (normalised to zero mean and unit variance), where ϕ indicates month of the year [Imbrie and Imbrie, 1980]:

$$F(t) = A\epsilon + Be\sin(\omega + \phi),$$

and A and B are constants. Therefore, insolation curves can by-and-large be characterised by their relative proportions of obliquity and precession.

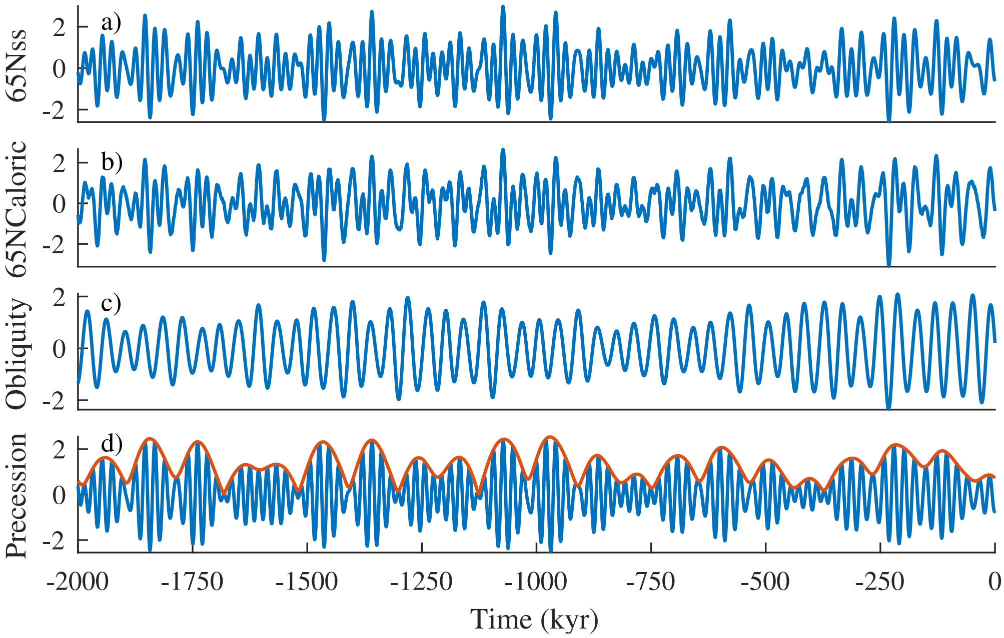


Figure 3.3: Insolation and astronomical parameter curves. a) Summer solstice insolation at 65 degrees North. b) Caloric summer insolation at 65 North, the insolation integrated over the summer, defined as the half year such that every summer day receives more insolation than any winter day. c) Obliquity, the tilt of Earth relative to its plane of rotation around the sun. d) Precession, an astronomical quantity related to how close the Earth is to the Sun during different seasons. Eccentricity, which is the ovalness of Earth's orbit around the Sun, forms an envelope of precession which is represented as a red curve. All curves are normalised to zero mean and unit variance

Generally, precession plays a bigger role closer to the equator and obliquity plays a bigger role closer to the poles. However, it also matters whether one considers the maximum insolation in a year, or integrated insolation over part of a year or above some threshold. In general, the larger part of the year one integrates over, the larger role does obliquity play. The reason for this is Kepler's second law, which implies that summers are shorter when Earth' is closer to the sun and longer when it is far away, negating the precessional effect of seasonal contrasts [Huybers, 2006].

Different reasoning motivates the use of different insolation curves. If insolation extremes are most important for ice sheet stability, then insolation at summer solstice should be more appropriate. On the other hand, if the integrated insolation over the year, or insolation above a low threshold intensity is most important, then obliquity should play a bigger role [Huybers, 2006]. Furthermore, the latitude of relevant ice sheets also matters; if ice sheets are more sensitive at low latitudes, then the relevant insolation curve contains more precession relative to obliquity.

In recent years, three insolation curves have seen the most use in models. One is the summer solstice insolation at 65 degrees North which is dominated by precession (e.g. [Imbrie et al., 2011, de Saedeleer et al., 2013, Ashwin and Ditlevsen, 2015]), another is the caloric summer insolation (e.g. [Tzedakis et al., 2017, Feng and Bailer-Jones, 2015]), and a third is pure obliquity [Huybers, 2007, Daruka and Ditlevsen, 2015]. We show these curves in Figure 3.3, together with orbital parameters.

Moreover, simplifications or approximations to insolation curves are sometimes used. For instance, [de Saedeleer et al., 2013] approximated the summer solstice insolation at 65 degrees North as a trigonometric sum containing 35 frequencies. Le Treut and Ghil [1983] used a combination of sines at periods that dominate precession. Daruka and Ditlevsen [2015], Ditlevsen and Ashwin [2018], Ashkenazy [2006] used a single sine curve to approximate obliquity, which closely resembles a modulated sine curve with dominant period 41 kyr. Using a simplified insolation curve can elucidate certain mechanisms like frequency locking, which would still be valid but not at clear with more complicated forcing.

3.3 Summary

Three insolation parameters obliquity, longitude of perihelion, and eccentricity govern the insolation received at different latitudes on Earth over different time periods of the year. These orbital parameters vary over time with dominant periods of approximately $21, 41, 100$ and 400 thousand years. Insolation variations at Northern latitudes influence the mass balance and stability of ice sheets in the Northern hemisphere, and as such are believed to play a major role for the glacial cycles. For different reasons, researchers often consider scalar insolation time series, even though the insolation field is defined across the globe. These insolation curves can to a good approximation be expressed as sums of obliquity and precession, both modulated sine curves, which conceptually simplifies the interpretation of insolation variations.

Chapter 4

Dynamical systems

A dynamical system is a rule describing how a state x evolves in time t. This broad definition makes it clear that dynamical systems are suitable models for a range of natural phenomena, including the glacial cycles.

The aim of this section is to provide an introduction to dynamical systems theory and provide a basis for later chapters. The first part will be rather informal and focus on concepts, whereas the later part will be somewhat more rigorous in preparation for later chapters.

4.1 Definition and basic concepts

We define a dynamical system as a tuple (f, X, T), where X is the state space containing states $x \in X$ that evolve in time $t \in T$ according to an evolution rule $f : T \times X \to X$ that maps states at time instants to new states. This definition is not very constrained. This is on purpose, since models of glacial cycles come in many different kinds. We begin with smooth ordinary differential equations, a widely studied class of systems, and then introduce more exotic ones.

An Ordinary Differential Equation (ODE) is a dynamical system with continuous space and time. Typically $T = \mathbb{R}$ (the real numbers) and $X = \mathbb{R}^n$, n-dimensional Euclidean space, in which the state only depends on a time derivative

$$\dot{x} = f(t, x), \tag{4.1}$$

where $\dot{x} := \frac{d}{dt}x$ and where the evolution rule $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is a real function. Note that x and f may be vector valued. Typically, f is assumed to be continuous and smooth too some degree such that $f \in C^k(X)$, the space of functions with continuous derivatives up to order k.

Autonomous systems are ODE:s for which the right hand side of (4.1), sometimes called the vector field, does not depend explicitly on time $f(t, x) = f(x)$. The qualitative behaviour of such systems is much better understood compared to non-autonomous systems.

4.2 The fold system

A classic example of an infinitely smooth, autonomous, system is the fold equation

$$\dot{x} = f(x; \lambda) = x^2 - \lambda, \tag{4.2}$$

where the dot denotes derivative with respect to time. In fact $f(x; \lambda)$ is a family of functions parametrised by a parameter λ. Such parametrised families of functions are central for Chapter 8 of this thesis. Let us study the system behaviour for a fixed value $\lambda = 1$. In this

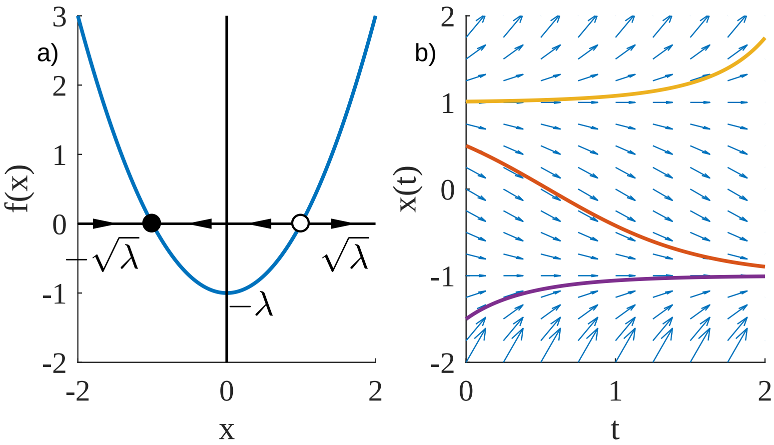


Figure 4.1: Graphical representation of the fold system $\dot{x} = x^2 - \lambda$, for $\lambda = 1$. a) The sign of the vector field $f(x)$ dictates on which segments the flow flows to the right or left on segments of the real line. The filled and open dots are stable and unstable equilibria respectively. b) Three solutions for different initial conditions. The blue arrows show the direction of the vector field in (x, t) space.

case, we can solve the *initial value problem*, which is (4.2) combined with an initial condition (t_0, x_0), analytically for $x(t)$. Separating variables, we have that

$$\begin{aligned}\frac{dx}{dt} &= x^2 - \lambda \\ \int_{x_0}^{x} dx \frac{1}{x^2-\lambda} &= \int_{t_0}^{t} dt, \\ \left| \frac{(x-\sqrt{\lambda})(x_0+\sqrt{\lambda})}{(x+\sqrt{\lambda})(x_0-\sqrt{\lambda})} \right| &= 2\sqrt{\lambda}(t-t_0) \\ x(t) &= \sqrt{\lambda}\frac{x_0\left(1+e^{2\sqrt{\lambda}(t-t_0)}\right)+\sqrt{\lambda}\left(1-e^{2\sqrt{\lambda}(t-t_0)}\right)}{x_0\left(1-e^{2\sqrt{\lambda}(t-t_0)}\right)+\sqrt{\lambda}\left(1+e^{2\sqrt{\lambda}(t-t_0)}\right)},\end{aligned} \tag{4.3}$$

unless $x = \pm\lambda$, in which case $x(t) = x_0$, and unless the denominator is zero, in which the solution blows up to $\pm\infty$.

We can understand the solution behaviour by plotting the graph of solutions over time, as in Figure 4.1. The direction and magnitude of the vector field is represented by blue arrows. We see that all three solution trajectories follow the vector field. Figure 4.1 indicates that for initial values $x_0 < \sqrt{\lambda}$, trajectories approach $x = -\sqrt{\lambda}$, whereas for $x_0 > \sqrt{\lambda}$ solutions diverge upwards. At $x = \pm\sqrt{\lambda}$ the change in x vanishes.

We can confirm these visual suspicions by inspecting the differential equation (4.2). Firstly, at points $x_0 = \pm\sqrt{\lambda}$ the solution does not change in time since $\dot{x} = 0$. These points are called equilibria, or fixed points. The stability of an equilibrium can typically be found from the eigenvalues of the linearisation of the vector field near the equilibrium x^*

$$f(x) \approx f(x^*) + \frac{\partial f}{\partial x}(x^*)(x - x^*) + \text{h.o.t}, \tag{4.4}$$

where h.o.t. stands for higher order terms. If the Jacobian (first derivative in the one-dimensional case) $\frac{\partial f}{\partial x}$ only has eigenvalues λ_i with strictly negative real part $\Re(\lambda_i) < 0$, $\forall i$,

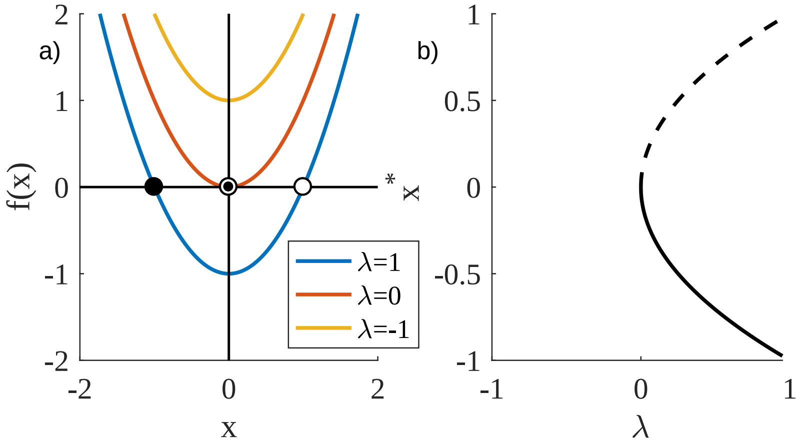


Figure 4.2: Saddle-node (fold) bifurcation in the fold system $\dot{x} = x^2 - \lambda$. a) As λ changes, the number of equilibria (solutions to $f(x) = 0$) changes. The solid and hollow circles are stable and unstable equilibria, and the half-filled circle is a saddle-node with a zero eigenvalue. b) Bifurcation diagram, showing the equilibria $x^* = \sqrt{\lambda}$ versus the bifurcation parameter λ. Solid and dashed line indicate stable and unstable equilibria respectively.

then the fixed point is stable, such that nearby points are attracted to that point. If at least one real part is positive on the other hand, $\Re(\lambda_i) > 0,\ \forall i$, then the point is unstable. For (4.2), we have that

$$\left.\frac{\partial f(x)}{\partial x}\right|_{x=\pm\sqrt{\lambda}} = \pm 2\sqrt{\lambda}, \tag{4.5}$$

confirming that $x^* = \sqrt{\lambda}$ is a stable equilibrium and $x^* = \sqrt{\lambda}$ is an unstable equilibrium. The result that equates the stability of an equilibrium with that of its linearisation is called the Hartman-Grobman theorem [Hartman, 1960, Grobman, 1959], and only holds as long the equilibrium is hyperbolic, meaning that none of the eigenvalues have zero real part. If $\Re(\lambda_i) = 0$ for some i, then the equilibrium is called non-hyperbolic and the stability cannot be inferred from the linearisation [Strogatz, 1994].

4.3 Bifurcation: Saddle-node bifurcation

Loosely speaking, a bifurcation occurs when the qualitative behaviour of a system changes as one (or more) parameters are varied gradually. We explore this concept by returning to the the fold system (4.2).

If we consider (4.2) for different values of λ between 1 and -1, we note that the number of equilibrium points change. For $\lambda > 0$ there are two equilibria $x^* = \pm\sqrt{\lambda}$ solving $x^2 - \lambda = 0$, for $\lambda = 0$ there is a single double root $x^* = 0$, and for $\lambda < 0$ there are no equilibria. This change can be visualised in Figure 4.2 a). As λ decreases, the quadratic function moves up the vertical axis such that the equilibria move closer to each other until they meet at $x = 0$ and eventually disappear. Since the number of solutions change, we say that the system undergoes bifurcation.

The parameter λ is called the bifurcation parameter. A typical way of visualising bifurcation is to plot the equilibria on the vertical axis and the bifurcation parameter on the horizontal axis as in Figure 4.2. In Figure 4.2 the line of equilibria "folds" over itself, which gives the bifurcation its name. The bifurcation is also called saddle-node bifurcation. This is because in higher dimensions the unstable equilibrium generally has some negative eigenvalues, making it a so-called saddle point, whereas the stable equilibrium is called a node. At a saddle-node a saddle and a node merge.

4.4 Numerical solutions

We were able to solve the initial value problem (4.2) analytically, but for general systems we do not have that luxury. We then have to resort to numerical solutions for estimates to the trajectories. For the remainder of this thesis, we will use the Euler forward numerical scheme in most of the cases. For numerical solutions, time must be discretised such that the i:th time point is given by $t_i = t_{i-1} + \Delta_i$, where Δ_i is the i:th time increment. We use constant time step, such that $\Delta_i = \Delta$ for all i. The Euler forward scheme estimates the future value of the solution trajectory as

$$x_{t_{i+1}} = x_{t_i} + \Delta f(t_i, x_{t_i}).$$

Although simple, this method has many drawbacks; errors grow fast relative to other methods and it is ill suited for problems that are difficult to integrate, such as stiff (a kind of difficult-to-integrate) problems. However, the method works well when accuracy or stability is not an issue, and it is convenient to use.

For difficult-to-integrate problems, we use integrators built into the software MATLAB, version 2016b. These have error control and adaptive step-size, and some are suited for stiff problems. The van der Pol system in the following section is one such system.

4.5 A fast-slow system in two dimensions

The van der Pol oscillator is a mathematical model of a type of electronic circuit, proposed and popularised in the 1920'ies [van der Pol, 1926]. The model has subsequently been applied to other fields, such as neuroscience and the theory of glacial cycles (e.g. [Crucifix, 2013, de Saedeleer et al., 2013, Ditlevsen and Ashwin, 2018]). The model

$$\ddot{x} - \frac{1}{\sqrt{\epsilon}}(1 - 3x^2)\dot{x} - x = 0, \tag{4.6}$$

for $0 < \epsilon \ll 1$, where $\ll$ means much smaller than, is the archetype of a relaxation oscillator. The reason why it is called a relaxation oscillator is that the state x undergoes periodic oscillations, consisting of slowly changing “relaxing” epochs combined with fast jumps, as can be seen by integrating the following transformed version of (4.6). By the Liénard transformation

$$y = x^3 - x + \sqrt{\epsilon}\dot{x},$$

combined with time rescaling $\sqrt{\epsilon}t \to t$, Equation (4.6) can be written as a system of equations

$$\begin{aligned} \epsilon\dot{x} &= -x^3 + x - y := g(x, y) \\ \dot{y} &= x := h(x, y), \end{aligned} \tag{4.7}$$

where we have named the vector fields g and h respectively. The solutions $x(t)$ to (4.7) for $\epsilon = 0.01$ can be plotted in time, as for the one-dimensional fold system (Figure 4.3 b,c)). We see that the *fast* variable x remains almost constant over time intervals but sometimes jump nearly instantaneously, whereas the *slow* variable y progresses slowly. The fast variable is named so because for small ϵ the time derivative $\dot{x}$ is very large compared to $\dot{y}$, whenever the right hand side is non-zero. The first line of (4.7) suggests that if y changes much slower than x, then $\dot{x}$ will quickly adjust x so that the right hand side is nearly zero. Because of this separation into a fast and slow variables, the system is called a *fast-slow* system.

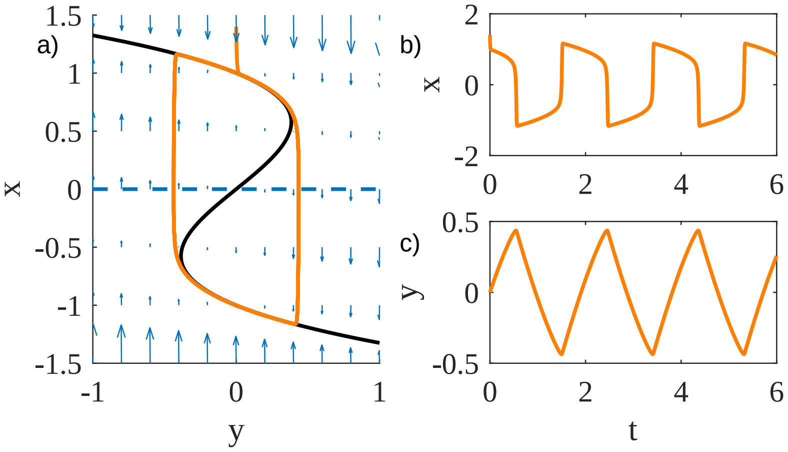


Figure 4.3: The van der Pol oscillator (4.7) with fast variable x and slow variable y. a) Phase portrait. The black curve is the fast nullcline (the critical set) to which trajectories attract. The blue dashed line is the slow nullcline; above it solutions move right and below it they move left. The orange curve is an example solution and the arrows show the direction of the vector field. b,c) Time series of variables x and y for the solution in a).

In the *singular limit* $\epsilon \to 0$ the adjustment of x to the critical set becomes “infinitely” fast almost everywhere, such that x is constrained to the *critical set* defined by $g(x, y) = 0$. The motion of the slow variable constrained to the critical set is given by the *reduced* system

$$\begin{aligned} 0 &= -x^3 + x - y \\ \dot{y} &= x, \end{aligned} \tag{4.8}$$

Plotting the critical set in the *state space* (also called *phase space*, the (x, y) plane) together with a sample trajectory, we see that for $\epsilon = 0.01$ trajectories attract rapidly towards the critical set almost everywhere, although note that near the folds ($x = \pm\sqrt{1/3}$) the attraction is slower (Figure 4.3 a)). We can understand why this is so by considering (4.7) in the *fast time* $\tau = t/\epsilon$. Writing $x' = dx/d\tau$, and setting $\epsilon = 0$, we get the *layer* equations

$$\begin{aligned} x' &= -x^3 + x - y \\ y' &= 0, \end{aligned} \tag{4.9}$$

which say that in the fast time, the slow variable is viewed as constant. The fast variable obeys a differential equation, whose equilibria lie on the critical set. As usual, we can infer the stability of these equilibria through stability analysis. Differentiating the right hand side we get

$$\frac{\partial g}{\partial x} = -3x^2 + 1,$$

which indicates that for $x = \pm\sqrt{1/3}$ any equilibria are nonhyperbolic and the stability thus is undecided from linear stability. Solving $g(x, y) = 0$ for y shows that there indeed are two such equilibria $(\pm 1/\sqrt{3}, \pm 2/\sqrt{3})$ at the folds in Figure 4.3. The derivative also confirms Figure 4.3, which indicates that equilibria are hyperbolically attracting for $|x| > 1\sqrt{3}$, and hyperbolically repelling for $-1/\sqrt{3} < x < 1/\sqrt{3}$. The similarity of the folds in Figure 4.3 and the fold system in Figure 4.2 is not coincidental. Heuristically, we suspect that because the equations are smooth, the local behaviour around a point (x_0, y_0) should be well captured by low order terms in the Taylor expansion

$$\begin{aligned} g_{loc}(x,y) \approx g(x_0,y_0) + g_x(x_0,y_0)(x-x_0) + g_y(x_0,y_0)(y-y_0) + \\ \frac{1}{2}g_{xx}(x_0,y_0)(x-x_0)^2 + \frac{1}{2}g_{yy}(x_0,y_0)(y-y_0)^2 + \\ \frac{1}{2}g_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \text{h.o.t.}, \end{aligned}$$

where h.o.t. again means higher order terms, and derivatives are written as subscripts, e.g. $\partial g/\partial x = g_x$. Evaluating near the positive fold point $(x_0, y_0) = (1/\sqrt{3}, 2/\sqrt{3})$ and shifting the origin $(\tilde{x}, \tilde{y}) = (x - x_0, y - y_0)$ we find that

$$g_{loc}(\tilde{x}, \tilde{y}) \approx -\tilde{y} - \sqrt{3}\tilde{x}^2, \tag{4.10}$$

which is identical to (4.2) save for a change of signs, reflecting the fold in the x-axis and reversing the stability of the branches. Importantly, the "constant" variable y takes on the role of λ, the bifurcation parameter. This viewpoint is central to our proposed classification of bifurcations in fast-slow systems in Chapter 8.

Not only (4.9) can be transformed to an equation on the form (4.2). Any smooth parametrised single-variable function $g(x; \lambda)$ for which $g(x_0; \lambda_0) = g_x(x_0; \lambda_0)$ at a point $(x_0; \lambda_0)$, which satisfies the *nondegeneracy conditions*

$$\begin{cases} \text{C.1:} \quad g_{xx}(x_0, \lambda_0) \neq 0 \\ \text{C.2:} \quad g_{\lambda}(x_0, \lambda_0) \neq 0, \end{cases} \tag{4.11}$$

can be transformed to (4.2) by a smooth change of coordinates in an open neighbourhood around $(x_0; \lambda_0)$ (see e.g. [Kuznetsov, 2004, Kuehn, 2015]). For this reason (4.2) is called a *normal form* for the fold, or saddle-node, bifurcation. There exist similar normal forms for other bifurcation problems.

The rate of convergence to the equilibrium at the fold is found by solving $\tilde{x}' = g_{loc}$ from (4.10) at bifurcation $\tilde{y} = 0$. We find that

$$\tilde{x}(\tau) = \frac{\tilde{x}_0}{1 + \sqrt{3}\tilde{x}_0(\tau - \tau_0)},$$

for initial fast time τ_0, indicating that convergence to the equilibrium is only algebraic, as opposed to exponential (4.3) away from the fold point. Moreover, (4.3) shows that the convergence rate decreases for decreasing λ. This explains why the convergence is weak near the fold point.

We end this section by introducing a few new terms, and by commenting on the behaviour of the van der Pol oscillator in the singular limit.

Firstly, the set $g(x, y) = 0$ is called the *x-nullcline*, in addition to the critical set. For a general differential equation with variables x_i, the set for which $\dot{x_i} = 0$ is called the *x_i-nullcline.* Hence $h(x, y) = x = 0$ defines the y-nullcline of the van der Pol system (4.7) (see Figure 4.3). Equilibria occur at the intersections of all nullclines. Hence, for the system (4.7) the only equilibrium is at $(0, 0)$, and it is unstable for every $\epsilon > 0$. Nullclines provide additional information; for smooth systems, $\dot{x_i}$ can only change sign at a nullcline. For example, the system (4.7) the slow variable moves rightward for $x > 0$ and leftward for $x < 0$. Note, however that the sign does not change at the nullcline if there is a zero of odd order on the nullcline (e.g. $0 = y - x^3$).

Sometimes, the set $g(x, y) = 0$ is called the *critical manifold*, which carries the assumption that the set is locally homeomorphic to a real hyperplane $\mathbb{R}^n$, for some $n \in \mathbb{N}$. However, we shall encounter situations when the critical set intersects itself, such that it is not a manifold. Therefore, we stick with the term critical set. Recall that a homeomorphism is a continuous, invertible mapping between spaces, which preserves topological properties.

The layer equation (4.9) and the reduced equation (4.8) suggest that whenever the singular system is not on the critical set, the layer equations bring trajectories close to an attracting part of the critical set, after which the slow dynamics get the time to evolve along the critical set. But what happens at singular points (where $g_x = 0$), such as folds? In general, this a complicated question with no straightforward answer. However, if the slow flow crosses a quadratic fold point satisfying the nondegeneracy conditions (4.11), the answer is clear: the fast subsystem kicks in again, producing a rapid vertical jump off the fold point as for the van der Pol system. If the destination point is a stable, regular part of the critical set (that is, a point where there is no fold) and the slow flow is nonzero at that point, then the slow subsystem kicks in again until, if ever, the fast system is activated upon reaching another fold. Trajectories satisfying these requirements (and some more) are in a sense robust to perturbation, and play a special role in our study of bifurcations of singular relaxation oscillations in Chapter 8.

Relaxation oscillations are special cases of a more general object called *limit cycles.* A limit cycle is an invariant, closed orbit with finite period, which is the limit of at least one trajectory not in the orbit. That limit cycles are invariant means that if you start on the limit cycle, you remain there for all forward and backward time. That the limit cycle is a closed orbit with finite period means that for any initial $x(t_0)$ on the limit cycle, there exists some finite T such that $x(t_0 + T) = x(t_0)$. That the cycle is the limit of a trajectory means that there exists some point, not on the limit cycle, such that the limit of the trajectory in forward or backward time equals the limit cycle.

Another question is what trajectories of the singular system $\epsilon = 0$ tells us about the behaviour of the nonsingular system for which $\epsilon > 0$. Although we do not give that concern much attention in this thesis, we provice a brief description of this topic.

4.5.1 Singular to non-singular

What does the singular system tell us about the non-singular system? The field of Geometric Singular Perturbation Theory is dedicated to this question. For an introduction, see e.g. [Jones, 1995, Kuehn, 2015]. The cornerstone of modern developments is a set of theorems due to Fenichél [Fenichel, 1979, Kuehn, 2015], which roughly states that under certain conditions there exists a slow manifold ϵ-nearby the critical manifold which shares properties such as smoothness with the critical manifold, thus corroborating the intuition that trajectories remain close to the critical manifold when the scale separation parameter ϵ is small enough, at least away from folds.

The most important necessary condition for Fenichél's theorem to hold is that the critical set is locally a *normally hyperbolic manifold*, meaning that the contraction normal to the manifold dominates the contraction tangential to it. In practice this is guaranteed as long as the Jacobian of the fast subsystem has no eigenvalues with zero real part, because ϵ can make g_x/ϵ arbitrarily large in magnitude as long as $g_x \neq 0$, dominating the expansion or contraction of the slow subsystem [Kuehn, 2015].

This guarantees that there is a *slow manifold* near the critical manifold on any regular part of the manifold, that is, away from folds. Near folds, interesting *canard* solutions exist, however. These are trajectories that stay ϵ-close to the critical manifold for a time independent of ϵ. The classic example due to Benoît et al. [1981] is the biassed van der Pol oscillator

$$\begin{aligned} \epsilon\dot{x} &= -x^3 + x - y \\ \dot{y} &= x - a, \end{aligned} \tag{4.12}$$

where a is a constant bias parameter. The system has one equilibrium at $p = (a, -a^3 + a)$. We perform linear stability analysis by evaluating the Jacobian at this point

$$J_p = \begin{pmatrix} \frac{\partial g}{\partial x}(p)/\epsilon & \frac{\partial g}{\partial y}(p)/\epsilon \\ \frac{\partial h}{\partial x}(p) & \frac{\partial h}{\partial y}(p) \end{pmatrix} = \begin{pmatrix} (-3a^2+1)/\epsilon & -1/\epsilon \\ 1 & 0, \end{pmatrix}$$

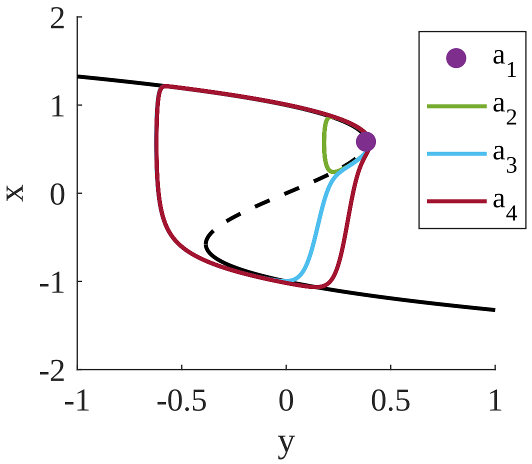


Figure 4.4: Canard explosion in the biassed van der Pol equation (4.12). For $a_1 = 1/\sqrt{3}$ there is a non-hyperbolic equilibrium on the fold (an Andronov-Hopf bifurcation). At $a_2 = a_1 - 0.003755$, a small amplitude limit cycle is born which follows the unstable part of the critical manifold (dashed black curve) and then returns to the upper stable branch (solid black curve). For $a_3 = a_1 - 0.0037652$, the trajectory follows the unstable branch for some time until it jumps to the lower stable branch, completing a much bigger limit cycle (hidden in part behind the deep red curve). For $a_4 = a_1 - 0.01$ the limit cycle closely resembles the large amplitude relaxation oscillation of the unbiassed van der Pol system. The system was integrated for 100 time units with $\epsilon = 0.05$ and using Matlab's stiff solver `ode23`, with absolute and relative error tolerances of 10^{-9} and 10^{-12} respectively. To remove transitents, only the solutions for the last 40 time units are plotted

and compute its eigenvalues to be $\lambda_{1,2} = (-3a^2 + 1)/\epsilon \pm \sqrt{(3a^2 + 1)^2/\epsilon^2 - 4}$, which has zero real part for $a = \pm 1/\sqrt{3}$ and $\epsilon \neq 0$. Hence, the equilibria are not hyperbolic at these points, and bifurcation occurs. In fact, since the eigenvalues at bifurcation are an imaginary pair (and some other non-degeneracy conditions hold), the bifurcation is an *Andronov-Hopf bifurcation*, at which a limit cycle is born from an equilibrium [Kuznetsov, 2004]. This can be appreciated from Figure 4.4, in which there is a stable equilibrium for $a = 1/\sqrt{3}$ (the fast nullcline and the slow nullcline $x - a = 0$ intersect). For $a = 1/\sqrt{3} - 0.003755$ a small limit cycle is born, which follows the unstable branch of the critical set for some time before jumping to the top stable branch. For just a further slightly decreased a, the solution instead jumps to the lower stable branch, causing the size of the limit cycle to increase explosively. This phenomenon is called *canard explosion*. *Canard* because the blue curve in Figure 4.4 with some imagination can be thought to resemble a duck (canard is duck in French), and *explosion* because the limit cycles rapidly grows in size for a small change in a. The range of a for which canards exist is exponentially small in ϵ [Kuehn, 2015]. An alternative explanation for the name "canard" is that the solutions following the unstable manifold are not expected exist, since solutions should be repelled. Hence the solutions are "fake news", which is called "canards" in French.

We sum up with two points, 1. that the behaviour of solutions to the nonsingular system near a singularity, such as a fold, can be very different from solutions to to the singular system, but also 2. that the set of parameters for which the systems behave very differently is exponentially small in ϵ. Thus, small enough ϵ, solutions to most nonsingular systems should behave similarly to solutions of the corresponding singular system.

4.6 Forced relaxation oscillations and synchronisation

What happens if a relaxation oscillator is influence by periodic stimulus, called forcing? Interestingly, this is one of the first questions that [van der Pol and van der Mark, 1927] investigated with the van der Pol oscillator. They found that the period of the relaxation oscillation was a fraction of the forcing period for a range of parameters, a phenomenon they termed "frequency demultiplication". Nowadays the phenomenon is called frequency locking, phase locking, entrainment or mode locking, depending on the context and precise meaning. To this day, there has been ample research both on the forced van der Pol oscillator (see e.g. [Cartwright and Littlewood, 1945, Mettin et al., 1993, Guckenheimer et al., 2003]), and on frequency locking as a particular case of synchronisation between oscillators (see [Pikovsky et al., 2001] for an introduction to the topic).

We shall start from the forced van der Pol oscillator and successively simplify it to an integrate-and-fire model that contains the essentials for explaining driven synchronisation, the phenomenon when one oscillator influences the period of another, but not the other way around.

4.6.1 From van der Pol to integrate-and-fire models

The forced van der Pol equation can be written

$$\begin{aligned} \epsilon\dot{x} &= -x^3 + x - y \\ \dot{y} &= x - A\sin(\omega_f t + \phi), \end{aligned} \tag{4.13}$$

where A is a forcing amplitude, $\omega_f = 2\pi/T_f$ is the angular frequency of the forcing, where T_f is the forcing period, and ϕ is the phase of the forcing. To get an appreciation for the system behaviour, note how the solution for $A = 3$ and $\omega_f = 2\pi/0.4$ loops around the fold in Figure 4.5 a). The forcing makes the solution move back and forth along the critical set; this occurs not only near the fold, but also in the interior of the stable sheet, as can be inferred from Figure 4.5 c). Note also, that the solution is periodic, with a period that is five times the period of the forcing (Figure 4.5 b)). This type of frequency locking we denote $1:5$ locking, or more generally $M:N$ locking in case the forcing completes N revolutions per M revolutions of the forced oscillator.

In Figure 4.5 c) we identify a sawtooth wave pattern with ripples from the forcing superimposed. Is it possible to capture this behaviour in an even simpler model that is easier to analyse? This was done first by Levinson [1949] and later by Storti and Rand [1988], who approximate the critical set as a piecewise linear function (Figure 4.6 a)).

We can simplify the system even more by letting the stable critical set be a pair of straight lines, as in Figure 4.6 b). Assuming furthermore that we are in the singular limit ϵ, we get a non-autonomous *hybrid* dynamical system of threshold type with two states G and D

$$\begin{aligned} &\text{In G: } \dot{x} = \mu - A\sin(\omega_f t + \phi) \text{ until } x(t) = x_{c,G}, \text{ then } G \to D \\ &\text{In D: } \dot{x} = -\lambda - A\sin(\omega_f t + \phi) \text{ until } x(t) = x_{c,D}, \text{ then } D \to G. \end{aligned} \tag{4.14}$$

Here, the states G (glaciating) and D (deglaciating) encode which stable branch the oscillator is on, $x(t)$ is the single (slow) variable, $\mu < 0$ is a constant growth rate, $\lambda > 0$ is a constant decay rate, $x_{c,G} > x_{c,D}$ are upper and lower thresholds respectively, and A, ω_f and ϕ are defined as before. This kind of model is called an *integrate-and-fire* model in the neuroscience community, at least under the typical assumption $\lambda \to \infty$. It is called so because some "potential" (here x) is accumulated/integrated slowly and then *fires* when the built up potential reaches a threshold. See e.g. [Knight, 1972, Glass and Mackey, 1979, Keener et al., 1981, Alstrøm et al., 1990] for examples and analyses of such models.

Note that although some model dynamics are lost in each simplification step from the van der Pol oscillator, the model (4.14) represents well the modelling objective "grow slowly until threshold is reached, then decay slowly until lower threshold is reached, repeat".

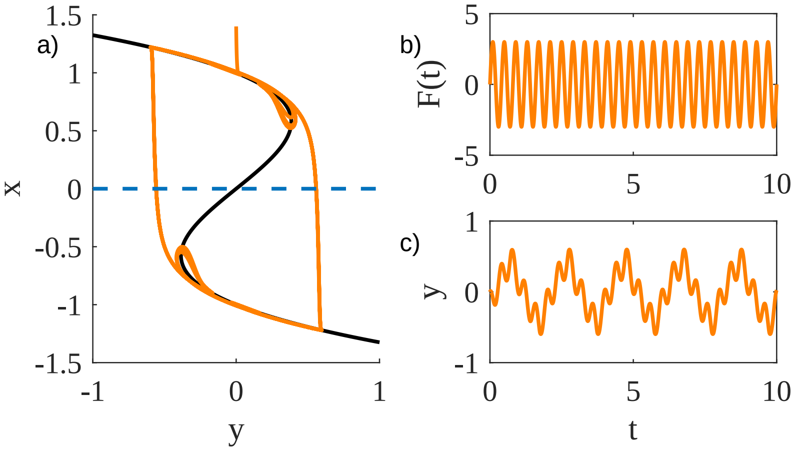


Figure 4.5: The periodically forced van der Pol oscillator (4.13) with fast variable x and slow variable y. a) Phase portrait. The black curve is the fast nullcline (the critical set) to which trajectories attract. The blue dashed line is the slow nullcline; above it solutions move right and below it they move left. The orange curve is an example solution and the arrows show the direction of the vector field. b) The forcing $F(t) = 3\sin(2\pi t/0.4)$, c) Time series of y

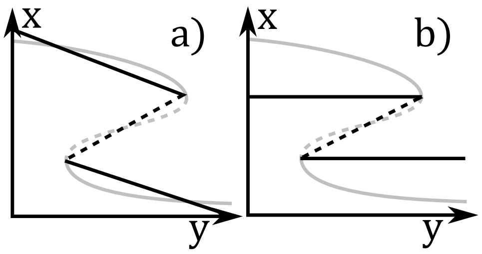


Figure 4.6: Simplifications of the critical set of the van der Pol oscillator. a) Standard s-shaped critical set. a) Piecewise linear set and grey standard s-shaped critical set. b) Piecewise constant set.

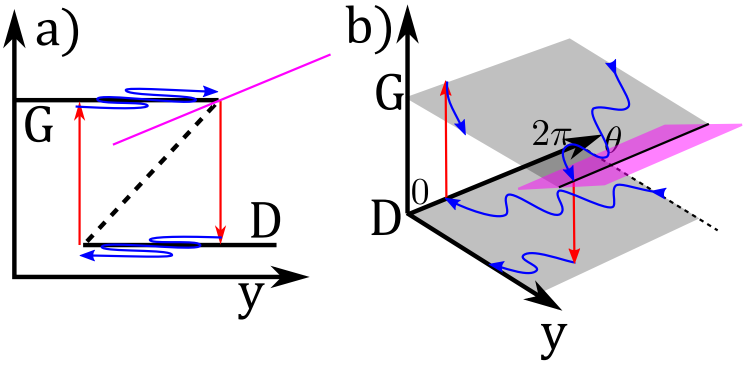


Figure 4.7: Poincaré section (magenta line segment/planar section) of the piecewise constant threshold switching system (4.14). a) Cross section in the space of states G/D and x. Solutions exist on the black solid branches, blue lines denote solutions on these branches and red arrows denote jumps between the branches, triggered when the solution crosses the Poincaré section at the end of a branch transversally. b) Poincaré section in the full G/D, x, θ space, where $\theta(t) = \omega_f t + \phi$ is the argument (phase) of the forcing sinusoid. The solution branches are transparent gray. The phase θ is defined on the circle $S = \mathbb{R}/2\pi\mathbb{Z}$, such that $\theta = 2\pi$ is identified with $\theta = 0$ (the phase space "wraps around itself")

The model (4.14) comes in a few variations: $\mu = \lambda$ gives equal growth and decay rates, corresponding to the unforced van der Pol oscillator, and $\lambda \to \infty$ corresponds to instantaneous resetting that is typical for neuronal modelling. Choosing λ such that the decay occurs over a constant amount of time, regardless of the threshold $x_{c,G}$ and forcing, was used to model glacial cycles [Huybers, 2007]. Moreover, instead of applying the forcing on the slow variable growth, one can instead apply it on the upper threshold [Glass and Mackey, 1979, Huybers, 2007, Alstrøm et al., 1990], the lower threshold or both thresholds [Pikovsky et al., 2001].

In the following we illustrate some important ideas and concepts of frequency locking, using the model (4.14).

4.7 Frequency locking and Arnold tongues

We say that a forced oscillator is $M : N$ frequency locked, if the oscillator completes M revolutions when the forcing completes N revolutions. This generically occurs on open sets of parameters, which implies that it is persistent to small perturbations. These open sets are called Arnold tongues, a name derives from the appearance of the sets in two-dimensional parameter space. We illustrate driver frequency locking and Arnold tongues with the integrate-and-fire *branch switching* oscillator model 4.14. A similar exposition can be found in [Glass and Mackey, 1979, Storti and Rand, 1988, Keener et al., 1981]. We are not aware of any treatment of is particular system, however, which offers some interesting insights.

Our goal is to find parameters which enable periodic solutions of (4.14). To do so we construct a *Poincaré map*, a section in state space that solutions cross transversally (Figure 4.7 a)). We choose a section which is crossed every time the solution switches from the G to the D state (think deglaciation for glacial cycles).

We can define a function f mapping the initial phase θ_n (starting right after crossing the Poincaré section) to the phase at the next return to the Poincaré section

$$\theta_{n+1} = f(\theta_n), \tag{4.15}$$

where n indexes the number of iterations and θ_n is understood to be defined modulo 2π (θ_n

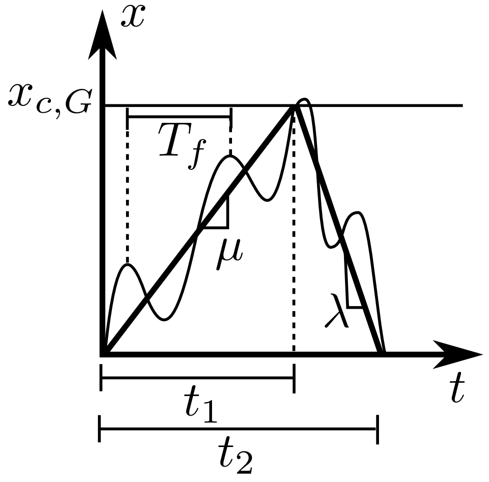


Figure 4.8: Derivation of the return map for the branch switching oscillator (4.14). The parameter $x_{c,G}$ is the upper threshold which triggers the switch to the decay state. The parameters μ and λ are the growth and decay rates, T_f is the forcing period, and t_1 and t_2 are the smallest times until reaching the upper and lower thresholds respectively.

is defined on the circle $S = \mathbb{R}/2\pi\mathbb{Z})$. This is the Poincaré map, an example of a *discrete time* dynamical system. Solutions to (4.14) are periodic if and only if the map (4.15) maps an initial phase θ_0 to itself after a finite number of iterations

$$\theta_0 = f \circ f \circ \cdots \circ f(\theta_0) = f^n(\theta_0), \tag{4.16}$$

where $\circ$ denotes function composition. If (4.16) holds for $n = 1$, the map has a fixed point which means that the oscillator completes one period per return to the Poincaré section. If (4.16) holds for $n > 1$ then the map has an n-cycle. Equivalently, the map $f_n(\theta) = f^n(\theta)$ has a fixed point.

Next, we find an expression for the Poincarè map of (4.14), and determine for which parameters periodic solutions exist. If a periodic solution persists on an open set of parameters, then the solution is frequency locked, and the parameter sets are called Arnold tongues.

Denote by t_i the i:th time the solution reaches a threshold ($x_{c,G}$ or $x_{c,D}$), starting from $t = t_0$ in the glaciating state. Thus, t_i for even i corresponds to decay state terminations. We naturally define the phase of the forcing at the start of a growth state G to be the argument of the cosine at such a time:

$$\theta_i = \omega_f t_{2i} + \phi. \tag{4.17}$$

Next, we seek an expression for the unknown times t_i. We assume without loss of generality that $x_{c,D} = 0$ and that the oscillator starts at $x(t_0) = 0, t_0 = 0, \theta_0 = \phi$ and in the growth state G. Then the solution is given by

$$x(t) = \mu t + \frac{A}{\omega_f}\left(\cos\left(\omega_f t + \phi\right) - \cos\phi\right),$$

until $x(t)$ reaches the upper threshold at some time t_1 such that

$$\begin{aligned} x(t_1) &= x_{c,G} \\ x_{c,G} &= \mu t_1 + \frac{A}{\omega_f}(\cos\left(\omega_f t_1 + \phi\right) - \cos\phi). \end{aligned} \tag{4.18}$$

Thereafter, the solution switches to the D state, and the time evolution is given by

$$x(t) = x_{c,G} - \lambda(t - t_1) + \frac{A}{\omega_f}(\cos(\omega_f t + \phi) - \cos(t_1\omega_f + \phi)),$$

valid until $t = t_2$ when $x(t) = x_{c,D} = 0$ such that

$$0 = x_{c,G} - \lambda(t_2 - t_1) + \frac{A}{\omega_f}(\cos(\omega_f t_2 + \phi) - \cos(t_1\omega_f + \phi)). \tag{4.19}$$

Now, we have reached an implicit expression for t_2, and thus the phase at first return

$$\theta_1 = \omega_f t_2 + \phi. \tag{4.20}$$

In order for $1 : N$ locked periodic solutions to exist, we require that $\theta_1 = \theta_0 + 2\pi N$, such that

$$2\pi N = \omega_f t_2,$$

or

$$t_2 = NT_f,$$

where we define the forcing period $T_f = 2\pi/\omega_f$. We can use this information to find frequency locking regions. We first add and subtract (4.18) and (4.19) seperately to get

$$\begin{aligned} 0 &= \mu t_1 - \lambda(t_2 - t_1) + \frac{A}{\omega_f}(\cos(\omega_f t_2 + \phi) - \cos\phi) \\ 2x_{c,G} &= \mu t_1 + \lambda(t_2 - t_1) + \frac{A}{\omega_f}(2\cos(\omega t_1 + \phi) - \cos(\omega_f t_2 + \phi) - \cos\phi). \end{aligned} \tag{4.21}$$

Now, since $\omega_f t_2 + \phi = N2\pi + \phi$, we get that

$$\begin{aligned} 0 &= \mu t_1 - \lambda(t_2 - t_1) \\ 2x_{c,G} &= \mu t_1 + \lambda(t_2 - t_1) + \frac{A}{\omega_f}(2\cos(\omega_f t_1 + \phi) - 2\cos\phi). \end{aligned}$$

The first line gives that

$$t_1 = \frac{\lambda}{\mu + \lambda} t_2 = \frac{\lambda}{\mu + \lambda} NT_f,$$

which inserted into the second line gives

$$2x_{c,G} = \frac{2\mu\lambda}{\mu + \lambda} NT_f + 2\frac{A}{\omega_f}(\cos(\omega_f t_1 + \phi) - \cos\phi),$$

which in turn after rearrangement and defining the internal (unforced) period to be $T_o = x_{c,G}(\mu + \lambda)/\mu\lambda$, gives that

$$(T_o - NT_f)\frac{2\pi\mu\lambda}{AT_f(\mu + \lambda)} = \left(\cos\left(\frac{N2\pi\lambda}{\mu + \lambda} + \phi\right) - \cos\phi\right).$$

For the equality to hold, the left hand side must not exceed the maximum of the right hand side. This produces bounds on the regions where $1 : N$ periodic solutions can exist, since the difference of cosines is bounded. The right hand side can be rewritten as

$$\begin{aligned} &C(\phi) = 0, && \text{if } \nu \neq 0, \text{ or} \\ &C(\phi) = \sqrt{2(1 - \cos\nu)}\cos\left(\pi + \arctan\left(\frac{\sin\nu}{\cos\nu - 1}\right) + \phi\right) && \text{otherwise,} \end{aligned} \tag{4.22}$$

where arctan is the inverse tangent function with domain $[-\pi/2, \pi/2]$ and

$$\nu = N2\pi\frac{\lambda}{\mu + \lambda}.$$

From (4.22) we learn that

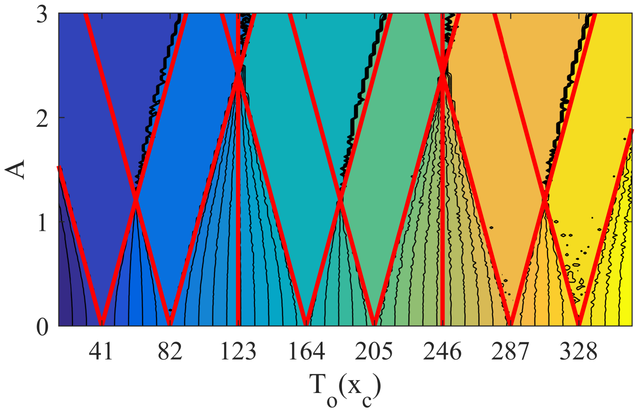


Figure 4.9: Arnold tongues in the $(A, T_o(x_{c,G}))$-plane of amplitude and internal period, as a function of the upper threshold of deglaciation $x_{c,G}$, for the branch switch model (4.14). The red lines enclose cone-shaped regions (Arnold tongues) in which $1 : N$ frequency locking can exist, meaning that the period of the oscillator is N times that of the forcing. Model parameters are $\mu = 1, \lambda = 2, T_f = 41$ and $x_{c,D} = 0$. The thin contour lines separate coloured regions of similar average duration, where brighter colours means longer duration. The average duration was estimated numerically for 2000 time units and the Euler forward scheme with a time step 0.05 time units. The glacial terminations for the first 500 time steps are ignored.

1. If $\nu = 0$, then the left hand side can only be satisfied if $(T_o = NT_f)$,
2. the maximum/minimum C_{max} and C_{min} of $C(\phi)$ occurs for $\nu = (2n+1)\pi$ and $\nu = 2n\pi$ respectively, for $n \in \mathbb{N}$,
3. C_{max} and C_{min} have the same absolute value $C(\phi)$, which depends on N, λ and μ,
4. $C(\phi)$ assumes every value in $[-1, 1]$, for every $\nu \neq 0$ and for some $\phi \in [0, 2\pi]$.

We shall examine these observations further, but first we state the boundaries of sets in the $A, T_o(x_c)$-plane, where T_o is a function of x_c, for fixed μ, λ, T_f and N:

$$\begin{cases} (T_o - NT_f)\frac{2\pi\mu\lambda}{T_f(\mu+\lambda)C_{min}} \leq A \leq 0 & \text{if } T_o \leq NT_f \\ (T_o - NT_f)\frac{2\pi\mu\lambda}{T_f(\mu+\lambda)C_{max}} \geq A \geq 0 & \text{if } T_o \geq NT_f, \end{cases} \tag{4.23}$$

for $C_{min}, C_{max} \neq 0$. In Figure 4.9 we show these sets for $N = 1, 2, \ldots, 8, \mu = 1, \lambda = 2, T_f = 41$ and $x_{c,G}$ ranging from 10 to 240. Red lines (the boundaries from (4.23)) enclose cone-shaped sets within which $1 : N$ frequency locked solutions exists. That the sets look like tongues explains the name “Arnold tongues”. Note that the tongues start at $T_o = NT_f$ (multiples of $T_f = 41$) at $A = 0$, and then widen as the forcing strength increases. The $N = 3$ and $N = 6$ tongues are exceptions, however. These tongues are “infinitely thin”, such that the set for which the forcing period T satisfies $T = NT_f$ is not open. The frequency locking vanishes because the argument of the cosine in (4.22), $\nu = 4\pi N/3$, is a multiple of 2π for N a multiple of 3. Similarly, for $\mu = 1, \lambda = 3$ every fourth $1 : N$ Arnold tongue vanishes. In the symmetric case $\mu = \lambda = 1$ all even N tongues vanish, as has been noted frequently for the unbiassed forced van der Pol oscillator [Levi, 1990, Guckenheimer et al., 2003].

The phenomenon of vanishing tongues has practical importance, even though tongues only vanish on parameter sets of Lebesgue measure zero. This is because the widths of tongues change continuously with parameters, such that for e.g. $\mu = 1, \lambda = 2.1$ the $N = 3, 6, \ldots$ tongues are considerably thinner than the other ones. We suspect the tongues to vary continuously also for other continuous model perturbations.

More generally, the widths of individual Arnold tongues can be inferred from (4.22). An interesting case is the limit as $\lambda \to \infty$, in which case the widths of all tongues tend to zero, preventing all frequency locking.

So far, we have neglected a few points. Firstly, there are $M : N$ Arnold tongues for $M > 1$ corresponding to fixed points of higher iterates of the the phase map (4.16). However, these do not share the neat analytical expressions of the $M = 1$ case. Regardless, these *higher order tongues* are generally much thinner than the $M = 1$ tongues, and are therefore arguably of less importance.

Secondly, the condition (4.23) is only a necessary condition for $1 : N$ frequency locking. This is evident from the overlapping tongues in Figure 4.9, since there cannot be simultaneously $1 : N_1$ and $1 : N_2$-frequency locking ($N_1 \neq N_2$) for the same parameters and initial conditions. The non-sufficiency of (4.23) comes from that the thresholds $x_{c,G}$ and $x_{c,D}$ can be reached for shorter times t_1, t_2 than required for frequency locking to occur.

Thirdly, we dodged the question of if there exists an initial phase ϕ inside each frequency locking region which is a fixed point of the return map. However, equation (4.22) guarantees that this is the case, as long as there exist t_1 and t_2 assumed by the necessary conditions.

Fourthly, we have neglected the behaviour of the first return map (the Poincaré map) and ways to diagnosticise whether there is frequency locking in the system. This can be done numerically, as we shall see.

4.7.1 The first return map

It is helpful to visualise the first return map to the Poincaré section (4.20) from one phase of the forcing θ_n to the next θ_{n+1}. Since there is no analytic expression for the first return time t_2, the map has to be evaluated numerically, either by a root finding algorithm or direct numerical integration of the full system (4.14). We choose the former option, increasing time by increments of 0.05 time units and calculating $x(t)$ until a threshold is reached.

The return map for weak forcing $A = 0.3$ and three different values of $x_{c,G} = 28, 30.724, 33$ can be seen in Figure 4.10a), where $\mu = 1, \lambda = 2$ and $T_f = 41$. We see that for each $x_{c,G}$ the graph of the map forms a continuous curve that wraps around the edges of the domain. For maps, fixed points occur at intersection with the identity line $\theta_{n+1} = \theta_n$. Two examples can be seen in Figure 4.10a) for $x_{c,G} = x$.

There is a bifurcation and stability theory for discrete time systems, analogous to, but distinct from the bifurcation theory for continuous time systems. For example, the continuous fold bifurcation (encountered in Section 4.2) has a discrete time analogue. In the discrete case, the defining condition is that $f(\theta_n) = \theta_n, \partial f / \partial \phi = 1$, such that the graph is tangent to the identity line at bifurcation. Figure 4.10 a) shows an example of cobwebbing, which visualises how the phase evolves in time, in this case towards a fixed point.

Figure 4.10a) shows how one stable and one unstable fixed point first merge at a saddle-node bifurcation, and then are annihilated as $x_{c,G}$ is increased from 28 to 33. The saddle-node bifurcation a bifurcation occurring on the boundary of Arnold tongues in a wide range of systems for relatively weak forcing. We can verify analytically that the bifurcation in Figure 4.10a) is indeed a saddle-node bifurcation, and a generic one on top of that. First, it can be shown (in Section 4.7.8) that the slope of the return map on the boundary of an Arnold tongue

$$\frac{\partial \theta_1}{\partial \phi} = \frac{(\lambda + A \sin(\nu + \phi))(\mu - A \sin \phi)}{(\lambda + A \sin \phi)(\mu - A \sin(\nu + \phi))},$$

indeed equals 1, where $\nu = 2\pi N \lambda / (\lambda + \mu)$. Note that the return map is singular if any of the factors in the denominator are zero, but that it does not occur as long as $A \leq \mu$ and $A \leq \lambda$, which is the case if $A = 0.3, \mu = 1, \lambda = 2$ as in Figure 4.10 a).

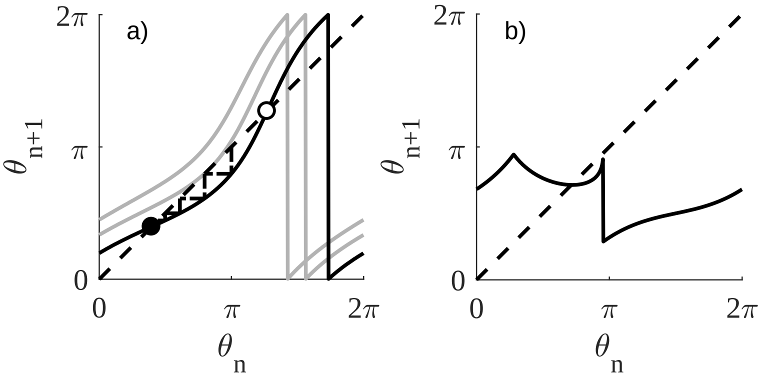


Figure 4.10: First return phase maps (4.20) for the branch switching oscillator (4.14). a) The solid black curve is the graph of the phase return map for $x_{c,G} = 28$. The filled/hollow dots are stable/unstable fixed points. A cobweb illustrating the evolution of the phase, starting at $\theta_0 = \pi$ is shown with dash-dotted line segments. The grey lines show the graph of the map for $x_{c,G} = 30.7240$ and $x_{c,G} = 33$. When the graph becomes tangent to the identity line (dashed diagonal) for $x_{c,G} = 30.7240$ a saddle-node bifurcation of maps occurs. Other parameters are $A = 0.3$, $\mu = 1$, $\lambda = 2$ and $T_f = 41$ b) The phase map contains discontinuities for $A = 1.3$ and $x_{c,G} = 41$

The genericity conditions for the saddle node $\partial^2 f/\partial\phi^2 \neq 0$ and $\partial f/\partial\alpha$, where α is a bifurcation parameter, can be shown to be satisfied for f for most parameters and choice of bifurcation parameter [Kuznetsov, 2004]. The last condition means that the parameter curve is non-tangent to the bifurcation surface in the parameter space.

Before moving on, we note that for small $A < \mu$ the map is invertible and well behaved, and is homeomorphic (can be continuously and invertibly deformed) to the famous and well studied standard circle map

$$\theta_{n+1} = \theta_n + \Omega + A\sin(\theta_n),$$

where Ω is a natural frequency and $A < 1$ an amplitude [Arnold, 1961, Pikovsky et al., 2001]. For larger A the boundaries of frequency locking regions may be different. For instance, there can be an abrupt shift from one frequency locking region to another as observed in Figure 4.9. The phase return map for $A = 1.3$, $x_c = 41$ (Figure 4.10 a)) has a discontinuity, where the slope is unbounded. This can lead to abrupt loss of frequency locking. Guckenheimer et al. [2003] studied several such bifurcations for the singular van der Pol oscillator.

4.7.2 Rotation number and average duration

A common way to quantify frequency locking is by the *rotation number*, sometimes called *winding number*,

$$\rho(\phi_0) = \lim_{n\to\infty} \frac{\phi_n - \phi_0}{2\pi n},$$

where now $\phi_n \in \mathbb{R}$ is defined on the entire real line through the natural *lift* [Kuehn, 2007]

$$\Phi(\phi + n2\pi) = \phi + n2\pi.$$

The rotation number measures the average number of revolutions of the forcing phase per revolution of the oscillator as time goes to infinity. For homeomorphisms (continuous maps with continuous inverse) on the circle, the rotation number is well defined (meaning it converges), and it is independent of initial conditions. However, for non-invertible maps, such as the branch switching oscillator (4.14) with large forcing, different initial conditions can converge to solutions with different rotation numbers in parameter regions of overlapping Arnold tongues, as suggested by the Arnold tongue diagram in Figure 4.9.

If the rotation number of a periodically forced system is rational

$$\rho(\phi_0) = \frac{M}{N}$$

then the oscillator completes M revolutions in the time that the forcing completes N revolutions. This indicates $M : N$ frequency locking, as long as this relation between periods persist under perturbation.

Sometimes, it is stated that a rotation number M/N implies that the period T of the oscillator is N/M times the forcing period T_f. This is somewhat inaccurate, since T refers to the *average* time it takes for a solution to return to a Poincaré section. But this is not necessarily the period of the oscillator, unless $M = 1$. If $M > 1$, then the period of the oscillator is generally MT. We call the time until return to a Poincaré section a *duration*, and we denote the n:th duration of a solution D_n.

In the context of glacial cycles, we use an alternative measure of frequency locking, *average duration* [Ashkenazy, 2006]

$$\overline{D} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} D_i,$$

which naturally measures the average duration between e.g. glacial terminations. Frequency locking for a periodically forced system implies that $\overline{D}$ is a multiple of the forcing period.

4.7.3 Devil's staircases

We can use the average duration to represent frequency locking as a function of one or more system parameters. Plotting the average duration versus internal period as a function of $x_{c,G}$ for parameters $A = 0.9, \mu = 1, \lambda = 2, T_f = 41$, produces a Devil's staircase, so called because it has an interesting structure (Figure 4.11). The average duration is constant on stairs, intervals of finite width, corresponding to Arnold tongues. Although it is not clear from the figure, the staircase consists of an infinite number of steps, each corresponding to a rational number associated with an Arnold tongue. Therefore the staircase is not continuous, and it is non-constant on a set of measure zero [Pikovsky et al., 2001]. However, as the order (M) of tongues increases, the widths of steps tends to zero [Strogatz, 1994]. This makes the staircase fractal, similar to the famous Cantor set [Strogatz, 1994].

This holds for map homeomorphism, like the weakly forced circle map and the weakly forced branch switching oscillator. For large forcing amplitude, the rotation number may become non-monotonic, ill defined, and the space between steps of the Devil's staircase may have positive length.

4.7.4 Numerical detection of frequency locking

In this thesis, we identify frequency locking by estimating the average duration by solving the system equations numerically and averaging the durations between deglacial transitions. Only the solution after a transient time is used to estimate the average duration. This is in order to give the system time to get near an attractor of the system. Frequency locking regions are taken to exist on regions in parameter space where the average duration remains constant under perturbation.

A drawback of this method is that it is difficult to resolve the precise locations of Arnold tongue boundaries. Furthermore, computational resources are spent in regions of parameter

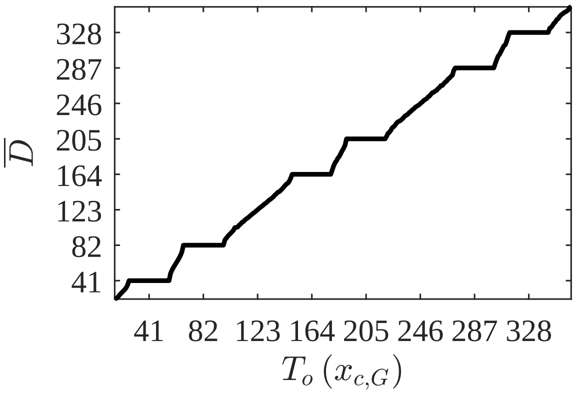


Figure 4.11: Devil's staircase for the branch switching oscillator (4.14), showing an estimate of the average duration of the oscillator as a function of internal period (in turn as function of upper threshold $x_{c,G}$). There is an infinite number of "steps" of constant average duration, each having positive width. Only the widest steps are visible

space where little interesting occurs. An alternative method to find bifurcations is through numerical continuation. In numerical continuation, an algorithm tracks a bifurcation hypersurface (typically a one-dimensional curve) in parameter space, based on defining conditions for bifurcations. Thus, only system equations close to a bifurcation surface needs to be computed. However, even though there advanced software for numerical continuation, such as AUTO [Doedel, 1997] and COCO [Dankowicz and Schilder, 2009], continuation of exotic problems involving e.g. more than two bifurcation parameters, hybrid dynamical systems, and singular fast-slow systems [Kuehn, 2015] is still not completely developed.

Since we are only interested in $1 : N$ frequency locking regions in this thesis, the naïve estimation of the average duration suffices for our purposes.

4.7.5 Quasiperiodic motion and forcing

If the average duration of a solution of a periodically forced system is *not* an integer of the forcing period, then we say that the motion is *quasiperiodic*, meaning that solutions are not periodic but in some sense characterised by a small set of frequencies. Quasiperiodic motion generically takes place in between Arnold tongue diagrams for small forcing [Pikovsky et al., 2001]. We say that the motion of a quasiperiodic solution lives on the (two-dimensional) torus $S_1 \times S_1$, where S_1 denotes the circle. This is because the motion must be described by two angle variables, one corresponding to the oscillator phase (which we have not cared to define) and the other to the forcing phase (Figure 4.12). Quasiperiodic motion *densely* covers the torus, meaning that every point on the torus wither belongs to a solution trajectory in angle space, or is the limit of some other point on the trajectory [Pikovsky et al., 2001].

If the forcing of the system contains multiple frequencies $T_1, T_2, ..., T_m$, then frequency locking can occur if a multiple of the average duration is a linear combination of the constituent periods

$$M\overline{D} = N_1 T_1 + N_2 T_2 + ...N_m T_m,$$

for non-negative (but possibly zero) integers M and N_m. This is possible even if the forcing itself is quasiperiodic, meaning that $T_1, T_2, \ldots$ are incommensurate, such that there does **not** exist integers $N_1, N_2, \ldots$ such that $N_i T_i = \sum_{i \neq j} N_j T_j$.

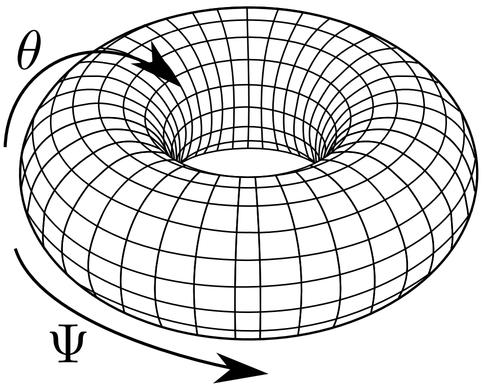


Figure 4.12: Illustration of a torus. The angle θ represents the forcing phase and the angle Ψ the oscillator phase. The figure is adapted from [YassineMrabet, 2007] under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license

However, two remarks are in place. The first is that the average duration only provides limited information. It disregards temporal information since it only quantifies an average of system behaviour in the asymptotic limit as time goes to infinity. Furthermore, it disregards information about the distribution of durations. In Chapter 8 we aim to shed light on these questions for the case of amplitude and frequency modulated sinusoidal forcing. The second remark is that quasiperiodic forcing can give rise to strange dynamics, in the most literal sense.

4.7.6 Strange behaviour: deterministic chaos

In the previous sections we studied systems which either tend to a fixed point, a limit cycle, or evenly fill a torus. However, systems can exhibit more exotic behaviour.

Perhaps the most commonly known exotic behaviour is deterministic *chaos*, popularised by the analogy to the "butterfly effect" analogy which suggests that a butterfly in Brazil could cause a hurricane in Texas. Although there exist multiple definitions of chaos, they all have a few aspects in common Strogatz [1994]. One is sensitive dependence on initial conditions, meaning that trajectories starting nearby each other diverge exponentially. Another is that the system is topologically mixing, meaning that dynamics eventually revisit every open neighbourhood of every other point in the set. This precludes dynamics which diverge to infinity from being classified as chaotic. Hence, the dynamics must be confined to a compact set, even though nearby trajectories diverge exponentially.

The solution to this conundrum are strange chaotic attractors; set in phase space on which dynamics are chaotic. The word "strange" comes from that the set has a fractal structure, meaning that it in a certain sense has a dimension between the integer dimensions. The Lorenz attractor, a strange chaotic attractor in $\mathbb{R}^3$ found in a simplified model of atmospheric convection, looks like it could consist of a complex of surfaces [Lorenz, 1963]. However, in fact it in covers slightly more volume in the sense of the box counting dimension, which is 2.06 [McGuinness, 1983]. Trajectories can diverge locally on the attractor, but since trajectories are bounded globally, the maximum distance between trajectories remain bounded.

Chaotic motion can occur in forced oscillators, including the branch switching oscillator (4.14) and the periodically forced van der Pol oscillator [Haiduc, 2008]. The "noise" heard by van der Pol and van der Mark [1927] when converting model output to sound was in fact deterministic chaos. Although chaotic motion generally occurs for a small set of parameters in periodically forced low-dimensional dynamical systems, Ashwin et al. [2018] found that quasiperiodic forcing in low-dimensional models of glacial cycles made chaotic behaviour more common.

One important aspect of strange chaotic attractors is what they *are not*; in particular the behaviour is neither static nor periodic. Furthermore, motion is fundamentally irregular, unlike quasiperiodic motion. We note, however, that there is no characteristic "appearance" of chaotic trajectories. They can look oscillatory as for the forced van der Pol oscillator in the chaotic regime [de Saedeleer et al., 2013], just that no cycle is identical to the other, and of course that nearby trajectories diverge.

One way to identify and quantify chaos are Lyapunov exponents. For a one-dimensional map $f(\theta)$, there is only one Lyapunov exponent, and it is defined as

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log \left| \frac{\partial f^i(\theta)}{\partial \theta} \right|,$$

for an initial θ, where log denotes the natural logarithm and the exponent denotes repeated function composition. The expression can be understood as the average expansion rate along a trajectory of the system, since $\partial f^i(\theta)/\partial\theta$ is the linearised growth rate of the map at each iterate. If the Lyapunov exponent is positive on an attractor, then the attractor is chaotic. If the exponent is negative, then nearby trajectories converge, and if it is zero nearby trajectories neither diverge nor converge. As an example, a two-cycle $\theta_1, \theta_2, \theta_1, \ldots$ with slopes $f'(\theta_1) = e$ and $f'(\theta_2) = e^{-2}$ has a Lyapunov exponent $\lambda = (1/2)(1 + (-2)) = -1/2 < 0$, so the attractor is non-chaotic (which of course also follows from it being both an attractor and a cycle). This example shows that a system can be locally sensitive to perturbations even though the Lyapunov exponent is negative (since $|f'(\theta_1)| > 0$); the Lyapunov exponent only encodes the average long-term net expansion of nearby trajectories. Note that the Lyapunov exponent does not encode stability of an attractor; otherwise there could not be chaotic attractors.

In the n-dimensional case, an attractor has a set of n Lyapunov exponents, forming the Lyapunov spectrum. In the continuous time case, Lyapunov exponents are determined from singular values of solutions to the variational equation, which describes the evolution of vectors in the tangent spaces of points along the trajectory. In continuous systems limit cycles have at least one zero Lyapunov exponent, corresponding to the phase along the trajectory.

4.7.7 Strange non-chaotic attractors

Yet another type of attractors are strange non-chaotic attractors. As the name implies, the attractor has a fractal structure, but unlike chaotic attractors nearby trajectories converge asymptotically. Strange non-chaotic attractors occur in quasiperiodically forced systems, and were discovered at late as in the 1980'ies [Grebogi et al., 1984]. In the context of glacial cycles, Mitsui et al. [2013, 2015], de Saedeleer et al. [2013] found that strange non-chaotic attractors appear in several models under quasiperiodic forcing, and also noted that strange non-chaotic attractors can be sensitive to noise, even though trajectories converge asymptotically in the noise-free case.

