

---

# Data-based Master Equations for the Stratosphere

Mauro Dall'Amico

---



München 2005



---

# Data-based Master Equations for the Stratosphere

Mauro Dall'Amico

---

Dissertation  
an der Fakultät für Physik  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Mauro Dall'Amico  
aus Thiene

München, im Mai 2005

Erstgutachter: Prof. Dr. J. Egger

Zweitgutachter: PD Dr. M. Dameris

Tag der mündlichen Prüfung: 28. Juni 2005

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>Preface</b>	<b>viii</b>
<b>Zusammenfassung</b>	<b>ix</b>
<b>Sommario</b>	<b>x</b>
<b>Abstract</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Master equations in climate research . . . . .	1
1.2 Stratospheric modes and the troposphere . . . . .	3
<b>2 Data-based master equations</b>	<b>5</b>
2.1 Architecture of a master equation . . . . .	5
2.2 Estimating transition coefficients from time series . . . . .	8
2.3 Probability density and dynamics in phase space . . . . .	9
2.4 The Markovian assumption . . . . .	10
2.5 Correlation functions . . . . .	11
2.5.1 Estimate of correlations from time series . . . . .	11
2.5.2 Correlations given by a master equation . . . . .	11
<b>3 On the numerical properties of master equations</b>	<b>13</b>
3.1 The Lorenz model extended with Gaussian white noise . . . . .	13
3.2 Numerical parameters . . . . .	17
3.2.1 Grid size . . . . .	17
3.2.2 Time series length . . . . .	19
3.2.3 Time resolution . . . . .	24
3.2.4 Dimension of the master equation . . . . .	27
<b>4 Master equations for stratospheric time series</b>	<b>29</b>
4.1 Time series of stratospheric climate indices . . . . .	29
4.2 The QBO and the arctic stratosphere . . . . .	33
4.3 The role of the 11-year solar cycle . . . . .	41
4.4 The Arctic Oscillation in the stratosphere and in the troposphere . . . . .	48

<b>5</b>	<b>Conclusions</b>	<b>58</b>
<b>A</b>	<b>List of acronyms</b>	<b>61</b>
<b>B</b>	<b>List of symbols</b>	<b>62</b>
	<b>Bibliography</b>	<b>64</b>
	<b>Acknowledgements</b>	<b>69</b>
	<b>Curriculum Vitae of the author</b>	<b>70</b>

# List of Figures

2.1	Phase-line intervals and transition coefficients . . . . .	6
3.1	The Lorenz attractor . . . . .	15
3.2	Time series of the Lorenz model with stochastic forcing (LMSF) . . . . .	16
3.3	Trajectory of the LMSF . . . . .	17
3.4	Choice of an adequate grid size . . . . .	18
3.5	Noise-to-signal ratio for growing time series length . . . . .	19
3.6	Case study. Evolution of an ensemble of initial conditions . . . . .	20
3.7	PDF forecasts delivered by master equations (role of time series length) . . . . .	21
3.8	Observed state density $\rho$ for short time series of the LMSF . . . . .	22
3.9	Skills of some master-equation forecasts (role of time series length) . . . . .	23
3.10	Convergence of transition matrix $\mathbf{T}$ (role of the time series length) . . . . .	24
3.11	PDF forecasts delivered by master equations (role of time resolution) . . . . .	25
3.12	Understanding the role of time resolution . . . . .	25
3.13	Correlation functions (role of time resolution) . . . . .	26
3.14	A two-dimensional master equation (role of the number of variables) . . . . .	27
4.1	Preparation of a time series of $T'$ from the ERA-40 (daily) daily means . . . . .	32
4.2	Time series of $Q_I$ and $U_{e20}$ . . . . .	34
4.3	Time series of $T'$ . . . . .	35
4.4	Variable set $(Q_I, T', U_{e20})$ : distribution of states, phase-space partition . . . . .	36
4.5	Correlation functions, $Q_I$ , $T'$ , and $U_{e20}$ . . . . .	37
4.6	Evolution of a PDF cloud delivered by a master equation . . . . .	38
4.7	Mean trajectories in the $(Q_I, T', U_{e20})$ phase space . . . . .	39
4.8	Standard deviation of some PDF clouds . . . . .	40
4.9	A two-dimensional master equation for the stratosphere . . . . .	41
4.10	Time series of the solar radio flux at a wave length of 10.7 cm $S_{10.7}$ . . . . .	43
4.11	Variable set $(Q_I, T', S_{10.7})$ : distribution of states, phase space partition . . . . .	44
4.12	Correlation functions, $Q_I$ , $T'$ , and $S_{10.7}$ . . . . .	45
4.13	Mean trajectories $(Q_I, T', S_{10.7})$ , QBO East and solar maximum . . . . .	46
4.14	Mean trajectories $(Q_I, T', S_{10.7})$ , QBO West and solar maximum . . . . .	46
4.15	Mean trajectories $(Q_I, T', S_{10.7})$ , QBO East and solar minimum . . . . .	47
4.16	Mean trajectories $(Q_I, T', S_{10.7})$ , QBO West and solar minimum . . . . .	47
4.17	Time series of $A_{10}$ , $A_{100}$ , and $A_{850}$ , 1957 - 1972 . . . . .	50
4.18	Time series of $A_{10}$ , $A_{100}$ , and $A_{850}$ , 1972 - 1987 . . . . .	51
4.19	Time series of $A_{10}$ , $A_{100}$ , and $A_{850}$ , 1987 - 2002 . . . . .	52
4.20	Variable set $(A_{10}, A_{100}, A_{850})$ : distribution of states, phase space partition . . . . .	53

---

4.21	Correlation functions, Arctic Oscillation indices . . . . .	54
4.22	Mean trajectories in the $(A_{10}, A_{100})$ phase plane . . . . .	55
4.23	Mean trajectories in the $(A_{10}, A_{850})$ phase plane . . . . .	55
4.24	Mean trajectories in the $(A_{100}, A_{850})$ phase plane . . . . .	56



# List of Tables

3.1	Time series of the Lorenz model with stochastic forcing . . . . .	16
4.1	Time series of the variable set $(Q_I, T', U_{e20})$ . . . . .	34
4.2	Time series of the variable set $(Q_I, T', S_{10.7})$ . . . . .	42
4.3	Time series of the variable set $(A_{10}, A_{100}, A_{850})$ . . . . .	49

# Preface

The doctorate leading to this Ph.D. thesis began in January 2002 at the Meteorological Institute of the Ludwig-Maximilian University of Munich (LMU), Germany, and has been supervised by Professor J. Egger. The aim of this doctoral thesis is to improve the understanding of stratospheric climate variability by means of data-based master equations.

A reader coming from the north of the Alps (like Prof. Egger) will find an abstract in German on [page ix](#), which is required at a German University. A reader from the south of the Alps (like the author) will enjoy its Italian translation on [page x](#). The abstract in English follows on [page xi](#). The motivation and issues behind this work and the scientific context are presented in [chapter 1](#). [Chapter 2](#) is devoted to the architecture of data-based master equations. [Chapter 3](#) is dedicated to a study on the numerical properties of master equations. In [chapter 4](#) master equations are derived from stratospheric time series and the results discussed. The conclusions are presented in [chapter 5](#).

This thesis was written with L<sup>A</sup>T<sub>E</sub>X using the `hyperref` package. If you decide to read this thesis on a computer screen, you will reach all hypertexted references with a mouse-click. When citing this thesis, please make reference to the Munich University Library web page <http://edoc.ub.uni-muenchen.de/> where the thesis can be found among the physics faculty dissertations.

# Zusammenfassung

Dreidimensionale datengestützte Mastergleichungen (Englisch: *data-based master equations*) werden entwickelt und anschließend angewendet, um die Klimavariabilität der Stratosphäre zu untersuchen. Mastergleichungen werden eingesetzt um Verständnis für beobachtete Systeme zu entwickeln, wo keine dynamischen Gleichungen vorliegen.

In dieser Arbeit werden Mastergleichungen als prognostische Gleichungen für die Wahrscheinlichkeitsdichte in einem von Klimavariablen aufgespannten Phasenraum verwendet. Die Entwicklung der Wahrscheinlichkeitsdichte kann dann Auskunft über den Zusammenhang zwischen den Variablen geben. Der Phasenraum wird in etliche hundert Boxen mit gleichgroßer Maschenweite partitioniert, die jeweils Zuständen entsprechen, die vom System angenommen werden können. In dieser diskretisierten Version des Phasenraumes lassen sich die Koeffizienten der Mastergleichung aufgrund der relativen Häufigkeiten der Übergänge schätzen, die in einer aus Beobachtungen oder Modellläufen gewonnenen Zeitreihe der Variablen beobachtet werden.

Datengestützte Mastergleichungen sind numerische Strukturen, deren Erfolg u.a. von Auflösung und Umfang der verfügbaren Zeitreihen abhängt. Diese Abhängigkeiten werden anhand des bekannten dreidimensionalen Lorenzmodells der Konvektion untersucht, das hier mit einem stochastischen Antrieb erweitert wird. Zeitreihen mit beliebiger Länge und zeitlicher Auflösung können leicht generiert werden. Ferner können die Resultate gut verglichen werden. Die besten Ergebnisse werden bei langen Zeitreihen mit grober zeitlicher Auflösung erzielt. Weiterhin spielen die Wahl der Variablen und deren Anzahl eine ausschlaggebende Rolle.

Zu diesen letzten Ergebnissen führen auch aus den ERA-40 Reanalysen abgeleitete Zeitreihen stratosphärischer Klimaindizes, wobei die Stratosphäre als Anwendungsgebiet dient. Die Mastergleichung zeigt, dass die arktische Stratosphäre während der Ostphase der quasi-zweijährigen Schwingung (QBO) des äquatorialen stratosphärischen Zonalwindes etwa 2 K wärmer als während der Westphase ist. Somit kann der Zusammenhang zwischen QBO und arktischer Stratosphäre quantifiziert werden. Der Einfluß des 11-jährigen Sonnenzyklus wird von der Mastergleichung beschrieben. So stellt sich heraus, dass der Zusammenhang zwischen QBO und Temperaturanomalie der arktischen Stratosphäre eine Abhängigkeit von der Sonnenvariabilität zeigt. Die Auswirkungen von stratosphärischen Prozessen auf das Klima in der Troposphäre werden mit Hilfe einer Mastergleichung für Zeitreihen eines Index der arktischen Oszillation (AO) in stratosphärischen und troposphärischen Druckniveaus analysiert. Die Mastergleichung erfasst dabei die Hauptzüge dieser Wechselwirkung zwischen Stratosphäre und Troposphäre. So wird gezeigt, dass sich Anomalien der AO in der mittleren Stratosphäre tief in die Troposphäre hinein mit einer Zeitskala von 4 Wochen fortpflanzen. Ferner zeigt die Mastergleichung, dass der Einfluss starker troposphärischer AO-Anomalien auf die untere Stratosphäre beschränkt bleibt.

# Sommario

Equazioni master tridimensionali basate su dati (inglese: *data-based master equations*) vengono sviluppate e successivamente utilizzate per studiare la variabilità del clima nella stratosfera. Le equazioni master vengono implementate per sviluppare comprensione su sistemi che vengono osservati e per cui non esistono equazioni dinamiche.

In questa tesi le equazioni master vengono utilizzate come equazioni prognostiche per la densità di probabilità in uno spazio delle fasi tracciato da variabili climatiche. Lo sviluppo della densità di probabilità può poi dare informazioni sulle relazioni tra le variabili. Lo spazio delle fasi viene discretizzato in varie centinaia di box con maglia ugualmente grande, le quali rappresentano a loro volta degli stati che possono essere assunti dal sistema. In questa versione discretizzata dello spazio delle fasi i coefficienti dell'equazione master possono essere stimati sulla base delle frequenze relative delle transizioni osservate in una serie temporale ottenuta da osservazioni o dall'uso di modelli.

Le equazioni master basate su dati sono strutture numeriche il cui successo dipende tra altre cose dalla risoluzione e dal volume delle serie temporali disponibili. Queste interdipendenze vengono studiate tramite il famoso modello tridimensionale della convezione di Lorenz, che viene esteso introducendo una forzante stocastica. Serie temporali con una qualsiasi lunghezza e risoluzione possono essere generate facilmente e i risultati possono essere confrontati direttamente. I migliori risultati vengono ottenuti con lunghe serie temporali a bassa risoluzione temporale. La scelta ed il numero delle variabili giocano altresì un ruolo determinante.

A questi ultimi risultati portano anche serie temporali di indici climatici stratosferici derivate dalle reanalisi ERA-40. La stratosfera è dunque presa come dominio di applicazione. L'equazione master mostra che durante la fase est dell'oscillazione quasi biennale (QBO) del vento zonale equatoriale nella stratosfera, la stratosfera artica è circa 2 K più calda che durante la fase ovest. Così può essere quantificata la relazione tra la QBO e la stratosfera artica. L'influsso del ciclo solare undecennale viene descritto dall'equazione master, facendo emergere come la relazione tra la QBO e l'anomalia della temperatura della stratosfera artica mostrino una dipendenza dalla variabilità solare. Le implicazioni di processi stratosferici sul clima nella troposfera vengono studiate con l'aiuto di un'equazione master per serie temporali di un indice dell'oscillazione artica (AO) a livelli di pressione stratosferici e troposferici. L'equazione master rileva le caratteristiche di questa interazione tra stratosfera e troposfera, mostrando come anomalie della AO nella stratosfera media si propagano profondamente nella troposfera con una scala temporale di 4 settimane. Inoltre l'equazione master mostra che l'influsso di forti anomalie troposferiche della AO rimane circoscritto alla stratosfera inferiore.

# Abstract

Three-dimensional data-based master equations are developed and subsequently used to study climate variability in the stratosphere. Master equations are used to develop understanding of observed systems where no dynamic equations are available.

Master equations are used in this thesis as prognostic equations for the probability density in a discretized phase space spanned by climate variables. The evolution of the probability density may then reveal information about the relationship between these variables. The phase space is partitioned into several hundred boxes of equal grid size representing at any one time states that the system can assume. In this discretized version of the phase space, the coefficients of a master equation may be estimated from the relative frequencies of transitions observed in a time series of the variables obtained from observations or numerical model runs.

Data-based master equations are numerical structures whose success depends among other things on the resolution and volume of the available time series. These dependencies are studied on the basis of data from the famous three-component Lorenz convection model extended with a stochastic forcing. Time series of the desired length and time resolution can thus be generated easily. Furthermore, the results can be compared directly. Best results are obtained through the combination of a long data record and a coarse time resolution. The choice of the variables and their number also play a crucial role in the success of a master equation.

Time series of stratospheric climate indices obtained from the reanalyses ERA-40 lead also to these last results. The stratosphere serves now as an implementation area. The master equation shows that during the eastern phase of the quasi-biennial oscillation (QBO) of equatorial zonal wind the arctic stratosphere is about 2 K warmer than during the western phase. Thus the relationship between QBO and arctic stratosphere can be quantified. The influence of the 11-year solar cycle is described by the master equation. It emerges that the relationship between QBO and temperature anomaly of the arctic stratosphere shows a dependence on solar variability. The implications of stratospheric processes on the climate in the troposphere are analysed with a master equation for a time series of an index of the Arctic Oscillation (AO) at stratospheric and tropospheric pressure levels. The master equation captures the main features of this interaction between stratosphere and troposphere. It is shown that anomalies of the AO in the middle stratosphere propagate deeply into the troposphere with a time scale of 4 weeks. Furthermore the master equation shows that the influence of strong tropospheric AO-anomalies remains confined to the lower stratosphere.



# Chapter 1

## Introduction

The master equation is a prognostic equation for the probability density function (PDF) in the discretized phase space spanned by a few variables describing the state of a system. The master equation uses the probabilities for grid-box (i.e. state) transitions to predict changes of the PDF in the phase space (e.g. [Gardiner, 1983](#), p. 8-11). The evolution of the PDF may be investigated in order to deepen the understanding of the underlying system. State and evolution of the system are now described by master equations. Time series of a set of variables obtained from observations or from numerical model runs may be used to derive the coefficients of data-based master equations. These equations are powerful tools when long time series of processes are available but their properties are not fully understood. The analysis of master equations may also help to understand observed systems when no dynamic equations are available or when models of these systems are not able to describe all phenomena. The idea in this thesis is to study first the numerics of data-based master equations and subsequently apply them to time series of stratospheric and tropospheric climate indices. The aim is to improve the understanding of major stratospheric modes of oscillation which are highly relevant for the troposphere but neither fully understood nor correctly described by current models. Finally, the correctness of the master equation approach may be assessed by testing to which extent the inherent Markovian assumption<sup>1</sup> holds. The success of this approach may be qualitatively evaluated by checking whether the observed phenomena are described by the master equation.

### 1.1 Master equations in climate research

In climate research it is common practice to fit a statistic or stochastic model to time series of observed variables<sup>2</sup>. Familiar examples are the projection of data on a regression model and the principal oscillation analysis (e.g. [von Storch and Zwiers, 1999](#)). Fokker-Plank equations also predict the PDF and are related to master equations (see [Zwanzig, 2001](#)). The drift and diffusion terms of a Fokker-Plank equation may be estimated from data (see [Siegert et al., 1998](#)). This technique was applied, for example, by [Ditlevsen \(1999\)](#) to

---

<sup>1</sup>Master equations make use of the Markovian assumption (see [section 2.4](#)) which postulates that the probability distribution at time step  $n$  depends only on the probability distribution at the precedent time step  $n - 1$ .

<sup>2</sup>Stochastic models are becoming widely used not only in the natural sciences but also in industry (e.g. [Gradišek et al., 2000](#)) and in the economy.

ice-core data and by Egger and Jönsson (2002) to meteorological observations in the Icelandic region. Master equations are also not new in climate research. Spekat et al. (1983) analysed the zonal, mixed and meridional weather regimes (German: *Großwetterlagen*) from a 100-year-long time series on the basis of a first-order Markov chain<sup>3</sup>. Fraedrich (1988) applied a Markov chain model to the problem of estimating predictability time scales from annual time series of El Niño/Southern Oscillation (ENSO). Nicolis (1990) suggested the use of master equations for carrying out long-term statistical predictions of systems undergoing chaotic dynamics. Nicolis et al. (1997) discussed the possibility of using a master equation in the problem of predicting transitions between weather regimes. Egger (2001) derived master equations from time series of the equatorial components of the global angular momentum of the atmosphere and related torques obtained from the European reanalyses ERA-15<sup>4</sup> of the ECMWF<sup>5</sup> ([www.ecmwf.int](http://www.ecmwf.int)). A mean velocity and a diffusion coefficient in the phase plane were hence derived. An analysis of the dynamics in the phase plane of two variables at a time could be based on the relative master equation. Pasmanter and Timmermann (2002) developed the theory of cyclic Markov chains and applied it to the ENSO predictability problem. Thus they inspired the work of Crommelin (2004), who addressed the issue of atmospheric circulation regimes in northern hemisphere winter with a similar Markov chain description.

In chapter 2, the approach of Egger (2001, 2002) is elaborated and extended to a set of three key climate variables. However, data-based master equations are numerical structures whose numerical features are not yet well known. For instance, Egger (2001) pointed out that the number of variables must be small and that master equations with more than three variables may hardly be solved. In fact, fairly long data records are needed to correctly estimate the coefficients of a master equation, and by increasing the number of variables involved (i.e. by increasing the phase space dimension) the required amount of data increases dramatically. Limitations are also imposed by the available computer resources. Moreover, the success of a master equation depends not only on the amount of data but also on their time resolution. Further, there are several approaches towards partitioning the phase space (e.g. Nicolis, 1990, Egger, 2001, Pasmanter and Timmermann, 2002), and there exist only rules-of-thumb to assess an adequate level of partitioning. By introducing the Lorenz convection model with white noise in chapter 3, an appropriate way is found to deal with these questions. Time series of the desired characteristics can in fact be generated easily. Moreover, one knows the model equations and may thus compare directly the evolution of the PDF as predicted by the master equation to the evolution of ensembles of initial conditions. In a real system this would not be possible because there is just one realization of the system. By means of the extended Lorenz model, a comprehensive approach towards the numerics of data-based master equations is developed, which encompasses discretisation level, time series characteristics, and choice and number of the variables.

The knowledge gained on the numerics of data-based master equations is applied in chapter 4 to the stratosphere. The coefficients of several data-based master equations will be derived from time series of stratospheric and tropospheric climate indices<sup>6</sup>. Master

---

<sup>3</sup>First-order Markov chains are closely related to the master equations used in this thesis.

<sup>4</sup>ERA-15 covers the period January 1979 - December 1993.

<sup>5</sup>Acronym for European Centre for Medium-Range Weather Forecasts.

<sup>6</sup>These climate indices are thus the variables of the analysed system.



equations for three variables at any one time will be derived and studied in order to learn more about important modes of variability of the stratosphere that are highly relevant for the climate in the troposphere.

## 1.2 Stratospheric modes and the troposphere

The stratosphere plays a crucial role in the climate system through its radiative and dynamical processes. In recent years there has been a growing realization that stratospheric effects might influence tropospheric climate (see [Houghton et al., 2001](#), p. 432-435). In order to correctly model and interpret climate change, it is therefore essential to understand and reproduce the stratospheric modes of variability. Moreover, in order to extend the range and improve the skill of tropospheric weather prediction, a great effort has been devoted to study how the troposphere is coupled with the stratosphere (e.g. [Baldwin and Dunkerton, 1999](#), [Baldwin et al., 2003](#)) and the oceans. The stratosphere undergoes a great degree of intraseasonal and interannual climate variability. A part of the climate variability in the stratosphere depends on internal mechanisms within the atmosphere, such as the quasi-biennial oscillation (QBO) of stratospheric equatorial zonal wind. Another part of the stratospheric climate variability is externally forced, for example by anomalies in solar radiation or volcanic eruptions (see [Labitzke, 1998](#), [Labitzke and van Loon, 1999](#)).