4.7.8 Stability of the first return map

The stability of the first return map can be inferred from the derivative of the first return map (4.20)

$$\frac{\partial \theta_1}{\partial \theta_0} = \omega_f \frac{\partial t_2}{\partial \theta_0} + \frac{\partial \phi}{\partial \theta_0} = \omega_f \frac{\partial t_2}{\partial \phi} + 1,$$

since $\theta_0 = \phi$. We find $\frac{\partial t_2}{\partial \phi}$ by implicit differentiation of the first line of (4.21)

$$0 = \mu \frac{\partial t_1}{\partial \phi} - \lambda \left(\frac{\partial t_2}{\partial \phi} - \frac{\partial t_1}{\partial \phi} \right) - \frac{A}{\omega_f} \left(\left(\omega_f \frac{\partial t_2}{\partial \phi} + 1 \right) \sin(\omega_f t_2 + \phi) - \sin\phi \right).$$

The derivative $\frac{\partial t_2}{\partial \phi}$ can be isolated to give

$$\frac{\partial t_2}{\partial \phi} = \frac{1}{\lambda + A\sin(\omega_f t_2 + \phi)}\left(\frac{\partial t_1}{\partial \phi}(\mu + \lambda) + \frac{A}{\omega_f}(\sin\phi - \sin(\omega_f t_2 + \phi))\right), \tag{4.24}$$

whenever the denominator is nonzero. Similarly, $\frac{\partial t_1}{\partial \phi}$ can be found by differentiating (4.18)

$$0 = \mu\frac{\partial t_1}{\partial \phi} - \frac{A}{\omega_f}\left(\left(\omega_f\frac{\partial t_1}{\partial \phi} + 1\right)\sin(\omega_f t_1 + \phi) - \sin\phi\right).$$

Isolating $\frac{\partial t_1}{\partial \phi}$ gives that

$$\frac{\partial t_1}{\partial \phi} = \frac{1}{\mu - A\sin(\omega_f t_1 + \phi)}\frac{A}{\omega_f}\left(\sin(\omega_f t_1 + \phi) - \sin\phi\right),$$

whenever the denominator is non-zero. We could substitute this expression directly into (4.24) to obtain a general form of the derivative of the return map. But instead, we consider immediately the case of $1 : N$ frequency locking, in which case $t_2 = NT_f$ and $t_1 = NT_f\lambda/(\mu + \lambda)$. Then (4.24) simplifies, and upon substitution we get

$$\frac{\partial t_2}{\partial \phi} = \frac{A}{\omega_f}\frac{\mu + \lambda}{(\lambda + A\sin\phi)(\mu - A\sin(\nu + \phi))}\left(\sin(\nu + \phi) - \sin\phi\right),$$

where we recall the definition $\nu = 2\pi N\lambda/(\lambda + \mu)$. Substituting into the expression for the first return map and simplifying gives that

$$\frac{\partial \theta_1}{\partial \phi} = \frac{(\lambda + A\sin(\nu + \phi))(\mu - A\sin\phi)}{(\lambda + A\sin\phi)(\mu - A\sin(\nu + \phi))}. \tag{4.25}$$

The magnitude of (4.25) for general ϕ is generally cumbersome to work out. However, we can show that the absolute value of (4.25) is 1 at the borders of Arnold tongues (assuming a frequency locked solution exists there). By showing that the second derivative $\partial^2\theta_1/\partial\phi^2 \neq 0$, it follows from the classical theorem of saddle-node bifurcations of maps [Kuznetsov, 2004] that for parameters in a neighbourhood of the bifurcation point, that there exist two nearby fixed points of different signs.

On the border of Arnold tongues the difference of cosines (4.22) obtains an extremum value, which gives a fixed point phase

$$\phi^* = -n\pi - \arctan\left(\frac{\sin\nu}{\cos\nu - 1}\right),$$

for some integer n. Inserting this ϕ^* into (4.25) and using standard trigonometric identities, one can show that indeed

$$\frac{\partial \theta_1}{\partial \phi}|_{\phi=\phi^*} = 1.$$

The second derivative at the boundary of an Arnold tongue can be found to be

$$\frac{\partial^2 \theta_1}{\partial \phi^2} = \frac{-A\cos(\phi)(\lambda + \mu)}{(\lambda + A\sin(\nu + \phi))(\mu - A\sin\phi)}. \tag{4.26}$$

This is only zero if $\cos\phi \neq 0$, which only occurs at the bottom of tongues, where $T_o = NT_f$. There, the return map is highly nondegenerate, with a continuum of equilibria along the identity line. Hence, if for $T_o \neq NT_f$ there exists an equilibrium on a tongue boundary, then it is a generic saddle-node bifurcation, and under generic perturbation of parameters there is one stable and one unstable fixed point. This is of course under the assumption that the denominator of (4.26) is non-zero, and that there exist t_1 and t_2 as assumed for $1 : N$ frequency locking.

4.8 Summary

This chapter has provided a brief introduction to dynamical systems, differential equations, bifurcation and synchronisation. The fold system was taken as an example of a dynamical system exhibiting bifurcation when a vector field defined by $\dot{x} = g(x)$ is perturbed by a bifurcation parameter λ. We saw that as λ passed through zero, the number of solutions to the equation $g(x) = 0$ changed, marking a qualitative change: a bifurcation. We then encountered relaxation oscillations, a form of limit cycle, in the van der Pol oscillator. The van der Pol oscillator has a scale separation parameter ϵ, which when small makes one variable much faster than the other. This is a form of fast-slow system, which is called singular in the limit as $\epsilon \to 0$.

Next, we applied periodic external stimulus to the van der Pol oscillator, called forcing. We then simplified the model to one of integrate-and-fire type, with explicit thresholds that govern switches between two distinct growth and decay states. This model we saw can become frequency locked to the forcing, meaning that the period of the oscillator is a rational number times the forcing frequencies. Frequency locking persists on open sets in parameter space; in two dimensions these sets are called Arnold tongues, while in one dimension they are called Devil's staircases. Frequency locking was defined through a Poincaré map, a section in phase space which solutions return to once per "revolution". Frequency locking manifests as fixed points of that map, and we learned that the creation of frequency locking solutions (for small enough forcing strength) occurs through a saddle-node bifurcation of maps.

Lastly, we mentioned that forcing with multiple frequencies can give rise to strange chaotic attractors, bounded sets with sensitive dependence on initial conditions and a fractal geometry. Moreover, yet another possible type of attractor has a strange geometry, but no sensitive dependence on initial conditions.

Chapter 5

Bifurcation theory

Bifurcation theory is the study of qualitative change under small perturbations. Famous examples are the instantaneous buckling of a vertical beam at a critical load [Levien, 2008] and the clearing of turbid lakes with the introduction of small amounts of predatory fish [Scheffer, 2009].

The precise meaning of bifurcation depends on context. However, every definition builds on two concepts: equivalence and bifurcation parameters. Bifurcation of an object X occurs when neighbouring objects, perturbed by a parameter λ are in some sense non-equivalent to X. The fold bifurcation of Section 4.2 serves as an introductory example.

Consider the function $f(x) = x^2$ in the space of infinitely smooth functions $C^\infty(\mathbb{R})$ (such that partial derivatives of all orders exist). Take as equivalence relation $\sim$ the number of zeros of $f(x)$. This fulfils all requirements of an equivalence relation, because $f \sim f$ (reflexivity), if $f \sim g$ then $g \sim f$ (symmetry), and if $f \sim g$ and $g \sim h$ then $f \sim h$ (transitivity). Under the simplest possible perturbation of a constant λ

$$f(x, \lambda) = x^2 - \lambda,$$

it is clear that the number of solutions to $f(x, \lambda) = 0$ goes from 0 to 1 to 2 as λ increases from negative to positive values through the critical parameter value $\lambda = \lambda_0 = 0$. Thus, the perturbed neighbours of $f(x, \lambda_0)$ are inequivalent under the “number of zeros” equivalence for arbitrarily small deviations from the critical value $\lambda - \lambda_0$. We say that $f(x, \lambda)$ undergoes bifurcation at $\lambda = \lambda_0$ under the “number of zeros” equivalence. Note that $f(x, \lambda)$ is a family of functions parametrised by the bifurcation, or unfolding, parameter λ.

Next, we give an introduction to a form of equivalence due to [Golubitsky and Schaeffer, 1985] which will serve as a basis for Paper 2 in Chapter 8. Focus will be on an intuitive understanding of the concepts. Lastly, we mention other commonly used senses of bifurcation in autonomous (not explicitly dependent on time) systems, and briefly mention bifurcation in nonautonomous systems.

5.1 Singularity theory

Singularity theory concerns the zeros of functions and has a long history [Andronov and Pontryagin, 1937, Whitney, 1955, Mather, 1969b,a, Thom, 1972]. Although initiated by Andronov and Pontryagin [1937] in the 1930’ies, the formulation of catastrophe theory by Thom [1972] captured the interest of a broad audience. Here, we restrict ourselves to infinitely smooth functions of one variable $g \in C^\infty(\mathbb{R})$. Consider the zero set of the function

$$g(x) = 0.$$

In one-dimensional singularity theory (c.f. the contact equivalence in [Golubitsky and Schaeffer, 1985, p. 166]), we say that a function $\hat{g}$ is equivalent ($\sim$) to a function g if there exists a diffeomorphism $X(x)$ and a positive function $S(x) > 0$ such that

$$\tilde{g}(x) = S(x)g(X(x)). \tag{5.1}$$

Recall that a diffeomorphism of functions is a differentiable map between functions with a differentiable inverse. The transformation (5.1) preserves the number of zeros of g since $S(x)$ is a positive rescaling of time, and $X(x)$ is a smooth rescaling of the x axis.

For example, the function

$$\hat{g}(x) = e^x(x + x^3))$$

is equivalent to $g(x) = x$, since $S(x) = e^x$ is continuous and positive, and $X(x) = x + x^3$ is smooth with a smooth inverse. However, e.g. $X(x) = x - x^3$ is not equivalent to x because it is not invertible. Note that the invertibility of the functions g and $\tilde{g}$ do not matter; only the invertibility of the mapping between them. Thus e.g. $g(x) = x - x^3$ and $g(x) = 2x - 8x^3$ are obviously equivalent.

Typically, the equivalence is not given globally, but locally to a point (without loss of generality restricted to the origin). For $\tilde{g}$ to be locally equivalent to g, we only ask that they are equivalent according to (5.1) in some neighbourhood U around the origin. Defining explicit neighbourhoods on which equivalence holds is cumbersome, since one has to keep track how neighbourhoods are transformed under (5.1). To avoid this, it is typical to introduce the concept of *germs* [Golubitsky and Schaeffer, 1985]. We say that f and g are *equal as germs* if there exists a neighbourhood U, however small, on which $f(x) \sim g(x)$.

It may occur to the reader that if *analytical* functions G and g are equivalent on some proper subset U of $\mathbb{R}$, then they are equal on all of $\mathbb{R}$ since the Taylor series converges to the function everywhere. Therefore, the need for germs may seem superfluous. However, non-analytic functions may have identical Taylor series at a point and yet be nonidentical. A classic example is the infinitely smooth but non-analytical (cf [Golubitsky and Schaeffer, 1985])

$$f(x) = \begin{cases} e^{-1/(1-x)}, & \text{if } x > 1 \\ 0, & \text{if } x \leq 1. \end{cases}$$

The Taylor coefficients of this function are all 0 at the origin, just as the Taylor coefficients of function $g(x) \equiv 0$. However, the two functions are clearly not identical on all of $\mathbb{R}$.

5.2 Introduction to unfoldings

Golubitsky and Schaeffer [1985] state two relevant subproblems regarding equivalence and bifurcation. Given functions g and $\tilde{g}$,

1. How can we tell if $g \sim \tilde{g}$?
2. In what "ways" can g and $\tilde{g}$ be non-equivalent?

The first problem is called the recognition problem. Rather spectacularly, it is possible to tell from a small number of partial derivatives whether g and $\tilde{g}$ are equivalent. For example, if $\tilde{g} = \tilde{g}_x = 0$ and $\tilde{g}_{xx}$ at the origin (where $\tilde{g}_x = d\tilde{g}/dx$), then $\tilde{g} \sim g(x) = \pm x^2$, for one choice of sign. We shall return to this problem shortly.

The second problem, the classification problem, is how to systematically classify ways for equivalent systems g and $\tilde{g}$ to become inequivalent. This motivates the introduction of concepts such as perturbation, unfolding, and codimension.

We say that $G(x, \alpha)$, where $\alpha \in \mathbb{R}^k$ is a k-parameter *unfolding* of $g(x)$ if $G(x, 0) = g(x)$. An unfolding $G(x, \alpha)$ can be thought of as a parametrised perturbation $p(x)$ to $g(x)$, since

$$G(x, \alpha) = g(x) + (G(x, \alpha) - G(x, 0)).$$

For example, $G(x, \alpha) = x^2 - \alpha$ is a one-parameter unfolding of $g(x) = x^2$. These k-parameter unfoldings enable a classification of perturbations based on the number k of unfolding parameters. However, the number of parameters in the most obvious sense is insufficient, since e.g. $G(x,\alpha) = x^2 + \alpha_1 + \alpha_2$ should belong to the same class as $G(x,\alpha) = x^2 + \alpha_1$. We want a classification of perturbations based on the minimum necessary number of parameters included. The concept of universal unfolding provides this.

A *versal unfolding* of g is an unfolding $G(x,\alpha)$ such that every other unfolding of g factors through G [Golubitsky and Schaeffer, 1985, p. 121]. That $H(x,\beta)$ *factors* through $G(x,\alpha)$ means that there exist smooth functions X, S and W such that

$$H(x,\beta) = S(x,\beta)H(X(x,\beta), W(\beta)),$$

and $S(x,0) = 1$, $X(x,0) = x$ and $W(0) = 0$. A *universal unfolding* is an unfolding with a minimal number of parameters. For example,

$$H(x,\beta) = x^2 + 2\beta_1 x + \beta_2$$

factors through

$$G(x,\beta) = x^2 + \beta_1 + \beta_2,$$

since the smooth transforms $S(x,\beta) \equiv 1$, $X(x,\beta) = x + \beta_1$ and $W(\beta) : (\beta_1,\beta_2) = (-\beta_1^2, \beta_2)$ transforms $H(x,\beta)$ into $G(x,\beta)$. However, $G(x,\beta)$ is not universal since it factors into yet another one-parameter unfolding

$$K(x,\beta) = x^2 + \beta_1,$$

defined by $S(x,\beta) \equiv 0$, $X(x,\beta) = x$ and $W(\beta) : (\beta_1,\beta_2) \mapsto \beta_1$.

If g has a universal unfolding containing a finite number of parameters, then the number of such parameters is called the *codimension* of the unfolding. If no such unfolding exists, we say that it has infinite codimension. In the previous example the universal unfolding $g(x,\alpha) = x^2 + \alpha$ is codimension one. We do not prove this classic result (see e.g. [Arnold et al., 1999]).

We note that codimension under equivalence (5.1) corresponds to the number of equality conditions imposed on $g(x)$, in addition to the assumed one $g(x) = 0$. It is typical for codimension to coincide with the number of constraints.

5.3 Singular classification and normal forms

The concept of codimension allows us to classify perturbations by codimension. It can be shown that under the equivalence (5.1) the only codimension one universal unfolding of $g(x)$ is $g(x,\alpha) = x^2 + \alpha$, recognised as the fold bifurcation in Section 4.2. A classic example of a codimension two unfolding is the pitchfork bifurcation $g(x) = x^3 + \alpha_1 x + \alpha_2$.

Since every unfolding of $g(x) = x^2$ factors through $G(x,\alpha) = x^2 + \alpha$, we can take that universal unfolding as representative for all unfoldings of $g(x) = x^2$. Such a representative is called a *normal form.*

We have used the word bifurcation to describe the perturbed family $g(x,\alpha) = x^2 + \alpha$. This is because for $g(x,0)$ the function is nonequivalent to nearby perturbations, and the parameter α "unfolds" the bifurcation. If the codimension of some $g(x)$ is 0, then there is no bifurcation. Such a function is called *non-degenerate* or *generic*, meaning that the function persists on an open set in parameter space (it is equivalent to nearby perturbations). In contrast, *degenerate* functions have at least some non-equivalent neighbours.

Classification of unfoldings of vector fields have been performed under slightly different equivalences than (5.1) (see e.g. [Arnold et al., 1999, Thom, 1972]). However, in the next section we introduce the notion of distinguished parameters due to Golubitsky and Schaeffer [1985], which gives another classification of unfoldings which is of interest in Chapter 8.

5.4 One distinguished parameter

Treating a subset of parameters as *distinguished parameters* instead of unfolding parameters gives rise to a new classification, of interest for fast-slow systems (Chapter 8). Rather than perturbing zero sets of vector fields $g(x) = 0$, one perturbs *bifurcation diagrams* $g(x, y) = 0$, where y is now a distinguished parameter. For example, with one variable and one distinguished parameter, the zero set of $g(x, y) = x^2 - y$ defines a bifurcation diagram (Figure 4.2 b) in Section (4.3)). An example perturbation of this $g(x, y)$ is $G(x, y, \alpha) = x^2 - y + \alpha_1 x + \alpha_2 xy$. We say that two bifurcation diagrams $\tilde{g}$ and g are equivalent if there exists a diffeomorphism $\Phi(x, y) = (X(x, y), Y(y)) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}\times\mathbb{R}$ and a continuous positive function $S(x, y) : \mathbb{R} \to \mathbb{R}$ such that

$$\tilde{g} = S(x, y)g(X(x, y), Y(y)). \tag{5.2}$$

We do not go into details into how this equivalence differs from that of (5.1); this we cover in Chapter 8.

5.5 Bifurcation under other equivalences

The most common notion of bifurcation of autonomous ordinary differential (or difference) equations (without distinguished parameters) is based on *topological equivalence* [Kuznetsov, 2004]. We say that two (continuous or discrete time) systems are topologically equivalent if there exists a homeomorphism $h : \mathbb{R}^n \to \mathbb{R}^n$ that maps the solutions of one system $x_g(t)$ (governed by $\dot{x} = g(x)$) to solutions of another system $x_f(t)$ (governed by $\dot{x} = f(x)$). Intuitively, topological equivalence is an equivalence of phase portraits of two systems. Notably, since h is not a diffeomorphism, but only a homeomorphism, this equivalence allows for transformations with "kinks". For instance, a spiral equilibrium which locally attracts trajectories in a spiralling fashion is topologically equivalent to a node equilibrium which attracts without "rotation", but it is not smoothly equivalent since a transformation between the vector fields at the origin necessarily is non-smooth. See for instance [Kuznetsov, 2004, Chapter 2].

We do not go into details of classification of bifurcation under topological equivalence, but just mention one important bifurcation called *Hopf bifurcation* (mentioned briefly in Section (4.5.1)). At a Hopf bifurcation an equilibrium changes stability, and (at least one) limit cycle is born form the equilibrium. Hopf bifurcation is thus one way through which self-sustained oscillations can form (of relevance for some glacial cycle models, e.g. [Maasch and Salzman, 1990]). Hopf bifurcation comes in a supercritical and a subcritical form. In the former case, a stable equilibrium becomes unstable, and a small stable limit cycle with amplitude on the order $\lambda^{1/2}$ emerges from the equilibrium. In the latter case, an unstable equilibrium becomes stable and an unstable unstable limit cyde is born. However, often (as for [Maasch and Salzman, 1990]) there is a stable limit cycle far from the equilbrium, such that upon varying the bifurcation parameter, the system jumps rapidly from a stable equilibrium to a distant attracting limit cycle.

Bifurcation of non-autonomous systems is mostly treated bifurcation in discrete-time autonomous systems defined through Poincaré maps, as we saw in Section (4.6). However, there is a parallel movement towards defining bifurcation and notions of attractivity for nonautonomous systems directly, see for instance [Rasmussen, 2006, Kloeden and Rasmussen, 2011]. In the context of glacial cycle models, the notion of pullback attractor garnered attention in [Crucifix, 2013, de Saedeleer et al., 2013]. Loosely speaking, a pullback attractor is an object at present time to which trajectories starting from an open set of initial conditions at time negative infinity converge.

Bifurcation in nonsmooth and hybrid dynamical systems have their own theoretical framework, and consequently bifurcations, see e.g. [di Bernardo et al., 2008, Guckenheimer and Johnson, 1995]. We do not encounter such bifurcations in this thesis.

5.6 Summary

We have given a brief introduction to bifurcation theory, the study of qualitative changes under perturbation. We learned that bifurcation occurs when the neighbours of an object are inequivalent to the object, according to an equivalence which must be specified. Perturbations are parametrised by a bifurcation parameter, and for particular values of the bifurcation parameter bifurcation occurs. We took as an example the fold system $g(x) = x^2$, and found that under contact equivalence the system is degenerate, meaning that it is inequivalent to its perturbed neighbours. A one-parameter perturbation λ is makes all systems $g(x; \lambda) = 0$ equivalent for any $\lambda \neq 0$. Therefore, we call $x^2 + \lambda$ is a universal unfolding of $g(x) = x^2$ under contact equivalence, and the codimension of the bifurcation is one.

Under a different *distinguished parameter* equivalence due to [Golubitsky and Schaeffer, 1985], the fold $g(x, y) = x^2 + y$ is equivalent to all its perturbed neighbours. Therefore, it is codimension one and we call it persistent. This equivalence we find is appropriate in the context of singular fast-slow systems in Chapter 8.

We finished the section by briefly introducing other notions of equivalence. We mentioned the Hopf bifurcation which marks a transition between equilibria and limit cycles. This is an important bifurcation under a commonly used topological equivalence.

Chapter 6

Modelling of glacial cycles

Equipped with notions of dynamical systems and bifurcation from the previous sections, we now review the development of models of the glacial cycles. First, we broadly sketch how glacial cycle models have cycles evolved through time. Then, we discuss these models in the context of the MPT.

6.1 Conceptual models of glacial cycles through time

Weertman [1961] showed with a model that stable ice sheets can shrink rapidly once a critical value of some parameter is crossed. This provided the mathematical foundation for models of glacial cycles involving ice sheet instability which had so far only been phrased phenomenologically. As a next step, Weertman [1976] tested whether insolation variations as proposed by Milanković could cause ice sheets to rise and collapse, producing glacial cycles. This was shown to be the case, although some assumptions had to be stretched far, and the resulting time series differed markedly from proxy data.

In 1979 Källén et al. [1979] coupled an unforced version of Weertman's model with an energy balance model of Budyko-Sellers type [Budyko, 1969, Sellers, 1969]. They found that the model possessed self sustained oscillations due to a temperature-precipitation feedback, which caused more snowfall when temperatures were high, but the internal oscillations were on the order of 10 kyr, much shorter than the observed ∼100 kyr. A proposed solution, evaluated by [Le Treut and Ghil, 1983], was that insolation variations with longer period such as the ∼100 kyr period of eccentricity could interact with the shorter internal period to produce ∼100 kyr long glacial cycles. The forced model did show ∼100 kyr variations at a stronger amplitude than in the forcing (compared to the Fourier coefficients of eccentricity). However, these variations were stills smaller than those at the forcing and internal periods.

A couple of years earlier, [Imbrie and Imbrie, 1980] had deviced a simple rule based model

$$\dot{x} = (x_{target}(t) - x)/\tau_i, \tag{6.1}$$

where x represents global ice volume, $x_{target}(t)$ is an equilibrium ice volume which varies with the negative of insolation variations $x_{target} \propto -F(t)$, and τ_i is a time scale of ice sheet response, assumed to be longer in glacial times ($\tau_i = \tau_g$) than in "deglacial" times ($\tau_i = \tau_d$), $\tau_g \gg \tau_d$. The system switches from the glaciating state to the deglaciating state when ice volume exceeds the target ice volume $x \geq x_{target}(t)$ and back to the glaciating state once $x < x_{target}(t)$. A simulation of the model is shown in Figure 6.1. The $\sim$ 100 kyr variations occur because of the amplitude modulation of precession (with dominant period ∼21 kyr) by eccentricity (period ∼100 kyr). The solution "surfs" the 100 kyr lower envelope of the forcing, since the system adjusts more rapidly to the insolation curve in the deglaciating state than the glaciating state. If ice volume is relatively high and insolation is strong, then there can be a rapid deglaciation. One shortcoming of the [Imbrie and Imbrie, 1980] model

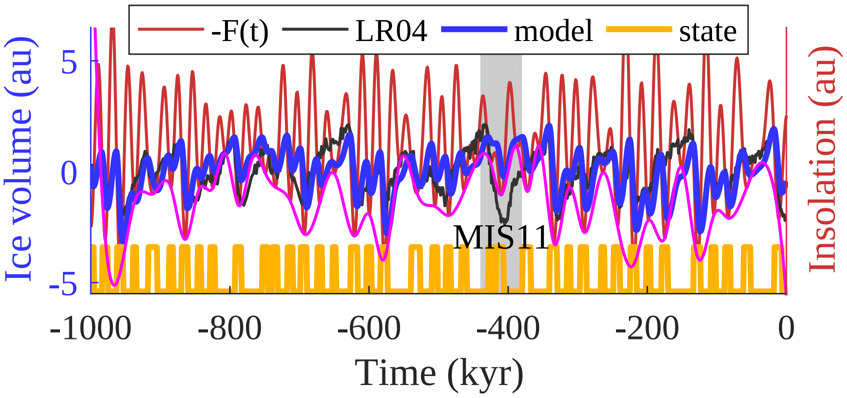


Figure 6.1: Simulation of the Imbrie and Imbrie [1980] model. The thick blue curve shows the model simulation and the black curve shows the LR04 stack, a proxy for global ice volume [Lisiecki and Raymo, 2005]. Both curves are normalised to zero mean and unit variance.The red curve is negative insolation $-F(t)$ with ~2/3 amplitude from precession and 1/3 amplitude from obliquity (see [Imbrie and Imbrie, 1980] for details). The yellow curve at the bottom shows whether the solution is in a glaciating (high) or deglaciating (low) state. The magenta curve is a spline (piecewise third order polynomial) lower envelope of the forcing. The shaded area highlights the "Marine Isotope Stage 11" (MIS 11) problem, which is the inability of the model to produce a deep interglacial observed in the LR04 stack around -400 kyr. Insolation has been slightly rescaled and clipped to improve the readability of the figure.

is that it does not explain the large glacial cycle around -400 kyr which occurs at a time when precessional forcing is weak. The problem takes the name "Stage 11" problem, from the interglacial (warm) epoch named Marine Isotope Stage 11. Solutions to this problem is addressed by some later models. Another shortcoming, called the "400"-kyr problem, is the the model shows strong variability at the 400 kyr period, a modulating frequency of precession, while this period is absent from proxy data. The presence of 400 kyr power is again due to "surfing" the envelope.

Another line of models was conceived by Barry Salzman and coworkers throughout the 1980'ies and into the 1990'ies. Central to several of these (e.g. [Saltzman and Sutera, 1984, Saltzman et al., 1984, Maasch and Salzman, 1990]) is the interaction between global ice volume, atmospheric CO_2, and sequestration of CO_2 by deep water formation. We present here a set of equations from [Maasch and Salzman, 1990] to illustrate the kind of model:

$$\begin{aligned} \dot{x} &= -x - y - uF(t) \\ \dot{y} &= -pz + ry + sz^2 - z^2y \\ \dot{z} &= -q(x+z). \end{aligned} \tag{6.2}$$

Here, x represents global ice volume, y represents atmospheric CO_2, z represents intensity of North Atlantic deep water formation, and $F(t)$ are insolation variations at 65 degrees North (slightly different from the July insolation in the publication). The parameters u, q, s are all positive constants, and $p = p(t)$ and $r = r(t)$ are positive parameters allowed to increase over time to reflect slow tectonic changes. A simulation of the model can be seen in Figure 6.2.

Although a lot of processes and assumptions go into (6.2), the model has two dynamic regimes which can be understood heuristically as follows. If $p > r$ and small enough, the

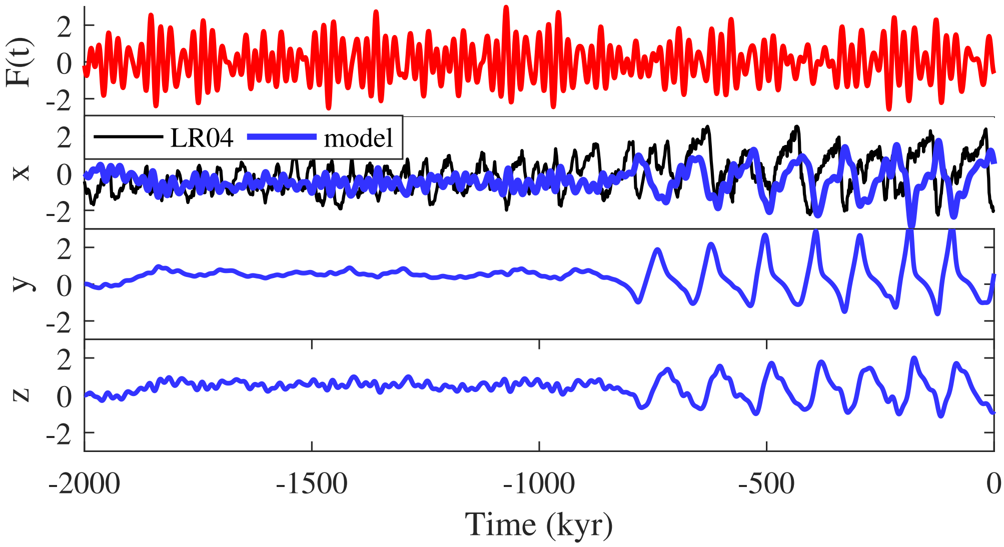


Figure 6.2: Simulation of the Maasch and Salzman [1990] model. The top panel shows summer solstice insolation variations at 65 degrees North. Ice volume x, atmospheric carbon dioxide y, and deep water formation z, all normalised, are shown in the lower panels. In the second top panel (for x) is the LR04 stack plotted for comparison (thin black curve). The parameters p and r in (6.2) are ramped linearly, as in the original publication.

(unforced $u = 0$) model has a stable equilibrium since both atmospheric CO_2 and ice volume feed back negatively to each other, reflected by the negative signs in (6.2). If $p < r$, the equilibrium becomes unstable, and the dynamics are attracted to a limit cycle as seen in Figure 6.2. In this regime, ice volume growth causes reduction of CO_2 due to increased weathering, which in turn feeds back to more ice volume growth [Saltzman and Maasch, 1988]. At the same time, the North Atlantic deep water formation which draws down CO_2 progressively weakens. However, at a critical ice volume level the ocean convection becomes unstable, leading to an upsurge of CO_2 and associated loss of ice volume, until deep water formation resumes once enough ice is gone. The model thus produces self-sustained glacial cycles explicitly as a set of feedbacks in the climate system. The period of these unforced cycles is close to 100 kyr, indicating that the glacial cycles can occur even without insolation variations.

Dynamically, changing p and r causes a stable equilibrium to turn unstable in a subcritical Hopf bifurcation, destroying at the same time an unstable limit cycle surrounding equilibrium. Thereafter, the dynamics tend to a large stable limit cycle which exists both before and after the bifurcation. This transition is one of the first dynamical mechanisms proposed to have caused the MPT.

Adding insolation variations has two effects. Firstly, they produce linearly forced glacial cycles prior to ∼800, and secondly they pace the self-sustained cycles in the late Pleistocene such that glacial terminations depend on the forcing. The directly forced glacial cycles have a few problems: As pointed out in [Maasch and Salzman, 1990], they create more 21 kyr precessional variability than is found in data. Furthermore, [Ashkenazy and Tziperman, 2004] showed that such linearly forced cycles are too symmetric compared to glacial cycles in proxy data. Nonetheless, the hypothesis that the Early Pleistocene glacial cycles are linearly forced lives on.

Further steps towards understanding the middle Pleistocene transition was taken by [Paillard, 1998]. Paillard constructed a phenomenological model (PP98 henceforth) for ice volume growth x, containing three states: an interglacial i, mild glacial g, and deep glacial

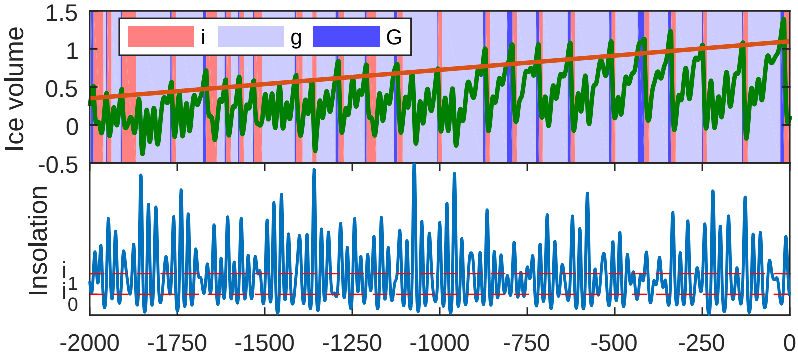


Figure 6.3: Simulation of the PP98 model Paillard [1998]. Ice volume x (green curve) is shown in the top panel, along with the ice volume threshold $x_{max}(t)$ (red line). The background colours indicate climate state i, g or G as per the legend. The lower panel shows the truncated insolation forcing function, with thresholds governing the $i \rightarrow g$ and $G \rightarrow i$ climate state switches. The parameters are as in the original publication.

G state. In each of the states $j \in \{i, g, G\}$, ice volume growth is governed by the linear differential equation

$$\dot{x} = \frac{(x_{target,j} - x)}{\tau_j} - A(t)F(t), \tag{6.3}$$

where $x_{target,j}$ is a state dependent target ice volume, τ_j is a state dependent response time, $A(t)$ is a time dependent forcing amplitude, and $F(t)$ is a truncated forcing function. The state switch rules are

$$\begin{aligned} i \rightarrow g \text{ if } A(t)F(t) < i_0 \\ g \rightarrow G \text{ if } x \geq x_{max}(t) \\ G \rightarrow i \text{ if } A(t)F(t) > i_1 \end{aligned} \tag{6.4}$$
$$,$$

where $x_{max}(t)$ is a critical ice volume threshold that increases linearly over time, and i_0 and i_1 are insolation thresholds. A typical simulation of the model is shown in Figure 6.3. For a detailed description of the model, see Chapter 10. Here, we only discuss it briefly.

The equation (6.3) is superficially similar to the Imbrie model (6.1), but it is fundamentally different. In PP98 ice volume grows even without insolation variations $A(t) \equiv 0$, unlike the Imbrie model. This makes the glacial cycles throughout the Pleistocene (also before the MPT) almost self-sustained, and their amplitude largely decoupled from insolation variations. We write almost self-sustained, since two of the state switch rules are ill defined (or never triggered) in absence of insolation variations. Still, in [Paillard, 1998] parameters are chosen such that a switches typically occur within zero to twenty kyr (with a narrow distribution), so dynamically the model resembles a self-sustained oscillator.

The strengths of the PP98 model is its conceptual simplicity, its ability to reproduce the MPT without strong assumptions, as well as its ability to simulate the sequence of late Pleistocene glacial terminations reasonable well. On the last point, we note that PP98 simulates the termination ∼400 kyr ago because ice volume grows even without insolation variations; in fact, when insolation variations are weak (as they are around −400 kyr ago) ice volume can grow undisturbed by insolation variations that would otherwise trigger a state shift.

The model reproduces the MPT by increasing the threshold of glacial termination and the forcing influence gradually, combined with frequency locking (in some sense at least) to the forcing. This mechanism has been used in other models of the MPT, and we define it in Chapter 7 where we name it *ramping with frequency locking.* However, the way the model reproduces the MPT has subtleties which to our knowledge have not been highlighted before. These are the topic of Chapter 10.

Later, Paillard and collaborator Parrenin followed up with similar models to PP98, one of which is mentioned in Paper A (Chapter 7) [Paillard and Parrenin, 2004, Parrenin and Paillard, 2003, 2012].

In year 2000 Gildor and Tziperman [2000] formulated a model of glacial cycles relying on a "sea-ice" switch (and the classic temperature-precipitation feedback) to cause glacial cycles. It works roughly as follows. Imagine a warm world without sea ice and with small terrestrial ice sheets. Much precipitation falls onto Northern hemisphere land masses, building up land ice, which through the albedo feedback cools the surrounding oceans, facilitating the spread of sea ice. However, as sea ice spreads, the ice sheets receive less precipitation, effectively suffocating them such that ablation (loss of ice mass) exceeds accumulation. An abrupt collapse of sea ice leads to warmer ambient oceans and an abrupt loss of sea ice cover, and so the cycle repeats.

Dynamically, the model is similar to the Paillard 98 model in that ice volume grows spontaneously until a threshold is reached, after which ice sheets collapse, and then slowly build up again. This is made even more obvious in [Ashkenazy, 2006], where a simplification of the model is presented.

The MPT is reproduced differently in different spin-offs of the [Gildor and Tziperman, 2000] model. In [Tziperman and Gildor, 2003] the sea ice switch is inactive for a warm deep ocean. The sea ice switch is then activated due to a slow ocean cooling. In Ashkenazy and Tziperman [2004], a threshold for glacial termination is ramped up like in P98. However, the ramp is not linear over the entire past 2 Myr, but piecewise linear and ramped only over 200 kyr. Thus, one can argue, that the MPT is prescribed rather than explained. This we discuss in Paper A in Chapter 7.

In 2003 Wunsch [2003] suggested an even simpler phenomenological model:

$$\begin{aligned} x_{n+1} &= \alpha x_n + \eta_n \text{ until } x > x_{max}, \text{ then} \\ x_n &\to 0, \end{aligned} \tag{6.5}$$

where x_n is ice volume evolving in discrete time n, x_{max} is a threshold of glacial termination, $0 < \alpha < 1$ is an autocorrelation parameter and η_n is a random process with elements independent of previous times. The process in the growth phase is called an AR(1) process if the random part is zero-mean, and the pattern of building up "potential" until resetting as a threshold is reached is called a integrate-and-fire model, see e.g. [Knight, 1972]. Wunsch used this model, which was in a way formulated graphically in [Broecker and van Donk, 1970, Macayeal, 1979], to propose that the frequency spectrum of the late Pleistocene glacial cycles can be reproduced without astronomical forcing. Huybers [2007] modified the model to read

$$\begin{aligned} x_{n+1} = x_n + \mu \text{ until } x > \theta(t), \text{ then} \\ \text{decrease } x \text{ to } 0 \text{ linearly over10 kyr}, \end{aligned} \tag{6.6}$$

where $\theta(t) = \theta_0 + \theta_1(t - t_0) - AF(t)$ is a threshold that increases linearly over time and also is modulated by forcing $AF(t)$ with amplitude A, and μ is a constant growth rate of ice volume. The forcing here has a pacing effect, meaning that it does not influence ice volume growth directly, but only when terminations occur. This pacing is crucial for how the model reproduces the MPT; over ranges of mean threshold values $\theta_0 + \theta_1(t - t_0)$ the system frequency locks to the forcing. As the threshold increases gradually, the system switches from one frequency locking region to another. This is another example of a model reproducing the MPT using ramping with frequency locking, the topic of Paper A (Chapter 7). However, if the forcing is modulated and if parameters vary over time, frequency locking as measured by asymptotic time-averaged quantities such as average duration, does not provide a complete description of system behaviour. Overcoming this limitation is the topic of Chapter 9.

In recent years, new dynamic perspectives have entered the modelling of glacial cycles. One of them is delay. Rial and Anaclerio [2000] used a delay logistic equation to explain the late Pleistocene glacial cycles. Huybers [2009] showed that a simple delay difference equation can exhibit long streaks of $\sim$40 kyr long cycles, followed by chaotic switches to much longer cycles. According to this hypothesis, the MPT could have occured by chance, and the "40 kyr world" would be highly unstable. Later, Huybers and Langmuir [2017] hypothesised that the depressing effect of ice sheets on Earth's crust can cause a delayed emission of CO_2 due to volcanic activity. This leads to oscillatory behaviour of ice volume and CO_2, and it is possible to go from quasi-linearly forced $\sim$40 kyr oscillations when forced with obliquity forcing, which by means of small fluctuations in the forcing can lead to longer, larger cycles. Thus also in this model the MPT occurs spontaneously. Lastly, [Quinn et al., 2018] reinterpreted an old Saltzman model as a delay differential equation, and showed that also for this model insolation variations can cause a shift from quasi-linearly forced solutions to self-sustained oscillations.

These delay models are examples of multistability, the coexistence of multiple solutions which attract different sets of initial conditions, and for which spontaneous switches can occur. Although known since long, Tziperman et al. [2006] were perhaps among the first to emphasise this, showing that self-sustained models of glacial cycles can synchronise to different oscillatory solutions. Crucifix [2013], de Saedeleer et al. [2013], Mitsui et al. [2013] further explored this topic, bringing to attention the concepts pullback attractors (a form of nonautonomous attractor), and nonchaotic strange attractors which synchronise solutions but yet are sensitive to perturbations. In 2016 Omta et al. [2015] proposed a simplified model in which the carbon cycle plays a special role. When forced by periodic forcing, it contains multiple frequency locked attractors at different orders. Small perturbations were shown to be able to cause switches between these frequency locked attractors, providing yet another mechanism for the MPT.

A new take on the Paillard [1998] hypothesis of discrete climate states were taken by [Ditlevsen, 2009], who modelled these states as different equilibria of a parametrised polynomial function. The MPT was then marked by the introduction of a deep glacial state, previously unavailable. Later, in Ashwin and Ditlevsen [2015] framed such a mechanism in the context of fast-slow systems, which served as inspiration for Paper B in Chapter 8.

Statistical approaches to modelling were taken by [Imbrie et al., 2011] and [Tzedakis et al., 2017]. Imbrie et al. [2011] fitted a weakly parametric model depending on insolation, ice volume x and rate of change of ice volume $\dot{x}$ to the LR04 stack fitted to a late part of the record. The authors claim to see a spontaneous MPT in the model, but upon inspection of the power spectrum over the last 3 Myr it rather seems like a temporary lull of $\sim$100 kyr power, not in agreement with data. Tzedakis et al. [2017] devised a statistical rule to determine from insolation and the time from the last interglacial whether another interglacial would be triggered or not. To model the MPT they, like [Mitsui et al., 2015], ramp the threshold for deglaciation which assumes rather than explains the MPT.

In later years, models of intermediate complexity have reproduced glacial cycles. These models are more detailed than the conceptual models reviewed so far, but less detailed than high-resolution models used e.g. for centennial climate predictions. Ganopolski and Calov [2011], Ganopolski and Brovkin [2017] reproduce the late Pleistocene glacial cycles rather well. However, the model suffers from two complications. One is that it has so many adjustable parameters that it can be made to fit proxy records better than the model perhaps deserves. This is the problem of overfitting, which can be difficult to circumvent given that many model parametrisations have to be fitted. The second problem is that the model relies on a rather tentative dust, volcanic outgassing and calving mechanisms for producing the glacial cycles. Thus, although the model includes more physical processes and is spatially resolved, it requires ad hoc assumptions just like low-dimensional conceptual models. That this is the case is evident from that Abe-Ouchi et al. [2013] also reproduce glacial cycles, but attribute them to isostatic rebound and ice sheet dynamics. Ideally, a highly detailed model with well understood and well modelled physics should be able to reproduce the entire climate record at high detail. This is very expensive, however, and it is unknown whether some day enough can be learned about past and present climate processes to avoid the need of ad hoc assumptions. Presently, high complecity models have only focussed on subquestions, such as whether a change in insolation can cause a glacial inception [Vettoretti and Peltier, 2011, Jochum et al., 2012].

On 3 April 2019, shortly before the submission of this thesis, a simulation from a model similar to that in [Ganopolski and Calov, 2011, Ganopolski and Brovkin, 2017] was published, reproducing climate over the past 3 million years, including the MPT[Willeit et al., 2019]. An in-depth analysis of these results is outside the scope of this thesis, so we restrict ourselves to a few comments. The model uses a combination of a trend in volcanic outgassing of CO_2 and erosion of regolith to reproduce the MPT (model hypotheses discussed in the next section). The study provokes many interesting, but unanswered questions: for instance, why do the reproduced glacial cycles before the MPT contain more obliquity than precession power, what is the dynamical mechanism causing the MPT when ramping either CO_2 or regolith erosion, and what causes the specified trends in climate variables? We touch on the second question in the discussion of Chapter 7.

We emphasise that [Willeit et al., 2019] does not settle the question of the mechanisms behind either the glacial cycles or the MPT, since the model shares the best and worst of conceptual and complex models. It requires plenty of poorly constrained assumptions, common with conceptual models, but suffer from overfitting and some dynamic obscurity common with complex models. The presence of precession power in the early record and problems reproducing a deglaciation around -400 kyr when precession is weak are two major shortcomings of the model. The latter shortcoming, shared by [Imbrie and Imbrie, 1980], might come from that the model has no self-sustained oscillations. We forego Section 6.3, and state that the ability of a model to reproduce the general pattern of glacial cycles does not prove that the model is correct. The causes and dynamical mechanisms behind glacial cycles and the MPT remain open questions.

We end this section by stating that we have left out a large amount of models from this review. This is purely because of constraints of space and to some extent limited awareness, not because the models are inadequate or inconsequential. For further reviews of models, see e.g. Crucifix [2012], Imbrie and Imbrie [1980], Dijkstra [2013].

6.2 Modelling the MPT

We now summarise different hypotheses of the MPT. According to one hypothesis, there was no external trigger for the MPT, except possibly internal noise or a particular state of insolation forcing. This hypothesis is espoused by [Huybers, 2009, Huybers and Langmuir, 2017, Quinn et al., 2018, Imbrie et al., 2011, Omta et al., 2015]. Another option is that the climate system entered a fundamentally new mode of operation as a result of a bifurcation. This is the view of e.g. [Maasch and Salzman, 1990, Ditlevsen, 2009, Ashwin and Ditlevsen, 2015, Tziperman and Gildor, 2003]. A third alternative is that the system remained in essentially the same mode of operation, but that its response changed due to a more-or-less gradual change of a control parameter. On the end of the spectrum invoking a slow change of parameter we find [Huybers, 2007, Paillard, 1998, Paillard and Parrenin, 2004] which use a linear change over 2 and 5 Myr respectively. On the other end of the spectrum we find [Ashkenazy and Tziperman, 2004]and [Feng and Bailer-Jones, 2015] with ramps over a couple hundred thousand year. In between are [Mitsui et al., 2015, Daruka and Ditlevsen, 2015, Tzedakis et al., 2017].

Models invoking no external change are naturally appealing explanations for the MPT. Models entering a new mode of operation due to a slowly changing bifurcation parameter both have the burden of explaining the changing bifurcation parameter, and to defend the claim of the system entering a new regime. In contrast, an abrupt change in model behaviour with unchanged physical mechanisms, requires less of an explanation. The problem of explaning the underlying bifurcation parameter remains, however. In this subclass of explanations, those invoking an abrupt change of external parameter have a stronger burden of explanation than those invoking a more gradual change of parameter.

We have excluded a few hypotheses for the MPT which did not fit in the timeline of model development. A well known one is the erosion of regolith hypothesis [Clark and Pollard, 1998]. Regolith is a layer of soft sediment hypothesised to have lain under major ice sheets prior to the MPT, causing basal sliding that prevented ice sheets from growing large. With time, sliding ice sheets would erode the regolith bed, thus exposing rigid ground on which ice sheets post-MPT could grow large. The hypothesis is appealing, but difficult to verify.

Another explanation for the MPT is that prior to ∼1 Myr ago ice volume variations on both hemispheres varied out of phase due to the anti-symmetric action of precession on the hemispheres [Raymo et al., 2006]. At some point the Antarctic ice sheet margin would extend into the sea, negating the variations in Antarctica and contributing to global cooling, which in turn could aid the formation of large Northern hemisphere ice sheets. This explanation relies on an unexplained global cooling, however, and the changes to Northern hemisphere glaciation are unexplained.

6.3 Constraints from data and model comparison

With so many models to choose from, can we say which one is correct? First of all, few models are incompatible, so an explanation for the glacial cycles and the MPT can rely on a combination of them. Furthermore, we will argue that “correct” might be the wrong word to use. Regardless, there have been attempts to evaluate how well a model explains observed data.

One strategy is to focus on a single feature in data, such as the timing of major terminations, temporal asymmetry, asymmetry about the mean, or the noise spectrum. Such a feature can be compared quantitatively or qualitatively with model output. The most basic and common way to compare model with data is to show model output on top of proxy data of whatever is compared (e.g. global ice volume), as we did in e.g. Figure 6.2. This obviously only allows for a qualitative comparison.

Another feature to evaluate is the timing of major glacial terminations or the times of interglacials. This was used in Huybers and Wunsch [2005], Huybers [2011] to test whether obliquity and precession influences the timing of glacial terminations, in [Feng and Bailer-Jones, 2015] to test which of a family of models and forcing fit the proxy records the best. Concurrence of glacial terminations with model terminations is central in [Parrenin and Paillard, 2012]. In Paper A in Chapter 7 we study how the durations between terminations evolve over time, and how the durations relate to the ramping with frequency locking hypothesis for the MPT.

Other features are the temporal asymmetry [Ashkenazy and Tziperman, 2004], asymmetry around the mean [Lisiecki and Raymo, 2007], modulation of the envelope of glacial cycles [Lisiecki and Raymo, 2007, A.Rial et al., 2013], and Fourier and wavelet spectra. A “perfect” model of the glacial cycles should replicate each of these features perfectly.

However, even if a model fails to replicate one or more “feature” in data, it does not mean that the model lacks utility or merit. For instance, Huybers and Langmuir [2017] note that although their glacial cycles are symmetric, the model probably can be modified to produce asymmetric cycles, as seen in data, while using the same physical and dynamical mechanism.

It is important to remember that the objective of simple climate models rarely is to reproduce the ice volume record perfectly, but rather to serve as easily understandable representatives of mechanisms, dynamical or physical. Therefore, it may not be useful to ask a model to replicate all conceivable features or an observed proxy record. This also raises the question of what we mean by a "correct" model. It could well be that a conceptual model captures the correct physical and dynamical mechanisms, and yet fails to replicate some "features" in proxy records due to abstraction of details.

Nevertheless, in some cases quantitative comparison between models is useful for excluding mechanisms. Roe and Allen [1999] compared residuals of a set of models with added noise fitted to an oxygen isotope record of the last 600 kyr. They found that they could not distinguish statistically between the models based on their test. Later, [Feng and Bailer-Jones, 2015] used Bayes factors to assess the relative ability of a family of models to reproduce a sequence of glacial terminations. Their test favoured a model forced by an equal amount of precession and obliquity, but only few model scenarios gave very low Bayes factors, indicative of poor performance. It is not clear how to interpret Bayes factors unless one of the models are "the truth", however; there could be a large Bayes factor between two obviously wrong models, and yet the Bayes factor between either model and the truth could be enormous. Does this mean that one model is more correct than the other? Carson et al. [2018] also used a Bayesian approach to distinguish between models, but using the one-step prediction of the model rather than termination times. They found conflicting evidence for different models, depending on whether they were forced astronomically, and whether a tuned proxy record was used.

We summarise by noting that there are features and statistical information in proxy records of glacial cycles which should be explained. This can be done through modelling and/or statistical analysis, but which features should be expected to be reproduced by which model is not clear.

6.4 Summary

The glacial cycles have mainly been modelled as conceptual models. This means that they contain few variables and often lack spatial considerations. In the 1980s when model development took off computational restriction was one reason for the use of conceptual models. But another reason was that these simple models were able to reproduce global variations in ice volume rather well, hinting at a dynamically simple underlying mechanism. As models using different physical and dynamical assumptions all were able to reproduce the last eight glacial cycles rather well, questions shifted more to which of them are "correct", or the most appropriate.

From here, research went along different routes. One route was to propose even more models using different assumptions, demonstrating that hitherto proposed models did not exhaust all possibilities. Parallel to, and joint with this route, modellers focussed one the ability of models to reproduce specific aspects of proxy records, such as glacial cycle asymmetry, carbon cycle modelling, and the ability to reproduce the MPT. Yet another route was try to exclude or choose between models based on statistical considerations. To this date, no conclusive answer has been reached. Recently, steps have been taken to reproduce the glacial cycles with more complex models. The idea is that models with more detailed physics require fewer ad hoc assumptions, and thus are more convincing. So far, only models of intermediate complexity have been able to reproduce full glacial cycles, while truly complex models are limited by computational constraints and thus can only investigate aspects of glacial cycles occurring on shorter time scales, such as glacial inception.

Chapter 7

Paper A: The MPT through frequency locking and a slow change in variable

In Paper A we aim to clearly describe a dynamical hypothesis for the middle Pleistocene transition (MPT) which has been used in many models, but not given a thorough treatment in the climatic literature. This mechanism, which we call *ramping with frequency locking* (RFL) relies on the slow ramping of a climate variable, which triggers switches between frequency locking regions, or approximations of such regions in case of complex multi-frequency forcing. Frequency locking and ramped bifurcation parameter have already been encountered in previous sections, and the paper (found in Appendix A) is mostly self-contained. Therefore, in this chapter we simply summarise the main results, discuss them, and then explain wavelet spectra.

7.1 Main results

The main contribution of Paper B is to clearly define and describe the mechanism (RFL) for causing the MPT in glacial cycle models. RFL is responsible for the MPT in several models (e.g.[Paillard, 1998, Paillard and Parrenin, 2004, Huybers, 2007, Ashkenazy and Tziperman, 2004]), but from discussions with researchers and from reviewer comments, it seems like this is not widely known. As a corollary, we point out that these models, although superficially different, rely on the same dynamic mechanism for causing the MPT. Thus we aid the classification of different mechanisms for the MPT, which is one way of getting a clearer view of the problem.

Secondly, we emphasise that the rate of ramping a bifurcation parameter matters. A slow, progressive ramping of a climate variable is not so remarkable physically, and through frequency locking it can explain the rapid increasing in durations between glacial terminations at the MPT. In contrast, a rapidly ramped bifurcation parameter does not explain the increase in duration dynamically, but rather presupposes it. Furthermore, we emphasise that the functional relations between (average) duration, internal period, ramping parameter, and time, all play a role in how abrupt change in durations is seen in a model. This helps understanding what influences the ability of a certain model to reproduce the MPT through RFL.

Thirdly, we emphasise the rarely mentioned fact that durations between glacial terminations increase over time, from ~ -1200 kyr until the present. It is often acknowledged that late Pleistocene glacial cycles vary in length between $\sim$80 and $\sim$120 kyr [Raymo, 1997], but not that there is a trend towards longer cycles. One exception is Huybers [2007], who claim that the mean period of a moving windowed Fourier spectrum increases over time.

Our claim is stronger however; we point out that individual durations between major glacial terminations increase over time, starting from -1200 kyr, and that the durations first cluster near to 80 kyr, and then have a propensity towards longer, upwards 120 kyr long durations.

Fourthly, we argue that such an increase in durations, shifting from $\sim$40 to $\sim$80 kyr is consistent with frequency locking mechanisms, as opposed to mechanisms for which long cycles emerge at the MPT with a period that is at or above 100 kyr.

Fifthly, we highlight a conundrum: while the durations between major glacial terminations and wavelet power is focussed around 80 kyr at the MPT, there is a strong anti-phase coherence with eccentricity, first at periods close to -120 kyr, and then decreasing to $\sim$100 kyr. This hints that there is a sub-dominant mode of variability from the early MPT and through the late Pleistocene which is distinct from the mode of variability that causes glacial terminations. We do not offer any solution to this conundrum, but leave it as an open question to solve. The MPT (~ -1200 to -800 kyr) should not just be considered an epoch of disorganisation, but harbours clues to what mechanism caused the drastic change in durations.

7.2 Discussion

Since the Discussion in Paper A is rather self-contained, we only discuss the model runs of [Willeit et al., 2019] in light of Paper A.

It is not clear what dynamical mechanisms underlie the glacial cycles and the MPT in the recently published CLIMBER-based Earth model of intermediate complexity in [Willeit et al., 2019]. However, the lack of self-sustained oscillations and difficulties reproducing a glacial termination around -400 kyr hint that works in a fashion similar to the model [Imbrie and Imbrie, 1980]. Self-sustained oscillators (or more specifically, models with glaciation in absence of forcing) typically do not have that problem. The MPT can be independently reproduced in[Willeit et al., 2019] by either ramping regolith erosion or atmospheric CO_2. For the case of ramped regolith, the area of exposed bedrock does not plateau when $\sim$100 kyr durations appear. Nonetheless, in the terminology of Paper A, the abruptness of the MPT according to the regolith hypothesis is probably best described as due to an abrupt change in “internal period” $T_o(p)$ resulting from of a slowly changing parameter p. However, $T_o(p)$ in this model should rather be thought of as “preferred” period, as the model seems to be either resonant to forcing or excitable [Marchionne et al., 2018]. The MPT under increased volcanic outgassing (causing a trend in CO_2 is somewhat more puzzling however. An interesting venue for future work could be to find a dynamical explanation and a representative conceptual model for this mechanism.

It would be interesting to learn how well the model in Willeit et al. [2019] agrees with the observed pattern of increasing durations between glacial terminations and wavelet coherence with eccentricity. From the wavelet spectrum in Figure 4 of Willeit et al. [2019], dominant long period amplitude seems concentrated below 100 kyr until approximately -250 kyr, consistent with data and resonance with a period shorter than the dominant periods of eccentricity variations. A more detailed analysis is required, however.

7.3 Errata

Since publication we have found one minor error and one unclarity in the published paper in Appendix A.

On page eleven, the last sentence in the first paragraph of column two should read “Whether solutions are truly frequency locked or depend [strongly] on initial conditions is irrelevant, as long as durations undergo abrupt change and tend to cluster.” The sentence, written as in the publication, it is tautological. With the added word “strongly” we wish to cover both sensitive dependence on initial conditions typical of chaos, as well as the dependence on initial conditions associated with initialising solution near boundaries of basins of attraction.

In Figure 1b on page one, the blue dashed lines are at multiples of 41 kyr, the dominant period of obliquity (informations missing from caption).

Chapter 8

Paper B: Classification of bifurcations in singular fast-slow systems

The second paper (Paper B) is more mathematical than the first and is the result of collaboration with Peter Ashwin and Peter Ditlevsen. In it we propose a classification of bifurcations in fast-slow systems in the singular limit, in which the fast variable is "infinitely" faster than the slow one, as well as bifurcations of singular relaxation oscillations that result from bifurcation in the fast-slow system. The paper (Appendix B) is self-contained and can be read in its entirety without particular preliminaries. However, we aim in this chapter to provide background, context, and clarification.

First, we describe how the paper was conceived and motivate why the results are important. Then, we summarise the main results and how they relate to previous research. At this point we strongly recommend the reader to read Paper B in Appendix B before proceeding to the rest of this chapter (save for the glossary at the end). Thereafter, we make some clarifications of points which were not given sufficient attention in the paper. We then discuss some additional results which were not included in the paper, whereafter we discuss directions for future work. In the final section, we provide a glossary of some terms which were not explicitly defined in the thesis.

8.1 Motivation and background

The initial motivation for the research in this paper came from a paper on a model of glacial cycles by Peter Ashwin and Peter Ditlevsen from 2015 [Ashwin and Ditlevsen, 2015]. The model describes the time evolution of one fast "state" variable y and one slow "global ice volume" variable v, whose fast subsystem depends on a bifurcation parameter λ that increases slowly over the Pleistocene. As λ changes, the critical set defined by the fast subsystem is deformed (Figure 8.1). Before a critical parameter value λ_c, relaxation oscillations live near two stable branches of an s-shaped critical set, much like the van der Pol oscillator. At a critical value λ_c, an unstable branch merges with one of the stable branches of the s-curve, such that for $\lambda > \lambda_c$ the two branches to split off into two different continuous curves. As a result, the amplitude of the relaxation oscillation increases rapidly.

The observed phenomenon was interesting for many reasons: firstly, it has practical relevance as a mechanism for the middle Pleistocene transition. Secondly, it occurred robustly under variation of a single parameter. Thirdly, there did not seem to be a systematic treatment of it in the literature. Surely, it described a bifurcation in some sense, but it was not clear which one.

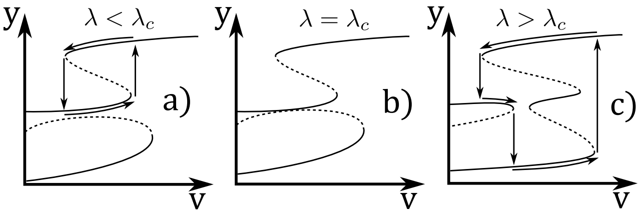


Figure 8.1: Sketch of the critical set and relaxation oscillations in the Ashwin and Ditlevsen 2015 model [Ashwin and Ditlevsen, 2015]. Solid/dashed black curves show stable/unstable parts of the critical set (equilibria of the singular fast subsystem). Arrows show relaxation oscillations composed of slow segments on the critical set and jumps from fold points. As a parameter λ changes from left to right, the critical set deforms such that at the critical value λ_c the relaxation oscillation changes from small amplitude to large amplitude.

Peter Ashwin recalled a section of Arnold et al. [1999], in which canards on degenerate manifolds similar to that of Figure 8.1 were briefly sketched and described. There was no detailed treatment, however, neither in the book nor in the list of references. After some literature search, it turned out that a description of the phenomenon had been tackled from two directions.

By 1985 [Golubitsky and Schaeffer, 1985] had extended work of John Mather and René Thom on singularity theory and the related catastrophe theory [Mather, 1969b,a, Thom, 1972], to zero sets of smooth families of vector fields depending on one parameter. They framed the theory in terms of bifurcation diagrams, since the parameter is naturally interpreted as a bifurcation parameter. However, they did not, as we do in Paper B (Appendix B), seem to realise that their bifurcation parameter could also be interpreted as the slow variable in a fast-slow system with one fast and one slow variable.

John Guckenheimer took direct interest in bifurcation of relaxation oscillations in singular systems. In a book chapter from 1996 [Guckenheimer, 1996] he presented ideas for extending the study of singularly perturbed (almost singular) fast-slow systems to a complete theory of singular and singularly perturbed relaxation oscillations. Again, in 2002 and 2003 [Guckenheimer, 2002, Guckenheimer et al., 2003] he refined and extended the outline for such a theory, but it he took most interest in bifurcation in systems of one or two fast, and two slow variables. That he insisted the critical set to be a manifold, like Takens before him [Takens, 1976], caused him not to recognise e.g. the bifurcation in the Ashwin and Ditlevsen model. Furthermore, considering the slow variables as non-distinguished unfolding parameters, rather than distinguished parameters, he assigned codimension for some bifurcations that is much higher than what we consider in our paper, and thus appears "rarer" in some sense. For instance, the cusp for one fast and two slow variables is codimension two to Guckenheimer, but codimension zero (persistent) in our setting. In this paper, we are taking steps towards the theory of singular relaxation oscillations (and other attractors) that Guckenheimer envisioned.

The importance of a systematic study of bifurcation in fast-slow systems should be evident from the number of fields in which they occur, such as neuroscience, climate, and reactional kinetics (see e.g. [Kuehn, 2015, Ashwin and Ditlevsen, 2015, Rinaldi and Scheffer, 2000, Harvey et al., 2011, Krischer et al., 1992, M. et al., 2008]). We are not aware of publications in applications, besides [Ashwin and Ditlevsen, 2015], reporting bifurcation due to the critical set, one of our main contributions. Reasons for this can be that the systems were not sufficiently scale separated, that the researchers simply did not have a name for it, or the bifurcations are rare for a reason not yet well understood. Nonetheless, we give examples of low order polynomial systems with such bifurcations in Paper B (Appendix B), indicating that they are not incredibly exotic objects. In any case, being able to enumerate all possible bifurcations up to a given codimension is a rather strong result. Given that a

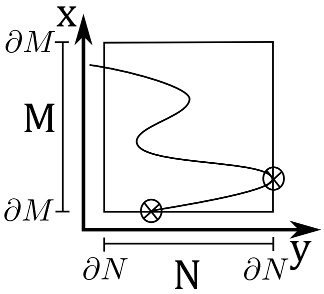


Figure 8.2: Example domain $M \times N$ considered for global equivalence (5.2) of bifurcation diagrams $g(x, y)$ (black solid curve). Boundaries are denoted ∂, and crossed circles show forbidden types of intersection with the zero set of the $g(x, y)$ and the boundary.

model in an application satisfies the assumptions of our classification, the classification tells us all the ways that model can bifurcate at codimension one.

8.2 Statement of main results

Our main contribution in this paper is to extend the *bifurcation theory* results of Golubitsky and Schaeffer [1985] to the context of fast-slow systems in the singular limit for one fast and one slow variable. Under a reasonable hypothesis and equivalence we state which fast-slow systems are persistent, and in what ways they can bifurcate at codimension one. Some bifurcations of fast-slow systems also give rise to bifurcation of singular relaxation oscillations. We list these. Our classification is a step towards completing the mission of Guckenheimer [Guckenheimer, 2002], to classify all bifurcations of singularly perturbed relaxation oscillations, although there is much more to be done. A particular contribution is the discovery of bifurcations of the critical set for which it loses its manifold structure; these bifurcations are absent from e.g. [Takens, 1976, Guckenheimer, 2002, Thom, 1972].

8.3 Clarifications

In this section we elaborate on a few topic which did not given much attention in the paper.

8.3.1 The choice of domains for the equivalence

In Section 2.2 of Paper B (Appendix B), we introduce equivalence on a compact domain in $M \times N \subset \mathbb{R} \times \mathbb{R}$, following [Golubitsky and Schaeffer, 1985, p. 143]. We do this to extend results on local persistence and bifurcation of germs to “globally” defined vector fields. We impose restrictions on g and $M \times N \subset \mathbb{R} \times \mathbb{R}$, however. By requiring that the vector field is inward flowing at the top and bottom boundaries ∂M, and that there are no folds on the right boundary, we avoid certain degenerate (atypical) intersections of the critical set with the boundary (recall that the critical set is the zero set of $g(x, y)$) (Figure 8.2). Likewise, we assume in the fast-slow context that the slow flow defined by $\dot{y} = h(x, y)$ is strictly inward on the right and left boundaries. By excluding some degeneracies by assumption, only the more meaningful ones remain to be classified.

These constraints also have a physical interpretation; in applications it is natural to define closed system boundaries such that the system state is contained within the system for all forward time.

8.3.2 Theorems of Golubitsky and Schaeffer

In the paper, we refer to results from [Golubitsky and Schaeffer, 1985] on persistence and codimension one bifurcation of critical sets (in their context called bifurcation diagrams). Here, we first restate their Theorem 10.1 on global persistence of zero sets, and then a corollary of their Theorem 2.1 which lists local bifurcation up to codimension three.

Theorem 1 (Thm 10.1, Chapter III in [Golubitsky and Schaeffer, 1985]) *Let* $G : M \times N \times W \to \mathbb{R}$ *be a family of bifurcation problems satisfying*

$$G(x, y, \lambda) \neq 0 \text{ for } \forall x \in \partial M, \ \forall y \in N, \ \forall \lambda \in W.$$

Let λ *and* β *be in the same connected component of* $W \setminus \Sigma$. *Then* $G(\cdot, \cdot, \lambda)$ *and* $G(\cdot, \cdot, \beta)$ *are globally equivalent (according to* (5.2) *on* $M \times N$.

Here, $N \times M$ is a closed rectangle as in Section 8.3.1, W is a closed disk in the space of unfolding parameters, $\lambda \in W$ and $\beta \in W$ are unfolding parameters, and Σ is the bifurcation set, the union of the following unfolding parameter sets

$$\begin{aligned} \mathcal{L}_1 &= \{\lambda \in W : \exists (x, y) \in M \times N \text{ such that } G = G_x = G_y = 0\}, \\ \mathcal{L}_2 &= \{\lambda \in W : \exists (x, y) \in M \times N \text{ such that } G = G_x = G_{xx} = 0\}, \\ \mathcal{G}_1 &= \{\lambda \in W : \exists (x_1, y), (x_2, y) \in M \times N \text{ such that } G = G_x = 0 \text{ at } (x_1, y), (x_2, y)\}. \end{aligned} \tag{8.1}$$

Subscripts of G denote partial derivatives, $\mathcal{L}_1$ is the set leading to degenerate fold tangency (isola and simple bifurcation in [Golubitsky and Schaeffer, 1985]), $\mathcal{L}_2$ is the set leading to hysteresis bifurcation, and $\mathcal{G}_1$ is the set leading to multiple limit point (double limit point in [Golubitsky and Schaeffer, 1985]). Thus $\Sigma = \mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{G}_1$. Illustrations of systems at bifurcation can be found in Figure 2 and Figure 3 in Paper B, Appendix B. It follows from the above theorem, that perturbations of any $F(x, y, \lambda)$ with λ away from the bifurcation boundary are persistent.

The following corollary lists all codimension one local bifurcations of vector fields $g(x, y)$:

Corollary 1 (From Thm 2.1, Chapter IV in [Golubitsky and Schaeffer, 1985]) *Let* $g(x, y)$ *be a germ in* $\mathcal{E}_{x,y}$ *satisfying* $g = g_x = 0$ *at* $(0, 0)$. *If the codimension of* g *is* 1, *then* g *is equivalent to one of the following bifurcation problems:* $\delta_1(x^2 - y^2)$ *(hyperbolic fold tangency),* $\delta_1(x^2 + y^2)$ *(elliptic fold tangency) or* $\delta_1 x^3 + \delta_2 y$ *(hysteresis). If the codimension of* g *is* 0, *then* g *is equivalent to* $\delta_1 x^2 + \delta_2 y$ *(nondegenerate quadratic fold).*

Here, $\mathcal{E}_{x,y}$ is the vector space of germs (locally defined smooth functions) in two variables and $\delta_1, \delta_2 \in \{-1, 1\}$. Note that non-degenerate quadratic fold is called limit point in [Golubitsky and Schaeffer, 1985]. Note also, that since the corollary only concerns local bifurcation, double limit points are not covered. While [Golubitsky and Schaeffer, 1985] do not give double limit points a rigorous treatment, it is not difficult to convince oneself that any non-local bifurcations at codimension one can only involve local codimension zero objects, that is: non-degenerate quadratic folds. It is clear from the equivalence (5.2) that singular points sharing y-coordinate are not persistent to perturbation, thus for codimension one bifurcation to occur exactly two non-degenerate quadratic folds must share y-coordinate.

We have tried to adapt the notation in this section to the one used in Paper B (Appendix B) as far as possible, but note that W, Σ, $\mathcal{G}$ and $\mathcal{E}$ denote different things here and in the paper.

8.3.3 Ideas behind the theorems

A considerable mathematical machinery goes into proving the theorems in the previous section. Reproducing this machinery is beyond the scope of this thesis, but below we summarise the main ideas behind it.

We begin with Theorem 2.1 in [Golubitsky and Schaeffer, 1985]. Let $\mathcal{E}_{x,y}$ be a vector space of germs under addition and scalar multiplication, and consider $g \in \mathcal{E}_{x,y}$. The ultimate goal is to decompose $\mathcal{E}_{x,y}$ into the tangent space of g, $T(g)$, and its complementary space $T(g)^{\perp}$, whose direct sum is the full space $\mathcal{E}_{x,y} = T(g) \bigoplus T(g)^{\perp}$. The tangent space $T(g)$ contains all perturbations p which leave $g+p$ equivalent to g (as germs). The complementary space contains only perturbations which leave $g+p$ inequivalent to p. The complementary space can be written as a scalar linear combination of basis elements of $T(g)^{\perp}$ and the number of such basis elements is the codimension of the universal unfolding of the germ.

Instead of determining $T(g)$ directly, Golubitsky and Schaeffer [1985] determine the restricted tangent space $RT(g)$, which contains the set of perturbations p which leave $g+p$ equivalent to g under the more restrictive *strong equivalence*

$$\tilde{g} = S(x,y)g(X(x,y),y). \tag{8.2}$$

This equivalence does not leave the distinguished parameter y fixed, unlike the equivalence (5.2). The perturbations $p \in RT(g)$ have a special structure (they are ideals), which makes it possible to express $RT(g)$ in terms of other, easily expressed, ideals in $\mathcal{E}_{x,y}$. Golubitsky and Schaeffer [1985] then show how to obtain $T(g)$ from $RT(g)$, given that $RT(g)$ has finite codimension. Having obtained $T(g)$, they compute $T(g)^{\perp}$ for all universal unfoldings up to codimension three.

A key to establishing that their list of universal unfoldings is complete, is the result that every g of codimension k is defined by a certain set of k independent equality conditions involving partial derivatives of g. The more equality conditions there are, the higher the codimension. This result also has practical importance, since it makes it possible to tell which normal form a given g is equivalent to (if any), without having to find explicit deformations in the equivalence (5.2).

The proof of Theorem 10.1 in Golubitsky and Schaeffer [1985] uses the fact that all degeneracies of singularities (points for which $g = g_x = 0$) are defined by the equality conditions in (8.1) (and possible more). Thus, by avoiding these sets, persistence is guaranteed. The remainder of the proof shows that the non-singular points $g \neq 0$ of the critical set ($g = 0$) can be mapped smoothly to other functions $\tilde{g}$ in the same connected component in parameter space as g.

8.3.4 The quantity $W[g]$

The quantity $W[g]$ is mentioned in Section 3.2 of Paper B (Appendix B), but we did not motivate it there. The reasoning behind it is similar to our expression for scalar quadratic curvature $Z[g]$, where g is the fast vector field. Then, we defined the scalar quadratic curvature in a coordinate system with the gradient along the x axis, and assumed that certain terms in the Taylor series expansion of g at a point do not affect the curvature. Thereafter we rotated the coordinate system of a vector field with gradient in a general direction such that the gradient was in the positive y_1 direction, after which we read off the relevant coefficients.

In the case of $W[g]$, we do not base our heuristic on any “scalar cubic curvature”. Instead, we start from a feasible normal form for the cusp tangency bifurcation

$$g(x, y_1, y_2) = \delta_1 x^3 + \delta_2 x y_1^2 + x y_1 \lambda + y_2,$$

where δ_2 is an unfolding parameter, and the sign of $\delta_1\delta_2 \neq 0$ gives the type of bifurcation ($\delta_1\delta_2 < 0$ gives beaks and $\delta_1\delta_2 > 0$ gives lips). Numerical approximations of critical sets corresponding to perturbations of the hypothesised normal form suggest that only the coefficients of the xy_1^2 and xy_2^2 terms perpendicular to the gradient govern how the normal form unfolds with λ (as “beaks” or “lips”, see Section 3.3.2 of Paper B in Appendix B). Starting from this assumption, the derivation of $W[g]$ is as for $Z[g]$ in Paper B in Appendix B.

Adding coefficients a, b, c, d, e to the above normal form at bifurcation ($\lambda = 0$), with terms of relevant order only, we get that

$$g(x, y, \lambda) = \delta_1 x^3 + cxy_1^2 + dxy_2^2 + exy_1y_2 + ay_1 + by_2.$$

Rotating the coordinate system to have gradient in the positive y_1 direction by the transformation

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \operatorname{sign}(g_{xxx}) \frac{1}{a^2 + b^2} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \end{bmatrix}$$

gives that the coefficient of the xy_2^2 term is

$$\delta_2 = \frac{1}{a^2 + b^2} \begin{bmatrix} -b & a \end{bmatrix} \begin{bmatrix} c & e/2 \\ e/2 & d \end{bmatrix} \begin{bmatrix} -b \\ a \end{bmatrix} = \frac{1}{2|\nabla_y^\perp g|^2} \nabla_y^\perp g^T D_y^2(g_x) \nabla_y^\perp g,$$

where the gradient is denoted ∇, $\perp$ means perpendicular, superscript T means transpose, and D_y^2 is the slow Hessian with elements $[D_y^2(g)]_{i,j} = \partial^2 g_i / \partial y_i \partial y_j$ for $i, j \in \{1, 2\}$. The gradients and the Hessian are expressed in the original slow coordinates (y_1, y_2).

Since only the sign of $\delta_1\delta_2 = 6g_{xxx} \frac{1}{2|\nabla_y^\perp g|^2} \nabla_y^\perp g^T D_y^2(g_x) \nabla_y^\perp g$ matters, we can define the more easily written quantity

$$W[g] = \operatorname{sign}(g_{xxx}) \nabla_y^\perp g^T D_y^2(g_x) \nabla_y^\perp g,$$

whose sign indicates whether the cusp tangency is of beaks ($W[g] < 0$) or lips ($W[g] > 0$) type.

8.3.5 Motivation for the fold tangency nondegeneracy condition

For one fast and two slow variables we conjecture a non-degeneracy condition for the fold tangency bifurcation defined by $\nabla_y g = 0$, namely the number of positive eigenvalues of the modified Hessian of g

$$\operatorname{sign}(g_{xx}) D^2(g), \tag{8.3}$$

where $D^2(g)$ has elements $[D_y^2(g)]_{i,j} = \partial^2 g / \partial x_i \partial x_j$ for $i, j = 1, 2, 3$ and $x_1 = x$, $x_2 = y_1$ and $x_3 = y_2$. The reasoning behind this expression is as follows.

Consider the Taylor expansion of some function g at a fold tangency, perturbed by a parameter λ

$$g(x, y_1, y_2) = ax^2 + 2bxy_1 + 2cxy_2 + dy_1^2 + 2ey_1y_2 + fy^2 + \lambda, \tag{8.4}$$

where a, b, c, d, e, f are constants. The linear and terms are zero since $\nabla_y g = 0$ and $g_x = g = 0$ at bifurcation.

First, we note that (8.4) is a quadratic form and can be written

$$g(x, y_1, y_2) = \begin{bmatrix} x & y_1 & y_2 \end{bmatrix} \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{bmatrix} x \\ y_1 \\ y_2 \end{bmatrix} + \lambda.$$

Since the matrix is symmetric with real coefficients, it is diagonalisable and has real eigenvalues. Thus, there exists an eigenbasis in which (8.4) can be written

$$g(x, y_1, y_2) = \bar{a}\bar{x}^2 + \bar{d}\bar{y_1}^2 + \bar{f}\bar{y_2}^2 + \lambda,$$

where $\bar{a}$, $\bar{b}$ and $\bar{c}$ are constants and $\bar{x}, \bar{y_1}$ and $\bar{y_2}$ are axis coordinates in a new basis. Here, we note three things: Firstly, degeneracy occurs if at least one of the eigenvalues $\bar{a}$, $\bar{b}$ and $\bar{c}$ are zero, which is covered by the equality condition $\det D^2(g) = 0$; the question is which qualitatively different subcases exists away from degeneracy. Secondly, $\bar{a} \neq 0$ is necessary for the fold to be quadratic. Thirdly, multiplying g by a nonzero constant rescales and/or reverses time, since $\epsilon\dot{x} = g$ in a fast-slow system. Therefore, we can rescale time by $\bar{a}$ to get the equivalent

$$g(x, y_1, y_2) = \bar{x}^2 + \frac{\bar{d}}{\bar{a}}\bar{y_1}^2 + \frac{\bar{f}}{\bar{a}}\bar{y_2}^2 + \frac{\lambda}{\bar{a}}. \tag{8.5}$$

From this expression, we identify subcases distinguished by the number of positive eigenvalues ($\mu_1 = 1, \mu_2 = \bar{d}/\bar{a}, \mu_3 = \bar{f}/\bar{a}$). One eigenvalue μ_1 is positive by construction. If all three eigenvalues are positive, $\mu_2, \mu_3 > 0$, then g describes the equation of an ellipsoid (isola bifurcation). If only one eigenvalue is positive, $\mu_2, \mu_3 < 0$, then the critical set upon unfolding by λ is given by

$$\bar{x} = \pm\sqrt{|\mu_2|\bar{y_1}^2 + |\mu_3|\bar{y_2}^2 - \frac{\lambda}{\bar{a}}},$$

and thus solutions for $\lambda/\bar{a} > 0$ only exists for sufficiently *large* y_1, y_2, in contrast to the isola case for which solutions exist for sufficiently *small* y_1, y_2. This case gives us wormhole fold tangency. In case two eigenvalues are positive, we get the tube fold tangency. The equation for the fold set is given by a pair of hyperbolae, e.g.

$$0 = \sqrt{|\mu_2|\bar{y_1}^2 - |\mu_3|\bar{y_2}^2 - \frac{\lambda}{\bar{a}}},$$

unlike the previous cases for which the fold set was given by an ellipse.

We comment on the multiplication of g_{xx} in (8.3). As noted previously, reversing time (multiplying g, and this its Hessian, by a negative constant) should give equivalent systems. We can remove this degeneracy by rescaling time t as $s = at$, such that if a happens to be negative, then we also rescale time. In doing so, we divide every element of g by a to get

$$\tilde{g}(x, y_1, y_2) = x^2 + \frac{2b}{a}xy_1 + \frac{2c}{a}xy_2 + \frac{d}{a}y_1^2 + \frac{2e}{a}y_1y_2 + \frac{f}{a}y^2 + \frac{\lambda}{a}. \tag{8.6}$$

The Hessian $D^2(\tilde{g})$ has at least one positive eigenvalue. This is because the Hessian cannot be negative definite by Sylvester's criterion [Gilbert, 1991], since its top left diagonal element is positive. Consequently, there are only three options for the remaining eigenvalues: either none, one, or two of them positive.

Next, we motivate why the signs of the eigenvalues of (8.6) tell which subtype of fold tangency $\tilde{g}$ is. Completing the square of (8.6) and defining new coefficients $\alpha, \beta, \gamma, \tilde{\lambda}$ gives that

$$\tilde{g}(x, y_1, y_2) = \left(x + \frac{b}{a}y_1 + \frac{c}{a}y_2\right)^2 + \alpha y_1^2 + \beta y_2^2 + \gamma y_1 y_2 + \tilde{\lambda}.$$

Defining the variable $z = x + \frac{b}{a}y_1 + \frac{c}{a}y_2$ and rotating the slow variables to $(\tilde{y_1}, \tilde{y_2})$, we get that

$$\tilde{g}(z, \tilde{y_1}, \tilde{y_2}) = z^2 + \tilde{\alpha}\tilde{y_1}^2 + \tilde{\beta}\tilde{y_2}^2 + \tilde{\lambda},$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are new constant coefficients. Now we note, in complete analogy with (8.5) that the way $\tilde{g}$ unfolds in λ is determined by the signs of the coefficients $\tilde{\alpha}$ and $\tilde{\beta}$. If $\tilde{\alpha}, \tilde{\beta} > 0$ there is an isola etc. We do not show it, but $\tilde{\alpha}$ and $\tilde{\beta}$ indeed have the same signs as the eigenvalues in (8.5).

In the discussion above, we divided g by a. Multiplying by a yields the same result. Furthermore, $D^2(ag) = aD^2(g)$, and if $a\mu$ is an eigenvalue of $aD^2(g)$ then μ is an eigenvalue of $D^2(g)$. Therefore, and since $a = 2g_{xx}$ and we only care about signs of eigenvalues, we deduce the criterion we claimed from the start: if $\operatorname{sign}(g_{xx})D^2(g)$ has one, two, or three eigenvalues, the fold tangency is of type wormhole, tube, or isola respectively.

8.3.6 Appropriateness of the equivalence

Since the classification of bifurcations at a certain codimension depends on the choice of equivalence, we spend a few lines to justify the equivalence for one fast and one slow variable. We start with the equivalence for the fast subsystem (that of critical sets).

The distinguished parameter is what sets the global equivalence (5.2) apart from equivalences used in standard singularity theory [Mather, 1969a, Arnold et al., 1999], catastrophe theory [Thom, 1972], and constrained systems [Takens, 1976]. A heuristic motivation for a distinguished parameter comes from consideration of relaxation oscillations; the slow variable generically "sweeps" the critical set and therefore should not raise codimension. In the other settings listed above, the slow variables are considered perturbations, but this raises codimension unnecessarily. Thus, we believe that the equivalence for the fast subsystem is well motivated.

The global singular equivalence in Section 4 of Paper B necessarily covers the fast and slow subsystems simultaneously. Since slow trajectories one exists only on the critical set, it makes sense to consider the slow subsystem only on the critical set, as we do. This also reflects that the slow dynamics are essentially one-dimensional, although they live in two-dimensional space. As a consequence, degeneracies of the slow subsystem which are not on the critical set, e.g. folds, are ignored.

The global singular equivalence gives rise to some uninteresting degeneracies at codimension one, such as fold umbra-fold umbra limit point and double slow equilibrium. These can be interesting at higher codimension, however.

8.4 Implications for models of glacial cycles

Proposition 5 in Paper B has an immediate implication for the modelling of glacial cycles and the middle Pleistocene transition. It says that if the glacial cycles can be modelled as a two-dimensional fast-slow system with smooth evolution functions, considered in the singular limit, there is a limited number of ways that the system can bifurcate (change qualitatively) under one-parameter variation. These seven ways are aligned and opposed double limit point, hyperbolic fold tangency, hysteresis, Singular Hopf, SNIC, and singular homoclinic. Only three of these can cause bifurcation of limit cycle to limit cycle with an abrupt change in amplitude and/or period. These are shown in Figure 8.3.

This theorem addresses the question of what might have caused the MPT from a dynamical point of view. Regardless of the physical mechanism, the dynamical mechanism must be exactly one of those in Figure 8.3 (see also Figure 11 in B). Hopefully, the knowledge of these dynamical mechanisms can inspire modellers to find physical ones for the MPT.

The above conclusion is only true if the assumptions of the model are appropriate of course. For example, non-autonomous, higher dimensional, or non-singular models may exhibit other phenomena. However, even if a model does not strictly speaking fulfil the necessary conditions, the bifurcations in Figure 8.3 may still be relevant. For example, the MPT in the Ashwin and Ditlevsen 2015 model essentially occurs due to hyperbolic fold tangency of the critical set, even though the model is astronomically forced. Another example are non-singular systems; at (singular) bifurcation, the behaviour of these systems can be complicated due to canard phenomena. Before and after bifurcation, however, the systems are generically well described by the singular systems for small enough scale separation. Hence, bifurcation in the singular system is also reflected in the nonsingular system.

There is still no classification of codimension one bifurcations for one fast and two slow variables. However, if our conjectured list of codimension one bifurcations of the critical set is correct, for some suitable equivalence, then no bifurcation of the critical set can cause bifurcation of limit cycles at codimension one. This is because limit cycles are one-dimensional curves which do not generically intersect points in phase space where bifurcation of the critical set occurs. However, we expect bifurcation due to non-generic intersection

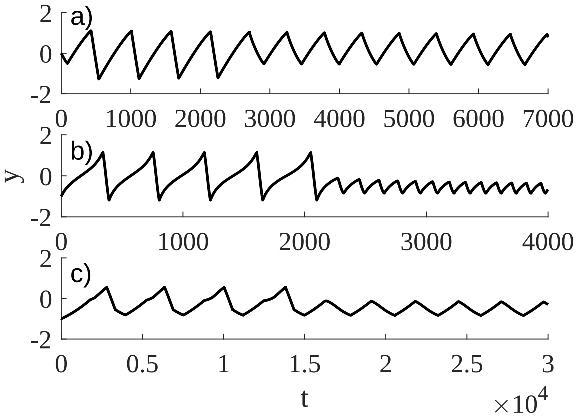


Figure 8.3: Examples of bifurcation in fast-slow systems which lead to solutions qualitatively similar to the middle Pleistocene transition. Bifurcation due to: a) aligned double limit point, b) fold tangency, c) opposed double limit point. Parameters are as in the respective cases in Figure 11 in Paper B, except that the integration times and scale separation parameters ϵ have been changed and the bifurcation parameter has been ramped linearly. The vertical axis (y) can be interpreted as global ice volume.

of limit cycles with the fold set and its umbra [Guckenheimer, 2002, Guckenheimer et al., 2003], and generic intersection of limit cycles with cusp points e.g. [Broer and Krupa, 2013].

Paper B is a step towards enumerating all possible dynamical mechanisms for the MPT, given a set of assumptions on the system. This approach is a systematic way to narrow down the cause of the MPT. This is both because the assumptions on the model are made explicit, and because models using different physical mechanisms may be dynamically equivalent (e.g. [Ashkenazy, 2006] and [Huybers, 2007]). Furthermore, models with the same physical motivation may be modelled dynamically differently (e.g. [Quinn et al., 2018] and [Maasch and Salzman, 1990], [Tziperman and Gildor, 2003] and [Ashkenazy, 2006]).

8.5 Future perspectives

Paper B presents several directions for future study. An obvious one is to generalise the fast-slow equivalence for one fast and one slow variables to more slow variables. A first step would be to extend the equivalence of critical sets for one fast variable and one distinguished parameter to multiple distinguished parameters. This apparently is challenging, however, since the shortcut through strong equivalence taken by Golubitsky and Schaeffer does not work [Montaldi, 1994]. Another route is to extend the theory to multiple (fast) variables. This is partly done in catastrophe theory [Thom, 1972], and is aided by the fact that potential systems do not contain limit cycles. The presence of limit cycles and other non-equilibrium asymptotic behaviour makes it difficult to use the theory of zero sets used by Golubitsky and Schaeffer, and probably requires a different approach.

A classification of bifurcations in one fast and two slow variables requires more than a classification of bifurcation of critical sets. It also needs to consider bifurcations purely in the slow subsystem (in the interior of stable subsets of the critical set) and degenerate interaction between the fold set of the critical set and attractors. [Guckenheimer, 1996, 2004] began sketching what bifurcations of limit cycles should be included in such a classification, but a rigorous treatment has so far not been made.

Another direction is to study how bifurcations in singular systems translate to bifurcation in almost singular (singularly perturbed) systems (for scale separation $\epsilon > 0$). Specific local cases have been studied widely with respect to canards (see e.g. [Kuehn, 2015, Jardón-Kojakhmetov and Broer, 2014] and references), but how these local bifurcations translate to global attractors has seen little attention so far. A notable exception is [Guckenheimer, 1996] and [Guckenheimer et al., 2003].

It would be interesting to learn if physically motivated glacial cycle models can be found that exhibit the bifurcations in Paper B. The bifurcations in singular fast-slow systems fits well with the assumption of multiple climate states used to motivate the models in e.g. [Paillard, 1998, Ditlevsen, 2009, Ashwin and Ditlevsen, 2015]. This assumption can be used as a starting point for physically motivated modelling.

We finally mention an interesting extension brought to our attention by Vivien Kirk, namely that of multiple time scales. For instance, one can consider bifurcation in systems with one fast, one slow, and one intermediate time scale.

8.6 Summary and Conclusions

In this chapter, we have presented and discussed Paper B in B, which concerns fast-slow systems: systems with components that vary on vastly different timescales. In this paper we prove persistence and classify bifurcations of the critical set, the set on which slow dynamics typically take place, up to codimension one. Our classification applies to smooth systems with one fast and one slow variable, but we further conjecture a classification up to codimension one for one fast and two slow variables.

Additionally, we define a global singular fast-slow equivalence which allows us to prove persistence of fast-slow systems and a classification of their bifurcations up to codimension one. This equivalence is used to derive similar results for singular relaxation oscillations, which in turn have relevance for the problem of glacial cycles and the MPT. In addition to the already published model in [Ashwin and Ditlevsen, 2015], we give examples of potential glacial cycle models having bifurcations which give solutions resembling the MPT.

The work presented here can be extended in a multitude of directions.

8.7 Glossary of concepts

8.7.1 Smooth and analytic functions

A *smooth function* $f(x)$ is a function which has well defined partial derivatives of all orders. An analytic function is a function $f(x)$ whose Taylor series converges to the function at any point. Every analytic function is smooth, but the converse is not true.

Consider for instance the classic example

$$\begin{cases} f(x) = e^{-1/x}, & \text{if } x > 0 \\ f(x) = 0, & \text{if } x \leq 0, \end{cases}$$

with n:th order derivative

$$\begin{cases} f^{(n)}(x) = e^{-1/x}\left(\sum_{i=1}^{n} \frac{i!}{x^{i+1}}(-1)^{i+1}\right) & \text{if } x > 0 \\ f^{(n)}(x) = 0, & \text{if } x \leq 0, \end{cases}$$

The function is smooth, since derivatives of all order exist, are smooth away from zero, and converges to 0 both from the left and the right. The function is not analytic, however, since it converges to the zero function $F(x) \equiv 0$ on a neighbourhood around $x = 0$ while $f(x) \neq F(x)$.

8.7.2 Compact set

A set U is *compact* if every cover of U has a finite subcover. Differently put: If U is compact if for every collection C of subsets of U such that

$$X = \bigcup_{x \in C} x,$$

then there is a finite subset F of C such that

$$X = \bigcup_{x \in F} x.$$

A similar definition holds for subsets.

By the Heine-Borel theorem, a subset of Euclidean space (as well as manifolds) is compact if and only if it is closed and bounded [Kirillov and Gvishiani, 1982].

8.7.3 Open set

The notion of open set depends on the chosen topology. We only consider subsets of Euclidean space, which inherit the Euclidean topology based on Euclidean distance.

We say that a set $D \in \mathbb{R}^n$ for some $n \in \mathbb{N}$ is *open*, if for every point y there exists an ϵ such that every point within an ϵ-distance of y is also in D.

8.7.4 Neighbourhood

Let X be a topological space. We say that $V \subset X$ is a neighbourhood of a point $p \in X$, if it contains an open set $U \subset V$ containing p.

Also, let X be a set of points. Assign to each $x \in X$ a non-empty collection of open subsets of $N(x)$ of x containing x. The elements of $N(x)$ constitute the neighbourhoods of x.

8.7.5 Topological space

A topological space is a space with a notion of whether elements are nearby to each other or not. Only if X is a metric space, there is a notion of *how* close two elements are. We use the definition due to Hausdorff.

A topological space X is a set of elements and a set of neighbourhoods $N(x)$ associated to each point in x, satisfying the following:

1. If $n \in N(x)$, then $x \in n$
2. If $m \in N(x)$ and $m \subset n$, then also $n \in N(x)$
3. If $m \in N(x)$ and $n \in N(x)$, then $m \cap n \in N(x)$
4. For every $n \in N(x)$, there exists an $m \subset n$ such that $n \in N(y)$ for every $y \in m$

Consequently, two points x and y are considered "close" if x is in a neighbourhood of y (and if y is in a neighbourhood of x).

8.7.6 Manifold

A manifold is loosely speaking an n-dimensional hypersurface that is locally homeomorphic to n-dimensional Euclidean space. As an example, the two-dimensional sphere in three dimensions is a manifold, since at every point on the sphere is can be locally approximated by a plane. A surface with a self-intersection is not a manifold at the intersection, since the set is not homeomorphic by Euclidean space there.

8.7.7 Homeomorphism and diffeomorphism

A homemorphism is a map $\phi : U \to V$ that is continuous and has a continuous inverse. A diffeomorphism a homeomorphism which is smooth with smooth inverse. In our setting we assume smoothness of infinite order.

Chapter 9

Paper 3: Temporal response of a self-sustained oscillator to modulated forcing

This chapter presents preliminary results exploring how forced oscillators respond to modulated forcing. The question that motivates this research is: “How do forced self-sustained oscillators respond to amplitude and frequency modulated forcing?”. The question was inspired by models of glacial cycles, but has potential implications for other applications. More precisely, starting from the assumption that we understand the response of forced oscillators to pure sinusoidal (also known as harmonic) forcing, what can we infer about the response to modulated sinusoidal forcing?

We build our analysis on the concept of frequency locking, previously described in Section 4.7. We take the Huybers 2007 model as an example, which was also used in Paper A (Chapter 7, [Huybers, 2007]). The new (in this thesis) concepts of instantaneous amplitude and frequency, as well as information entropy, are introduced and explained in this chapter.

First, we introduce and motivate the problem, then we describe the methods used. Thereafter, we present results from simulations of the Huybers 2007 model, after which we summarise and conclude.

9.1 Background and motivation

The long time, asymptotic, response of oscillating systems to periodic forcing has been well understood since the time of Poincaré. For weak to moderate strength of the periodic forcing, the forced system assumes a period that is a rational of the forcing period, or is quasiperiodic, meaning that it oscillates with an average period that is irrationally related to the forcing. The periodic response persists on open sets of parameters, called Arnold tongues, and is therefore called frequency locking, phase locking, or mode locking, a phenomenon with relevance for a variety of applications.

The response to aperiodic forcing, or periodic forcing that is not a simple sine, but still has some similarity to periodic forcing is not as well understood, however. An example of such forcing is the variations in insolation to different latitudes on Earth, which influences the timing of glacial cycles. This forcing can be closely described as the sum of two modulated sine waves with central periods approximately 41 kyr (obliquity) and 21 kyr (precession).

It is known that such multi-frequency quasiperiodic forcing can produce frequency locking to one or a combination of the constituent frequencies [Pikovsky et al., 2001, Feudel et al., 1997]. Furthermore, it can give rise to strange nonchaotic attractors, asymptotic states which are not periodic, but which have a peculiar recurrence structure [Grebogi et al., 1984, Mitsui et al., 2015]. However, none of these results relate to the short-time response to modulated forcing as they are concerned with the limit as time goes to infinity.

Previous studies of synchronisation under modulated forcing considered mainly weakly non-linear oscillators [Jensen, 2002, Davies and Nandlall, 1990, García-Álvarez et al., 2008, Lucas et al., 2018]. Davies and Nandlall [1990] considered a sinusoidally amplitude modulated weakly nonlinear van der Pol oscillator, and found unsurprisingly that the weak and strong forcing regimes yielded a nearly unperturbed solutions and forcing dominated solutions. Jensen [2002] found that frequency-modulated sine forcing can cause temporary frequency locking, if the rate of convergence of the nonlinear oscillator is fast enough relative to the change in forcing frequency. He derived analytical results for a weakly nonlinear oscillator but claimed that the results remain valid for moderate nonlinearity (taking the forced van der Pol oscillator as an example). Lucas et al. [2018] recently emphasised the need to not only focus on asymptotically defined quantities such as Lyapunov exponents and winding number in the case that the forcing has time varying parameters. Building on [Jensen, 2002], Lucas et al. [2018] showed that frequency modulation can cause the system to move in and out of synchronisation over a modulating cycle, but still remain synchronised on average.

We focus on a more complicated class of strongly nonlinear oscillators forced by one and two irregularly amplitude and frequency modulated sines. Additionally, we increase the internal period of the oscillator over time. This complicated scenario is motivated by models of glacial cycles, in particular models reproducing the abrupt change in amplitude and period called the middle Pleistocene transition (MPT).

To understand the behaviour of such systems, we go beyond stating that oscillators can in principle temporarily frequency lock, and instead assess how well they do so in practice. To accomplish this, we adopt a probabilistic viewpoint under lack of information. As a part of this approach, we consider time-varying distributions of durations between returns to a threshold (glacial terminations), which contain more information than just the average duration or winding number. We furthermore devise a score (a descriptive statistic) to assess how well the quasistatic (frozen) view, valid for infinitely slowly changing parameters, informs us about the non-quasistatic (non-frozen) dynamics. Additionally, we compute the information entropy of the distribution of durations to assess how specific knowledge we have about the likelihood of observing a certain duration at a given time.

The outline of this chapter is as follows. First, we introduce the concepts of frequency locking, instantaneous amplitude and period, information entropy and score. Thereafter, we present simulation results for the Huybers 2007 model of glacial cycles, forced obliquity and precession (astronomical parameters) in different proportions. Finally, we conclude our findings.

9.2 Methods

We study the phenomenon by making use of three fundamental concepts: frequency locking, instantaneous frequency and instantaneous amplitude from the Hilbert transform, and information entropy. We define and describe them in order. Thereafter, we define the model, present modulated forcing curves, and describe the study design.

9.2.1 Frequency locking

As was shown in Section 4.7, a self sustained oscillator — an oscillator that oscillates in absence of external stimulus — can become frequency locked to periodic forcing. By this we mean that the average duration $\overline{D}$, defined as the average time until return to a suitably defined Poincaré section, equals a fraction of the forcing period T_f:

$$\overline{D} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} D_i = \frac{N}{M} T_f,$$

where D_i is the i:th return to the Poincaré section, and $N, M \in \mathbb{N}$ are integers. This relation only holds in the asymptotic limit that time goes to infinity, unless the system is initialised precisely on the attractor of the system (a fixed point of an iterate of the Poincaré map). However, for trajectories near the attractor the finite-time average duration converges to the asymptotic average duration. Note, that we also require (by definition) that the relation holds on an open set of parameters in order to call it frequency locking.

Frequency locking and the average duration is one way to characterise a solution to a differential or difference equation; it tells us on average how many revolutions the oscillator undergoes per revolution of the forcing oscillator. Importantly, frequency locking tends to imply that durations cluster into groups.

9.2.2 Instantaneous amplitude and period

A modulated sine wave is a function of the form

$$F(t) = A(t)\cos\left(\int_{t_0}^{t} \frac{2\pi}{T_f(s)} ds + \phi_0\right), \tag{9.1}$$

where $A(t)$ is a time-varying amplitude, $T_f(t)$ is a time-varying period and ϕ_0 is a constant initial phase. Note that for any given signal $F(t)$ there is an infinite number of choices of $A(t)$ and $T_f(s)$ satisfying (9.1). One choice of $A(t)$ and $T_f(s)$ with nice properties is defined through the Hilbert transform [Mélice et al., 2001, Vakman, 1996].

The Hilbert transform of $F(t)$ is

$$H[F(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{F(s)}{t-s} ds, \tag{9.2}$$

where the integral denotes the Cauchy principal value [Ktonas and Papp, 1980, Mélice et al., 2001, Boashash, 1992]. Equation (9.2) is less cryptic in frequency space, since the Hilbert transform of the Fourier transform $\hat{F}(\omega)$, of $F(t)$ is

$$H[\hat{F}(\omega)] = i\hat{F}(\omega)\,\mathrm{sign}\,(\omega),$$

where i denotes the imaginary unit and $\mathrm{sign}\,(\omega)$ is the sign of the angular frequency $\omega = 2\pi/T_f$. Thus, the Hilbert transform shifts the phase of the Fourier transform by 90 degrees (through multiplication of i) and further reverses the sign of the negative part of the Fourier spectrum (which is symmetric for real signals).

The usefulness of the Hilbert transform will be evident from a sequence of steps (see also [Ktonas and Papp, 1980]). First, note that we can express $\hat{F}(\omega)$ as

$$\hat{F}(\omega) = \frac{1}{2}(\hat{F}_+^*(-\omega) + \hat{F}_+(\omega)),$$

where superscript $*$ denotes complex conjugate, and $\hat{F}_+(\omega)$ is twice the positive part of $\hat{F}(\omega)$

$$\hat{F}_+(\omega) = 2\hat{F}(\omega)\theta(\omega),$$

where $\theta(\omega)$ is the Heaviside step function $\theta(\omega) = 0$ for $\omega < 0$ and $\theta(\omega) = 1$ for $\omega > 0$. Consequently,

$$\hat{F}(\omega) = \frac{1}{2}(\hat{F}_+^*(-\omega) + \hat{F}_+(\omega)),$$

and

$$F(t) = \frac{1}{2}(F_+^*(t) + F_+(t)) = \Re(F_+(t)) = \Re(A(t)e^{i\phi(t)}) = A(t)\cos(\phi(t)), \tag{9.3}$$

where $\Re$ denotes the real part of a complex number, and $A(t)$ and $\phi(t)$ are the amplitude and phase of the polar representation of $F_+(t)$. Now, the Hilbert transform demonstrates its utility, since $\hat{F}_+(\omega)$ can be expressed as

$$\hat{F}_+(\omega) = \hat{F}(\omega) + \hat{F}(\omega)\,\mathrm{sign}\,(\omega) = \hat{F}(\omega) - iH[\hat{F}(\omega)],$$

and thus

$$F_+(t) = F(t) - iH[F(t)]. \tag{9.4}$$

Combining (9.4) with (9.3), gives the instantaneous amplitude

$$A(t) = \sqrt{(F(t) + \Re(-iH[F(t)]))^2 + \Im(-iH[F(t)])^2},$$

and the instantaneous phase

$$\phi(t) = \arctan\left(\frac{\Im(-iH[F(t)])}{F(t) + \Re(-iH[F(t)])}\right) + N\pi,$$

where N is an appropriate choice of integer and $\Im$ is the imaginary part of a complex number. From the instantaneous phase, we can define the instantaneous period $T_f(t)$ through

$$\phi(t) = \int_{t_0}^{t} \frac{2\pi}{T_f(s)} ds + \phi_0,$$

where t_0 is an initial time and ϕ_0 the phase at $t = t_0$. Inserting these into (9.3), we arrive at (9.1).

Although this choice of instantaneous amplitude and period is defined for all real signals, it only is guaranteed to be “meaningful” or “physical” [Ktonas and Papp, 1980, Cohen et al., 1999] if the signal is narrow-band. That a signal is narrow-band means that its Fourier transform is focussed on a single frequency away from 0. There is no sharp rule for when a signal is well decomposed through the Hilbert transform; the niceness of the instantaneous amplitude and period has to be assessed visually case by case. Typically, one wants the instantaneous amplitude to form a smooth envelope which intersects the maxima of the signal. The instantaneous period then follows from (9.3).

9.2.3 Information entropy

The Shannon information entropy of a discrete probability distribution with n outcome probabilities $P = \{p_i\}_{i=1}^n$ is given by

$$S(P) = \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right),$$

where $\sum_{i=1}^n p_i = 1$, $\log_2$ is the two-logarithm [Shannon, 1948], and we make the definition that $0 \cdot \log_2(1/0) := 0$. This quantity, like the thermodynamic definition of entropy, measures “disorderedness” of the distribution, which can be appreciated by noting that S is maximal for a uniform distribution $p_i = 1/n, \ \forall\ i \in \{1, 2, \ldots, n\}$, and zero for an atomic distribution $p_i = 1$ for some i, and $p_j = 0, \ \forall\ j \neq i$. Information entropy can also be thought of as *average surprise*, since if we define the surprise of observing outcome i as $I(p_i) = \log_2(1/p_i)$, then the average surprise in the distribution is $S = \sum_{i=1}^n p_i I(p_i)$ [Freiberger, 2015]. High surprise indicates that the distribution is uninformative, like the uniform distribution.

We want a measure of informativeness of a distribution of durations that is normalised between 0 and 1. To this end, we define the *normalised information entropy* as

$$S_n(P) = \frac{1}{\log_2(n)} \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right). \tag{9.5}$$

We note that the precise choice of measure of informativeness is rather arbitrary; the Shannon entropy has a few nice properties, but almost any function which gives higher scores for more concentrated probability distributions serve the purpose [Kumar et al., 1986]. The choice of measure is not important for us, since we use it only qualitatively.

9.2.4 Model and forcings

We use the Huybers 2007 model as an example of a forced self-sustained oscillator. The state variable (ice volume) x increases linearly in time at a rate μ until a threshold $\theta(t)$ (different from the Heaviside step function) is hit, whereafter x decreases linearly to 0 over a fixed time $T_{decay} = 10$. Once x reaches 0, it starts to increase once more:

$$\begin{cases} \dot{x}(t) = \mu, \text{ until } x(t) = \theta(t), \text{ then,} \\ \text{linearly decrease } x \to 0 \text{ over time } T_{decay}, \text{ repeat.} \end{cases} \tag{9.6}$$

The threshold function

$$\theta(t) = R(t) + F(t) \tag{9.7}$$

consists of a secular, aperiodic part $R(t)$ and a part consisting of periodic components $F(t)$, the forcing. If $F(t)$ is not identically zero, we say that (9.6) is forced. We always use $\mu = 1$ and $T_{decay} = 10$. For all our experiments we choose a linear ramp $R(t) = 26 + 0.05(t - t_0)$, where $t_0 = -2000$ is the initial simulation time.

We force the model with three forcings. The first two are obliquity (Figure 9.1 a,b,c)) and precession (Figure 9.1 d,e,f)), which are amplitude and frequency modulated sines. The third forcing consists of a sum of obliquity with standard deviation 5 and precession with standard deviation 10 (Figure 9.1 g)). The envelope and instantaneous phase was estimated from the Hilbert transform for obliquity, which was first linearly interpolated to a time step time step 0.1 kyr from a time step 1 kyr. The instantaneous period was estimated from differences of the instantaneous phase. The envelope for precession was taken to be eccentricity, and the instantaneous phase to be the longitude from perihelion [Berger, 1978, Laskar et al., 2004]. We chose them as such, since the Hilbert transform estimates are qualitatively similar, but have some numerical artefacts.

9.2.5 The frozen distribution of durations

Our method for understanding the response of the forcing relies on fixing a set of model parameters at a time τ, either just the secular ramp $R(\tau)$ in (9.7) or both $R(\tau)$ and the instantaneous amplitude $A(\tau)$ and period $T_f(\tau)$ of the forcing. We then run each of these *frozen* systems forward in time t for a fixed τ, keeping track of all the durations $d_{i,\tau}$ between glacial terminations, where i indexes the n durations. We will investigate how much information the τ-dependent empirical distributions $\tilde{\rho}_\tau(D)$ of durations $D_\tau = \{d_{\tau,i}\}_{i=1}^{n}$ contain about the *non-frozen* system. As $n \to \infty$, $\tilde{\rho}_\tau(D)$ approaches a theoretical, asymptotic distribution $\rho_\tau(D)$.

The ice volume evolution part of the frozen system (Eq. (9.6)) is identical to the non-frozen system except that the threshold $\theta(t;\tau)$ is parameterised by τ. The threshold for the *ramp-only frozen* system is

$$\theta_{ro}(t;\tau) = R(\tau) + A(t)\cos\left(\int_{t_0}^{t} \frac{2\pi}{T_f(s)} ds + \phi_0\right), \tag{9.8}$$

while the threshold for the *fully frozen* system (without subscript) is

$$\theta(t;\tau) = R(\tau) + A(\tau)\cos\left(\frac{2\pi}{T_f(\tau)}(t - t_0) + \phi_0\right). \tag{9.9}$$

The experimental procedure is as follows. Fix a time range of interest $I = [t_0, t_1]$ for the non-frozen system. In our example, we choose -2000 kyr to 0 kyr since the Huybers 2007 model (9.6) models the MPT, an abrupt increase in average duration, over that time. Pick a uniform sample of frozen times in this time range, e.g. $T_{sample} = \{\tau_0, \tau_1, ..., \tau_m\} = \{-2000, -1990, ..., 0\}$. For each $\tau_j \in T$, run the frozen system with threshold $\theta(t;\tau_j)$ (or $\theta_{ro}(t;\tau_j)$) over another time range $I_{integrate} = [t_{start}, t_{end}]$ which may but does not have to

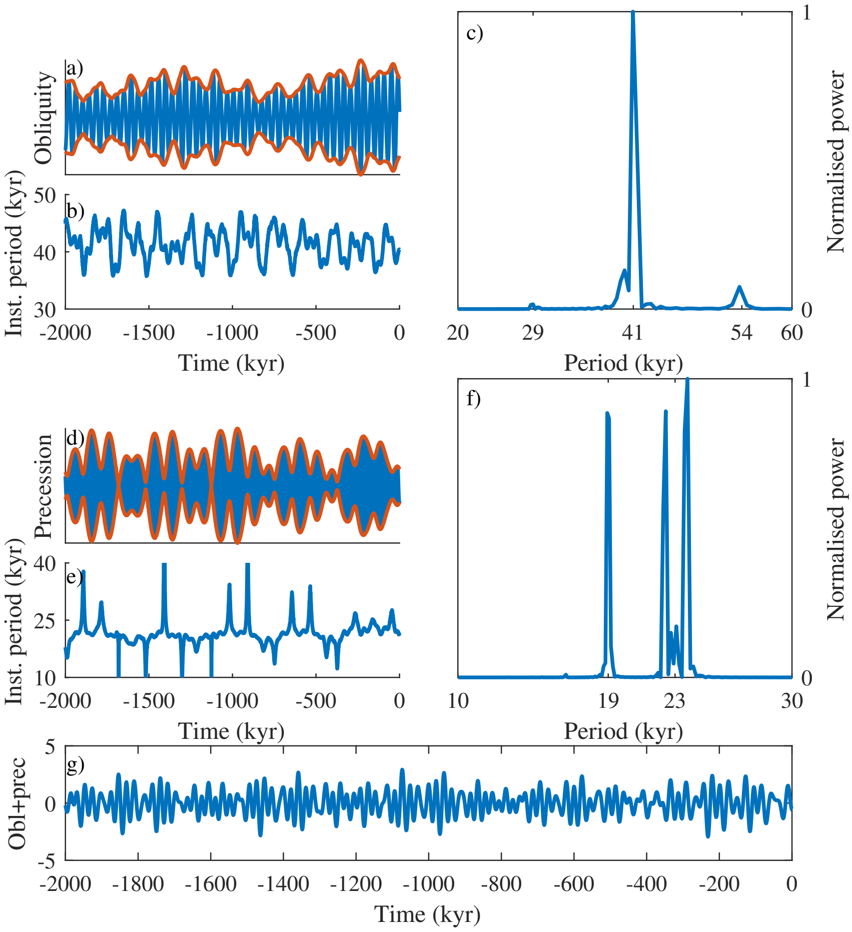


Figure 9.1: Forcing functions. Panels a,b,c) show obliquity, panels d,e,f) show precession, and panel g) shows an equal sum of obliquity and precession, normalised to zero mean and unit variance. Panels a) and d) show obliquity and precession curves over time, with instantaneous amplitude $A(t)$ envelopes in red. Panels b) and e) show instantaneous period $T_s(t)$ over time, and panels c) and f) show normalised Fourier power spectra. Panel g) shows a time series of a sum of 33% obliquity and 67% precession. Note that scales and plot ranges vary between panels.

equal the time range of interest I. Record a sample of durations $\tilde{\rho}_{\tau_j}(D) = \{D_{i,\tau_j}\}_{i=1}^{n}$. If the forcing is stationary, this sample of durations approaches some stationary distribution $\rho_{\tau_j}(D)$ as the time of integration t_{end} tends to infinity $t_{end} \to \infty$.

If we consider the ramp-only frozen system with threshold (9.8), then we neglect knowledge about the local amplitude and period of the forcing. In this case, we define $\tilde{\rho}_{\tau_j}(D)$ to consist of durations between glacial terminations of (9.6) under non-frozen forcing $F(t)$, and take $I_{integrate} = I$. We choose this integration interval to ensure that the forcing reflects the forcing of the non-frozen system on the interval I. We denote both the empirical distributions for the fully frozen and *ramp only* frozen systems by $\tilde{\rho}_{\tau_j}(D)$; which one we mean should be clear from the context.

9.2.6 Evaluating informativeness

We engage in a thought experiment where we do not know the past history or present state of the frozen system, but only assume that the system behaves as in the asymptotic limit, that is: it is near an attractor. Now we may ask how much knowledge we have about what the next observed duration, or equivalently, how surprised we would be on average when observing the next duration. This information is contained in the asymptotic distribution of durations $\rho_{\tau_j}(D)$. Note that the sequence of durations is perfectly deterministic, given perfect knowledge about the initial state. However, if we give up knowledge about past states and the dynamics of the system (apart from the asymptotic distribution), then $\rho_{\tau_j}(D)$ is properly a probability measure, and samples from it are random.

We can measure the surprise or disorderedness using relative entropy $S(P)$ (9.5), where P is the probability distribution associated with the "random process" D. We run into a couple of problems, however. First, the asymptotic distribution $\rho_{\tau_j}(D)$ could be continuous or fractal, and thus not discrete. We address this problem by binning the durations in bins of fixed width (we use a width of 3 kyr). This decreases the resolution of the distribution, but does not alter it essentially.

The second problem is the support of the distribution; if we define the support to be between the maximum and minimum observed duration, then a single-valued and a broad uniform distribution have the same relative entropy 1, which does not rhyme well with intuition. Thus, we clearly have to impose bounds on the supports of distributions of durations. To do this, we run the fully frozen system (9.6) with threshold (9.9) for a dense uniform sample of initial phases $\phi_0 \in [0, 2\pi]$ of the sine forcing function. This gives us all possible durations for the frozen system, given any initial condition. The boundaries of this distribution then defines the support of the asymptotic distribution $\rho_{\tau_j}(D)$.

In case of the ramp-only frozen system with threshold (9.8), the above method for defining bounds on the distribution fails, since then the forcing not a pure sine and only defined by its phase. In this case, we take the support of the distribution to be the maximum and minimum observed duration for each τ, for lack of better options. Ideally, one should record durations for solutions initialised on a fine grid of times $t \in I$ for each $R(\tau)$, but we consider this too time consuming. We expect the bounds of this distribution to be wide enough, but there is a risk that the bounds are underestimated.

Thus, we can evaluate the informativeness of the asymptotic distribution over frozen time by calculating $S(P)$ for each τ. If the informativeness is high, which means that entropy is low, then frequency locking effects increase the knowledge about which durations are likely to be observed. If informativeness is low, then little information is gained compared to random durations.

9.2.7 Score: Testing for agreement between frozen and non-frozen systems

It is false to assume that the asymptotic distribution of durations for the frozen system is representative for the non-frozen system. The question is "how" much it matters that the assumption is false; perhaps the frozen system still provides information about the non-frozen system.

To evaluate this, we define a score Σ which compares a sample of durations of the non-frozen system with the (estimated) asymptotic distribution of durations of the frozen system. Let $\tilde{d} = \{d_i\}_{i=1}^k$ denote the set of k sample durations at termination times $\tilde{t} = \{t_i\}_{i=1}^k$ from a single run of the non-frozen system over the time interval of interest $I = [t_0, t_1]$. In a first naive attempt, we normalise the empirical distribution of durations $\tilde{\rho}_\tau(D)$ for each $\tau = t_i$ such that the maximum is one. This gives a score for t_i equal to

$$\sigma_i = \tilde{\rho}_{t_i}(d_i).$$

The total score Σ is the average of all subscores

$$\Sigma = \frac{1}{n} \sum_{i=1}^{k} \sigma_i. \tag{9.10}$$

The maximum score $\Sigma = 1$ is obtained by a set of durations which coincide with the most likely duration for every t_i. The minimum score $\Sigma = 0$ is obtained for a set of durations which occur where $\rho_{t_i}(d_i)$ has zero density for every t_i. Intermediate scores are expected if a subset of durations occur at high density of $\rho_{t_i}(d_i)$ and the rest at low density, or if all durations occur at where $\rho_{t_i}(d_i)$ has intermediate density.

Note that an everywhere uniform distribution gives a score $\Sigma = 1$ for every pair (t_i, d_i), random or otherwise. Conversely, a narrow deltoid distribution produces a score of zero for nearly almost every random set of durations and times $\tilde{d}$ and $\tilde{t}$. Thus, the more informative a distribution is, the better Σ can distinguish between sequences of durations which are random and those that are well explained by the asymptotic distributions of durations.

However, the score (9.10) fails to work in practice because it penalises almost correct durations too severely. This is because the asymptotic distributions are discrete in general, and therefore almost every real duration, including durations d_i which are very close to peaks of the asymptotic distribution, get scores $\sigma_i = 0$, which goes against our intention. Therefore, we adopt smoothing in both duration and time space.

We choose a Gaussian smoothing filter with bandwidths $w_t = 10$ kyr in the time direction and $w_d = 3$ kyr in the duration direction, such that if writing $\rho_\tau(d) := \rho(\tau, d)$, the smoothed distribution $\overline{\rho}(\tau, d)$ is given by

$$\overline{\rho}(\tau, d) = \frac{1}{p w_t w_d} \sum_{i=1}^{p} K\left(\frac{t - t_i}{w_t}, \frac{d - d_i}{w_d}\right),$$

where p is the total number of points in all empirical distributions $\tilde{\rho}(\tau_i, d_i),\ i \in \{1, 2, ..., m\}$, and the Gaussian kernel is

$$K(x, y) = \frac{1}{2\pi} e^{-\frac{x^2 + y^2}{2}}.$$

Unfortunately, the smoothing introduces arbitrary parameters w_t, w_d and the arbitrary choice of smoothing kernel. Therefore, evaluation of the score Σ should be done for a range of kernel bandwidths before any conclusions can be drawn. Precise choice of kernel should not be important. The choice of bandwidth should reflect the amount of slack that we are willing to give sample durations which coincide well, but not perfectly, with the true distribution $\rho_\tau(d)$. A too wide bandwidth produces excessively high scores for durations which are clearly not well approximated by the true distribution, and a too narrow distribution penalises small errors too much. A tradeoff between these considerations must be made.

We now define a smoothed score

$$\overline{\Sigma} = \frac{1}{n} \sum_{i=1}^{k} \overline{\sigma_i},$$

where

$$\overline{\sigma_i} = \overline{\rho}(t_i, d_i).$$

The values of $\overline{\Sigma}$ are scaled towards larger values relative to those of Σ. Notably, the minimum is greater than zero. A set of sample durations $\tilde{d}_i$ are well explained by the asymptotic distribution $\tilde{\rho}(\tau, d)$ if $\tilde{d}_i$ produces a score $\overline{\Sigma}$ that is high compared to uniformly distributed random samples of the same size.

9.3 Results

We first present non-frozen model simulations and frozen distributions of durations for the Huybers 2007 model under obliquity, precession, and combined obliquity and precession forcing. Thereafter, we show calculated scores and empirical entropies.

9.3.1 Obliquity forcing: Durations in the fully frozen system

In Figure 9.2 we show an example run for the Huybers 2007 model forced by obliquity. Figure 9.2 a) illustrates how the empirical distribution of durations $\tilde{\rho}_\tau(f)$ for the fully frozen system is generated. The instantaneous amplitude $A(\tau)$ and period $T_f(\tau)$ of the forcing, as well as the linear ramp $R(\tau)$, are fixed at the values in the non-frozen system at time $\tau = -500$ kyr. This sinusoidally forced system with constant threshold is integrated for a long time (here between -2000 kyr and 0 kyr), and the durations between transitions d_i constitute the empirical distribution. In this case, there are two clusters of shorter and longer durations. This is also seen in Figure 9.2 c), where the empirical distribution of durations $\tilde{\rho}_\tau(d)$ expressed as black dots, has two clusters.

Figure 9.2 b) shows a simulation of the non-frozen model, which both displays frequency locking before -1200 kyr and a rapid shift to longer durations from -1350 kyr to -1000 kyr.

Figure 9.2 c) contains much interesting information.

1. Firstly, the frozen distribution of durations typically covers only part of the possible range of durations (white band). This means that the distribution is informative.
2. Secondly, the distribution of durations is clustered, typically near multiples of the instantaneous period $T_f(t)$. This is typical for frequency locking.
3. Thirdly, the clusters of the frozen distribution meander along with the instantaneous amplitude and period. When the instantaneous amplitude is large, durations tend to be shorter, and when the instantaneous period is long, durations tend to be longer.
4. Fourthly, clusters of long duration appear and disappear as the instantaneous parameters vary; examples of this are the clusters around -820 kyr, -650 kyr and -580 kyr. Thus, a monotonically increasing ramp $R(t)$ does **not** imply that the average duration increases monotonically.
5. Fifthly, the clusters of durations are not fixed at multiples of the average forcing frequency, as is the case for perfectly harmonic forcing. For instance, at -820 kyr the longest cluster is focussed on 105 kyr and at -650 kyr the cluster is focussed at 110 kyr, neither multiples of the average forcing frequency 41 kyr.
6. Sixthly, the non-frozen durations importantly tend to be close to the clusters of the frozen distribution. However, upon close inspection, some durations are not very close to the distribution. See for instance the duration around -426 kyr.

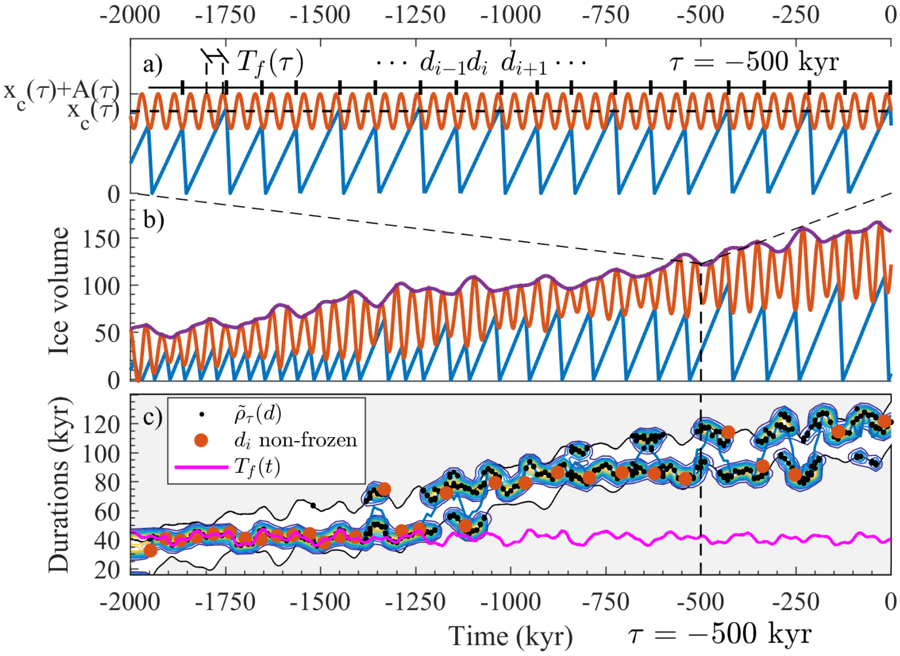


Figure 9.2: Huybers 2007 model forced by obliquity. Panel b) shows a simulation of ice volume in the non-frozen system (blue sawtooth), with obliquity forcing (red sinusoidal curve). The instantaneous amplitude $A(t)$ is shown in purple. Panel a) shows a simulation of the frozen system, with instantaneous amplitude $A(t)$ and period $T_f(t)$, as well as ramp function $R(t) = x_c(t) = 26+0.05(t-t_0)$, fixed at their values in the non-frozen system at time $\tau = -500$ kyr. The sample durations between glacial terminations in the frozen system are indexed by i and denoted d_i. Panel c) shows the durations from the non-frozen simulation (red dots) and the empirical distribution of durations for the frozen system $\tilde{\rho}_\tau(d)$. Contours show the smoothed distribution $\overline{\rho}_\tau(d)$. The magenta curve shows the instantaneous period $T_f(t)$, and the blue curve shows the mean of the distribution of durations $\tilde{\rho}_\tau(d)$. The white band on gray background shows the range of durations that is possible in the frozen system. Obliquity forcing is scaled to have standard deviation 20 between -2000 kyr and 0 kyr

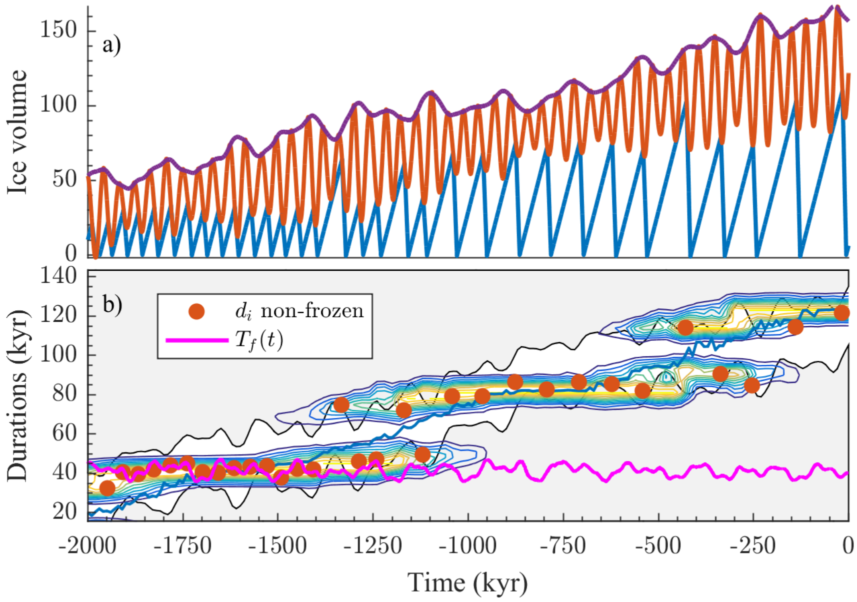


Figure 9.3: *Ramp-only* frozen distribution of durations between glacial terminations in the Huybers 2007 model forced by obliquity. a) Model ice volume (blue sawtooth) and forcing (red curve) with envelope (purple) as in Figure 9.2. b) Ramp only frozen distribution, menaning that the distribution of durations is estimated from a 2000 kyr simulation at each frozen time $\tau = -2000, -1990, ..., 0$, for which the ramp $R(\tau)$ is fixed, but the insolation amplitude and period $A(t)$ and $T_f(t)$ is free to vary. Contours are levels sets of the smoothed distribution of durations (see Figure 9.2). The forcing is scaled to standard deviation 20.

9.3.2 Obliquity forcing: Durations in the ramp only frozen system

Figure 9.3 shows the empirical distribution of durations between glacial terminations for the *ramp-only* frozen system. This distribution lacks information about the local amplitude and period of the forcing, and instead averages the response over time, as is evidenced by the monotonicity of its time evolution. Hence, one might expect the distribution to be a weighted average of the fully frozen distributions with $R(\tau)$ fixed, and $A(\tau)$ and $T_f(\tau)$ varied over their full ranges. This is not the case, however, as evidenced by the lack of ∼100 kyr durations around −620 kyr which are present for the fully frozen system in Figure 9.2. In Figure 9.3 the clusters of durations are broad at every time instant, unlike in Figure 9.2, where they typically consist of few durations.

9.3.3 Precession: Durations in the fully frozen system

The response to strong precession forcing, shown in Figure 9.4 is different from the response to obliquity forcing. Due to the more rapid 21 kyr period oscillations, the range of possible durations for each τ is shrunk. Consequently, the durations follow the undulations of the envelope more closely. If only the ramp (and not the forcing parameters) had been frozen, as in Figure 9.3, then the temporal tracking of the envelope would have been obscured. Note, that although the possible durations are confined to a narrow band, the distribution is rather uninformative, since the smoothed distribution of durations covers nearly the entire band (see Section 9.2.6).

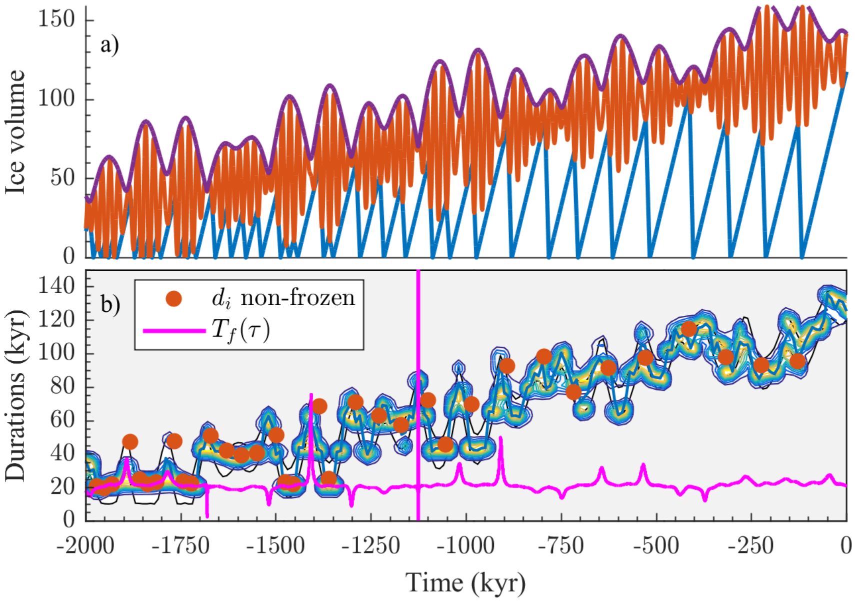


Figure 9.4: Fully frozen distribution of durations between glacial terminations for Huybers 2007 with precession forcing. a) Non-frozen model simulation, as in Figure 9.2. b) Smoothed frozen distribution as level curves. Durations from the model run as red dots. The instantaneous period $T_f(t)$ spikes where the instantaneous amplitude is weak. The forcing is scaled to have standard deviation 20.

9.3.4 Obliquity and precession: Durations in the fully frozen system

Forcing consisting of 33% obliquity and 67% precession differs from the previous cases as it is a sum of two modulated sines. The envelope of the forcing is therefore less regular, and dependent on the relative phase of the modulated sines in the two forcing components. When precession is strong relative to obliquity, durations sometimes track the envelope, as around -1850 kyr. When obliquity is strong as around -1600 kyr, however, there is local frequency locking to obliquity. From -1000 and onward there is no noticeable frequency locking of durations, which gradually increase monotonically.

9.3.5 Informativeness of frozen distributions

In Figure 9.6, we compare the informativeness of the frozen distributions generated by obliquity forcing, precession, a combination of obliquity and precession, respectively. We show informativeness both the for non-smoothed and smoothed distributions $\tilde{\rho}_\tau(d)$ and $\overline{\rho}(\tau; d)$.

For both obliquity and precession forcing the non-smoothed entropy is zero at several times. This means that all durations are contained in a single 3 kyr wide bin, which does not cover the entire interval of possible durations. Typically, this is indicative of local frequency locking, as can be seen by comparing Figure 9.6 a,c) with Figure 9.2 and Figure 9.4. The entropy for precession is nearly one where the interval of valid durations is narrow. At most other times, the entropy is approximately 0.4, which indicates intermediate informativeness, typically due to the presence of two clusters of durations for the same frozen time τ.

In contrast, the entropy for the non-smooth combined obliquity and precession forcing is zero only rarely; typically the entropy stays between 0.25 and 0.75 with an average of

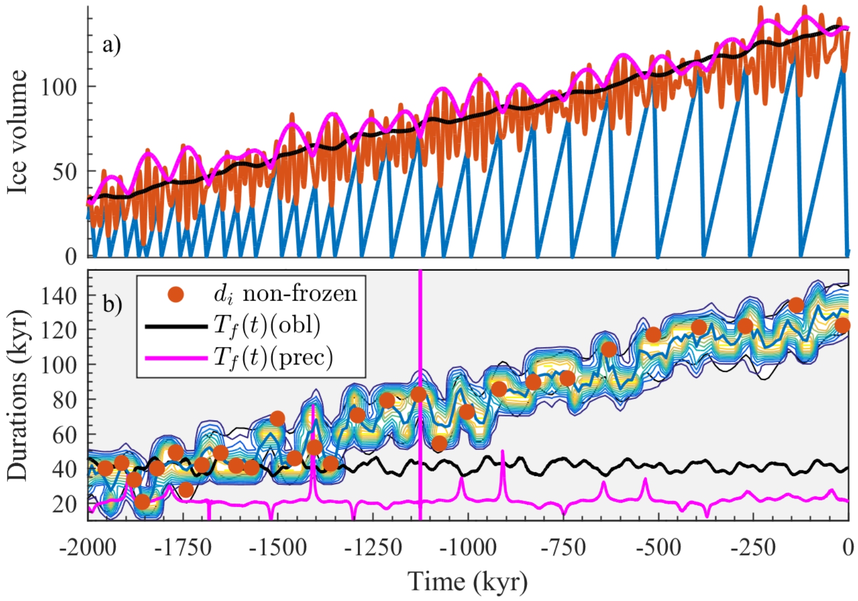


Figure 9.5: Fully frozen distribution of durations between glacial terminations for Huybers 2007 forced by a sum of 33% obliquity and 67% precession amplitude. a) Non-frozen model simulation, as in Figure 9.2. The black and magenta curves are the instantaneous amplitudes of obliquity and precession respectively. b) The frozen distribution is shown as level curves. Durations from the model run are shown as red dots. Obliquity is scaled to have standard deviation 5 and precession to have standard deviation 10

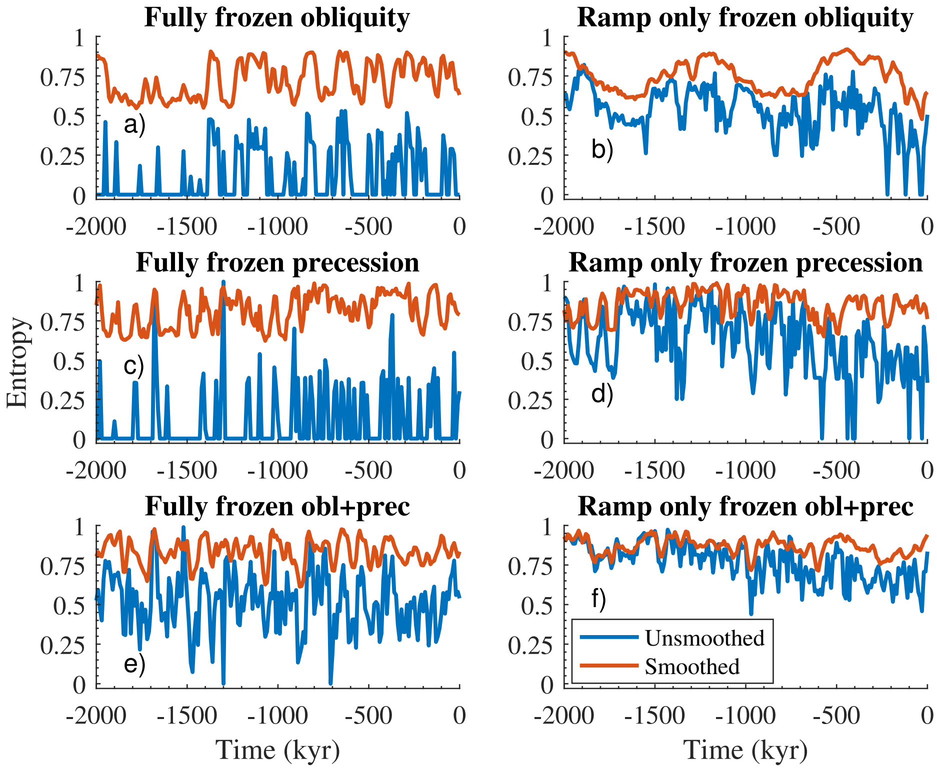


Figure 9.6: Normalised information entropy (9.5) for frozen systems. Panels a,c,e) show entropy for the fully frozen system and panels b,d,f) show entropy for the *ramp only* frozen system in which the ramp $R(\tau)$, but not instantaneous insolation parameters $A(t)$ and $T_f(t)$ are frozen at time τ. Blue curves show entropy for the empirical non-smoothed frozen distribution of durations $\tilde{\rho}_\tau(d)$ binned in 3-kyr intervals between the bounds of possible durations (see Section 9.2.6). Red curves show the entropy for the smoothed frozen distribution of durations $\overline{\rho}(\tau; d)$ after binning large samples of it in 3-kyr bins over the range of allowed durations

0.5. This reflects that the non-smoothed distribution of durations is rather diffuse as can be inferred from Figure 9.5 c).

The entropy for all smoothed distributions are focussed around 0.75 and never strays under 0.5. This is natural, since local peaks are attenuated. Still, local variations in the entropy of non-smoothed distributions on time scales longer than the smoothing (10 kyr) can still be discerned in the entropy of the corresponding smoothed distributions.

Turning to the *ramp only* frozen system, the entropy of the non-smoothed distributions of durations are considerably higher than for the fully frozen system. This is expected since the distribution of durations for fixed τ is naturally broader as the instantaneous amplitude and period are allowed to vary. Comparing Figure 9.6 and Figure 9.3 (obliquity forcing) shows that the entropy is higher when the distribution of durations is multimodal, which is natural. This is seen to a lesser extent for precession. Here, the entropy can be high despite the distribution being unimodal, wherever the range of possible durations is narrow.

The entropies of smoothed and non-smoothed distributions of durations are more similar than for the fully frozen system. This is natural, since smoothing an already diffuse distribution has little effect. The difference for combined obliquity and precession forcing is remarkably small, which is expected since the distribution is rather diffuse to begin with.

9.3.6 Comparing the frozen and non-frozen durations

The blue histograms in Figure 9.7 show empirical distributions of scores $\overline{\Sigma}$, calculated for random sequences of durations from the set of possible termination time-duration pairs (within white bands in e.g. Figure 9.2), having the same length as the sequence of duration from the non-frozen system. The first column a,c,e) shows scores under the fully frozen system for obliquity, precession and mixed forcing (Figure 9.2,9.4,9.5). The second column b,d,f) shows corresponding scores for the *ramp only* frozen system (empirical distribution shown only for obliquity in Figure 9.3). Each red vertical line shows a score for a sample run of the non-frozen model, using either a smoothed score function $\overline{\sigma}$ generated from the fully frozen distribution (Figure 9.7 a,c,e)) or the *ramp only* frozen function (Figure 9.7 b,d,f)). Yellow histograms show distributions of scores for sequences of time-duration pairs sampled randomly from the non-smoothed empirical distribution of durations for the frozen system.

Figure 9.7 shows that the non-frozen score exceeds all of the scores for random samples from a uniform distribution defined on the range of possible durations for obliquity forcing, both for the fully frozen and *ramp only* frozen system (Figure 9.2a) and b)). Thus, it is highly unlikely that a random sequence of durations would generate the observed score. The implies a significant association between the frozen distribution of durations and the sequence of durations in the non-frozen system.

On the other hand, the non-frozen score is much smaller than scores for sequences sampled from the non-smoothed frozen distribution of durations (yellow histogram). This indicates that the non-frozen durations agree significantly less with the fully frozen distribution of durations than if they actually were samples from it.

In contrast, the distribution of scores for durations sampled from the non-smoothed empirical *ramp only* distribution of durations $\tilde{\rho}_\tau(d)$, agrees better with the score for the non-frozen model run. The non-frozen score under the smoothed *ramp only* frozen distribution is thus typical for an “optimal” sequence of durations (we cannot expect higher scores than for random samples from the distribution of durations itself). Hence, based on score, the *ramp only* distribution of durations is a good explanatory model for the sequence of non-frozen durations.

For precession, the fully frozen distribution exceeds the non-frozen score in ∼7% of random samples (Figure 9.7 c)). Note that the tails of the distribution are not well sampled, so the precise percentage should be interpreted with caution. Nevertheless, it is unlikely that the sequence of non-frozen durations agree with the frozen distribution merely by chance.

The *ramp only* frozen distribution under precession forcing (Figure 9.7 d)), on the other hand, does not overlap with the score for the non-frozen system. This is both because the

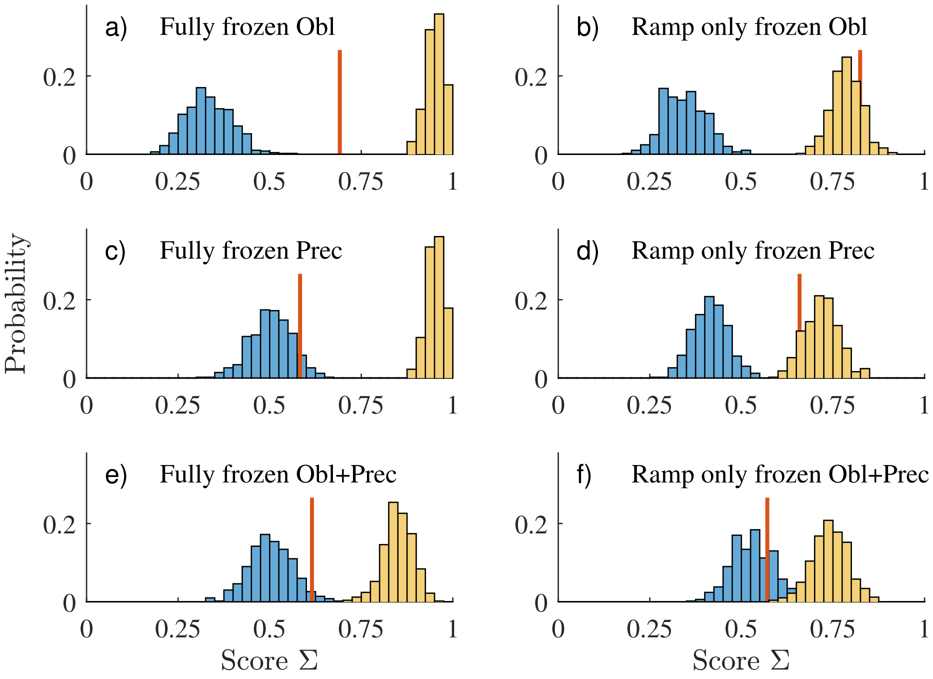


Figure 9.7: Scores $\overline{\Sigma}$ for three forcings: obliquity, precession, and a sum of 33% obliquity and 67% precession. Histograms show the empirical distribution (500 samples) of score under two different frozen scenarios; a,c,e) is for the fully frozen system, and b,d,f) is for the ramp only frozen system. Blue histograms show distributions of score for random sequences of durations sampled uniformly from the set of possible durations. Yellow histograms show distributions of scores for sequences of durations sampled from the smoothed frozen distributions of durations. Red vertical lines show the score of a typical simulation of the non-frozen system. If the red line intersects a point of high density of the blue histogram, then the simulated sequence of durations fits the distribution of durations (NB: not distribution of score) as well as a random sequence. On the other hand, intersection with the yellow histogram implies as good a fit as durations sampled from the frozen distribution of durations itself

mean of the frozen distribution is shifted towards lower values, and because the non-frozen score is higher. The non-frozen score overlaps the *ramp only* frozen distribution of scores, but not as much for obliquity.

For a combination of 33% obliquity and 67% precession forcing, the fully frozen distribution of scores exceeds the non-frozen score in ~2% of cases, comparable to for precession forcing. The *ramp only* distribution of scores, however, exceeds the non-frozen score in 27% of cases, so it is a rather likely that a randomly drawn sequence of durations produces a score greater than that for the non-frozen sequence of durations.

For combined obliquity and precession forcing, the distribution of scores under samples from the unsmoothed distribution of *fully* frozen durations has a markedly lower median than for pure obliquity or precession forcing (0.85 vs 0.95). This is because the pure forcing frozen distributions are typically unimodal, whereas the combined forcing distributions are broader.

In the discussion we explain why the scores for the *ramp only* frozen system are higher than for the fully frozen system, and how to interpret the differences in the panels of Figure 9.2.

9.4 Discussion

We being by interpreting the results of the previous section, in particular those of score and information entropy. We then discuss some methodological and epistemological questions.

9.4.1 Interpreting the scores

Scores for sequences of durations from the non-frozen system must always be interpreted relative to distributions of scores from the frozen system. This is because of three reasons. The first is the bandwidth of the smoothed distribution of durations; as the bandwidth approaches zero, the score for almost any sequence of real durations approaches zero. Conversely, a large bandwidth makes all scores close to one. The second reason is due to the width of the support of the distribution of durations; a wide support makes random durations produce small subscores, while a narrow support makes them produce large subscores. The third is that distributions of durations which are focussed on narrow peaks of equal height naturally produce higher subscores than other distributions; e.g. a distribution with probabilities 0.5 assigned to two events produce a subscore score of 1 for every observable event, while a distribution with probabilities 0.25 and 0.75 only has a 75% chance to produce a subscore 1 (and a 25% probability to produce a subscore 1/3).