These modes of oscillation of the stratosphere are relevant also for the troposphere. For example, the QBO influences not only an important part of the stratospheric circulation ([Holton and Tan, 1980](#)), but by influencing also the atmospheric ozone it affects the intensity of the ultraviolet light reaching the biosphere (e.g. [Stolarski et al., 1991](#)). Great efforts are being made in trying to understand the role of the 11-year solar cycle (SC), whose effects are seen in the stratosphere and in the troposphere (see [Labitzke, 1998](#), [van Loon and Labitzke, 1998](#), [Labitzke and van Loon, 1999](#)). The interaction of the SC with the QBO was highly debated in the late 1990s (e.g. [Salby and Callaghan, 2000](#), and references therein). Recent studies suggest possible mechanisms through which the SC interacts with the QBO influencing the global circulation of the atmosphere (e.g. [Kodera and Kuroda, 2002](#), [Labitzke, 2004a](#), and references therein). Their combined effect tends to modify the Brewer-Dobson circulation and the behaviour of the polar vortices (e.g. [Labitzke, 2004a,b](#)). The occurrence and the timing of stratospheric sudden warmings are also influenced by the SC and the QBO ([Gray et al., 2004](#)), and pronounced changes in the stratospheric north polar vortex have shown to be followed by anomalies in the surface air temperature ([Thompson et al., 2002](#)) and changes in surface weather patterns ([Thompson and Wallace, 2001](#)). The dynamical coupling of the troposphere with the stratosphere<sup>7</sup> is a debated issue (e.g. [Haynes et al., 1991](#), [Egger, 1996](#), [Haynes et al., 1996](#)). In this thesis we focus on the Arctic Oscillation (AO), which is one of the major modes of oscillation in the northern extra-tropics. In the stratosphere the AO is linked to the temperature in the arctic region. In the troposphere over the North Atlantic the AO appears as the North Atlantic Oscillation (NAO) ([Wallace, 2000](#)). [Baldwin and Dunkerton \(1999\)](#) suggested that wintery anomalies of the AO propagate downward into the troposphere. These works motivate an improved understanding of the AO, also referred to as

---

<sup>7</sup>The issue of the stratosphere-troposphere exchange, highly relevant in topics concerned with atmospheric chemistry, is not addressed in this thesis.

the northern annular mode (NAM). The phase of the NAO/AO is a very important climate index for middle latitudes in the northern hemisphere, especially for Europe and the western United States, and Canada (Qian et al., 2000, Thompson et al., 2002).

These long-term modes of oscillation account for a part of the natural climate variability and may indeed change as a consequence of anthropogenic climate change. Thus it is highly important to know and understand these modes while evaluating past or even modelled climate change. In this light it appears quite critical that most current atmospheric general circulation models do not simulate such phenomena as the QBO (Houghton et al., 2001). It is still not clear which mechanisms are actually acting (e.g. Labitzke, 1998, Labitzke and van Loon, 1999, Rind, 2002). For these reasons, the stratosphere provides an excellent test bed for data-based master equations. Besides the QBO we will investigate the role of the 11-year SC, and the AO at stratospheric and tropospheric pressure levels.

These modes are large-scale phenomena. However, data-based master equations are not suited for data fields. As discussed in section 1.1, the number of variables to be included in a master equation must be small. For this reason, several indices are defined in chapter 4 in order to reduce the main features of an observed phenomenon to a single figure, which indeed varies in time. This will be achieved with spatial averaging of meteorological variables such as wind components, temperature, geopotential height, and their difference between selected regions. For example, the phase of the QBO may be expressed with the normalised and de-seasonalised zonal wind at 30 hPa above the equator. This is the QBO index  $Q_I$  which we will use in chapter 4. Time series of these indices will be obtained from the European reanalyses ERA-40 of the ECMWF (Uppala et al., 2005), which cover the period September 1957 - August 2002, and from a time series of the solar radiation at a wavelength of 10.7 cm. Master equations for three indices at any one time will be derived and discussed in order to bring more insight into stratospheric climate variability.

# Chapter 2

## Data-based master equations

Master equations describe the dynamics of a system in terms of transitions between states. Instead of considering, say, velocities, momenta, or chemical concentrations, they forecast the PDF in the discretized phase space spanned by (some of) the system's variables. Depending on the variables, master equations may thus be used at different degrees of abstraction (Zwanzig, 2001, p. 61-63). This thesis is aimed to a deeper understanding of stratospheric climate variability. The distance between the master equations and the dynamical equations describing state and evolution of the stratosphere and, consequently, the level of abstraction are quite high. In this chapter, data-based master equations are presented, and the significance problem is dealt with. Some tools to visualise the development of the PDF are developed. The Markovian assumption inherent in the master equation is discussed, and methods to estimate the correlation functions are presented in order to test its fulfilment.

### 2.1 Architecture of a master equation

Let us consider a phase space spanned by three meteorological variables, say, a wind component at some location, a temperature averaged over a given area or volume and a geopotential height difference. To fully describe the state of the atmosphere in that particular area one should introduce (and know) many more variables. This is, however, not possible in reality. Instead one can describe state and evolution of the atmosphere with regards to a few variables, leading to a trajectory in phase space. One will note, for instance, that a northerly wind precedes temperature drops and that the two show some connection with the geopotential height difference at some time. Even though these variables and time are continuous quantities, time series of measurements will indeed be treated as discrete in time. Moreover, when these quantities are recorded on some digital medium, they will inevitably undergo discretisation at some accuracy level. The final user may nevertheless want to know whether the temperature was between  $0^{\circ}\text{C}$  and  $-10^{\circ}\text{C}$  or else lower than  $-10^{\circ}\text{C}$  so that the temperature axis will be *partitioned* into, say, 8 to 10 intervals<sup>1</sup> and the measurements grouped according to the interval they fall in. In some other case one may be interested in the statistical distribution of atmospheric states. In this case a probability

---

<sup>1</sup>For example, following temperature intervals may be considered:  $[-40^{\circ}\text{C}, -30^{\circ}\text{C}[$ ,  $[-30^{\circ}\text{C}, -20^{\circ}\text{C}[$ ,  $[-20^{\circ}\text{C}, -10^{\circ}\text{C}[$ ,  $\dots$ ,  $[+20^{\circ}\text{C}, +30^{\circ}\text{C}[$ , and  $[+30^{\circ}\text{C}, +40^{\circ}\text{C}[$ .

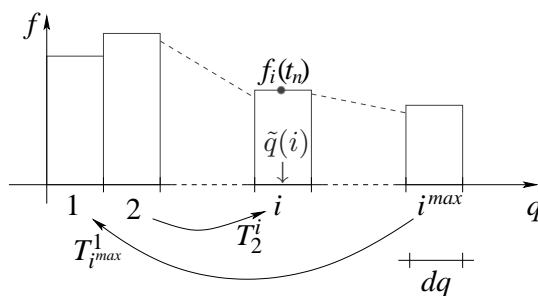


Figure 2.1: The transition coefficient  $T_2^i$  gives the probability that the variable  $q$  leaves segment (2) to enter ( $i$ ) at the next time step.

distribution, or better, a PDF value will be calculated on the base of observations or model runs. Each temperature interval will be assigned a value of the PDF.

For the sake of simplicity, let us first limit our considerations to the case of one variable  $q$  and then extend results to the multivariate case. The phase space is thus reduced to an axis. Phase space partitions may of course be done in many different ways. In this thesis we will consider partitions into intervals with equal grid size  $dq$ . Other possibilities will be discussed in [section 2.4](#). In [Fig. 2.1](#), for instance, the region of the  $q$ -axis where the PDF is non-zero is partitioned into  $i^{max}$  intervals with equal width  $dq$ ; this region extends between  $q^{min} = 0$  and  $q^{max} = i^{max} dq$ . The centre of interval ( $i$ ) is located at  $\tilde{q}(i) = q^{min} + (i - \frac{1}{2}) dq$  (see [Fig. 2.1](#)). Let us assume that observations are available at discrete points in time  $t_n = n dt$ , where  $dt$  is the time interval; time is also considered as a discrete variable. Any state satisfying  $q^{min} + (i - 1) dq \leq q < q^{min} + (i) dq$  is located in the interval ( $i$ ). The PDF is defined in the following way:  $f_i(t_n) dq$  is the probability that the state variable  $q$  is in the interval ( $i$ ) at time  $t_n$ . It is normalised so that

$$\sum_{i=1}^{i^{max}} f_i(t_n) dq = 1$$

for every  $n$ . Obviously  $0 \leq f_i(t_n) \leq \frac{1}{dq}$ . The master equation describes dynamics in this discretized version of the phase space ([Zwanzig, 2001](#)):

$$\frac{d}{dt} f_i(t) = \sum_{i'} \mathcal{T}_{i'}^i f_{i'}(t) - \sum_{i'} \mathcal{T}_i^{i'} f_i(t) \quad (2.1)$$

where  $\mathcal{T}_i^{i'}$  is the transition rate from interval ( $i$ ) to interval ( $i'$ ). After time discretisation, (2.1) is replaced by:

$$\frac{f_i(t_{n+1}) - f_i(t_n)}{dt} = \sum_{i'} \mathcal{T}_{i'}^i f_{i'}(t_n) - \sum_{i'} \mathcal{T}_i^{i'} f_i(t_n)$$

In the rest of this thesis we will refer to the following form:

$$f_i(t_{n+1}) = f_i(t_n) + \sum_{i'} \mathcal{T}_{i'}^i f_{i'}(t_n) - \sum_{i'} \mathcal{T}_i^{i'} f_i(t_n) \quad (2.2)$$

where  $\mathcal{T}_i^{i'} = dt T_i^{i'}$ . The transition coefficients  $T_i^{i'}$  give the probability that the state variable  $q$  leaves the interval ( $i$ ) to enter ( $i'$ ) at the next time step (see [Fig. 2.1](#)). Gains

[losses] in PDF from [to] other intervals are described by the second [third] term on the right. Boundary conditions are not needed provided the domain of solution contains all available observations and these adequately represent all those states which the system may possibly assume.

Besides this gain-loss equation form, master equations may also be written in operator form:

$$f_i(t_{n+1}) = f_i(t_n) + dt \sum_{i'} \mathcal{D}_{i'}^i f_{i'}(t_n) \quad (2.3)$$

where

$$\mathcal{D}_{i'}^i = \mathcal{T}_{i'}^i - \delta_{ii'} \sum_p \mathcal{T}_{i'}^p \quad (2.4)$$

and again the first [second] term in (2.4) refers to gains [losses] from [to] other cells. Conservation of probability means that  $\sum_i \mathcal{D}_{i'}^i = 0$ . Further, transition rates cannot be negative, so that all off-diagonal terms of  $\mathcal{D}$  will be positive or zero. Thus, according to standard matrix theorems, matrix  $\mathcal{D}$  will have at least one zero eigenvalue and all other eigenvalues will have negative real parts describing an approach to equilibrium. Matrix  $\mathcal{D}$  could have more than one equilibrium state, so that different initial states can lead to different stationary solutions at infinite time (Zwanzig, 2001). From (2.3) follows:

$$\begin{aligned} f_i(t_{n+1}) &= \sum_{i'} \left[ dt \mathcal{T}_{i'}^i + \delta_{ii'} \left( 1 - dt \sum_p \mathcal{T}_i^p \right) \right] f_{i'}(t_n) \\ &= \sum_{i'} U_{i'}^i f_{i'}(t_n) \end{aligned}$$

so that the master equation may be written in matrix form<sup>2</sup>:

$$\mathbf{f}(t_{n+1}) = \mathbf{U} \cdot \mathbf{f}(t_n). \quad (2.5)$$

In this thesis we assume that the underlying systems are ergodic, so that there is only one equilibrium state (see also Spekat et al., 1983). The eigenvector associated with the biggest eigenvalue of matrix  $\mathbf{U}$ , i.e. the unit eigenvalue<sup>3</sup>, represents the ‘‘climatological’’ equilibrium distribution  $\bar{\mathbf{f}}$ . The PDF will tend for great time steps  $n$  to this climatological mean (Spekat et al., 1983). This equilibrium distribution is expected, at least for adequately large data records<sup>4</sup>, to be very similar to the observed state density  $\boldsymbol{\rho}$ :

$$\bar{f}_i \simeq \rho_i = \frac{N_i}{\sum_{i=1}^{i^{max}} N_i} \cdot \frac{1}{dq} \quad (2.6)$$

where  $N_i$  is the number of events when an observation falls in the interval  $i$ , and  $\boldsymbol{\rho}$  is normalised to  $\sum_{i=1}^{i^{max}} \rho_i dq = 1$ . Stationarity may be guaranteed by checking that no time

<sup>2</sup>Form (2.5) is sometimes referred to in literature as *first-order Markov chain* (e.g. Spekat et al., 1983, Pasmanter and Timmermann, 2002, Crommelin, 2004).

<sup>3</sup>See also the discussion on stochastic matrices and the Perron-Frobenius theorem in Pasmanter and Timmermann (2002).

<sup>4</sup>A few data-based master equations were derived only from data from the winter months. In some cases the equilibrium distribution was highly different from the observed climate. This could happen because while considering just a few ‘‘bits’’ of a time series, transition *to* some states away from the mean happen to be more frequent than transitions *from* these states.

series ends in grid boxes from where no transition is observed<sup>5</sup>. The value of the second biggest eigenvalue of  $\mathbf{U}$  describes how fast the system tends to the mean climate state. Thus  $(\ln 1/\lambda_2)^{-1}$  gives information about the ‘‘predictability’’ time scale (Spekat et al., 1983). However, it is not the aim of this thesis to discuss the predictability of stratospheric phenomena.

## 2.2 Estimating transition coefficients from time series

The transition coefficients  $T_i^{i'}$  (see Fig. 2.1) may be estimated from time series obtained from observations or numerical model runs. This technique was applied by Spekat et al. (1983) while estimating transition probabilities (i.e. conditioned probabilities) for the change from one weather regime to another. In Nicolis et al. (1997) a straightforward counting process within a daily weather regimes data set classified into three main clusters also led to the conditional probability matrix giving the estimated transition probabilities from one weather regime to another. Egger (2001, 2002) counted transitions between phase space grid boxes to estimate the transition coefficients of several master equations. Pasmanter and Timmermann (2002) used time series from an intermediate ENSO model run to compute transition matrices. Crommelin (2004) also calculated transition probability matrices by counting transitions.

In this thesis, transition coefficients are estimated from the relative frequencies of transitions observed in the time series. In (2.6)  $N_i$  represents the number of times an observation falls in  $(i)$ .  $M_i^{i'}$  is the number of transitions from  $(i)$  to  $(i')$  observed in the time series<sup>6</sup>. The transition coefficients are estimated as follows:

$$T_i^{i'} \sim \frac{M_i^{i'}}{N_i}. \quad (2.7)$$

Indeed, the reliability of this statistical estimate will be higher the longer the time series is. Crommelin (2004) introduced an approach to statistical significance which may be applied to equipartitions into very few cells<sup>7</sup> whenever one is interested in a few most meaningful state transitions. Vautard et al. (1990) introduced a significance test based on Monte Carlo simulations while examining a few atmospheric circulation pattern transitions<sup>8</sup>. However, in this thesis partitions into  $10^2$  to  $10^3$  and more cells are considered, so that even  $10^6$  transitions are possible. An individual consideration of the significance level of even a few selected transitions would therefore be senseless. Thus the significance problem is dealt with in the following way. The frequencies of transition from an interval  $(i)$  towards an interval  $(i')$  are expected to be approximately binomially distributed (see Spekat et al., 1983). This means that out of  $N_i$  transitions from interval  $(i)$ , we expect  $N_i \cdot T_i^{i'}$  to result in transitions to  $(i')$ .  $N_i \cdot T_i^{i'}$  is the expectation value from  $N_i$  independent binomial Bernoulli trials, where at each trial  $T_i^{i'}$  is the probability of ‘‘success’’ [i.e. transition to interval  $(i')$ ],

<sup>5</sup>Should a time series end in a grid box from where no transition is observed, the PDF would then slowly, but constantly accumulate in this grid box. The system would never reach stationarity and the PDF would eventually reach high values in this grid box.

<sup>6</sup>To conserve probability,  $N_i = \sum_{i'} M_i^{i'}$ . Therefore  $N_i$  is not increased when a time series *ends* in  $(i)$ .

<sup>7</sup>Crommelin (2004) considered phase-space partitions into up to 6 cells.

<sup>8</sup>In their analysis, Vautard et al. considered transitions between 7 clusters, so that their transition matrix included 49 elements (i.e. possible transitions).

and  $1 - T_i^{i'}$  the probability of “no success” (see [Spekat et al., 1983](#), [Crommelin, 2004](#)). For the central limit theorem, the distribution will tend, for great sample size, to a normal distribution. The 90% confidence intervals of the transition coefficients may be easily calculated [recall (2.7)]:

$$c_i^{i'} = 1.645 \sqrt{\frac{T_i^{i'} (1 - T_i^{i'})}{N_i}}.$$

In general, there exist only rules-of-thumb with respect to the length of a data record in order for such confidence intervals to be small<sup>9</sup> (e.g. [Spekat et al., 1983](#), [Crommelin, 2004](#)). We now want to obtain a figure that describes the overall significance of the transition coefficients in phase space. Thus use is made of a single weighted confidence interval ratio

$$R_w = \frac{\sqrt{\sum_{i,i'} \rho_i (c_i^{i'})^2}}{\sqrt{\sum_{i,i'} \rho_i (T_i^{i'})^2}} \quad (2.8)$$

where  $\rho_i$  [recall (2.6)] acts as a weight.  $R_w$  is thus a “noise-to-signal” ratio, which will be “small” if the time series is adequately long, and consequently the statistical significance of the transition coefficients is high. Should  $R_w$  be close to unity, this would mean that there are too few observations and as such, a coarser grid size  $dq$  should be chosen. An approach based on the (expected) convergence of matrix  $\mathbf{T}$  with increasing time series length, is developed in [chapter 3](#). Indeed, at the margins of the “cloud of observed states” in phase space there will be several cells with very few observations, whereas cells located close to the climatological mean will be frequently visited<sup>10</sup>. A constant grid size, as chosen by the author, has the advantage of keeping the same resolution for close-to-average and for extreme events<sup>11</sup>. It also allows for easy programmability<sup>12</sup> despite great phase-space cell numbers. The comparison of correlations described in [section 2.4](#) will in any case reveal whether the master equation correctly describes the underlying phenomena.

Depending on the system, transition coefficients may depend on time. Seasonal dependence was introduced for instance by [Pasmanter and Timmermann \(2002\)](#), who estimated transition matrices for each month of the year on the base of a 640 year long ENSO model run. However, if one is working with observational records instead of numerical model runs, the amount of data available is in most cases insufficient to introduce such a time dependence. In this thesis the mean and yearly cycle will be deducted from the data on the base of a Fourier-analysis (see [Fig. 4.1](#)). Transition matrices are then estimated on the base of time series of de-seasonalised anomalies.

## 2.3 Probability density and dynamics in phase space

The dynamics in the phase space spanned by three stratospheric key climate variables is studied in this thesis in order to learn more about the relationship between the variables.

<sup>9</sup>[Chapter 3](#) deals among other things with this drawback of master equations, i.e. the need for long data sets.

<sup>10</sup>See also the discussion in [Egger \(2001, section 5\)](#).

<sup>11</sup>In the case of equipartitions the cells corresponding to extreme values occupy a large interval of the variable’s physical range ([Pasmanter and Timmermann, 2002](#)).

<sup>12</sup>Special algorithms have to be developed, for instance, in the case of equipartitions (see [Pasmanter and Timmermann, 2002, subsection 2.1](#)).

The PDF evolves in phase space according to the master equation (2.2).

In a two-dimensional case isolines of the PDF may be easily visualised so that one may follow the evolution of a “probability cloud” starting in one grid box. To extract information on the mean motion in the phase plane of two climate variables at a time, Egger (2001, 2002) determined transition velocities and diffusion coefficients in analogy to the drift and diffusion terms of Fokker-Plank equations. However, such transition velocities and diffusion coefficients may hardly be visualised and understood in a three- or more-dimensional phase space.

In the three-dimensional case the PDF may be integrated along one axis and the evolution in the phase-plane perpendicular to that axis studied. A more quantitative way to study dynamics in phase space is to compute the mean position  $\hat{\mathbf{q}} = (\hat{q}_1, \dots, \hat{q}_V)$  of the cloud:

$$\hat{q}_v = \sum_i q_v f_i(q_1, \dots, q_V) \prod_w dq_w \quad v = 1, \dots, V \quad (2.9)$$

and its standard deviation:

$$\sigma = \sqrt{\sum_i f_i(q_1, \dots, q_V) [\sum_v (q_v - \hat{q}_v)^2] \prod_w dq_w} \quad (2.10)$$

where  $V = 3$  and  $dq_w = dq$ . The evolution of the mean position describes a trajectory in phase space.

## 2.4 The Markovian assumption

In (2.1) it is assumed that the knowledge of the state of the system at time  $n$  is sufficient for knowing the state of the system at time  $n + 1$ . This very important assumption is referred to as the Markovian assumption. A test of this assumption is needed.

Mainly in the 1960s, a time in which computer performance was not comparable to today’s, elegant methods were developed for treating Markovian processes in a discrete time and space environment (e.g. Billingsley, 1961). In Spekat et al. (1983) weather regimes from a 100-year-long time series were grouped into three possible states: zonal, mixed and meridional (recall page 2). In this three-cell partition the Markovian assumption was confirmed by using a  $\chi^2$ -test. Nicolis (1990) made use of “Markov partitions<sup>13</sup>” for dividing the Z-component axis in the Lorenz convection model (Lorenz, 1963) into two cells, and checked with a generalised  $\chi^2$  test that some Markovian conditions were satisfied. Nicolis et al. (1997) suggested that exponential decay of exit time distributions is a characteristic signature of a first order Markov process. Egger (2001, 2002) partitioned two-dimensional phase planes into squares of constant grid size and tested the Markovian assumption by comparing auto- and cross-correlation functions of the variables as obtained from the master equations to those obtained from the data. Pasmanter and Timmermann (2002) used “equipartitions”, i.e. partitions where each cell contains an equal number of data points, in a two-variable phase plane and advised to compute correlations to check

<sup>13</sup>Nicolis (1990) stated that the definition of Markov partitions is purely topological and that the mere choice of these partitions does not guarantee the validity of the conditions for a process to be Markovian.



whether the Markovian assumption holds. They also suggested a finer partition or an enlargement of the variable set to eliminate possible violations. [Crommelin \(2004\)](#) also made use of equipartitions and, recalling that a Markovian system has no memory of its history, used 7-day means instead of daily data, so that consecutive data points were largely uncorrelated.

In this thesis the phase space will be partitioned into cubes of constant grid size and the Markovian assumption tested by comparing correlations as done by [Egger \(2001, 2002\)](#).

## 2.5 Correlation functions

Correlations represent a crucial quantity while working with master equations. As stated in [section 2.4](#), one may use them to test the Markovian assumption inherent in the master equation. Moreover, correlations give information about the overall performance of a master equation and may as well give hints about dynamics in phase space ([Egger, 2001, 2002](#)). We now consider three-dimensional systems,  $i_1$ ,  $i_2$ , and  $i_3$  being the grid box indices. The grid size  $dq$  is constant for all variables, so that we consider cubic grid boxes.

### 2.5.1 Estimate of correlations from time series

In this section  $q_v(t_0), \dots, q_v(t_n), \dots, q_v(t_N)$  represents a time series of variable  $v$  sampled by observing an ergodic process (see [von Storch and Zwiers, 1999](#), p. 29 and chapter 12) at  $N + 1$  consecutive times, beginning at some arbitrary time  $t_0$ ,  $dt$  being again the time interval. Correlation functions may be estimated directly from these data:

$$r_{q_k q_l}(\tau) = \frac{C_{q_k q_l}(\tau)}{\sqrt{C_{q_k}(0) C_{q_l}(0)}} \quad k, l = 1, 2, 3 \quad (2.11)$$

where  $C_{q_k q_l}(\tau)$  is the sample covariance function for time lag  $\tau$  between variables  $q_k$  and  $q_l$ :

$$C_{q_k q_l}(\tau) = \frac{1}{t_N - |\tau|} \sum_{t=t_0}^{t_N-\tau} q'_k(t) q'_l(t + \tau) \quad \forall \tau \geq 0 \quad (2.12)$$

$$C_{q_k q_l}(\tau) = \frac{1}{t_N - |\tau|} \sum_{t=t_0-\tau}^{t_N} q'_k(t) q'_l(t + \tau) \quad \forall \tau < 0$$

where  $q'_k(t)$  and  $q'_l(t + \tau)$  are the anomalies of the  $k^{\text{th}}$  and  $l^{\text{th}}$  variables respectively,  $q'_l$  is considered with a time lag  $\tau$ , and the summation runs over all data values. For significance tests, refer to [Schönwiese \(1985\)](#) and/or to [von Storch and Zwiers \(1999\)](#).

### 2.5.2 Correlations given by a master equation

In this thesis it is assumed that underlying systems are ergodic (see [von Storch and Zwiers, 1999](#), p. 29). Covariance functions as given by a master equation may be computed by considering spatial contributions. The contribution  $[c_{q_k q_l}(\tau)]_{i_1^0, i_2^0, i_3^0}$  at time lag  $\tau$  to the total

estimate of the covariance function  $C_{q_k q_l}(\tau)$  for variables  $q_k$  and  $q_l$  from a master equation in which the PDF is initially ( $t = 0$ ) concentrated in box  $(i_1^0, i_2^0, i_3^0) \equiv (\mathbf{i}^0)$  is

$$\begin{aligned} [c_{\tilde{q}_k \tilde{q}_l}(\tau)]_{i_1^0, i_2^0, i_3^0} &= \sum_{i_1, i_2, i_3} \tilde{q}_k^{\prime 0} q_l' f(\mathbf{i}^0, t=0; \mathbf{i}, t=\tau) dq^3 dq^3 \\ &= \sum_{i_1, i_2, i_3} \tilde{q}_k^{\prime 0} \tilde{q}_l' f(\mathbf{i}^0, t=0) g(\mathbf{i}^0, t=0; \mathbf{i}, t=\tau) dq^3 dq^3 \end{aligned}$$

where lags are non-negative,  $\tilde{q}_v'$  refers to the  $v^{\text{th}}$  coordinate of the centre of the particular box (recall Fig. 2.1),  $f(\mathbf{i}^0, t=0)$  takes into account the observed mean probability in the starting box and  $g$  the consequent conditional probability to be calculated from a master equation run with  $g_{\mathbf{i}^0}(t=0) = 1/dq^3$ ; the *zero* superscripts  $^0$  indicate that the corresponding quantity refers to or derives from the initial condition in  $(\mathbf{i}^0)$ . The time mean PDF in this initial box  $\bar{f}_{\mathbf{i}^0} \simeq \rho_{\mathbf{i}^0} \equiv \rho^0$  so that

$$\begin{aligned} [c_{q_k q_l}(\tau)]_{i_1^0, i_2^0, i_3^0} &\simeq \rho^0 \tilde{q}_k^{\prime 0} dq^3 \left( \sum_{i_1, i_2, i_3} g(\mathbf{i}^0, t=0; \mathbf{i}, t=\tau) \tilde{q}_l' dq^3 \right) \\ &= \rho^0 \tilde{q}_k^{\prime 0} dq^3 (\hat{q}_l^{\prime 0}(\tau)) \end{aligned} \quad (2.13)$$

where the second equality derives from (2.9),  $\hat{q}_l^{\prime 0}(\tau)$  indicating the mean  $q_l$ -coordinate of the cloud. These contributions sum up to the total estimate of the covariance function:

$$C_{q_k q_l}(\tau) = \sum_{i_1^0, i_2^0, i_3^0} [c_{q_k q_l}(\tau)]_{i_1^0, i_2^0, i_3^0}. \quad (2.14)$$

To summarise, the summation in (2.13) determines at lag  $\tau$  the mean  $q_l$ -coordinate of the cloud that had started in box  $(\mathbf{i}^0)$ . The summation in (2.14) puts together all these contributions. Covariances for negative lags may be derived according to

$$C_{q_k q_l}(\tau) = C_{q_l q_k}(-\tau).$$

Finally, correlation functions are derived according to (2.11).

# Chapter 3

## On the numerical properties of master equations

There are many factors that influence the success of a data-based master equation. Some of these depend on choices made by the user such as the *choice of the variables and their number*, and the *type and degree of phase-space partition*. Other factors are somehow given, such as the *available computer resources*. In this thesis the transition coefficients of a master equation are estimated from time series of the variables. Thus the *length and resolution of the time series* greatly influence the success of a master equation forecast. All these factors might of course act together. For instance, the amount and type of data will influence the choice of how to partition the phase space, and the available computer resources set a limit to the size of the array representing transition matrix  $\mathbf{T}$ . Moreover, while analysing data from observations, one has just one realization of the phenomenon. For these reasons these questions are addressed here on the base of the famous Lorenz convection model (Lorenz, 1963) extended with a stochastic forcing. This approach has two assets; firstly, time series of the desired characteristics may be generated easily, secondly, the model equations are known and thus the PDF forecast by a master equation may be directly compared to the distribution of an ensemble of initial conditions. Moreover, correlation functions will also reveal information about the quality of the master equation.

### 3.1 The Lorenz model extended with Gaussian white noise

The three-component Lorenz convection model (Lorenz, 1963) is extended in this section with a stochastic forcing and taken as a base for a study on the numerical properties of data-based master equations to be conducted in section 3.2. The idea of using a three-component system has at least two assets. Three-dimensional systems are of great interest because a continuous-time set of ordinary differential equations showing chaotic<sup>1</sup> behaviour must involve at least three equations (see Kaplan and Glass, 1995, p. 308). Moreover, it is the aim of this thesis to study climate variability in the stratosphere with three-dimensional master equations<sup>2</sup>. Lorenz's paper "Deterministic Nonperiodic Flow" (Lorenz, 1963) attracted the

---

<sup>1</sup>Chaotic systems offer interesting predictability studies.

<sup>2</sup>An extension to, say, a four or five-dimensional master equation is, in the author's opinion, feasible.

attention of not only meteorologists but also other branches of the natural sciences and in particular mathematicians. [Palmer \(1993\)](#) took the Lorenz model as a conceptual representation of the chaotic extra-tropical circulation and used it to investigate the roles of time averaging and ensemble forecasting in order to interpret results from weather prediction models<sup>3</sup>. The Lorenz model equations ([Lorenz, 1963](#)) with stochastic forcing

$$\begin{aligned}\dot{X} &= -\sigma X + \sigma Y + \alpha \xi \\ \dot{Y} &= -XZ + rX - Y + \alpha \xi \\ \dot{Z} &= XY - bZ + \alpha \xi\end{aligned}\tag{3.1}$$

describe a three-dimensional dynamical system. Variables  $X$ ,  $Y$ ,  $Z$ , and time are nondimensional. The stochastic forcing  $\alpha \xi$  is the product of a factor  $\alpha$  times a Gaussian white noise  $\xi$ . For  $\xi$  to be a white noise it is required that it is uncorrelated:

$$\langle \xi \rangle = 0 \quad , \quad \langle \xi(t) \xi(\tilde{t}) \rangle = \delta(t - \tilde{t})$$

where  $\delta(\tau)$  is the Dirac function. A consequence of this is that the spectrum<sup>4</sup> of a white noise has the same constant value for all frequencies. New values for  $\xi$  are continuously generated throughout the numerical integration (with a time step  $dt$ ) of the model equations:

$$\xi_j = \frac{1}{\sqrt{dt}} \gamma_j$$

([Levy, 1948](#)) where  $\gamma_j$  is a random number with Gaussian deviate<sup>5</sup>:

$$\langle \gamma_j \rangle = 0 \quad , \quad \langle \gamma_j \gamma_h \rangle = \delta_{jh} \quad .$$

Hence:

$$\langle \xi_j \rangle = 0 \quad , \quad \langle \xi_j \xi_h \rangle = \frac{1}{dt} \delta_{jh} \quad .$$

The Lorenz model is of the first order in time; therefore the Markovian assumption is satisfied<sup>6</sup> for the model<sup>7</sup>.

Let us first consider Eq.set (3.1) without stochastic forcing ( $\alpha = 0$ ). For suitable values of the parameters  $\sigma$ ,  $r$ , and  $b$ , the dynamics of the Lorenz model show deterministic chaos<sup>8</sup>. A standard choice in order to obtain chaotic dynamics is the following:  $\sigma = 10$ ,  $r = 28$ , and  $b = 8/3$ . Hereafter we will refer to Eq.set (3.1) with these parameter values<sup>9</sup>. The evolution of the state vector  $q = (q_1, q_2, q_3) = (X, Y, Z)$  describes then the famous Lorenz attractor with inhomogeneous ‘‘butterfly-wings’’ shape. [Figure 3.1](#) shows the distribution

<sup>3</sup>That was the time when the first embryos of ensemble forecasting models were appearing on the meteorological scene (e.g. [Molteni et al., 1996](#), and references therein).

<sup>4</sup>The spectrum is the Fourier transform of the auto-covariance function.

<sup>5</sup>See the Box-Muller algorithm in [Press et al. \(1999\)](#).

<sup>6</sup>Most equations describing the circulation of the atmosphere include first-order derivatives in time, and first and higher-order derivatives in space.

<sup>7</sup>This does not assure that the Markovian assumption holds for a master equation derived from a time series obtained via numerical integration of (3.1).

<sup>8</sup>Deterministic chaos is defined as aperiodic, bounded, and deterministic dynamics with sensitive dependence on initial conditions (see [Kaplan and Glass, 1995](#)).

<sup>9</sup>These values appear for instance also in [Sparrow \(1982\)](#), [Palmer \(1993\)](#), [Kaplan and Glass \(1995\)](#), [Pichler \(1997\)](#).

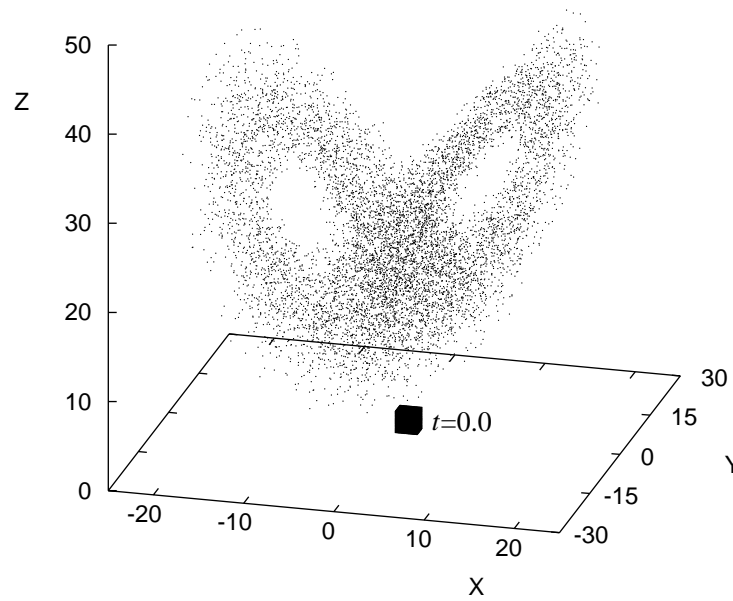


Figure 3.1: Distribution at  $t = 0$  and at  $t = 70$  (nondimensional time units) of a cloud of 8000 points; calculated from Eq. set (3.1) without white noise. Variables  $X$ ,  $Y$ , and  $Z$  are nondimensional.

of an ensemble of 8000 initial conditions at time  $t = 0$  and  $t = 70$  (nondimensional time units), computed according to Eq. set (3.1) without white noise ( $\alpha = 0$ ). At  $t = 0$  the points are located within a cube of size  $dq = 2.50$ . At  $t = 70$  the points occupy the entire butterfly-wings shaped Lorenz attractor. A trajectory outside the Lorenz attractor is rapidly caught into the attractor, whereas trajectories in the attractor diverge (e.g. Pichler, 1997). The dynamics on the attractor show sensitive dependence on initial conditions<sup>10</sup>. Hence arises most of the unpredictability of the Lorenz model. However, as shown by Palmer (1993), there are portions of the Lorenz attractor where predictability is higher than average and there is little evidence of the chaotic nature of the system's dynamics. In other cases an ensemble of points reaches the splitting part of the attractor and adjacent trajectories diverge towards the two different attractor wings. In other words, there are regions of the attractor that are relatively more sensitive to initial conditions.

For the above mentioned parameter values, the divergence of the Lorenz model without stochastic forcing is  $\nabla \cdot (\dot{X}, \dot{Y}, \dot{Z}) = -\sigma - 1 - b$ , i.e. negative and constant. This suggests that the phase space is continuously shrinking onto the Lorenz attractor, which has a fractal dimension of 2.06 (Pichler, 1997). The amplitude of the Gaussian white noise is now set to  $\alpha = 2.5$ . This forcing acts against this frictional contraction. Noise and damping are the two elements required for the interpretation of climate as a stationary stochastic system (see von Storch and Zwiers, 1999, p. 1-2). This is actually always assumed when master equations with time-constant transition coefficients are considered. The choice of this value for  $\alpha$  implies a moderate dispersion of the initial conditions without drastically altering

<sup>10</sup>Fig. 3.2 illustrates that just slightly different initial conditions lead to completely different results. This behaviour may be seen also while running the Lorenz model with no stochastic forcing (see Palmer, 1993, Fig. 2). Moreover, several numerical experiments have confirmed that the Lorenz model with white noise is sensitive not only to different computational accuracies, but also to the choice of the compiler or even compiler version. Statistics don't change indeed for long time series.

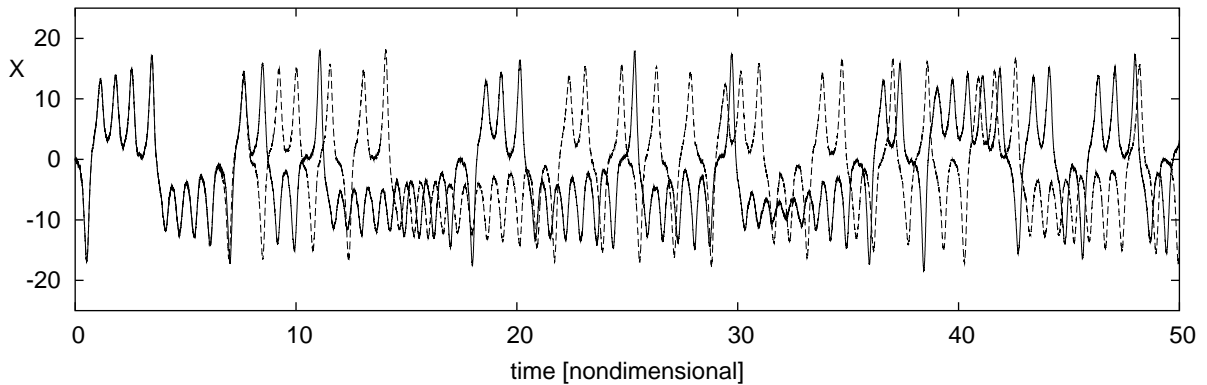


Figure 3.2: Time series of the X-component of the Lorenz model with white Gaussian noise (3.1). The initial conditions of the *dashed* time series are just slightly different from the initial conditions of the *solid* time series.

character and shape of the trajectories [see Fig. 3.6 (a, b) and Fig. 3.3].

Two time scales are associated with the model. The one describing the evolution of a trajectory about the (weakly) unstable fixed point at the centre of each attractor wing is  $t_w \approx 0.7$ , whereas the time scale related to the residence time in a wing  $t_r$  varies between 1 and 7 nondimensional time units. These time scales can be clearly seen in Fig. 3.2 (*solid* line), where the evolution of the X-component from an arbitrary initial state is shown as a function of time; the dependence from the initial conditions is evident by looking at the evolution of the *dashed* line, whose initial conditions are very slightly different from those of the solid line (their initial distance is  $10^{-3}$ ). A three-dimensional plot of the trajectory corresponding to the solid time series of Fig. 3.2 is shown in Fig. 3.3, where the circled dot marks the initial condition. Table 3.1 reports for the sake of completeness a first few lines of the time series describing this trajectory. The perturbation due to the Gaussian

Lorenz mod. alpha= 2.5 - nondimensional variables			
t	X	Y	Z
0.00000	-0.05774	-0.23136	14.13964
0.02000	0.10746	-0.45110	13.57726
0.04000	-0.20393	-0.48820	13.14308

Table 3.1: The first few lines of the time series describing the trajectory shown in Fig. 3.3.

white noise is hardly noticeable in Fig. 3.2 and Fig. 3.3. On the base of several numerical experiments it may be stated that the presence of noise in the time series does not sensibly affect the evolution of the PDF according to the corresponding master equation. The forecasts delivered by a master equation constructed with a long time series will resemble better the evolution of an ensemble of trajectories according to (3.1) with a larger noise amplitude  $\alpha$ .

Time series are generated easily by numerical integration of (3.1). This is carried out with a Runge-Kutta method of the fourth order (see Gambolati, 1994). The time step for this scheme is set to  $dt_{RK} = 0.001$ . This allows a smooth representation of trajectories. The evolution of a cloud of points is not altered by reducing the time step by a factor 10. Time series are then sampled from the integration run by just “picking” variable values every time interval  $dt$ , which represents the time interval of an atmospheric observation.

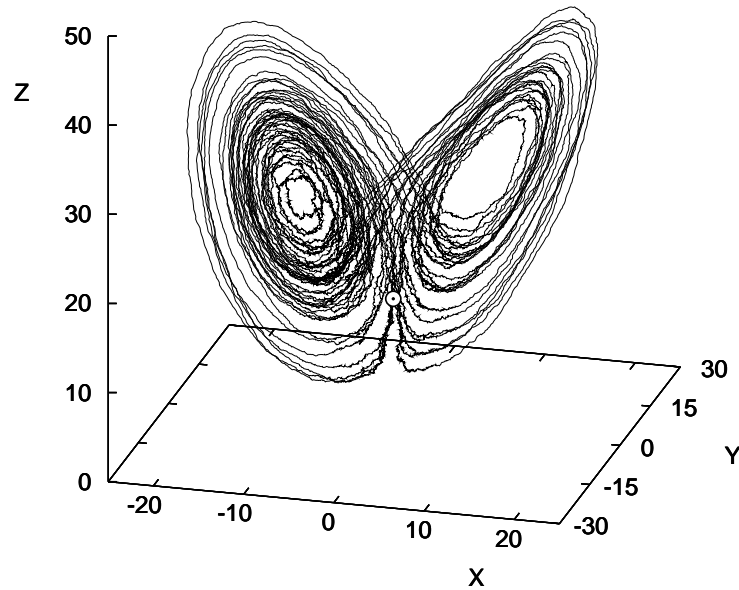


Figure 3.3: Three-dimensional plot of the trajectory of the Lorenz model with white Gaussian noise whose time series of the X-component is shown in Fig. 3.2. The initial condition is marked by the circled dot  $\odot$ .

This sampling interval afterwards becomes the time step of the master equation. The transition coefficients of the master equation are then estimated from these time series as explained in section 2.2.

The calculation of transition coefficients deserves a mention here. As the model equations (3.1) are known, the transition coefficients could be precisely calculated. One could in fact consider a cloud of initial conditions homogeneously distributed within a grid box in the discretized phase-space. By solving (3.1) one obtains the distribution of this cloud after a time  $dt$ . The transition coefficients for the initial box will be the relative frequencies of points in the grid boxes reached by the cloud. A drawback of this approach is that grid boxes intersecting the Lorenz attractor also include portions of the phase space whose evolution is not directly relevant to that of points in the attractor. Therefore the quality of the forecast would not necessarily improve. Moreover, such a procedure could not be applied to real systems where only observations are available.

## 3.2 Numerical parameters

The main numerical parameters influencing the numerics of data-based master equations are discussed in the following subsections, where the transition coefficients of several master equations will be derived from time series as described in section 2.2.

### 3.2.1 Grid size

The phase space is partitioned into cubic boxes of equal grid size  $dq$ . The choice of an adequate grid size depends first on how accurately one wants to resolve “processes” in phase space, and on the available computer resources. For the Lorenz model with white

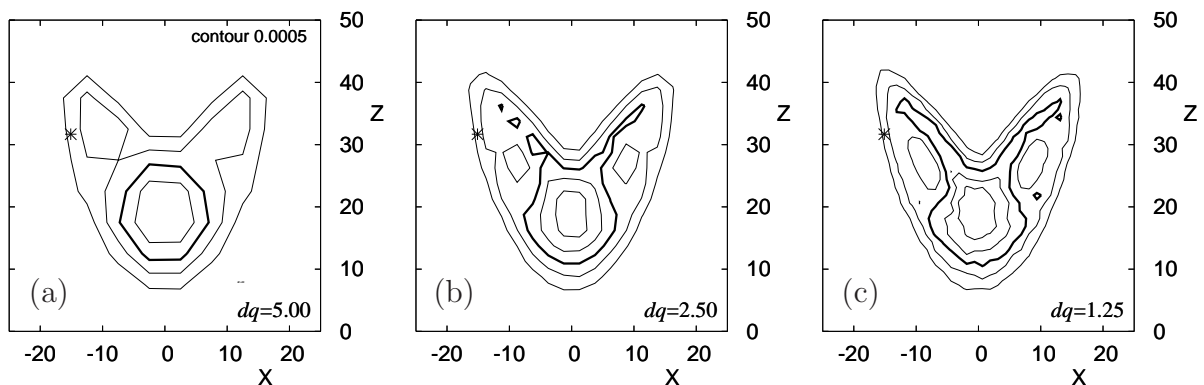


Figure 3.4: Normalised density of ensemble members at  $t = 70$  obtained with different grid sizes, integrated along the  $Y$ -axis, from (3.1); isoline contour interval is  $5 \times 10^{-4}$ , and every third isoline is bold. The star  $*$  marks the centre of a grid box with lateral size of 2.50 [nondimensional] where 125000 initial conditions were uniformly distributed. At  $t = 70$  the ensemble members occupy a phase space region which corresponds to the modified Lorenz attractor. (a) The phase space has been discretized with  $dq = 5.00$  in order to determine the normalised density, and this grid size is too coarse to resolve the “holes” in the attractor wings. In (b) and (c), where  $dq = 2.50$  and  $dq = 1.25$  respectively, these may be seen.

noise, one has to decide how accurately to resolve the “modified” attractor. With a grid size  $dq = 5.00$ , 1200 grid boxes are needed to describe the phase space region displayed in Fig. 3.4 (a) (with  $-30 \leq Y < 30$ ). Almost 250 of these boxes intersect the attractor; this is the minimum number of boxes needed to represent it. As may be seen in Fig. 3.4 (a), the choice of  $dq = 5.00$  does not resolve the “holes” in each of the butterfly wings. This is possible with  $dq = 2.50$ , as may be seen in Fig. 3.4 (b). The overall picture contains more details when  $dq = 1.25$  [see Fig. 3.4 (c)]. However, the required computer performance grows dramatically by reducing the grid size. With  $dq = 2.50$ , approximately 1000 boxes are needed to represent the attractor, and with  $dq = 1.25$  almost 4750 boxes. Moreover, the time needed for computing correlation functions increases with approximately the *eighth* power of the inverse of the grid size. For these reasons, use of an even smaller grid size, like  $dq = 1.25$ , is not advisable<sup>11</sup>. Therefore the grid size is set to  $dq = 2.50$ . Correlation functions as delivered by the master equation, which are used in this thesis to test the Markovian assumption, confirm the assets of this choice in terms of quality and of needed computer resources<sup>12</sup> (not shown).