Having said this, we return to the results of Section 9.3.6. Judging by the scores, all frozen distributions except for the *ramp only* frozen distribution for combined obliquity and precession forcing provide significant information about the sequences of durations which occur in the non-frozen system, and thus have explanatory value.

However, for pure obliquity and precession forcing, the non-frozen score is higher for the *ramp only* distribution of durations than for the fully frozen distribution. Furthermore, the non-frozen score agrees better with the “ideal” distribution of scores for durations which are sampled from the non-smoothed distribution of durations. Thus we conclude that the fully frozen distribution of durations is a too restrictive model of durations in the non-frozen system.

A methodological shortcoming might be the cause of this, however. For the *ramp frozen* distribution of random scores, we have assumed the same bounds on durations as for the fully frozen system (white band in Figure 9.3). Figure 9.3 clearly shows, that the *ramp only* empirical distribution of durations sometimes (e.g. at −620 kyr and −220 kyr) exceeds the supposed bounds and sometimes is hardly overlaps the band of “possible” durations. Thus, the subscores of random durations sampled from the white band at these frozen times τ inevitably produce low scores, although it should be possible to obtain the maximum score of 1 for every τ. Hence, if correct bounds on durations were used, we would expect the median of the *ramp frozen* distribution of random scores (blue histogram) to be higher, and thus would make the non-frozen score less remarkable.

The distribution of scores from sampling of the empirical distribution of durations (yellow histogram), for combined obliquity and precession forcing, is comparable for the fully frozen and *ramp only* frozen distributions of durations. This is because both the non-smoothed fully frozen and *ramp only* distributions of durations cover the entire band of valid durations rather evenly. This is reflected in the rather low informativeness of the fully frozen distribution of durations (Figure 9.6).

Contrary to what Figure 9.7 c) suggests, that the score for the non-frozen sequence of durations is more similar to a random distribution than an ideal distribution, Figure 9.4 suggests that the non-frozen durations coincide rather well with the frozen distribution of durations. In particular, the sequence of durations follows the envelope of the insolation well. This apparent contradiction has two sources. The first is that the distribution of random scores of durations has a high median score since the random samples from the narrow strip are likely to coincide with a peak of the distribution of durations. The second is that many non-frozen durations fall outside of the frozen distribution, even though the overall undulations follow the distribution well. For instance, the time series resulting from sampling the mean of the frozen distribution whenever a duration occurs in the non-frozen distribution likely is highly linearly correlated with the sequence of durations, even after removing the linear trend.

We note that if the distributions overlap, then it is not possible to tell a random sequence of durations from a "true" sequence of durations, samples from the distribution used to define the score. We also note that for weak forcing the band of allowed durations shrinks to a line, and frequency locking effects diminish. In this case, we expect both random and non-frozen scores to be very high.

Lastly, we point out that the frozen distribution of durations, and consequently the score, does not contain any information about which particular sequence of durations will occur. In order to learn of this, the knowledge of the complete dynamics is generally required.

9.4.2 Interpreting informativeness

It is non-trivial to interpret the results in Section 9.3.5 and Figure 9.6. First, recall from Section 9.2.6 that low normalised entropy implies high informativeness. For instance, when entropy is close to 1, the distribution is nearly uniform and all outcomes are equally likely. On the contrary, when entropy is close to 0 the distribution is unimodal and narrow (a discrete distribution has all probability assigned to a single event), and the distribution gives full information about observed events. Intermediate entropy can result from different causes.

We note that zero entropy implies that all durations fall onto the same value, up to the 3 kyr binning of the histogram. This follows from the $1:1$ frequency locking observed in Figure 9.2, but the entropy can be zero also without frequency locking.

An entropy close to one can occur for distribution with a narrow support. Thus, a high entropy does not necessarily imply that the distribution is τ is broad and uniform, it could just be that the support of the distribution is narrow.

We note that the entropy may be sensitive to bin choice and bin width when the support of the distribution is narrow. For instance, different choices of bin width of position can change the entropy for the sample durations $\{1,1,1,1,2,2,2,2\}$ on the interval $[0,3]$ can lead to entropy either zero or one. Taking bin edges $\{0,1.5,3\}$ gives entropy 1, whereas bin edges {-0.75,0.75,2.25,3.75} gives entropy 0. Therefore, entropy can fluctuate rapidly without reflecting true changes in informativeness.

9.4.3 Discussion about score

There are several issues with Σ and $\overline{\Sigma}$, may which can be addressed. Firstly, the smoothing in $\overline{\Sigma}$ introduces arbitrariness in the form of the bandwidths w_t and w_d. One option is to choose them as to minimise the overlap with the positive tail of the random distribution of scores. This is insensitive to overfitting, since almost certainly the optimum for score is found for intermediate bandwidths; large bandwidths likely give a thick tailed distribution of random scores and small bandwidths makes random scores vanishingly small. Intermediate bandwidths should be insensitive to small tweaking of the bandwidth.

Another criticism may be that single runs tend to contain more short durations (e.g. before -1400 kyr in Figure 9.2), making the average score depend more on shorter than longer durations. To reduce this problem, one can either sample durations only at termination times in the single run, which introduces another bias, or one can estimate a continuous rate of terminations function based on the single run terminations, and sample termination times from this rate function.

Thirdly, the non-frozen model run subscores $\overline{\sigma}_i$ are not independent. Therefore, it is strictly speaking not appropriate to compare the single run score to random scores which have independent subscores, as we do. It is difficult to overcome this potential source of bias, however.

Fourthly, one might question the appropriateness of sampling durations randomly when calculating random score; after all, sequences of durations are deterministic. However, once again we reiterate our premise that we assume that we do not know the system dynamics or previous history of the system. We only ask: "If a duration were to occur at this time instant assuming that dynamics are quasistatic, what is the probability that the duration falls within a small interval Δd of some particular d^*?"

The normalisation of $\rho(\bar{\tau}, d)$ when calculating score $\overline{\Sigma}$ might seem unnatural: why not just normalise to probability and let each subscore be the probability (density) of observing d_i given t_i? The reason for this is twofold: firstly, it is desirable to have an easily interpretable number that equals 1 if the most likely duration occurs at each time τ_i. Secondly, we wish to put unimodal and bimodal distributions on equal footing; obtaining the most likely duration at any time instant should count equally towards the total score.

9.4.4 More knowledge required than gained?

The technique described in this chapter can be criticised for requiring complete knowledge of the system in order to derive incomplete knowledge. Specifically, the dynamics of the system must be fully known in order to generate the asymptotic distributions of durations $\rho_\tau(d)$ and the bounds on possible durations. On the other hand, the distributions we derive only contain probabilistic knowledge. We have two responses to this.

Our first response is that many other indicators, such as Lyapunov exponents, require complete knowledge about the dynamics but yet provide valuable insight that is not evident from observing the system equations alone. Secondly, we do not intend these methods to be used for inference in unknown systems, but rather to provide additional understanding to systems whose governing equations are known, and which display interesting behaviour such as temporal clustering of durations.

9.4.5 The bounds on possible durations

The ability to define bounds on possible durations in H07 is a luxury available only to a few systems. Many systems have an unbounded phase space which can lead to arbitrarily long durations. Instead of imposing hard bounds on such systems, one can instead simulate the system from a suitably sampled subspace, for different phases of the instantaneous forcing to get an approximate (unbounded) distribution of durations. An idea could be to sample random durations from this distribution when generating a null distribution for the score $\overline{\Sigma}$, rather than from a uniform distribution between the bounds, as we do now.

9.5 Conclusions

We can now answer the question posed in the introduction. The response of the Huybers 2007 oscillator model to modulated forcing can be understood from the time-dependent distribution of durations between glacial terminations for the frozen system, for which some model parameters are frozen. Roughly speaking, termination time-duration pairs in simulations of the non-frozen system tend to occur where the distribution of durations in the frozen system has high density. This distribution reveals clustering of durations, a frequency locking effect, and local time variations of the distribution of durations due to modulation of the amplitude and period of the forcing.

We quantify the agreement between non-frozen model run termination time-duration pairs and the distribution of durations in the frozen system with a score proportional to the local density of the distribution. This score reveals an association between durations in non-frozen model runs and the distribution of durations, which is stronger than for random sequences of durations. However, the score is higher when using a distribution of durations from the frozen system for which only a slowly ramped parameter is fixed, compared to fixing both the ramp and modulation parameters. This suggests that accounting for local amplitude and period variations may be too restrictive.

However, our conclusions change depending on our assessment criteria and assumed information. For instance, the score for precession forcing is rather low compared to random durations, even though the distribution of durations and durations in the non-frozen simulation follow each other closely (Figure 9.4). This is in part because we assume knowledge that random durations can only be sampled from a narrow distribution. Assuming less such knowledge about sampling distribution would lead to a different conclusion about the utility of the frozen system distribution of durations to explain durations in the non-frozen system.

Information entropy quantifies how informative a distribution of durations is. For instance, if all durations are concentrated to a narrow cluster, then the entropy is low and much is known about which durations will be observed. This is typical for frequency locking to periodic forcing. Monitoring entropy over time tracks the amount of information conveyed by the frozen distribution at a given time.

Several methodological problems require further investigation. Smoothing, sampling, and alternative measures of informativeness are a few. It would be interesting to extend the method to stochastic dynamics, and to evaluate it for more models and forcings. We do not expect the method to work well in systems for which convergence to attracting solutions is very slow, or for which the history of forcing modulations prior to termination events play a significant role.

In conclusion, the methods presented in this chapter advance the understanding of the response of nonlinear oscillators to by modulated sine forcing, but much further research is needed.

Chapter 10

Dynamics behind the middle Pleistocene transition in the Paillard 1998 model

10.1 Introduction

We learned in Chapter 7 that the middle Pleistocene transition (MPT) can occur due to ramping of a system parameter and frequency locking, a mechanism we called *ramping with frequency locking* (RFL). The Paillard 1998 model (henceforth P98) was cited as one of the first published models of glacial cycles to explain the MPT with RFL.

However, this description faces a problem due to P98 being forced by summer solstice insolation at 65 North. This forcing is dominated by precessional periods 19 and 23 kyr. Under RFL, we generically expect the average duration between glacial termination to increase stepwise, moving through all multiples of the dominant forcing frequencies $1 \times 19, 1 \times 23, 2 \times 19, 2 \times 23$ etc. Such a progression of durations is not supported by data (see Paper A).

In Paper A we suggest that the durations between glacial terminations increased from ca 40 kyr to ca 80 kyr around -1250 kyr. This can be explained by frequency locking to twice the dominant obliquity period 41 kyr, as is probably the case in the Paillard 2004 model [Paillard and Parrenin, 2004]. In P98, however, the durations immediately become centered on period $\sim$100 kyr around -1000 kyr. This cannot be explained by frequency locking to obliquity, but requires another explanation.

The late Pleistocene $\sim$100 kyr glacial cycles is frequently attributed to frequency locking to the eccentricity envelope of precession (e.g. [Raymo, 1997, Lisiecki, 2010]). However, we are not aware of any model of glacial cycles that explicitly demonstrates how such locking can arise. Here, we provide such a model, which is closely related to P98.

Next, we explain how the glacial cycle period changes abruptly from $\sim$41 kyr to $\sim$100 kyr. We show that the durations between glacial terminations in fact do not increase abruptly, but vary considerably and on average increase gradually over time. The abrupt transition in glacial cycle period is only apparent from frequency analysis or visual inspection. We show that P98 has a particular set of features which allows for this abrupt change in apparent period, even if the durations between terminations do not change abruptly. Removing one or more of these features gives the model simulations more generic behaviour with a less abrupt increase in perceived glacial cycle period. This indicates that the mechanism is not very robust. First of all, however, we introduce the model.

10.2 The P98 model

P98 has three states i, g and G corresponding to interglacial, mild glacial and deep glacial conditions. The evolution rule of ice volume v is given by

$$\dot{v} = \frac{(v_{R,0} - v)}{\tau_R} - \frac{F(t)}{\tau_F}, \tag{10.1}$$

where $v_{R,0} = [0, 1, 1]$ is an equilibrium ice volume, $\tau_R = [5, 80, 80]$ is a state dependent rate (for $R \in \{i, g, G\}$ respectively), and $\tau_F = 28$ is a constant rate. $F(t)$ is a truncated insolation function and relates to the Laskar summer solstice insolation at 65 N $I(t)$ as

$$F(t) = \text{normalise}(f(\text{normalise}(I(t) + kt))), \tag{10.2}$$

where $k = 3 \cdot 10^{-3} W/m^2 kyr$, normalise$(\cdot)$ subtracts the mean and divides by the standard deviation on a considered interval $t \in [t_0, t_1]$, and $f(x)$ is a truncation function

$$f(x) = \frac{1}{2}(x + \sqrt{4a^2 + x^2}), \tag{10.3}$$

with $a = 1$.

The state switch rules are

$$\begin{cases} i \to g & \text{if } F(t) < i_0 \\ g \to G & \text{if } v \geq v_{max}(t) \\ G \to i & \text{if } F(t) > i_1, \end{cases} \tag{10.4}$$

where i_0 and i_1 are lower and upper thresholds for $F(t)$ and $v_{max}(t) = 0.35 + 0.75(t - t_0)/(t_1 - t_0)$.

P98 acts much like a forced integrate-and-fire oscillator; in the g state, ice volume grows linearly until a threshold is reached. Then, in the G and i states, ice volume decreases towards 0 according to an insolation rule which gives typical deglaciation times of about 6-12 kyr. Thereafter the system enters the g state and the cycle repeats.

The linear ramping of v_{max} causes an increase of the internal time scale of oscillation, which is counteracted in part by the linear reduction in insolation.

10.3 Comparison between model and climate data

We characterise the change in period across the MPT using two tools, the durations between $G \to i$ transitions and a wavelet spectra. The former focusses on the durations between abrupt terminations of glacial conditions and can be identified as a Poincaré subsection, a mathematical tool used to measure periodicity of solutions. Wavelet spectra instead measure the local frequency composition of a signal.

Comparing durations between glacial terminations and wavelet spectra for the P98 model and the LR04 stack, a proxy for past ice volume and deep sea temperatures, reveals striking differences (Figure 10.1 and Figure 10.2).

Durations between glacial terminations in the LR04 stack, identified manually, are close to 40 kyr until about −1250 kyr (with an exception at −1600 kyr) and do not exceed 100 kyr until −400 kyr. Durations in the simulation run of P98, on the other hand, exceed 100 kyr multiple times over the past 2000 kyr, most notably within the so-called “40 kyr world” prior to −1250 kyr. Based only on the progression of durations in P98, the model clearly misrepresents the climate evolution across the MPT.

However, the wavelet spectra of model and proxy data agree better. Both spectra show a marked increase in long, approximately 100 kyr, period amplitude around −800 kyr. Between −2000 and −1200 kyr, both spectra contain ample 40 kyr amplitude, and little

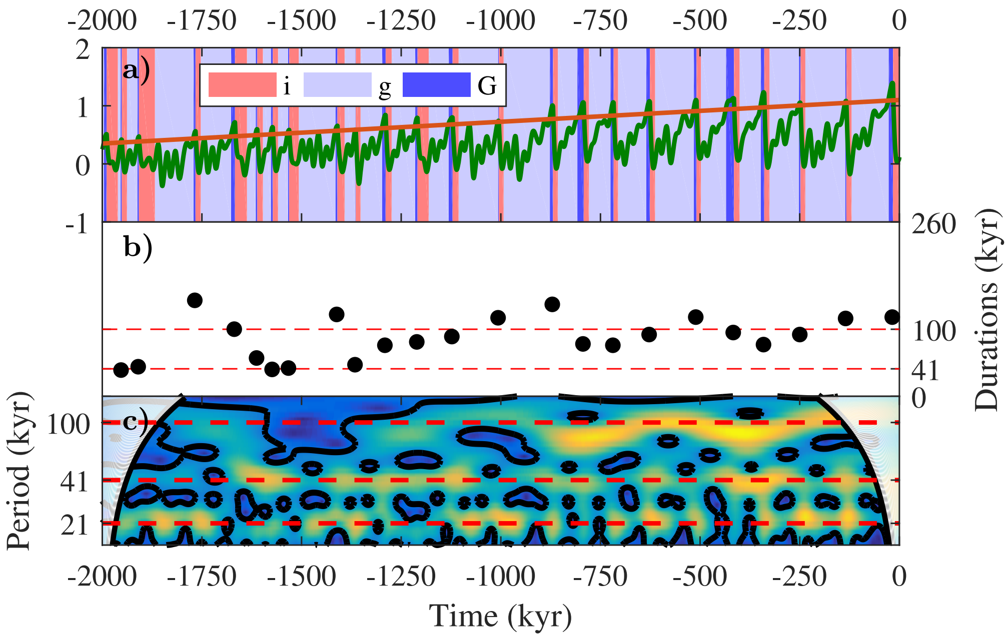


Figure 10.1: Simulation of the Paillard 1998 model((10.1),[Paillard, 1998]). a) Model run. Simulated ice volume is shown in green, and the threshold for $g \to G$ transitions is shown as a red line. The state (i, g or G) is indicated by background colour. b) Durations between $G \to i$ transitions (glacial terminations) are shown as black dots. Red dashed lines show historically signifcant periods. c) Morlet wavelet amplitude spectrum (square root of power) with bandwidth parameter $\omega_0 = 6$ kyr [Torrence and Compo, 1998]. Lighter colors means more wavelet amplitude. Above the thick black contour, wavelet amplitude is at least 30% of the maximum wavelet amplitude in the plot. The opaque regions on the edges show the cone of influence, outside which edge effects are non-negligible. The period scale is logarithmic. All parameters are as in [Paillard, 1998], and the forcing is Summer solstice insolation at 65 N (see Fig 10.3)

amplitude at longer periods. Both the model and LR04 stack see an increase in approximately 80 kyr amplitude around −1200 and −800 kyr, with a reduction around ca −1000 kyr. The model produces more precessional ∼21 kyr amplitude than in data, as noted in [Paillard, 1998], but we do not focus on this discrepancy here.

In this chapter, we aim to explain 1. how P98 manages to reproduce on average 100 kyr durations and ∼100 kyr wavelet amplitude after ca ∼1000 kyr and 2. why the wavelet spectrum contains little 100 kyr amplitude before −1000 kyr, despite there being long durations between glacial terminations.

10.4 Modulated insolation curves

P98 is forced by insolation at 21 June at 65 degrees North, which can be well described as a sum of 71% precession and 29% obliquity (in amplitude) (Figure 10.3). Obliquity is the tilt of Earth's axis of rotation, and precession is a parameter describing the orientation of Earth's axis of rotation relative to the sun at different times of the year.

Importantly, both obliquity and precession can be well described as amplitude and frequency modulated sine waves with average periods 41 kyr (obliquity) and 21 kyr (precession). Eccentricity, the ovalness of Earth's orbit around the sun, forms the amplitude envelope of precession, with dominant periods ∼100 kyr and ∼400 kyr. The envelope of obliquity is less regular, but it varies on time scales of about 150 kyr and 1200 kyr [Mélice et al., 2001].

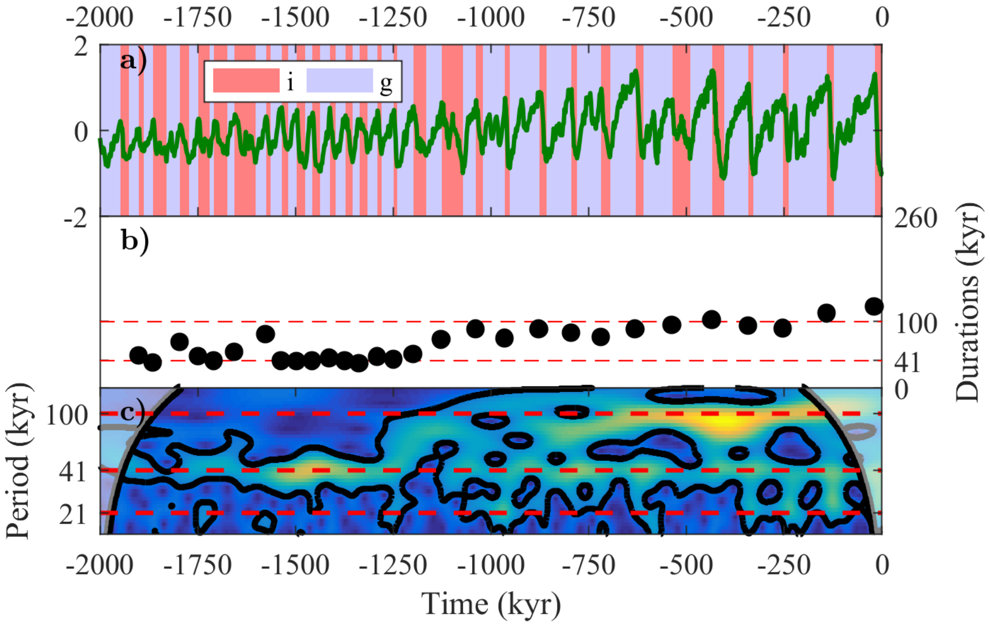


Figure 10.2: The LR04 stack of benthic foraminifera, a record describing variations in global ice volume and deep ocean temperatures. Description as in the caption of Figure 10.1, except that there are only two states i and g, which are manually identified. Transitions in b) are defined as $g \to i$ switches

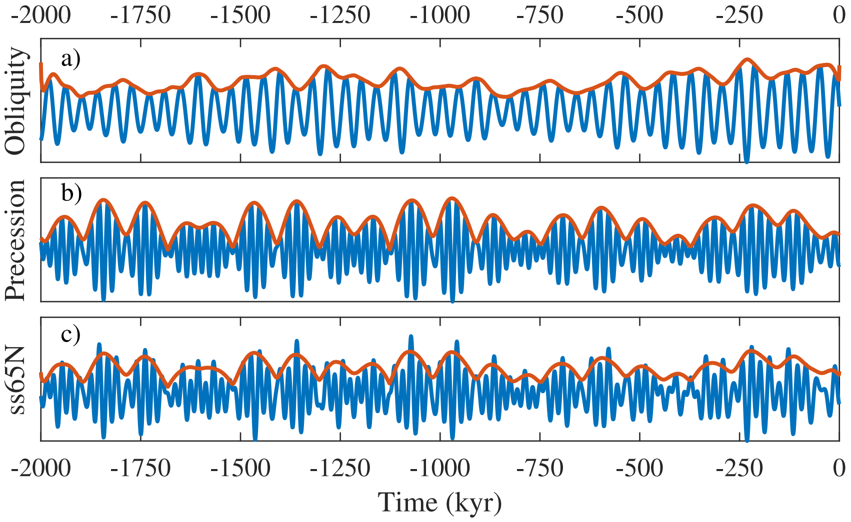


Figure 10.3: Obliquity, precession, eccentricity and summer solstice insolation at 65 degrees North [Berger, 1978, Laskar et al., 2004]. Summer solstice insolation at 65N is approximately a linear combination of 29% obliquity and 71% precession.
Blue curves show insolation curves and red curves show envelopes. In a) the envelope is estimated from the Hilbert transform (see Chapter 9), in b) it is eccentricity, and in c) it is a sum of 29% obliquity and 71% precession envelope

The amplitude modulations cause the system to respond differently to the forcing at different times, depending on the local size of its envelope, in contrast to pure sine forcing which has a constant amplitude. Unlike a model forced purely by the envelope, however, the sinusoid structure makes the influence of insolation average out over one period of the *carrier* (rapidly oscillating) wave as long as the dynamics are linear (no thresholds etc.). These two effects combined will help us understand why the system prefers durations which are on average close to 100 kyr, even though the internal time scale of oscillation changes drastically.

10.5 Frequency locking to a beat frequency

It has been suggested that the eccentricity envelope of precession is responsible for the late Pleistocene ice ages, both in the Paillard 98 model [Huybers, 2007], and more generally [Raymo, 1997, Ridgwell et al., 1999, Lisiecki, 2010]. So far, however, no models in the literature clearly show how an oscillator can lock to the period of an envelope, even though the forcing contains no or little amplitude at that period, as is the case for precession forcing. Here, we create and analyse such a model, which approximates the P98 model.

The 100 kyr precession envelope can roughly be described as a beat frequency between the periods 19 and 23 of precession, since $T_{envelope} \approx \frac{1}{2}1/((\frac{1}{19} - \frac{1}{23})/2) \approx 100$ kyr. (The factor $\frac{1}{2}$ comes from the envelope is defined to be positive). The theory of two-frequency locking [Pikovsky et al., 2001, Feudel et al., 1997] predicts that frequency locking can occur to such a linear combination of forcing frequencies, but it does not describe on how large parameter regions locking to combination frequencies occur. We expect frequency locking to an amplitude modulated sine to occur ubiquitously in models satisfying the following assumptions.

Assume that growth of ice volume $x(t)$ is described as

$$\begin{gathered}\dot{x}(t) = g(x) - f(t),\\ x \to 0, \text{ when } x(t) = x_c,\end{gathered} \tag{10.5}$$

where x_c is a constant threshold ice volume, $g(x)$ is a monotonically increasing function describing internal ice volume growth and $f(t) = h(t)k(t)$ a zero-mean forcing function. The function $h(t)$ is non-negative with period T_g (g for group). The function $k(t)$ is locally symmetric and zero-mean, with maximum amplitude 1 and the power focussed on a single period, making it almost periodic with a period T_p (p for phase, also called carrier). Assume (A1) that $T_p \ll T_g$, such that $h(t) \approx$ constant over one carrier period. Consequently, $h(t)$ can be interpreted as the upper envelope of $f(t)$. We also assume (A2) that $g(x)$ is weakly nonlinear, such that $x(t + T_p) - x(t) \approx x_{int}(t + T_p) - x_{int}(t)$ given a constant envelope and given that the threshold in (10.5) is not reached, where $x_{int}(t)$ is the evolution of ice volume without forcing variations. This assumption says that the effect on ice volume from forcing approximately averages out over one forcing period.

It is possible to justify why frequency locking to the envelope should occur in the rather general class of models just described; here we only illustrate frequency locking in one model in this class, however.

10.5.1 An example model

One model that (almost) conforms to the model assumptions is

$$\begin{aligned}\dot{x}(t) &= \mu - A \sin\left(\frac{2\pi t}{T_g}\right) \sin\left(\frac{2\pi t}{T_p}\right)\\ x &\to -\frac{AT_p}{2\pi} \sin\left(\frac{2\pi t_c}{T_g}\right) \cos\left(\frac{2\pi t_c}{T_p}\right), \text{ at time } t_c \text{ when } x(t_c) = x_c,\end{aligned} \tag{10.6}$$

from which we identify by comparing with (10.5):

$$\begin{aligned} g(t) &= \mu, \\ h(t) &= A\left|\sin\left(\frac{2\pi t}{T_g}\right)\right|, \\ k(t) &= \operatorname{sign}\left(\sin\left(\frac{2\pi t}{T_g}\right)\right)\sin\left(\frac{2\pi t}{T_p}\right), \end{aligned} \tag{10.7}$$

where sign denotes the sign of the argument, or zero if the argument is zero. The peculiar resetting condition is chosen to make the internal ice volume growth (defined in the next subsection) reset to 0, rather than the forced ice volume. This is to avoid a degeneracy which makes frequency locking impossible, while still keeping the model simple and analytically tractable. A simulation of this model is shown in Figure 10.4 b).

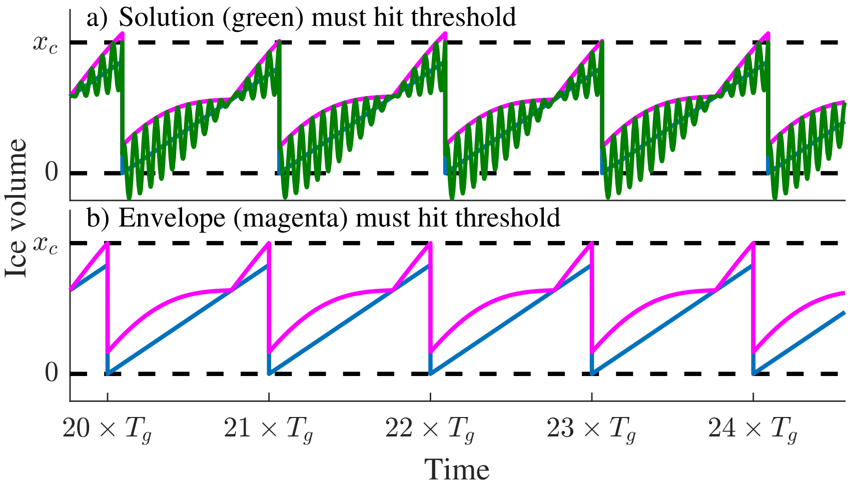


Figure 10.4: Simple models illustrating locking to a beat period. a) A termination occurs when the integrated insolation (green) superimposed on internal ice volume growth (blue) reaches a threshold (dashed black). The envelope (magenta), due to beating of adjacent frequencies, entrains the oscillator to its approximate period. b) An approximate model illustrating the role of the envelope, for which transitions occur when the envelope, and not the insolation curve, reaches a threshold.

10.5.2 Deriving a simpler model

Assuming that (A1) $T_p \ll T_g$, and given that (A2) $g(x) = \mu$ is linear, the forcing contribution to ice volume over an integer number of periods T_p is approximately zero, since over a single period:

$$\begin{aligned} x(t_0 + T_p) - x(t_0) &= \int_{t_0}^{t_0+T_p} \dot{x}(s)ds \\ &= \int_{t_0}^{t_0+T_p} \mu ds - \int_{t_0}^{t_0+T_p} A\sin\left(\frac{2\pi s}{T_g}\right)\sin\left(\frac{2\pi s}{T_p}\right)ds \\ &\approx \mu T_p + A\sin\left(\frac{2\pi t_0}{T_g}\right)\int_{t_0}^{t_0+T_p} \sin\left(\frac{2\pi s}{T_p}\right)ds \\ &= \mu T_p. \end{aligned} \tag{10.8}$$

Hence, to learn when the system crosses the threshold x_c we only need to consider how the envelope $h(t)$ affects the extrema of ice volume. We see that $h(t)$ dictates not only the extrema of the rate of change $\dot{x}$, but also the extrema of the solution, as is seen by integrating (10.6) over a general time t:

$$\begin{aligned} x(t) - x(t_0) &= \int_{t_0}^{t} \dot{x}(s) ds \\ &= \mu(t - t_0) - \int_{t_0}^{t} A \sin\left(\frac{2\pi s}{T_g}\right) \sin\left(\frac{2\pi s}{T_p}\right) ds \\ &\approx [x_{int}(s)]_{t_0}^{t} + \left[\frac{T_p}{2\pi} A \sin\left(\frac{2\pi s}{T_g}\right) \cos\left(\frac{2\pi s}{T_p}\right)\right]_{t_0}^{t}, \end{aligned} \tag{10.9}$$

where we identify new functions

$$\begin{aligned} \tilde{h}(t) &= \tilde{A}\left|\sin\left(\frac{2\pi t}{T_g}\right)\right|, \\ \tilde{K}(t) &= -\operatorname{sign}\left(\sin\left(\frac{2\pi t}{T_g}\right)\right)\left(\cos\left(\frac{2\pi t}{T_p}\right)\right), \\ x_{int}(t) &= \mu t \end{aligned}$$

and the new constant amplitude $\tilde{A} = T_p A/2\pi$, where x_{int} stands for *internal* ice volume. We explain the approximation in the last line of (10.9) in Section 10.8.4. Since we have assumed that (A1) $T_p \ll T_g$ and (A2) $g(x) = \mu$ is weakly nonlinear, we conclude that whenever $x_{int}(t) + \tilde{h}(t) = x_c$, within a short amount of time $\Delta t \in [0, T_p]$, $x(t)$ will also reach x_c. For frequency locking purposes we therefore ask when the system

$$\begin{aligned} \dot{x}_{int}(t) &= \mu, \\ x_{int} \to 0, &\text{ when } x_{int}(t) + \tilde{h}(t) = x_c, \end{aligned} \tag{10.10}$$

admits periodic solutions (Figure 10.4). Hence, we have approximated the problem as one of periodic frequency locking to the period $T_g/2$ (since $\tilde{h}(t)$ has period $T_g/2$).

10.5.3 Finding frequency locking regions

In what follows, we use the techniques from Section 4.7 to find that the average duration remains constant at multiples of the envelope period for a large set of parameters. These sets are called Arnold tongues; an example is shown in Figure 10.5. Below follows a derivation of conservative bounds for $1 : N$ tongues, parameters sets on which the oscillator completes one oscillation per N forcing envelope period, on average.

Index the epochs between transitions by n and denote by $t_{c,n}$ the n:th transition of the x_c threshold. Let $x_{int,n}(t)$ denote the internal ice growth function in epoch n and introduce the phase $\phi_n = \frac{2\pi}{T_g}(t - t_{c,n-1}) + \phi_{n-1}$ of the sine in $\tilde{h}(t)$ at time $t_{c,n}$. Then

$$\begin{aligned} x_{int,n+1}(t) &= \mu(t - t_{c,n}) \\ x_{int,n+1} \to 0, \text{ when } x_{int,n+1}(t) &+ \tilde{A}\left|\sin\left(\frac{2\pi}{2\tilde{T}_g}(t_{c,n+1} - t_{c,n}) + \phi_n\right)\right| = x_c, \end{aligned} \tag{10.11}$$

where $\tilde{T}_g = T_g/2$ is the period of the envelope.

For $1 : N$ frequency locking we require that the phase at termination is the same as the starting one, plus N half-rotations of the sine in (10.11), $\phi_{n+1} = \phi_n + N\pi$. This gives that

$$\begin{aligned} \frac{2\pi}{2\tilde{T}_g}(t_{c,n+1} - t_{c,n}) + \phi_n &= \phi_n + \pi N \\ (t_{c,n+1} - t_{c,n}) &= N\tilde{T}_g. \end{aligned} \tag{10.12}$$

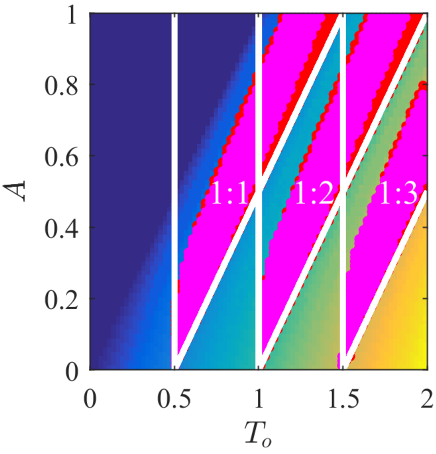


Figure 10.5: Arnold tongue (frequency locking) diagram for (10.10), an approximate version of the model (10.6) which exemplifies frequency locking to the envelope of amplitude modulated forcing. Lighter colours mean higher average duration. Red dotted regions indicate 1 : N frequency locking tongues for the approximated model (10.10), and magenta dots indicate corresponding regions for the non-approximated model (10.6). Parameters are $\mu = 1, T_g = 1, T_p = 1/30$, and internal period T_o is changed by varying x_c from 0 to 2. Major tongues are defined as all pairs $(T_o, \tilde{A})$ for which $|\bar{D} - NT_g/2| < 0.01$

We find for which parameters this condition is consistent with the threshold rule by inserting the expression for $t - t_{c,n}$ into (10.11):

$$\begin{aligned} \mu(t_{c,n+1} - t_{c,n}) + \tilde{A}\left|\sin\left(\frac{2\pi}{2\tilde{T}_g}(t_{c,n+1} - t_{c,n}) + \phi_n\right)\right| &= x_c \\ \mu N\tilde{T}_g + \tilde{A}\left|\sin\left(\frac{2\pi}{2\tilde{T}_g}(N\tilde{T}_g) + \phi_n\right)\right| &= x_c \\ \mu N\tilde{T}_g - \tilde{A}\left|\sin\phi_n\right| &= x_c \\ \frac{\tilde{A}}{\mu}\left|\sin\left(\phi_n\right)\right| &= T_o - N\tilde{T}_g, \end{aligned} \tag{10.13}$$

where $T_o = x_c/\mu$ is the internal period of the oscillator. From this expression we see that 1 : N frequency locking is possible only if

$$0 \leq \frac{\mu}{\tilde{A}}(T_o - N\tilde{T}_g) \leq 1. \tag{10.14}$$

Condition (10.14) gives us conservative bounds on synchronisation regions, within which 1 : N frequency locking is possible. We stress that this is only a necessary but not sufficient condition for locking since the threshold may be reached for an earlier time than the assumed $t_{c,n+1} = t_{c,n} + N\tilde{T}_g$. This is evident from Figure 10.5, which shows that numerically estimated Arnold tongues cover only a proportion of the area spanned by the conservative limits from (10.14). Higher order $M : N$, $M \neq 1$ Arnold tongues are thinner and can be numerically approximated, given sufficiently high numerical accuracy and resolution in parameter space.

10.5.4 Comments on approximations

Even though the model (10.11) is an approximation to (10.6), we observe that if (10.11) has a fixed point of the forcing phase at glacial termination, then solutions to (10.6) tend to have an average duration close to that of (10.11) at the fixed point (see Figure 10.4). The agreement improves with decreasing T_p relative to T_g and T_o.

The Arnold tongue diagram (Figure 10.5) demonstrates that the frequency locking characteristics of the original and approximated systems agree rather well.

The strange resetting condition in (10.6) was chosen to avoid resetting x to 0, which makes frequency locking impossible. This is a degeneracy, however; frequency locking occurs typically for models with resetting $x \to 0$, if the model as also perturbed with non-constant intrinsic growth (as is the case for P98), positive time to deglaciate (as in [Huybers, 2007]), or a rule that glaciation can only resume for a certain phase of the forcing (as in P98).

10.5.5 Relation to P98

P98 differs from the ideal model (10.5) in a number of ways. Firstly, (10.5) has only two states, while P98 has three. This makes little qualitative difference, however, since the deep glacial G state and the interglacial i states in P98 function as a delay T_{delay} of insolation-dependent length. Delays of constant length generally do not impair the frequency locking capabilities of (10.5), and neither do we expect the delays in P98 to.

Secondly, the forcing in P98 is not an ideal modulated sine, but has an additional 413 kyr modulation and other variations on top of it. We have not studied the consequences of this in detail, but we conjecture that these modulations influence the precise timing of the glacial cycles, which would otherwise be ~100 kyr long if the forcing were modulated only at 100 kyr.

Thirdly, the forcing in P98 is truncated according to (10.3). However, numerical experiments indicate that this has a minor effect on frequency locking to the forcing envelope. This is demonstrated in Figure 10.6.

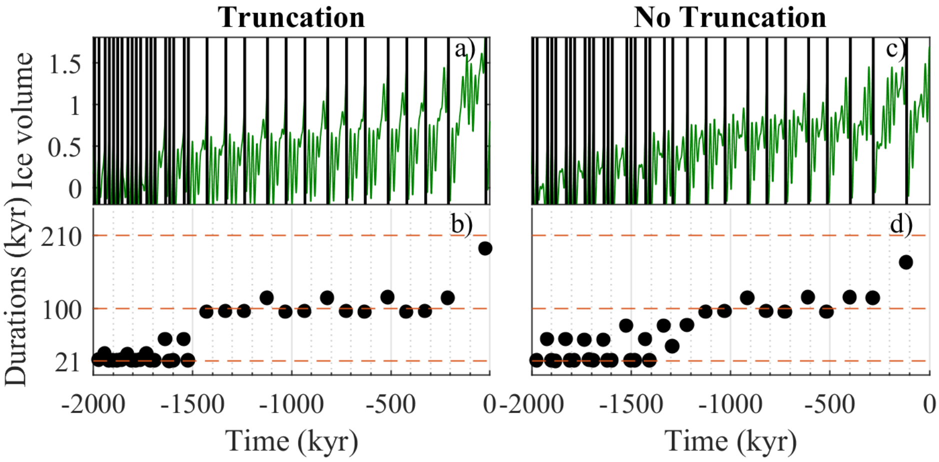


Figure 10.6: Frequency locking in a modified P98 to the 100 kyr envelope of a sum of two sines with periods 19 and 23.45 kyr. The parameter $\tau_F = 15.56$, and v_{max} increases linearly between 0.2 and 1.7. Vertical lines indicate $G \to i$ transitions. a) and b) solution and durations for P98 with modulated forcing. The $\sim$ 100 kyr durations alternate between $\sim 95 = 5 \times 19$ kyr and $\sim 115 = 5 \times 23.45$ kyr but are on average $\sim 100 = (1/19 + 1/23.45)^{-1}$ kyr long. c) and d) same as a) and b) but for non-truncated forcing

10.5.6 Summary

We have shown that a simple threshold model forced by a modulated sine can be approximated as a model that is forced periodically at the period of the envelope, assuming that the phase period T_p is much shorter than the envelope period T_g, and that the internal growth function $g(x)$ is only weakly nonlinear. This approximation tells us how a system can synchronise to the period of an envelope for a wide range of parameter values, even if that period is only a combination tone of two similar periods and therefore is not prominent in the Fourier spectrum.

The just described mechanism for frequency locking to a forcing envelope might apply to P98, even though P98 differs from the example model (10.6).

10.6 The absence of 100 kyr power prior to -1000 kyr

Having observed in Figure 10.1 that there are long durations between glacial terminations in P98 prior to -1000 kyr, we ask why these durations do not give rise to long period (> 40 kyr) wavelet amplitude.

Unfortunately, there does not seem to be a single simple explanation for this. Rather, it is the cause of multiple features of P98. We go through some of these in order: truncation of the forcing, the $g \to G \to i$ transition rule, and changes in envelope amplitude over time due to long 413 kyr modulations. We study these effects by simulating modified versions of the model, where we have changed one of these features individually. First, however, we study the distribution of durations between $G \to i$ transitions of the frozen system in Figure 10.7 in search of clues.

10.6.1 Frozen distribution of durations

The frozen system at time t_{freeze} is the system obtained by fixing $v_{max}(t)$ and the secular ramp of the forcing $F(t)$ to their values at time $t = t_{freeze}$. For each t_{freeze}, the frozen-time varying (empirical) distribution of durations is the set of durations recorded over a 2000 kyr run of the frozen system. Note that this frozen system was called *ramp only* frozen system in Chapter 9.

Figure 10.7 interestingly shows that the average duration increases gradually from ∼50 kyr at -2000 kyr to ∼105 kyr at 0 kyr without any signature of the MPT. However, the durations form major clusters with means at approximately $45, 59, 75, 94, 115, 134, 142$ kyr, so the increasing average duration is due to redistribution of durations between clusters and not because a single most typical duration increases gradually. The latter point is confirmed by that the mode of durations changes in jumps, from 45 kyr to 75 to 115, back to 75, and then finally to 94. The centres of clusters of durations are close to multiples of the main precession periods 19 and 23 kyr, e.g. $75 \approx 2 \times 19$ and $45 \approx 2 \times 23$. This is a signature of frequency locking.

Indicative of the MPT is the sudden disappearance of the ∼45 kyr cluster around -1200 kyr, and concurrent increase in the density of the 94 and 115 kyr clusters. Since the 75, and to some extent, 94 and 115 kyr clusters, are present before -2000 kyr, the increase in apparent (and wavelet) ∼100 kyr amplitude is characterised by a loss of short durations, rather than gain of long durations.

We have not been able to explain why the ∼45 kyr durations disappear abruptly at -1200 kyr. In the following, we do however attempt to explain why the long durations which occurred prior to -1000 kyr did not produce much ∼100 kyr amplitude, and why the amplitude increases abruptly around -1000 kyr.

Finally, we note that durations in the non-frozen, original, system do not have to coincide with the clusters of durations for the frozen system in Figure 10.7. For instance, in Figure 10.1, the long ∼140 and ∼100 kyr durations around -1750 kyr do not correspond to high density in the distribution of frozen durations.

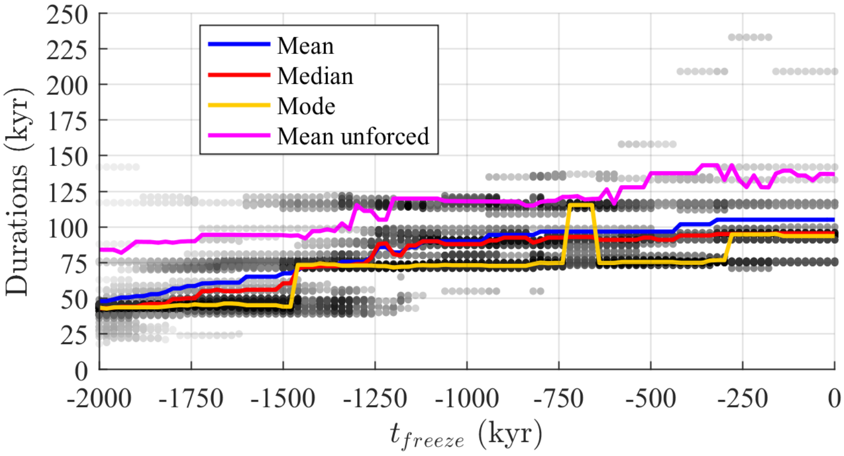


Figure 10.7: Frozen ramp distribution of durations in P98. Greyscale dots indicate sample durations from 2000 kyr long simulations with fixed v_{max} and secular ramp of the forcing $F(t)$. Solid curves show measures of centrality for the distribution of durations. The magenta curve shows the mean of unforced solutions, unforced meaning that forcing does not influence ice volume growth rates (the limit as $\tau_F \to \infty$, although forcing still applies for state switch rules)

10.6.2 Heuristic description of long durations prior to 1000 kyr

The sequence of durations between glacial terminations prior to −1000 kyr can be described heuristically in terms of the forcing and ice volume dynamics. Although the description is rather imprecise, it is useful as a basis for the discussion why they do not generate noticeable ∼100 kyr wavelet amplitude.

At time −2000 kyr the threshold of deglaciation is sufficiently low, and the insolation (Figure 10.3) is sufficiently weak, that durations could occur every precession cycle. However, the $i \to G$ resetting rule prevents this, making them roughly two precession cycles long.

Around −1750 kyr, however, the 100 kyr modulations of the precession envelope is locally strong (due to the 413 kyr period). Because of the truncation of the forcing, the net ice volume growth over one ∼21 kyr precession cycle is weakened when insolation envelope is strong, and strengthened when the insolation envelope is weak (since the average of the truncated forcing is zero). Thus, for almost 100 kyr since the glacial inception at −1873 kyr, ice volume growth is suppressed, preventing ice volume from reaching the threshold. Then, around −1800 kyr the precession envelope weakens, and ice volume grows rapidly to reach the threshold for $g \to G$ transition around −1766 kyr. The same mechanism repeats for the subsequence ∼100 kyr cycle of the precession envelope.

Between −1670 and −1510 kyr, the precession envelope weakens, and the contribution from obliquity causes locally strong ∼40 kyr oscillations in the forcing, which directly lead to ∼40 kyr glacial cycles. Then, the precession envelope becomes stronger again, leading to a ∼100 kyr cycle as described before. During the following strong 100 kyr precession envelope cycle, however, suppression of ice volume growth is not strong enough to prevent a 46 kyr long glacial termination midway through the precession envelope cycle. This effectively splits the potential ∼100 kyr cycle into one 46 and one 75 kyr cycle.

Thereafter, between ~ -1300 and ~ -1100 kyr, the precession envelope weakens and two "normal" 80 and 88 kyr cycles follow. Then, as the precession envelope strengthens, a 116 kyr long cycle occurs. However, visually this cycle looks like one 43 plus one 73 kyr cycle, because of an anomalously strong forcing at −1073 kyr. Finally, another 136 kyr cycle due to a strong precession envelope follows, terminating at -869 kyr.

To summarise, before -1000 kyr, the durations between glacial terminations are either due to ice volume growth suppression by the precession envelope when precession is strong, or a "normal" frequency locking response, as exemplified by the Huybers 2007 model, when precession and obliquity have comparable amplitude. These "normal" cycles have periods coinciding with multiples of the constituent forcing frequencies, depending on the the local v_{max}; between -1670 and -1510 cycles are $57, 39$ and 41 kyr long, and between ~ -1300 and ~ -1100 cycles are $75, 80$ and 88 kyr long. At -1364 kyr, circumstances splits the would-be $\sim$100 kyr cycle into two.

On a final note, we mention that the temporal dominance of the precession part of the forcing is reflected in the wavelet spectrum: Very regularly, 21 kyr amplitude grows whenever the precession amplitude is strong, and declines in favour of $\sim 40-80$ kyr amplitude when it is weak. Such a pattern is not discernible in the wavelet spectrum of the LR04 stack (Figure 10.2).

10.6.3 Model without truncation

Simulating the P98 model with non-truncated forcing (taking the limit as $a \to \infty$ in (10.3)), gives a similar solution as for truncated forcing, with a few exceptions (Figure 10.8).

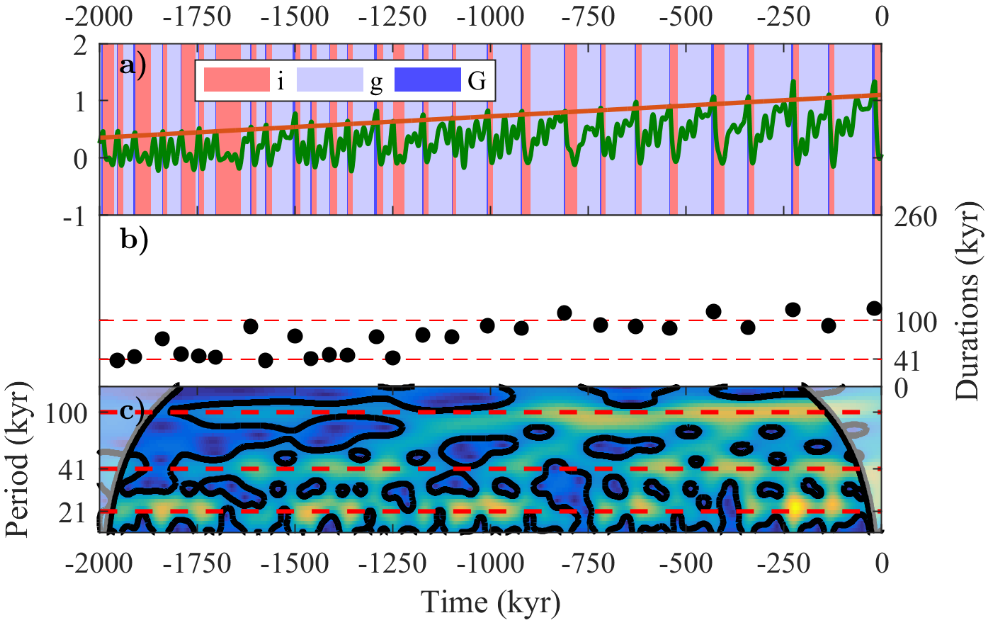


Figure 10.8: P98 without truncation of the insolation function $F(t)$. See Figure 10.1 for an explanation

Firstly, durations are more tightly concentrated around few clusters; before -1050 kyr fifteen out of sixteen durations are within 4 kyr of either 41 or 74 kyr, and all eleven durations after -1050 kyr are within 15 kyr of 100 kyr. Moreover, there are no modulations of durations due to the 413 kyr component, unlike for truncated forcing. Compared to truncated forcing the progression of durations is closer to monotonic increasing. However, the increase of $\sim$100 kyr wavelet amplitude is also more gradual compared to truncated forcing, making the MPT seem less abrupt.

The differences for truncated and non-truncated forcing reported above make sense, since the truncation of the forcing amplifies large absolute values. Hence, the long durations of first suppressed and then accelerated ice volume growth early in the Pleistocene present for truncated forcing (e.g. Figure 10.1 a) around -1750 kyr), are absent for non-truncated forcing. Furthermore, the truncated forcing enhances the contrast between forcing envelope sizes at -1000 and -600 kyr, such that durations around -1000 kyr are very long with suppressed wavelet amplitude, and durations around -600 kyr can overshoot the threshold v_{max}, increasing the perceived long period amplitude. This increased contrast in wavelet amplitude is not present for non-truncated forcing, as is seen in Figure 10.8.

Thus, we conclude that truncated forcing makes the MPT more abrupt compared to for non-truncated forcing, but that it is not necessary for reproducing the MPT.

10.6.4 Model with reversed time forcing

To test the effect of local time peculiarities in the forcing we apply forcing reversed in time, such that at -2000 kyr the system is forced with the orbital configuration at 0 kyr and vice versa. This forcing obviously has the exact same spectral power as the original forcing, and it is locally (over one precession period) approximately time reversal symmetric. Therefore, if local peculiarities were unimportant, then the time reversed and non-reversed forcings should produce similar solutions.

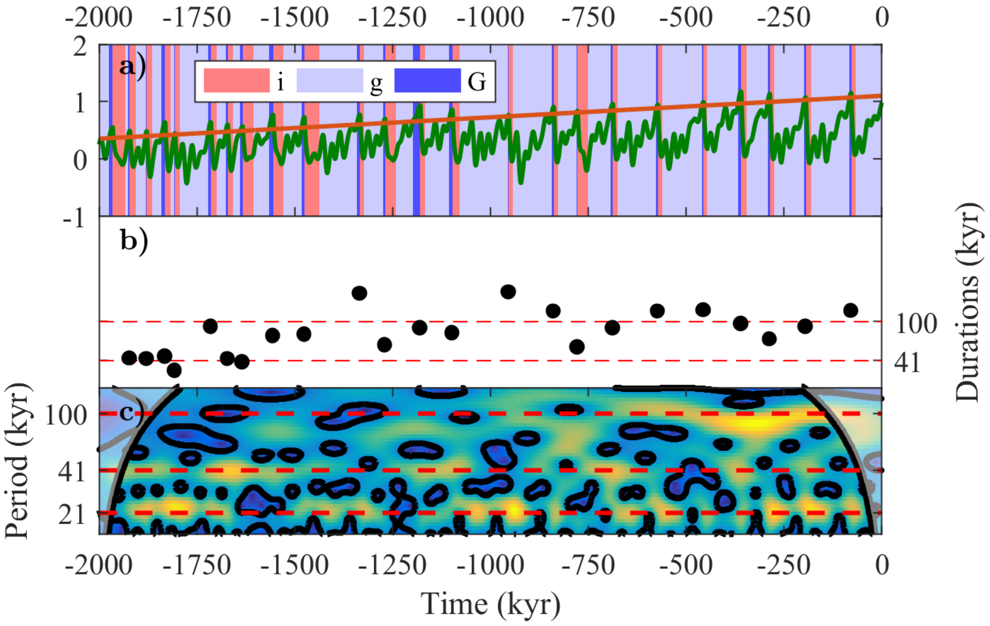


Figure 10.9: P98 with time-reversed insolation forcing function $F(t)$. See Figure 10.1 for an explanation

Figure 10.9 shows that the large precession envelope amplitude before -1000 kyr compared to the relatively weaker amplitude after -1000 kyr (for the non-time reversed forcing) plays a role for the abruptness of the MPT in P98. This is because when reversing the insolation, the weaker precession envelope before -1000 kyr allows ice sheets to grow interrupted by strong insolation, as hypothesised by [Lisiecki, 2010], thus producing long period ($64-92$ kyr) wavelet amplitude associated with long durations. Similarly, stronger insolation envelope between -1000 and 500 kyr gives weaker 100 kyr amplitude. Altogether, the stronger long period amplitude before -1000 kyr and weaker amplitude after -1000 kyr, resulting from reversing the direction of the forcing, weakens the apparent abruptness of the MPT. Thus the local forcing envelope of precession is crucial for obtaining an abrupt MPT in P98.

10.6.5 Model with time shifted forcing

Continuing to explore the role of local time variations in the forcing, we force the system with forcing shifted 500 kyr forward in time, or equivalently, we initialise the system at time -2500 kyr (Figure 10.10).

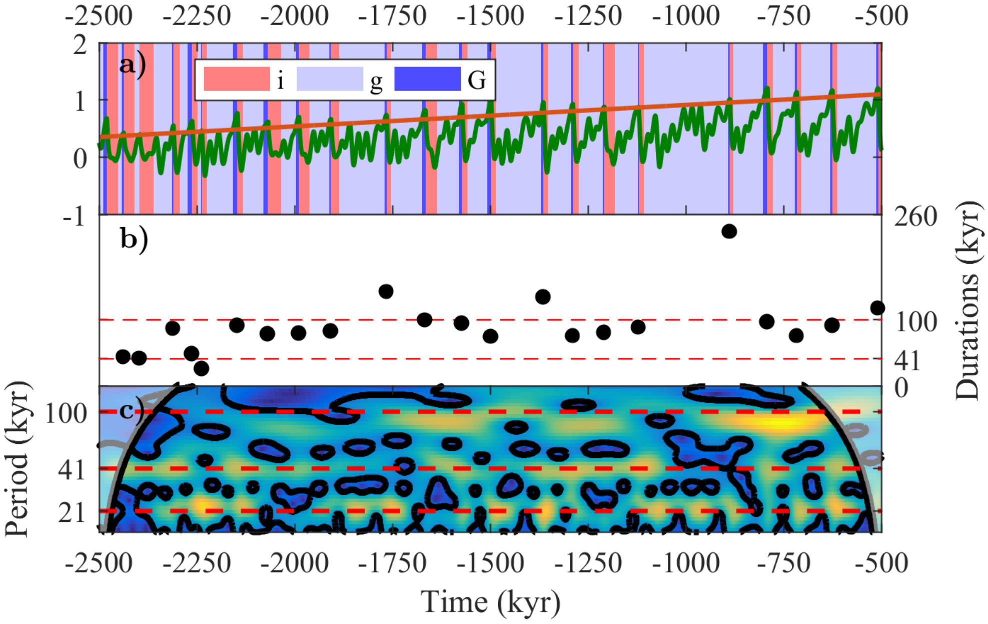


Figure 10.10: P98 with insolation forcing $F(t)$ shifted 500 kyr into the future (or equivalently, the other initial conditions are shifted 500 kyr backward). See Figure 10.1 for an explanation

Several features from the non-shifted time are familiar, for example: around -1750 kyr there are two long durations due to strong precession envelope, followed by two long uninterrupted growth cycles. Around -1000 kyr there is a very long duration, again due to the suppressing effect of strong precession envelope.

However, since the threshold for deglaciation v_{max} is larger around -1600 kyr compared to for non-shifted forcing, the durations are 94 and 74 kyr instead of $\sim$40 kyr, giving rise to more long period wavelet amplitude. This, combined with the temporary loss of 100 kyr amplitude around -1000 kyr makes the shift from short to long durations and dominant wavelet amplitude drawn out and staggered. This constitutes further evidence that the local amplitude variations of the precession envelope aids the perceived abruptness of the MPT.

10.6.6 Different transition rule

Modifying the transition rules to restart the insolation after a 2-kyr lag (by letting $g \to G$ and $G \to i$ after 1 kyr each) does not prevent a strong increase in $\sim$100 kyr amplitude around -700 kyr, with an early increase around -1250 kyr (Figure 10.11). There is comparatively less $\sim$40 kyr wavelet power before -600 kyr.

The main difference from the original transition rules appears aestethical; the "overshoots" above v_{max} present for the original system Figure 10.1 after -1000 kyr, are absent in Figure 10.11 where the modified transition rule is used. Changing the transition rule prevents the deep glacial state G in P98 from allowing "excess ice volume" to build up at times of weak precession envelope, which incorporates an idea expressed in [Raymo, 1997]. Figure 10.11 indicates that the overshoot of ice volume above v_{max} enables an increase in

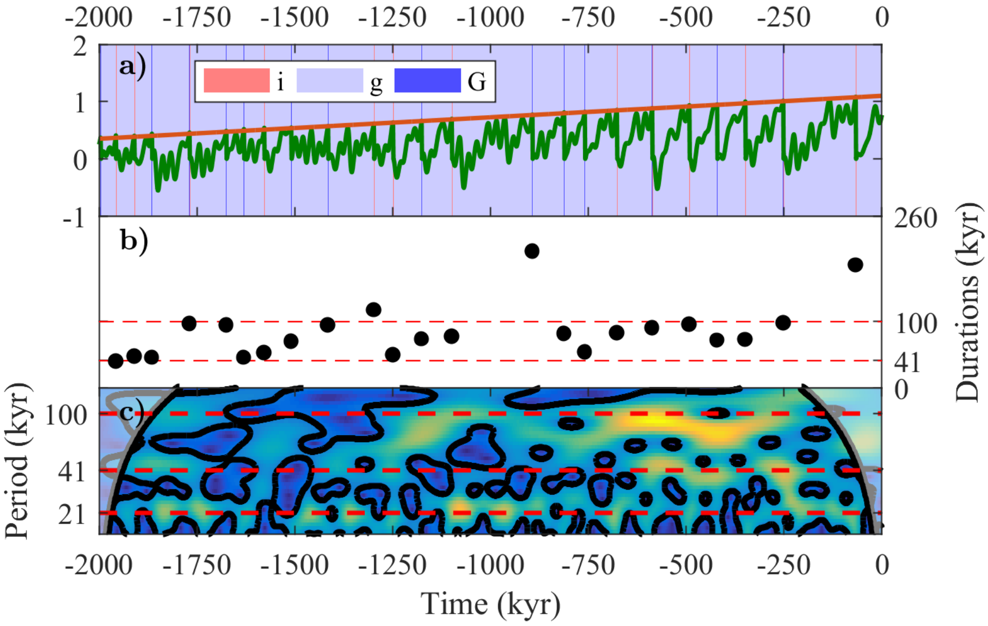


Figure 10.11: P98 with $g \to G$ and $G \to i$ transitions occuring unconditionally after one kyr. See Figure 10.1 for an explanation

maximum ice volume around -1000 kyr, but that is not crucial for initiating a shift to dominant 100 kyr amplitude.

The $g \to G \to i$ transition rule also influences the precise timing of glacial terminations, a focus in [Paillard, 1998], but this is of minor importance for explaining the dynamics behind the MPT.

10.6.7 Summary: Abrupt increase in 100 kyr durations

Based on the frozen distribution of durations in Section 10.6.1, our heuristic understanding of glacial cycles in P98 in Section 10.6.2, and our analysis of perturbed models in Section 10.6.3 to 10.6.6, we summarise the causes the abrupt increase in ~100 kyr durations around -1000 kyr observed in P98 as follows.

Firstly, there are both ~100 kyr durations and ~100 kyr wavelet amplitude prior to -1000 kyr, it is just that the amplitude is smaller compared to after -1000 kyr. The question is what causes the rather abrupt increase in amplitude.

The abrupt increase in ~100 kyr wavelet amplitude is due to multiple factors. One is the rapid decrease in precession envelope amplitude around -1000 kyr, which suppresses glacial cycles near 100 kyr, and lets them grow more freely over the remaining 1000 kyr. Thus, the locally strong precession envelope plays a major role in reproducing the MPT in P98. The rather particular $g \to G \to i$ transition rule, allowing ice volume overshoot over v_{max}, increases the amplitudes of glacial cycles after -1000 due to weakened precession envelope, but is not dynamically crucial. The forcing truncation decreases the amplitude of long duration glacial cycles at times of strong precession by suppressing ice volume growth over one ~100 kyr period of the envelope, and furthermore increases the contrasts between the local maxima of insolation during times of stronger and weaker precession envelope.

The above discussion does not address why the cluster of ~45 kyr durations in the frozen distribution disappears around -1250 kyr. This may have to do with the rather abrupt increase in internal durations at the same time, and is potentially is related to the increase in ∼100 kyr durations during the MPT. We have not explored the cause of the abrupt disappearance of the cluster of ∼45 kyr durations, but an explanation probably includes the idea that a minimum required threshold v_{max} is necessary for short durations to be possible.

10.7 Conclusions

We have explained the dynamical mechanism of the abrupt increase in wavelet amplitude around −1000 kyr in the Paillard 1998 model (P98), one of the first models to provide an explanation for the middle Pleistocene transition. The mechanism is essentially due to ramping of the internal period and frequency locking (Chapter 7), but two specific questions require answers: 1. why are durations between glacial terminations and wavelet amplitude ∼100 kyr after −1000 kyr, and 2. why is there an abrupt increase in wavelet amplitude around −1000 kyr?

We answer the first question by constructing a transparent model reminiscent of P98, forced by a modulated sine function. This model shows clearly how an integrate-and-fire type self-sustained oscillator can frequency lock to the period of the envelope of the forcing, in this case the ∼100 kyr modulations of the precession by eccentricity. This idea has been expressed earlier (e.g. [Lisiecki, 2010]), but to our knowledge no explicit model demonstrating the mechanism has been shown previously.

The answer to the second question is more subtle. By estimating the distribution of durations of the *frozen system* (which characterises the model under quasistatic parameter ramping), we find that the average duration between glacial terminations increases gradually, whereas the mode (most common) duration increases rather rapidly around −1250 kyr, concurrent with the loss of short ∼40 kyr durations. Thus the apparent shift in glacial cycle period can in part be understood as a rapid change in the most common duration, which due to a limited number of durations in a given time window can give the impression of a rapid increase in average duration.

This explanation is insufficient, however. By running perturbed versions of P98, we find that an anomalously strong precession envelope around −1000 kyr, combined with an anomalously weak precession envelope around −600 kyr is crucial for the abrupt increase in wavelet amplitude at the MPT. As such, the MPT in P98 is dependent on the particular time evolution of the forcing amplitude, and can not be explained purely by periodic frequency locking. Furthermore, rather particular model assumptions of truncated forcing and rules for switching between climate states makes the MPT appear even more abrupt. This casts doubt on the generality of the model; dynamically similar but non-identical models forced by similar forcing might have a hard time reproducing the abruptness of the MPT seen in P98.

We reiterate this last point. It is interesting to know which dynamical mechanisms are likely to be present in extensive climate models which contain few explcit assumptions to reproduce the MPT.

Given that the Paillard 1998 describes the glacial system adequately, some interesting implications follow. One is that not all ∼40 kyr glacial cycles in the early Pleistocene are proper transitions between interglacial, glacial, and deep glacial cliamte states, but may arise from direct forcing within the glacial climate state. It might be possible to evaluate whether this is the case in the climate system, either by statistical analysis or by seeking physical signatures of “proper” climatic state shifts in proxy records.

Finally, we note that the MPT can be replicated in P98 under obliquity-heavy forcing (Takahito Mitsui, personal communication), but in this case the transition is dynamically explained by the typical ramping with frequency locking to an almost periodic signal (Chapter 7).

10.8 Auxiliary results

10.8.1 Derivation for more general oscillators

Returning to (10.5), we show that the derivations for the example in the previous subsection extend to more general internal growth functions $g(t)$ and forcing functions $f(t)$. We begin by showing that forcing does not affect the mean growth, under certain assumptions.

Consider the system at a time t_1 and position x_1 far from the threshold x_c. By expanding $g(t)$ to first order we can show (Section 10.8.2) that the local discrepancy between forced and non-forced systems scales quadratically in the inverse time scale $\beta := g'(x(t_1)) := \tau^{-1}$ (quadratic Taylor coefficient) and in the carrier wave period T_p. Neglecting the net influence from forcing, assuming β and T_p small, allows us to solve (10.5) between transitions, starting at $t_0, x(t_0) = 0$ as

$$x(t) = G(x(t_0)) - G(x(t_0)) + \int_{t_0}^{t} f(s)ds. \tag{10.15}$$

Knowing the internal ice volume evolution, we only need to know how the forcing influences extremes in ice volume. This information we get from the primitive function $F(t) = \int_{t_0}^{t} f(s)ds + F(t_0)$. Assuming that $k(t)$ changes much more rapidly than $h(t)$, which is natural if $T_p \ll T_g$, then the extremes of $F(t)$ are proportional to $h(t)K(t)$, with $K(t)$ the zero-mean primitive function of $k(t)$. This is because if $k(t)$ and $h(t)$ depend on T_p and T_g as $k(t) = k_0(\frac{t}{T_p})$ and $h(t) = h_0(\frac{t}{T_g})$ only, then the time derivative of

$$\tilde{F}(t) = T_p h(t)K(t), \tag{10.16}$$

namely

$$\tilde{F}'(t) = h'(t)K(t) + h(t)K'(t) = c_1 \frac{T_p h_0'\left(\frac{t}{T_g}\right)}{T_g} K(t) + c_2 h(t) \frac{T_p k_0\left(\frac{t}{T_p}\right)}{T_p}, \tag{10.17}$$

is dominated by the second term as $T_p \ll T_g$. Here, c_1 and c_2 are constants. Hence, $\tilde{F}(t) = T_p h(t)K(t)$ is an approximate primitive function of $f(t)$, which converges to the true primitive function of $f(t)$ in the limit as $\frac{T_p}{T_g} \to 0$. We define for convenience $E(t) = c_1 T_p h(t)$ as the (approximate) envelope of the integrated insolation.

Therefore, defining $x_{int}(t)$ to be the solution of the internal ice volume evolution

$$x_{int}(t) = G(x(t_0)) - G(x(t_0)), \tag{10.18}$$

the threshold is hit approximately at the smallest time t^* for which $x_{int}(t^*) + F(t^*) = x_c$. If $K(t)$ oscillates rapidly $T_p \ll T_g$, then the threshold is reached shortly after

$$x_{int}(t^*) + \frac{c_1}{T_p} h(t^*) = x_c, \tag{10.19}$$

since $F(t) = \frac{1}{T_p} h(t)K(t)$ and $K(t)$ is defined to have maximum amplitude 1 (true for a pure sine). Compare with the example in (10.11).

Equation (10.19) together with the internal evolution equation $x_{int}(t) = G(x(t)) - G(x(t_0))$ defines an oscillator forced at the period at the envelope, making it amenable to frequency locking.

10.8.2 Net ice volume growth for nonconstant internal growth rates

Expand $g(x(t))$ in a Taylor series around a point (t_0, x_0):

$$g(x(t)) = g(x_0) + g'(x_0)(x(t) - x_0) + \text{h.o.t.}, \tag{10.20}$$

where h.o.t. stands for higher order terms and prime denotes derivative with respect to x. If $g(x)$ is slowly varying, the higher order terms are small and the two-term truncation $g(x_0) + g'(x_0)(x - x_0)$ works well for a wide range of x. We note in particular that the $g(x)$ of Huybers 2007, Ashkenazy 06 and and Paillard 98 models have vanishing derivatives of order two and higher [Huybers, 2007, Ashkenazy, 2006, Paillard, 1998].

Thus, we can estimate the ice volume growth locally around (t_0, x_0). In general, no explicit solution exists for arbitrary $f(t)$, but if $k(t)$ is a sine, we can solve (10.5) locally as

$$\dot{x}(t) = g(x_0) + g'(x_0)(x - x_0) - h(t_0)\sin\left(\frac{2\pi t}{T_p}\right) \tag{10.21}$$

This differential equation, with initial condition (t_0, x_0) has the solution

$$\begin{aligned} x(t) &= x_0 e^{\beta(t-t_0)} + \frac{\alpha}{\beta}(e^{\beta(t-t_0)} - 1) + \\ &\frac{h(t_0)}{\omega_p^2 + \beta^2}[\omega_p \cos(\omega_p t) + \beta \sin(\omega_p t) - e^{\beta(t-t_0)}(\omega_p \cos(\omega_p t_0) + \beta \sin(\omega_p t_0))], \end{aligned} \tag{10.22}$$

where $\alpha = g(x_0) - g'(x_0)x_1$, $\beta = g'(x_0)$ and $\omega_p = \frac{2\pi}{T_p}$. Taking the difference between this solution and the unforced solution $x_{unf}(t)$ started at $x_{unf,0} = x_0 + F(t_0) \approx x_0 - \frac{h(t_0)}{\omega_p}\cos(\omega_p t_0)$

$$x_{unf}(t) = x_{unf,0} e^{\beta(t-t_0)} + \frac{\alpha}{\beta}(e^{\beta(t-t_0)} - 1), \tag{10.23}$$

gives the discrepancy

$$\begin{aligned} \Delta x(t) &= e^{\beta(t-t_0)}(x_0 - x_{unf,0}) + \frac{h(t_0)}{\omega_p^2 + \beta^2}[\omega_p \cos(\omega_p t) + \beta \sin(\omega_p t) - \\ &e^{\beta(t-t_0)}(\omega_p \cos(\omega_p t_0) + \beta \sin(\omega_p t_0))] \\ &= e^{\beta(t-t_0)} h(t_0) \cos(\omega_p t_0)\left(\frac{1}{\omega_p} - \frac{\omega_p}{\omega_p^2 + \beta^2}\right) - e^{\beta(t-t_0)} \frac{\beta}{\omega_p^2 + \beta^2}\sin(\omega_p t_0) + \\ &\frac{h(t_0)}{\omega_p^2 + \beta^2}[\omega_p \cos(\omega_p t) + \beta \sin(\omega_p t)] \\ &= e^{\beta(t-t_0)} h(t_0)\left(\frac{\beta^2}{\omega_p(\omega_p^2 + \beta^2)}\cos(\omega_p t_0) - \frac{\beta}{\omega_p^2 + \beta^2}\sin(\omega_p t_0)\right) + \\ &\frac{h(t_0)}{\omega_p^2 + \beta^2}[\omega_p \cos(\omega_p t) + \beta \sin(\omega_p t)] \end{aligned} \tag{10.24}$$

By assuming that $\beta \gg T_p \Rightarrow \beta \ll \omega_p$, the discrepancy can be approximated as

$$\Delta x(t) \approx \frac{h(t_0)}{\omega_p}\cos(\omega_p t) = -F(t), \tag{10.25}$$

such that $x_{unf}(t) + F(t)$ approximates the solution $x(t)$ well. In particular, at times $t^* = \frac{\pi}{2\omega_p} + NT_f, N \in \mathbb{N}$, $\Delta x(t^*) \approx 0$, implying that the forced and unforced solutions approximately co-incide periodically. Note that the initial value $x_{unf,0}$ of the internal, unforced, solution depends on the initial phase $\omega_p t_0$ and only (approximately) coincides with x_0 only when the carrier phase is $\pm\pi$.