<sup>11</sup>One may try to limit the analysis to transitions to, say,  $B$  boxes on either side of any box in order to reduce the size of the array needed to represent  $\mathbf{T}$  while working with a very fine grid size. Let us now call  $v_{max}$  the maximum value among all velocity components in phase space. It must be assured that

$$B \geq \frac{v_{max} \cdot dt}{dq} . \quad (3.2)$$

For instance, let us assume that it will suffice to consider transitions up to three boxes either side. Thus in a one-dimensional system one has to consider for each box transitions to [and from] 6 boxes (3 on the left and 3 on the right). Hence an array with up to  $7 \cdot i^{max} = (2B + 1) i^{max}$  elements is needed to represent transition matrix  $\mathbf{T}$  in the one-dimensional case. For a three-component system one must thus consider up to 343 transitions per box. Despite the strong limit imposed by (3.2), one would still have to deal with huge array sizes. Besides other possible algorithms and approaches it is therefore reccomandable to carefully select a grid size which is not “unnecessarily” fine.

<sup>12</sup>Other grid sizes were also tested through several numerical integrations. The holes in the centre of each attractor wing are clearly visible with  $dq = 2.50$ . A grid size  $dq = 2.00$ , for instance, does not



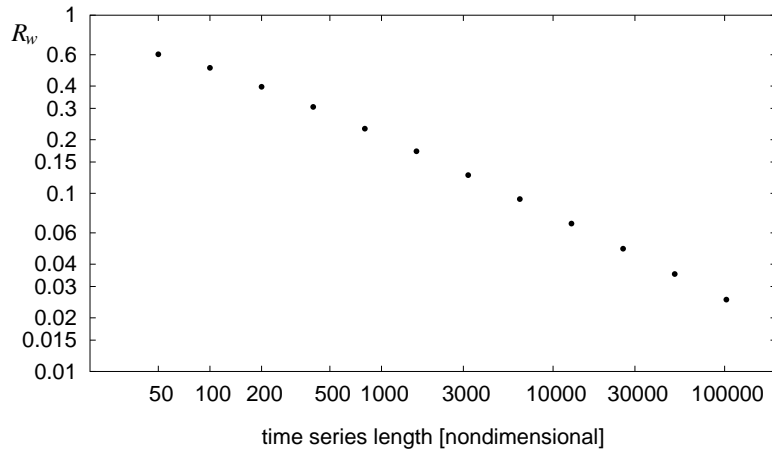


Figure 3.5: The noise-to-signal ratio  $R_w$  defined in (2.8) is shown here as a function of the time series length in nondimensional time units; time resolution  $dt = 0.020$ ; grid size  $dq = 2.50$ . The time series are generated by extending the numerical integration of (3.1) which led to the trajectory shown in Fig. 3.3.

### 3.2.2 Time series length

Following the discussion in section 2.2, transition coefficients have to be estimated from the time series. Values of the noise-to-signal ratio  $R_w$  [defined in (2.8)] are shown in Fig. 3.5 as a function of the length of the time series for a time resolution  $dt = 0.020$  and a grid size  $dq = 2.50$ . The choice of  $dt = 0.020$  is arbitrary; one wants to “smoothly” describe the evolution of the trajectory around the centre of each butterfly wing. In the following figures, the time series length will be indicated using a capital Greek Delta:  $\Delta t$ . The values of  $R_w$  in Fig. 3.5 are quite high for short time series. For  $\Delta t < 200$ , for instance,  $R_w > 0.5$ , which means that the confidence intervals of the estimated transition coefficients are of the same magnitude as the transition coefficients themselves, and consequently their statistical significance is quite low. This is no more the case for time series lengths  $\Delta t > 3000$ , where  $R_w$  is of order 0.1 and below.

In order to understand better the role of the time series length, a case study is carried out. Figure 3.6 (a) shows the distribution at  $t = 0$ ,  $t = 0.2$ ,  $t = 0.4$ , and  $t = 0.6$  of a cloud of 8000 initial conditions. Their distribution at  $t = 0.8$  is shown in Fig. 3.6 (b). This cloud of states is initially located in a grid box of the phase space discretized with  $dq = 2.50$ . This grid box includes part of the attractor. The cloud evolves, slightly “touching” the splitting region of the attractor. Some forecasts of poor quality should disperse quickly enough to occupy this region causing loss in prediction skill. Forecasts from such an initial condition may therefore be considered as a quite difficult task for a master equation. Forecasts are considered at time  $t = 0.8$ . This time is close to the predictability limit (see Palmer, 1993). Figure 3.6 (c) shows the reference density of ensemble members<sup>13</sup>. The PDFs forecast by master equations derived from time series of different lengths are shown in Fig. 3.7. With time series starting from the chosen initial coordinates (see Fig. 3.3) having a length shorter than 50 time units no forecast is possible, since no observation falls into the starting grid

---

bring great advantages in terms of correlation functions. It does, however, increase the needed computer resources.

<sup>13</sup>In order to directly compare with the PDF, the density of ensemble members is normalised just like the PDF and the observed state density  $\rho$  (see section 2.1).

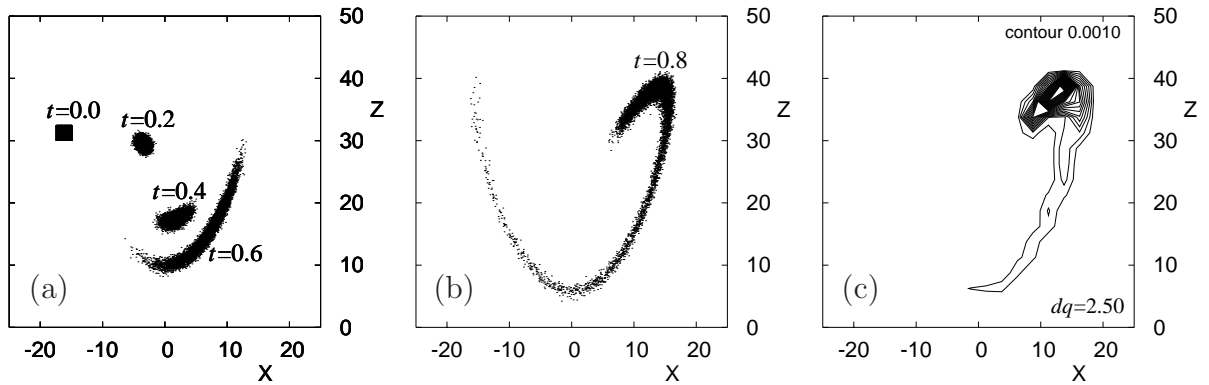


Figure 3.6: (a) The positions at  $t = 0$ ,  $t = 0.2$ ,  $t = 0.4$  and  $t = 0.6$  of a cloud of 8000 points according to (3.1); the positions at  $t = 0.8$  are shown in (b); of a cloud of 8000 points according to (3.1). (c) Isolines (integrated along the Y-axis) of the normalised density of ensemble members (points) at  $t = 0.8$ , to be compared with the master-equation forecasts shown in Fig. 3.7.

box. The forecasts shown in Fig. 3.7(a-c) appear to be of poor quality. The forecasts shown in Fig. 3.7(g-l) are of better quality and are hardly distinguishable from another; in these forecasts the PDF is higher on the right wing of the attractor, just as is the case for the density of ensemble members [Fig. 3.6(c)] obtained directly from the model equations (3.1). The PDF on the left wing is a result of the smearing, which leads to nonzero PDF for time  $t = 0.4$  [see Fig. 3.6(a)] on a wider region, a part of which evolves to the left attractor wing. The high variance between the forecasts in Fig. 3.7(a-f) is not surprising. Figure 3.8 shows the observed state density  $\rho$  seen in the time series used to obtain the forecasts shown in Fig. 3.7(a-f). The observations are far too dense on the left side of the attractor even for a time series length of 400 time units, as shown in Fig. 3.8(d); the corresponding forecast is shown in Fig. 3.7(d). For a time series length  $\Delta t = 800$  the observations are just slightly too dense on the left wing, as is shown in Fig. 3.8(e), and the forecast delivered by the corresponding master equation in Fig. 3.7(e) is definitely better than the latter.

In order to give quantitative support to the statements made in the last paragraph, a measure for the forecast skill  $S$  is developed. This is a weighted measure of the mean square deviation between the reference and the PDF predicted by the master equation. The reference is the observed density of ensemble members as given by numerical integration of the model equations (3.1). Moreover, the density of ensemble members is used as a weight (see Gardiner, 1983, p. 40). Then the measure is normed with the outcome in the case that the PDF is zero in all grid boxes where the density of ensemble members is non-zero, i.e. in the case where the two patterns are fully apart. The skill score  $S$  is therefore negative oriented, with perfect forecasts scoring zero. In integral notation

$$S = \frac{\int (f - \rho)^2 \rho dq^3}{\int (\rho)^2 \rho dq^3}$$

where the integration is done on the whole phase space. In the discretized phase space at time step  $t_n$ :

$$S(\mathbf{f}, \boldsymbol{\rho}) = \frac{\sum_{\mathbf{i}} [f_{\mathbf{i}}(t_n) - \rho_{\mathbf{i}}(t_n)]^2 \rho_{\mathbf{i}}(t_n) dq^3}{\sum_{\mathbf{i}} (\rho_{\mathbf{i}}(t_n))^2 \rho_{\mathbf{i}}(t_n) dq^3} \quad (3.3)$$

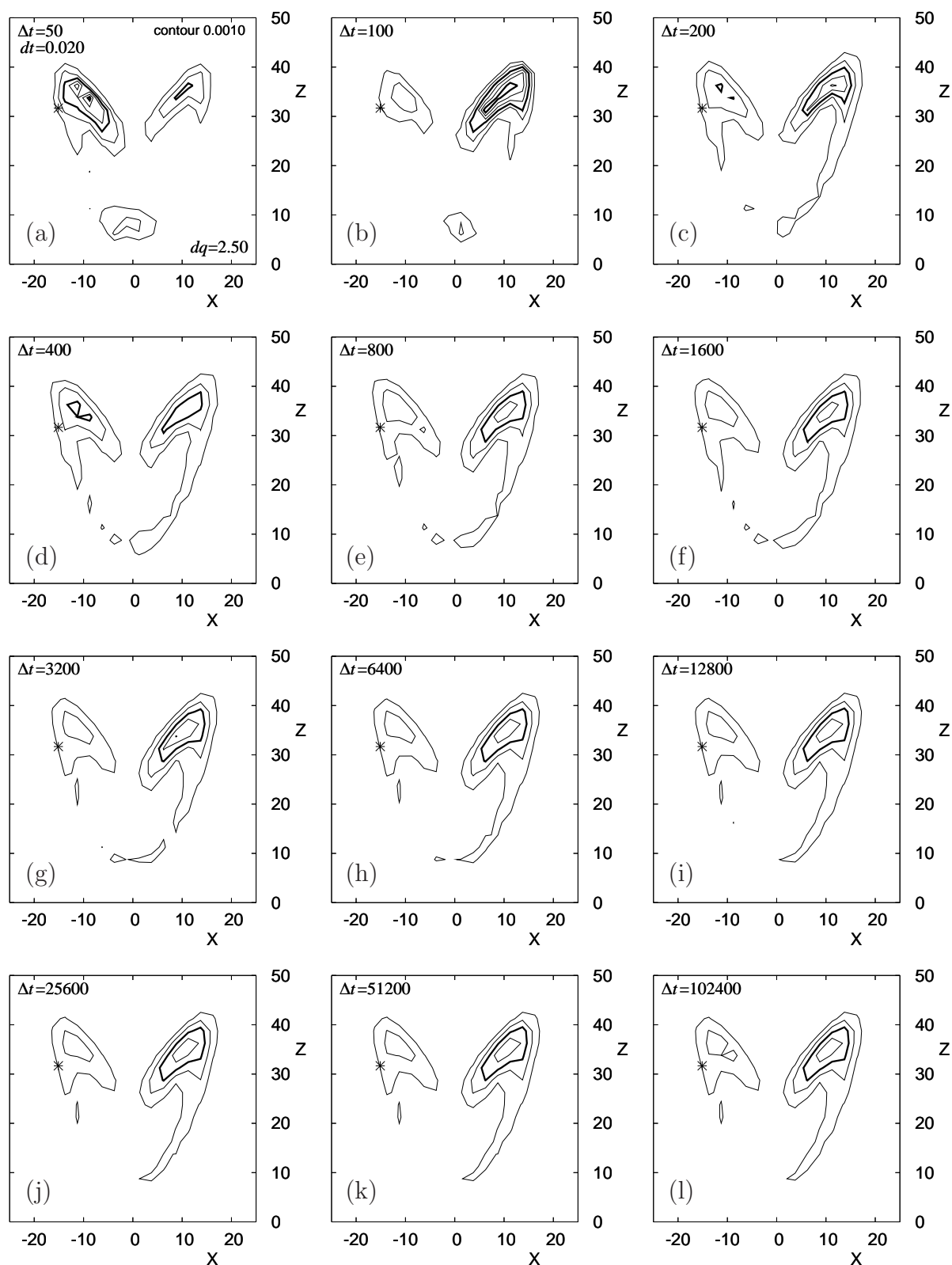


Figure 3.7: PDF forecasts for time  $t = 0.8$  as delivered by master equations derived from time series of various lengths; time resolution  $dt = 0.020$ ; grid size  $dq = 2.50$ . The star \* marks the sharp initial condition [see Fig. 3.6 (a)]. The PDF has been integrated along the Y-axis. The isoline contour interval is  $1 \times 10^{-3}$ , and every third isoline is bold. The reference density of ensemble members is shown in Fig. 3.6 (c).

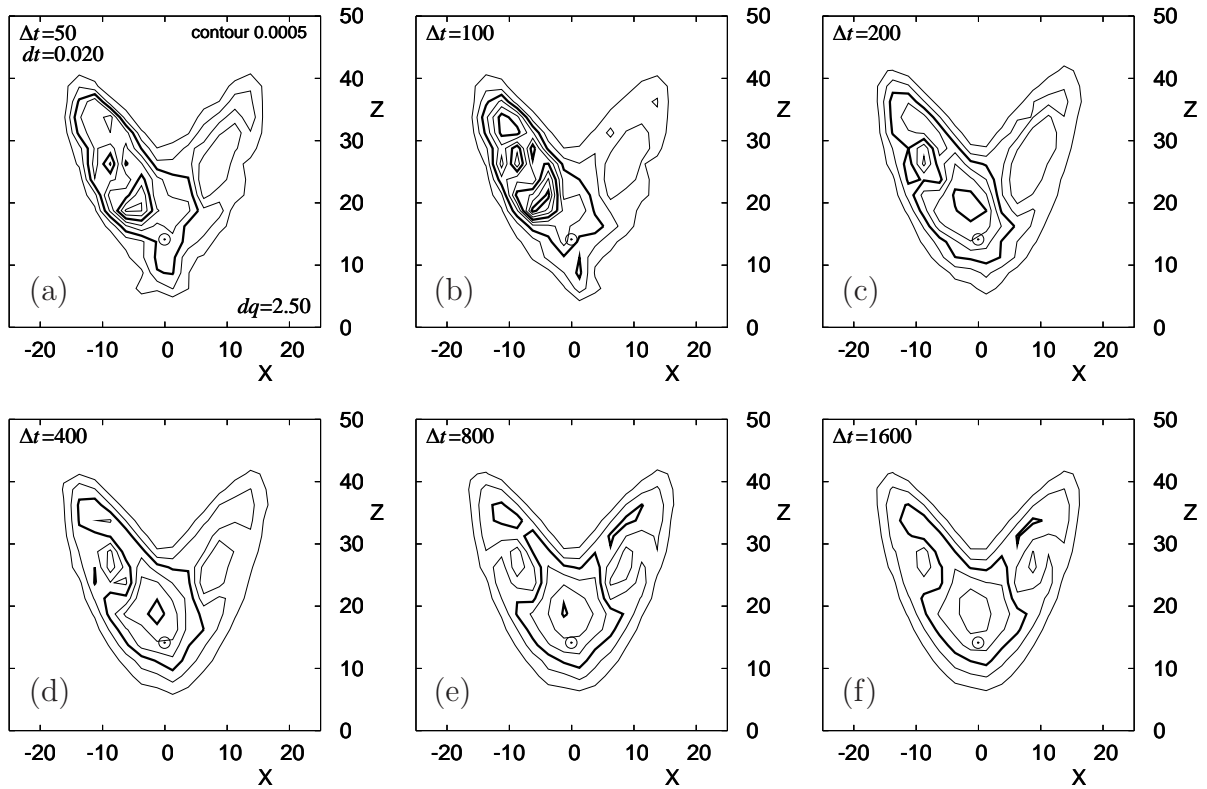


Figure 3.8: Observed state density  $\rho$  (integrated along the Y-axis) for time series of different lengths; time resolution  $dt = 0.020$ ; grid size  $dq = 2.5$ . The isoline contour interval is  $5 \times 10^{-4}$ , and every third isoline is bold. The beginning of the time series is marked by a circled dot  $\odot$ .

where the summation runs over all grid boxes. Values of  $S$  for the forecasts for  $t = 0.8$  shown in Fig. 3.7 are shown in Fig. 3.9 as a function of the time series length. The reference density of ensemble members is shown in Fig. 3.6 (c). This skill measure is again  $S = 0$  when density of ensemble members and PDF values are equal in every grid box of the phase space; when the two patterns are clearly separated a unit score  $S = 1$  is expected. A value of the skill measure greater than unity,  $S > 1$ , may occur if the PDF largely overestimates the point density in some grid boxes. This could be the case when very short time series are used. The PDF would in fact move in “chunks” along the few trajectories intersecting the initial grid boxes<sup>14</sup>. This is the case, for instance, for Fig. 3.7 (b), which fortuitously scores the best skill. In Fig. 3.7 (a), however, the PDF is erroneously led to the left side of the attractor. Moreover, the skill hardly improves for time series lengths  $\Delta t \gg 1 \times 10^3$ . Thus  $3 \times 10^3$  may be considered as an optimal length for a time series with resolution  $dt = 0.020$  and for this phase space partition<sup>15</sup> ( $dq = 2.50$ ). The ratio  $R_w$  for these time series is smaller than 0.2, which could cautiously be taken here as a maximum “noise-to-signal” ratio to work with. If time series are obtained from observations, no skill may be computed as just one realisation occurs in reality.

An indefinitely long time series should not be expected to lead to an ever increasing

<sup>14</sup>With small values of  $\alpha$ , this could lead to chunks ending up in regions where the point density is very high. This could also result in a better skill against what the visual impression would suggest.

<sup>15</sup>Good results were obtained with time series lengths  $\Delta t \geq 3 \times 10^3$  even while using smaller grid sizes (not shown).

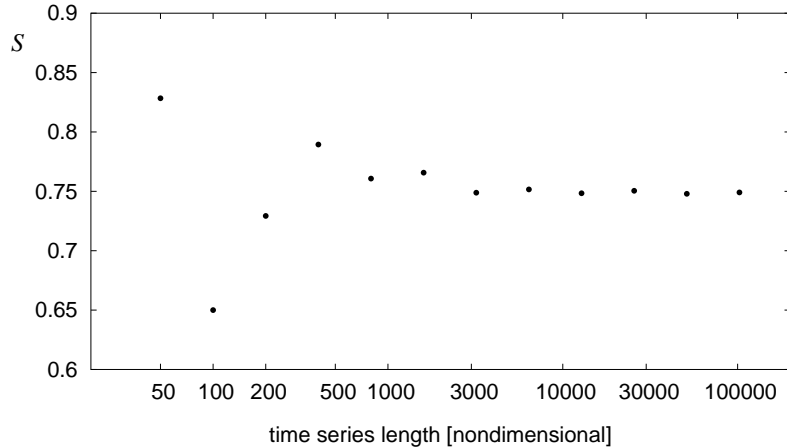


Figure 3.9: Values of the skill  $S$  [defined in (3.3)] for the master equation forecasts shown in Fig. 3.7 are shown as a function of the time series length; time resolution  $dt = 0.020$ ; grid size  $dq = 2.50$ . The reference density of ensemble members is shown in Fig. 3.6 (c).

forecast precision. One will at most achieve a correct estimate for the transition coefficients relative to the used time resolution and the chosen grid size. To further investigate this point, a different approach is introduced. A convergence study is conducted to examine how the estimate of matrix  $\mathbf{T}$  improves with growing time series length. One could think of matrix  $\mathbf{T}$  as of a high-dimensional vector, whose estimate converges to some limit vector as the time series length goes to infinity. A weighted convergence coefficient  $C_w$  is introduced. This is the normalised vectorial distance between two matrices  $\mathbf{T}(l)$  and  $\mathbf{T}(s)$  estimated on the base of the shorter and the longer of two time series respectively. The absolute value of the one deriving from the longer time series is used as the norm. Further, the elements of the matrices are weighted with the observed state density  $\rho_{\mathbf{i}}(l)$  for the longer time series in the grid box where the transition originates:

$$C_w(l, s) = \frac{\sqrt{\sum_{\mathbf{i}, \mathbf{i}'} \rho_{\mathbf{i}}(l) \left[ \left( T_{\mathbf{i}}^{\mathbf{i}'}(l) - T_{\mathbf{i}}^{\mathbf{i}'}(s) \right) \right]^2}}{\sqrt{\sum_{\mathbf{i}, \mathbf{i}'} \rho_{\mathbf{i}}(l) \left[ T_{\mathbf{i}}^{\mathbf{i}'}(l) \right]^2}}. \quad (3.4)$$

The weight  $\rho_{\mathbf{i}}(l)$  reduces the effect of “outliers” occurring seldom, i.e. only in very long time series<sup>16</sup>. Figure 3.10 shows the weighted convergence coefficient  $C_w$  as a function of the longer time series length. Here the longer time series is always twice as long as the shorter. As expected, matrix  $\mathbf{T}$  converges. The normalised weighted difference between transition matrices tends to zero for increasing time series length. The results shown in Fig. 3.10 confirm that time series of arbitrary length would not produce a noticeable gain in forecast reliability beyond what can be considered as an optimal length. As an effect of using a finer grid size  $dq = 1.25$ , there is a greater margin of improvement. The forecast precision may surprisingly profit from using longer time series also in the case of a coarser time resolution (not shown).

<sup>16</sup>These outliers occur at the very borders of the attractor and influence portions of  $\mathbf{T}$  which are not relevant for the evolution of the PDF. Calculations carried out without weighting do not lead to convergence.

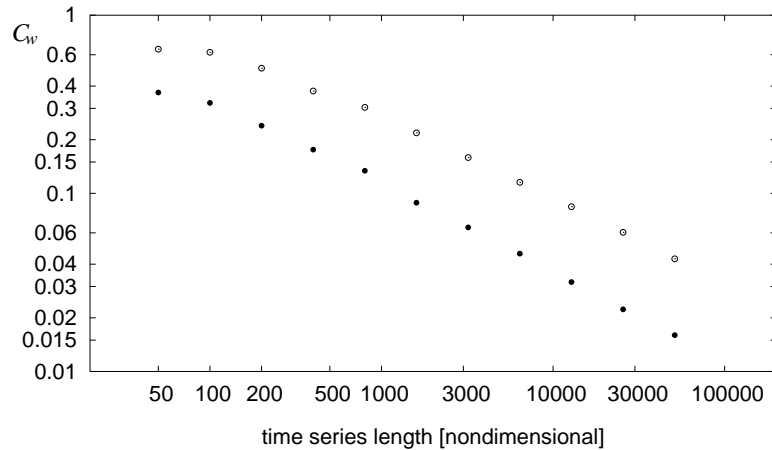


Figure 3.10: The convergence coefficient  $C_w$  is shown as a function of the length of the longer time series, which is twice as long as the shorter one; time resolution  $dt = 0.020$ . Points  $\bullet$  mark the results obtained with a grid size  $dq = 2.50$ , circles  $\circ$  correspond to  $dq = 1.25$ .

### 3.2.3 Time resolution

Time resolution strongly influences the predictions delivered by a master equation. On one hand a fine time resolution is desirable in order to achieve good estimates of the transition coefficients (Egger, 2001). Any statistical significance analysis will profit from a voluminous data set, as is the case when time resolution is fine. On the other hand the time scales of the investigated phenomena are usually known and the time resolution should not be too fine. For example, using a time step of 30 seconds while modelling inter-annual climate variability would unnecessarily increase the computing time. The time resolution sets a limit to the highest systems' frequency  $\nu_h$  that one will be able to reproduce with a master equation. For the Nyquist-Shannon sampling theorem  $\nu_h$  is half of  $1/dt$ . In order to smoothly reproduce an oscillation,  $dt$  should be some 3 to 7 times smaller than the period of the shortest oscillation to be resolved.

Figure 3.11 (b) shows for reference the forecast obtained with a time series length  $\Delta t = 3200$  and time resolution  $dt = 0.020$  [same as Fig. 3.7 (g)]. Fig. 3.11 (a) shows the forecast obtained with a master equation when the time resolution is increased by a factor 5,  $dt = 0.004$ . In Fig. 3.11 (c) the time resolution is coarser,  $dt = 0.100$ ; this time resolution is approximately 7 times smaller than the time scale describing the evolution of a trajectory around the centre of each butterfly wing ( $t_w \approx 0.7$ ). All three master equations have been derived from a time series with length  $\Delta t = 3200$ , which corresponds to the optimal length discussed in subsection 3.2.2. The best forecast of the three is Fig. 3.11 (c), where the PDF on the right wing of the attractor, though not quite as high as the density of ensemble members in Fig. 3.6 (c), exceeds a value of 0.0100. Surprisingly, the best forecast is the one obtained with the *coarser* time resolution  $dt = 0.100$ ! The skill scores  $S$  computed for the forecasts in Fig. 3.11 [reference density of ensemble members in Fig. 3.6 (c)] confirm this. For the finest time resolution the skill is the worst at 0.83 [Fig. 3.11 (a)], and improves to 0.75 with a coarser time resolution [Fig. 3.11 (b)]; the skill is best at 0.5 for the coarsest time resolution [Fig. 3.11 (c)].

This result deserves a deeper study. The two drawings in Fig. 3.12 represent an admittedly simple example which may help understand the surprisingly rapid smearing of

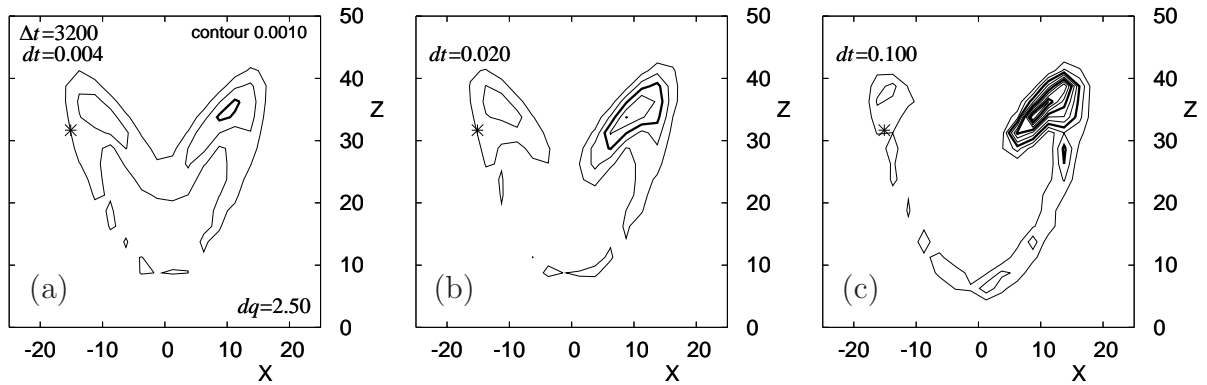


Figure 3.11: PDF forecasts for time  $t = 0.8$  delivered by master equations derived from time series of various time resolutions; time series length  $\Delta t = 3200$ ; grid size  $dq = 2.50$ . The star  $*$  marks the sharp initial condition [see Fig. 3.6 (a)]. The PDF has been integrated along the Y-axis. (a)  $dt = 0.004$ ; (b)  $dt = 0.020$  [same conditions as in Fig. 3.7 (g)]; (c)  $dt = 0.100$ . The isoline contour interval is  $1 \times 10^{-3}$ , and every third isoline is bold. The reference density of ensemble members is shown in Fig. 3.6 (c).

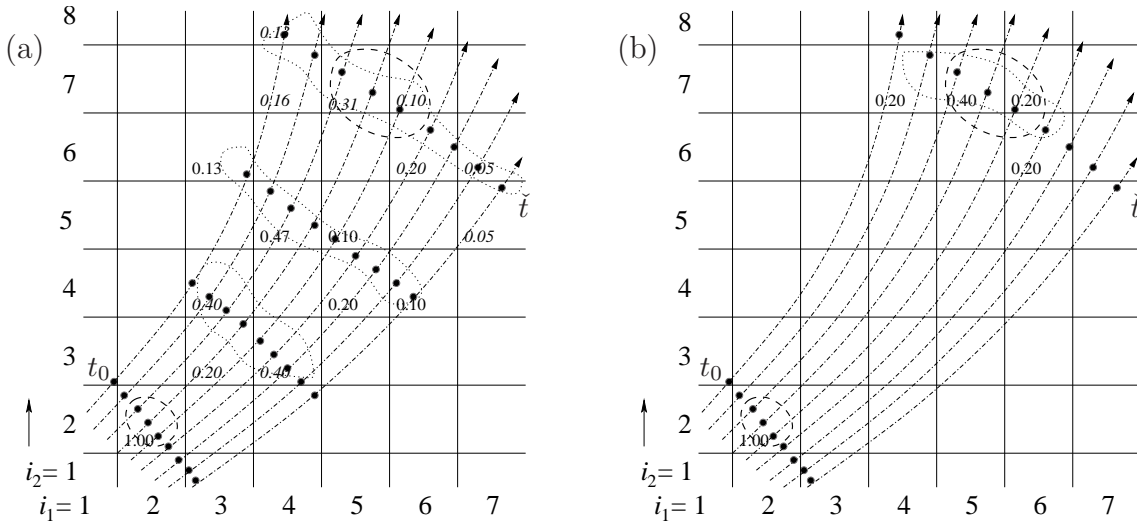


Figure 3.12: Influence of time resolution  $dt$  on a master equation. (a)  $dt = (\tilde{t} - t_0)/3$ , whereas in (b)  $dt = (\tilde{t} - t_0)$ . The PDF smears faster with finer  $dt$  as shown in (a). Refer to the text for further details.

the PDF observed in connection with a finer time resolution. The phase-plane in Fig. 3.12 is partitioned into squares of unit grid size ( $dq = 1$ ). The dash-dotted lines indicate trajectories of the analysed system, the arrows show the direction of motion. Bold points indicate recorded observations. The dashed closed curves describe the domain of a cloud of states at two different times  $t_0$  (initial time) and  $\tilde{t}$  (a later time). The PDF is set equal to  $1/dq = 1$  in box (2, 2) at the beginning ( $t = t_0$ ); PDF values are written in all boxes where the PDF is non-zero (italics are used where necessary to distinguish among successive time steps). In Fig. 3.12 (a)  $dt = (\tilde{t} - t_0)/3$ , whereas the time resolution in Fig. 3.12 (b) is coarser:  $dt = (\tilde{t} - t_0)$ . The domain where the PDF as predicted by the corresponding master equation is non-zero is described by the dotted lines. The PDF smears definitely faster in the case with finer  $dt$  as shown in Fig. 3.12 (a). This surprising result is probably the effect of discretisation, divergence of trajectories, and stronger diffusion in the case of

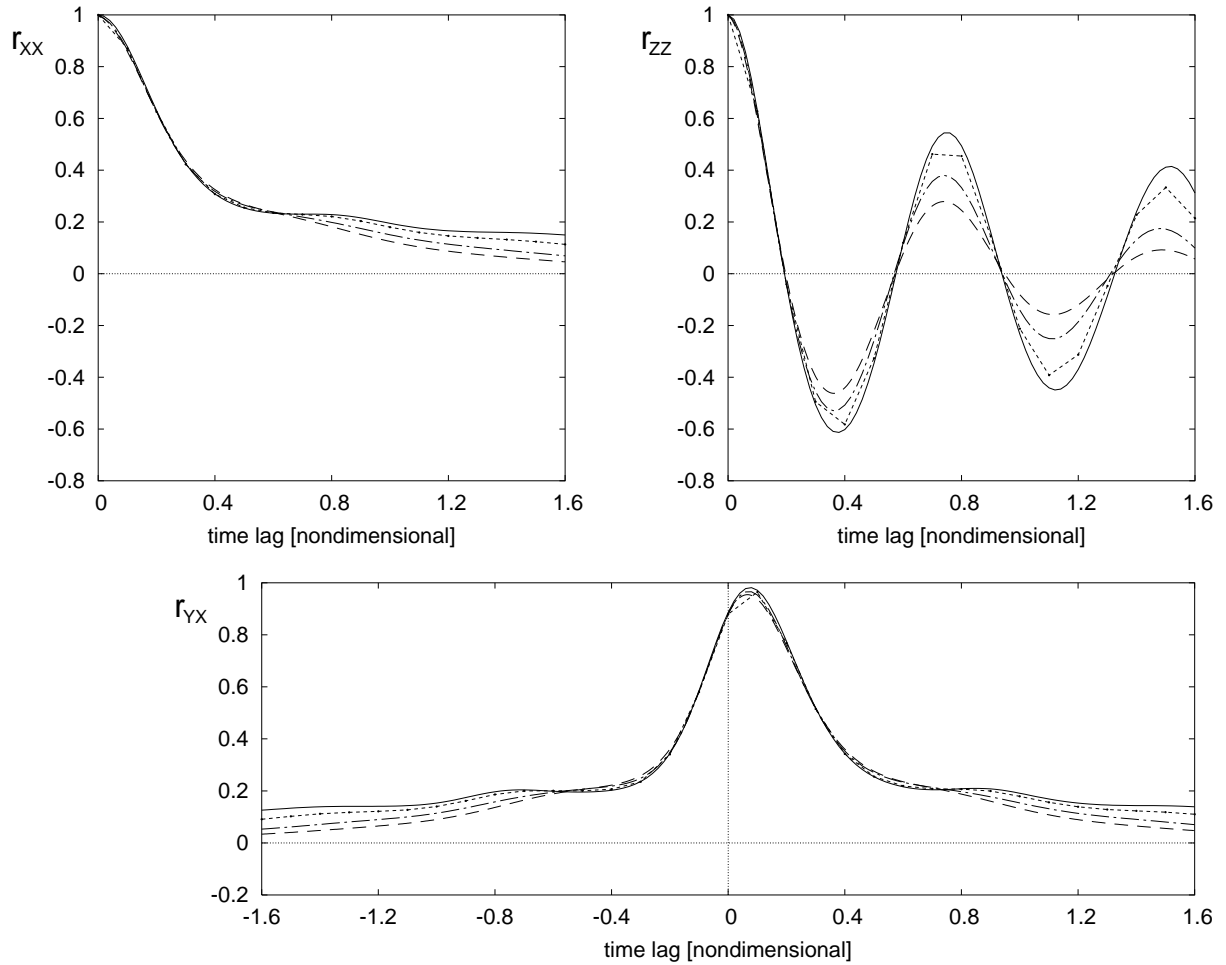


Figure 3.13: Correlation functions  $r_{xx}$ ,  $r_{zz}$ , and  $r_{yx}$  as observed (*solid*), and as reproduced by master equations derived from time series of various time resolutions, *dotted* for  $dt = 0.100$ , *dash-dotted* for  $dt = 0.020$ , and *dashed* for  $dt = 0.004$ ; time series length:  $\Delta t = 3200$ ; phase space grid size:  $dq = 2.50$ .

finer time resolutions.

Fig. 3.13 shows some correlation functions<sup>17</sup> for the master equations seen in Fig. 3.11. Correlation functions are derived from the whole phase space. Thus they allow statements which do not depend on the particular initial condition. The master equations approximate the decay of the correlation functions extremely well. The master equation based on the time series with the coarsest time resolution does the best job in reproducing the correlations observed in the data. Moreover, these results confirm the fulfilment of the Markovian approximation. It seems that a coarser time step increases the memory of the transition coefficients.

At this point it is indeed interesting to find out whether there is a time resolution, beyond which forecasts start becoming worse. Several numerical integrations have shown that for time resolutions  $dt \gg 0.100$  the forecast just slightly improves. Therefore, in order to avoid aliasing while plotting correlation functions, it is recommendable to use a time resolution  $dt < t_w/2$ .

<sup>17</sup>Correlation functions were not introduced in the discussion in subsection 3.2.2 because they highly depend on the length of the observation, which is kept constant here.



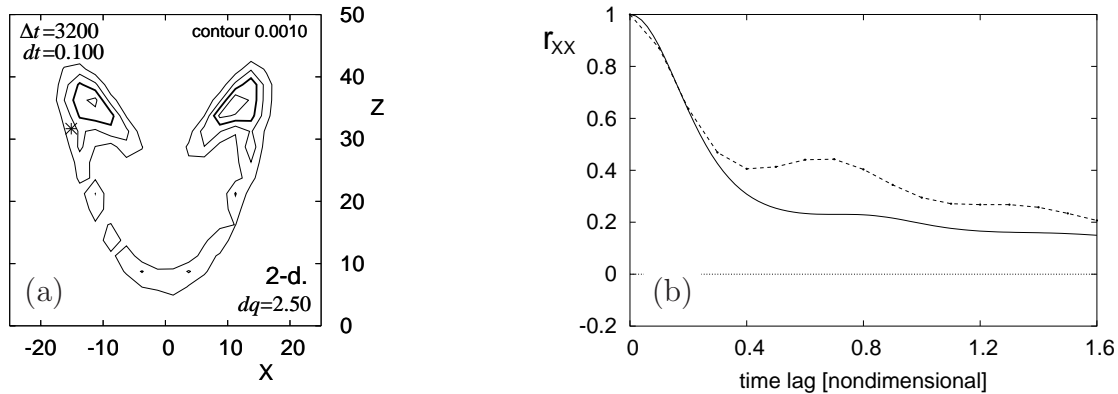


Figure 3.14: (a) PDF forecast for time  $t = 0.8$  delivered by a master equation derived from a time series of the  $X$  and  $Z$ -components of the Lorenz model with stochastic forcing; time series length  $\Delta t = 3200$ ; time resolution  $dt = 0.100$  ; grid size  $dq = 2.50$ . The star  $*$  marks the sharp initial condition [see Fig. 3.6 (a)]. The isoline contour interval is  $1 \times 10^{-3}$ , and every third isoline is bold. The reference point density is shown in Fig. 3.6 (c). The corresponding forecast, obtained with a three-dimensional master equation is shown in Fig. 3.11 (c). (b) The auto-correlation function of the first component  $r_{XX}$  as observed (*solid*) and as reproduced by the master equation (*dotted*).

### 3.2.4 Dimension of the master equation

The choice of the variables and their number also plays a crucial role. In reality it is quite hard to make use of a variable set which includes all possible variables that influence a system.

Let us now imagine a curious observer of the evolution of a trajectory on the modified Lorenz attractor. He may not immediately suspect that he is dealing with a three-component system as he is in front of a screen projecting the trajectory on the  $(X, Z)$  plane. He would just pick  $X$  and  $Z$ -component values while writing down a time series. The trajectory, however, is generated from the complete equation set (3.1). The forecast delivered by a master equation in the phase plane of the  $X$  and  $Z$ -components of the Lorenz model is shown in Fig. 3.14 (a). In order to compare the two-dimensional forecast with the best three-dimensional one [see Fig. 3.11 (c)]. The auto-correlation function  $r_{XX}$  of the  $X$ -component is shown in Fig. 3.14 (b). Our curious observer will note from Fig. 3.14 (b) that this master equation derived in the phase plane of these two variables may not capture the behaviour of the Lorenz model beyond a time scale of  $t = 0.3$ . He would conclude (correctly) that the Markovian assumption is hardly fulfilled. This could be a hint to introduce another variable<sup>18</sup>.

The lesson of Fig. 3.14 is that it is desirable to work with all three components. While deriving master equations for atmospheric data sets, however, one must keep in mind that the atmosphere is a very complex system. It is practically impossible to consider the complete variable set determining its evolution. However, while two-dimensional dynamics in a phase plane cannot represent chaotic behaviour (Kaplan and Glass, 1995), this may

<sup>18</sup>In Kaplan and Glass (1995) is shown how it is possible to reconstruct the geometry of a chaotic system from a time series even if only one of the variables is measured. This is done with a technique called time-lag embedding.

be possible in a three-dimensional phase space. Moreover, any extension of the analysis to some further variable may help in reducing violations of the Markovian approximation (Pasmanter and Timmermann, 2002).

# Chapter 4

## Master equations for stratospheric time series

The investigation of stratospheric climate variability is an interesting implementation field for the three-dimensional data-based master equation. In this chapter we present time series of stratospheric climate indices derived from the European reanalyses ERA-40 (Uppala *et al.*, 2005) and from a time series of the solar radiation flux at a wavelength of 10.7  $\mu\text{m}$ . The results from chapter 3 will be used here. Three-dimensional master equations will be derived from such time series in order to learn more about the relationship between the variables. Thus it is crucial that the master equation faithfully reproduces the behaviour of the observed system. The main modes of oscillation in the stratosphere and their relevance for the troposphere have been discussed in section 1.2. In this chapter we first investigate the roles of the QBO and the 11-year SC on the stratospheric temperature in the north polar region, and then the stratosphere-troposphere coupling in terms of the AO.

### 4.1 Time series of stratospheric climate indices

The ECMWF reanalyses ERA-40 (Uppala *et al.*, 2005) produced a comprehensive set of global analyses describing the state of the atmosphere, and land and ocean-wave conditions from 1957-09-01 to 2002-08-31. The atmospheric part of ERA-40 was obtained by driving the atmospheric model of the ECMWF with the most complete observational records available for the covered time period. The atmospheric model used for ERA-40 has 60 levels in the vertical located between the surface and 0.1 hPa (approx. 65 km), a T159 spherical-harmonic representation for basic dynamical fields, and a reduced Gaussian grid with approximately uniform 125 km spacing for surface and other grid-point fields. The ERA-40 dataset (available at [www.ecmwf.int/research/era](http://www.ecmwf.int/research/era)) includes, among other things, observations, 6-hourly analysis and forecast fields from the atmospheric model<sup>1</sup>, and monthly-means. The reanalyses have been examined at standard pressure levels and on a regular  $7.5^\circ \times 7.5^\circ$  latitude/longitude grid<sup>2</sup>. This resolution is sufficient for the investigated

---

<sup>1</sup>Analysis and forecast fields are available at the full model resolution, at standard pressure levels, and on isentropic and  $PV = \pm 2$  surfaces.

<sup>2</sup>Section 4.2 discusses some tests carried out with time series of monthly daily means, which were examined at standard pressure levels and on a regular  $2.5^\circ \times 2.5^\circ$  latitude/longitude grid.

large-scale phenomena, and keeps the global data set at a reasonable volume.

A time series of the observed solar flux at 10.7 cm is introduced in order to investigate the role of the 11-year SC on climate variability in the stratosphere<sup>3</sup>. The radio wave irradiated from the sun at a wavelength of 10.7 cm (2800 MHz) has been measured at various observatories since 1947 and is known to be a good measure of the extreme ultraviolet solar radiation (EUV). The EUV is supposed to influence the climate in the stratosphere, but may not be directly measured from the ground (Labitzke, 1998). The time series used in this thesis has been produced by the Dominion Radio Astrophysical Observatory of the Herberg Institute of Astrophysics, Canada (available at [www.drao.nrc.ca/icarus/www/sol\\_home.shtml](http://www.drao.nrc.ca/icarus/www/sol_home.shtml)).

Master equations are derived from time series of three variables at any one time, all of which must of course have a consistent time resolution. As only daily measurements of the solar flux are available<sup>4</sup>, daily means have been derived from the 6-hourly reanalyses<sup>5</sup> ERA-40. Altogether, there are 16436 data points.