In case the initial condition $x_{unf,0}$ deviates from the one above, the difference between solutions still decays over time. Hence, errors between solutions are reduced, potentially nullifying discrepancies between solutions.

Note, however, that starting the internal, unforced oscillation at a different value than 0 effectively changes the internal period.

10.8.3 Beat frequency from sum of sines

$$
\begin{aligned}
f(x,y) = \sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right) &= (\sin(x/2)\cos(y/2) + \sin(y/2)\cos(x/2))\times \\
&\quad (\sin(x/2)\cos(y/2) - \sin(y/2)\cos(x/2)) \\
&= \sin^2(x/2)\cos^2(y/2) - \sin^2(y/2)\cos^2(x/2) \\
&= \sin^2(x/2)\cos^2(y/2) - \sin^2(y/2)(1-\sin^2(x/2)) \\
&= \sin^2(x/2) - \sin^2(y/2) \\
&= \frac{(1-\cos(x))}{2} - \frac{(1-\cos(y))}{2} \\
&= \frac{1}{2}(\cos(x) - \cos(y)).
\end{aligned}
\tag{10.26}
$$

Letting $\alpha = (x+y)/2, \beta = (x-y)/2$ gives that

$$
\sin(\alpha)\sin(\beta) = \frac{1}{2}(\cos(\alpha-\beta) - \cos(\alpha+\beta)) \tag{10.27}
$$

10.8.4 Primitive function of product of sines

The zero-mean primitive function of $g(t) = \sin(\alpha t)\sin(\beta t)$ is

$$
\begin{aligned}
F(t) &= \frac{1}{2}\left(\frac{1}{(\alpha-\beta)}\sin(\alpha-\beta)t - \frac{1}{(\alpha+\beta)}\sin(\alpha+\beta)t\right) \\
&= \frac{1}{2(\alpha^2-\beta^2)}((\alpha+\beta)\sin(\alpha-\beta)t - (\alpha-\beta)\sin(\alpha+\beta)t) \\
&= \frac{1}{2(\alpha^2-\beta^2)}(\alpha(\sin(\alpha-\beta)t - \sin(\alpha+\beta)t) + \beta(\sin(\alpha-\beta)t + \sin(\alpha+\beta)t) \\
&= \frac{1}{2\left(\alpha^2-\beta^2\right)}\Big(\alpha\left(\cos\left(\alpha+\frac{\pi}{2}-\beta\right)t - \cos\left(\alpha+\frac{\pi}{2}+\beta\right)t\right) + \\
&\quad \beta\left(\cos\left(\alpha-\left(\beta-\frac{\pi}{2}\right)\right)t - \cos\left(\alpha+\left(\beta-\frac{\pi}{2}\right)\right)t\right)\Big) \\
&= \frac{1}{2(\alpha^2-\beta^2)}\left(2\alpha\sin\left(\alpha+\frac{\pi}{2}\right)t\sin(\beta t) + 2\beta\sin(\alpha t)\sin\left(\beta-\frac{\pi}{2}\right)t\right) \\
&= \frac{1}{\alpha^2-\beta^2}(-\alpha\cos(\alpha t)\sin(\beta t) + \beta\sin(\alpha t)\cos(\beta t))
\end{aligned}
\tag{10.28}
$$

This is the sum of two anti-phase beat waves. If $\alpha \gg \beta$, then the first term dominates and

$$
F(t) \approx \frac{1}{\alpha^2-\beta^2}(-\alpha\cos(\alpha t)\sin(\beta t)) \approx -\frac{1}{\alpha}\cos(\alpha t)\sin(\beta t)
$$

Chapter 11

Summary and conclusions

The cause of the abrupt change in amplitude and period of glacial cycles at the middle Pleistocene transition (MPT) approximately 1 Million years ago remains one of the great unsolved problems of paleoclimate. In this thesis we have advanced the understanding of the possible *dynamical* causes behind the MPT, that is, we elucidate how a transition such as the MPT can occur in mathematically formulated models, regardless of the underlying physics.

We have approached the problem from two directions. The first is to show how an abrupt change in system period can arise from frequency locking of internal glacial cycles to insolation variations. One general mechanism for this, involving a slow trend of some climate variable, we defined and termed *ramping with frequency locking* in Chapter 7. In Chapter 9, we further explored how variations in the intensity and local period of forcing influences which durations between glacial terminations ("periods" of glacial cycles) are observed at what times. On the same theme, we found in Chapter 10 that the Paillard 1998 model of glacial cycles reproduces the MPT through ramping and frequency locking to the eccentricity envelope, but that the apparent abrupt increase in glacial cycle period is due to rather specific model assumptions.

A second direction was taken In Chapter 8, where we classified all the typical ways in which relaxations oscillations can bifurcate in certain systems where one fast model variable changes much faster than another slow one. This has an immediate application to glacial cycles, where ice volume growth and decay can take the role of a slow variable and switches between distinct climate states (glaciating/deglaciating) can take the role of fast variable. Our results imply that if the glacial cycles are effectively described by one fast and one slow variable, subject to a typical slow parameter perturbation, then there is a limited number of ways that an abrupt shift in period and/or amplitude of self-sustained limit cycles can occur, which we enumerated.

Furthermore, we showed in Section 4.7 how the relative rates of glaciation and deglaciation in a periodically forced integrate-and-fire oscillator model influences the width of Arnold tongues, or frequency locking regions. Such knowledge is important for the hypothesis that the MPT is due frequency locking to forcing and a ramped parameter, since it tells us the relative time that the system spends in different Arnold tongues, assuming a certain rate of increase of an external parameter influencing the internal period.

Our main contributions can be summarised as follows. Firstly, we clearly defined the mechanism of *ramping with frequency locking* (RFL) as a cause of the MPT. This dynamical mechanism is used in multiple models of glacial cycles, but the dynamical commonality among models has not been appreciated in the literature. Considering the large number of conceptual glacial cycle models with the potential to reproduce the MPT, it is valuable to classify and organise different ways in which a shift in amplitude and period can arise dynamically. Such knowledge can hopefully be used to understand mechanisms for the MPT in more comprehensive Earth system models.

Secondly, we observed that the notion of a stationary ~100 kyr world after the MPT may be misleading, since the durations between major glacial terminations are shown to increase over time from about 80 kyr to about 120 kyr from −1250 kyr until the present. Such a progression of durations, supported by wavelet spectra, is consistent with a gradually increasing intrinsic period of glacial cycles, either through *ramping with frequency locking* or another mechanism like that of [Ashwin and Ditlevsen, 2015]. On the other hand, an increase in duration does not come as naturally for some mechanisms of glacial cycles.

Thirdly, we found a significant association between eccentricity and ice volume data stretching back to −1200 kyr, through wavelet coherence. This association at periods ~120 kyr at the MPT stand in contrast to the ~80 kyr durations between glacial termination at that time. None out of five tested conceptual models reproduce both coherence with eccentricity and a sequence of glacial terminations consistent with data. It thus remains a challenge to understand the cause of these associations. Exploration of how the conflicting evidence can be reconciled can start with conceptual models.

Fourthly, we classified bifurcations in certain fast-slow systems with one fast and one slow variable up to codimension one. This, we have not seen done before. Although this classification helps organising dynamical ways through which the MPT can occur, we expect the classification to be useful for problems other than glacial cycles in addition to having intrinsic mathematical interest.

Fifthly, we added to the understanding of how modulated sinusoidal forcing influences the dynamical behaviour of an integrate-and-fire model of glacial cycles proposed by Huybers [2007]. We found that the clustering of durations between glacial terminations in *frozen* systems in which either system parameters and/or forcing amplitude and period is fixed, is in a sense predictive of which durations are observed in transient runs of the original, non-frozen system.

Sixthly, we constructed an explicit model showing how integrate-and-fire glacial cycle models can synchronise to the eccentricity envelope, as is likely the case in the Paillard 1998 model. We further found how, unexpectedly, the model could go from showing negligible 100 kyr wavelet amplitude before ∼ − 1000 kyr to strong amplitude thereafter. This abrupt increase can to a great part be attributed to the local change in the strength of eccentricity around −1000 kyr, which together with some model assumption diminished ~100 kyr amplitude before the MPT, and amplifies it afterwards. This implies that if other models are to reproduce the MPT like Paillard 1998 did, they likely require rather specific assumptions.

11.1 Future perspectives

There are plenty of directions for future work. Relating to Chapter 7, an explicit challenge for modellers is to reconcile the association between eccentricity and ice volume proxies with the durations across the MPT. Another open question is if RFL arises naturally in models higher up in the model complexity hierarchy, such as Earth system models of intermediate complexity and comprehensive Earth system models. Ganopolski and Calov [2011] found a negative answer in a model of intermediate complexity, since glacial cycles were not self-sustained in that model. However, as mentioned in Chapter 7, the system might still be excitable so that a mechanism analogous to RFL still applies. Furthermore, it would be interesting to learn if other models of intermediate complexity or even comprehensive Earth system models built on different assumptions share the conclusions of [Ganopolski and Calov, 2011, Ganopolski and Brovkin, 2017, Willeit et al., 2019].

The are many ways of extending the work in Chapter 8. An immediate one is to classify bifurcations in singular systems up to codimension higher than one. This includes cases when the conditions for two codimension one bifurcations occur simultaneously. Another direction is to classify bifurcations of the critical set for one fast and two variables; we presented a conjectured such classification in Chapter 8. Extending this to a global singular classification for fast-slow systems is likely a major challenge. Yet another possibility is extension to multiple fast variables. This likely requires a completely new equivalence since the asymptotic behaviour of fast subsystem is not limited to equilibria. Moving the focus to applications, one can instead study specific examples of bifurcations of limit cycles for more fast and slow variables.

We believe that more can be learned regarding the finite time response of oscillators to modulated forcing. The methods presented in Chapter 9 should be tested on a range of systems, such as oscillators without explicit thresholds. If they work better for some systems than others, perhaps it can be understood when the methods work better or worse. Furthermore, other measures of how much the frozen distribution says about sequences of durations in the non-frozen system should be considered. For instance, linear correlation could probably well capture the meandering of durations due to variations of the eccentricity envelope.

The range of dynamical mechanisms by which the MPT can occur has likely not been completely charted. It might be possible to amend the mechanism for the MPT in Paillard 1998, in which the climate system frequency locks to the precession envelope. Such improved models should requires less stringent assumptions and reproduce better the glacial cycles prior to the MPT. It is also interesting to explore the role of the local precession envelope around -1000 kyr in other models, by running them with reversed or time-shifted forcing.

11.2 Final comments

The dynamical and physical causes of the MPT remain unknown at the present year 2019. We have advanced the knowledge about possible dynamical mechanisms, but we cannot conclude which one is more correct than another. To make progress on the MPT, we believe that a combination of approaches is needed; to advance knowledge about past climate states through new and improved proxies as in [Farmer et al., 2019], to run increasingly more complex models for a longer time as in [Willeit et al., 2019], and to seek to understand the dynamics of the transition in these complex models by constructing and analysing transparent and easily understood conceptual models. At the same time, we must be humble about the limits of both proxies and models. A proxy can only be related to a physical quantity through strong assumptions and imperfect modelling, and models can easily seem correct for the wrong reasons. These problems are exacerbated by the eternal bane of the paleosciences: that we cannot run controlled and repeated experiments on the past. If the truth about the MPT will ever be known with high certainty, we are likely far from that point.

Bibliography

Ayako Abe-Ouchi, Fuyuki Saito, Kenji Kawamura, Maureen E.Raymo, Jun'ichi Okuno, Kunio Takahashi, and Heinz Blatter. Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume. *Nature*, 500:190, Aug 2013. doi: 10.1038/nature12374.

Preben Alstrøm, Bo Christiansen, and Mogens T. Levinsen. Characterization of a simple class of modulated relaxation oscillators. *Phys. Rev. B*, 41:1308–1319, Jan 1990. doi: 10.1103/PhysRevB.41.1308.

Aleksandr A. Andronov and Lev S. Pontryagin. Coarse systems. *Doklady Akademii Nauk SSSR*, 14(5):247–250, 1937.

José A.Rial, Jeseung Oh, and Elizabeth Reischmann. Synchronization of the climate system to eccentricity forcing and the 100,000-year problem. *Nature Geoscience*, 6:289, Mar 2013. doi: 10.1038/ngeo1756.

V. I. Arnold. Small denominators. I. mapping the circle onto itself. *Izv. Akad. Nauk SSSR Ser. Mat.*, 25(1):21–86, 1961.

V. I. Arnold, V. S. Afrajmovich, Yu. S. Ilyashenko, and L. P. Shilnikov. *Bifurcation theory and catastrophe theory*. Springer-Verlag, Berlin, 1999. ISBN 3-540-65379-1. Translated from the 1986 Russian original by N. D. Kazarinoff, Reprint of the 1994 English edition from the series Encyclopaedia of Mathematical Sciences [Dynamical systems. V, Encyclopaedia Math. Sci., 5, Springer, Berlin, 1994].

Yosef Ashkenazy. The role of phase locking in a simple model for glacial dynamics. *Climate Dynamics*, 27:421–431, 2006. doi: 10.1007/s00382-006-0145-5.

Yosef Ashkenazy and Eli Tziperman. Are the 41 kyr glacial oscillations a linear response to Milankovich forcing? *Quaternary Science Reviews*, 23:1879–1890, 2004. doi: 10.1016/j.quascirev.2004.04.008.

Peter Ashwin and Peter Ditlevsen. The middle Pleistocene transition as a generic bifurcation on a slow manifold. *Climate Dynamics*, 24(9-10), 2015. doi: 10.1007/s00382-015-2501-9.

Peter Ashwin, Charles David Camp, and Anna 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*, 1(20), 2018. doi: 10.1093/climsys/dzy002.

Eric Benoît, Jean Louis Callot, Francine Diener, and Marc Diener. Chasse au canard. *Collectanea Mathematica*, 32:37–74, 1981.

Andre L Berger. Long-Term Variations of Daily Insolation and Quaternary Climatic Changes. *Journal of the Atmospheric Sciences*, 35, 1978. doi: 10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2.

André Berger, Marie-France Loutre, and Qiuzhen Yin. Total irradiation during any time interval of the year using elliptic integrals. *Quaternary Science Reviews*, 29(17):1968 – 1982, 2010. doi: 10.1016/j.quascirev.2010.05.007.

Richard Bintanja, Roderik S. W. van de Wal, and Johannes Oerlemans. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature*, 437: 125–128, Sep 2005. doi: 10.1038/nature03975.

B. Boashash. Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proceedings of the IEEE*, 80(4):520–538, April 1992. doi: 10.1109/5.135376.

Wallace S. Broecker and Jan van Donk. Insolation changes, ice volumes, and the δ^{18}O record in deep-sea cores. *Reviews of Geophysics*, 8(1):169–198, 1970. doi: 10.1029/RG008i001p00169.

H.W. Broer and T.J. Kaper M. J. Krupa. Chasse au canard. *Journal of Dynamics and Differential Equations*, 25(925):37–74, 2013. doi: 10.1007/s10884-013-9322-5.

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

Jake Carson, Michel Crucifix, Simon Preston, and Richard D. Wilkinson. Bayesian model selection for the glacial–interglacial cycle. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(1):25–54, 2018. doi: 10.1111/rssc.12222.

M.L. Cartwright and J.E. Littlewood. On non-linear differential equations of the second order. *Journal of the London Mathematical Society*, 1-20(3):180–189, 1945. doi: 10.1112/jlms/s1-20.3.180.

Peter U Clark and David Pollard. Origin of the middle Pleistocene transition by ice sheet erosion of regolith. *Paleoceanography*, 13(1):1–9, 1998. doi: 10.1029/97PA02660.

Leon Cohen, Patrick Loughlin, and David Vakman. On an ambiguity in the definition of the amplitude and phase of a signal. *Signal Processing*, 79(3):301–307, dec 1999. ISSN 0165-1684. doi: 10.1016/S0165-1684(99)00103-6.

Michel Crucifix. Oscillators and relaxation phenomena in Pleistocene climate theory. *Philosophical transactions of the Royal Society of London A*, 370(1962):1140–1165, 2012. doi: 10.1098/rsta.2011.0315.

Michel Crucifix. Why could ice ages be unpredictable? *Climate of the Past*, 9:2253–2267, 2013. doi: 10.5194/cp-9-2253-2013.

Harry Dankowicz and Frank Schilder. An extended continuation problem for bifurcation analysis in the presence of constraints, 2009.

Istvan Daruka and Peter Dalager Ditlevsen. A conceptual model for glacial cycles and the middle Pleistocene transition. *Climate Dynamics*, 46:29–40, 2015. doi: 10.1007/s00382-015-2564-7.

H.G. Davies and D. Nandlall. The response of a van der Pol oscillator to a modulated amplitude sinusoidal input. *Journal of Sound and Vibration*, 140(3):499 – 512, 1990. doi: 10.1016/0022-460X(90)90763-P.

Bernard de Saedeleer, Michel Crucifix, and Sebastian Wieczorek. Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual study. *Climate Dynamics*, 40: 273–294, 2013. doi: 10.1007/s00382-012-1316-1.

Mario di Bernardo, Chris J. Budd, Alan R. Champneys, Piotr Kowalczyk, Arne Nordmark, Gerard Olivar Tost, and Petri T. Piiroinen. Bifurcations in nonsmooth dynamical systems. *SIAM Review*, 50(4):629–701, 2008. doi: 10.1137/050625060.

Henk A. Dijkstra. *Nonlinear Climate Dynamics*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139034135.

Peter D. Ditlevsen and Peter Ashwin. Complex Climate Response to Astronomical Forcing: The Middle-Pleistocene Transition in Glacial Cycles and Changes in Frequency Locking. *Frontiers in Physics*, 6:62, 2018. doi: 10.3389/fphy.2018.00062.

Peter Dalager Ditlevsen. Bifurcation structure and noise-assisted transitions in the Pleistocene glacial cycles. *Paleoceanography*, 24, 2009. doi: 10.1029/2008PA001673.

Eusebius J. Doedel. Nonlinear numerics. *Journal of the Franklin Institute*, 334(5):1049 – 1073, 1997. doi: 10.1016/S0016-0032(97)00027-6.

Cesare Emiliani. Pleistocene temperatures. *The Journal of Geology*, 63(6):538–578, 1955. doi: 10.1086/626295.

EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature*, 429: 623–628, 2004. doi: 10.1038/nature02599.

J. R. Farmer, B. Hönisch, L. L. Haynes, D. Kroon, S. Jung, H. L. Ford, M. E. Raymo, M. Jaume-Seguí, D. B. Bell, S. L. Goldstein, L. D. Pena, M. Yehudai, and J. Kim. Deep Atlantic Ocean carbon storage and the rise of 100,000-year glacial cycles. *Nature Geoscience*, 2019. doi: 10.1038/s41561-019-0334-6.

Fabo Feng and C A L Bailer-Jones. Obliquity and precession as pacemakers of Pleistocene deglaciations. *Quaternary Science Reviews*, 122:166–179, 2015. doi: 10.1016/j.quascirev.2015.05.006.

Neil Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of Differential Equations*, 31:53–98, 1979. doi: 10.1016/0022-0396(79)90152-9.

Ulrike Feudel, Celso Grebogi, and Edward Ott. Phase locking in quasiperiodically forced systems. *Physics Reports*, 290:11–25, 1997. doi: 10.1016/S0370-1573(97)00055-0.

Marianne Freiberger. Information is surprise. `https://plus.maths.org/content/information-surprise`, 2015. Accessed: 11 March 2019.

A. Ganopolski and V. Brovkin. Simulation of climate, ice sheets and CO_2 evolution during the last four glacial cycles with an Earth system model of intermediate complexity. *Climate of the Past*, 13(12):1695–1716, 2017. doi: 10.5194/cp-13-1695-2017.

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. doi: 10.5194/cp-7-1415-2011.

David García-Álvarez, Aneta Stefanovska, and Peter V. E. McClintock. High-order synchronization, transitions, and competition among Arnold tongues in a rotator under harmonic forcing. *Phys. Rev. E*, 77, 2008. doi: 10.1103/PhysRevE.77.056203.

George T. Gilbert. Positive definite matrices and Sylvester's Criterion. *The American Mathematical Monthly*, 98(1):44–46, 1991. doi: 10.2307/2324036.

Hezi Gildor and Eli Tziperman. Sea ice as the glacial cycles' climate switch: role of seasonal and orbital forcing. *Paleoceanography*, 15(6):605–615, 2000. doi: doi.org/10.1029/1999PA000461.

Leon Glass and Michael C. Mackey. A simple model for phase locking of biological oscillators. *Journal of Mathematical Biology*, 7:339–352, 1979. doi: 10.1007/BF00275153.

Martin Golubitsky and David Schaeffer. *Singularities and Groups in Bifurcation Theory, Volume I*, volume 51 of *Applied Math Sci.* Springer, 1985.

Celso Grebogi, Edward Ott, Steven Pelican, and James A. Yorke. Strange attractors that are not chaotic. *Physica D*, 13:261–268, 1984. doi: 10.1016/0167-2789(84)90282-3.

D M. Grobman. Homeomorphisms of systems of differential equations. *Doklady Acad Nauk SSR*, 128:880–881, 1959. doi: 10.1177/0959683610394885.

John Guckenheimer. Towards a global theory of singularly perturbed dynamical systems. In *Progress in Nonlinear Differential Equations and Their Applications*, volume 19. Birkhauser, 1996.

John Guckenheimer. Nonlinear dynamics and chaos: Where do we go from here? In *Progress in Nonlinear Differential Equations and Their Applications.* IoP Publishing, 2002. ISBN 9780750308625.

John Guckenheimer. Bifurcations of relaxation oscillations. In *Bifurcations, Normal Forms and Finiteness Problems in Differential Equations*, volume 137 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 295–316. Springer, 2004.

John Guckenheimer and Stewart Johnson. Planar hybrid systems. In Antsaklis P., Kohn W., Nerode A., and Sastry S., editors, *Hybrid systems, II (Ithaca, NY, 1994)*, volume 999 of *Lecture Notes in Comput. Sci.*, pages 202–225. Springer, Berlin, 1995. doi: 10.1007/3-540-60472-3_11.

John Guckenheimer, Kathleen Hoffman, and Warren Weckesser. The Forced van der Pol Equation I: The Slow Flow and Its Bifurcations. *SIAM Journal on Applied Dynamical Systems*, 2(1):1–35, 2003. doi: 10.1137/S1111111102404738.

Radu Haiduc. Horseshoes in the forced van der Pol system. *Nonlinearity*, 22(1):213–237, dec 2008. doi: 10.1088/0951-7715/22/1/011.

Philip Hartman. A lemma in the theory of structural stability of differential equations. *Proc. Amer. Math. Soc.*, 11:610–620, 1960. doi: 10.2307/2034720.

E. Harvey, V. Kirk, M. Wechselberger, and J. Sneyd. Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics. *J. Nonlinear Sci.*, 21:639–683, 2011.

J. D. Hays, John Imbrie, and N. J. Shackleton. Variations in the Earth's orbit: Pacemaker of the ice ages. *Science*, 194(4270):1121–1132, 1976. doi: 10.1126/science.194.4270.1121.

Peter Huybers. Early Pleistocene glacial cycles and the integrated Summer insolation forcing. *Science*, 313:508–510, 2006. doi: 10.1126/science.1125249.

Peter Huybers. Glacial variability over the last two million years: an extended depth-derived agemodel, continuous obliquity pacing, and the Pleistocene progression. *Quaternary Science Reviews*, 26:37–55, 2007. doi: 10.1016/j.quascirev.2006.07.013.

Peter Huybers. Pleistocene glacial variability as a chaotic response to obliquity forcing. *Climate of the Past*, 5:481–488, 2009. doi: 10.5194/cp-5-481-2009.

Peter Huybers. Combined obliquity and precession pacing of late Pleistocene deglaciation. *Nature*, 480:229–231, 2011. doi: 10.1038/nature10626.

Peter Huybers and Charles H Langmuir. Delayed CO_2 emissions from mid-ocean ridge volcanism as a possible cause of late-Pleistocene glacial cycles. *Earth and Planetary Science Letters*, 457:238–249, 2017. doi: 10.1016/j.epsl.2016.09.0.

Peter Huybers and Carl Wunsch. A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography*, 19 (1), 2004. doi: 10.1029/2002PA000857.

Peter Huybers and Carl Wunsch. Obliquity pacing of the late Pleistocene glacial terminations. *Nature*, 434:491, Mar 2005. doi: 10.1038/nature03401.

John Imbrie. Astronomical theory of the Pleistocene ice ages: A brief historical review. *Icarus*, 50(2):408 – 422, 1982. doi: 10.1016/0019-1035(82)90132-4.

John Imbrie and John Z. Imbrie. Modeling the climatic response to orbital variations. *Science*, 207:943–953, 1980. doi: 10.1126/science.207.4434.943.

John Z Imbrie, Annabel Imbrie-Moore, and Lorraine Lisiecki. A phase-space model for Pleistocene ice volume. *Earth and Planetary Science Letters*, 307:94–102, 2011. doi: 10.1016/j.epsl.2011.04.018.

IPCC. Summary for policymakers. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, Maycock, M. Tignor, and T. Waterfield, editors, *An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, pages 1–32. World Meteorological Organization, Geneva, Switzerland, 2018.

Hildeberto Jardón-Kojakhmetov and Henk W. Broer. Polynomial normal forms of constrained differential equations with three parameters. *Journal of differential equations*, 257:1012–1055, 2014. doi: 10.1007/BF03026731.

R. V. Jensen. Synchronization of driven nonlinear oscillators. *American Journal of Physics*, 70(6):607–619, 2002. doi: 10.1119/1.1467909.

Markus Jochum, Alexandra Jahn, Synte Peacock, David A. Bailey, John T. Fasullo, Jennifer Kay, Samuel Levis, and Bette Otto-Bliesner. True to Milankovitch: Glacial inception in the New Community Climate System Model. *Journal of Climate*, 25(7):2226–2239, 2012. doi: 10.1175/JCLI-D-11-00044.1.

Christopher K.R.T Jones. Geometric singular perturbation theory. In Johnson R., editor, *Dynamical Systems*, volume 1609 of *Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg, 1995. ISBN 978-3-540-60047-3. doi: 10.1007/BFb0095239.

Amitai Katz. The interaction of magnesium with calcite during crystal growth at 25–90°c and one atmosphere. *Geochimica et Cosmochimica Acta*, 37(6):1563 – 1586, 1973. ISSN 0016-7037. doi: 10.1016/0016-7037(73)90091-4.

J. P. Keener, F. C. Hoppensteadt, and J. Rinzel. Integrate and fire models of nerve membrane response to oscillatory input. *SIAM Journal on Applied Mathematics*, 41(3):503–517, 12 1981. ISSN 0036-1399.

A. A. Kirillov and A. D. Gvishiani. *Problems and theorems in functional analysis*. Springer-Verlag, New York, 1 edition, 1982. ISBN 978-1-4613-8153-2. Translated by H.H. McFaden.

Peter E. Kloeden and Martin Rasmussen. *Nonautonomous dynamical systems*. Providence, Rhode Island : American Mathematical Society, 2011. ISBN 9781470414030.

Bruce W. Knight. Dynamics of encoding in a population of neurons. *The Journal of General Physiology*, 59(6):734–766, 1972. doi: 10.1085/jgp.59.6.734.

K. Krischer, M. Eiswirth, and G. Ertl. Oscillatory co oxidation on pt(110): modeling of temporal self-organization. *J. Chem. Phys.*, 96, 1992. doi: 10.1063/1.462226.

Periklis Y. Ktonas and Nicola Papp. Instantaneous envelope and phase extraction from real signals: Theory, implementation, and an application to eeg analysis. *Signal Processing*, 2 (4):373 – 385, 1980. ISSN 0165-1684. doi: 10.1016/0165-1684(80)90079-1.

Christian Kuehn. An introduction to rotation theory. Online resource, 2007. http://tutorials.siam.org/dsweb/circlemaps/circle.pdf.

Christian Kuehn. *Multiple time scale dynamics*, volume 191 of *Applied Mathematical Sciences*. Springer, Cham, 2015. ISBN 978-3-319-12315-8; 978-3-319-12316-5. doi: 10.1007/978-3-319-12316-5.

U. Kumar, V. Kumar, and J.N. Kapur. Normalized measures of entropy. *International Journal of General Systems*, 12(1):55–69, 1986. doi: 10.1080/03081078608934927.

Yuri A. Kuznetsov. *Elements of applied bifurcation theory*, volume 112 of *Applied Mathematical Sciences*. Springer-Verlag, New York, third edition, 2004. ISBN 0-387-21906-4. doi: 10.1007/978-1-4757-3978-7.

E. Källén, C. Crafoord, and M. Ghil. Free oscillations in a climate model with ice-sheet dynamics. *Journal of the Atmospheric Sciences*, 36(12):2292–2303, 1979. doi: 10.1175/1520-0469(1979)036<2292:FOIACM>2.0.CO;2.

J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A. C. M. Correia, and B. Levrard. A long-term numerical solution for the insolation quantities of the Earth. *Astronomy and Astrophysics*, 428:261–285, 2004. doi: 10.1051/0004-6361:20041335.

Herve Le Treut and Michael Ghil. Orbital forcing, climatic interactions, and glaciation cycles. *Journal of Geophysical Research*, 88(C9):5167–5190, 1983. doi: 10.1029/JC088iC09p05167.

C. H. Lear, H. Elderfield, and P. A. Wilson. Cenozoic deep-sea temperatures and global ice volumes from mg/ca in benthic foraminiferal calcite. *Science*, 287(5451):269–272, 2000. doi: 10.1126/science.287.5451.269.

Mark Levi. A period-adding phenomenon. *SIAM Journal on Applied Mathematics*, 50(4): 943–955, 1990. doi: 10.1137/0150058.

Raph Levien. The elastica: a mathematical history. Technical Report UCB/EECS-2008-103, EECS Department, University of California, Berkeley, Aug 2008. URL `http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-103.html`.

Norman Levinson. A second order differential equation with singular solutions. *Annals of Mathematics*, 50(1):127–153, 1949. doi: 10.2307/1969357.

Lorraine E. Lisiecki. Links between eccentricity forcing and the 100,000-year glacial cycle. *Nature Geoscience*, 3:349, Apr 2010. doi: 10.1038/ngeo828.

Lorraine E Lisiecki and Maureen E Raymo. A Plioene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography*, 20:437–440, 2005. doi: 10.1029/2004PA001071.

Lorraine E. Lisiecki and Maureen E. Raymo. Plio–Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, 26(1):56 – 69, 2007. doi: 10.1016/j.quascirev.2006.09.005.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Maxime Lucas, Julian Newman, and Aneta Stefanovska. Stabilization of dynamics of oscillatory systems by nonautonomous perturbation. *Phys. Rev. E*, 97, 2018. doi: 10.1103/PhysRevE.97.042209.

Dieter Lüthi, Martine Le Floch, Bernhard Bereiter, Thomas Blunier, Jean-Marc Barnola, Urs Siegenthaler, Dominique Raynaud, Jean Jouzel, Hubertus Fischer, Kenji Kawamura, and Thomas F. Stocker. High-resolution carbon dioxide concentration record 650,000-800,000 years before present. *Nature*, 453:379, 2008. doi: 10.1038/nature06949.

Mikikian M., M. Cavarroc, L. Couëdel, Y. Tessier, and L. Boufendi. Mixed-mode oscillations in complex-plasma instabilities. *Phys. Rev. Lett.*, 100, 2008. doi: 10.1103/PhysRevLett.100.225005.

Kirk A Maasch and Barry Salzman. A low-order dynamical model of global climatic variability over the full Pleistocene. *Journal of Geophysical Research*, 95(D2):1955–1963, 1990. doi: 10.1029/JD095iD02p01955.

Douglas R Macayeal. A catastrophe model of paleoclimate. *Journal of Glaciology*, 24(30): 245–257, 1979.

Arianna Marchionne, Peter Ditlevsen, and Sebastian Wieczorek. Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual study. *Physica D*, 380-381: 8–16, 2018. doi: 10.1016/j.physd.2018.05.004.

John Mather. Stability of C^∞ mappings, IV. classification of stable germs by r-algebras. *Publications Mathématiques de l'IHÉS*, 37:223–248, 1969a.

John N. Mather. Stability of C^∞ mappings: II. infinitesimal stability implies stability. *Annals of Mathematics*, 89(2):254–291, 1969b. doi: 10.2307/1970668.

Charles Edward Mathews, T. G Bonney, Louis Kurz, William Windham, and Peter Martel. *The annals of Mont Blanc: a monograph.* T. Fisher Unwin, Paternoster Square, London, 1898. Available at https://archive.org/details/annalsofmontblan00math/page/348, accessed 9 April 2019.

Mark J. McGuinness. The fractal dimension of the Lorenz attractor. *Physics Letters A*, 99 (1):5 – 9, 1983. ISSN 0375-9601. doi: 10.1016/0375-9601(83)90052-X.

R. Mettin, U. Parlitz, and W. Lauterborn. Bifurcation structure of the driven van der Pol oscillator. *International Journal of Bifurcation and Chaos*, 03(06):1529–1555, 1993. doi: 10.1142/S0218127493001203.

Takahito Mitsui, Michel Crucifix, and Kazuyuki Aihara. Dynamics between order and chaos in conceptual models of glacial cycles. *Climate dynamics*, 42:3087–3099, 2013. doi: 10.1007/s00382-013-1793-x.

Takahito Mitsui, Michel Crucifix, and Kazuyuki Aihara. Bifurcations and strange nonchaotic attractors in a phase oscillator model of glacial–interglacial cycles. *Physica D*, 306:25–33, 2015. doi: 10.1016/j.physd.2015.05.007.

Alan C. Mix and William F. Ruddiman. Oxygen-isotope analyses and Pleistocene ice volumes. *Quaternary Research*, 21(1):1 – 20, 1984. doi: 10.1016/0033-5894(84)90085-1.

James Montaldi. The path formulation of bifurcation theory. In *Dynamics, Bifurcation and Symmetry*, pages 259–278. Springer, 1994.

J. L. Mélice, A. Coron, and A. Berger. Amplitude and frequency modulations of the Earth's obliquity for the last million years. *Journal of Climate*, 14(6):1043–1054, 2001. doi: 10.1175/1520-0442(2001)014<1043:AAFMOT>2.0.CO;2.

Anne Willem Omta, Bob W Kooi, George A K van Voorn, Rosalind E M Rickaby, and Michael J Follows. Inherent characteristics of sawtooth cycles can explain different glacial periodicities. *Climate Dynamics*, 46:557–569, 2015. doi: 10.1007/s00382-015-2598-x.

Didier Paillard. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature*, 391:378–381, 1998. doi: 10.1038/34891.

Didier Paillard and Frederic Parrenin. The Antarctic ice sheet and the triggering of deglaciations. *Earth and Planetary Science Letters*, 227:263–271, 2004. doi: 10.1016/j.epsl.2004.08.023.

F. Parrenin, J.-M. Barnola, J. Beer, T. Blunier, E. Castellano, J. Chappellaz, G. Dreyfus, H. Fischer, S. Fujita, J. Jouzel, K. Kawamura, B. Lemieux-Dudon, L. Loulergue, V. Masson-Delmotte, B. Narcisi, J.-R. Petit, G. Raisbeck, D. Raynaud, U. Ruth, J. Schwander, M. Severi, R. Spahni, J. P. Steffensen, A. Svensson, R. Udisti, C. Waelbroeck, and E. Wolff. The EDC3 chronology for the EPICA Dome C ice core. *Climate of the Past*, 3(3):485–497, 2007. doi: 10.5194/cp-3-485-2007.

Frederic Parrenin and Didier Paillard. Terminations VI and VIII (530 and 720 kyr BP) tell us the importance of obliquity and precession in the triggering of deglaciations. *Climate of the Past*, 8(6):2031–2037, 2012. doi: 10.5194/cp-8-2031-2012.

Frédéric Parrenin and Didier Paillard. Amplitude and phase of glacial cycles from a conceptual model. *Earth and Planetary Science Letters*, 214(1):243 – 250, 2003. doi: 10.1016/S0012-821X(03)00363-7.

J. R. Petit, J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. PÉpin, C. Ritz, E. Saltzman, and M. Stievenard. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399:429–436, Jun 1999. doi: 10.1038/20859. Article.

Arkady Pikovsky, M Rosenblum, and Jurgen Kurths. *Synchronization: A universal phenomenon in the nonlinear sciences*. Cambridge University Press, Cambridge, 1 edition, 2001.

N.G. Pisias and T.C. Moore. The evolution of Pleistocene climate: A time series approach. *Earth and Planetary Science Letters*, 52(2):450 – 458, 1981. doi: 10.1016/0012-821X(81)90197-7.

Courtney Quinn, Jan Sieber, Anna S. von der Heydt, and Timothy M. Lenton. The Mid-Pleistocene transition induced by delayed feedback and bistability. *Dynamics and Statistics of the Climate System*, 3(1):1–17, 2018. doi: 10.1093/climsys/dzy005.

Martin Rasmussen. Towards a bifurcation theory for nonautonomous difference equations. *Journal of Difference Equations and Applications*, 12(3-4):297–312, 2006. doi: 10.1080/10236190500489400.

M. E. Raymo, L. E. Lisiecki, and Kerim H. Nisancioglu. Plio-Pleistocene ice volume, Antarctic climate, and the global δ^{18}O record. *Science*, 313(5786):492–495, 2006. doi: 10.1126/science.1123296.

Maureen E Raymo. The timing of major climate terminations. *Paleoceanography*, 12:577–585, 1997. doi: 10.1029/97PA01169.

Frederique Remy and Laurent Testut. Mais comment s'ecoule donc un glacier apercu historique. *Comptes Rendus Geoscience*, 338(5):368–385, 2006. ISSN 1631-0713. doi: 10.1016/j.crte.2006.02.004.

J.A. Rial and C.A. Anaclerio. Understanding nonlinear responses of the climate system to orbital forcing. *Quaternary Science Reviews*, 19(17):1709 – 1722, 2000. ISSN 0277-3791. doi: 10.1016/S0277-3791(00)00087-1.

Andy J. Ridgwell, Andrew J. Watson, and Maureen E. Raymo. Is the spectral signature of the 100 kyr glacial cycle consistent with a Milankovitch origin? *Paleoceanography*, 14(4): 437–440, 1999. doi: 10.1029/1999PA900018.

S. Rinaldi and M. Scheffer. Geometric analysis of ecological models with slow and fast processes. *Ecosystems*, 3:507–521, 2000.

Gerard H. Roe and Myles R. Allen. A comparison of competing explanations for the 100,000-yr ice age cycle. *Geophysical Research Letters*, 26(15):2259–2262, 1999. doi: 10.1029/1999GL900509.

M. Sakanoue, K. Konishi, and K. Komura. *Stepwise determinations of thorium, protactinium and uranium isotopes and their applications in geochronological studies.* International Atomic Energy Agency (IAEA), 2 edition, 1967.

Barry Saltzman and Kirk 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. doi: 10.1029/GB002i002p00177.

Barry Saltzman and Alfonso Sutera. A model of the internal feedback system involved in late Quaternary climatic variations. *Journal of the Atmospheric Sciences*, 41(5):736–745, 1984. doi: 10.1175/1520-0469(1984)041<0736:AMOTIF>2.0.CO;2.

Barry Saltzman, Anthony R. Hansen, and Kirk A. Maasch. The late Quaternary glaciations as the response of a three-component feedback system to Earth-orbital forcing. *Journal of the Atmospheric Sciences*, 41(23):3380–3389, 1984. doi: 10.1175/1520-0469(1984)041<3380:TLQGAT>2.0.CO;2.

Marten Scheffer. *Critical transitions.* Princeton University Press, 2009. ISBN 9780691122045.

William D. Sellers. A global climatic model based on the energy balance of the Earth-atmosphere system. *Journal of Applied Meteorology*, 8(3):392–400, 1969. doi: 10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2.

Nicholas John Shackleton. Oxygen isotope analyses and Pleistocene temperatures reassessed. *Nature*, 215(15):39 – 55, 1967. doi: 10.1038/215015a0.

Nicholas John Shackleton and Neil D. Opdyke. Oxygen isotope and palaeomagnetic stratigraphy of Equatorial Pacific core V28-238: Oxygen isotope temperatures and ice volumes on a 105 year and 106 year scale. *Quaternary Research*, 3(1):39 – 55, 1973. doi: 10.1016/0033-5894(73)90052-5.

Nicholas John Shackleton and Neil D. Opdyke. Oxygen-isotope and paleomagnetic stratigraphy of pacific core v28-239: Late pliocene to latest Pleistocene. *Geological Society of America Memoir*, 145:449–464, 1976.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Sindia Sosdian and Yair Rosenthal. Deep-sea temperature and ice volume changes across the pliocene-Pleistocene climate transitions. *Science*, 325(5938):306–310, 2009. doi: 10.1126/science.1169938.

Alistair B. Sproul. Derivation of the solar geometric relationships using vector analysis. *Renewable Energy*, 32(7):1187 – 1205, 2007. doi: 10.1016/j.renene.2006.05.001.

Duane Storti and Richard H. Rand. Subharmonic entrainment of a forced relaxation oscillator. *International Journal of Non-Linear Mechanics*, 23(3):231 – 239, 1988. ISSN 0020-7462. doi: https://doi.org/10.1016/0020-7462(88)90014-5.

Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering.* Reading, Massachusets, Addison-Wesley Publishing, 1994.

Floris Takens. Constrained equations; a study of implicit differential equations and their discontinuous solutions. In Peter Hilton, editor, *Structural Stability, the Theory of Catastrophes, and Applications in the Sciences*, pages 143–234, Berlin, Heidelberg, 1976. Springer Berlin Heidelberg. ISBN 978-3-540-38254-6.

René Thom. *Stabilité structurelle et morphogénèse.* Benjamin, Advanced Book Program, Reading, Massachusets, 1972.

Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorologial Society*, 79(1):61–78, 1998. doi: 10.2307/1907187.

P C Tzedakis, Michel Crucifix, Takahito Mitsui, and E W Wolff. A simple rule to determine which insolation cycles lead to interglacials. *Nature*, 542:427–432, 2017. doi: 10.1038/nature21364.

Eli Tziperman and Hezi Gildor. On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. *Paleoceanography*, 18(1): 1–8, 2003. doi: 10.1029/2001PA000627.

Eli Tziperman, Maureen Raymo, Peter Huybers, and Carl Wunsch. Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing. *Paleoceanography*, 21:1–11, 2006. doi: 10.1029/2005PA001241.

David Vakman. On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. *IEEE Transactions on Signal Processing*, 44(4):791–797, 1996. doi: 10.1109/78.492532.

B. van der Pol. On "relaxation-oscillations". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):978–992, 1926. doi: 10.1080/14786442608564127.

B. van der Pol and J. van der Mark. Frequency demultiplication. *Nature*, 120:363–364, 1927. doi: 10.1038/120363a0.

G. Vettoretti and W.R. Peltier. The impact of insolation, greenhouse gas forcing and ocean circulation changes on glacial inception. *The Holocene*, 21(5):803–817, 2011. doi: 10.1177/0959683610394885.

Johannes Weertman. Stability of ice-age ice sheets. *Journal of Geophysical Research (1896-1977)*, 66(11):3783–3792, 1961. doi: 10.1029/JZ066i011p03783.

Johannes Weertman. Milankovitch solar radiation variations and ice age ice sheet sizes. *Nature*, 261(5555):17–20, 1976. doi: 10.1038/261017a0.

Hassler Whitney. On singularities of mappings of Euclidean spaces. I. mappings of the plane into the plane. *Annals of Mathematics*, 62(3):374–410, 1955. doi: 10.2307/1970070.

M. Willeit, A. Ganopolski, R. Calov, and V. Brovkin. Mid-Pleistocene transition in glacial cycles explained by declining CO_2 and regolith removal. *Science Advances*, 5(4), 2019. doi: 10.1126/sciadv.aav7337.

Carl Wunsch. Quantitative estimate of the Milankovitch-forced contribution to observed Quaternary climate change. *Climate Dynamics*, 20(353), 2003. doi: 10.1007/s00382-002-0279-z.

YassineMrabet. A simple torus. Online resource, 11 2007. `https://commons.wikimedia.org/wiki/File:Simple_Torus.svg`.

Appendix A

Paper A:The middle Pleistocene transition by frequency locking and slow ramping of internal period

Climate Dynamics
https://doi.org/10.1007/s00382-019-04679-3

The middle Pleistocene transition by frequency locking and slow ramping of internal period

Karl H. M. Nyman[1] · **Peter D. Ditlevsen**[1]

Received: 24 August 2018 / Accepted: 9 February 2019

Abstract

The increase in glacial cycle length from approximately 41 to on average 100 thousand years around 1 million years ago, called the middle Pleistocene transition (MPT), lacks a conclusive explanation. We describe a dynamical mechanism which we call ramping with frequency locking (RFL), that explains the transition by an interaction between the internal period of a self-sustained oscillator and forcing that contains periodic components. This mechanism naturally explains the abrupt increase in cycle length from approximately 40 to 80 thousand years observed in proxy data, unlike some previously proposed mechanisms for the MPT. A rapid increase in durations can be produced by a rapid change in an external parameter, but this assumes rather than explains the abruptness. In contrast, models relying on frequency locking can produce a rapid change in durations assuming only a slow change in an external parameter. We propose a scheme for detecting RFL in complex, computationally expensive models, and motivate the search for climate variables that can gradually increase the internal period of the glacial cycles.

Keywords Glacial cycles · Middle Pleistocene transition · Frequency locking · Internal period · Abrupt transition

1 Introduction

Since the beginning of major Northern hemisphere glaciation 2.7 million years (Myr) ago, Earth has undergone alternating epochs of icy and cold conditions on the one hand, and warm and ice-free conditions on the other (Fig. 1). While these glacial cycles are attested from geological records (EPICA Community Members 2004; Huybers 2007; Lisiecki and Raymo 2005), there is no single conclusive theory of their origins.

Milutin Milanković proposed in the 1920s that glacial cycles should repeat roughly every 40 thousand years (kyr), based on calculations of incoming solar radiation. Alas, in the 1970s accurately dated oxygen isotope data from ocean sediment cores cast doubt on his theory, by revealing that the glacial cycles over the past 800 kyr were closer to 100 kyr long (Imbrie and Imbrie 1979). Soon it became clear, however, that the ~ 100 kyr cycles were preceded by ~ 40 kyr long cycles, consistent with the theory of Milanković. This spawned three questions: what caused the ~ 100 kyr cycles, what caused the ~ 40 kyr cycles (was Milanković right?), and what caused the transition between them around 1 Myr ago, called the middle Pleistocene transition (MPT) (Clark et al. 2006). The last question is the focus of this paper.

The main strategy to address these questions has been to replicate the palaeoclimatic records using simple models of glacial cycles with few variables, referred to as conceptual models [see (Crucifix 2012) for a review]. One reason for this is that the rather regular and cyclic variations in data suggest that the main dynamics can be captured by a system of few degrees of freedom, even as the full climate system obviously has a large number of degrees of freedom. None of these models describe the climate system in detail, but they are useful for understanding underlying dynamics. Virtually all models involve insolation variations due to changes to Earth's orbital configuration relative to the sun, an idea heralded by Adhémar, Croll and Milankovitch (Imbrie and Imbrie 1979). But the specific role played by insolation variations is still unknown and debated.

✉ Karl H. M. Nyman
karl.nyman@nbi.ku.dk

Peter D. Ditlevsen
pditlev@nbi.ku.dk

[1] Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

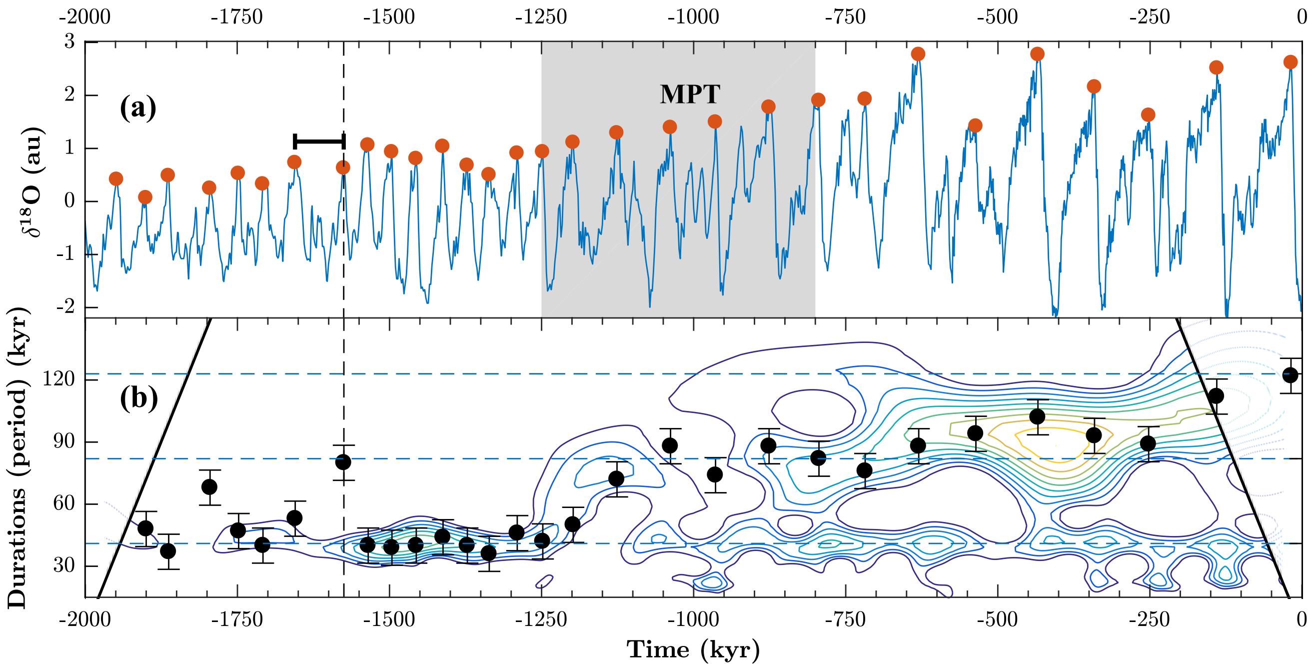


Fig. 1 **a** The LR04 stack of normalised isotopic oxygen anomalies in deep ocean sediment cores—a proxy for global ice volume and deep ocean temperatures (Lisiecki and Raymo 2005). Higher values means more ice. **b** Durations between successive major terminations (black), showing an increase at the MPT. Error bars indicate one standard deviation dating uncertainty. Contours (darker to lighter) show the amplitude of a wavelet spectral estimate. Each lighter contour corresponds to an increase of 7% of the maximum amplitude, starting at 30%. Outside the cone of influence (thick black line), edge effects are important. See "Appendix B and D" for details

Several solutions to the cause of the MPT have been presented within the context of conceptual models. Some mechanisms rely on a bifurcation occurring in the unforced climate system which fundamentally changes how the system operates (Ashwin and Ditlevsen 2015; Ditlevsen 2009; Huybers and Langmuir 2017; Maasch and Salzman 1990; Tziperman and Gildor 2003). Other mechanisms invoke a “spontaneous” change, such as a shift between attractors due to subtle changes in insolation (Omta et al. 2015; Quinn et al. 2018) or random fluctuations (Imbrie et al. 2011; Salzman and Verbitsky 1993), or as a coincidence (Huybers 2009). A third possible mechanism for the MPT assumes one essential mode of oscillation throughout the Pleistocene and relies crucially on the interaction between insolation variations and an increasing internal period. This mechanism, previously imprecisely referred to as phase/frequency locking and non-linear resonance—but here ramping with frequency locking (RFL)—is the focus of this paper.

The main appeal of this mechanism is that nothing special had to occur in the climate system over the MPT (Huybers 2007); it is only required that the internal period was ramped slowly—interactions with forcing are enough to cause an abrupt increase in durations between glacial terminations (Fig. 1, bottom panel).

The first publications where RFL was used (Paillard 1998; Paillard and Parrenin 2004) did not explain why the durations between glacial transitions increased abruptly over the MPT. Ashkenazy (2006) and Ashkenazy and Tziperman (2004) hinted how frequency locking (therein called phase locking) could produce an abrupt increase in duration, by showing diagrams of average duration as a function of a system parameter (Devil’s staircases). Huybers (2007) was first to both show a model trajectory of ice volume using the mechanism, and to attribute the effect to “skipping of obliquity cycles”, a frequency locking effect. Recently, Daruka and Ditlevsen (2015), Feng and Bailer-Jones (2015), Mitsui et al. (2015) and Tzedakis et al. (2017) alluded to the mechanism, but neither emphasised that frequency locking can explain the MPT assuming only a slow linear change in a climate parameter. Instead, by ramping some parameter in a way that mimics the rapid change in durations over the MPT, they prescribe an abrupt increase in period over the MPT rather than explaining it. Here, we for the first time properly define RFL and emphasise its generality.

Following (Huybers 2007), we question the common assumption that climate entered a stationary state in the late Pleistocene, and instead argue that the sequence of durations between glacial terminations is consistent with a slow increase of the internal period of the climate system until present (Fig. 1b). According to this view, the typical durations between transitions changed from ∼40 to ∼80 kyr around 1200 kyr ago, after which they increased gradually in the mean to present time, with the last duration being ~ 120 kyr long.

Here, we first aim to explain RFL in a clear way, using a harmonically forced simple model. We use harmonic (pure sine) forcing because it makes frequency locking concepts clearer, while still producing qualitatively similar behaviour to astronomical forcing curves. We should not expect model runs with such simplified forcing to agree well with data, however. We use forcing with period 41 kyr, corresponding to the main period of obliquity variations (Berger 1978), which determine the total insolation integrated over the summer at Northern latitudes (Huybers 2006).

We then define RFL, specify a class of models able to reproduce the MPT using the mechanism, and propose a decomposition of model components to understand the abruptness of the MPT. We consider evidence in data for a 40–80 kyr shift in durations between terminations and a subsequent gradual increase, and why this supports RFL in favour of some other mechanisms for the MPT. We then discuss how insights from harmonic forcing relate to non-harmonic forcing, how RFL can be detected in complex and computationally expensive models, and some climate variables that can cause an increase in the internal period of the glacial cycles.

2 The idea behind ramping with frequency locking

We illustrate RFL using a deterministic and continuous time version of the H07 model (Huybers 2007) (see Fig. 2). The model is arguably the simplest to represent alternating stages of intrinsic growth and decay of ice sheets, with the growth state ending abruptly as a critical ice volume is reached. It is is an integrate-and-fire threshold model conceptually very similar to the models in Ashkenazy and Tziperman (2004), Glass and Mackey (1979), Huybers (2007), Imbrie and Imbrie (1980), Imbrie et al. (2011), Paillard (1998), Parrenin and Paillard (2003, 2012), van der Pol and van der Mark (1927) and de Saedeleer et al. (2013).

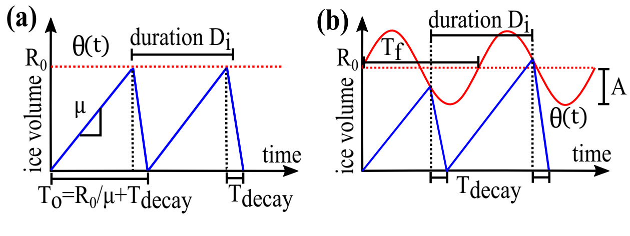


Fig. 2 The H07 model (Eq. 1), **a** unforced and **b** periodically forced. Ice volume (blue) grows linearly at a rate μ until a threshold (red) is hit, after which ice volume is reset to 0 over a time T_{decay}. In **a** the threshold of glacial termination $\theta(t)$ is constant $\theta(t) = R_0$, whereas in **b** it oscillates periodically as $\theta(t) = R_0 + A\sin(2\pi t/T_f)$. T_o is the internal (unforced) period of the model

Physically, sudden and rapid deglaciation has been explained e.g. with isostatic rebound, rapid CO_2 outgassing (Paillard and Parrenin 2004) and rapid loss of Northern hemisphere sea ice cover (Gildor and Tziperman 2000).

We assume that ice volume $x(t)$ grows at a constant rate μ in a glacial state until it reaches a threshold $\theta(t)$. Then deglaciation starts, whereby ice volume decays to 0 over a fixed time $T_{decay} = 10$ kyr:

$$\begin{aligned}&\dot{x} = \mu \text{ until } x(t) = \theta(t), \text{ then}\\&\quad \text{linearly decrease } x(t) \text{ to 0 over time } T_{decay}, \text{ repeat.}\end{aligned} \tag{1}$$

Small perturbations to the model, such as having a constant rate of decay instead of a fixed time, does not qualitatively affect its behaviour.

We split $\theta(t)$ into a *forcing* term $A \cdot F(t)$—a zero-mean sum of periodic components—and a *ramping* term $R(t)$: $\theta(t) = R(t) + A \cdot F(t)$.

In the limit of constant ramping $R(t) = R_0$ and zero forcing $A = 0$, the system has a constant *internal* period of oscillation $T_o = \frac{R_0}{\mu} + T_{decay}$ (subscript *o* for *oscillator*), see Fig. 2a. But if the threshold increases slowly over time, for instance linearly $\theta(t) = R(t) = R_0 + R_1 t$ as in Fig. 3ai, then the internal period $T_o(t) = \frac{R_0}{\mu} + T_{decay} + \frac{R_1}{\mu} t$ also increases slowly (Fig. 3bi). As there is no forcing, the durations between glacial terminations follow $T_o(t)$ closely.

However, with periodic forcing $\theta(t) = R_0 + R_1 t + A\sin(2\pi/T_f)$, durations D_i are near multiples of the forcing period $D_i \approx NT_f$, $N \in \mathbb{N}$ (Fig. 3bi). Roughly speaking, the multiple that is realised is the one closest to the internal period T_o. This phenomenon, called frequency locking (Pikovsky et al. 2001), has been studied extensively over the past century [e.g. Cartwright and Littlewood (1945); Glass and Mackey (1979); Le Treut and Ghil (1983); van der Pol and van der Mark (1927); Tziperman et al. (2006)].

In Fig. 3ai the durations D_i change abruptly from $1 \times T_f$ to $2 \times T_f$ and finally $3 \times T_f$. These abrupt changes in durations resulting from a gradual change in an underlying parameter is one possible dynamical mechanism behind the MPT.

We call the mechanism RFL, rather than non-linear resonance, phase locking or frequency locking as it has previously been called. This we do to emphasise both that an internal period must increase gradually over time (ramping), and that the internal oscillations must be locked to external forcing. This is opposed to e.g. the mechanism in Omta et al. (2015), which realises the MPT through jumps between coexisting frequency locked solutions.

We note that RFL is a special case of "slow passage through bifurcation" [e.g. (Baer et al. 1989; Do and Lopez 2012)], for

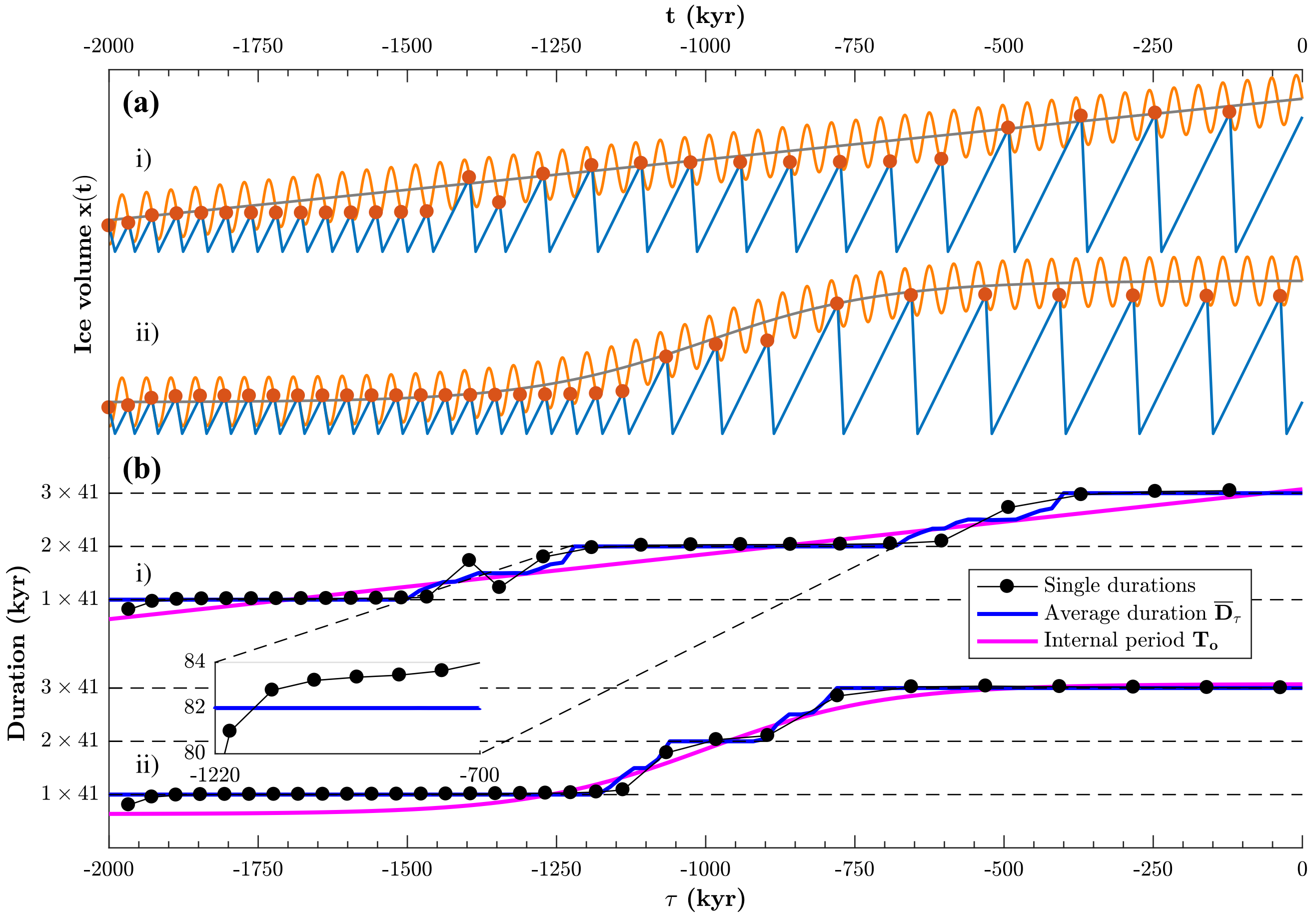


Fig. 3 The ramping with frequency locking (RFL) mechanism for the periodically forced H07 model. **a** Ice volume (blue sawtooth) over time for a i) linear and ii) sigmoidal ramp of the upper threshold $\theta(t)$. Periodic threshold in orange and unforced threshold in grey. **b** Average duration between glacial terminations $\overline{D}_\tau$ over frozen time τ (black solid) (known as Devil's staircases, Sect. 4.1) for the i) linearly and ii) sigmoidally ramped thresholds, and sample durations (red dotted lines) for the forced solutions in a). Magenta lines show the internal period $T_o(\tau)$. The inset shows that durations with time-varying ramping $R(t)$ do not agree perfectly with the average duration $\overline{D}_\tau$, computed for $R(t) = \text{const}$ (see Sect. 5). Parameters are $\mu = 1, T_{decay} = 10, T_f = 41, A = 20$ and ramp functions are i) $R(\tau) = 26 + 0.05 \times (\tau + 2000)$ and ii) $R(t) = 26 + 50(\tanh(\frac{t+1000}{300}) + 1)$ respectively

which the bifurcations typically are saddle-node bifurcations of limit cycles marking transitions in and out of frequency locking regions (Pikovsky et al. 2001). [However, see e.g. Guckenheimer et al. (2003) and Levi (1990) for other relevant bifurcations].

Finally, we note that H07 is an illustrative example of RFL, and not representative of all glacial cycle models. However, the rapid jumps between frequency locking regions occur generically in a broad class of models, defined next.

3 A formal description of RFL

H07 (Fig. 2) is just one particular model capable of realising the MPT through RFL. We could simply call these self-sustained oscillators, but we aim to be more precise and to establish notation.

First, we naturally require the model to be a dynamical system, such that there is an evolution rule $f(t, x)$ taking a state $x(t)$ forward in time t. We identify the model with the evolution rule and denote it f (without arguments) for brevity. $f(t, x)$, $x(t)$ and t can be very general, for instance; t can be continuous or discrete, x can be of any dimension, and $f(t, x)$ can e.g. be a piecewise smooth ODE paired with a switching rule, as for H07.

We also require f to be forced by a continuous zero-mean sum of periodic components $A(t)F(t)$ with an amplitude $A(t)$, called the *forcing*. We further require that f is parametrised by a set of parameters $p(t)$, whose time-varying subset $R(t)$ is called the *ramping*. Thus we can write $f = f(t, x, R(t), A(t)F(t))$.

We define the *frozen system* $f_\tau := f(t, x, R(\tau), A(\tau)F(t))$ as f with parameters frozen at time $t = \tau$. Importantly, we require that f is a *self-sustained* oscillator with internal period

$T_o(R(\tau))$, meaning that every solution to f_τ with $A(t) \equiv 0$ tends *asymptotically* (as $t \to \infty$) to a periodic solution with period $T_o(R(\tau))$. For RFL to be relevant we require that $T_o(R(\tau))$ increases as a function of τ. This is the *ramping* part of RFL.

The *frequency locking* part of RFL comes from the response of f to non-zero but constant forcing $A(\tau)$. For small and medium size $A(\tau)$, asymptotic solutions to the *frozen system* $f(t, x, R(\tau), A(\tau)F(t))$, are generically periodic with periods related rationally to the forcing periods (Pikovsky et al. 2001). The oscillator period can for instance be twice that of the forcing period. If so, the oscillator period (and therefore frequency) remains constant on open sets of parameters and we say that solutions are frequency locked to the forcing (we return to this in Sect. 4).

The essence of RFL is that the period of the frozen system can change rapidly as function of $T_o(R)$ when a ramped parameter causes the system to switch between frequency locking regions.

However, some remarks are in place. Firstly, the system with time varying parameters $f(t, x(t), R(t), A(t)F(t))$ is *not* the same as the frozen system $f(t, x(t), R(t), A(\tau)F(\tau))$ since solutions to the former cannot equilibrate to solutions of the latter in finite time. We return to differences between the two systems in Sect. 5 but until then we focus on the frozen system.

Secondly, the period of an oscillator is not the same as the length of individual "cycles". For instance, around − 1350 kyr in Fig. 3, short and long "cycles" alternate. This makes the average time between terminations 61.5 kyr, whereas the period (time until repetition, two large peaks) is 123 kyr. Therefore, we instead characterise local behaviour with the *average duration*:

$$\overline{D}_\tau = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} D_{i,\tau}, \tag{2}$$

where $D_{i,\tau}$ denotes the i:th duration between successive crossings of a fixed threshold for the frozen system f_τ, and the limit is taken as the number of crossings n goes to infinity. The threshold should be chosen such that a crossing occurs once per glacial cycle. For some models, like H07, the threshold defining glacial terminations can be used as a threshold to define durations $D_{i,\tau}$. For models without explicit thresholds, an appropriately chosen Poincaré section can be used instead (Pikovsky et al. 2001).

4 Breaking down the dependency of $\overline{D}_\tau$ on τ

Comparing Fig. 3bi and ii shows that $\overline{D}_\tau$ can rise steeply both from frequency locking effects under a gradual change of parameter (Fig. 3bi) and from ramping of a climate parameter rapidly (Fig. 3bii).

We wish to break down the contribution to the local change in average duration from these effects and do so by considering the change $\Delta\overline{D}_\tau$ under a small perturbation $\Delta\tau$:

$$\begin{aligned}\frac{\Delta\overline{D}_\tau(\tau)}{\Delta\tau} &\approx \frac{\overline{D}_\tau(\tau+\Delta\tau) - \overline{D}_\tau(\tau)}{\Delta\tau} \\ &\approx \frac{\Delta\overline{D}_\tau(T_o, A)}{\Delta T_o}\,\frac{\Delta T_o(R)}{\Delta R}\,\frac{\Delta R(\tau)}{\Delta\tau} \\ &\quad + \frac{\Delta\overline{D}_\tau(T_o, A)}{\Delta A}\,\frac{\Delta A(\tau)}{\Delta\tau},\end{aligned} \tag{3}$$

where e.g. $\frac{\Delta\overline{D}_\tau(T_o,A)}{\Delta T_o} := \frac{\overline{D}_\tau(T_o+\Delta T_o,A)-\overline{D}_\tau(T_o,A)}{\Delta T_o}$, and where we have neglected higher order terms. This approximation is generally better the smaller $\Delta\tau$ is. As $\Delta\tau \to 0$, (3) tends to the chain rule, but since $\frac{\Delta\overline{D}_\tau(T_o,A)}{\Delta T_o} = 0$ wherever differentiable (see Sect. 4.1) it is more appropriate to consider $\Delta\overline{D}_\tau$ over short intervals of time $\Delta\tau$. In what follows we restrict ourselves to $\frac{\Delta A(\tau)}{\Delta\tau} = 0$.

Equation (3) says that the rate of change (abruptness) in time of the average duration $\overline{D}_\tau$ is approximately the product of the rates at which $R(t)$ changes with time, T_o changes with R, and $\overline{D}_\tau$ changes with T_o. Our point is that each of these factors can contribute to an abrupt change of $\overline{D}_\tau$ at the MPT, but they have different interpretations from a modelling perspective. We discuss these factors next.

4.1 $\overline{D}_\tau(T_o, A)$: Arnold tongues and Devil's staircases

The average duration $\overline{D}_\tau(T_o, A)$ as a function of internal period T_o and forcing amplitude A describes the frequency locking contribution to changes to $\overline{D}_\tau$ over time τ.

Frequency locking can be visualised in Arnold tongue diagrams; Fig. 4a reveals regions of constant average duration $\overline{D}_\tau$ in (T_o, A) space called Arnold tongues (Crucifix 2013; Pikovsky et al. 2001; de Saedeleer et al. 2013). Inside major $1 : N$ tongues, solutions are periodic with period N times the forcing period $T_f = 41$ kyr, as evidenced in Fig. 4a. Minor tongues emanate at $A = 0$ from other rationals of T_f, and in between them are quasiperiodic solutions. (We show only $M : N$, $M = \{1, 2\}$ Arnold tongues, defined numerically as sets for which $|\overline{D}_\tau - T_f \frac{N}{M}| < 0.5$. $\overline{D}_\tau$ is estimated over 6 Myr.)

A change in τ that in turn leads to a change in $T_o(R(\tau))$ traces out a path in (A, T_o) space (black and magenta lines in Fig. 4a. Such a path represents the change in system state as one or more parameters change in time over the MPT. The paths in Fig. 4a pass through the major 1:1, 1:2 and 1:3 locking tongues, in which there are respectively 1, 2 and 3 forcing periods per oscillator period. We learn that for larger A, a larger portion of the path stays inside the major $1 : N$ tongues, an observation also made in (Ashkenazy 2006).

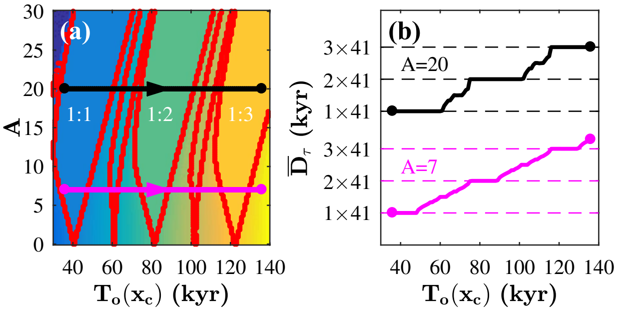


Fig. 4 **a** Arnold tongue diagram for increasing upper threshold R_0 and periodic forcing ($T_f = 41$ kyr) in the H07 model, showing regions in (A, T_o) space of constant average duration $\overline{D}_\tau$ (enclosed by red dots). Major $1 : N$ tongues, meaning that $\overline{D}_\tau = NT_f$, are labelled. Colour scale from blue (short) to yellow (long) reflects average duration. **b** Comparison between average duration $\overline{D}_\tau$ as a function of internal period T_o for strong ($A = 20$) and weaker ($A = 7$) forcing

Another way of visualising the change in average duration $\overline{D}_\tau$ as a function of T_o are Devil's staircases (Pikovsky et al. 2001) (Fig. 4b), in which the forcing amplitude A is fixed. We see that the average duration $\overline{D}_\tau$ is constant within Arnold tongues and that the staircase for larger A contains longer steps of constant duration, as predicted from Fig. 4a. Hence, stronger forcing influence tends to cause more abrupt changes to the average period.