As discussed in section 1.1, the number of variables to be included in a master equation must be small. In this chapter we consider therefore time series of a few meaningful climate indices, which are the variables of the master equations. These indices are defined on the base of spatial averaging of meteorological variables and of the difference of their values between selected regions. To investigate the role of the phase of the QBO, an index  $Q_I$  is defined in the following way: the zonally averaged zonal wind at 30 hPa above the equator is normalised to the standard deviation of its time series, and the seasonal cycle is deducted. A similar index is derived from the zonally averaged zonal wind at 20 hPa above the equator: after normalisation and de-seasonalisation we obtain the time series of  $U_{e20}$ . These levels are close to the height where the QBO signal is the greatest, i.e. about 25 km (Baldwin et al., 2001, p. 218). The solar activity index  $S_{10.7}$  is defined as the normalised and de-seasonalised solar radio flux at a wavelength of 10.7 cm. The index for the temperature anomaly in the arctic stratosphere  $T'$  is defined as the normalised and de-seasonalised temperature anomaly in 10 hPa, averaged for latitudes  $\phi \geq 75^\circ\text{N}$ .  $T'$  is chosen as “central” stratospheric variable in the first master equations for the stratosphere. This choice is not new (e.g. Labitzke, 1987) and may be easily explained. A temperature gradient between the north polar stratosphere and the middle latitude stratosphere induces, through thermal wind balance, a vertically sheared zonal vortex. As temperature variations in the mid-latitude stratosphere are relatively weak, these considerations imply that colder [warmer] temperatures  $T'$  suggest westerly [easterly] anomalies in the stratosphere at higher latitudes. The correlation coefficient for  $T'$  and the zonally averaged zonal wind at  $55^\circ\text{N}$  at 10 hPa is  $r = -0.6$ . This brings us to the considerations made in section 1.2 regarding the Arctic Oscillation (AO) and its importance for the climate over Europe. AO indices are defined at 10 hPa ( $A_{10}$ , middle stratosphere), 100 hPa ( $A_{100}$ , lowermost stratosphere), and 850 hPa ( $A_{850}$ , lower troposphere) as the normalised and de-seasonalised zonally averaged

<sup>3</sup>The objectively measured radio-wave at 10.7 cm is highly and positively correlated with the 11-year sun-spot cycle (Labitzke, 2001).

<sup>4</sup>The solar flux is usually measured at noon.

<sup>5</sup>Actually, it would be more correct to call these data “ERA-40 (daily) daily means”, as is usual, for instance, to use expressions like “ERA-40 monthly daily means” to distinguish between 6-hourly reanalyses and daily means. Here and in the rest of this thesis, however, the expression “ERA-40 daily means” will be used to refer to the averaging of values at 00:00 UTC, 6:00 UTC, 12:00 UTC, and 18:00 UTC for any date.

geopotential height difference between  $45^\circ\text{N}$  and  $67.5^\circ\text{N}$ , i.e. the geostrophic zonal wind in this latitude channel.

The normalisation of every index to the standard deviation of the respective time series is an important step in order to work with a grid size  $dq$  which is constant for all variables. The de-seasonalisation of the time series is also a very important point, because in this thesis master equations with constant transition coefficients<sup>6</sup> are considered. Thus the mean, and the annual and semi-annual harmonics are deducted in between a Fourier analysis and synthesis of any time series. The time series are now ready to be used for estimating the transition coefficients of a master equation. To visualise this time series generation, Fig. 4.1 shows the beginning of the time series of the temperature anomaly in the arctic stratosphere (averaged north of  $75^\circ\text{N}$ ) before [Fig. 4.1 (a)] and after normalisation [solid line in Fig. 4.1 (b)]. In Fig. 4.1 (b) the seasonal cycle (consisting of the mean plus the annual and the semi-annual harmonics) is shown by the dashed line. The (de-seasonalised) time series of  $T'$  is shown in Fig. 4.1 (c). A few stratospheric winter warmings are easily recognisable in Fig. 4.1 (c). Winter and spring are clearly the principal contributors of anomalous values. Several studies in the past focused primarily on the winter season (e.g. Holton and Tan, 1980, Labitzke, 1987, Baldwin and Dunkerton, 1999). The circulation conditions during the winter allow in fact the propagation of planetary waves from the troposphere to the stratosphere. A recent paper of Labitzke (2004a) suggests, however, the introduction of the QBO throughout the year. Moreover, master equations derived for time series from the winter season have turned out to be quite complicated to work with and to interpret. Data had to be manually processed in order to obtain a stationary solution for the master equation, and this solution differed substantially from the time mean PDF (i.e. the climate) observed in the data<sup>7</sup>. It is the author's opinion that no information should be left or filtered out of the data (besides of course the seasonal cycle). This applies also to the summer and the fall. Thus time series for the complete period (September 1957 to August 2002) are considered.

In subsection 3.2.3 we learnt that master equations better describe the underlying system when the time series have a coarse time resolution. In this light it is of course tempting to work with monthly daily means. However, large data sets are needed in order to correctly estimate the transition coefficients (Egger, 2001). In order to work with a coarse time resolution and to obtain significant estimates of the transition coefficients, time series from the ERA-40 daily means are analysed in the following way. Let  $dt_t$  be the time resolution of the time series, i.e.  $dt_t = 1$  [day], and  $dt$  designate the optimum time resolution, with  $dt = S \cdot dt_t$ , where  $S$  is integer and  $S \geq 1$ . A set of  $S$  time series is developed from the original time series. The  $s^{\text{th}}$  time series starts on the  $(s-1)$  day after beginning (i.e. 1957-09-01) and  $(S-1)$  days are "skipped" before writing down the following value. The same is repeated for  $s = 1, \dots, S$ . Grid box transitions are thus considered on the base of the value observed  $S$  days after the current value. The terms in (2.7) are counted on the base of the transitions observed in these  $S$  time series<sup>8</sup>. In this fashion low values of the

<sup>6</sup>Seasonal dependence may of course be an interesting issue which, as discussed in chapter 2, would require very long data sets. These may be easily generated from a model run, as is the case in Pasmanter and Timmermann (2002). In this thesis, however, observational data are considered, so that their length is limited in time.

<sup>7</sup>For ERA-40 monthly daily means it turned out that the QBO-West phase was overemphasised in the stationary solution.

<sup>8</sup>Again, the last observation of each of the  $S$  time series does not increment  $N_i$ .

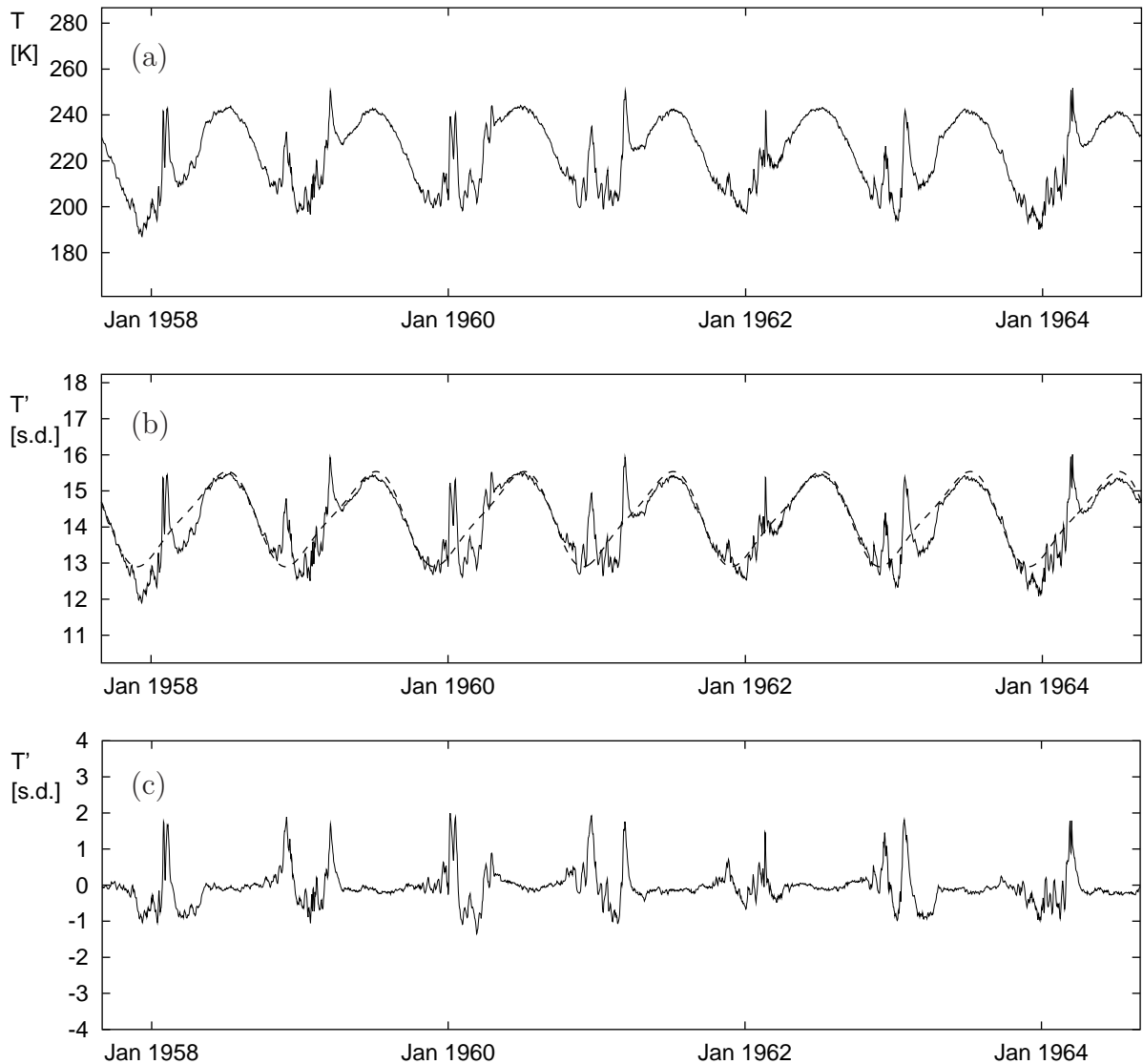


Figure 4.1: Normalisation and de-seasonalisation of a time series of the temperature of the arctic stratosphere in 10 hPa, averaged north of  $75^\circ\text{N}$ . (a) The time series as obtained from the ERA-40 daily means. The stratospheric warmings occurring in the winter months are clearly visible. (b) The same time series after normalisation with its standard deviation (*solid*), and its seasonal cycle (*dashed*, consisting of the mean, and the annual and semi-annual harmonics). (c) The time series after de-seasonalisation, now ready for use with a master equation. The time series for the entire period covered by ERA-40 is shown in Fig. 4.3.

noise-to-signal ratio  $R_w$  (and consequently a high statistical significance of the transition coefficients) can be obtained despite the coarser time step  $dt$ . A drawback of this technique is that stationarity of the master equation may not be guaranteed for very large values of  $S$ . Moreover, the quality increase of the master equation levels off with increasing  $S$ . In the following sections we will thus present and discuss results obtained for values of  $S$  ranging between a week ( $S = 7$ ) and a “month” ( $S = 30$ ). Some results obtained from the ERA-40 monthly daily means are mentioned in section 4.2.

## 4.2 The QBO and the arctic stratosphere

One of the main aims of climate studies is the search for atmospheric modes of oscillation with long time-scales. The QBO of the zonal wind in the stratosphere over the equator represents the most striking evidence of a long-term periodic atmospheric oscillation<sup>9</sup> besides the externally forced annual components (with their harmonics). There is increasing evidence that the QBO influences, among other things, the global circulation in the stratosphere, the breakdown of the polar vortices, and the atmospheric ozone, thus indirectly influencing the troposphere (e.g. Holton and Tan, 1980, Stolarski et al., 1991, Baldwin et al., 2001). This evidence motivates the analysis and discussion presented in this section, which in turn is an important base for the study on the role of solar variability presented in section 4.3.

The impact of the QBO on the temperature in the arctic stratosphere is investigated in this section with a three-dimensional master equation. The coefficients of this master equation are estimated from a time series of the variable set  $(Q_I, T', U_{e20})$  obtained as explained in section 4.1. Figure 4.2 shows the complete time series of  $Q_I$  (solid line) and  $U_{e20}$  (dashed line) for the period covered by ERA-40. Positive anomalies of approximately 1.0 [st. dev.] or above occur during the West phase of the QBO (see also Table 4.1). The two variables are needed in order to predict  $Q_I$ -values. The mere knowledge of  $Q_I$  would not suffice. For each  $Q_I$ -interval an approximately equal number of transitions is observed towards higher as towards lower values of  $Q_I$ . In fact, when the QBO West [East] phase is advancing,  $Q_I$  increases [decreases] with time and a transition will be observed towards an interval on the right [left] along the  $Q_I$ -axis. Mere diffusion towards the mean climate state can therefore be expected in the two-dimensional  $(Q_I, T')$  phase plane. By knowing  $Q_I$  and  $U_{e20}$  it is possible to predict a higher value of  $Q_I$  if  $U_{e20} > Q_I$ , as  $U_{e20}$  leads  $Q_I$  (see Fig. 4.2). This is a consequence of the downward propagation of wind anomalies in the equatorial stratosphere. Figure 4.3 shows the complete time series of  $T'$  for the period covered by ERA-40. Stratospheric warmings are easily recognisable with their temperature peaks, most of which are followed by short periods with negative anomalies. The first half of the 1990s is slightly warmer than the rest of the time series. Moreover, a temperature trend can hardly be seen in these “raw” data. The first few lines of the time series of these three variables are reported for completeness in Table 4.1, where the first two lines represent the mean and standard deviation of the three variables.

Fig. 4.4 (a) shows a projection on the  $(Q_I, T')$  phase plane of this time series’ distribution of observed states. In Fig. 4.4 (b) are shown isolines of the corresponding observed state density  $\rho$  (integrated along the  $U_{e20}$ -axis) for a phase-space discretisation with grid size  $dq = 0.5$  [standard deviations]. The long-term, stationary PDF prediction from a master equation with  $dt = 30$  [days] for an arbitrary initial condition is shown for reference in

---

<sup>9</sup>The QBO of the equatorial zonal wind is characterised by alternating phases with westerlies and easterlies respectively with an average period of 22 to 34 months. Wind regimes at a height of approx. 40 km move downwards with a speed of about 1.3 km/month, with west phases advancing more uniformly downwards compared with east phases. The amplitude of the QBO is nearly constant between 10 and 40 hPa and decreases towards ground. The QBO dominates the long-term variation of the zonal wind between  $-10S \leq \phi \leq 10N$ . Positive shear is associated with a local positive temperature anomaly. The QBO is driven by the propagation into the stratosphere of Kelvin and Rossby gravity waves (e.g. Holton, 1992, Andrews et al., 1987).

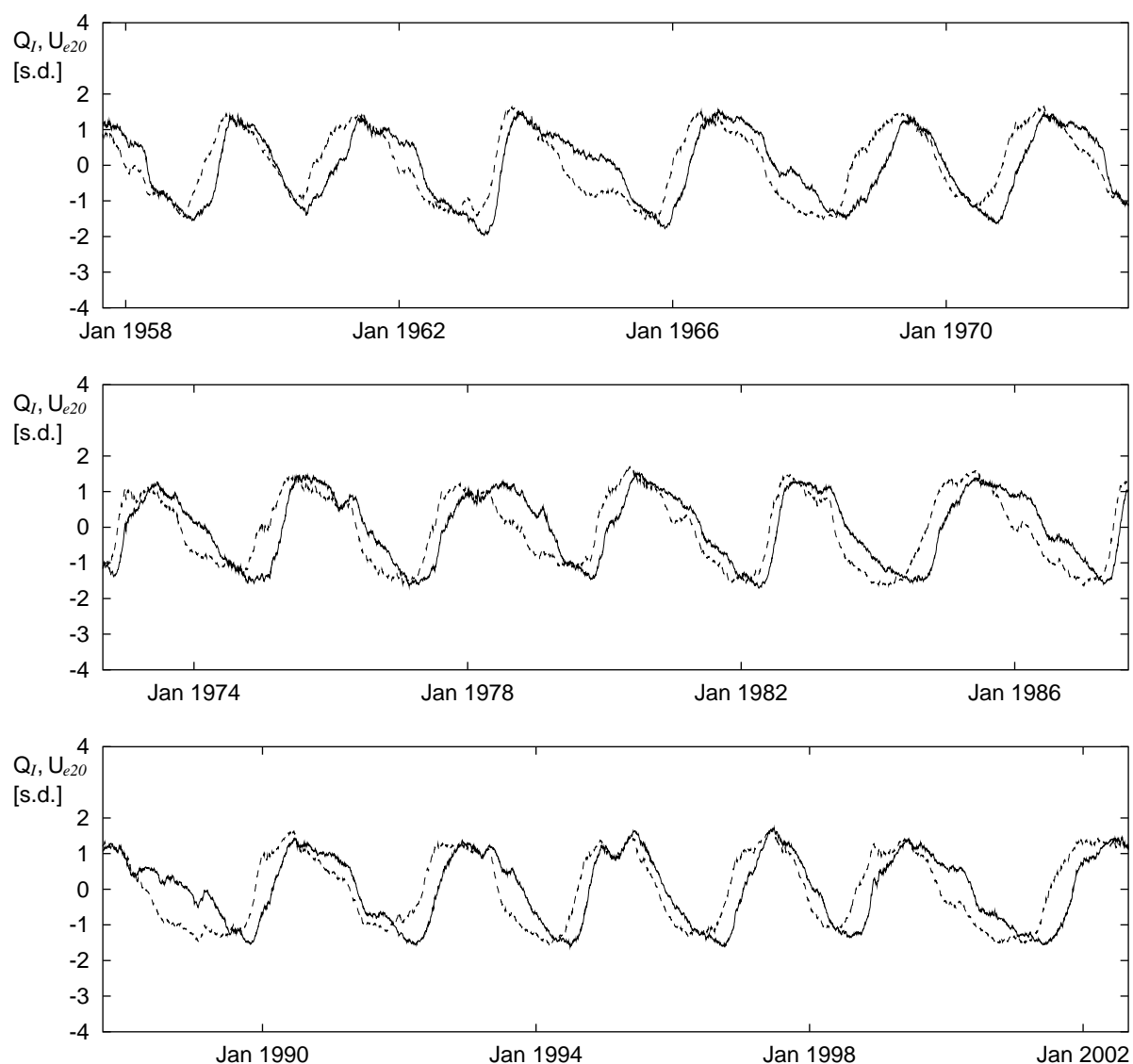


Figure 4.2: Time series of the normalised and de-seasonalised QBO index ( $Q_I$ , *solid*) and of  $U_{e20}$  (*dashed*) obtained from the ERA-40 daily means. Positive anomalies of 1.0 or above [st. dev.] indicate westerly phases of the QBO.

# - mean =	-7.31661	223.82137	-10.75571
# / stdev=	16.61775	15.72660	18.91079
1957-09-01	1.20998	-0.06652	0.87737
1957-09-02	1.19300	-0.04491	0.81894
1957-09-03	1.14083	-0.07391	0.83091

Table 4.1: The first few lines of a time series obtained from the ERA-40 daily means for  $Q_I$ ,  $T'$ , and  $U_{e20}$  respectively. In the first two lines, which don't yet belong to the time series, are reported the mean and the standard deviation of  $Q_I$ ,  $T'$ , and  $U_{e20}$  in [ $\text{m s}^{-1}$ ], [K], and [ $\text{m s}^{-1}$ ] respectively.

Fig. 4.4 (c) and is almost not discernible from the latter<sup>10</sup>. This is a first sign that the

<sup>10</sup>The stationary solution delivered by a master equation will tend to the observed state density for longer time series, but will not necessarily be in perfect agreement with it. Recall also the discussion in



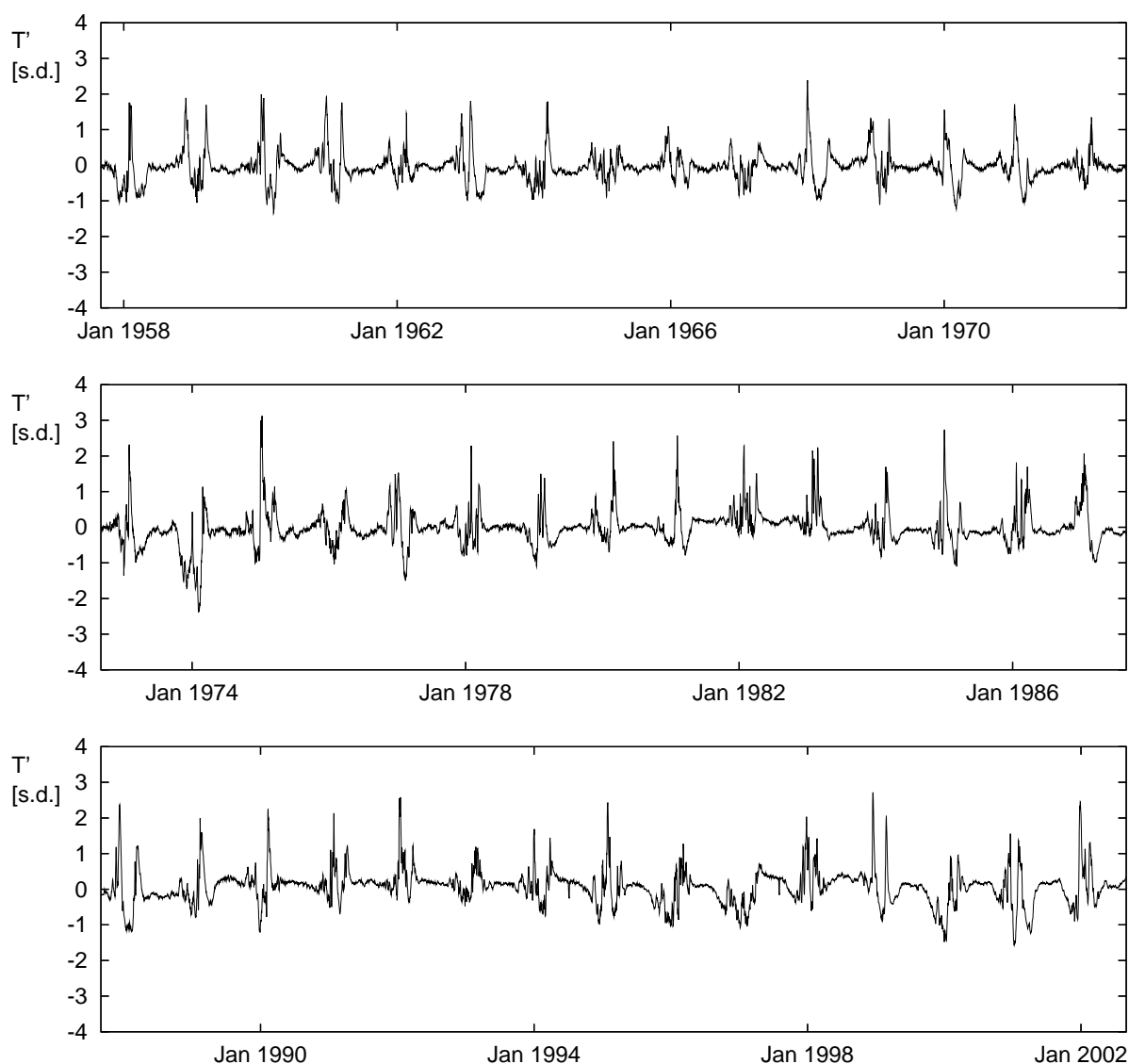


Figure 4.3: Time series of  $T'$  (normalised and de-seasonalised temperature anomaly in 10 hPa north of  $75^\circ\text{N}$ ), obtained from the ERA-40 daily means. Strong positive anomalies indicate the occurrence of a stratospheric warming.

master equation correctly describes the observed system. The noise-to-signal ratio  $R_w$  obtained under these conditions is  $R_w = 0.32$ , which, together with the otherwise needed computer resources, discourages the use of a finer grid size<sup>11</sup>.

Correlation functions for this variable set ( $Q_I$ ,  $T'$ ,  $U_{e20}$ ), as obtained from the considered time series and from the master equation, are shown in Fig. 4.5. These may be used to test the Markovian assumption. The auto-correlation of the QBO index  $r_{\text{QQ}}$  from the master

page 31, where it is pointed out that master equations derived for the winter months delivered stationary solutions which overemphasised the QBO West phase.

<sup>11</sup>In this case it could of course be argued that  $R_w$  would have been lower if 6-hourly values had been used (instead of daily means). In that case acceptable  $R_w$  values could have been obtained also for, say,  $dq = 0.4$ . However, time series of all variables must be consistent, and the solar flux at 10.7 cm is measured on a daily basis (recall section 4.1).

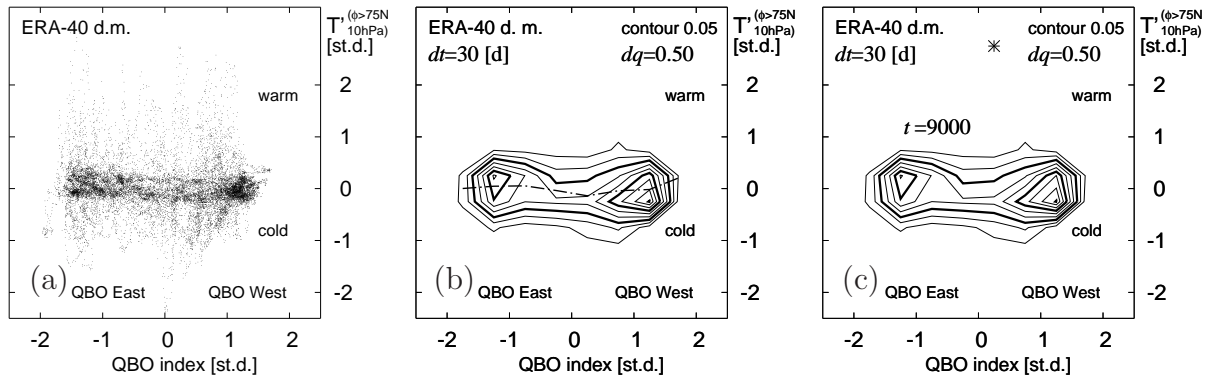


Figure 4.4: Daily mean values from the time series of the variable set  $(Q_I, T', U_{e20})$  obtained from ERA-40 as described in section 4.1 are projected in (a) onto the  $(Q_I, T')$  phase plane. After discretising the phase space with  $dq = 0.5$  and integrating along the  $U_{e20}$ -axis, one obtains the observed state density  $\rho$  whose isolines are shown in (b), where the time step  $dt = 30$  [days]. The *dashed-dotted* line indicates the mean  $T'$  as a function of the QBO index. (c) The stationary PDF (at  $t = 9000$ ) obtained from a master equation for the time series of the three variables. The sharp initial condition is marked by a star  $*$ ; time step  $dt = 30$  [days]; grid size  $dq = 0.50$  [st. dev.]; the PDF has been integrated along the  $U_{e20}$ -axis. The isoline contour interval in (b) and (c) is 0.05, and every third isoline is bold.

equation for a time step  $dt = 30$  [days] (*dashed* line) decays faster, i.e. the transition probabilities must have a longer memory of that captured by the estimated ones. This could be improved, for instance, by increasing the time step  $dt$ . The decay of  $r_{\text{QQ}}$  is in fact closer to that of the data if the time step  $dt = 90$  [days] is chosen (*dashed-dotted* line). The top right panel with  $r_{\text{TT}}$  suggests, however, that such a time step is not appropriate, as the time scale<sup>12</sup> of  $T'$  may not be described with a larger  $dt$ . The auto-correlation function  $r_{\text{TT}}$  delivered by the master equation for  $dt = 30$  is very encouraging. Altogether, this is a case where  $T'$  is just slightly correlated to the other variables, whereas  $Q_I$  and  $U_{e20}$  oscillate. All in all, this master equation well satisfies the Markovian assumption<sup>13</sup>.

In Fig. 4.4 (b-c) the local maximum for QBO East [West] has a slightly positive [negative] temperature anomaly. This is in qualitative agreement with the relationship suggested by Holton and Tan (1980), who analysed differences in composites of 50 hPa geopotential height and geostrophic wind in the winter months. Holton and Tan used monthly mean Northern Hemisphere data for the period 1962-1977 (16 years), and found that the zonal mean geopotential height at high latitudes was significantly lower during QBO West than during QBO East (see also Baldwin et al., 2001, p. 218-220). They suggested a mechanism whereby shifts in the latitude of the zero mean zonal wind line (critical line) were responsible for the extra-tropical QBO. The study shown in this section is based on the complete period (with no seasonal preference) covered by ERA-40 (45 years), so that a more robust quantitative analysis is possible. This  $Q_I - T'$  relationship may be seen also in the panel for  $r_{\text{TQ}}$  in Fig. 4.5. The cross-correlation is slightly negative for lag zero<sup>14</sup>, and

<sup>12</sup>The time scale for a variable may be defined as the lag for which its auto-correlation decays by a factor  $\frac{1}{e}$  (Baldwin et al., 2003).

<sup>13</sup>This is not the case for a master equation based on just the first two variables, where correlations decay monotonically to zero (see Fig. 4.9).

<sup>14</sup>Due to the volume of the data (16436 data points), even very low correlations are significant. A  $t$ -test (see Schönwiese, 1985, p. 138-143) suggests that a value of  $r = 0.02$  is significant at the 99% confidence level for lags up to 1500 [days]. It could be argued that the data are not completely independent. The number

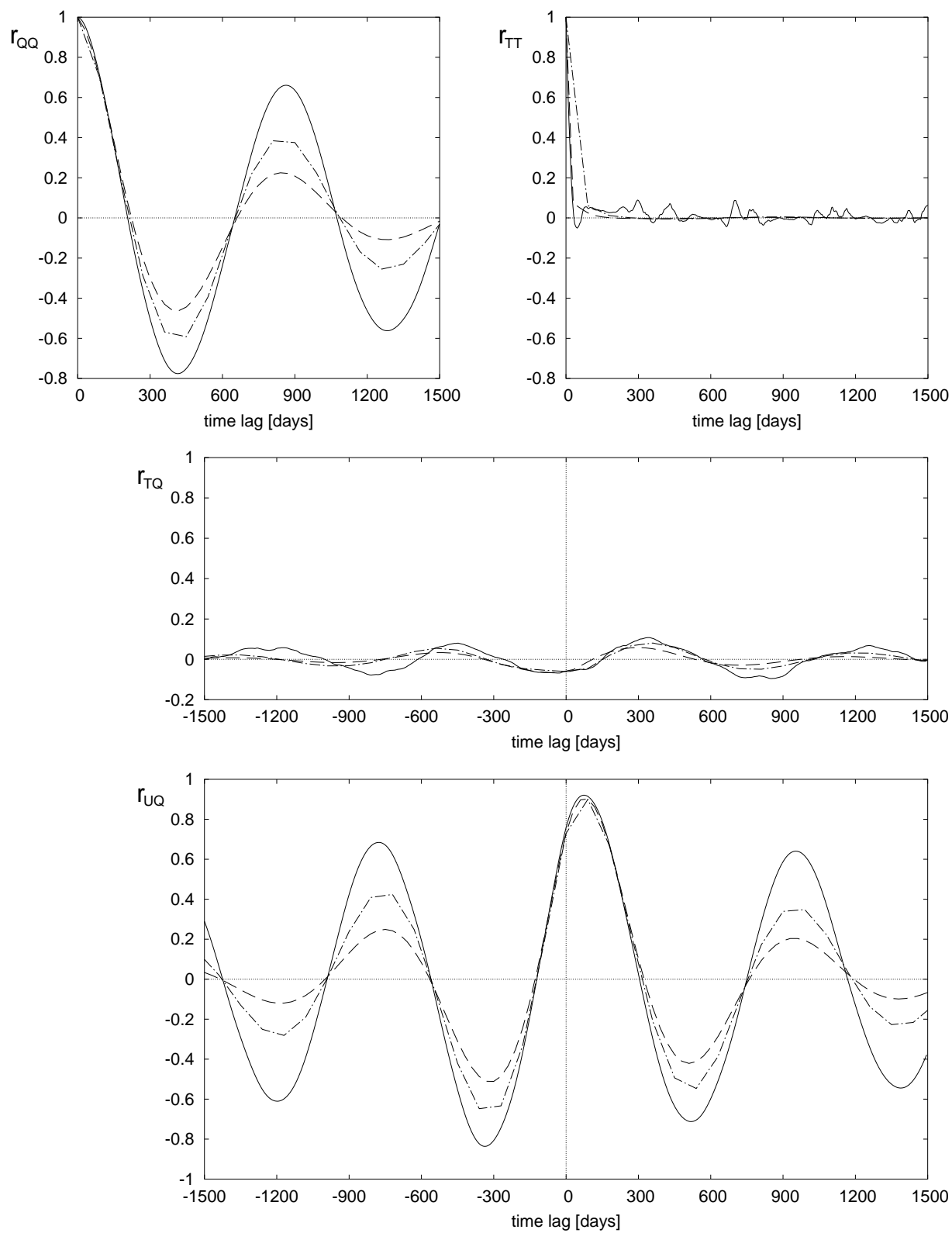


Figure 4.5: Correlation functions  $r_{QQ}$ ,  $r_{TT}$ ,  $r_{TQ}$ , and  $r_{UQ}$  as observed (*solid*), and as reproduced by master equations derived from a time series of the variable set  $(Q_I, T', U_{e20})$  obtained from ERA-40 daily means, *dashed* for  $dt = 30$ , and *dash-dotted* for  $dt = 90$  [days]; phase space grid size  $dq = 0.50$  [st. dev.].

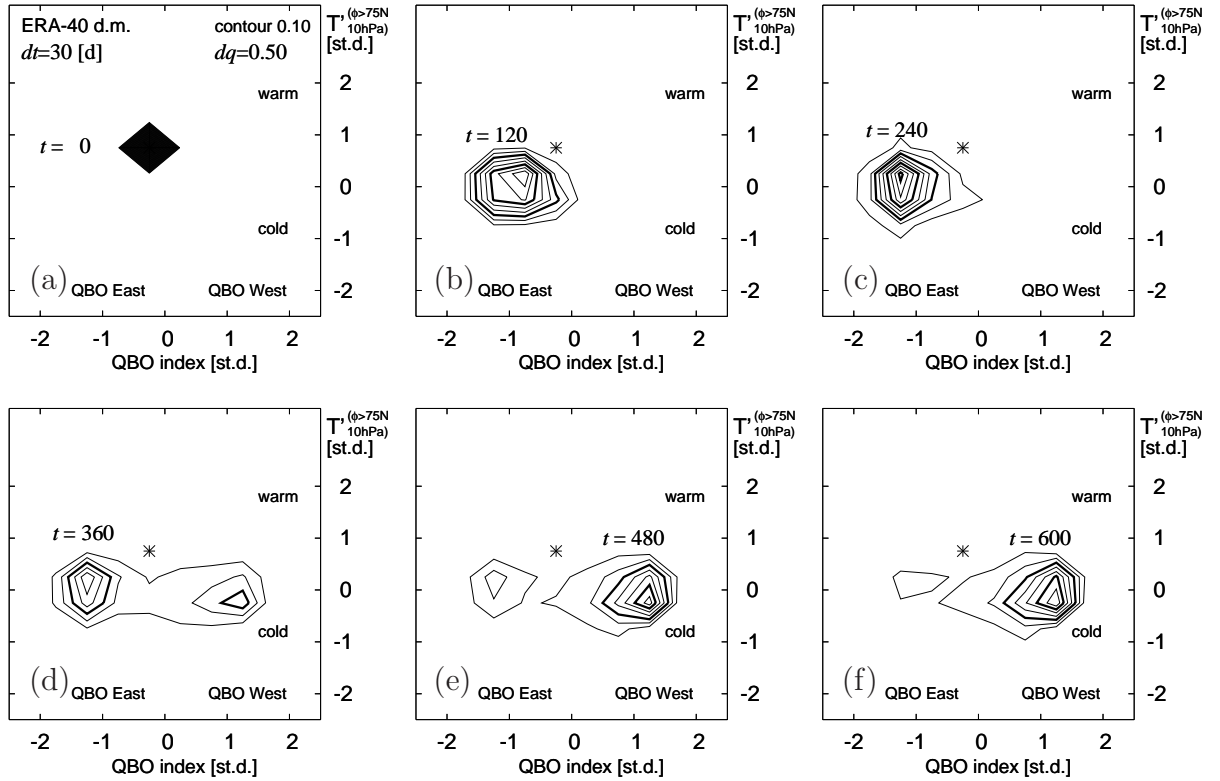


Figure 4.6: The development of a PDF cloud from a sharp initial condition (marked by a star  $*$ ) is shown at a few selected times. These PDF forecasts are delivered by a master equation derived from a time series for the variable set  $(Q_I, T', U_{e20})$  obtained from the ERA-40 daily means; time step  $dt = 30$  [days]; grid size  $dq = 0.50$  [st. dev.]. The PDF has been integrated along the  $U_{e20}$ -axis. The isoline contour interval is 0.10, and every third isline [(b) through (f)] is bold.

displays otherwise a wave-like behaviour with a period corresponding to that of the QBO. If the grid size is decreased, however, the position of the maxima in Fig. 4.4 (b-c) shifts towards  $T' = 0$  (not shown). Besides the maxima, the mean  $T'$  for a given  $Q_I$ -interval is given by the *dashed-dotted* line in Fig. 4.4 (b). Surprisingly, a clear inclination of the dashed-dotted line between QBO East and QBO West may hardly be identified in this time series obtained from the ERA-40 daily means.

The stationary PDF in Fig. 4.4 gives just a static picture of the distribution of states. To understand more about dynamics in phase space, the master equation is integrated for several initial conditions. PDF clouds are initially of grid box size (sharp initial condition). The evolution of a PDF cloud from an arbitrary initial condition delivered by a master equation with time step  $dt = 30$  for a grid size  $dq = 0.5$  is shown in Fig. 4.6. The slightly positive temperature anomaly abates within the first time step as may be seen from Fig. 4.6 (a, b). The high PDF values on the left side of the phase space in Fig. 4.6 (b, c) indicate that a QBO East phase is reached within  $t = 120$  and  $t = 240$ . At  $t = 600$  the system has developed towards a QBO West phase. The master equation thus describes the QBO. Some selected mean trajectories [recall Eq. (2.9)] are shown in Fig. 4.7. The stars mark the initial conditions, and are located in the centre of the grid box ( $\mathbf{z}^0$ ) where the PDF

---

of data could be divided by the number of days needed for  $r_{QQ}$  to decay to  $e^{-1}$  to obtain approximately 100 independent data. In this case a value  $r = 0.15$  would still be significant at the 90% confidence level.

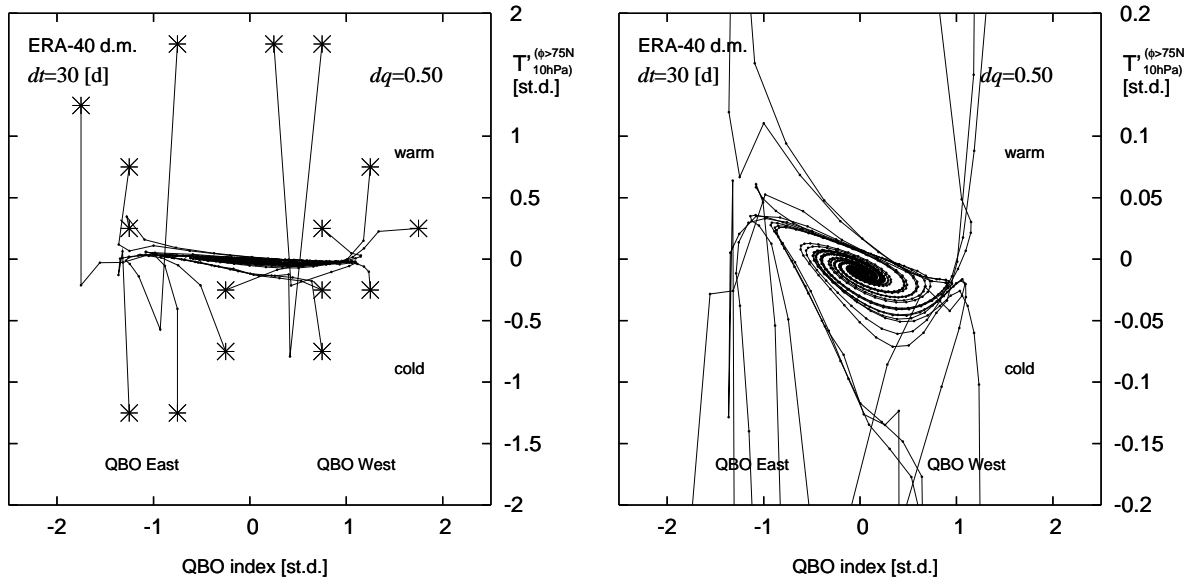


Figure 4.7: A few selected mean trajectories delivered by a master equation for the variable set  $(Q_I, T', U_{e20})$ ; time step  $dt = 30$  [days]; grid size  $dq = 0.5$  [st. dev.]; time series from ERA-40 daily means. The trajectories have been projected on the  $(Q_I, T')$  phase plane. The stars  $*$  mark the initial conditions, and the dots  $\cdot$  mark the mean position every 30 days. In the right panel the  $T'$ -axis has been expanded by a factor 10 to better study the behaviour of trajectories beyond the time scale of temperature anomalies.

is initially set  $f_{2,0} = 1/dq^3$ . As may be seen on the left panel of Fig. 4.7, the trajectories show impressive bends at their beginning. Great temperature anomalies abate within 30 days and great positive anomalies reverse at the next time step. In the right panel of Fig. 4.7 the  $T'$ -axis has been expanded to better visualise the behaviour of trajectories beyond the time scale of temperature anomalies. On the longer time scale, the circulation in phase space is anticyclonic and, as expected, contracting towards the origin. The origin is in fact the centre of the observed state density  $\rho \simeq \bar{f}$  and represents the mean climate state. Moreover, this (long-term) behaviour of a damped oscillator is accompanied with higher [lower] temperatures during the QBO East [West] phase. This is in qualitative accordance with the results suggested by Holton and Tan<sup>15</sup>. Furthermore, a slight cooling takes place during the West QBO phase. The inclination of trajectories is  $-0.03$  [nondimensional] from QBO East to QBO West (which corresponds to  $-0.03$  [K m<sup>-1</sup> s]); as the trajectories move from QBO West to QBO East the inclination is approximately  $-0.05$  (corresponding to  $-0.05$  [K m<sup>-1</sup> s]). This suggests a temperature difference between the two QBO phases of 2 [K]. Please recall that we are working with the complete de-seasonalised data, and not with a particular season of the year; this figure may undergo a seasonal cycle. The master equation provides therefore an interesting visualisation of complex dynamics in the phase space. It also allows an objective, comprehensive and robust quantification of the relationship between the QBO and the temperature in the arctic stratosphere<sup>16</sup>. The

<sup>15</sup>Holton and Tan (1980) found in the zonal mean at 50 hPa for November-December a geopotential height difference of 100 [m] north of 75°N, and a geostrophic wind difference of 5 [m s<sup>-1</sup>] at 65°N (significance level of 5 to 10%).

<sup>16</sup>Behind standard approaches like compositing of data there is always the choice of catalogising *winters* in “East” or “West” (see also Fig. 1 in Thompson et al., 2002); this is not possible when zonal winds in the equator are in between the two phases. The PDF forecasts delivered by the master equation are more objective and general.

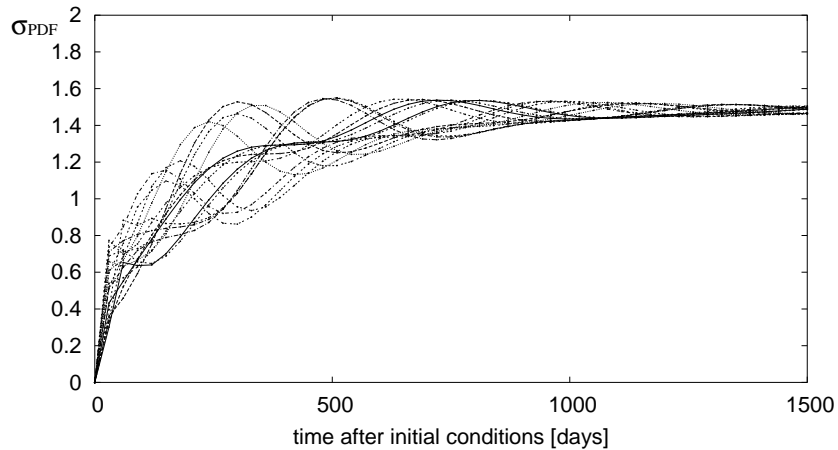


Figure 4.8: The standard deviation of the PDF clouds for the initial conditions considered in Fig. 4.7 shown as a function of time.

standard deviation  $\sigma$  of the position of the clouds [recall (2.10)] is shown in Fig. 4.8 as a function of the time after the initial condition. After about 500 days  $\sigma$  has grown to 1.3 [st. dev.], which is almost the standard deviation of the data themselves.

The use of ERA-40 monthly daily means deserves a mention here. If a master equation is derived from monthly daily means of the variables, the noise-to-signal ratio for the transition coefficients is  $R_w = 0.85$ . This means that confidence intervals are almost of the same order of magnitude as the transition coefficients themselves and, consequently, their statistical significance low. This supports the idea of working with time series having a finer time resolution, like, for instance, daily means<sup>17</sup>.

Let us turn our attention to a two-dimensional master equation for the variable set  $(Q_I, T')$ . The phase space grid size is unchanged ( $dq = 0.5$ ). The PDF forecast delivered for time  $t = 600$  [days] by a master equation in the phase plane of this variable set is shown in Fig. 4.9 (a). This master equation has been derived from the same time series of  $Q_I$  and  $T'$  as the three-dimensional one, whose forecast for  $t = 600$  is shown in Fig. 4.6 (f). Two-dimensional forecasts cannot reproduce the QBO, as may be seen by comparing Fig. 4.9 (a) with Fig. 4.4 (f). In Fig. 4.9 (a) is also given the mean trajectory for this master equation run; the square gives the mean position at  $t = 600$ . Beyond the time scale of temperature anomalies a monotonic approach to the mean climate state is observed. The auto-correlation function  $r_{QQ}$  of the QBO index is shown in Fig. 4.9 (b). The monotonic decay of  $r_{QQ}$  confirms that the two-dimensional master equation describes a monotonic approach to the mean climate state characterised by strong diffusion. However, it is interesting to note that the trajectory approaches the origin along a line with the same inclination observed in the right panel of Fig. 4.7; the same applies to trajectories approaching the origin from positive  $Q_I$  values<sup>18</sup> (not shown). Caution is needed as the Markovian assumption is barely fulfilled beyond a time scale of  $\sim 100$  days. This discussion shows that a minimum of two variables are needed just to reproduce a single oscillation, in our case  $Q_I$  and  $U_{e20}$ . To study the relationship of an oscillation with another variable

<sup>17</sup>The correlation functions delivered by a master equation for ERA-40 monthly daily means show, however, about the same progression as those seen for ERA-40 daily means with  $dt = 30$ .