4.2 $T_o(R)$ and $R(\tau)$

The function $T_o(R)$, if continuous and monotonic, stretches and squeezes Arnold tongues by scaling the independent variable T_o of $\overline{D}_\tau(T_o)$. In the Ashkenazy model (Ashkenazy 2006), for instance, a faster-than-linearly increasing $T_o(R)$ makes the 1:2 and 1:3 Arnold tongues, as a function of ice volume threshold, narrower and more closely spaced than the 1:1 tongue.

Ramping $R(\tau)$ continuously and monotonically, similarly stretches and squeezes Arnold tongues. For instance, in Fig. 5a sigmoidal ramping $R(t)$ makes the 1:2 Arnold tongue narrower compared to a linear change of $R(t)$. Figure 3 further illustrates this, showing model runs for either a sigmoidally ($R(t) = 26 + 50 \times (\tanh((t + 1000)/300) + 1)$) or a linearly $R(t) = 26 + 0.05 \times (t + 2000)$ ramped threshold. The sigmoidal ramping accelerates the increase in average duration around -1000 kyr, making the transition more abrupt. Note that the parameters in the functions $R(\tau)$ in Figs. 3 and 5 are different.

4.3 The roles of $\overline{D}_\tau(T_o, A)$, $T_o(R)$ and $R(\tau)$ in reproducing the MPT

All of $\overline{D}_\tau(T_o, A)$, $T_o(R)$ and $R(\tau)$ govern the average duration $\overline{D}_\tau$ and are able to cause an abrupt change of it, like the one observed at the MPT. From a modelling point of view, however, the functions carry different assumptions and are compatible with different hypotheses.

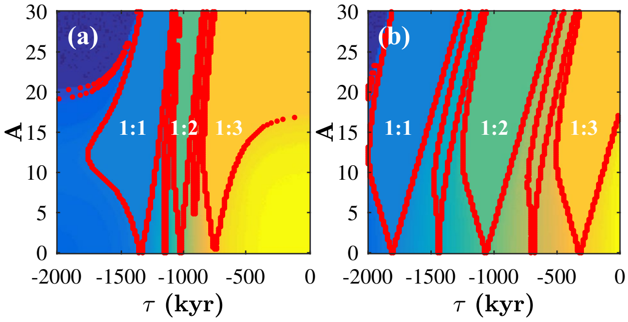


Fig. 5 Stretching of Arnold tongues by ramping the threshold parameter $R(\tau)$ in H07 at different rates. In **a** $R(\tau)$ is ramped sigmoidally $R(\tau) = 20 + 55 \times \left(\tanh\left(\frac{\tau+1000}{300}\right) + 1\right)$, while in **b** $R(\tau) = 20 + 0.055 \times (\tau + 2000)$. For further details, see Fig. 4 and the text

A model having an abrupt change due to $\overline{D}_\tau(T_o, A)$ relies on frequency locking properties, and assumes only slowly varying functions $T_o(R)$ and $R(\tau)$. Hence, the internal period is assumed to change slowly with model parameters, and parameters are assumed to change slowly in time. Such a model, relying on few assumptions about the climate system, makes full use of the RFL mechanism. The models in Huybers (2007), Paillard (1998) and Paillard and Parrenin (2004) and H07 are of this kind.

A model which relies predominantly on $T_o(R)$ for an abrupt change in average duration $\overline{D}_\tau$ is also consistent with a slowly changing external parameter $R(\tau)$, but a particular function $T_o(R)$ requires a physical explanation.

A model relying on a rapidly changed external parameter $R(\tau)$ does not need frequency locking properties of $\overline{D}_\tau(T_o, A)$ or a non-linear response of internal dynamics to the parameter $T_o(R)$. However, such a model prescribes the abrupt change in average duration at the MPT rather than explaining the dynamics behind it. Therefore, such an explanation requires justification for the rapidly changed external parameter. The models in Ashkenazy and Tziperman (2004), Daruka and Ditlevsen (2015), Mitsui et al. (2015) and Tzedakis et al. (2017) can be said to fall under this category, although they also achieve some abruptness through $\overline{D}_\tau(T_o, A)$.

5 Validity of the quasi-static approximation $f \sim f_\tau$

The quasistatic approximation is the approximation that parameters $R(\tau)$ change so slowly that the local average duration of f at time $t = \tau$:

$$\overline{D}_{\tau,loc} = \frac{1}{|I(\tau)|} \sum_{i \in I(\tau)} D_{i,\tau}, \tag{4}$$

is equal to $\overline{D}_\tau$. $I(\tau)$ is the set of indices of durations $D_{i,\tau}$ within a time interval $[\tau - \tau_0, \tau + \tau_1]$ around τ, with $\tau_0, \tau_1 > 0$. If $I(\tau) = \emptyset$, then we define $\overline{D}_{\tau,loc} = 0$.

If the quasistatic approximation holds, then the average duration, Arnold tongue diagrams and Devil's staircases calculated for the frozen system f_τ provide accurate information about local dynamics of f.

However, if $R(t)$ and/or $T_o(R(t))$ change rapidly around $t = \tau$, then there are two sources of discrepancy between $\overline{D}_{\tau,loc}$ and $\overline{D}_\tau$.

The first comes from that the length of the interval of time needed for a good average may be long relative to the local change of $\overline{D}_\tau$ for the system f_τ. Figure 6 illustrates that for a $9 \times 41 = 369$ kyr-periodic solution, a long interval is needed to get a local average duration $\overline{D}_{\tau,loc}$ in agreement with $\overline{D}_\tau$. At the same time, a long averaging interval fails to capture abrupt changes to $\overline{D}_\tau$.

The second is that solutions to f may fail to track solutions to f_τ. This occurs if the "frozen" attractor of f_τ changes (in some sense) at a fast rate, and if solutions attract to the frozen attractor at a slow rate. Quantifying these rates in a coordinate- and model-independent way seems difficult, however.

A candidate measure of rate of attraction is the maximal Lyapunov exponent of the return map mapping one transition time to another (Pikovsky et al. 2001). This can be normalised to a common time scale between models and is coordinate independent. However, since it is only a local measure it neglects the time it takes to enter a small neighbourhood of the attractor. This time can in practice dominate, as is the case in the standard circle model (not shown, model described in Pikovsky et al. 2001).

The local change in average duration $\left|\frac{\Delta \overline{D}_\tau}{\Delta \tau}\right|$ is a candidate measure of rate of change of an attractor of f_τ, since it exists in all models f and is coordinate-independent. It is ambiguous how large $\Delta\tau$ should be, however. Furthermore, the average duration $\overline{D}_\tau$ is only a proxy for the position of an attractor in phase space; an attractor can move even if $\frac{\Delta \overline{D}_\tau}{\Delta \tau} = 0$. This explains the consistent deviation of single durations from the predicted and locally constant $\overline{D}_\tau = 82$ kyr in the inset of Fig. 3bi.

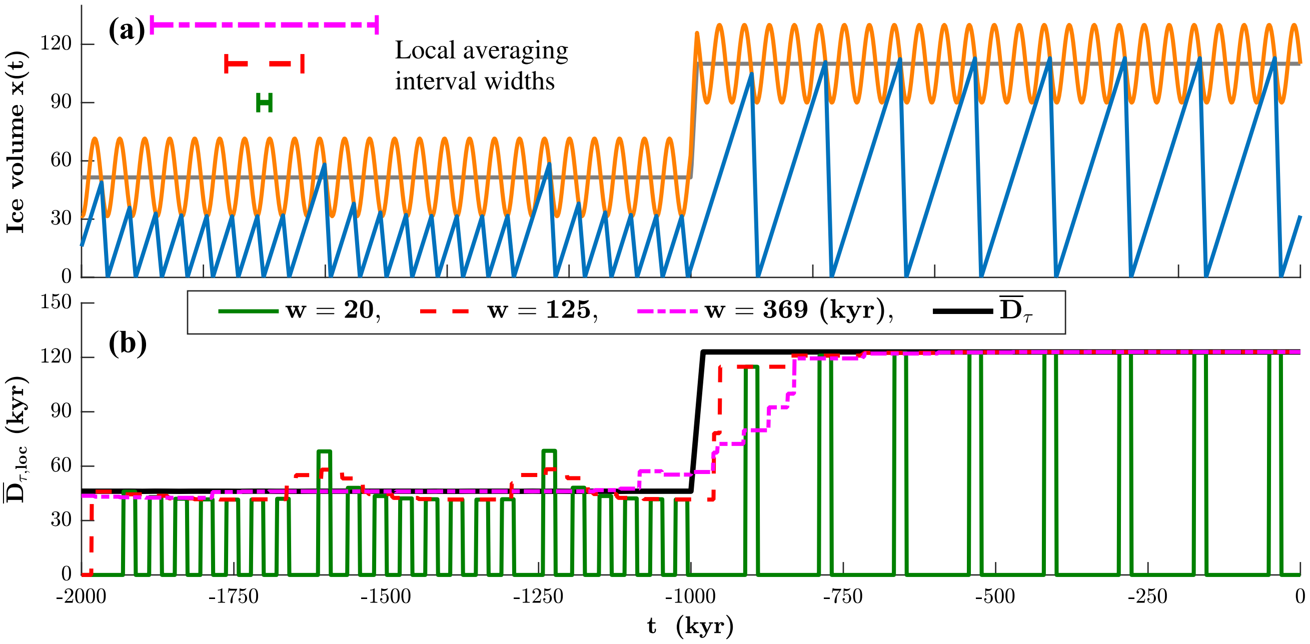


Fig. 6 Illustration of the difficulty estimating a local average frequency. **a** Ice volume from H07 (blue sawtooth), threshold of glacial termination (orange) and mean threshold of glacial termination $R(t)$ (grey). **b** Shows estimates of the local average duration $\overline{D}_{\tau,loc}$ between glacial terminations in the H07 model with periodic forcing ($T_f = 41$ kyr), for different window widths. **b** $\overline{D}_{\tau,loc}$ (green solid, red and magenta dashed) are running averages of durations in sliding windows of width $w = [20, 125, 369]$ kyr, or 0 kyr if there are no durations in a window. The frozen time average duration $\overline{D}_\tau$ (black solid) is shown for reference. Prior to − 990 kyr the steady state solution has a period of $9 \times 41 = 369$ kyr, but an average duration $\overline{D}_\tau = 46.125$ kyr. The mean threshold of glaciation $R(t)$ is ramped from 51.5 to 110 over 10 kyr (− 990 to − 1000 kyr). Other parameters are $\mu = 1, T_{decay}$ and $A = 20$

6 Is there a 100 kyr world?

The late Pleistocene (∼ − 800–0 kyr) is sometimes referred to as the “100 kyr world”, carrying the implicit notion that the Earth system has settled in a stationary mode with a dominant time scale of 100 kyr (Fig. 1a). This view, originating from the closeness to the 100 kyr component of eccentricity (an astronomical parameter), is supported by the rate of increase of mean ice volume seemingly levelling off (Clark et al. 2006; Mudelsee and Schulz 1997), and that the Fourier spectrum over the last ∼ 800 kyr is centred around 100 kyr.

We propose on the contrary, following (Huybers 2007), that the glacial period increased gradually from ∼ 80 kyr around − 1200 kyr to ∼ 120 kyr at present day. The change from ∼ 40 to ∼ 80 kyr long cycles at − 1200 kyr can be a shift from 1 × 41 to 2 × 41 kyr obliquity frequency locking, and/or 2 × 21 to 4 × 21 kyr precession locking. We base this claim on durations between major glacial terminations and a wavelet spectrum of the LR04 stack (see Fig. 1b); both quantities increase rather rapidly around − 1200 kyr and show a steady but irregular increase towards present time.

6.1 Identifying the shift to longer periods

While Huybers (2007) observed that the mean period of global ice volume variations increases over time, we make the stronger claim that an abrupt shift from 40 to 80 kyr long durations occurred around − 1200 kyr. We base this claim on our identification of major glacial terminations, which unlike spectral decomposition ignores glacial cycle shape and is unaffected by time–frequency resolution.

A disadvantage of using glacial termination events is that it is unclear what constitutes a major termination, and whether it is meaningful to characterise glacial cycles by termination events. Nevertheless, we believe that our identification of major terminations is sufficiently robust to support the claim that the duration shifted abruptly from ∼ 40 to ∼ 80 kyr around − 1200 kyr.

6.2 Testing for trend after the MPT

It appears that the durations between successive glacial terminations are increasing over time starting at the onset of the MPT around − 1200 kyr.

We evaluate whether this trend is statistically significant, using a variation on the Mann-Kendall test (Kendall 1955; Mann 1945). Our null hypothesis H0 is that the sequence of thirteen durations from − 1126 kyr until present is generated by a process with stationary mean, and that any observed monotonicity is by chance. Since $\tilde{D}_i = D_i - D_{mean}$, successive deviations from the mean duration $D_{mean} = 91$ kyr are correlated, we immediately reject a white noise process as assumed in the standard Mann-Kendall test. Instead, we model them as an AR(1) process, such that $\tilde{D}_{i+1} = \alpha\tilde{D}_i + \sigma_d\xi_i$, where ξ_i are independent Gaussian zero mean and unit variance elements. The parameters $\alpha = 0.6$ and $\sigma_d = 14.5$ kyr are the standard estimates of lag 1 and 0 autocorrelation coefficients, respectively.

We test the hypothesis using the Kendall τ test statistic for monotonicity τ_K, based on the number of ordered and disordered pairs in a sequence. $\tau_K = 1$ for a perfectly ordered sequence and $\tau_K = 0$ for sequence with equally many ordered and disordered pairs (see "Appendix A" for a definition of τ_K). We evaluate τ_K for $2 \cdot 10^4$ samples of the AR(1) process. As indicated in Fig. 7, it is unlikely ($p < 0.05$) to observe the test statistic in durations from data, assuming that the durations follow an AR(1) process. Therefore, we reject the null hypothesis of no trend.

Adding age model uncertainty to the Monte Carlo sequences of durations only makes it more difficult to reject H0. Furthermore, slightly different choices of major glacial terminations, or the use of an untuned record, does not influence the conclusion of the test.

6.3 Consequences for modelling the MPT

Some explanations for the MPT do not reproduce the sequence of successively longer durations between glacial terminations in data as naturally as RFL. Instead, they produce long period cycles at the onset of the MPT which shorten towards the present as a parameter is ramped.

The Maasch and Salzman 1990 model in Fig. 8 is one such model (Maasch and Salzman 1990). The inconsistency with data is evident when comparing the model durations with those in the LR04 stack (Fig. 1). Another such model

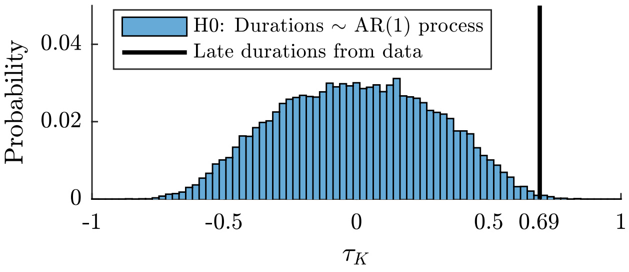


Fig. 7 Histogram shows a Monte Carlo distribution of the Kendall tau (τ_K) test statistic, under the null hypothesis H0 that the sequence of durations between glacial terminations from − 1126 kyr follow an AR(1) process. Larger τ_K indicates a more monotonic sequence. Black line shows the test statistic τ_K for durations in an ice volume proxy (Fig. 1). τ_K under H0 exceeds the observed τ_K only in 5% of the cases. For details, see Sect. 6.2

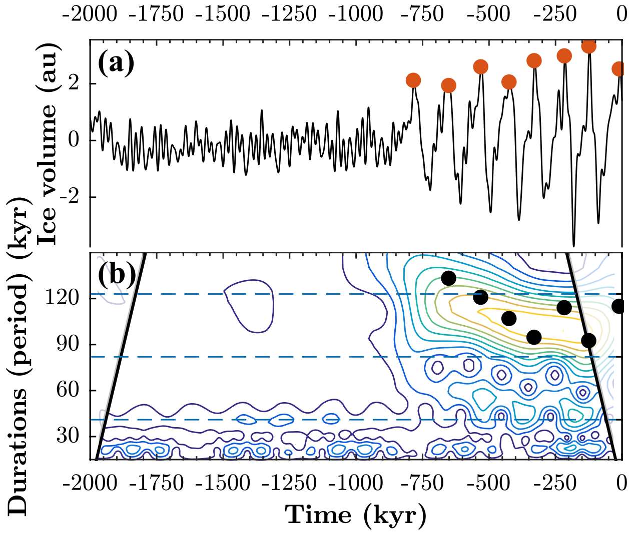


Fig. 8 Simulation of the Maasch and Salzman model in Maasch and Salzman (1990) forced by Summer solstice insolation at 65 degrees North. **a** Global ice volume over time (black), with glacial terminations (red dots) at peaks chosen for simplicity to be above 1.2 normalised ice volume units and spaced at least 60 kyr apart. Self-sustained cycles emerge around − 800 kyr and shorten towards the present. **b** Durations and wavelets as in Fig. 1, except that contours start at 10% of the maximum wavelet amplitude

is the Tziperman and Gildor 2003 model (Tziperman and Gildor 2003).

Although different dynamical mechanisms are at play in these models, they have in common that a long period limit unforced cycle emerges near a region of slow motion in phase space. As a parameter is varied, the limit cycle moves farther from this region, shortening the internal period.

RFL on the other hand naturally explains both a sudden shift from 40 to 80 kyr cycles and a gradual increase towards longer cycles, since the system can respond both smoothly and abruptly to an increasing internal period, due to the Devil's staircase structure (e.g. Fig. 4). We interpret the progression of durations as evidence against models like Maasch and Salzman 1990 and Tziperman and Gildor 2003, and for mechanisms that naturally produce increasing glacial cycle length, such as RFL.

6.4 Eccentricity and RFL

According to one view, the "100"-kyr world constitutes the time when δ^{18}O is strongly associated with eccentricity, an astronomical parameter that modulates the amplitude of precession at periods ~ 100 kyr and ~ 400 kyr. Previous studies have found evidence both for Lisiecki (2010), Rial (1999) and Rial et al. (2013) and not for Huybers (2007) a significant relationship between eccentricity and δ^{18}O over the past approximately 1000 kyr. We investigate whether such a relation is consistent with RFL.

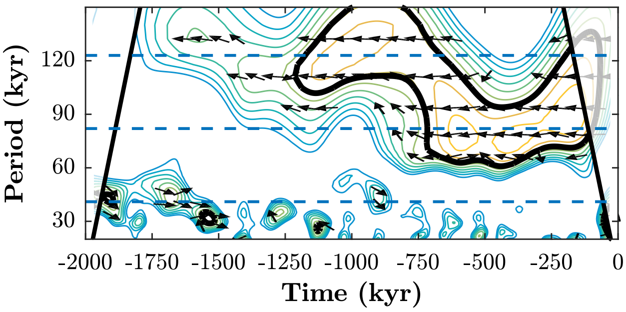


Fig. 9 Wavelet coherence between eccentricity and the LR04 stack (see "Appendix C"). Rightward arrows indicate that the signals are in-phase and leftward that they are anti-phase. Thin contour lines indicate a 7% increase in coherence, starting at 0.3 of the maximum coherence, which is 1. The thick contour is the 95% confidence interval against red noise processes (see "Appendix C"). Outside the cone of influence (thick black side lines), edge effects are important

Wavelet coherency offers a way to assess phase relations between the LR04 stack and eccentricity without apriori choosing frequency bands to compare ("Appendix C"). It is a smoothed measure of local correlation in time-frequency space and is normalised by the local power of the time series (Grinsted et al. 2004; Torrence and Compo 1998). Thus, large cross-wavelet power does not result in large coherency only by virtue of sharing a time scale, which trivially results from two time series varying strongly on a ~ 100 kyr time scale (Maraun and Kurths 2004). The wavelet coherence shows a clear anti-phase association at a central period, which meanders between ~ 80 and ~ 120 kyr starting ~ 1500 to ~ 1200 kyr ago (Fig. 9). Note that the wavelet amplitude of the LR04 stack is weak at periods longer than 80 kyr around time − 1000 kyr (Fig. 1), so the coherence is not with the dominant mode of variability. The general picture is similar for the untuned H07 stack (Huybers 2007).

Is the anti-phase coherence between ice volume and eccentricity compatible with RFL? Judging from Fig. 9, the coherence around − 1200 kyr ago is the most puzzling, since it is centred on period ~ 120 kyr, while durations between glacial terminations are only ~ 80 kyr long. Therefore, the coherence must have another source, perhaps related to glacial cycle shape. Evaluating the cross-wavelet transform for two non-RFL models (Ashwin and Ditlevsen 2015; Maasch and Salzman 1990), and three RFL models (Paillard 1998; Paillard and Parrenin 2004; Feng and Bailer-Jones 2015) (the best-fit), we find that only the (Paillard 1998) model somewhat faithfully reproduces the phase coherence with eccentricity (not shown). The non-RFL models show negligible coherence with eccentricity after the MPT, which does not improve even if the forcing amplitude is increased relative to the published values. The RFL models (Paillard and Parrenin 2004; Feng and Bailer-Jones 2015) show some coherence, but only strongly after − 500 kyr. Although the model in Paillard (1998) is anti-phase coherent with eccentricity, it fails to reproduce the sequence of durations between glacial terminations across the MPT. It remains a challenge to formulate a model which both reproduces the sequence of durations and coherency with eccentricity. This should be possible for a model using RFL, but such a model must produce ice volume variability on the ~ 120 kyr time scale at the MPT, in addition to the variations with period ~ 80 kyr associated with glacial terminations.

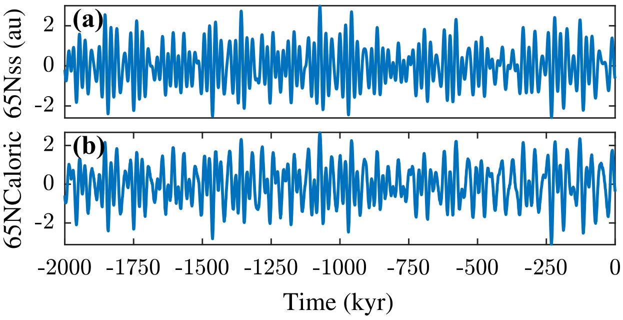


Fig. 10 Astronomical insolation curves. **a** Summer solstice insolation at 65 degrees North (65Nss), normalised to zero mean and unit variance (Laskar et al. 2004). The signal is approximately a linear combination of 33% normalised obliquity and 77% normalised precession, two modulated sinusoidal signals with central frequencies 41 and 22 kyr (Crucifix 2013). **b** Caloric summer insolation at 65N, consisting of roughly 50% obliquity and 50% precession. Precession is amplitude modulated by eccentricity, a signal with dominant periods ~ 100 and ~ 400 kyr

7 Non-harmonic forcing

RFL is not restricted to harmonic forcing, but occurs also for astronomical, non-harmonic forcing. This is for instance the case for the Paillard and Parennin 2004 model (Fig. 11), forced by summer solstice insolation at 65 degrees North (65Nss, Fig. 10a). As a parameter is increased linearly, durations first cluster around 41 kyr, then shift abruptly to cluster around 82 kyr at − 1000 kyr, after which they increase gradually until present. The shift to 80 kyr durations is later than in proxy data (Fig. 1) and there are some short and long durations not clear in the proxy record, but overall the glacial terminations coincide well.

Multi-frequency forcing generally produces Devil's staircases with shorter steps of constant duration, making them look "smooth" (e.g. Fig. 12). This is apparently a problem for RFL since it relies on rapid jumps in durations. However, RFL can still be relevant as demonstrated by H07 forced by caloric summer insolation at 65N consisting of an equal amount of obliquity and precession (Figs. 10b, 12, 13) (Tzedakis et al. 2017). The median and mode of the distribution of durations change more abruptly than the mean, which reflects that the gradual increase in average duration is caused by a gradual redistribution of durations between

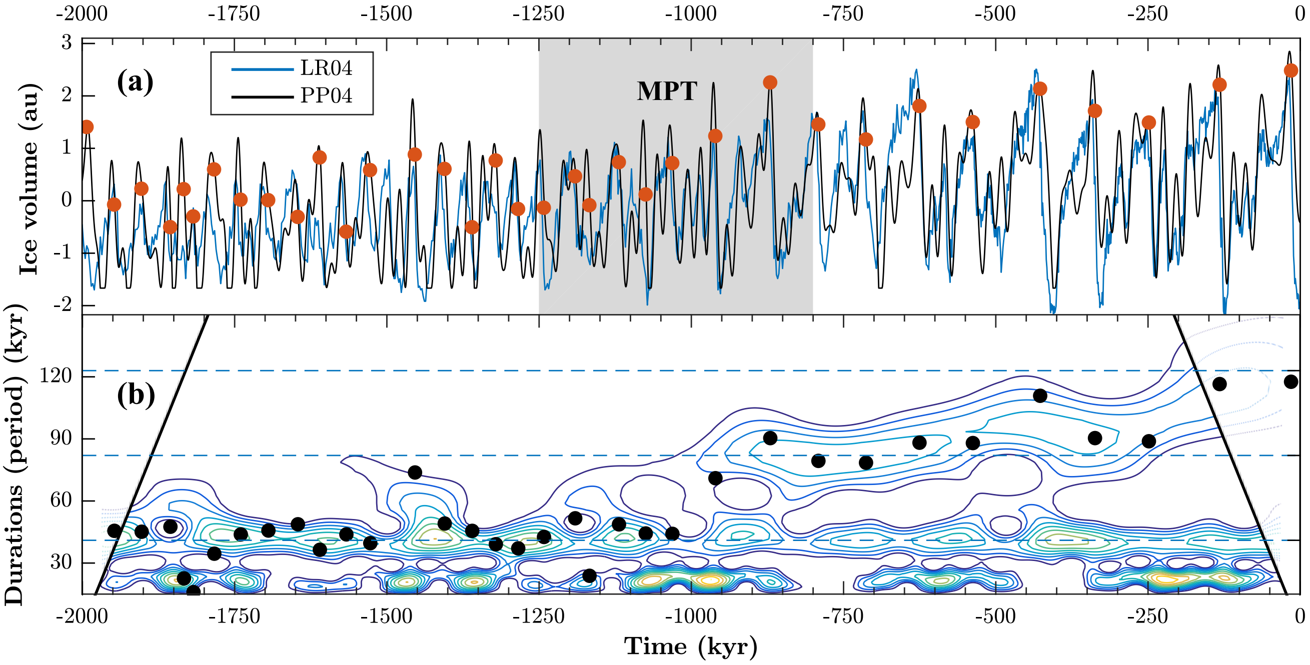


Fig. 11 Simulation of the Paillard and Parennin 2004 model in Paillard and Parrenin (2004) forced by Summer solstice insolation at 65 degrees North (Fig. 10a). **a** Model ice volume over time (black) contrasted with the LR04 stack (blue) (Fig. 1), with glacial terminations (red dots) at times when a switch in Southern ocean circulation occurs. **b** Durations and wavelets as in Fig. 1

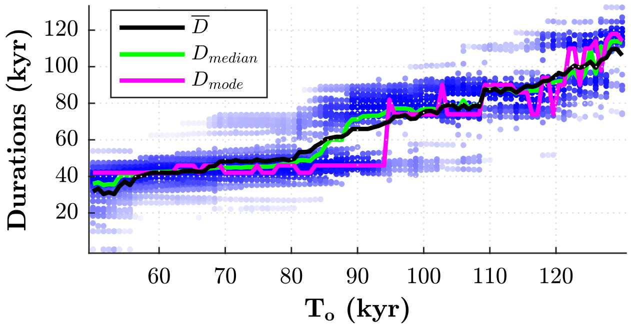


Fig. 12 Devil's staircases of H07 forced by caloric summer insolation (equal amounts of obliquity and precession, Fig. 10b). The average duration $\overline{D}$ (black line) as function of internal period T_o is gradually increasing, whereas the median D_{median} and the mode D_{mode} are more step-like. Blue dots are the population of durations for fixed internal period; darker colours indicate higher density of durations. D_{mode} is defined from binning the durations; D_{mode} is the mean of the edges of the 4-kyr bin with the highest frequency. All quantities are evaluated from − 2000 kyr to the present. Model parameters as in Fig. 13

clusters, rather than a gradual increase of the most typical durations. In a simulation with time-dependent ramping parameter, the local-in-time distribution of durations cannot be sampled well. Therefore the majority of the realised durations come from the dominant clusters of durations, which can give the impression that durations shift rapidly, in spite of the average duration changing gradually (Figs. 12, 13).

Multi-frequency forcing gives rise to many interesting phenomena regarding predictability of solutions, see for instance (Ashwin et al. 2018; Crucifix 2013; Grebogi et al. 1984; Imbrie and Imbrie 1980; Le Treut and Ghil 1983; Mitsui et al. 2015; de Saedeleer et al. 2013; Tziperman et al. 2006). Importantly, however, these phenomena are not essential to RFL. Whether solutions are truly frequency locked or depend on initial conditions is irrelevant, as long as durations undergo abrupt change and tend to cluster.

We conclude that RFL, clearly understood under periodic forcing, also is relevant for astronomical forcing. Indeed, recent studies provide evidence for the long-standing hypothesis that a combination of precession and obliquity paces the glacial cycles (Feng and Bailer-Jones 2015; Huybers 2011; Tzedakis et al. 2017). Differences between periodic and multi-frequency forcing exist, but are not crucial for modelling the MPT with RFL.

8 Relevance for complex models and physical mechanisms

We see two practical uses of our description of RFL: to guide modelling of the MPT in complex models, and to drive the search for slowly changing climate variables.

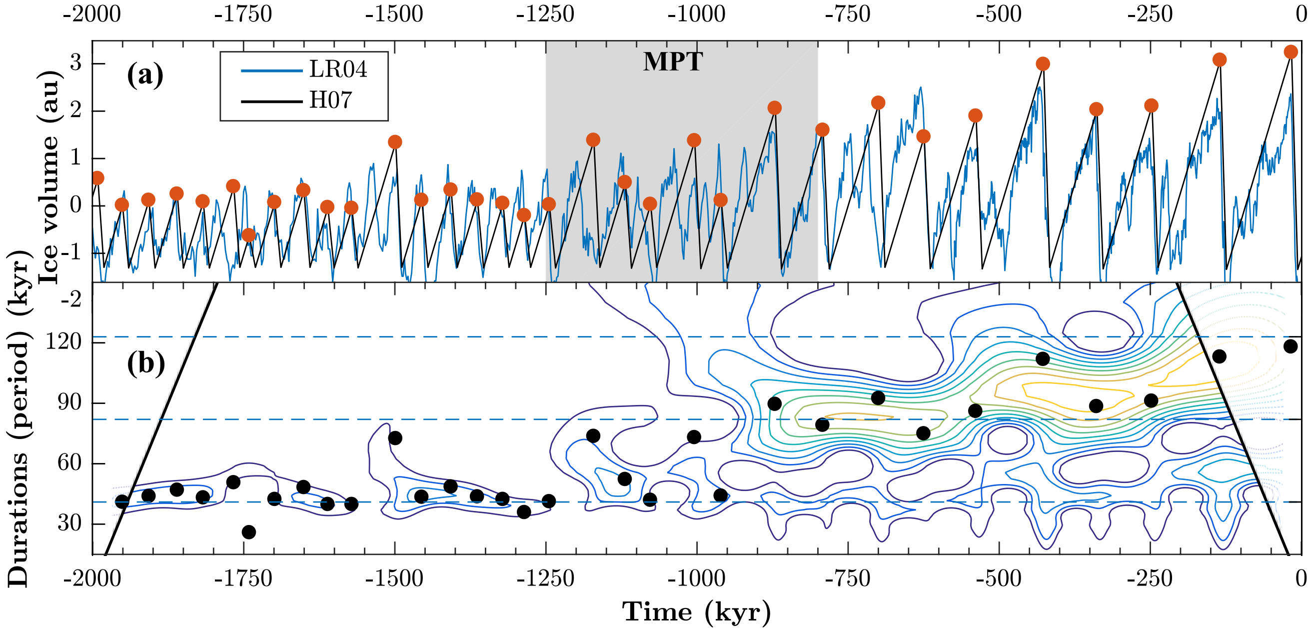


Fig. 13 Simulation of the H07 model forced by caloric summer insolation at 65N, a sum of equal amounts obliquity and precession (Fig. 10b). **a** Model ice volume (black) shown with the LR04 stack (blue) (Fig. 1), with glacial terminations (red dots) at times when a threshold of deglaciation is reached. **b** Durations and wavelets as in Fig. 1. The threshold of deglaciation increases linearly as $R(t) = 40 + 0.04(t + 2000)$, and forcing amplitude is $A = 26$. All other parameters are as in Sect. 2

8.1 Relevance for complex models

While climate physics are highly simplified in conceptual models like H07, their dynamics are well understood. The opposite holds true for earth system models (ESMs), which resolve multiple processes of climate in detail. To learn if the dynamical mechanism of RFL applies to such a model, we could in theory produce an Arnold tongue diagram as for H07 (Fig. 4). However, since running ESMs is computationally expensive this is presently not possible. Nevertheless, it might be possible to detect signatures of RFL from only few model runs.

First, one should investigate whether the glacial cycles are self-sustained by fixing model parameters at plausible values and fixing the insolation field at its mean value. This is the case if, after a transient time, variations in ice volume on the order of 10–100 kyr persist.

The next step is to sparsely sample an Arnold tongue diagram. First, a ramping parameter must be chosen. This does not have to be a scalar, but can be a function like a parametrisation, as long as its change over time is well defined. The parameter should be one that feasibly could influence the internal period of glacial cycles.

If changing the parameter changes the internal period, then one can compare the average duration of (insolation variation) forced and unforced solutions. If the average duration of the forced solutions is close to either 40 or 80 kyr and remains close even under parameter perturbations that change the internal period, then this is an indication that the system is frequency locked to insolation in a way relevant for the glacial cycles. In that case, there is good reason to research RFL more closely in the model.

Ashkenazy (2006) suggested that synchronisation can be detected by running the system from multiple initial conditions and see if solutions converge. This procedure is not enough for us; we need to know if the internal period can be shifted appropriately with a change in parameter, and we need to know if the durations can robustly cluster on 40 and 80 kyr.

8.2 Ramped climate variables

To evaluate whether RFL caused the MPT one must identify slowly changing climate variables. Two such candidate variables are atmospheric CO_2 and atmospheric or oceanic temperatures. Since less CO_2 leads to a generally cooler atmosphere, it can be viewed as a proxy for global average atmospheric temperature. Local cooling can occur for other reasons, however.

There are currently no direct measurements of atmospheric CO_2 across the MPT, but a recent reconstruction back to – 2000 kyr suggests that the mean CO_2 did not change in the mean until at least – 1300 kyr (Hönisch et al. 2009). Since the reconstruction implies that CO_2 fell 31 ppm by – 700 kyr, the decrease in CO_2 must either have been rapid and driving the MPT, or a consequence of it. A rapid change in CO_2 is still consistent with RFL, but in that case RFL does not explain the abrupt increase in cycle length at the MPT, and instead one must find an explanation for the rapid increase in CO_2. However, the planned European BEOIC deep ice core drilling in Antarctica can hopefully improve estimates of CO_2 across the MPT.

There is evidence of a gradual deep ocean cooling since the onset of northern Hemisphere glaciation 2.7 Myr ago (Lisiecki and Raymo 2005). How much of this cooling occurred across the MPT is not known, however. The reconstruction of deep water temperatures by Elderfield et al. (2012) indicates a gradual cooling in the mean from – 1300 kyr until present, but also a puzzling warming from – 1500 to – 1300 kyr. Therefore, glacial cycle length does not appear to have a direct relation with mean deep ocean temperature. However, it may be that sea surface temperatures in the vicinity of major ice sheets are more relevant for glacial dynamics. If so, detailed and reliable reconstruction of such temperatures is necessary to evaluate whether they act as ramped climate variables in RFL.

Another slowly varying parameter could be the erosion of regolith. According to this hypothesis soft material under ice sheets eroded throughout the Pleistocene, enabling them to grow larger before collapsing Clark and Pollard (1998). The hypothesis is difficult to test empirically, however.

In addition to the candidate ramping climate variables mentioned, there may be others that are relevant for the MPT. RFL motivates the search for other such climate variables. These might not only be relevant for RFL, but for any mechanism of the MPT invoking deterministic bifurcation.

9 Criteria for RFL

Having demonstrated RFL in H07 and Paillard and Parennin 2004, we ask in which models RFL is most likely to be relevant.

We expect RFL in all models similar to H07, that is, models with a critical threshold of deglaciation (explicit or not), two intrinsic growth and decay states, additive forcing, and a climate variable that naturally controls the internal period.

Furthermore, RFL is facilitated by dynamics focussed on a single strongly attracting limit cycle. This is because solutions to the system f with ramped parameters then track frozen solutions of f_τ well, and because it is difficult for perturbations to bring solutions away from the neighbourhood of the attractor.

Crucially, a model using RFL needs a parameter that can increase the internal period by 100 kyr. The models in e.g. (Le Treut and Ghil 1983 and Maasch and Salzman 1990) are therefore difficult to reconcile with RFL since the internal periods are on the order of 10 and 100 kyr respectively, and do not change much within the physical range of model parameters.

Lastly, we note that e.g. excitable systems and dissipative resonant oscillators (Crucifix 2012) also can undergo a rapid change in durations due to frequency locking related phenomena, although they are not self-sustained oscillators. Self-sustained oscillators are distinguished by having an internal period $T_o(R)$ through which we can define Arnold tongue diagrams and Devil's staircases; for non-self-sustained oscillators we have to define these through parameters R directly. Furthermore, it has been argued that the term frequency locking should be restricted to self-sustained oscillators (Marchionne et al. 2018; Pikovsky et al. 2001), why it makes sense to define RFL for self-sustained oscillators only.

10 Conclusions

The glacial cycles did not enter a stationary 100 kyr world at the MPT; instead, durations between glacial terminations shifted abruptly from ~ 40 to ~ 80 kyr around − 1200 kyr, followed by a gradual increase (Fig. 1). The dynamical mechanism ramping with frequency locking (RFL) naturally explains this progression of durations. As the internal period of a model glacial cycle model increases gradually, frequency locking to insolation variations causes the durations between glacial terminations to increase sometimes abruptly and sometimes gradually.

The RFL mechanism is rather general and explains the behaviour of a range of models describing glacial cycles and the MPT (Ashkenazy 2006; Crucifix et al. 2011; Feng and Bailer-Jones 2015; Huybers 2007; Mitsui et al. 2015; Paillard 1998; Paillard and Parrenin 2004; Tzedakis et al. 2017).

Here we described how RFL can be understood in terms of a dynamical system f and a frozen system f_τ with parameters $R(t)$ fixed at times $t = \tau$. The average duration $\overline{D}_\tau$ defined for f_τ provides some information about single durations in solutions $x(t)$ to f around $t = \tau$, but since $\overline{D}_\tau$ is defined asymptotically, one must interpret solutions to f in terms of f_τ with care.

Model behaviour can be understood from considering parameter paths through Arnold tongue diagrams and corresponding Devil's staircases (Figs. 3, 4, 5). These diagrams as functions of frozen time τ depend on

- the change in average duration $\overline{D}_\tau(T_o)$ as function of T_o,
- the change in internal period $T_o(R)$ as function of R,
- the change in parameters $R(\tau)$ as function of time τ, and
- the amplitude A of the forcing.

This decomposition clarifies different ways in which the average duration can change abruptly in models of the class f. For instance, the abruptness of the change in $\overline{D}_\tau$ can be adjusted either by changing the forcing amplitude A or the ramping of $R(t)$. While the effects of changing A or the ramping of $R(t)$ are typically easy to guess, we are not aware of any general rules dictating the widths of particular Arnold tongues. Such understanding may be researched further.

RFL is relevant also for multi-frequency astronomical forcing. Multi-frequency forcing tends to make Devil's staircases less abrupt, but durations can still increase rapidly when a model parameter is slowly ramped.

It remains a challenge to reconcile the phase coherence between eccentricity and ice volume proxies at the MPT with the ~ 80 kyr durations between glacial terminations. The RFL model Paillard 1998 reproduces well the phase coherence, but not the durations. Two RFL models and two non-RFL models with a strong self-sustained limit cycles forced by precession-heavy forcing (modulated by eccentricity) fail to reproduce the phase coherency.

The RFL mechanism provides an explanation for the MPT without the climate system entering a new mode of operation. A shift from ~ 40 kyr long to ~ 80 kyr long cycles due to frequency locking to obliquity and precession, is consistent with data (Fig. 1), and is used in models (Huybers 2007; Paillard and Parrenin 2004). This warrants further study of frequency locking characteristics of models throughout the model hierarchy, as well as a search for gradually increasing climate parameters. Some models use a rapidly ramped parameter to accelerate the increase in durations between glacial terminations at the MPT, but such a ramping begs for justification that a model relying solely on frequency locking does not require.

Acknowledgements We thank Peter Ashwin for valuable discussions and three anonymous reviewers for constructive comments. Insolation codes were adapted from the `palinsol` R package by M. Crucifix. Matlab code for evaluating wavelet coherency was provided by A. Grinsted. This research has been funded by the European Union's Horizon 2020 innovation and research programme for the ITN CRITICS under the Marie Skłodowska-Curie Grant agreement no. 643073.

Appendix A: Kendall's tau

Kendall's tau, here denoted τ_K, when testing for monotonicity of a sequence $\{D_i\}_{i=1}^n$ is defined as:

$$\tau_K = \frac{n_c - n_d}{\sqrt{n_0^2 - n_0 n_1}},$$

where $n_c - n_d = \sum_{i<j} \operatorname{sign}(D_j - D_i)$ is the number of pairs (D_i, D_j) that are ordered $(D_j > D_i)$ minus the number that is disordered, $n_0 = n(n-1)/2$ is the total number of pairs and $n_1 = \sum_k t_k(t_k - 1)/2$ is the sum of the number of tied elements t_k in the k:th group of tied elements. For example, the sequence $\{1, 2, 2\}$ has two ordered pairs (1, 2) and (1, 2), zero disordered pairs, and one tied pair (2, 2). Hence, $k = 1$ such that $n_c - n_d = 2$, $n_0 = 3$ and $n_1 = 1$, giving $\tau_K = 2/\sqrt{6} \approx 0.82$.

Appendix B: Wavelets

Wavelet spectra are estimated with the MATLAB function `cwt`, using Morlet basis functions with bandwidth parameter $\omega_0 = 6$ (Torrence and Compo 1998), softwaree by. Contours show wavelet amplitude (square root of variance) relative to the maximum, incremented in evenly spaced percentage units. The cone of influence marks the e-folding time of the amplitude of a discontinuity at the edge of the time interval. Inside the cone of influence edge effects are negligible (Torrence and Compo 1998).

Appendix C: Cross-wavelet transform and wavelet coherency

The cross-wavelet transform of two time series X and Y is given by $W_n^{XY}(s) = W_n^X(s)W_n^{Y*}(s)$, where $W_n^X(s)$ is the wavelet transform of X at discrete time n and scale s, and $W_n^{Y*}(s)$ is the complex conjugate of the wavelet transform of Y. The wavelet squared coherency is defined as:

$$R_n^2(s) = \frac{|S(s^{-1}W_n^{XY}(s))|^2}{|S(s^{-1}W_n^X(s))|^2 |S(s^{-1}W_n^Y(s))|^2},$$

where S is a smoothing function in space and time, found in Grinsted et al. (2004) and Torrence and Webster (1999), and s is scale (approximately equal to period).

Significance levels for wavelet coherency between X and Y was estimated Monte Carlo by estimating the wavelet coherency between a large number (1000) of red noise processes with parameters fitted from X and Y. The confidence levels are reliable if at least one of X and Y can be modelled as a red noise process (Maraun and Kurths 2004), which holds decently for the LR04 stack (Fig. 9). (Maraun and Kurths (2004) considered a white noise process, but we expect the conclusion to hold also for a red noise process). Large coherent regions in time-period space are more likely non-spurious than small ones (Maraun and Kurths 2004).

Appendix D: Glacial terminations in LR04

Major glacial terminations in the LR04 stack (Fig. 1) are identified at times $t = -$ [1948, 1900, 1863, 1795, 1748, 1708, 1655, 1575, 1535, 1496, 1456, 1412, 1372, 1336, 1290, 1248, 1198, 1126, 1038, 964, 876, 794, 718, 630, 536, 434, 341, 252, 140, 18] kyr. We assume conservatively an age model uncertainty with constant standard deviation 6 kyr over the past 2000 kyr (Lisiecki and Raymo 2005), which gives a standard deviation $\sqrt{2} \cdot 6$ kyr on the durations between terminations, assuming somewhat wrongly that errors are independent and normally distributed.

References

Ashkenazy Y (2006) The role of phase locking in a simple model for glacial dynamics. Clim Dyn 27:421–431. https://doi.org/10.1007/s00382-006-0145-5

Ashkenazy Y, Tziperman E (2004) Are the 41 kyr glacial oscillations a linear response to Milankovich forcing? Quat Sci Rev 23:1879–1890. https://doi.org/10.1016/j.quascirev.2004.04.008

Ashwin P, Ditlevsen P (2015) The middle Pleistocene transition as a generic bifurcation on a slow manifold. Clim Dyn. https://doi.org/10.1007/s00382-015-2501-9

Ashwin P, Camp CD, von der Heydt AS (2018) Chaotic and non-chaotic response to quasiperiodic forcing: limits to predictability of ice ages paced by Milankovitch forcing. Dyn Stat Clim Syst. https://doi.org/10.1093/climsys/dzy002

Baer S, Ernaux T, Rinzel J (1989) The slow passage through a hopf bifurcation: delay, memory effects, and resonance. SIAM J Appl Math 49(1):55–71. https://doi.org/10.1137/0149003

Berger AL (1978) Long-term variations of daily insolation and quaternary climatic changes. J Atmos Sci. https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2

Cartwright M, Littlewood J (1945) On non-linear differential equations of the second order. J Lond Math Soc 1–20(3):180–189. https://doi.org/10.1112/jlms/s1-20.3.180

Clark PU, Pollard D (1998) Origin of the middle Pleistocene transition by ice sheet erosion of regolith. Paleoceanography 13(1):1–9. https://doi.org/10.1029/97PA02660

Clark PU, Archer D, Pollard D, Blum JD, Rial JA, Brovkin V, Mix AC, Pisias NG, Roy M (2006) The middle Pleistocene transition: characteristics, mechanisms, and implications for the long-term changes in atmospheric pCO$_2$. Quat Sci Rev 25:3150–3184. https://doi.org/10.1016/j.quascirev.2006.07.008

Crucifix M (2012) Oscillators and relaxation phenomena in Pleistocene climate theory. Philos Trans R Soc Lond A 370(1962):1140–1165. https://doi.org/10.1098/rsta.2011.0315

Crucifix M (2013) Why could ice ages be unpredictable? Clim Past 9:2253–2267. https://doi.org/10.5194/cp-9-2253-2013

Crucifix M, Lenoir G, de Saedeleer B (2011) The mid-Pleistocene transition and slow fast dynamics. In: EGU2011-3629-1, EGU general assembly 2011, geophysical research abstracts, vol 13, poster

Daruka I, Ditlevsen PD (2015) A conceptual model for glacial cycles and the middle Pleistocene transition. Clim Dyn 46:29–40. https://doi.org/10.1007/s00382-015-2564-7

Ditlevsen PD (2009) Bifurcation structure and noise-assisted transitions in the Pleistocene glacial cycles. Paleoceanography. https://doi.org/10.1029/2008PA001673

Do Y, Lopez JM (2012) Slow passage through multiple bifurcation points. Am Inst Math Sci 18(1):95–107. https://doi.org/10.3934/dcdsb.2013.18.95

Elderfield H, Ferretti P, Greaves M, Crowhurst S, McCave IN, Hodell D, Piotrowski AM (2012) Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition. Science 337(6095):704–709. https://doi.org/10.1126/science.1221294

EPICA Community Members (2004) Eight glacial cycles from an Antarctic ice core. Nature 429:623–628. https://doi.org/10.1038/nature02599

Feng F, Bailer-Jones CAL (2015) Obliquity and precession as pacemakers of Pleistocene deglaciations. Quat Sci Rev 122:166–179. https://doi.org/10.1016/j.quascirev.2015.05.006

Gildor H, Tziperman E (2000) Sea ice as the glacial cycles' climate switch: role of seasonal and orbital forcing. Paleoceanography 15(6):605–615. https://doi.org/10.1029/1999PA000461

Glass L, Mackey MC (1979) A simple model for phase locking of biological oscillators. J Math Biol 7:339–352. https://doi.org/10.1007/BF00275153

Grebogi C, Ott E, Pelican S, Yorke JA (1984) Strange attractors that are not chaotic. Phys D 13:261–268. https://doi.org/10.1016/0167-2789(84)90282-3

Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Process Geophys 11(5/6):561–566. https://doi.org/10.5194/npg-11-561-2004

Guckenheimer J, Hoffman K, Weckesser W (2003) The forced van der pol equation I: the slow flow and its bifurcations. SIAM J Appl Dyn Syst 2(1):1–35. https://doi.org/10.1137/S1111111102404738

Huybers P (2006) Early Pleistocene glacial cycles and the integrated summer insolation forcing. Science 313:508–510. https://doi.org/10.1126/science.1125249

Huybers P (2007) Glacial variability over the last two million years: an extended depth-derived agemodel, continuous obliquity pacing, and the Pleistocene progression. Quat Sci Rev 26:37–55. https://doi.org/10.1016/j.quascirev.2006.07.013

Huybers P (2009) Pleistocene glacial variability as a chaotic response to obliquity forcing. Clim Past 5:481–488. https://doi.org/10.5194/cp-5-481-2009

Huybers P (2011) Combined obliquity and precession pacing of late Pleistocene deglaciation. Nature 480:229–231. https://doi.org/10.1038/nature10626

Huybers P, Langmuir CH (2017) Delayed CO_2 emissions from mid-ocean ridge volcanism as a possible cause of late-Pleistocene glacial cycles. Earth Planet Sci Lett 457:238–249. https://doi.org/10.1016/j.epsl.2016.09.0

Hönisch B, Hemming G, Archer D, Siddall M, McManus JF (2009) Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. Science 324(5934):1551–1554. https://doi.org/10.1126/science.1171477

Imbrie J, Imbrie JZ (1980) Modeling the climatic response to orbital variations. Science 207:943–953. https://doi.org/10.1126/science.207.4434.943

Imbrie J, Imbrie KP (1979) Ice ages: solving the mystery, 1st edn. MacMillan, London

Imbrie JZ, Imbrie-Moore A, Lisiecki L (2011) A phase-space model for Pleistocene ice volume. Earth Planet Sci Lett 307:94–102. https://doi.org/10.1016/j.epsl.2011.04.018

Kendall M (1955) Rank correlation methods, 2nd edn. Hafner Publishing Co., Oxford

Laskar J, Robutel P, Joutel F, Gastineau M, Correia ACM, Levrard B (2004) A long-term numerical solution for the insolation quantities of the earth. Astron Astrophys 428:261–285. https://doi.org/10.1051/0004-6361:20041335

Le Treut H, Ghil M (1983) Orbital forcing, climatic interactions, and glaciation cycles. J Geophys Res 88(C9):5167–5190. https://doi.org/10.1029/JC088iC09p05167

Levi M (1990) A period-adding phenomenon. SIAM J Appl Math 50(4):943–955. https://doi.org/10.1137/0150058

Lisiecki LE (2010) Links between eccentricity forcing and the 100,000-year glacial cycle. Nature Geosci 3:349–352. https://doi.org/10.1038/ngeo828

Lisiecki LE, Raymo ME (2005) A Plioene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. Paleoceanography 20:437–440. https://doi.org/10.1029/2004PA001071

Maasch KA, Salzman B (1990) A low-order dynamical model of global climatic variability over the full Pleistocene. J Geophys Res 95(D2):1955–1963. https://doi.org/10.1029/JD095iD02p01955

Mann HB (1945) Nonparametric tests against trend. Econometrica 13(3):245–259. https://doi.org/10.2307/1907187

Maraun D, Kurths J (2004) Cross wavelet analysis: significance testing and pitfalls. Nonlinear Process Geophys 11(4):505–514. https://doi.org/10.5194/npg-11-505-2004

Marchionne A, Ditlevsen P, Wieczorek S (2018) Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual study. Phys D 380–381:8–16. https://doi.org/10.1016/j.physd.2018.05.004

Mitsui T, Crucifix M, Aihara K (2015) Bifurcations and strange nonchaotic attractors in a phase oscillator model of glacial-interglacial cycles. Phys D 306:25–33. https://doi.org/10.1016/j.physd.2015.05.007

Mudelsee M, Schulz M (1997) The mid-Pleistocene climate transtion: onset of 100 ka cycles lags ice volume build-up by 280 ka. Earth Planet Sci Lett 151:117–123. https://doi.org/10.1016/S0012-821X(97)00114-3

Omta AW, Kooi BW, van Voorn GAK, Rickaby REM, Follows MJ (2015) Inherent characteristics of sawtooth cycles can explain different glacial periodicities. Clim Dyn 46:557–569. https://doi.org/10.1007/s00382-015-2598-x

Paillard D (1998) The timing of Pleistocene glaciations from a simple multiple-state climate model. Nature 391:378–381. https://doi.org/10.1038/34891

Paillard D, Parrenin F (2004) The Antarctic ice sheet and the triggering of deglaciations. Earth Planet Sci Lett 227:263–271. https://doi.org/10.1016/j.epsl.2004.08.023

Parrenin F, Paillard D (2003) Amplitude and phase of glacial cycles from a conceptual model. Earth Planet Sci Lett 214(1):243–250. https://doi.org/10.1016/S0012-821X(03)00363-7

Parrenin F, Paillard D (2012) Terminations vi and viii (530 and 720 kyr bp) tell us the importance of obliquity and precession in the triggering of deglaciations. Clim Past 8(6):2031–2037. https://doi.org/10.5194/cp-8-2031-2012

Pikovsky A, Rosenblum M, Kurths J (2001) Synchronization: a universal phenomenon in the nonlinear sciences, 1st edn. Cambridge University Press, Cambridge

van der Pol B, van der Mark J (1927) Frequency demultiplication. Nature 120:363–364. https://doi.org/10.1038/120363a0

Quinn C, Sieber J, von der Heydt AS, Lenton TM (2018) The mid-Pleistocene transition induced by delayed feedback and bistability. Dyn Stat Clim Syst 3(1):1–17. https://doi.org/10.1093/climsys/dzy005

Rial JA (1999) Pacemaking the ice ages by frequency modulation of earth's orbital eccentricity. Science 285(5427):564–568. https://doi.org/10.1126/science.285.5427.564

Rial JA, Oh J, Reischmann E (2013) Synchronization of the climate system to eccentricity forcing and the 100,000-year problem. Nat Geosci 6:289–293. https://doi.org/10.1038/ngeo1756

de Saedeleer B, Crucifix M, Wieczorek S (2013) Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual study. Clim Dyn 40:273–294. https://doi.org/10.1007/s00382-012-1316-1

Salzman B, Verbitsky MY (1993) Multiple instabilities and modes of glacial rhythmicity in the Plio–Pleistocene: a general theory of late Cenozoic climatic change. Clim Dyn 9:1–15. https://doi.org/10.1007/BF00208010

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79(1):61–78. https://doi.org/10.2307/1907187

Torrence C, Webster PJ (1999) Interdecadal changes in the enso-monsoon system. J Clim 12(8):2679–2690. https://doi.org/10.1175/1520-0442(1999)012<2679:ICITEM>2.0.CO;2

Tzedakis PC, Crucifix M, Mitsui T, Wolff EW (2017) A simple rule to determine which insolation cycles lead to interglacials. Nature 542:427–432. https://doi.org/10.1038/nature21364

Tziperman E, Gildor H (2003) On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. Paleoceanography 18(1):1–8. https://doi.org/10.1029/2001PA000627

Tziperman E, Raymo M, Huybers P, Wunsch C (2006) Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to milankovitch forcing. Paleoceanography 21:1–11. https://doi.org/10.1029/2005PA001241

Appendix B

Paper B

Bifurcation of critical sets and relaxation oscillations in singular fast-slow systems

Karl Nyman[1], **Peter Ashwin**[2] **and Peter Ditlevsen**[1]

1. Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark
2. Department of Mathematics, University of Exeter, Exeter EX4 4QF, UK

,

E-mail: `karl.nyman@nbi.ku.dk`

Abstract. Fast-slow dynamical systems have subsystems that evolve on vastly different timescales, and bifurcations in such systems can arise due to changes in any or all subsystems. We classify bifurcations of the critical set (the equilibria of the fast subsystem) and associated fast dynamics, parametrized by the slow variables. Using a distinguished parameter approach we are able to classify bifurcations for one fast and one slow variable. Some of these bifurcations are associated with the critical set losing manifold structure. We also present a conjectured a list of generic bifurcations of the critical set for one fast and two slow variables. We further consider how the bifurcations of critical set can be associated with generic bifurcations of attracting relaxation oscillations under and appropriate singular notion of equivalence.

Keywords: Fast-slow dynamics, Relaxation oscillation, Bifurcation, Singularity

1. Introduction

Many natural systems are characterized by interactions between dynamical processes that run at very different timescales. These can often be modelled as fast-slow systems, where system dynamics can be separated into interacting fast and slowly changing variables. This has been applied to a wide range of natural phenomena, from plasma oscillations[28], surface chemistry [22] and cell physiology [20] to ecology[32] and climate[3]. The dynamical behavior of such systems can often be understood in a common mathematical framework. See [24] for a recent monograph that summarizes both techniques and applications, and [2, 5, 6, 12, 14, 17, 19, 35] for examples of related work.

The analysis of fast-slow systems is built around a *geometric singular perturbation theory* (GSPT) approach, perturbing from a singular limit where timescales decouple. In the singular limit, on the slow timescale there are instantaneous jumps (determined by the fast dynamics) between periods of slow evolution. The slow evolution typically takes place on stable sheets of a *critical set* where the fast dynamics is in stable balance, interspersed by fast transitions. If the fast dynamics is one dimensional, then it is typically confined a manifold (and hence the set is often called a *critical manifold*), though at bifurcation it may lose its manifold structure at isolated points. The fast transitions are determined by what we call the *umbral map* defined as the map from a fold point on the critical set to another part of the critical set. In the case of stable periodic behaviour, the resulting limit cycles are referred to as *relaxation oscillations*. Many examples of bifurcations of relaxation oscillations have been considered [2], including some associated with bifurcations of the critical set [3] although it seems that no exhaustive list of scenarios has been proposed.

Although singular perturbation theory has been developed to explain many aspects of behaviour near the singular limit, there is still no full understanding of generic bifurcations of limit cycles even in the singular limit. Guckenheimer [14, 15, 16] suggests an approach and several results along these lines, but, as far as we are aware, these conjectures are yet to be framed, let along proved, in rigorous terms. The main aim of this paper is to present an approach to doing precisely this, using singularity theory with distinguished parameters and appropriate notions of equivalence. We classify local and global transitions in the critical set by codimension and consider the consequences for the *umbral map*. In doing so, we find a variety of scenarios that give bifurcation of attractors in such singular systems.

We show that it is possible to split the problem of bifurcations of relaxation oscillations into two aspects: bifurcations of the critical set, and bifurcations caused by singularities of the slow flow (possibly interacting with the critical set). In the simplest case of one fast and one slow variable, bifurcations of the critical set can be directly tackled using a global version of the singularity theory with distinguished parameter in [13]. We highlight that this theory needs extension to make it suitable for systems with multiple fast and slow variables. We are able to identify a large number of scenarios that can lead to bifurcation of relaxation oscillations. Note that we only consider fast-slow systems where the fast dynamics is constrained to a subset of the variables; we suggest that it will be useful to extend the theory developed here to a more general fast-slow systems that are not in standard form: see for example [23, 25, 38].

We structure the paper as follows: in Section 2 we briefly introduce the singular limit of fast-slow systems, critical sets, singular trajectories and global equivalence of critical sets. In Section 3 we explore persistence and bifurcation of critical sets by examining versal unfoldings of the fast dynamics parametrized by the slow variables, using a notion of global equivalence of the fast dynamics. In the case of one fast and one slow variable we classify persistence (Proposition 1) and bifurcations of the critical set up to codimension one (Proposition 2) using the theory of [13]. For one fast and several slow variables we highlight the need for an improved theory of bifurcations with multiple distinguished parameters. We present conjectured statements of persistence and of codimension one bifurcations of the critical set (Conjectures 1 and 2 respectively) for one fast and two slow variables. We find a rich variety of distinct mechanisms for typical codimension one bifurcations of the critical set which includes local and global bifurcations in the fast variable.

Section 5 turns to the question of bifurcation of attractors in fast-slow systems and in particular bifurcation of stable relaxation oscillations. We introduce a global singular equivalence for the singular trajectories and use this to classify persistence (Proposition 4) and codimension one bifurcations (Proposition 5) of these simple relaxation oscillations. These codimension one bifurcations naturally split into those caused by bifurcations of the critical set, and those caused by interaction of singularities of the slow flow with the critical set: in Section 5.3 we present some numerical examples of various types. Finally, Section 6 is a discussion of some of the challenges for GSPT to describe the unfolding of such bifurcations, and relation to other singularity theory approaches. We include several Appendices that give more details of the tools used for the classification and the examples.

2. Singular trajectories of fast-slow systems

limitintro

A *fast-slow system* is a system of coupled ODEs for $z = (x, y) \in \mathbb{R}^{m+n}$ of the form

$$\begin{cases} \epsilon\dot{x} &= g(x,y) \\ \dot{y} &= h(x,y) \end{cases} \tag{1}$$

eq:mainsystem

where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, $\epsilon > 0$ is a small constant and t is a time measure relative to the slow dynamics. The functions $g(x, y)$ and $h(x, y)$ are C^∞ functions of their arguments (they are well approximated by Taylor series to arbitrary order). We refer to x as the *fast* and y as the *slow* subsystems respectively. These systems have a singular limit $\epsilon \to 0$, where a typical trajectory can remain close to an equilibrium of the fast system, except at isolated points where it "jumps" along a trajectory of the fast subsystem. The singularly perturbed system with $\epsilon > 0$ will have trajectories that typically remain near a trajectory of the singular limit, although especially near bifurcations, trajectories may also explore unstable parts of the slow dynamics along so-called *canard solutions* (see e.g. [5, 2, 24, 35]).

In order to understand such systems it useful to consider the *reduced* or *slow equations* in slow time t:

$$\begin{cases} 0 &= g(x,y) \\ \dot{y} &= h(x,y). \end{cases} \tag{2}$$

eq:reduced

Solutions of (2) are constrained to the *critical set*

$$\mathcal{C}[g] = \left\{(x, y) \in \mathbb{R}^{m+n} : g(x, y) = 0\right\}. \tag{3}$$

eq:critman

Note that by the regular value theorem [27], this critical set is a manifold at all points where the derivative of g has maximal rank, and this is true for an open dense set of g. The set is often called a critical manifold, however we do not use this notation as at bifurcation points the set may lose its manifold structure.

The flow g may have singular equilibria of the fast dynamics within $\mathcal{C}[g]$. We define the *regular points* of the critical set $\mathcal{C}[g]$

$$\mathcal{C}_{reg}[g] = \{(x,y) \in \mathcal{C}[g] \ : \ \partial_x g(x,y) \text{ is hyperbolic}\} . \tag{4}$$

The remaining non-hyperbolic (fold) points form the *fold set* of the critical set

$$\mathcal{F}[g] = \{(x,y) \in \mathcal{C}[g] \ : \ \partial_x g(x,y) \text{ is non-hyperbolic}\} . \tag{5}$$

At all regular points the reduced equations (2) can be used to describe the flow. At fold points, however, we need to consider the fast dynamics and there will be *jumps* in slow time determined by the fast subsystem only.

Changing variable to a *fast time* $\tau = t/\epsilon$ and taking the limit $\epsilon \to 0$ gives quite a different set of equations: the *layer* or *fast equations*:

$$\begin{cases} x' &= g(x,y) \\ y' &= 0, \end{cases} \tag{6}$$

`eq:layer`

were we write x' to denote $\frac{d}{d\tau}x$ and note that (a) the constant slow variable y now acts as a parameter for evolution of the fast variable x and (b) the layer equation, when restricted $\mathcal{C}[g]$ consists entirely of equilibria for $m = 1$ (for $m > 1$ there may be other objects, such as limit cycles). We split the regular set into a disjoint union of attracting/repelling/saddle points

$$\mathcal{C}_{att}[g], \ \mathcal{C}_{rep}[g], \ \mathcal{C}_{sad}[g]$$

so that $\mathcal{C}_{reg}[g] = \mathcal{C}_{att}[g] \cup \mathcal{C}_{rep}[g] \cup \mathcal{C}_{sad}[g]$. Note that $\mathcal{F}[g]$ is the union of the set of boundaries of these sets, and also that the set $\mathcal{C}_{sad}[g]$ only exists for $m \geq 2$. Note that the regular set is the union of all non-singular points

$$\mathcal{C}_{reg}[g] = \mathcal{C}[g] \setminus \mathcal{F}[g].$$

We now make the notion of *jumps* more precise: For any point $p = (x,y) \in \mathcal{F}[g]$ we define the (possibly set-valued) *umbral map*‡

$$U[g](p) = \{\omega\text{-limits of a non-trivial trajectories in (6) with } \alpha\text{-limit } p\} \tag{7}$$

(where the ω and α-limits are the limit sets for the trajectory as $t \to \infty$ and $t \to -\infty$ respectively) to be the solutions of the layer equations not equal to p. The ω-limits are always non-empty since we will assume bounded global attractors. However, if all ω limits equal p then the umbral map is empty. The *umbra* or *drop set* is the image of the folds under the umbral map:

$$\mathcal{U}[g] = U[g](\mathcal{F}[g]).$$

We define the projection onto the slow variable as

$$\pi : \mathbb{R}^{m+n} \to \mathbb{R}^n, \quad \pi(x,y) = y$$

---

‡ The image of this umbral map is called the *umbra* (meaning shadow) [14]. The umbra is also called *drop set* [24].

and for $p = (x, y)$ we define the set of all *co-folds* to p as

$$\Pi(p) = \{q \in \mathcal{F}[g] \ : \ \pi(p) = \pi(q)\} = \pi^{-1}(\pi(p)) \cap \mathcal{F}[g], \tag{8}$$

`eq:cofolds`

i.e. all fold points with the same slow coordinate as p. Similarly, we define the set of folds sharing a given slow (y) coordinate to be:

$$P(y) = \mathcal{F}[g] \cap \pi^{-1}(y). \tag{9}$$

`eq:Pydef`

2.1. Trajectories in the singular limit

Note that typical points in $\mathbb{R}^{m+n}$ are not on $\mathcal{C}[g]$: starting at $(x, y) \notin \mathcal{C}[g]$ there will be fast motion governed by the layer equations (6). If this settles to a limit we will typically have arrived at a point on the critical set that is a stable equilibrium, i.e. on $\mathcal{C}_{att}[g]$. The slow dynamics then carries the trajectory around $\mathcal{C}_{att}[g]$ until (possibly) it hits a fold point $p = (x, y) \in \mathcal{F}[g]$. If $U[g](p)$ is a single point then there is a unique non-trivial trajectory of the layer equations from this point, the fast motion will take the dynamics to $U[g](p)$.

Hence typical trajectories in the singular limit (which can be viewed as trajectories of a constrained differential equation [33]) are composed of segments of slow trajectories on $\mathcal{C}_{att}[g]$ interspersed with fast jumps. Depending on the nature of the slow segments and fast jumps, characterised by the umbral maps, there may be a trajectory of the $\epsilon > 0$ system that remains close to the singular trajectory. More precisely, we define a singular trajectory [4, 24] (also called a candidate trajectory) as follows:

`ngulartraj`

Definition 1 (Singular trajectory) *A singular trajectory is a homeomorphic image under* $\gamma_0 : [a, b] \to M \times N$ *with* $a < b$*, where*

- *The interval is partitioned as* $a = t_0 < t_1 < \ldots < t_{n-1} < t_m = b$ *into a finite number of subintervals.*
- *The image of each subinterval* $\gamma_0(t_{j-1}, t_j)$ *is a trajectory of either the fast subsystem or the slow subsystem.*
- *The image* $\gamma_0(a, b)$ *is oriented consistently with the orientations on each subinterval induced by the fast or slow flows.*

Note that t is a parametrisation of the curve rather than fast or slow time. Consequently, the image of a subinterval can be a complete homoclinic or heteroclinic orbits of the fast or slow subsystem. In typical cases, the attractor will consist of subintervals that alternate between fast and slow segments, but this may not be the case at bifurcation. In the case that all slow segments are on $\mathcal{C}_{att}[g]$, this will typically perturb [10] to similar trajectories $\epsilon \to 0$. If the slow segments explore other hyperbolic points on the critical manifold, canard trajectories may appear. Several possible cases of fast/slow trajectories are considered in [10].

2.2. Global equivalence of critical sets

In order to define persistence and bifurcation of critical sets we need a notion of equivalence between the parametrized fast vector fields. We define global equivalence of fast vector fields for the case $m = n = 1$, and leave the generalisation of this equivalence to $n, m > 1$ open.

We consider *fast* and *slow* vector fields $g : \mathbb{R}^{m+n} \to \mathbb{R}^m$ and $h : \mathbb{R}^{m+n} \to \mathbb{R}^n$, with the generic assumptions of smoothness and being in the singular limit $\epsilon \to 0$. More precisely we consider

$$V_f' := C^\infty(\mathbb{R}^{m+n}, \mathbb{R}^m), \quad V_s' := C^\infty(\mathbb{R}^{m+n}, \mathbb{R}^n). \tag{10}$$

`eq:restrictedsy`

We furthermore consider only systems (1) with bounded global attractors.

We restrict ourselves to vector fields defined on some fixed compact regions $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ with open interior and smooth boundaries having outward normals $m(x)$ and $n(y)$ respectively. Implicitly fixing these M and N, we define the open sets

$$V_f := \left\{ g \in V_f' \ : \ \begin{array}{c} g(x,y) \cdot n(x) < 0 \ \forall (x,y) \in (\partial M, N) \ \text{ and} \\ g(x,y) = 0 \Rightarrow g_x(x,y) \neq 0 \ \forall (x,y) \in (M, \partial N) \end{array} \right\} \tag{11}$$

`eq:Xf`

and

$$V_s := \{ h \in V_s' \ : \ h(x,y) \cdot m(y) < 0 \ \forall (x,y) \in (M, \partial N) \}. \tag{12}$$

`eq:Vs`

Note that if $(g, h) \in M \times N$ then the forward dynamics must remain in $M \times N$ and that there are no tangencies of the flow, (or the critical set) with the boundary. These conditions ensure that the properties persist under small perturbations of the vector field.

Recall now that the critical set and the fast dynamics depend only on g, and suppose that $g, \tilde{g} \in V_f$. We say that $g \sim \tilde{g}$ and the critical set $\mathcal{C}[g]$ is *globally equivalent on* $M \times N$ (c.f. [13, p144]) to $\mathcal{C}[\tilde{g}]$ if there are functions $Y(y) : N \to \mathbb{R}^n$, $X(x,y) : M \times N \to \mathbb{R}^m$ and $S : M \times N \to (0, \infty)$ such that:

$$\tilde{g}(x,y) = S(x,y) g(X(x,y), Y(y)) \tag{13}$$

`eq:globalequiva`

i.e. we only consider changes in coordinate that map fast dynamics to fast dynamics up to a possible change in timescale. More precisely, we assume that:

- The map $\Phi(x,y) := (X(x,y), Y(y))$ is a diffeomorphism
- The map $S(x,y) > 0$ is smooth on $M \times N$

The requirement that $S(\cdot, \cdot) > 0$ ensures that trajectories preserve their time orientation under equivalence.

3. Persistence and bifurcation of critical sets

`lmanifolds`

Assume we define V_f as in (11) for some compact regions M and N. In order to define persistence of the critical sets we define the unfolding of the slow dynamics following [13, Section III]. We say a smooth function $G(x, y, \lambda)$ for $\lambda \in \mathbb{R}^r$ is a r-parameter *unfolding* of $g(x, y)$ if

$$G(x, y, 0) = g(x, y)$$

for all $(x, y) \in M \times N$. Reference [13] mostly assumes G and g are germs of vector fields, though in [13, Theorem III.6.1] the equivalence is global within a compact region, as we consider here.

If G and H are both unfoldings of g, we say that H *factors through* G if there exists smooth mappings S, X, Y, L and $W \subset \mathbb{R}^r$, a neighbourhood of 0, such that

$$H(x, y, \lambda) = S(x, y, \lambda)G(X(x, y, \lambda), Y(y, \lambda), L(\lambda)), \text{ for all } \lambda \in W \text{ and } (x, y) \in (M, N),$$

and $S(x, y, 0) = 1, X(x, y, 0) = x, Y(y, 0) = y, L(0) = 0$ (see [13]). We define G to be a *versal* unfolding if every unfolding H of g factors through G. We say it is *persistent* if it is its own unfolding, i.e. for any unfolding $G \in C^\infty(M \times N \times \mathbb{R}^r)$ such that $G(x, y, 0) = g(x, y)$, on $M \times N$ there is a neighbourhood W of 0 in $\mathbb{R}^r$ such that

$$G(x, y, \lambda) \sim g(x, y), \ \forall \lambda \in W,$$

where, as before, $\sim$ denotes global equivalence on $M \times N$.

If the unfolding is versal and contains a minimum number of parameters, we call it a *universal unfolding* [13]. The number of parameters λ in such a universal unfolding G is the *codimension* of g. In particular, if g is *persistent* then it is its own universal unfolding, and in this case we say it has *codimension zero.* We say a *bifurcation* of critical sets occurs for g if $\mathcal{C}[g]$ is non-persistent, and the *codimension* of the bifurcation is that of the universal unfolding of g.

3.1. Persistence and codimension one bifurcation of critical sets for one fast and one slow variable

The case $m = n = 1$ can be directly treated using the global bifurcation theory with distinguished parameter approach of [13, Section III], considering the slow variable y as the distinguished parameter. Consider some $g \in V_f'$ and note that the critical set is

$$\mathcal{C}[g] = \{p = (x, y) \in \mathbb{R} \times \mathbb{R} \ : \ g(p) = 0\},$$

and that the fold set is

$$\mathcal{F}[g] = \{p \in \mathcal{C}[g] \ : \ g_x(p) = 0\}.$$

Table 1 lists the three *degenerate fold* sets $\mathcal{D}_i[g], i = \{1, 2, 3\}$ for $m = n = 1$: fold tangency, hysteresis point, and multiple limit point. The term limit point is a historical term for fold point. The set of all degenerate folds is then

$$\mathcal{D}[g] = \mathcal{D}_1[g] \cup \mathcal{D}_2[g] \cup \mathcal{D}_3[g], \tag{14}$$

eq:defDg11

and any point in $\mathcal{F}[g] \setminus \mathcal{D}[g]$ is a *non-degenerate fold point.* Note that [13] refers to the fold tangency as a "simple bifurcation" and a multiple limit point as a "double limit point" but our notation offers easier generalization to higher n. The following theorem characterizes the persistent critical sets, using a result from [13].

Table 1. Degenerate folds for $m = n = 1$: Proposition 1 states that if $\mathcal{D}[g] = \mathcal{D}_1[g] \cup \mathcal{D}_2[g] \cup \mathcal{D}_3[g]$ defined in (14) is empty then $g \in V_f$ is persistent on $M \times N$.

Fold tangency:	$\mathcal{D}_1[g] \;=\; \{p \in \mathcal{F}[g] : g_y(p) = 0\},$
Hysteresis point:	$\mathcal{D}_2[g] \;=\; \{p \in \mathcal{F}[g] : g_{xx}(p) = 0\},$
Multiple limit point:	$\mathcal{D}_3[g] \;=\; \{p \in \mathcal{F}[g] : \; \lvert\Pi(p)\rvert \geq 2\}.$

persistent

Table 2. Subsets of $\mathcal{D}_i[g]$ for $m = n = 1$ whose union contains all codimension one bifurcations. Note that $\det(D^2 g) = g_{xx}g_{yy} - g_{xy}^2$ and that the first two are local degeneracies.

Quadratic fold tangency:	$\mathcal{D}_1^1[g] \;=\; \{p \in \mathcal{D}_1[g] : \lvert\Pi(p)\rvert = 1 \text{ and } \det(D^2 g(p)) \neq 0\},$
Cubic hysteresis point:	$\mathcal{D}_2^1[g] \;=\; \{p \in \mathcal{D}_2[g] : \lvert\Pi(p)\rvert = 1 \text{ and } g_{xxx}(p) \neq 0\},$
Double limit point:	$\mathcal{D}_3^1[g] \;=\; \{p \in \mathcal{D}_3[g] : \lvert\Pi(p)\rvert = 2\}.$

persistent

prop:CM110

Proposition 1 (Codimension zero, $m = n = 1$) *In the case* $m = n = 1$*, if* $g \in V_f$ *has no degenerate folds (i.e. if* $\mathcal{D}[g] = \emptyset$*) then the critical set* $\mathcal{C}[g]$ *is persistent to smooth perturbations.*

Proof: We apply [13, Theorem 6.1]: this states that there is bifurcation equivalence if there are no (a) simple bifurcations (here called fold tangencies), (b) hysteresis points (c) double limit points (here called multiple limit points) or (d) codimension one interactions of equilibria or folds with the boundaries. The assumptions in (11) are open conditions that ensure that (d) does not happen and that any unfolding of g will remain within V_f for small enough perturbations. Hence the only obstructions are (a-c) which are avoided if $\mathcal{D}[g]$ is empty. $\square$

To aid the classification of codimension one bifurcations, we define $\mathcal{D}_i^1[g]$ in Table 2. These are open dense subsets of $\mathcal{D}_i[g]$ that avoids obvious further degeneracies. We then subdivide these cases further in Table 3. A vector field g is degenerate at codimension one if exactly one of these degeneracies $\mathcal{D}_{i,j}^1[g, y]$ occur for exactly one slow coordinate y (in exactly one fast fibre) which means we only need to compare points in $P(y)$ for some y. We avoid higher codimension fold tangency by precluding the cases $\det(D^2 g) = g_{xx}g_{yy} - g_{xy}^2 = 0$ or higher order hysteresis $g_{xxx} = 0$. The next result shows that Table 3 and Figure 1 give a complete list of codimension one bifurcations for this case.

Table 3. Complete list of degeneracies that lead to codimension 1 bifurcations listed in Proposition 2. We write $P(y) = \{p_i\}$ as the set of distinct singular points $p_i = (x_i, y)$ of the vector field g with slow coordinate y. Note that local degeneracies have $\lvert P(y)\rvert = 1$

Hyperbolic fold tangency	$\mathcal{D}_{1,1}[g,y] = \{P(y) \subset D_1^1[g] : \lvert P(y)\rvert = 1$ and $\det(D^2 g(p)) < 0\}$	Fig. 2 a,b,c)
Elliptic fold tangency	$\mathcal{D}_{1,2}[g,y] = \{P(y) \subset D_1^1[g] : \lvert P(y)\rvert = 1$ and $\det(D^2 g(p)) > 0\}$	Fig. 2 d,e,f)
Stable hysteresis:	$\mathcal{D}_{2,1}[g,y] = \{P(y) \subset D_2^1[g] : \lvert P(y)\rvert = 1$ and $g_{xxx}(p) > 0\}$	Fig. 2 g,h,i)
Unstable hysteresis:	$\mathcal{D}_{2,2}[g,y] = \{P(y) \subset D_2^1[g] : \lvert P(y)\rvert = 1$ and $g_{xxx}(p) < 0\}$	Fig. 2 j,k,l)
Aligned umbra-fold double limit:	$\mathcal{D}_{3,1}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 a,b,c)
Opposed umbra-fold double limit:	$\mathcal{D}_{3,2}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 d,e,f)
Aligned umbra-umbra double limit:	$\mathcal{D}_{3,3}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = U(p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 g,h,i)
Opposed umbra-umbra double limit:	$\mathcal{D}_{3,4}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = U(p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 j,k,l)
Aligned non-interacting double limit:	$\mathcal{D}_{3,5}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $(U[g](p) \cup p) \cap_{p_i \neq p} (U[g](p_i) \cup p_i)) = \emptyset$ and $\nu[g](p_1) \cdot \nu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 m,n,o)
Opposed non-interacting double limit:	$\mathcal{D}_{3,6}[g,y] = \{P(y) \subset D_3^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p) \cap (U(P(y)) \cup P(y) \setminus U[g](p) = \emptyset$, $\forall p \in P(y)$, and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. 3 p,q,r)

ntsubcases

prop:CM111

Proposition 2 (Codimension one, $m = n = 1$) *For $n = 1$ and $m = 1$ the codimension one bifurcations of critical sets $\mathcal{C}[g]$ for $g \in V_f$ are characterised in Fig. 1, such that one of the sets $\mathcal{D}_{j,k}[g,y]$ in Table 3 is non-empty for precisely one y. At such a bifurcation, precisely one of the following occurs:*

- *(i) Two folds merge at a fold tangency (e.g. Fig. 2 a,b,c) or d,e,f)).*
- *(ii) Two folds merge at a hysteresis bifurcation (e.g. Fig. 2 g,h,i) or j,k,l)).*
- *(iii) Two fold points share the same slow coordinate: there are six distinct ways this can occur (e.g. Fig. 3)*

Proof: To avoid persistence, at least one of the the degeneracies $\mathcal{D}_i[g]$ listed in Table 1

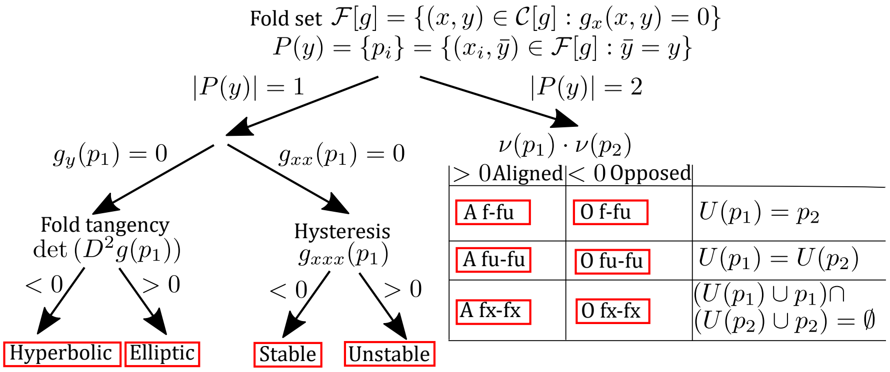


Figure 1. (Color online) Conditions that lead to codimension one degeneracies of the critical set for $m = n = 1$ (see also Table 3). Note that ν is the direction vector of a fold and $\det(D^2(g))$ is the Hessian of g. Similarly, A means aligned, O means opposed, f means fold, fu means fold umbra and fx means non-interacting fold. For ease of notation, $\nu[g](p) = \nu(p)$. For a persistent codimension one bifurcation exactly one branch must be followed for exactly one fast fibre (a single y), leading to one of the red boxes. See text for details

acytable11

must occur for some $i \in \{1, 2, 3\}$: as these are independently defined we can assume that only one will occur for an open dense set of unfoldings. Without loss of generality we can assume that the open conditions in Table 2 apply. □

The subcases of $\mathcal{D}_1^1[g]$ follow from examining the sign of $\det(D^2 g)$: the hyperbolic fold tangency $\mathcal{D}_{1,1}[g, y]$ is the simple bifurcation of [13] while the elliptic fold tangency $\mathcal{D}_{1,2}[g, y]$ is also called the isola. Similarly, the cubic hysteresis $\mathcal{D}_2^1[g]$ can be either stable or unstable, depending on the sign of the leading order term. These cases can be transformed into the normal forms of Table 4 (these are given in [13]). The cases $\mathcal{D}_1^1[g]$ unfold on varying a typical parameter λ as shown in Fig. 2 a,b,c) and d,e,f) respectively, while $\mathcal{D}_2^1[g]$ unfold as shown in Fig. 2 g,h,i) and j,k,l).

The double limit points $\mathcal{D}_3^1[g]$ can be split into several subsets according to the direction of the folds given by the signs of

$$\nu[g](p) = g_{xx}(p) g_y(p)$$

at the two limit points, and k, the number of regular sheets that separate them. The number k determines whether the umbrae and folds intersect. If $k = 0$, then the umbra of one fold intersects the other fold, if $k = 1$ then the umbrae of the folds intersect, and if $k \geq 2$ then the umbrae and folds do not intersect. The six distinct subcases of $\mathcal{D}_3^1[g]$ are shown in Figure 3.

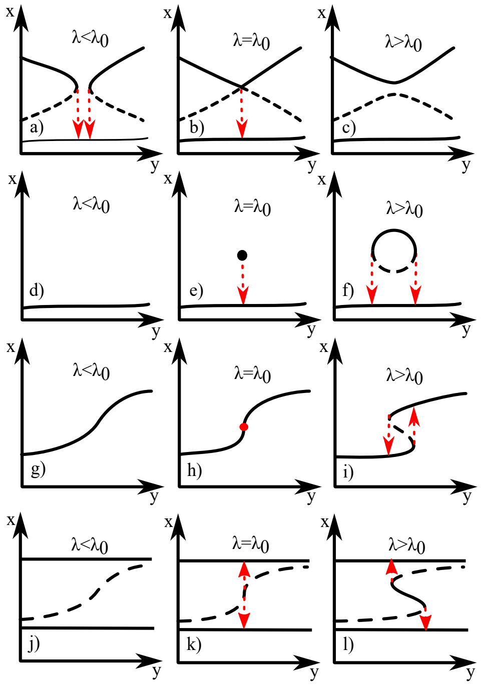


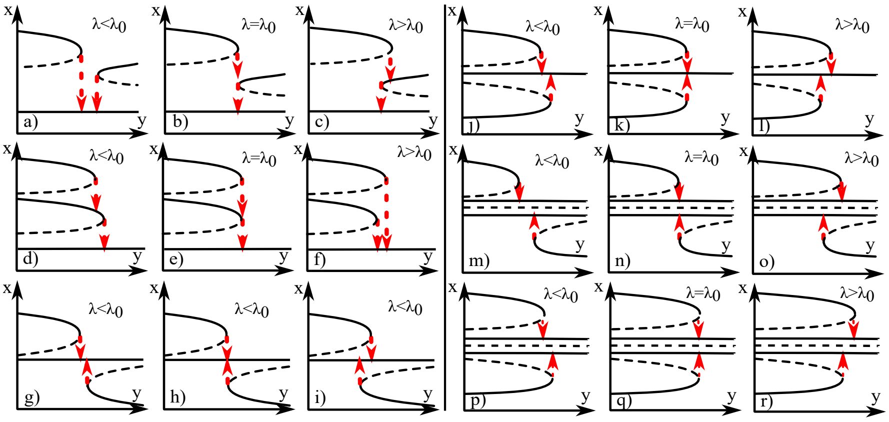
Figure 2. (Color online) Unfoldings of local codimension one structural bifurcations of the critical set for $m = n = 1$. Solid black lines show $\mathcal{C}_{att}[g]$, dashed black lines show $\mathcal{C}_{rep}[g]$ while red arrows show the umbral map from fold points.

eCM111D1D2

Table 4. Normal forms (for $m = n = 1$) and hypothesised normal forms (for $m = 1, n = 2$) of local codimension one bifurcations of the critical manifold. Different signs of $\delta_1, \delta_2 \neq 0$ give different subcases of bifurcation.

$m = n = 1$	
Fold tangency:	$g(x, y) = x^2 + \delta_1 y^2 + \lambda$,
Hysteresis:	$g(x, y) = \delta_1 x^3 + \lambda x + y$,
$m = 1,\ n = 2$	
Fold tangency:	$g(x, y_1, y_2) = x^2 + \delta_1 y_1^2 + \delta_2 y_2^2 + \lambda$,
Cusp tangency:	$g(x, y_1, y_2) = \delta_1 x^3 + \delta_2 x y_2^2 + \lambda x + y_1$,
Swallowtail:	$g(x, y_1, y_2) = x^4 + \lambda x^2 + y_1 x + y_2$,

ormalforms



ureCM111D3

Figure 3. (Color online) Unfoldings of subcases of double limit point, the global codimension one bifurcation of the critical set for $m = n = 1$. Bifurcation occurs when the unfolding (bifurcation) parameter λ equals the critical value λ_0. Solid black lines show $\mathcal{C}_{att}[g]$, dashed black lines show $\mathcal{C}_{rep}[g]$, and red arrows show the umbral map from fold points.

3.2. Persistence of critical sets for one fast and one slow variable

In analogy with the $m = n = 1$ case we give a conjectured list of all degeneracies that can cause non-persistency of vector fields with one fast and two slow variables, up to codimension one. First, we introduce some notation. For any $g \in V_f'$ we write

$$g_x = \frac{\partial g}{\partial x}, \;\; \nabla_y g = (g_{y_1}, g_{y_2}) \;\; \text{and} \;\; \nabla_y^{\perp} g = (-g_{y_2}, g_{y_1}).$$

By $u||v$ we mean that vectors u and v are parallel. For $v \neq 0$ we write the projection of v onto u as $\operatorname{proj}_u v = (u \cdot v)u/|u|^2$. $D^2(g)$ is the Hessian of g with respect to all components of $p = (x, y_1, y_2)$:

$$[D^2(g)]_{ij} = \frac{\partial^2 g}{\partial p_i p_j}, \;\; i, j \in \{1, 2, 3\}.$$

The slow Hessian $D_y^2(g)$ is defined analogously, but with $i, j \in \{2, 3\}$.

Recall that the critical set is

$$\mathcal{C}[g] = \{p = (x, y) \in \mathbb{R} \times \mathbb{R}^2 \;\; : \;\; g(p) = 0\}$$

and the fold set is

$$\mathcal{F}[g] = \{p \in \mathcal{C}[g] \;\; : \;\; g_x(p) = 0\}.$$

As folds are not typically isolated in this case, we also need to distinguish between quadratic folds, cubic cusps and higher order cusps (Table 5).

Table 5. Singularities of the critical set for one fast and two slow variables.

Quadratic fold:	$\mathcal{F}_0[g] = \{p \in \mathcal{F}[g] \ : \ g_{xx}(p) \neq 0\}$
Cubic cusp:	$\mathcal{F}_1[g] = \{p \in \mathcal{F}[g] \ : \ g_{xx}(p) = 0, \ g_{xxx}(p) \neq 0\}$
Higher order cusp:	$\mathcal{F}_2[g] = \{p \in \mathcal{F}[g] \ : \ g_{xx}(p) = 0, \ g_{xxx}(p) = 0\}$

larities12

Table 6. Possible degeneracies of the critical set for one fast and two slow variables, $m = 1$ and $n = 2$. As before, the co-fold set $\Pi(p)$ is the subset in $\mathcal{F}[g]$ sharing slow coordinate with the point p

Fold tangency	$\mathcal{D}_1[g] = \{p \in \mathcal{F}[g] : \ \nabla_y g(p) = 0\}$
Cusp tangency	$\mathcal{D}_2[g] = \{p \in \mathcal{F}[g] \setminus \mathcal{F}_0[g] : \ \nabla_y^{\perp} g(p) \cdot \nabla_y g_x(p) = 0\}$
High order cusp	$\mathcal{D}_3[g] = \{p \in \mathcal{F}[g] : \ g_{xxx}(p) = 0\}$
Tangency of quadratic fold projections	$\mathcal{D}_4[g] = \{p_1 \in \mathcal{F}_0[g] : \nabla_y g(p_1) \vert\vert \nabla_y g(p_2)$ for some $p_2 \in \Pi(p_1) \setminus p_1 \cap \mathcal{F}_0[g]\}$
Cusp projection intersection	$\mathcal{D}_5[g] = \{p \in \mathcal{F}[g] : \lvert\Pi(p)\rvert \geq 2$ and $\Pi(p) \cap \mathcal{F}_1[g] \neq \emptyset\}$
Triple fold projection intersection	$\mathcal{D}_6[g] = \{p \in \mathcal{F}[g] : \lvert\Pi(p)\rvert \geq 3\}$

neracies12

We believe that the list of degenerate sets $\mathcal{D}_i[g]$ of $\mathcal{F}[g]$ given in Table 6 is an exhaustive list of degeneracies under a suitable equivalence. These degeneracies are natural extensions from the degeneracies for $m = n = 1$; generic objects (quadratic fold lines and cubic cusps) can intersect ($\mathcal{D}_1[g]$, $\mathcal{D}_2[g]$, $\mathcal{D}_3[g]$), their projections can become tangent $\mathcal{D}_4[g]$ or intersect $\mathcal{D}_5[g]$ and $\mathcal{D}_6[g]$. More precisely, we define the set of degenerate points

$$\mathcal{D}[g] = \mathcal{D}_1[g] \cup \mathcal{D}_2[g] \cup \mathcal{D}_3[g] \cup \mathcal{D}_4[g] \cup \mathcal{D}_5[g] \cup \mathcal{D}_6[g].$$

Note that the umbral map is single valued for $p \in \mathcal{F}[g] \setminus (\mathcal{D}_2[g] \cup \mathcal{D}_3[g])$. If $p \in \mathcal{D}_2 \cup \mathcal{D}_3[g]$ then it can be zero, one or two-valued (see Figure C1). We now conjecture a persistence criterion for $m = 1$, $n = 2$ that is analogous to the $m = n = 1$ case in Proposition 1.

conj:CM120 **Conjecture 1 (Codimension zero, $m = 1$, $n = 2$.)** *For one fast and two slow variables, the critical set $\mathcal{C}[g]$ is persistent to perturbations for $g \in V_f$ if all folds are non-degenerate, i.e. if $\mathcal{D}[g] = \emptyset$.*

Unfortunately, the method of proof in [13, Thm III.6.1] used for Proposition 1 does not easily generalize to this case of "multiple distinguished parameters" (two slow variables in our case). This is because degenerate cases appear with codimension infinity [29], at least if we consider the restricted global equivalence where we require that $q(y)$ is the identity in (13). Hence proof of this result will require a less stringent (but still natural) form of global equivalence.

3.3. Codimension one bifurcations of critical sets for one fast and two slow variables

A variety of degeneracies can persistently occur for one parameter families, i.e. at codimension one. Table 7 lists degeneracies that we believe contain all persistent codimension one bifurcations for a suitable notion of global equivalence. We divide these into local or global degeneracies. The *local* degeneracies (Table 8) are denoted $\mathcal{D}^1_{i,j}[g,y]$ for $i \in \{1,2,3\}$, the others involve interaction of two or more points in the same fast fibre of y: these *global* degeneracies are denoted $\mathcal{D}^1_{i,j}[g,y]$ for $i \in \{4,5,6\}$ and are listed in detail in Appendix C.

For persistent fold tangency we require that $D^2(g)$ has no zero eigenvalues. For persistent cusp tangency we require that the direction of a cubic cusp (defined later) is non-zero and non-parallel to the gradient. This follows if we assume that the quantity

$$W[g](p) = g_{xxx}(p)\nabla_y^{\perp} g^T(p) D_y^2(g_x(p))\nabla_y^{\perp} g(p) \qquad (15)$$

eq:secondorder

is non-zero. For a typical fold projection tangency, we require that the quadratic terms of two quadratic folds are distinct. This can be guaranteed by considering the quadratic curvature vectors

$$z[g](p) = \left(\frac{\nabla_y g(p)^T D_y^2(g(p))\nabla_y g(p)}{2|\nabla_y g(p)|^4}\right)\nabla_y g(p) \qquad (16)$$

eq:curvaturevec

at fold points p_1 and $p_2 \in \Pi(p_1)$ (see Figure 4 c,d) and Appendix A for more details). We require that these curvature vectors are not both parallel and of equal magnitude:

$$z[g](p_1)\cdot z[g](p_2) \neq |z[g](p_1)|^2. \qquad (17)$$

eq:curvaturevec

The direction vector of a cubic cusp is given by

$$\mu[g] = \frac{g_{xxx}}{\mathrm{proj}_{\nabla_y^{\perp} g}\,\nabla_y g_x}\nabla_y^{\perp} g. \qquad (18)$$

eq:cuspdirecti

(See Figure 4 a) and Appendix B). If $\mu[g]$ is undefined or zero, there is a degeneracy. The former occurs if $\nabla_y^{\perp} g(p)\cdot\nabla_y g_x(p) = 0$ and the latter if $g_{xxx} = 0$. The degeneracy condition for fold tangency is analogous since the vanishing of quadratic fold vector (Fig. 4 b))

$$\nu[g] = g_{xx}\nabla_y g = 0 \qquad (19)$$

eq:folddirecti

also causes degeneracy. The quantities $\nu[g]$, $\mu[g]$, $W[g]$ and $z[g]$ are discussed more in Appendix A and Appendix B. Hypothesised normal forms for the local codimension one bifurcations are listed in Table 4.

We now go trough the persistent subcases of codimension the one degeneracies listed in Table 7.

3.3.1. Fold tangency Fold tangency occurs when a pair or continuum of folds intersect. Typical fold tangency $\mathcal{D}^1_1[g]$ is classified by $|\Sigma_+|$, the number of positive eigenvalues in Σ, the spectrum of $\mathrm{sign}\,(g_{xx}(p))D^2(g)(p)$. The cases of *wormhole* ($|\Sigma_+| = 1$), *tube* ($|\Sigma_+| = 2$) and *isola* ($|\Sigma_+| = 3$) are shown in Figure 5 a) to i). Note that $g_{xx}(p)\det(D^2(g)(p) \neq 0$ implies that all non-positive eigenvalues are negative and that at least one eigenvalue is positive.

Table 7. Subsets of $\mathcal{D}[g]$, for one fast and two slow variables, whose union we conjecture contains all codimension one bifurcation sets. Note that $\mathcal{D}^1_i[g]$ are local for $i = 1, 2, 3$ and global for $i = 4, 5, 6$. $\Pi(p)$ is the set of all singular points sharing slow coordinate with p. (see the text for details)

Typical fold tangency	$\mathcal{D}^1_1[g] = \{p \in \mathcal{D}_1[g] : \lvert\Pi(p)\rvert = 1 \text{ and } \det(D^2 g(p)) \neq 0\}$
Typical cusp tangency	$\mathcal{D}^1_2[g] = \{p \in \mathcal{D}_2[g] : \lvert\Pi(p)\rvert = 1 \text{ and } W[g](p) \neq 0\}$
Swallowtail	$\mathcal{D}^1_3[g] = \{p \in \mathcal{D}_3[g] : \lvert\Pi(p)\rvert = 1 \text{ and } g_{xxxx} \neq 0\}$
Typical double quadratic fold projection tangency	$\mathcal{D}^1_4[g] = \{p \in \mathcal{D}_4[g] : \lvert\Pi(p)\rvert = \lvert\Pi(p) \cap \mathcal{F}_0[g]\rvert = 2$ and $z[g](p) \cdot z[g](q) \neq \lvert z[g](p)\rvert^2, \ \forall q \in \Pi(p) \setminus p\}$
Quadratic fold - cubic cusp projection intersection	$\mathcal{D}^1_5[g] = \{p \subset \mathcal{D}_5[g] : \lvert\Pi(y)\rvert = 2$ and $\lvert\Pi(p) \cap \mathcal{F}_0[g]\rvert = 1$ and $\lvert\Pi(p) \cap \mathcal{F}_1[g]\rvert = 1\}$
Triple quadratic fold projection intersection	$\mathcal{D}^1_6[g] = \{p \in \mathcal{D}_6[g] : \lvert\Pi(p)\rvert = 3 \text{ and } \lvert\Pi(p) \cap \mathcal{F}_0[g]\rvert = 3\}$.

neracies12

Table 8. Subsets of *local* degeneracies, for one fast and two slow variables parametrized by the slow coordinate, conjectured to include all codimension one bifurcations. Note that $P(y)$ is the set of all singular points of the vector field g with slow coordinate y. The number of positive eigenvalues of the Hessian $\operatorname{sign}(g_{xx}) D^2[g]$ at p is written as $\lvert\Sigma_+\rvert$. Note that no eigenvalues are zero, since $\det(D^2 g(p)) \neq 0$. (see the text for details)

Wormhole fold tangency	$\mathcal{D}_{1,1}[g, y] = \{P(y) \subset \mathcal{D}^1_1[g] : \lvert\Sigma_+\rvert = 1\}$	Fig. 5 a,b,c)
Tube fold tangency	$\mathcal{D}_{1,2}[g, y] = \{P(y) \subset \mathcal{D}^1_1[g] : \lvert\Sigma_+\rvert = 2\}$	Fig. 5 d,e,f)
Isola fold tangency	$\mathcal{D}_{1,3}[g, y] = \{P(y) \subset \mathcal{D}^1_1[g] : \lvert\Sigma_+\rvert = 3\}$	Fig. 5 g,h,i)
Stable beaks cusp tangency	$\mathcal{D}_{2,3}[g, y] = \{P(y) \subset \mathcal{D}^1_2[g] : W[g](p) < 0 \text{ and } g_{xxx} < 0\}$	Fig. 5 j,k,l)
Unstable beaks cusp tangency	$\mathcal{D}_{2,4}[g, y] = \{P(y) \subset \mathcal{D}^1_2[g] : W[g](p) < 0 \text{ and } g_{xxx} > 0\}$	
Stable lips cusp tangency	$\mathcal{D}_{2,1}[g, y] = \{P(y) \subset \mathcal{D}^1_2[g] : W[g](p) > 0 \text{ and } g_{xxx} < 0\}$	Fig. 5 m,n,o)
Unstable lips cusp tangency	$\mathcal{D}_{2,2}[g, y] = \{P(y) \subset \mathcal{D}^1_2[g] : W[g](p) > 0 \text{ and } g_{xxx} > 0\}$	
Swallowtail:	$\mathcal{D}_{3,1}[g, y] = \{P(y) \subset \mathcal{D}^1_3[g]\}$	Fig. 5 p,q,r)

neracies12

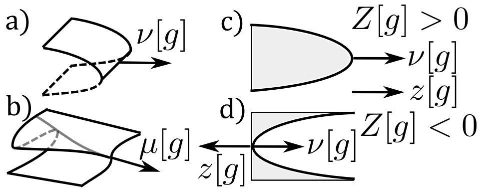


Figure 4. Illustration of fold and cusp alignment vectors. a) Fold (direction) vector $\nu[g]$, b) cusp (direction) vector $\mu[g]$, c) convex fold (viewed as a projection onto the slow subsystem) scalar quadratic curvature $Z[g]$ and curvature direction $z[g]$, d) same as c) but concave.

efinitions

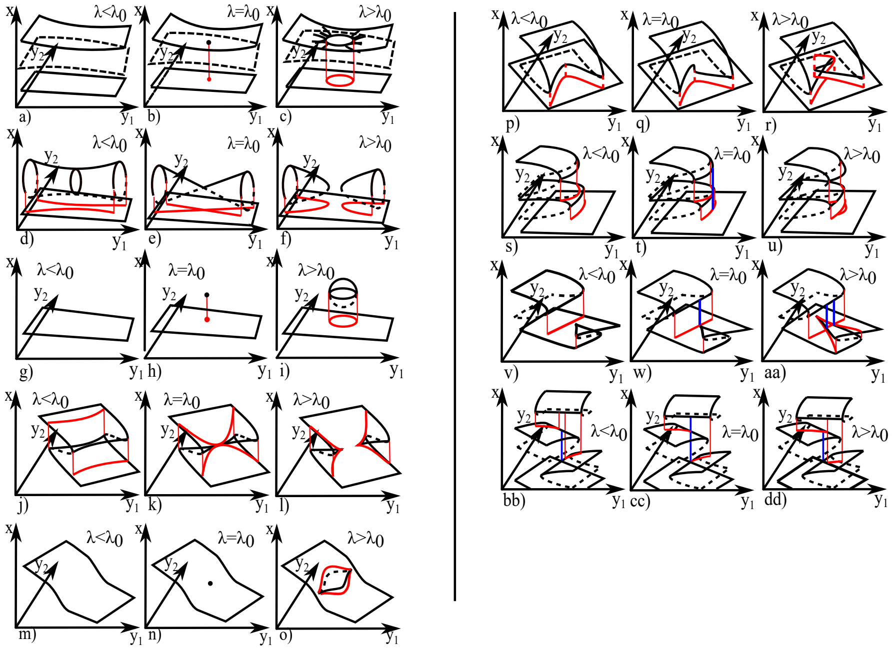


Figure 5. (Color online) Unfolding of examples of codimension one bifurcation of the critical set for $m = 1$, $n = 2$ (see Table 7). Bifurcation occurs when the bifurcation parameter λ equals the critical value $\lambda = \lambda_0$. Solid/dashed black lines show the stable/unstable sheets of the critical set while red lines show the image of the fold under the umbral map. Blue lines indicate special points of intersection.

121DSample

3.3.2. Cusp tangency At a cusp tangency, two cusps meet locally along a line. The subsets *beaks* and *lips* are distinguished by whether cusps are directed away from or towards each other before bifurcation (see Figure 5 and Figure C1). These degeneracies are named after the appearance of their projections onto the slow variables (Figure 9). The case $W[g] > 0$ gives “lips” and $W[g] < 0$ gives “beaks”. We further subdivide these cases depending on their stability, determined by the sign of g_{xxx}.

3.3.3. Swallowtail The swallowtail (Figure 5 m,n,o)) is well known from catastrophe theory and occurs when a fold “folds over itself” to create a degenerate fold that splits up into a pair of cusps.

3.3.4. Fold projection tangency At a fold tangency, the projection of two curves of folds onto the slow variables are tangent. We divide fold projection tangency $\mathcal{D}_4^1[g]$ into subcases depending on whether the folds at points p_1 and p_2 approach each other from the same direction (aligned) or opposite directions (opposed), captured by the sign of the inner product of the fold vectors $\nu[g](p_1)\cdot\nu[g](p_2)$. We further subdivide the opposed cases depending on the sign of the sum

$$Z[g](p_1) + Z[g](p_2), \tag{20}$$

where

$$Z[g](p) = \operatorname{sign}\left(g_{xx}(p)\right)\frac{\nabla_y^{\perp} g(p)^T D_y^2(g(p))\nabla_y^{\perp} g(p)}{|\nabla_y^{\perp} g(p)|^3} \tag{21}$$

is the *scalar quadratic curvature* of the fold projection at a point p (see Figure 4 c,d)). $Z[g](p) > 0$ correspond to a quadratically convex fold with respect to the fold direction and $Z[g](p) < 0$ corresponds to a quadratically concave fold. Hence

$$Z[g](p_1) + Z[g](p_2) < 0$$

means that the concave curvature dominates, and the degeneracy is called a *covering fold projection tangency* since the folds locally cover the slow plane (see Figure 6). Similarly, if

$$Z[g](p_1) + Z[g](p_2) > 0,$$

then the degeneracy is called a *non-covering fold projection tangency.* Accounting for whether the fold umbrae interact with each other, or one fold umbra interacts with a fold, or neither, we get six subcases of opposed fold projection tangency (Table C1). If the two fold projections are aligned the total curvature does not matter as long as $Z[g](p_1) \neq Z[g](p_2)$, with one exception. This exceptional case occurs if a fold umbra hits a fold, in which case it matters if the curvature of the umbral fold dominates the destination fold or not (Figure 6 a.i) and a.ii)). More details are listed in Appendix C.

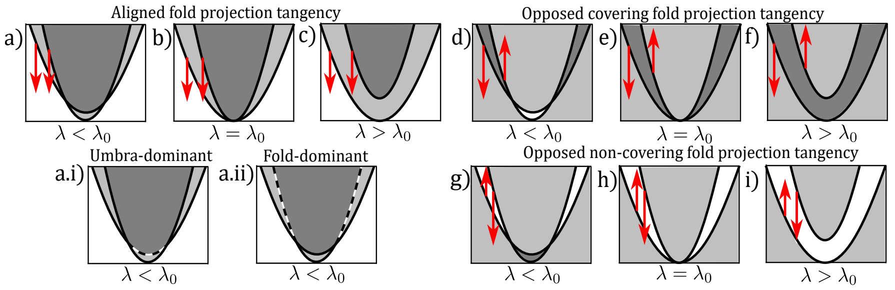


Figure 6. (Color online) Unfolding of a fold projection tangency, viewed in projection onto the slow plane. Three principal cases are shown in a,b,c), d,e,f) and g,h,i). Darker colours mean that more sheets of the critical manifold overlap. Red arrows show fold vectors at tangency points. The aligned fold-fold umbra subcase has umbra-dominant and fold-dominant subcases a.i) and a.ii) respectively. Dotted lines show parts of the destination fold covered by the umbral fold, seen from the stable side of the umbral sheet.

encysketch

3.3.5. Cusp-fold projection intersection At a cusp-fold projection intersection, the projections of a cusp and a fold line coincide in their projection onto the slow variables. We classify the intersection $\mathcal{D}_5[g]$ of a cusp and a fold projection into ten cases (Table C2), depending on the stability of the cusp (determined by g_{xxx}), the direction from which the cusp approaches the fold (determined by the sign of $\nu[g](p_2) \cdot \mu[g](p_2)$), and k, the number of regular sheets of equilibria separating the fold and cusp. If $k = 0$, then one umbra intersect directly with a fold or cusp point (e.g. Figure 5). If $k = 1$, then two umbrae intersect, and if $k \geq 2$ then none of the umbrae or folds intersect. The middle columns of Figures C5 and C6 show typical cases of these degeneracies. Note that no degeneracies involving the umbrae of a stable cusp exist, since stable cusps have no umbrae.

3.3.6. Triple fold projection intersection The projections of three fold lines $\mathcal{D}_6$ onto the slow variables can intersect transversally in two ways: as a *covering triple limit* or as a *non-covering triple limit* (see Figure 7 b)). For brevity we write $\nu_i := \nu[g](p_i)$. In the covering case, all folds are opposed in the sense that their direction vectors span a convex cone covering all of $\mathbb{R}^2$. Therefore, the zero vector can be written as a linear combination of the direction vectors using only non-negative coefficients $a_i \geq 0$, not all zero:

$$\nu_1 a_1 + \nu_2 a_2 + \nu_3 a_3 = 0. \tag{22}$$

eq:linearcombi

In the non-covering case the convex cone of the direction vectors does not cover $\mathbb{R}^2$, meaning that at least one coefficient has to be negative in order for the vector sum to be zero (see Figure 7 a)). Therefore, the two subcases are defined by the signs of the

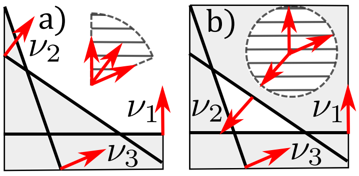


Figure 7. (Color online) Sheets of the critical manifold near a) Non-covering and b) covering triple limit points bifurcations, projected onto the slow variables. Solid black lines show folds. Red arrows indicate direction vectors of folds $\nu_i, i \in \{1,2,3\}$, and grey areas indicate overlapping folds. The convex cones spanned by the direction vectors are shown as striped regions.

`oncovering`

coefficients in (22)

$$\begin{cases} \text{Non-covering triple limit} & \text{if } \pm\,\text{sign}\,(a_1,a_2,a_3) = (+,+,-) \\ \text{Covering triple limit} & \text{if } \pm\,\text{sign}\,(a_1,a_2,a_3) = (+,+,+) \end{cases}, \tag{23}$$

for some choice of prefactor sign. Note that a higher codimension degeneracy will occur if $a_i = 0$ for at least one i. Interactions of umbrae of the folds with other folds or umbrae give additional subclasses of triple limit points: these cases are detailed in Appendix C in Table C3. Note that it is not possible for all three fold umbrae to intersect.

We summarise the discussion in this section with the following classification of codimension one bifurcations for the case of one fast and two slow variables, analogous to Proposition 2.

`conj:CM121`

Conjecture 2 (Codimenson one, $m = 1, n = 2$.) *For $m = 1$ and $n = 2$ the codimension one bifurcations of critical sets $\mathcal{C}[g]$ for $g \in V_f$ are characterised in Figure 8, such that precisely one of the sets $\mathcal{D}_{j,k}[g,y]$ in Table 3 is non-empty, for precisely one $y \in \mathbb{R}^2$. At such a bifurcation, precisely one of the following occurs:*

(i) A loop or pair of hyperbolae appears in the fold projections at a fold tangency $\mathcal{D}_{1,k}$. [e.g. Fig. 5 a) to i)]

(ii) Two cusps annihilate at a cusp tangency $\mathcal{D}_{2,k}$. [e.g. Fig. 5 j) to o)]

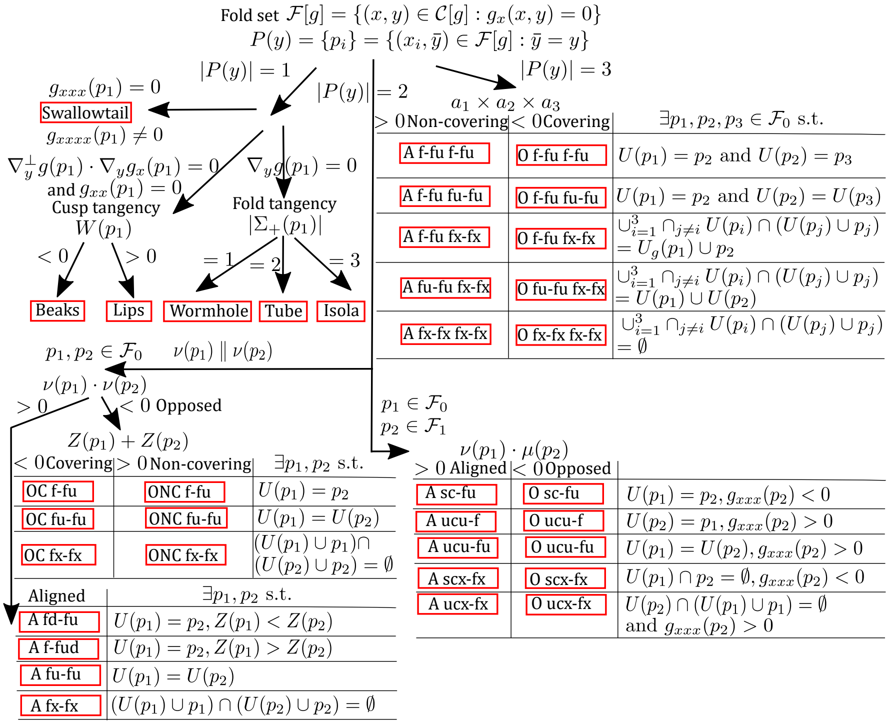
(iii) A quadratic fold line folds over to form two cusps in a swallowtail $\mathcal{D}_{3,k}$. [e.g. Fig. 5 p,q,r)]

(iv) The projections of two quadratic fold curves onto the slow variables become tangent $\mathcal{D}_{4,k}$. [e.g. Fig. 5 s,t,u)]

(v) The projections of a quadratic fold line and a cubic cusp intersect $\mathcal{D}_{5,k}$. [e.g. Fig. 5 v,w,aa)]

(vi) The projections of three fold lines intersect $\mathcal{D}_{6,k}$. [e.g. Fig. 5 bb,cc,dd)]

Figure 9 shows shows the projections of fold lines and cusps that correspond to the possible codimension one degeneracies of the critical set. Appendix C gives a detailed listing of inequivalent subcases of codimension one bifurcations associated with



acytable12

Figure 8. (Color online) Classification of codimension one degeneracies of the critical set for $m = 1$ and $n = 2$. The top left group correspond to local degeneracies $\mathcal{D}_{1,2,3}$. Note that ν and μ are direction vectors of folds and cusps in the projection. Z and W relate to quadratic terms of folds and cusp projections, and Σ_+ is the number of positive eigenvalues of the Hessian of g multiplied by sign (g_{xx}). Note that f means fold, sc/uc means stable/unstable cusp, fu/cu means fold/cusp umbra and fx/cx means non-interacting fold/cusp. Similarly, fd means fold dominant and fud means fold umbra dominant. For ease of notation we suppress dependence on g e.g. $\nu[g](p) = \nu(p)$. For a persistent codimension one bifurcation, exactly one branch must be followed for exactly one fast fibre (a single y), leading to one of the red boxes (see the text for details)

projection intersection: we do not attempt to suggest global normal forms for these cases.

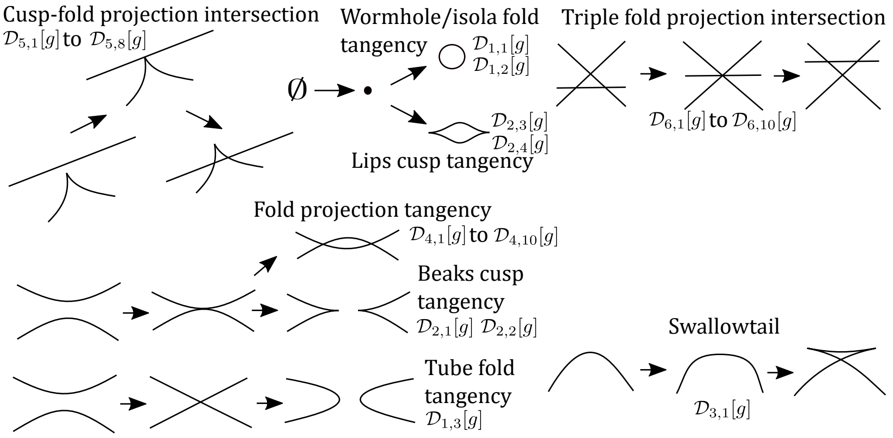


Figure 9. (Color online) Types of codimension one bifurcations of the critical set for one fast and two slow variables, shown in terms of changes to the fold set projected onto slow variables. The cases are enumerated more precisely in Figure 8.

:tionsketch

4. Global singular equivalence, persistence and bifurcation

4.1. Global singular equivalence of systems

To define a useful notion of global equivalence of system (1) in the singular limit, we fix compact regions M and N as above and suppose that $\{g,h\}$ and $\{\tilde{g},\tilde{h}\}$ are both in $V_f \times V_s$ where these are defined as in the previous section.

We say $\{g,h\}$ is *globally singularly equivalent* to $\{\tilde{g},\tilde{h}\}$ (on $M \times N$) if one can write

$$\left\{\begin{array}{ll} \tilde{g}(x,y) = S(x,y)g(X(x,y),Y(y)) & \text{for all } (x,y) \in M \times N \\ \tilde{h}(x,y) = T(x,y)h(X(x,y),Y(y)) & \text{for all } (x,y) \in \mathcal{C}[\tilde{g}] \end{array}\right. \tag{24}$$

eq:globalsinge

where:

- The map $\Phi(x,y) = (X(x,y),Y(y))$ is a diffeomorphism on $M \times N$.
- The function $S(x,y) > 0$ is smooth and positive on $M \times N$.
- The function $T(x,y) > 0$ is smooth and positive on $M \times N$.

Note that because we are only interested in equivalence of the singular systems, we allow independent re-parametrization of the fast and slow timescales. Note that $T(x,y)$ is globally defined but only evaluated on $\mathcal{C}[\tilde{g}]$. Clearly, if $\{g,h\}$ is globally singularly equivalent to $\{\tilde{g},\tilde{h}\}$ then g is globally equivalent to $\tilde{g}$ in the sense of (13). One can check that this is an equivalence relation - it is transitive and reflexive, and one can check it is symmetric by noting that if (24) holds then

$$\left\{\begin{array}{ll} g(x,y) = \tilde{S}(x,y)\tilde{g}(\tilde{X}(x,y),\tilde{Y}(y)) & \text{for all } (x,y) \in M \times N \\ h(x,y) = \tilde{T}(x,y)\tilde{h}(\tilde{X}(x,y),\tilde{Y}(y)) & \text{for all } (x,y) \in \mathcal{C}[g] \end{array}\right. \tag{25}$$

because $(x,y) \in \mathcal{C}[g]$ if and only if $\Phi(x,y) \in \mathcal{C}[\tilde{g}]$, and one can verify that:

- The map $\tilde{\Phi}(x,y) = (\tilde{X}(x,y), \tilde{Y}(y))$ is a diffeomorphism that is the inverse of Φ on $M \times N$.
- The function $\tilde{S}(x,y) = 1/S(\tilde{\Phi}(x,y))$ is smooth and positive on $M \times N$.
- The function $\tilde{T}(x,y) = 1/T(\tilde{\Phi}(x,y))$ is smooth and positive on $M \times N$.

Note that singular trajectories are mapped onto each other by global singular equivalence as expressed in the following result.

Lemma 1 *Suppose that $\{g,h\}$ is globally singularly equivalent to $\{\tilde{g},\tilde{h}\}$ on $M \times N$. Then the singular trajectories of these systems are equivalent via a diffeomorphism.*

Proof: To see this, suppose that Φ, S, T are found that satisfy (24) and suppose that $\gamma_0 : [a,b] \to M \times N$ is a singular trajectory for $\{g,h\}$ as in Definition 1 for $a = t_1 < \cdots < t_m = b$. If $J_j = (\tilde{t}_j, t_{j+1})$ is any fast trajectory segment then $\tilde{\Phi}(J_j)$ is a fast trajectory segment for $\{\tilde{g},\tilde{h}\}$ with the same orientation (and time scaled by $\tilde{S}$). If J_j is a slow segment then it lies within $\mathcal{C}[g]$ and so $\tilde{\Phi}(J_j)$ is a slow trajectory segment for $\{\tilde{g},\tilde{h}\}$ that lies within $\mathcal{C}[\tilde{g}]$ with the same orientation (and time scaled by $\tilde{T}$). □

4.2. Persistence under global singular equivalence

If system (1) is its own universal unfolding under global singular equivalence then we say the system is *persistent*. Clearly, in such a case the fast vector fields indexed by the slow variables must be persistent, but also we cannot have degeneracies of the slow system on the critical set. We expect (1) to be persistent under global singular equivalence if the fast subsystem is persistent and, in addition, the slow system has persistent behaviour on the critical set.

For the case $m = n = 1$ we can make this statement more precise. We define the slow nullcline

$$\mathcal{N}[h] = \{(x,y) \ : \ h(x,y) = 0\}.$$

For any regular point $p_0 = (x_0, y_0) \in \mathcal{C}_{reg}[g]$ there will be a curve $(X_{p_0}(y), y) \in \mathcal{C}[g]$ with $X_{p_0}(x_0) = y_0$ such that $g(X_{p_0}(y), y) = 0$. Implicitly differentiating this gives

$$\frac{dX_{p_0}}{dy}(y) = -\frac{g_y(X_{p_0}(y), y)}{g_x(X_{p_0}(y), y)}$$

for y close to y_0. Then we can locally reduce (2) to an equation on the critical set of the form

$$\dot{y} = H_{p_0}(y) := h(X_{p_0}(y), y).$$

If $p_0 = (x_0, y_0) \in \mathcal{N}[h] \cap \mathcal{C}_{reg}[g]$ then $H_{p_0}(y_0) = 0$ is an equilibrium and its linear stability is determined via

$$H'_{p_0} = \frac{dH_{p_0}}{dy} = -h_x \frac{g_y}{g_x} + h_y$$

evaluated at p_0. This highlights that the slow dynamics are essentially one-dimensional when restricted to $\mathcal{C}_{reg}[g]$. Before stating a result on persistence of fast-slow systems, we give some definitions. We define the restriction of the slow nullcline onto the critical manifold as

$$\mathcal{N}_r[g,h] = \mathcal{N}[h] \cap \mathcal{C}[g],$$

and define the slow degenerate set $\mathcal{E}[g,h]$ as the union of two subsets, defined shortly: $\mathcal{E}[g,h] = \mathcal{E}_1[g,h] \cup \mathcal{E}_2[g,h]$. The slow locally degenerate set is

$$\mathcal{E}_1[g,h] = \{p = (x,y) \in \mathcal{N}_r[g,h] \ : \ g_x(p)h_y(p) - h_x(p)g_y(p) = 0\},$$

which occurs when the fast and the slow nullclines intersect tangentially. Note that this condition is equivalent to the determinant of the Jacobian of the full system being zero, and for $p_0 \in \mathcal{C}_{reg}[g]$ this implies that $H'_{p_0}(y) = 0$.

Recalling that $\pi(p)$ is the projection onto the slow variables, we define the set of *slow co-equilibria*

$$\Xi[g,h](p) = \{\pi^{-1}(\pi(p)) \cap \mathcal{N}_r[g,h]\},$$

which we use to define the *multiple slow equilibrium* set

$$\mathcal{E}_2[g,h] = \{p \in \mathcal{N}_r[g,h] \ : \ |\Xi[g,h](p)| \geq 2\}.$$

We define the (mixed) fold-equilibrium multiple projection set as

$$\mathcal{M}[g,h] = \{p \in \mathcal{N}_r[g,h] \ : \ \pi(p) \cap \pi(\mathcal{F}[g]) \neq \emptyset\},$$

that is, the set of equilibria that share slow coordinate with at least one fold of the critical manifold. This finally allows us to define the fast-slow degenerate set as

$$\mathcal{G}[g,h] = \mathcal{D}[g] \cup \mathcal{E}[g,h] \cup \mathcal{M}[g,h].$$

Theorem 1 below establishes that this set contains all degeneracies under global singular equivalence.

ingpersist **Theorem 1** *In the case $m = n = 1$, if $g \in V_f$ and $h \in V_s$ then (1) is persistent under global singular equivalence for $\{g,h\}$ if and only if the all of the following hold (i.e. $\mathcal{G}[g,h] = \emptyset$):*

- *(i) The critical set $\mathcal{C}[g]$ has no degenerate folds (i.e. $\mathcal{D}[g] = \emptyset$).*
- *(ii) There is at most one equilibrium per slow coordinate y (i.e. $\mathcal{E}_2[g,h] = \emptyset$)*
- *(iii) There is no intersection of the slow nullcline and folds or co-folds, i.e. ($\mathcal{M}[g,h] = \emptyset$).*
- *(iv) There are no degenerate slow equilibria on the critical manifold ($\mathcal{E}_1[g,h] = \emptyset$)*

Proof: We begin with the "if" part. If the critical set has degenerate folds $\mathcal{D}[g] \neq \emptyset$ then g is non-persistent under global equivalence. Hence, we need that $\mathcal{D}[g] = \emptyset$. Assume by contradiction that $\mathcal{E}_2[g,h] \neq \emptyset$. Then at least two equilibria share slow coordinate y_1. Under a generic perturbation of $\{g,h\}$ these will have different slow coordinates, and since the base is preserved under global singular equivalence, $\{g,h\}$ cannot be deformed to make the equilibria share y coordinate. Hence, for persistence we need $\mathcal{E}_2[g,h] = \emptyset$. A similar argument implies that $\mathcal{M}[g,h] = \emptyset$ is required for persistence. Given that $\mathcal{M}[g,h] = \emptyset$, $\mathcal{E}_1[g,h] \neq \emptyset$ implies that there is a $p_0 \in \mathcal{C}_{reg}[g]$ and a y such that $H_{p_0}(y) = H'_{p_0}(y) = 0$. But this is a non-hyperbolic equilibrium, and thus is not persistent to perturbation. Hence, for persistence we need $\mathcal{E}_1[g,h] = \emptyset$.

For the "only if" part, we need to argue that there is no other way for (1) to be non-persistent than if $\mathcal{G}[g,h] \neq \emptyset$. Assume that $\mathcal{G}[g,h] = \emptyset$. Then there is a neighbourhood of every singularity of g and equilibrium of $\{g,h\}$ that has only one fold or equilibrium in the fast fibre. These are either quadratic folds of g or hyperbolic equilibria of $\{g,h\}$, both which are persistent under perturbation. Hence, $\{g,h\}$ must be persistent. ☐

Note that the global singular equivalence (24) does not depend on the nullcline $\mathcal{N}(x,y)$ away from the critical manifold. Bifurcation occurs if one of the assumptions in Theorem 1 is broken. Note that the assumptions that $g \in V_f$ implies the critical set does not intersect $\partial M \times N$, and that $h \in V_s$ implies that the nullcline does not intersect $M \times \partial N$; more generally there will be additional persistence conditions that require persistent intersection with these boundaries.

4.3. Generic bifurcations in singular fast-slow systems

We can understand generic codimension one bifurcations of fast-slow systems (1) by examining the ways that the persistence conditions of Proposition 1 are violated. For $m = 1$ and $n = 1, 2$ this means that the codimension one bifurcations of the critical manifold are possible bifurcations under global singular equivalence. In addition, there are many ways that a change in the slow subsystem can lead to a bifurcation.

For $m = n = 1$ we define, as for the critical manifold, subsets of $\mathcal{E}_1[g,h]$, $\mathcal{E}_2[g,h]$ and $\mathcal{M}[g,h]$ containing all codimension one degeneracies of $\mathcal{G}[g,h]$, which are not only due to degeneracy of the fast subsystem, in Table 9. We further define subsets of these, which give codimension one bifurcation if all except for one of the subsets are empty, and if the nonempty subset is nonempty for only one slow coordinate y (Table 3). Equipped with the subsets in Table 10, Proposition 3 lists the codimension one degeneracies of 1 for $m = n = 1$.

Proposition 3 *In the case $m = n = 1$, codimension one bifurcation of the fast-slow system (1) for $\{g,h\}$ occurs due to exactly one of the following reasons, for exactly one slow coordinate $y \in N$.*

:odimonero1 *(i) Two folds of $\mathcal{C}[g]$ merge at a quadratic fold tangency of the critical manifold at some (x,y), and $\mathcal{G}[g,h] \setminus \mathcal{D}^1_1[g,y] = \emptyset$.*

:odimonero2 *(ii) There is a cubic hysteresis of $\mathcal{C}[g]$ at some (x,y), and $\mathcal{G}[g,h] \setminus \mathcal{D}^1_2[g,y] = \emptyset$.*

:odimonero3 *(iii) There is a double limit point of $\mathcal{C}[g]$ for some y and $\mathcal{G}[g,h] \setminus \mathcal{D}^1_3[g] = \emptyset$*

:odimonero4 *(iv) There is a nondegenerate slow saddle-node equilibrium on the regular part of the critical manifold, and $\mathcal{G}[g,h] \setminus \mathcal{E}_{1,1}[g,h,y] = \emptyset$*

Table 9. Subsets of $\mathcal{E}_1[g,h]$, $\mathcal{E}_2[g,h]$ and $\mathcal{M}[g,h]$ for $m=n=1$ whose union contains all codimension one bifurcations, not only due to bifurcation in the fast subsystem. Note that $\det(D\{g,h\}) = g_x h_y - g_y h_x$ is the Jacobian of the full system (1) and that the first subset is a local degeneracy.

Saddle-node:	$\mathcal{E}_1^1[g,h] = \{p \in \mathcal{E}_1[g,h] : \lvert\Xi[g,h](p)\rvert = 1$ and $\det(D\{g,h\}(p)) \neq 0\}$
Double slow equilibrium:	$\mathcal{E}_2^1[g,h] = \{p \in \mathcal{E}_2[g,h] : \lvert\Xi[g,h](p)\rvert = 2\}$
Fold-equilibrium double projection set:	$\mathcal{M}^1[g,h] = \{p \in \mathcal{M}[g,h] : \lvert\Xi[g,h](p)\rvert = 1\}$

.persistent

Table 10. Degeneracies that lead to codimension one bifurcation of the singular fast-slow system (1) up to global singular equivalence. The sets $\mathcal{F}[g]$ and $\mathcal{U}[g]$ are the fold and the umbral sets and $\pi(p)$ is the projection map. $R(y)$ is the set of all equilibria sharing slow coordinate y. In saddle-node non-degeneracy condition, r and q are eigenvectors of the Jacobian of the full system and its adjoint respectively, and $B = \sum_{j,k}^{2} q_j q_k \frac{\partial^2}{\partial\xi_1 \partial\xi_2}(g,h)$, where $(\xi_1,\xi_2) = (x,y)$ [26, p.175]. The last column associates the degeneracy to a possible bifurcation of relaxation oscillations.

Non-degenerate saddle-node	$\mathcal{E}_{1,1}[g,h,y] = \{R(y) = \pi^{-1}(y) \cap \mathcal{N}_r[g,h] \subset \mathcal{E}_1^1[g,y] : \lvert R(y)\rvert = 1$ and $b = \frac{1}{2} < r, B(q,q) > \neq 0\}$	Fig. 10 d,e,f), SNIC
Double slow equilibrium:	$\mathcal{E}_{2,1}[g,h,y] = \{R(y) \subset \mathcal{E}_2^1[g] : \lvert R(y)\rvert = 1\}$	
Sink-fold intersection:	$\mathcal{M}_{1,1}[g,h,y] = \{R(y) \subset \mathcal{M}^1[g,h] : \lvert R(y) \cap \mathcal{F}[g]\rvert = 1$ and $H'_{p0}(y) < 0$ for $p_0 \in R(y)\}$	Fig. 10 a,b,c), Singular Hopf
Source-fold intersection:	$\mathcal{M}_{1,2}[g,h,y] = \{R(y) \subset \mathcal{M}^1[g,h] : \lvert R(y) \cap \mathcal{F}[g]\rvert = 1$ and $H'_{p0}(y) > 0$ for $p_0 \in R(y)\}$	
Sink-fold umbra intersection:	$\mathcal{M}_{1,3}[g,y] = \{R(y) \subset \mathcal{M}^1[g,h] : \lvert R(y) \cap \mathcal{U}[g]\rvert = 1$ and $H'_{p0}(y) < 0$ for $p_0 \in R(y)\}$	
Source-fold umbra intersection:	$\mathcal{M}_{1,4}[g,h,y] = \{R(y) \subset \mathcal{M}^1[g,h] : \lvert R(y) \cap \mathcal{U}[g]\rvert = 1$ and $H'_{p0}(y) > 0$ for $p_0 \in R(y)\}$	Fig. 10 g,h,i), Singular Homoclinic
Source-fold umbra intersection:	$\mathcal{M}_{1,5}[g,h,y] = \{R(y) \subset \mathcal{M}^1[g,h] : R(y) \cap (\mathcal{U}[g] \cup \mathcal{F}[g]) = \emptyset\}$	

.subcases11

(v) *There are exactly two hyperbolic equilibria that share the same slow coordinate* y *(i.e. there are exactly two points* $p_1, p_2 \in \mathcal{E}_2[g,h]$ *for which* $\pi(p_1) = \pi(p_2) = y$*, and* $\mathcal{G}[g,h] \setminus \mathcal{E}_{2,1}[g,h,y] = \emptyset$. :odimonero5

(vi) *The slow nullcline intersects the critical set transversally at exactly one point* (x,y) *that shares slow coordinate with a quadratic fold, and* $\mathcal{G}[g,h] \setminus \mathcal{M}^1[g,h] = \emptyset$:odimonero6 wbifcodim1

Proof: Note that all degeneracies are contained in the set $\mathcal{G}[x,y] = \mathcal{D}_1[g] \cup \mathcal{D}_2[g] \cup \mathcal{D}_3[g] \cup \mathcal{E}_1[g,h] \cup \mathcal{E}_2[g,h] \cup \mathcal{M}[g,h]$. Because of this, and because the defining conditions are independent, codimension one degeneracy will occur at a point that is in exactly one of those sets. Furthermore, bifurcation must occur for exactly one y since otherwise more than one equality constraint is imposed, raising the codimension.

Case i) describes the only subset $\mathcal{D}_1^1[g]$ of $\mathcal{D}_1[g]$ containing codimension one degeneracies exclusively in $\mathcal{D}_1[g]$, and therefore it produces codimension one degeneracy of $\{g, h\}$. The same is true for cases ii) and iii). Case iv) is codimension one since we impose just one equality condition and exclude higher codimension degeneracy with the non-degeneracy condition in Table 10. Case v) is codimension one since hyperbolic equilibria are persistent, and more than two hyperbolic equilibria sharing slow coordinate would impose more than one equality constraint. Case vi) is codimension one for the same reason. $\square$

Note that codimension two bifurcations may combine degeneracies in more than one of these sets.

5. Generic bifurcations of relaxation oscillations in singular fast-slow systems

`ifurcation`

Not all bifurcations of the singular fast-slow system listed in Proposition 3 will lead to bifurcation of singular relaxation oscillations, as the degeneracy in the fast-slow system must interact with a limit cycle. We focus on bifurcation of singular relaxation oscillations and *simple* relaxation oscillations, a generic subclass due to [14]. Several cases of these bifurcations have been considered in the literature, see for example [10].

5.1. Singular relaxation oscillations

Consider a fast-slow system with $m = 1$ fast variables (1) in the singular limit $\epsilon = 0$. A relaxation oscillation is a singular periodic trajectory $\gamma : [a, b] \to M \times N$ (i.e. such that $\gamma(b) = \gamma(a)$) where the slow segments are in $\overline{\mathcal{C}_{att}}$. If the oscillation consists of alternating stable slow segments on $\mathcal{C}_{att}[g]$ up to non-degenerate folds, fast segments from these folds to their umbra, and satisfies certain other non-degeneracy conditions then we say it is a simple relaxation oscillation. These are called *strongly common slow-fast cycles* in [10] where it is shown that these singular trajectories will be shadowed by a stable periodic orbit for small enough ϵ. Guckenheimer stated a similar persistence theorem in [14]; in Section 5.2 we state and prove a version of it.

We say a continuous curve $s_k : [0, 1] \to M \times N$ is a *slow segment* of a singular trajectory if there is a continuous and monotonic increasing $\theta : [0, 1] \to \mathbb{R}$ such that $s_k(\theta(t))$ is a trajectory of (2). We say a slow segment s_k has *slow time duration* $T_k > 0$ if it can be parameterised by $\theta(t) = t/T_k$. If not, and if $\theta(0) = \theta(1)$ then $T_k = 0$, otherwise $T_k = \infty$.

Up to equivalence of the fast segments joining the slow segment end-points, we define a *relaxation oscillation* in terms of its slow segments as

$$\mathcal{A} = \{s_k(\theta) \ : \ \theta \in [0,1]\}_{k=0}^{d-1} \tag{26}$$

eq:RA

a sequence of continuously parametrized slow segments $s_k : [0,1] \to M \times N$. We will assume that *either* $\mathcal{A}$ is a loop entirely within $\overline{\mathcal{C}_{att}}[g]$ *or*

- $s_k(\theta) \subset \overline{\mathcal{C}_{att}}[g]$ for all $\theta \in (0,1)$
- There is a trajectory $\phi(t)$ of the fast system such that $\alpha(\phi(0)) = s_k(1)$ and $\omega(\phi(0)) = s_{k+1}(0)$ for k modulo d.

for each k, where $\omega(p)$ and $\alpha(p)$ are the omega and alpha limits of a point p respectively. This equivalence class of relaxation oscillations has more than one member if there is more than one fast segment joining two consecutive slow segments.

We define the *slow period* $\mathcal{P}(\mathcal{A})$ of a relaxation oscillation $\mathcal{A}$ to be the total slow time duration of its slow segments. This is

$$\mathcal{P}(\mathcal{A}) = \sum_{k=0}^{d-1} T_k \tag{27}$$

which may be infinite, where orbits in the equivalence class of $\mathcal{A}$ clearly all have the same slow period. We allow the possibility that $\mathcal{A}$ is a loop on $\overline{\mathcal{C}_{att}}[g]$ without jumps, in which case $d = 1$ and $s_0(1) = s_0(0)$, or that the jumps are trivial and on $\mathcal{C}_{att}[g]$. Infinite slow period relaxation oscillation ($\mathcal{P}(\mathcal{A}) = \infty$) of a variety of types are covered by this definition. We define a *simple* relaxation oscillation (cf Guckenheimer [15, 18]) as follows:

Definition 2 (Simple relaxation oscillation ($m = 1, n \geq 1$)) *A relaxation oscillation $\mathcal{A}$ in (26) is* simple *if all of the following hold:*

i) The slow period $\mathcal{P}(\mathcal{A})$ is finite.

ii) The slow segments are on $\mathcal{C}_{att}[g]$, except possibly the last point.

itm:sro4 *iii) Either $s_k(1) \in \mathcal{F}[g] \setminus \mathcal{D}[g]$, or $d = 1$ and $s_0(1) = s_0(0)$.*

iv) The slow segments are not tangent to either fold set or umbral set.

v) The singular return map local to $s_k(1)$ is well-defined with a hyperbolic equilibrium at $s_k(1)$.

itm:sro3 ef:simplero

Note that the assumption $P(\mathcal{A}) < \infty$ implies that the slow segments do not limit to any equilibria of the slow flow. For one fast and one slow variable, our definition of a simple relaxation oscillation can be expressed in a simpler way:

Definition 3 (Simple relaxation oscillation ($m = n = 1$)) *A relaxation oscillation (26) with one fast and one slow variable ($m = n = 1$) is* simple *if*

i) The slow period $\mathcal{P}(\mathcal{A})$ is finite.

ii) We have $s_k(\theta) \in \mathcal{C}_{att}[g]$ for all $\theta \in [0,1)$.

iii) Either $s_k(1) \in \mathcal{F}[g] \setminus \mathcal{D}[g]$ or $d = 1$ and $s_0(1) = s_0(0)$.

simplero11

5.2. Persistence and bifurcation

rsistbif11

We say that a simple relaxation oscillation undergoes bifurcation if the relaxation oscillation ceases to be simple under perturbation of the singular fast-slow system. If not, we say that the relaxation oscillation is persistent. Note that as we only consider fast-slow systems on absorbing regions in $\mathbb{R}^2$, singular relaxation oscillations can bifurcate to either equilibrium points or other singular relaxation oscillations. The following proposition links bifurcation of simple relaxation oscillations to degeneracy in the fast-slow system (cf [14, 15, 16]):

iessropers

Proposition 4 *A simple relaxation oscillation* $\mathcal{A}$ *is persistent for* $n = m = 1$ *if the fast-slow system is persistent under global singular equivalence.*

Proof: Assume $\{g, h\}$ is persistent. Because the slow period is finite, no slow equilibrium can intersect a slow segment $s_k(\theta)$ in a degenerate way; intersection for $\theta \in \{0, 1\}$ implies that $\mathcal{M}[g, h] \neq \emptyset$ and intersection for $\theta \in (0, 1)$ implies $\mathcal{E}_1[g, h] \neq \emptyset$, since the flow must have the same direction on both sides of the equilibrium, which is not possible if the equilibrium is hyperbolic. Hence Condition i) is persistent. Condition ii) is persistent since $s_k(0) \notin \mathcal{C}_{att}[g]$ implies that both $s_{k-1}(1)$ and $s_k(0)$ are in $\mathcal{F}[g]$, which in turn would imply that $\mathcal{D}_3[g] \neq \emptyset$. Condition iii) is persistent since the fold set is non-degenerate everywhere, and since trivial relaxation oscillations $s_0(1) = s_0(0)$ coincide with hyperbolic equilibria in the interior of $\overline{C_{att}}[g]$. □

The converse is not true, however: only some of the degeneracies in the fast-slow system will give rise to a bifurcation of a simple relaxation oscillation. The reasons are listed in the following Proposition and in Table 11. Examples are portrayed in Figures 10 and Figure 11.

Proposition 5 *Bifurcation of a singular relaxation oscillation* $\mathcal{A}$ *for* $m = n = 1$ *due to codimension one bifurcation of the singular fast-slow system* $\{f, g\}$ *occurs for exactly one of the following reasons, at exactly one point* $(x, y) \in \mathcal{A}$.

(i) A source in the slow subsystem intersects the umbra of a quadratic fold in the fast subsystem: $(x, y) \in \mathcal{M}_{1,4}[g, h, y]$.

(ii) A sink in the slow subsystem intersects a quadratic fold in the fast subsystem: $(x, y) \in \mathcal{M}_{1,1}[g, h, y]$.

(iii) There is saddle-node bifurcation of the slow subsystem in the interior of $\mathcal{C}_{att}[g]$*:* $(x, y) \in \mathcal{E}_{1,1}[g, h, y]$.

(iv) There is a hyperbolic fold tangency of the critical manifold: $(x, y) \in \mathcal{D}_{1,1}[g, y]$.

(v) There is a stable hysteresis of the critical manifold: $(x, y) \in \mathcal{D}_{2,1}[g, y]$.

(vi) There is an aligned double fold-umbra limit point of the critical manifold: $(x, y) \in \mathcal{D}_{3,1}[g, y]$ *or* $(x, y) \in \mathcal{D}_{3,2}[g, y]$.

(vii) There is an opposed double fold-umbra limit point of the critical manifold: $(x, y) \in \mathcal{D}_{3,1}[g, y]$ *or* $(x, y) \in \mathcal{D}_{3,2}[g, y]$.

Table 11. Bifurcations of singular relaxation oscillations for $m = n = 1$ due to codimension one bifurcation of the fast-slow system. Definitions of the calligraphic sets are found in Table 3 and Table 10. It is understood that $\mathcal{A}(\lambda)$ is perturbed due to perturbed $g(\lambda) = g(x, y, \lambda)$ and $h(\lambda) = h(x, y, \lambda)$. The final column indicates whether the bifurcation is due to bifurcation of the critical set.

Type: Example	Conditions for a simple singular relaxation oscillation $\mathcal{A}(\lambda)$ to bifurcate at $\lambda = \lambda_0$	Bifurcation of $\mathcal{C}[g]$?
Singular homoclinic: Figure 10(g,h,i)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{E}_{1,1}[g(\lambda_0), h(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{E}_{1,1}[g(\lambda_0), h(\lambda_0), y] = \emptyset$,	No
Singular Hopf: Figure 10(a,b,c)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{M}_{1,1}[g(\lambda_0), h(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{M}_{1,1}[g(\lambda_0), h(\lambda_0), y] = \emptyset$,	No
Singular SNIC: Figure 10(d,e,f)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{M}_{1,4}[g(\lambda_0), h(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{M}_{1,4}[g(\lambda_0), h(\lambda_0), y] = \emptyset$,	No
Hyperbolic fold tangency: Figure 11(d,e,f)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{D}_{1,1}[g(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{D}_{1,1}[g(\lambda_0), y] = \emptyset$,	Yes
Hysteresis: Figure 11(j,k,l)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{D}_{2,1}[g(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{D}_{2,1}[g(\lambda_0), y] = \emptyset$,	Yes
Aligned fold-fold umbra double limit: Figure 11(a,b,c)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{D}_{3,1}[g(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{D}_{3,1}[g(\lambda_0), y] = \emptyset$,	Yes
Opposed fold-fold umbra double limit: Figure 11(g,h,i)	There exists a unique $y \in N$ such that $\mathcal{A}_{lim}(\lambda_0) \neq \emptyset$ and $\mathcal{D}_{3,2}[g(\lambda_0), y] \neq \emptyset$ and $\mathcal{G}[g(\lambda_0), h(\lambda_0)] \setminus \mathcal{D}_{3,2}[g(\lambda_0), y] = \emptyset$,	Yes

`:bifofRO11`

`obifcodim1`

Proof: We determine which of the codimension one degeneracies in Table 1 and Table 10 can cause bifurcation of singular relaxation oscillations by ruling out those that cannot, and by providing examples in the following section. Codimension one bifurcation of limit cycles requires that a regular stable part of the critical manifold exists in a neighbourhood of the bifurcation point, or else the limit cycle does not generically pass through that point. This excludes elliptic fold tangency $\mathcal{D}_{1,2}[g, y]$ and unstable hysteresis $\mathcal{D}_{2,2}[g, y]$. Furthermore, a non-interacting double limit point $\mathcal{D}_{3,5}[g, y]$, $\mathcal{D}_{3,6}[g, y]$, an equilibrium sharing slow coordinate with a fold point but not intersecting it or its umbra $\mathcal{M}_{1,5}[g,h,y]$, or two hyperbolic equilibria sharing y-coordinate $\mathcal{E}_{2,1}[g,h,y]$ are ruled out, because these degeneracies cannot break the simple property of limit cycles at codimension one. A source equilibrium intersecting a fold $\mathcal{M}_{1,2}[g,h,y]$, a sink interacting with a fold umbra $\mathcal{M}_{1,3}[g,h,y]$ are excluded since no relaxation oscillation can exist either at or in a neighbourhood of the bifurcation parameter at codimension one. A "fold umbra - fold umbra" double limit point cannot cause bifurcation at codimension one, since the umbra is generically on $\mathcal{C}_{reg}[g]$, and not intersecting any equilibria (making the period finite). Therefore, there is no way for a simple relaxation oscillation to be lost at such a point. □

The remaining codimension one bifurcations can break the simple property by violating one of the defining conditions: these are listed in Table 11. If a vector field is perturbed by a distinguished parameter $\lambda \in \mathbb{R}$, a simple singular relaxation oscillation may cease to exist for some critical λ_0 where we assume the limit is from below. In such cases we define a limit relaxation oscillation $\mathcal{A}_{lim}(\lambda_0)$ as the limit, in the Hausdorff distance, of a sequence of relaxation oscillations $\mathcal{A}(\lambda)$ parametrized by λ

$$\mathcal{A}_{lim}(\lambda_0) = \lim_{\lambda \to \lambda_0 -} \mathcal{A}(\lambda). \tag{28}$$

eq:limitro

The limit is well defined for simple relaxation oscillations but may be empty. The bifurcation of relaxation oscillations will unfold for the non-singular systems $\epsilon > 0$ to give a variety of canards that will appear on a case to case basis: see for example [24, 10]. Outside a small (in ϵ) range of parameters $\lambda(\epsilon)$ near the critical λ_c, many of the solutions will closely resemble those of the singular system. Hence, the bifurcations in Proposition 5 will give rise to a detectable qualitative change even for non-singular systems.

5.3. Examples of bifurcations of relaxation oscillations

xamplebifs

To illustrate how the bifurcations of limit cycles can be realised, we show in Figure 11 some examples of bifurcations of relaxation oscillations in fast-slow systems (1) with $m = n = 1$, near the singular limit. In all cases the critical manifold is expressed as a relatively low-order polynomial (Table 12). We explain how these were derived in Appendix D.

Table 12. Examples of fast subsystems g that undergo each of the codimension one bifurcations of the critical set $g(x, y) = 0$ for $m = n = 1$ at $\lambda = \lambda_c \approx 0$, shown in Figure 11. Indefinite integrals are taken to have zero constant term. $(\hat{x}, \hat{y})$ are scaled, rotated and translated coordinates. Details how $g(x, y)$ is constructed, and parameters for the opposed double limit degenerate case can be found in Appendix D

Hysteresis:	$g(x,y) = \int -a(x+x_1)(x+x_2)(x+x_3)^2 dx + \lambda x - b - y$, $(a, b, x_1, x_2, x_3) = (15/4, 6/10, -1, 1/25, -1)$,
Fold tangency:	$g(x,y) = -(g_1(x,y)g_2(x,y) + \lambda x + q)$, with $g_1(x,y) = x^3 - 2x + y$ and $g_2(x,y) = (x - x_c)^2 + (y - y_c)^2 - R^2$ and $(x_c, y_c, R, q) = (81/100, -1/4, 11/20, 1/100)$,
Aligned double limit:	$g(x,y) = \int -a(x+x_1)(x+x_2)(x+x_3)(x+x_4)dx + \lambda x - y$, $(a, x_1, x_2, x_3, x_4) = (640/49, -1, -13/40, 1/2, 5/4)$,
Opposed double limit:	$g(x,y) = -(g_1(x,y)g_2(\hat{x},\hat{y}) + \lambda x + q)$, with $g_1(x,y) = 0.5x^3 - x + y$, and $g_2(\hat{x},\hat{y}) = (\hat{x}^2 + \hat{y}^2)^3 - (\hat{x}^2 + (\hat{x}^2 + \hat{y}^2)^2\hat{y}^2)$,

nbifpoly11

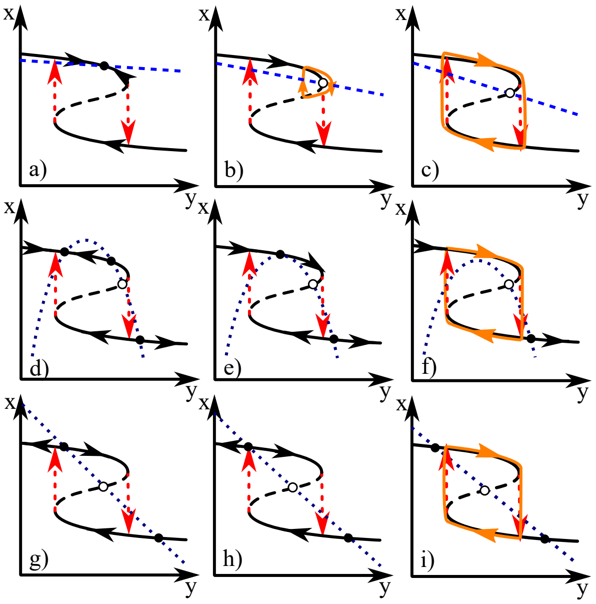


Figure 10. (Color online) The middle column shows typical examples of the classes of codimension one bifurcations of equilibria for $m = n = 1$ that are not due to bifurcation of the critical set (Table 11). In all cases solid black lines show the critical set, red lines show the image of the fold under the umbral map, blue lines show nullclines of the slow variable and orange lines show stylised solutions of the non-singular system. Filled/open dots are stable/unstable equilibria of the fast subsystem.

111slowsub

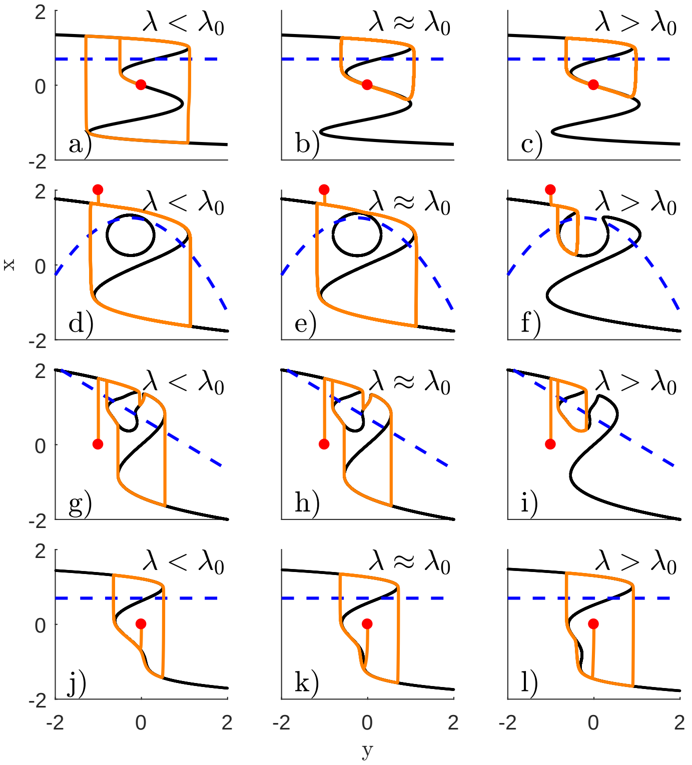


Figure 11. (Color online) Examples of bifurcation of relaxation oscillation due to bifurcation of the critical set for $m = n = 1$. Bifurcation occurs at the critical value λ_c of the bifurcation parameter λ. Black lines are the critical set, blue dashed lines are nullclines of the slow subsystem, and orange lines show example trajectories of a nearly singular system started at the indicated red point, that evolve towards a relaxation oscillation. Polynomial equations for the critical set are listed in Table 12, bifurcation conditions are listed in Table 11, and a detailed description of the systems is found in Appendix D

ifexamples

6. Discussion

discussion

Almost two decades ago, Guckenheimer [15] called for a classification of bifurcations of relaxation oscillations in fast-slow systems up to two slow and two fast variables. In this paper we have used bifurcation theory with distinguished parameters and singular equivalence to take some steps towards such a classification. Indeed, in [15] Guckenheimer gives the following list of codimension one degeneracies that we can relate to our classification:

- **G1:** A fast segment ends at a regular fold point. There are two cases depending on whether the slow flow approaches or leaves the fold near this point.
- **G2:** A slow segment ends at a folded saddle.
- **G3:** A fast segment encounters a saddle point.
- **G4:** There is a point of Hopf bifurcation at a fold.
- **G5:** A slow segment ends at a cusp.
- **G6:** The reduced system has a quadratic umbral tangency between projections of fold and umbra.

The degeneracy **G1** is a subcase of fold projection intersection for $m = 1, n = 1, 2$ and fold projection tangency for $m = 1, n = 2$. The degeneracy **G2** appears when the slow flow is tangent to a fold line: this can occur for $n \geq 2$. Degeneracy **G3** can appear at a saddle for $m \geq 2$ or at an unstable node for $m = 1$. Note that the formulation of **G3** is slightly modified from [15]. Degeneracy **G4** corresponds to a singular Hopf bifurcation, which we discussed in the context of $m = n = 1$. Degeneracy **G5** corresponds to a hysteresis bifurcation for $m = n = 1$, and to a limit cycle hitting a cusp on the slow manifold for $m = 1$, $n = 2$. Finally, degeneracy **G6** requires $n \geq 2$.

Guckenheimer states in [15] that the list is incomplete, and mentions the case that a slow segment ends at a folded node as an example. Degeneracy due to fold tangency of the critical set is missing from the list, since the slow variables were not regarded as distinguished parameters in [15]. We believe that Proposition 5 completes the list for $m = n = 1$.

For $n = 2$, the degeneracies involve tangencies of generic one-dimensional objects such as relaxation oscillations, fold lines and fold umbrae, as well as intersections of one-dimensional objects and generic zero-dimensional objects such as cusp points and equilibria in the slow subsystem. Some of these cases are in Guckenheimer's list. Note that degeneracies of the critical manifold do not cause codimension one bifurcation of relaxation oscillations since they occur at points, which do not generically intersect relaxation oscillations. They will be involved in bifurcation of invariant tori or more complex singular attractors or of relaxation oscillations at higher codimension however.

6.1. Relation to other singularity theory problems

There appears to be connections between global equivalence of critical sets with two distinguished parameters (for $m = 1$ and $n = 2$) and the equivalence of vector fields under projection to the slow plane. In particular, several hypothesised degeneracies under strong equivalence also appear as degeneracies of orthogonal projections of vector fields, see for instance [1, 8, 30, 34, 37] and references. Crucially, however, such equivalences do not produce degeneracies such as the fold tangency where the manifold structure is lost.

Our approach to bifurcation of the critical manifold in fast-slow systems uses a singularity theory approach with distinguished parameters from [13]. In the following we briefly indicate how this approach is related to singularity theory, catastrophe theory, the theory of constrained equations, and projections from manifolds to manifolds.

Much of singularity theory concerns the stability of zero sets of smooth functions $g(x) = 0$ under perturbation [36]. In the singularity theory approach to bifurcation theory of [13], there is a distinguished (bifurcation) parameter y that is not "mixed up" with the unfolding parameters. Hence, e.g. the quadratic fold $g = x^2 + y$ is a codimension one degeneracy of $g(x) = x^2$ in singularity theory, but is codimension zero in bifurcation theory with one distinguished parameter y. For our interpretation, y is identified with the slow variables.

Catastrophe theory [2, 31] classifies the changes to stationary points of potentials $V(x)$, with $\nabla V(x) = g(x)$, by codimension of deformation. Although different equivalences and objects are studied in catastrophe theory and singularity theory, for one (fast) variable, the classification of local singularities of the critical set is the same. Constrained systems [33, 21] correspond to singular fast-slow systems where equivalence of critical manifolds is defined by potential functions, as in catastrophe theory, together with a slow flow local to a point. Unlike our approach, there are no distinguished parameters and the unfolding parameters are identified with the slow variables. This means that some local bifurcations (notably fold tangencies) that are present for the distinguished parameter approach are missed, because "slow variables" never appear in powers higher than one in the local normal forms.

Singularity theory, catastrophe theory and constrained equations have been framed in terms of germs, which are local notions of functions. This means that global intersections of projections of singularities (such as double limit points which are important for bifurcations of relaxation oscillations) have not been widely studied in these contexts, some exceptions being [2, 7, 15, 10].

6.2. Further perspectives

Persistence and codimension one bifurcation of the critical set for one fast ($m = 1$) and two slow variables ($n = 2$) remains to be proved. This requires a suitable equivalence, which should give rise to the degeneracies that we have listed, but possibly more.

A full investigation of bifurcations of singular relaxation oscillations for $m = 1, n = 2$ is outside the scope of this paper. Some specific examples have been studied by Guckenheimer [17, 19, 16] who outlined a scheme for investigation of bifurcation of solutions to singular fast-slow systems in [14].

The general case of two fast variables $m = 2$ is considerably more complicated as the vector fields cannot be written as gradients of potentials, and hence there can be other asymptotic behaviour than fixed points. If $n = 2$ then for generic asymptotic fast dynamics, the system will approach a critical set that is a union of all equilibria, periodic orbits and homoclinic/heteroclinic cycles of the fast system. The persistence of bifurcations on the critical set will depend on the number of slow variables. For $n = 1$ then we expect persistence precisely when (a) All singularities of equilibria within the critical set are quadratic folds or Hopf points. (b) All singularities of limit cycles within the critical set are one of saddle-nodes of limit cycles, saddle node on a periodic orbit, or homoclinic bifurcation. (c) The slow flow has generic intersection with umbrae of the singularities. For $n = 2$ we will get in addition generic local and global bifurcation two singularities at isolated points in the slow variables; this will include, for example, cusp points, Bogdanov Takens points and Bautin points at singular equilibria, and a wide variety of possible generic codimension two bifurcations of homoclinic orbits [9].

We have ignored the phenomena that arise when the scale separation is imperfect, that is for $\epsilon > 0$. In that case, the fast and slow subsystems evolve at similar speeds close to singular points; this gives rise to *canards* and mixed mode oscillations [11]. Canard behaviour has been extensively studied, especially near regular values of the critical set see e.g. [5, 24, 14, 35]. Canards for degenerate critical sets are discussed in [2, 7, 10], but we are unaware of any systematic treatment.

Acknowledgements

We thank James Montaldi, Christian Kuehn, Hildeberto Jardón-Kojakhmetov, Berndt Krauskopf and Christian Bick for valuable discussions. This research has been funded by the European Union's Horizon 2020 innovation and research programme for the ITN CRITICS under the Marie Skłodowska-Curie grant agreement No. 643073.

References

[1] F. Alharbi and V.M. Zakalyukin. Quasi-projections of surfaces with boundaries. *Journal of Mathematical Sciences*, 199(5):473–480, 2014.

[2] V. I. Arnold, V. S. Afrajmovich, Yu. S. Ilyashenko, and L. P. Shilnikov. *Bifurcation theory and catastrophe theory.* Springer-Verlag, Berlin, 1999. Translated from the 1986 Russian original by N. D. Kazarinoff, Reprint of the 1994 English edition from the series Encyclopaedia of Mathematical Sciences [*Dynamical Systems. V*, Encyclopaedia Math. Sci., 5, Springer, Berlin, 1994; MR1287421 (95c:58058)].

[3] Peter Ashwin and Peter Ditlevsen. The middle pleistocene transition as a generic bifurcation on a slow manifold. *Climate Dynamics*, 45(9-10):2683–2695, 11 2015.

[4] Eric Benoît. Canards et enlacements. *Publications mathématiques de'l IHÉS*, 72:63–91, 1990.

[5] Eric Benoît, Jean Louis Callot, Francine Diener, and Marc Diener. Chasse au canard. *Collectanea Mathematica*, 32:37–74, 1981.

[6] Katherine Bold, Chantal Edwards, John Guckenheimer, Sabyasachi Guharay, Kathleen Hoffman, Judith Hubbard, Ricardo Oliva, and Warren Weckesser. The forced van der Pol equation. II. Canards in the reduced system. *SIAM J. Appl. Dyn. Syst.*, 2(4):570–608, 2003.

[7] H.W. Broer, T.J. Kaper, and M. J Krupa. Geometric desingularization of a cusp singularity in slow–fast systems with applications to zeeman's examples. *Journal of Dynamics and Differential Equations*, 25(925):37–74, 2013.

[8] J.W. Bruce. Seeing the mathematical viewpoint. *The Mathematical Intelligencer*, 6(4):18–25, 1984.

[9] A.R. Champneys and Yu. A. Kuznetsov. Numerical detection and continuation of codimension-two homoclinic bifurcations. *International Journal of Bifurcation and Chaos*, 04(04):785–822, 1994.

[10] P. de Maesschalck, F. Dumortier, and R. Roussarie. Cyclicity of common slow-fast cycles. *Indagationes Mathematicae*, 22(3):165 – 206, 2011. Devoted to: Floris Takens (1940-2010).

[11] M. Desroches, J. Guckenheimer, B. Krauskopf, C. Kuehn, H. Osinga, and M. Wechselberger. Mixed-mode oscillations with multiple time scales. *SIAM Review*, 54(2):211–288, 2012.

[12] G.B. Ermentrout and N. Kopell. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM Journal of Applied Math*, 46:233–253, 1986.

[13] Martin Golubitsky and David Schaeffer. *Singularities and Groups in Bifurcation Theory, Volume I*, volume 51 of *Applied Math Sci.* Springer, 1985.

[14] John Guckenheimer. Towards a global theory of singularly perturbed dynamical systems. In *Progress in Nonlinear Differential Equations and Their Applications*, volume 19. Birkhauser, 1996.

[15] John Guckenheimer. Nonlinear dynamics and chaos: Where do we go from here? In *Progress in Nonlinear Differential Equations and Their Applications.* IoP Publishing, 2002.

[16] John Guckenheimer. Bifurcations of relaxation oscillations. In *Bifurcations, Normal Forms and Finiteness Problems in Differential Equations*, volume 137 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 295–316. Springer, 2004.

[17] John Guckenheimer, Kathleen Hoffman, and Warren Weckesser. The forced van der Pol equation. I. The slow flow and its bifurcations. *SIAM J. Appl. Dyn. Syst.*, 2(1):1–35, 2003.

[18] John Guckenheimer and Stewart Johnson. Planar hybrid systems. In Antsaklis P., Kohn W., Nerode A., and Sastry S., editors, *Hybrid systems, II (Ithaca, NY, 1994)*, volume 999 of *Lecture Notes in Comput. Sci.*, pages 202–225. Springer, Berlin, 1995.

[19] John Guckenheimer and Phillip Meerkamp. Unfoldings of singular Hopf bifurcation. *SIAM Journal of Applied Dynamical Systems*, 11(4):1325–1359, 2012.

[20] E. Harvey, V. Kirk, M. Wechselberger, and J. Sneyd. Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics. *J. Nonlinear Sci.*, 21:639–683, 2011.

[21] Hildeberto Jardón-Kojakhmetov and Henk W. Broer. Polynomial normal forms of constrained differential equations with three parameters. *Journal of differential equations*, 257:1012–1055, 2014.

[22] K. Krischer, M. Eiswirth, and G. Ertl. Oscillatory co oxidation on pt(110): modeling of temporal self-organization. *J. Chem. Phys.*, 96, 1992.

[23] Christian Kuehn. Normal hyperbolicity and unbounded critical manifolds. *Nonlinearity*, 27(6):1351, 2014.

[24] Christian Kuehn. *Multiple time scale dynamics*, volume 191 of *Applied Mathematical Sciences.* Springer, Cham, 2015.

[25] Christian Kuehn and Christian Münch. Duck traps: Two-dimensional critical manifolds in planar systems. *Dynamical Systems*, 2018.

[26] Yuri A. Kuznetsov. *Elements of applied bifurcation theory*, volume 112 of *Applied Mathematical Sciences.* Springer-Verlag, New York, third edition, 2004.

[27] Tu W. Loring. *An Introduction to Manifolds.* Springer, 2010.

[28] Mikikian M., M. Cavarroc, L. Couëdel, Y. Tessier, and L. Boufendi. Mixed-mode oscillations in complex-plasma instabilities. *Phys. Rev. Lett.*, 100, 2008.

[29] James Montaldi. The path formulation of bifurcation theory. In *Dynamics, Bifurcation and Symmetry*, pages 259–278. Springer, 1994.

[30] Toru Ohmoto and Francesca Aicardi. First order local invariants of apparent contours. *Topology*, 45:27–45, 2006.

[31] Tim Poston and Ian N. Stewart. *Catastrophe Theory and Its Applications.* Dover, 1978.

[32] S. Rinaldi and M. Scheffer. Geometric analysis of ecological models with slow and fast processes. *Ecosystems*, 3:507–521, 2000.

[33] Floris Takens. Constrained equations; a study of implicit differential equations and their discontinuous solutions. In Peter Hilton, editor, *Structural Stability, the Theory of Catastrophes, and Applications in the Sciences*, pages 143–234, Berlin, Heidelberg, 1976. Springer Berlin

Heidelberg.

[34] F. Thomas and P. Wenger. On the topological characterization of robot singularity loci. a catastrophe-theoretic approach. In *2011 IEEE International Conference on Robotics and Automation*, pages 3940–3945, May 2011.

[35] Martin Wechselberger. à propos de canards (Apropos canards). *Trans. Amer. Math. Soc.*, 364(6):3289–3309, 2012.

[36] H. Whitney. On singularities of mappings of euclidean spaces i. *Annals of mathematics*, 62:374–410, 1955.

[37] Toshiki Yoshida, Yutaro Kabata, and Toru Ohmoto. Bifurcation of plane-to-plane map-germs of corank 2. *The Quarterly Journal of Mathematics*, 199(1):369–391, 2014.

[38] Antonios Zagaris, Hans G Kaper, and Tasso J Kaper. Fast and slow dynamics for the computational singular perturbation method. *Multiscale Modeling & Simulation*, 2(4):613–638, 2004.

Appendix A. The quadratic curvature

:curvature

We define the scalar quadratic curvature at a fold point p in the direction of a fold of the critical manifold as

$$\begin{aligned} Z[g](p) = &\quad \operatorname{sign}(g_{xx}) \tfrac{{\nabla_y^\perp g}^T D_y^2(g) \nabla_y^\perp g}{2|\nabla_y^\perp g|^3}. \\ = &\quad \operatorname{sign}(g_{xx}) \tfrac{(g_{y_1 y_1} g_{y_2}^2 + g_{y_2 y_2} g_{y_1}^2 - 2 g_{y_1 y_2} g_{y_1} g_{y_2})}{2(g_{y_1}^2 + g_{y_2}^2)^{3/2}}, \end{aligned} \tag{A.1}$$

eq:curvaturecriterion

where $D_y^2(g)$ is the Hessian of the slow subsystem, all functions are evaluated at a point p (e.g. $g = g(p)$), superscript T denotes transpose and $\perp$ denotes perpendicular. $Z[g] > 0$ implies convex and $Z[g] < 0$ implies concave. $Z = 0$ is a degenerate case which reads, to first order, that the fold line is locally straight, or not a quadratic fold.

Equation (A.1) implies that if $\operatorname{sign}(g_{xx}) D_y^2(g)$ is positive definite, then the fold line is locally convex, and if $\operatorname{sign}(g_{xx}) D_y^2(g)$ is negative concave, then the fold line is locally convex. Note however, that if $\operatorname{sign}(g_{xx}) D_y^2(g)$ is indefinite or has a zero eigenvalue, then the sign of $Z[g]$ can be either positive, negative or zero depending on the direction of the fold.

In the following we derive (A.1). We start with a fold in the origin, which is quadratic both in the fast and slow directions, and whose slow direction obviously is along the y_1-axis

$$g(x, y) = \xi x^2 + \kappa y_1 + \zeta y_2^2.$$

Recall that the direction vector of a fold is given by $\nu[g] = g_{xx} \nabla_y g = (2\xi\kappa, 0)$, so if $\xi\kappa > 0$ then the fold is directed rightward, while if $\xi\kappa < 0$ the direction is directed leftward. It makes sense to define the slow quadratic curvature of this fold as

$$Z[g] = \operatorname{sign}(\xi) \frac{\zeta}{2\kappa}, \tag{A.2}$$

eq:normalformcu

since then the curvature is independent of the magnitude of ξ, proportional to $\zeta = 0$, and inversely proportional to κ. In this case $Z[g]$ is independent of the point at which it is evaluated, but this is not generally the case. The fold is convex in the direction of the fold if $Z[g](p) > 0$ and concave if $Z[g](p) < 0$.

We now consider a general quadratic polynomial function of a quadratic fold with terms of relevant order

$$g(x, y_1, y_2) = \xi x^2 + a y_1 + b y_2 + \alpha y_1^2 + \beta y_2^2 + 2\gamma y_1 y_2 \tag{A.3}$$

eq:generalfold

and seek a rotation

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

which brings the direction vector $\nu[g] = \mathrm{sign}\,(g_{xx})[a, b]^T$ in the positive y_1 direction, that is: $R\nu[g] = [|\nu[g]|, 0]^T$. This gives that

$$R = \mathrm{sign}\,(g_{xx})\frac{1}{\sqrt{a^2+b^2}}\begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

and

$$R^{-1} = \mathrm{sign}\,(g_{xx})\frac{1}{\sqrt{a^2+b^2}}\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

The old slow coordinates (y_1, y_2) are expressed in the new ones $(\hat{y_1}, \hat{y_1})$ as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = R^{-1}\begin{pmatrix} \hat{y_1} \\ \hat{y_2} \end{pmatrix}.$$

In the new coordinates (A.3) becomes

$$\begin{aligned} g(x, \hat{y_1}, \hat{y_2}) = \; & \xi x^2 + \tfrac{1}{a^2+b^2}[\mathrm{sign}\,(g_{xx})\sqrt{a^2+b^2}a(a\hat{y_1} - b\hat{y_2}) + \\ & \mathrm{sign}\,(g_{xx})\sqrt{a^2+b^2}b(b\hat{y_1} + a\hat{y_2}) + \\ & \alpha(a\hat{y_1} - b\hat{y_2})^2 + \beta(b\hat{y_1} + a\hat{y_2})^2 + \\ & 2\gamma(a\hat{y_1} - b\hat{y_2})(b\hat{y_1} + a\hat{y_2})] \\ = \; & \xi x^2 + \tfrac{1}{a^2+b^2}[\mathrm{sign}\,(g_{xx})(a^2+b^2)^{3/2}\hat{y_1} + (\alpha b^2 + \beta a^2 + 2\gamma ab)\hat{y_1}^2 + \\ & (\alpha b^2 + \beta a^2 - 2\gamma ab)\hat{y_2}^2 + 2(-\alpha ab + \beta ab + (a^2 - b^2)\gamma)\hat{y_1}\hat{y_2}]. \end{aligned}$$

Hence, reading off the coefficients of $\hat{y}_1$ and $\hat{y}_2^2$ in analogy with (A.2), we get for non-zero g_{xx} and $\nabla_y g$ the scalar quadratic curvature

$$\begin{aligned} Z[g] = \; & \mathrm{sign}\,(g_{xx})\tfrac{(\alpha b^2+\beta a^2-2\gamma ab)}{(a^2+b^2)^{3/2}} = \\ = \; & \mathrm{sign}\,(g_{xx})\tfrac{(g_{y_1y_1}g_{y_2}^2 + g_{y_2y_2}g_{y_1}^2 - 2g_{y_1y_2}g_{y_1}g_{y_2})}{2(g_{y_1}^2+g_{y_2}^2)^{3/2}} \\ & \mathrm{sign}\,(g_{xx})\tfrac{\nabla_y^\perp g^T D_y^2(g)\nabla_y^\perp g}{2|\nabla_y^\perp g|^3}. \end{aligned} \tag{A.4}$$

We can now define the quadratic (non-scalar) curvature as

$$z[g](p) = \frac{Z[g](p)\nu[g]p}{|\nu[g](p)|},$$

where $|\nu[g](p)|$ is the length of the fold curvature vector. The quadratic curvature points in the direction of the fold if the fold is convex, and against the direction if it is concave (see Figure 4).

Appendix A.1. Persistent subcases of the fold projection tangency

We classify the fold projection tangency degeneracy (degeneracy subset $\mathcal{D}_4[g]$ for $m = 1$ fast and $n = 2$ slow variables) into a number of qualitatively different subcases. The subcases are separated by the scalar quadratic curvatures at the points of degeneracy, whether the umbrae interact with each other or a fold, and in the case of fold-umbra degeneracy, whether the dominant curvature belongs to the fold curve with the largest x-coordinate.

Aligned folds generate four distinct subcases. In analogy with the situation for $m = n = 1$, we have fold-fold, fold-umbra, and non-interacting fold cases. But the fold-umbra has two subcases, depending on whether the fold with interacting umbra also has the greatest scalar quadratic curvature $Z[g]$. Hence, there are four subcases.

Opposed folds have six distinct subcases, three for each case that either the sum of curvatures is positive (net convex) or negative (net concave). Opposed fold projection tangency does not have the two fold-umbra subcases of aligned fold projection tangency, since the dominant x-values are reversed by a rotation of the slow variables by half a turn. Hence there are six such subcases, and ten subcases in total.

The nonpersistence condition for fold projection tangency degeneracy at points p_1 and $p_2 \in \Pi(p_1)$ is

$$z[g](p_1) \cdot z[g](p_2) = |z[g](p_1)|^2 = |z[g](p_2)|^2$$

that is, when the folds have the same curvature vector (see Appendix A). Note that since the folds are tangent it is implicit that

$$z[g](p_1) \cdot z[g](p_2) = \pm |z[g](p_1)||z[g](p_2)|.$$

Inequality in the above expression cannot distinguish between typical subcases. For that purpose we will later use the scalar quadratic curvature $Z[g]$ and information about whether folds are aligned or opposed.

We now separate two subcases depending on whether folds are aligned or opposed. If $\nu[g](p_1) \cdot \nu[g](p_2) > 0$ the folds are aligned, and qualitatively indistinguishable unless the umbra of one fold (with a larger x component) hits the other fold. Therefore, we get the four subcases illustrated in Figure C3 and tabulated in Table C1. (Figure 6 a,b,c) shows a slow projection sketch of aligned fold projection tangency).

If, on the other hand $\nu[g](p_1) \cdot \nu[g](p_2) < 0$, then the folds are opposed and therefore distinguishable at bifurcation. (Figure 6 d,e,f) shows a slow projection sketch of opposed fold projection tangency).

Assuming that folds are opposed, the defining condition for covering opposed fold projection tangency is

$$Z[g](p_1) + Z[g](p_2) < 0,$$

that is, if at least one fold is concave, and the curvature of any convex fold is smaller in magnitude than that of the convex curve (see Figure 6 d,e,f)). If, on the other hand

$$Z[g](p_1) + Z[g](p_2) > 0,$$

then we have covering opposed fold projection tangency (see Figure 6 g,h,i)). All such vector fields have at least one locally convex curve, and if there is one concave curve then the magnitude of the curvature of the convex curve dominates.

Opposed fold curves do not have the umbra-dominant and fold-dominant subcases that aligned fold curves do, since a rotation of the slow variables by 180 degrees turns one such case into the other. Therefore, there are four aligned cases and six opposed cases, giving in total the ten subcases in Table C1.

Appendix B. Definition of the cusp direction vector

`pdirection`

We define/derive the direction of a cubic cusp similarly to how we defined the curvature of a quadratic fold. We stipulate what the direction of the cusp should be in a normal form, where mixed terms have been neglected and the gradient of the system is in the positive y_1-direction

$$g(x, y_1, y_1) = \xi x^3 + \kappa x y_2 + \zeta y_1 \tag{B.1}$$

In this case we define the direction vector as

$$\mu[g] = \left(\frac{\xi}{\kappa}\right)(-\zeta, 0), \tag{B.2}$$

where $(0, -\zeta)$ is a vector perpendicular to the gradient $(\zeta, 0)$.

Next, we consider a more general expression, still only with terms of relevant order

$$g(x, y_1, y_2) = \xi x^3 + c x y_1 + d x y_2 + a y_1 + b y_2, \tag{B.3}$$

`eq:generalcusp`

and rotate the slow subsystem as to make the gradient in the old coordinates (a, b) directed along the positive y_1 axis. We accomplish this with a rotation, mapping the new coordinates $\hat{y_1}, \hat{y_2}$ to the old ones

$$\begin{cases} y_1 = & a\hat{y_1} - b\hat{y_2} \\ y_2 = & b\hat{y_1} + a\hat{y_2}. \end{cases} \tag{B.4}$$

Equation (B.3) now becomes

$$\begin{aligned} g(x, \hat{y_1}, \hat{y_2}) = \; & \xi x^3 + \tfrac{1}{\sqrt{a^2+b^2}}[c(a\hat{y_1} - b\hat{y_2})x + d(b\hat{y_1} + a\hat{y_2})x \\ & + a(a\hat{y_1} - b\hat{y_2}) + b(b\hat{y_1} + a\hat{y_2})] \\ = \; & \xi x^3 + \tfrac{(ac+db)}{\sqrt{a^2+b^2}}\hat{y_1}x + \tfrac{(-bc+ad)}{\sqrt{a^2+b^2}}\hat{y_2}x + \sqrt{a^2+b^2}\hat{y_1}. \end{aligned} \tag{B.5}$$

The cusp vector in the new coordinates (reading off the coefficients of x^3, $y_2 x$ and y_1) is thus

$$\mu[g] = \frac{\xi\sqrt{a^2+b^2}}{-bc+ad}(\sqrt{a^2+b^2}, 0). \tag{B.6}$$

Rotating this vector back to the original coordinates we get that

$$\mu[g] = \frac{\xi\sqrt{a^2+b^2}}{-bc+ad}(-b, a), \tag{B.7}$$

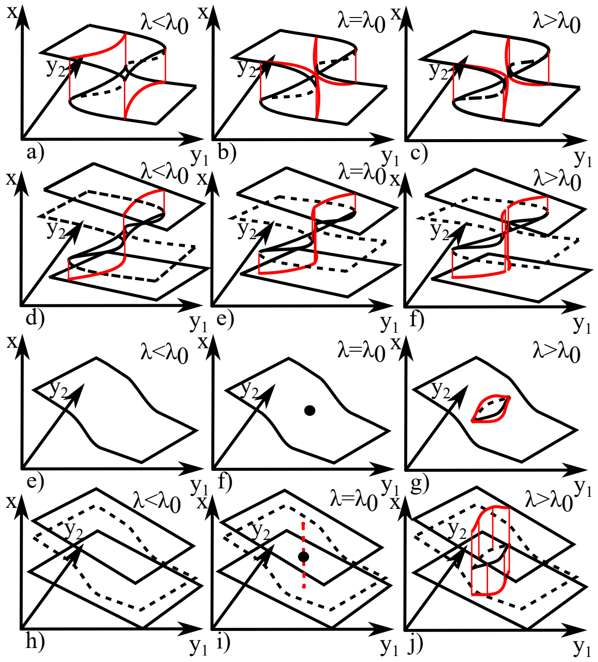


Figure C1. (Color online) The middle column shows typical examples of the class $\mathcal{D}_2$ codimension 1 bifurcations of the critical set for $m = 1$, $n = 2$. In all cases solid/dashed black lines show stable/unstable sheets of the critical set and red lines show the image of the fold under the umbral map.

ureCM121D2

which we recognise can be written (up to constant scaling)

$$\mu[g] = \frac{g_{xxx}}{\mathrm{proj}_{\nabla_y^\perp g} \nabla_y g_x} \nabla_y^\perp g, \tag{B.8}$$

where $\mathrm{proj}_u\, v = (u \cdot v)u/|u|^2$ is the projection of a vector v onto a vector u. Hence, the cusp is always perpendicular to the gradient, with magnitude inversely proportional to the projection of the gradient of g_x onto the gradient perpendicular. As a consequence, the magnitude of the cusp vector blows up (becomes undefined) if $\nabla_y g_x$ is parallel to the gradient (its perpendicular component vanishes).

Appendix C. Global codimension one bifurcations for one fast and two slow variables

al12tables

Tables C1, C2 and C3 list the various inequivalent subclasses of degeneracies $\mathcal{D}_{4,5,6}$ in Table 6. We also include a number of figures which illustrate these degeneracies.

Table C1. Subclasses of special *global* degeneracies for one fast and two slow variables with tangency of fold projection. $P(y)$ is the set of all singular points of the vector field g with slow coordinate y. $\nu[g](p)$ is the direction vector of a fold at a point p and $Z[g](p)$ is the scalar quadratic curvature of a fold (see the text for details). f=fold, *fu*=fold umbra, *fx*=non-interacting fold

Aligned umbra-dominant fu-f tangency:	$\mathcal{D}_{4,1}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) > 0$, and $Z[g](p_1) < Z[g](p_2)$ for some $p_1, p_2 \in P(y)\}$	Fig. C3 a,b,c)
Aligned fold-dominant fu-f tangency:	$\mathcal{D}_{4,2}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) > 0$, and $Z[g](p_1) > Z[g](p_2)$ for some $p_1, p_2 \in P(y)\}$	Fig. C3 a,b,c)
Aligned fu-fu tangency:	$\mathcal{D}_{4,3}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C3 d,e,f)
Aligned fx-fx tangency:	$\mathcal{D}_{4,4}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $\forall p_1$, $U[g](p_1) \neq p_2, U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C3 g,h,i)
Opposed covering fu-f tangency:	$\mathcal{D}_{4,5}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 a,b,c)
Opposed covering fu-fu tangency:	$\mathcal{D}_{4,6}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 d,e,f)
Opposed covering fx-fx tangency:	$\mathcal{D}_{4,7}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $\forall p_1$, $U[g](p_1) \neq p_2, U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 g,h,i)
Opposed non-covering fu-f tangency:	$\mathcal{D}_{4,8}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 j,k,l)
Opposed non-covering fu-fu tangency:	$\mathcal{D}_{4,9}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 m,n,o)
Opposed non-covering fx-fx tangency:	$\mathcal{D}_{4,10}[g,y] = \{P(y) \subset \mathcal{D}_4^1[g] : \lvert P(y)\rvert = 2$ and $\forall p_1$, $U[g](p_1) \neq p_2, U[g](p_2)$ and $\nu[g](p_1) \cdot \nu[g](p_2) < 0$ and $Z[g](p_1) + Z[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C4 p,q,r)

`s12P2Folds`

Table C2. Subclasses of special *global* degeneracies for one fast and two slow variables involving the intersection of projections of a fold and cusp. $P(y)$ is the set of all singular points of the vector field g with slow coordinate y. $\nu[g](p)$ and $\mu[g](p)$ are direction vectors of folds and cusps respectively (see the text for details). f=fold, *fu*=fold umbra, *uc*=unstable cusp, *sc*=stable cusp,*ucu*=unstable cusp umbra, *fx*=non-interacting fold, *cx*=non-interacting cusp

Aligned fu-sc intersection:	$\mathcal{D}_{5,1}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $g_{xxx}(p_2) < 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C5 d,e,f)
Opposed fu-sc intersection:	$\mathcal{D}_{5,2}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_1) = p_2$ and $g_{xxx}(p_2) < 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C5 a,b,c)
Aligned f-ucu intersection:	$\mathcal{D}_{5,3}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_2) \cap p_1 \neq \emptyset$ and $g_{xxx}(p_2) > 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 d,e,f)
Opposed f-ucu intersection:	$\mathcal{D}_{5,4}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_2) \cap p_1 \neq \emptyset$ and $g_{xxx}(p_2) > 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 a,b,c)
Aligned fu-ucu intersection:	$\mathcal{D}_{5,5}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_2) \cap U[g](p_1) \neq \emptyset$ and $g_{xxx}(p_2) > 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 j,k,l)
Opposed fu-ucu intersection:	$\mathcal{D}_{5,6}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $U[g](p_2) \cap U[g](p_1) \neq \emptyset$ and $g_{xxx}(p_2) > 0$ and $\nu[g](p_1) \cdot \mu[g](p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 g,h,i)
Aligned fx-scx intersection:	$\mathcal{D}_{5,7}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $(U[g](p_1) \cup p_1) \cap (U[g](p_2) \cup p_2) = \emptyset$ and $\nu[g](p_1) \cdot \mu[g](p_2) > 0$ and $g_{xxx}(p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 p,q,r)
Opposed fx-scx intersection:	$\mathcal{D}_{5,8}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $(U[g](p_1) \cup p_1) \cap (U[g](p_2) \cup p_2) = \emptyset$ and $\nu[g](p_1) \cdot \mu[g](p_2) < 0$ and $g_{xxx}(p_2) < 0$, for some $p_1, p_2 \in P(y)\}$	Fig. C6 m,n,o)
Aligned fx-ucx intersection:	$\mathcal{D}_{5,9}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $(U[g](p_1) \cup p_1) \cap (U[g](p_2) \cup p_2) = \emptyset$ and $\nu[g](p_1) \cdot \mu[g](p_2) > 0$ and $g_{xxx}(p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	
Opposed fx-ucx intersection:	$\mathcal{D}_{5,10}[g,y] = \{P(y) \subset \mathcal{D}_5^1[g] : \lvert P(y)\rvert = 2$ and $(U[g](p_1) \cup p_1) \cap (U[g](p_2) \cup p_2) = \emptyset$ and $\nu[g](p_1) \cdot \mu[g](p_2) < 0$ and $g_{xxx}(p_2) > 0$, for some $p_1, p_2 \in P(y)\}$	

s12P2Cusps

Table C3. Subclasses of special *global* degeneracies for one fast and two slow variables that involve $|P(y)| = 3$ singular points. $P(y)$ is the set of all singular points of the vector field g with slow coordinate y. $\nu[g](p)$ are direction vectors of folds (see the text for details). *f*=fold, *fu*=fold umbra, *fx*=non-interacting fold

Aligned fu-f f-fu intersection:	$\mathcal{D}_{6,1}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_2) = p_3$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) > 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 d,e,f)
Opposed fu-f f-fu intersection:	$\mathcal{D}_{6,2}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_2) = p_3$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) < 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 d,e,f)
Aligned fu-f fu-fu intersection:	$\mathcal{D}_{6,3}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_2) = U(p_3)$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) > 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 g,h,i)
Opposed fu-f fu-fu intersection:	$\mathcal{D}_{6,4}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_2) = U(p_3)$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) < 0\}$	Fig. C8 g,h,i)
Aligned fu-f fx-fx intersection:	$\mathcal{D}_{6,5}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_i) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p_i)) = \emptyset,\ \forall p_i \neq p_1$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) > 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 j,k,l)
Opposed fu-f fx-fx intersection:	$\mathcal{D}_{6,6}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = p_2$ and $U[g](p_i) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p_i)) = \emptyset,\ \forall p_i \neq p_1$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) < 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 j,k,l)
Aligned fu-fu fx-fx intersection:	$\mathcal{D}_{6,7}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = U[g](p_2)$ and $U[g](p_i) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p_i)) = \emptyset,\ \forall p_i \neq p_1$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) > 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 m,n,o)
Opposed fu-fu fx-fx intersection:	$\mathcal{D}_{6,8}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_1) = U[g](p_2)$ and $U[g](p_i) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p_i)) = \emptyset,\ \forall p_i \neq p_1$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) < 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 m,n,o)
Aligned fx-fx fx-fx intersection:	$\mathcal{D}_{6,9}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p_i) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p_i)) = \emptyset,\ \forall p_i$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) > 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 p,q,r)
Opposed fx-fx fx-fx intersection:	$\mathcal{D}_{6,10}[g,y] = \{P(y) \subset \mathcal{D}_6^1[g] : \|P(y)\| = 3$ and $U[g](p) \cap (P(y) \cup U[g](P(y)) \setminus U[g](p)) = \emptyset,\ \forall p \in P(y)$ and $(\nu[g](p_1) + \nu[g](p_2)) \cdot (\nu[g](p_1) + \nu[g](p_3)) < 0$ for some $p_1, p_2, p_3 \in P(y)\}$	Fig. C8 p,q,r)

racies12P3

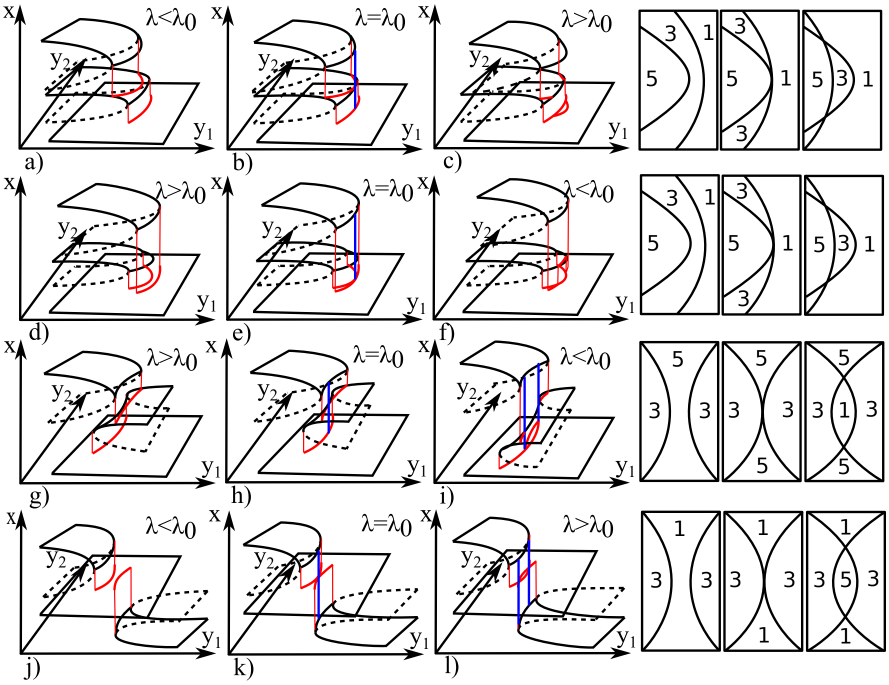
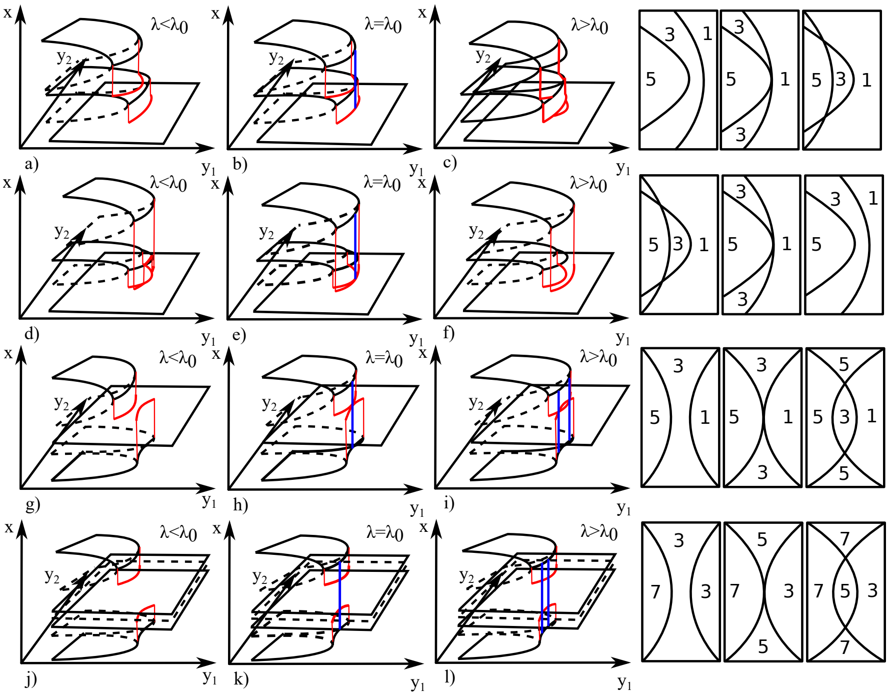


Figure C2. (Color online) The middle column shows global fold projection tangency bifurcations (class $\mathcal{D}_4$). Solid/dashed black lines show the stable/unstable sheets of the critical set, red lines show the image of the fold under the umbral map and blue lines indicate tangency of projections of fold sheets. Right: the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This encodes the direction from which the quadratic fold curves approach.

ureCM121D4



1D4Aligned

Figure C3. (Color online) The middle column shows global aligned fold projection tangency bifurcations (class $\mathcal{D}_4$). Solid/dashed black lines show the stable/unstable sheets of the critical set, red lines show the image of the fold under the umbral map and blue lines indicate tangency of projections of fold sheets. Right: the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This encodes the direction from which the quadratic fold curves approach.

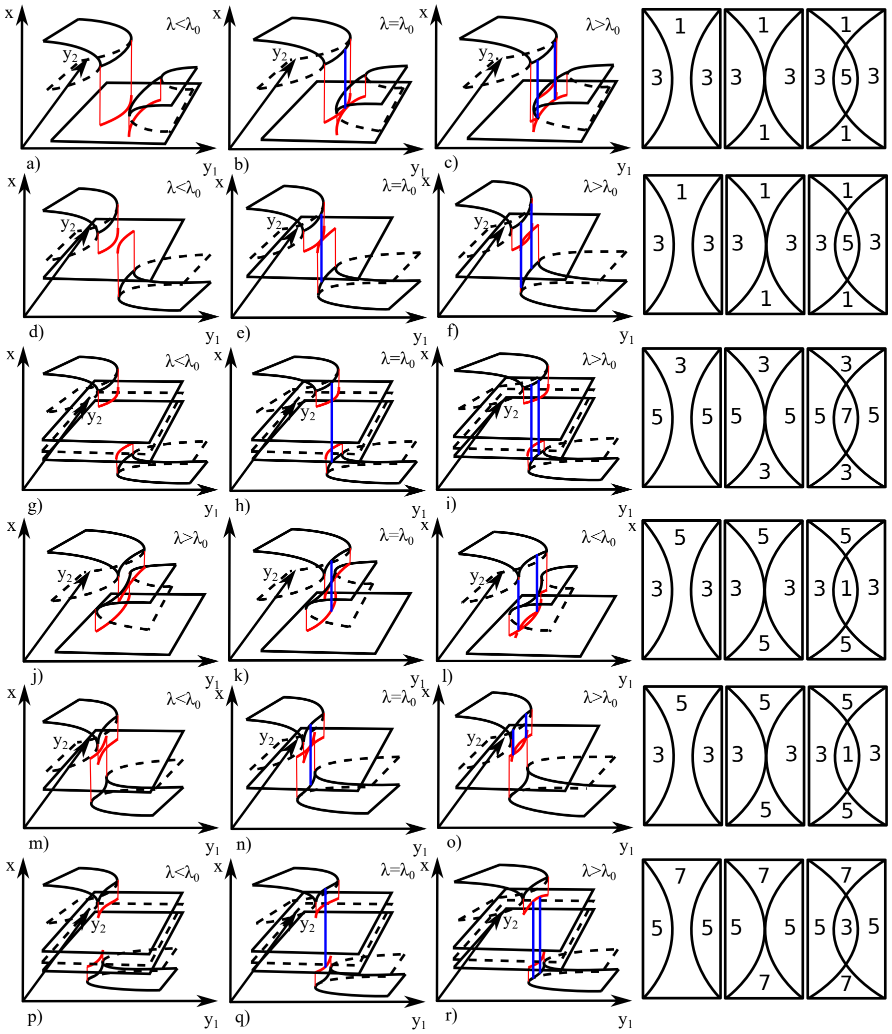


Figure C4. (Color online) The middle column shows global opposed fold projection tangency bifurcations (class $\mathcal{D}_4$). Solid/dashed black lines show the stable/unstable sheets of the critical set, red lines show the image of the fold under the umbral map and blue lines indicate tangency of projections of fold sheets. Right: the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This encodes the direction from which the quadratic fold curves approach.

1D4Opposed

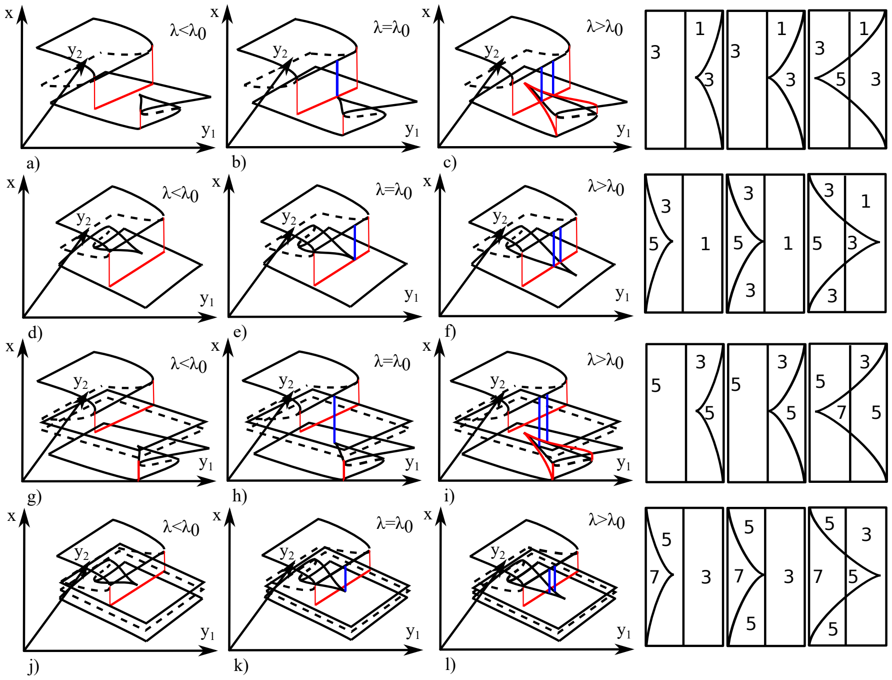


Figure C5. (Color online) The middle column shows global cusp-fold intersection bifurcations (class $\mathcal{D}_5$). Solid/dashed black lines show the stable/unstable critical set, red lines show umbrae, and blue lines show points where the slow projections of the cusp fold line intersect. Right: the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This encodes the direction from which the quadratic fold curves approach.

M121D5Stab

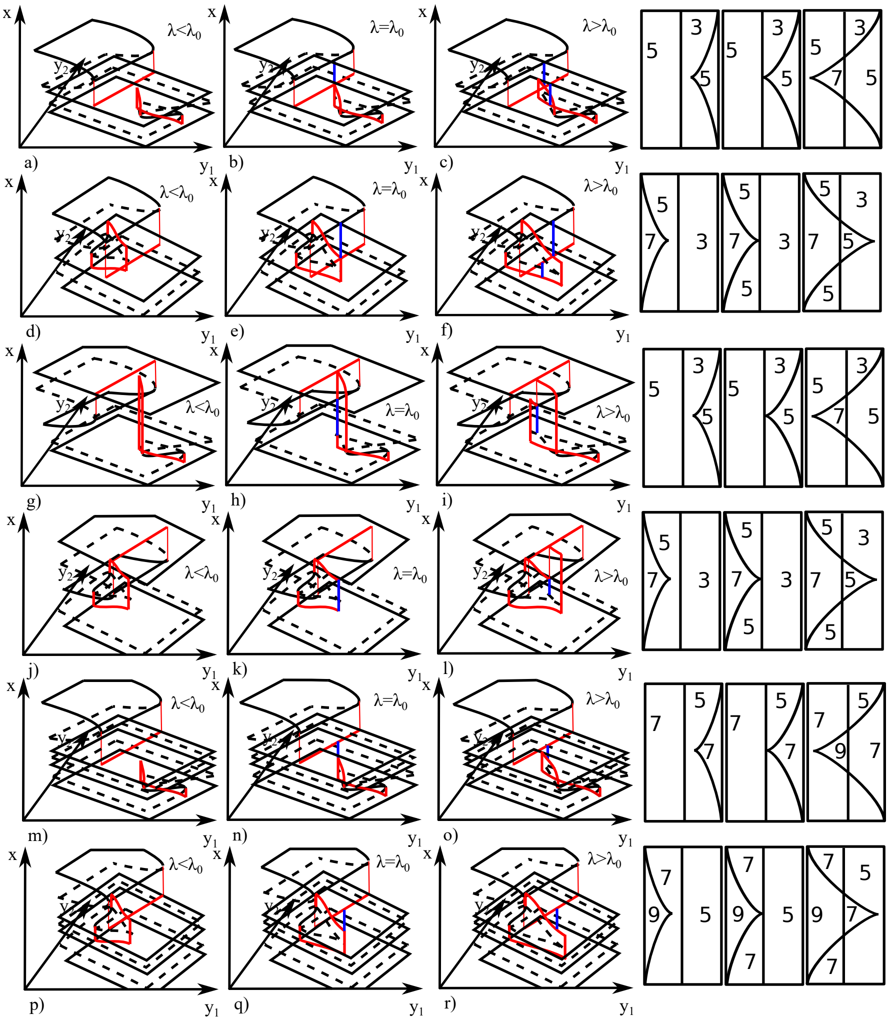


Figure C6. (Color online) The middle column shows typical examples of the unstable class D4 (2) of one-parameter ($p = 1$) structural bifurcations of the critical set for $m = 1$, $n = 2$. In all cases solid black lines show the stable/unstable critical set, red lines show the image of the fold under the umbral map, and blue lines show intersections of the images of folds for the umbral map. To the right of each triptych is a sketch of the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This sketch is a signature of the bifurcation, which encodes the direction from which the quadratic fold curves approach.

.21D5Unstab

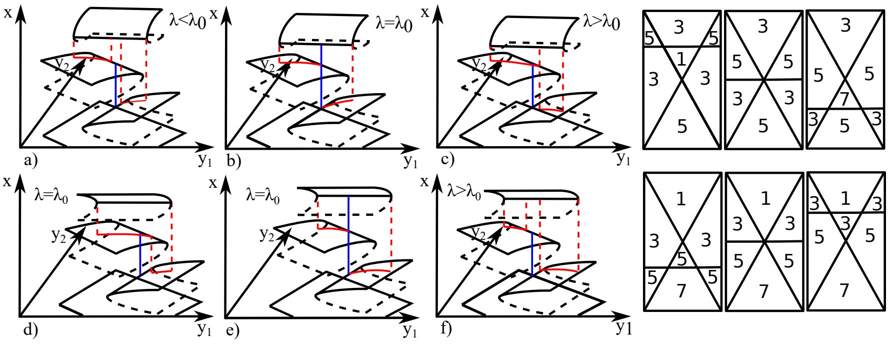


Figure C7. (Color online) The middle column shows an opposed a,b,c) and aligned d,e,f) fold-fumbra fold-fumbra triple limit point bifurcation of the critical set (class $\mathcal{D}_6$). Solid/dashed black curves show the stable/unstable critical set, red curves show umbrae, and blue curves show intersections of (at least) the lower two folds. Right: Sketch of the number of fixed points of $g(x, y)$ through the bifurcation, to be viewed as a projection on the slow variables. This sketch encodes the direction from which the quadratic fold curves approach.

ureCM121D6

Appendix D. Examples of bifurcations of relaxation oscillations for one fast and one slow variable

ıplesection

In this section we present the equations for the example fast-slow systems for $m = n = 1$, showing bifurcations of relaxation oscillations due to the critical manifold in Figure 11, how they were constructed, and how the figures were produced.

Appendix D.1. Bifurcation of relaxation oscillation due to hyperbolic fold tangency

We seek fast and slow subsystems $g(x, y)$ and $h(x, y)$ such that (1) displays bifurcation of singular relaxation oscillations due to hyperbolic fold tangency bifurcation (Figure 11 a,b,c)).

The fast subsystem $g(x, y)$ is written as a perturbed product of a hysteresis curve and a circle:

$$\begin{aligned} g_{hyst}(x, y) &= x^3 - 2x + y \\ g_{circ}(x, y) &= (x - \lambda)^2 + (y - y_c)^2 - R^2 \\ g(x, y) &= -(g_{hyst}(x, y) g_{circ}(x, y) + \lambda x + q). \end{aligned} \tag{D.1}$$

$(x_c, y_c) = (0.81, -0.25)$ is the centre and $R = 0.55$ the radius of the circle, λ is the bifurcation parameter and $q = 0.01$ is a genericity parameter. For some $\lambda \in [-0.02, 0.02]$ tangency bifurcation off the critical set occurs.

The slow subsystem is taken to be

$$h(x, y) = x - (-b(y - y_c)^2 + x_{max}),$$

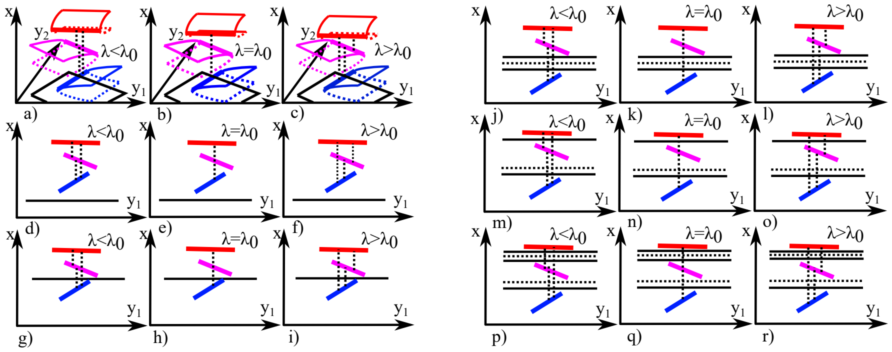


Figure C8. (Color online) All subclasses of triple limit point degeneracy (Table C3), to be viewed almost as a projection onto the (x, y_1) plane. Solid/dashed black horizontal curves show the stable/unstable critical set, and solid vertical lines show coordinates where folds intersect transversally. Red, blue and magenta lines are representations of three fold lines (with an illustration for the a,b,c) case). A bifurcation occurs as all three intersections of slow projections of folds merge. Each case can be either opposed or aligned, as shown in Fig. C7.

D6Sideview

where $b = 0.5$ and $x_{max} = x_c + R - 0.1$. This choice of $h(x, y)$ makes the nullcline $h(x, y) = 0$ intersect the critical set at where it is unstable, for $\lambda \in [-0.02, 0.02]$.

In Figure 11 a,b,c) we solve (1) with Matlab's stiff solver `ode23s` for 2000 time units, starting from initial conditions $(x, y) = (2, -1)$ and with scale separation parameter $\epsilon = 0.02$.

Appendix D.2. Aligned double limit point bifurcation of relaxation oscillations

We seek fast and slow subsystems $g(x, y)$ and $h(x, y)$ such that (1) displays bifurcation of singular relaxation oscillations due to aligned double limit point bifurcation (Figure 11 d,e,f)).

$g(x, y)$ is constructed such that the critical set $g(x, y) = 0$ is a fifth order polynomial in x as a function of y with extrema at points x_1, x_2, x_3 and x_4, that is, g_x satisfies:

$$g_x(x, y) = a(x - x_1)(x - x_2)(x - x_3)(x - x_4)$$

We fix $x_1 = -1$,$x_3 = 1/2, x_4 = 5/4$ and leave a and x_2 as free parameters. We take g to be the primitive function of g_x with zero constant term (default of Matlab's `int` command) such that

$$g(x, y) = \int a(x - x_1)(x - x_2)(x - x_3)(x - x_4)dx - y,$$

and $g(0, 0) = 0$. We solve the linear pair of equations

$$g(x_1, 1) = g(x_3, 1) = 0$$

for a and x_2, giving $x_2 = -13/40$ and $a = 640/49$. Then we add a bifurcation parameter λ breaking the degeneracy, giving

$$g(x, y) = \int a(x - x_1)(x - x_2)(x - x_3)(x - x_4)dx - \lambda x - y.$$

In Figure 11, $\lambda \in [-0.1, 0.1]$. Finally, we reverse the sign of x, such that

$$g(x, y) = \int -a(x + x_1)(x + x_2)(x + x_3)(x + x_4)dx + \lambda x - y.$$

The slow subsystem is set to be positive above the constant nullcline $x = x_{nc} = 1.5$ and negative below such that

$$h(x, y) = x - x_{nc}.$$

In Figure 11 d,e,f) we solve (1) with Matlab's stiff solver `ode23s` for 1000 time units, starting from initial conditions $x = y = 0$ and with scale separation parameter $\epsilon = 0.01$.

Appendix D.3. Opposed double limit point bifurcation of relaxation oscillations

We seek fast and slow subsystems $g(x, y)$ and $h(x, y)$ such that (1) displays bifurcation of singular relaxation oscillations due to opposed double limit point bifurcation (Figure 11 g,h,i)).

We construct the fast subsystem $g(x, y)$ the perturbed product of a hysteresis curve and a "bean" curve from `http://www.2dcurves.com/higher/highergb.html`

$$\begin{aligned} g_{hyst}(x, y) &= 0.5x^3 - x + y \\ g_{bean,base}(\hat{x}, \hat{y}) &= (\hat{x}^2 + \hat{y}^2)^3 - (\hat{x}^2 + (\hat{x}^2 + \hat{y}^2)^2\hat{y}^2) \\ g(x, y) &= -(g_{hyst}(x, y)g_{bean}(\hat{x}, \hat{y}) + \lambda x + q), \end{aligned} \tag{D.2}$$

where λ is a bifurcation parameter $\lambda \in [-0.003, 0.006]$, $q = 0.01$ is a genericity parameter, and $(\hat{x}, \hat{y})$ are scaled, rotated and translated coordinates (x, y):

$$(\hat{x}, \hat{y}) = (Mx \cos\theta + My \sin\theta - x_c, -Mx \sin\theta + My \cos\theta - y_c),$$

where $M = 1.5, \theta = 13/40\pi$ and $(x_c, y_c) = (0.97, -0.55)$.

The slow subsystem

$$h(x, y) = x - (ky + c),$$

with $k = (x_1 - x_2)/(y_1 - y_2)$, $c = x_1 - ky_1$, $x_1 = 0.7868$, $x_2 = 1.221$, $y_1 = -0.11$ and $y_2 = -0.74$ is chosen to make the nullcline $h(x, y) = 0$ pass through the unstable parts of the critical set and enable relxation oscillation.

In Figure 11 g,h,i) we integrate (1) with Matlab's stiff solver `ode23s` for 500 time units, starting from initial conditions $(x_0, y_0) = (2.3, -1)$ and with scale separation parameter $\epsilon = 0.001$.

Appendix D.4. Hysteresis bifurcation of relaxation oscillations

We seek fast and slow subsystems $g(x, y)$ and $h(x, y)$ such that (1) displays bifurcation of singular relaxation oscillations due hysteresis bifurcation (Figure 11 j,k,l)).

The function $g(x, y)$ is constructed such that the critical set $g(x, y) = 0$ at bifurcation $\lambda = 0$ is a fifth order polynomial in x as a function of y with quadratic extrema at points x_1, x_2 and a cubic double root at x_3, that is, g_x satisfies:

$$g_x(x, y) = a(x - x_1)(x - x_2)(x - x_3)^2 - \lambda$$

We fix $x_1 = 1$,$x_3 = 1$ and leave a and x_2 as free parameters. We take g to be the primitive function of g_x with constant term $q = -0.6$ and y-term $-y$ such that

$$g(x, y) = \int a(x - x_1)(x - x_2)(x - x_3)^2 dx - \lambda x - q - y,$$

and $g(0, 0) = 0$. For $\lambda = 0$ we solve the linear pair of equations

$$g(x_1, 1) = g(x_3, 1) = 0$$

for a and x_2, giving $x_2 = -13/40$ and $a = 640/49$. The system undergoes hysteresis bifurcation for $\lambda \in [-0.2, 0.2]$.

Finally, we reverse the sign of x, $x \mapsto -x$, such that

$$g(x, y) = \int -a(x + x_1)(x + x_2)(x + x_3)(x + x_4)dx + \lambda x - q - y.$$

The slow subsystem is set to be positive above the constant nullcline $x = x_{nc} = 0.7$ and negative below such that

$$h(x, y) = x - x_{nc}.$$

In Figure 11 j,k,l) we solve (1) with Matlab's stiff solver `ode23s` for 1000 time units, starting from initial conditions $x = y = 0$ and with scale separation parameter $\epsilon = 0.05$.

Appendix C

Co-authorship statements

PHD SCHOOL OF SCIENCE
UNIVERSITY OF COPENHAGEN

3A. Co-authorship statement

All papers/manuscripts with multiple authors which is part of a PhD thesis should contain a co-author statement, stating the PhD student's contribution to the paper

1. PhD student

Name:	Karl Hans Mikael Nyman	UCPH user id: *Or date of birth*	vrk331
Department:	Niels Bohr Institute		

2. Paper/Manuscript

This co-authorship declaration applies to the following:

Title:	The middle Pleistocene transition by frequency locking and slow ramping of internal period
Authors(s):	Karl H.M. Nyman and Peter D. Ditlevsen
Journal:	Climate Dynamics
Vol/page:	N/A
DOI:	10.1007/s00382-019-04679-3

3. Contributions to the paper/manuscript made by the PhD student

What was the role of the PhD student in designing the study?
Main designer
How did the PhD student participate in data collection and/or development of theory?
Performed analysis of models and data
Which part of the manuscript did the PhD student write or contribute to?
All parts
Did the PhD student read and comment on the final manuscript?
Yes

PHD SCHOOL OF SCIENCE
UNIVERSITY OF COPENHAGEN

3A. Co-authorship statement

All papers/manuscripts with multiple authors which is part of a PhD thesis should contain a co-author statement, stating the PhD student's contribution to the paper

1. PhD student

Name:	Karl Hans Mikael Nyman	UCPH user id: *Or date of birth*	vrk331
Department:	Niels Bohr Institute		

2. Paper/Manuscript

This co-authorship declaration applies to the following:

Title:	Bifurcation of critical sets and relaxation oscillations in singular fast-slow systems
Authors(s):	Karl Nyman and Peter Ashwin and Peter Ditlevsen
Journal:	Arxiv preprint
Vol/page:	N/A
DOI:	arXiv:1902.09203v2

3. Contributions to the paper/manuscript made by the PhD student

What was the role of the PhD student in designing the study?
Developed and expanded on a suggestion by Peter Ashwin, with input from Peter Ditlevsen
How did the PhD student participate in data collection and/or development of theory?
Developed theory in collaboration with Peter Ashwin, with input from Peter Ditlevsen
Which part of the manuscript did the PhD student write or contribute to?
Co-wrote all parts, including design of all figures and appendices
Did the PhD student read and comment on the final manuscript?
Yes