<sup>18</sup>This is an important result, because it allows some further interpretation whenever an oscillation may not be described because of limited computer resources.

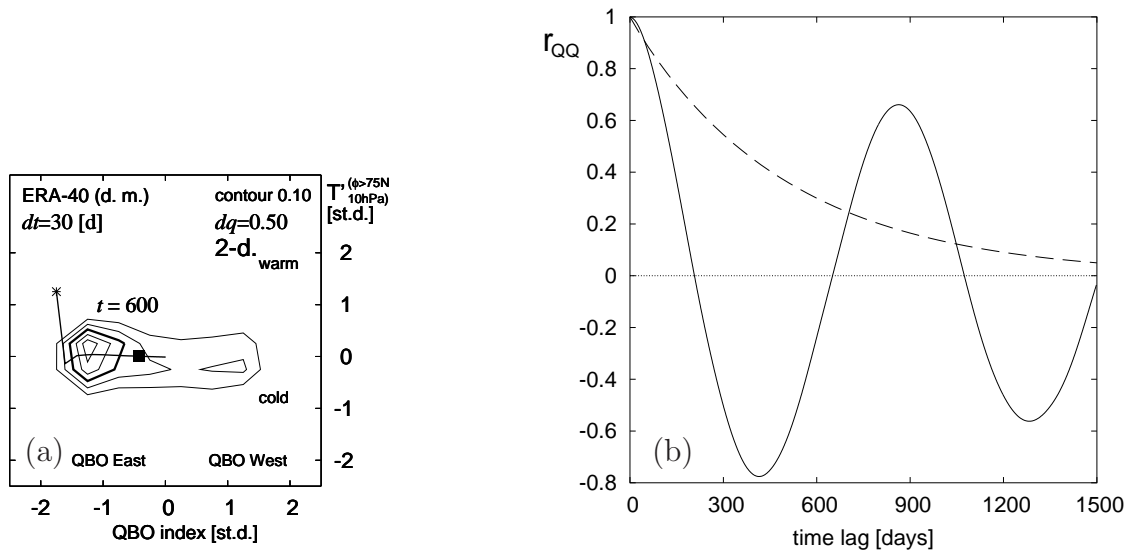


Figure 4.9: (a) The PDF forecast for time  $t = 600$  delivered by a master equation derived from a time series of the variable set  $(Q_I, T')$  obtained from ERA-40 daily means. Time step  $dt = 30$  [days]; grid size  $dq = 0.5$ . The star  $*$  marks the sharp initial condition. The mean trajectory is also shown, and the square  $\blacksquare$  marks the mean position at  $t = 600$ . The isoline contour interval is 0.10, and every third isoline is bold. The corresponding forecast, obtained with a three-dimensional master equation is shown in Fig. 4.4 (f). (b) The auto-correlation function of the QBO index  $r_{QQ}$  as observed (*solid*) and as reproduced by the master equation (*dashed*).

at least three variables are needed. This issue will be further treated in section 4.3, where the effects of two oscillations on  $T'$  are considered.

### 4.3 The role of the 11-year solar cycle

A decadal wave is observed in the stratosphere, which has been attributed to the 11-year SC<sup>19</sup> (e.g. van Loon and Labitzke, 1998, Labitzke, 1998, Labitzke and van Loon, 1999). This signal is strongest above the tropics in the summer hemisphere, and the QBO is believed to modulate its global signal during the northern winter (Labitzke, 2001). The mechanisms behind the relatively strong response of the stratospheric climate to small variations in the solar irradiance are not yet fully understood nor is the mechanism of interaction of the QBO and the 11-year SC (e.g. Labitzke, 1998, Labitzke and van Loon, 1999, Salby and Callaghan, 2000, Rind, 2002). It was suggested that the Hadley circulation intensifies during the maximum phase of the 11-year SC (e.g. Labitzke, 2001), and that the Brewer-Dobson circulation is also affected by the 11-year SC (e.g. Kodera and Kuroda, 2002, Labitzke, 2001, 2004a). Kodera and Kuroda (2002) investigated the strong dynamical response to the solar forcing with focus on the stratopause region and suggested that the solar influence is transmitted downwards through modulation of the oscillation of the polar night jet and a change in the Brewer-Dobson circulation. Gray et al. (2004) recently suggested that the timing of Northern Hemisphere stratospheric sudden warmings is influenced by the interaction between the 11-year SC and the QBO. They proposed a mechanism

<sup>19</sup>The period of the “11-year SC” actually varies between 9.5 and 12 years (Labitzke, 1998, Labitzke and van Loon, 1999).

in which zonal wind anomalies in the equatorial/subtropical stratopause region associated with the QBO and the 11-year SC either reinforce or cancel out. The data-based master equation, however, is a stochastic model, and no assumptions are made a priori about dynamics and physics. One hopes to learn more about the relationship between a few climate variables from the evolution of the PDF.

The interesting issue of the influence of solar variability and of the QBO on the temperature in the arctic stratosphere is investigated with a three-dimensional master equation for the variable set  $(Q_I, T', S_{10.7})$ . The transition coefficients of the master equations are estimated on the base of a time series of the variable set obtained as explained in section 4.1. The time series of  $Q_I$  and  $T'$  have been shown in Fig. 4.2 (solid line) and Fig. 4.3 respectively. Figure 4.10 shows the time series of the index of the solar flux<sup>20</sup> at a wavelength of 10.7 cm  $S_{10.7}$  for the period covered by ERA-40, which includes four cycles. These are well visible in the time series. The phase of higher activity beginning in 1999 is quite strong, showing several peaks. The first few lines of the time series of the three variables are reported for completeness in Table 4.2, where the first two lines represent the mean and standard deviation of the three variables.

# - mean =	-7.31661	223.82137	131.87372
# / stdev=	16.61775	15.72660	55.84401
1957-09-01	1.20998	-0.06652	2.61114
1957-09-02	1.19300	-0.04491	2.48451
1957-09-03	1.14083	-0.07391	2.57276

Table 4.2: The first few lines of a time series for the variable set  $(Q_I, T', S_{10.7})$  obtained as explained in section 4.1. In the first two lines, which don't yet belong to the time series, are reported the mean and the standard deviation of  $Q_I$ ,  $T'$ , and  $S_{10.7}$  in  $[m s^{-1}]$ ,  $[K]$ , and  $[10^{-22} W m^{-2} Hz^{-1}]$  respectively.

Fig. 4.11 (a) shows a projection on the  $(Q_I, S_{10.7})$  phase plane of this time series' distribution of observed states. Figure 4.11b shows isolines of the corresponding observed state density  $\rho$  (integrated along the  $T'$ -axis) for a phase-space discretisation with grid size  $dq = 0.5$  [st. dev.]. The long-term, stationary PDF prediction from a master equation with time step  $dt = 30$  [days] for an arbitrary initial condition is shown in Fig. 4.11 (c) and is almost not discernible from Fig. 4.11 (b). This is a first sign that the master equation correctly describes the observed system. The mean trajectory for the chosen initial condition is shown as well. Beyond an initial increase in the value of  $S_{10.7}$ , this trajectory approaches the origin (i.e. the mean climate state) without any oscillation<sup>21</sup>. The noise-to-signal ratio obtained under these conditions is  $R_w = 0.55$ , which is quite high. However, grid size and time step will be kept here to  $dq = 0.50$  and  $dt = 30$  respectively in order to guarantee consistence with the study in section 4.2. The two maxima seen in Fig. 4.11 (b-c) bring little information, as these are a consequence of the distribution of values of the QBO index  $Q_I$ . Interesting is, however, the slight local maximum centred at  $(-1.25, 0.25)$ , and the minimum located lower on its left. These latter features can be a hint that dynamics for high and low solar activity will show a different behaviour. We will see this while analysing dynamics in phase space with the master equation.

<sup>20</sup>Extreme values of  $+4.00$  [st. dev] and beyond have been replaced by  $+3.99$  in order to keep the array representing transition matrix  $\mathbf{T}$  at a reasonable size.

<sup>21</sup>The period of the 11-year solar cycle is approximately  $3.8 \cdot 10^3$  [days], so that within a time window of 9000 days at least two oscillations could take place.



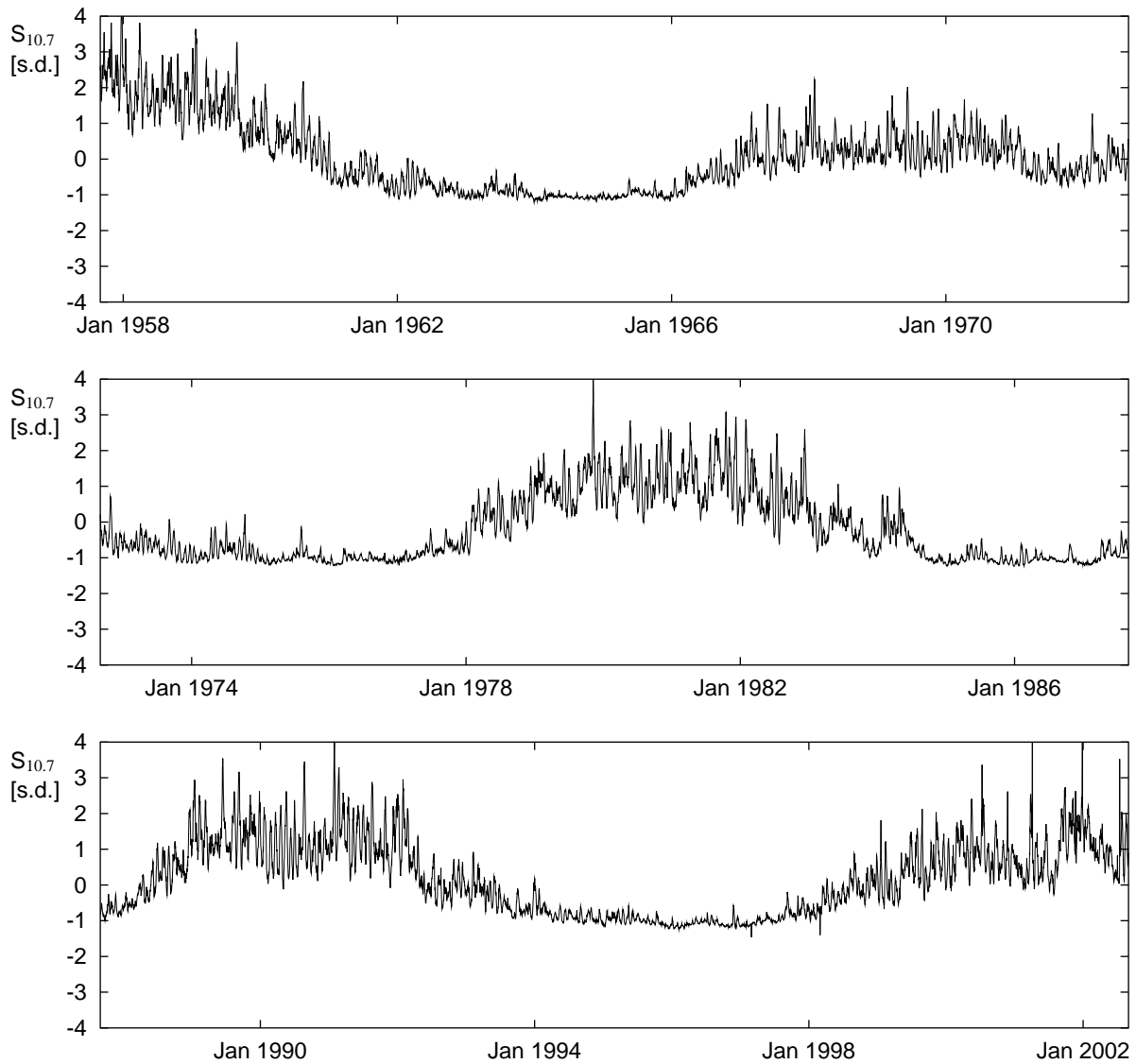


Figure 4.10: Normalised and de-seasonalised time series of the 10.7 cm solar radio flux, measured daily at 20:00 UTC by the Dominion Radio Astrophysical Observatory of the Herberg Institute of Astrophysics, Canada.

Before presenting dynamics in phase space, let us consider Fig. 4.12, where correlation functions for this variable set are shown as a function of time lag. Let us look at the *solid* line in Fig. 4.12, which shows the cross-correlation between  $S_{10.7}$  and  $T'$ ,  $r_{ST}$ , estimated directly from the data. A slight correlation between  $T'$  and  $S_{10.7}$  exists, which is  $r_{ST} \simeq 0.06$ , i.e. positive for lag zero, and grows beyond 0.1 when  $S_{10.7}$  leads  $T'$  by a few months. Moreover, a QBO-like frequency is not immediately visible in  $r_{ST}(\tau)$ . The top left panel of Fig. 4.12 confirms that this master equation describes a monotonic approach to the mean climate state. Caution is therefore needed in evaluating these results, as the Markovian assumption is barely fulfilled beyond short time scales. However, the decay of  $r_{QQ}$  as delivered by this master equation (*thin dashed line*) recalls the decay seen for  $r_{QQ}$  in Fig. 4.9 (b) while discussing the two-dimensional master equation. To reproduce both the QBO and the 11-year SC, we should work here with a coarser grid size and with a

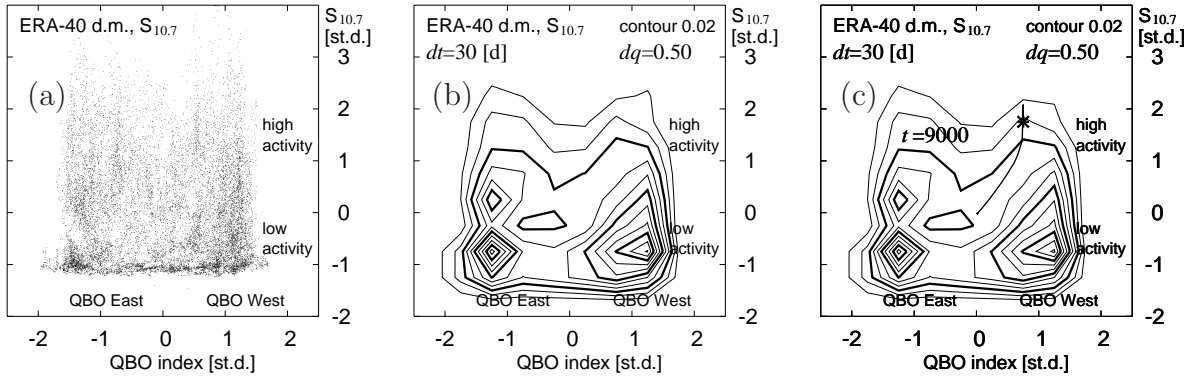


Figure 4.11: Daily mean values from the time series of the variable set  $(Q_I, T', S_{10.7})$  obtained as described in section 4.1 are projected in (a) onto the  $(Q_I, S_{10.7})$  phase plane. After discretising the phase space with  $dq = 0.5$  and integrating along the  $T'$ -axis, one obtains the observed state density  $\rho$  whose isolines are shown in (b), where the time step  $dt = 30$  [days]. (c) The stationary PDF (at  $t = 9000$ ) obtained from a master equation for the considered time series of the three variables. The sharp initial condition is marked by a star  $*$ , and the mean trajectory is shown as well; time step  $dt = 30$  [days]; grid size  $dq = 0.50$  [st. dev.]; the PDF has been integrated along the  $T'$ -axis. The isoline contour interval in (b) and (c) is 0.02, and every third isoline is bold.

five-dimensional master equation for variables  $Q_I, T', S_{10.7}, U_{e20}$ , and a last variable describing, for instance, the value of  $S_{10.7}$  say, 4 years before, according to the time-lag embedding technique (see Kaplan and Glass, 1995). We can expect such a master equation to deliver much better correlation functions, therefore allowing deeper insight. However, we will extend the considerations made at the end of section 4.2 to cautiously evaluate the PDF evolutions delivered by the current three-dimensional master equation.

Let us now turn our attention to dynamics in phase space. Several master equation runs are started from grid boxes where the state vector  $(Q_I, T', S_{10.7})$  resides 30 or more times within the time series. The transition coefficients for the first time step(s) will therefore have smaller confidence intervals. The symbols in Fig. 4.13, 4.14, 4.15, and 4.16 give the mean position every 30 days and can be used to identify a trajectory while comparing a left panel with the corresponding one on the right. The left panels show projections on the  $(Q_I, S_{10.7})$  phase plane. The right panels show projections on the  $(Q_I, T')$  phase plane. The  $T'$  axis is expanded for the same reasons and to allow direct comparisons with the right panel in Fig. 4.7. As mentioned at the end of section 4.2, if the  $T'$ -axis would be expanded in Fig. 4.9 (a) as well, we would see that the trajectory approaches the mean climate state (i.e. the origin) along a straight line having the inclination observed in the right panel of Fig. 4.7. This may be now compared to the inclination that trajectories show in the right panels of Fig. 4.13-4.16 during their approach to the mean climate state. Figure 4.13 shows selected mean trajectories which begin during a QBO East phase and solar maximum. Figure 4.14 shows selected mean trajectories which begin during a QBO West phase and solar maximum. With some abstraction, and recalling the quantitative discussion made on Fig. 4.7, we may interpret the dynamics seen during high solar activity in Fig. 4.13 and Fig. 4.14. The relationship seen in the right panel of Fig. 4.7 is here “shifted” to higher temperature anomalies of  $+0.1$  [st. dev.] (corresponding to  $+1$  K) and more. Moreover, without a five-dimensional master equation, it is hard to interpret the surprisingly high steepness in the  $(Q_I, T')$  phase plane of the trajectory marked by the diamonds in Fig. 4.13. Trajectories beginning with either  $Q_I = 0.25$  or  $S_{10.7} = -0.25$  [st. dev.] at times display

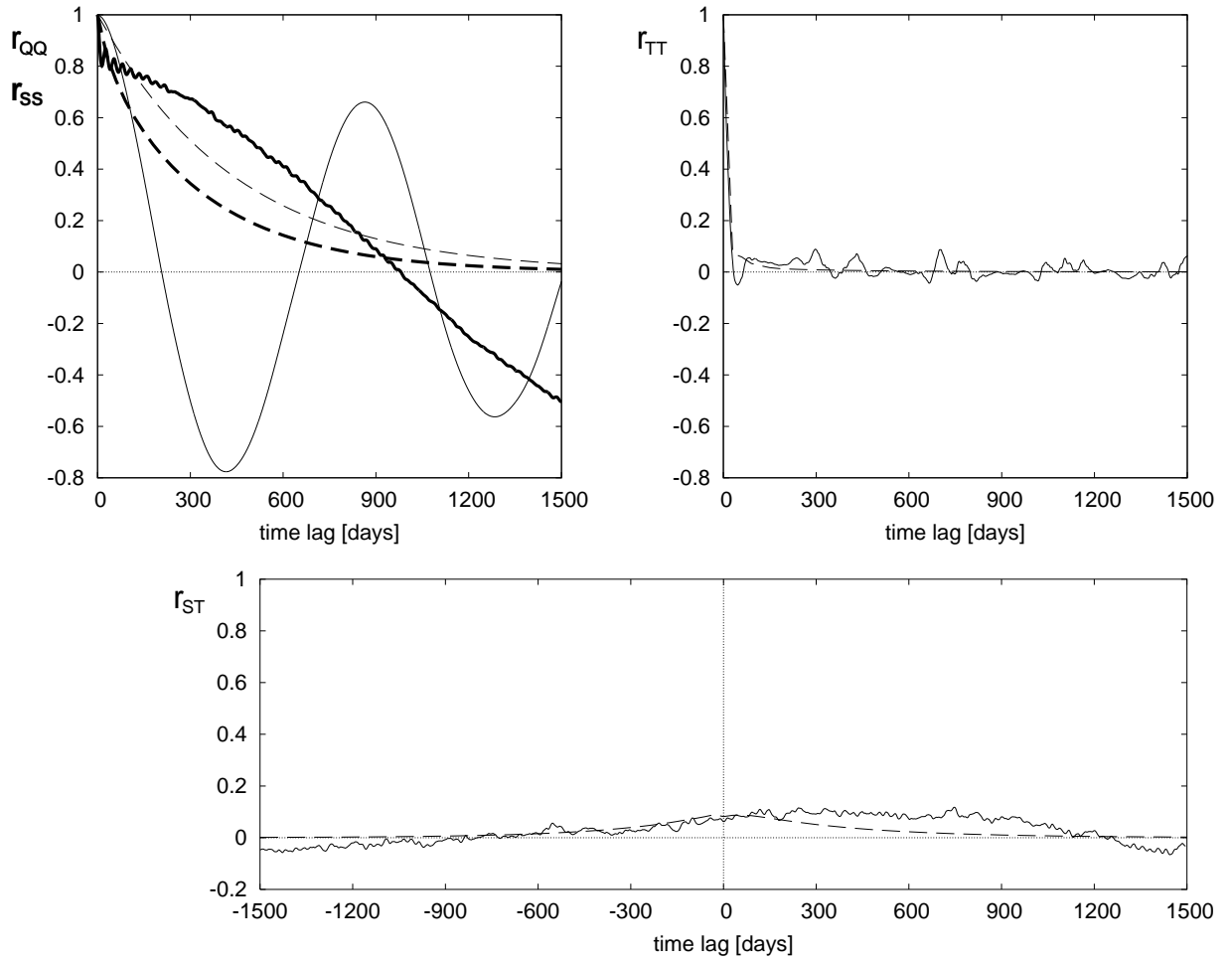


Figure 4.12: Correlation functions  $r_{QQ}$ ,  $r_{SS}$ ,  $r_{TT}$ , and  $r_{ST}$  as observed (*solid*), and as reproduced by the master equation (*dashed*) derived from the time series of the variable set  $(Q_I, T', S_{10.7})$  obtained as explained in section 4.1 ; time step  $dt = 30$ ; phase space grid size:  $dq = 0.50$  [st. dev.]. The *bold* lines in the top left panel refer to values of  $r_{SS}$ .

behaviour exhibited for lower or higher values of the respective variables. Let us now move on to the trajectories beginning with low  $S_{10.7}$  values. Figure 4.15 shows selected mean trajectories which begin during a QBO East phase and solar minimum. Figure 4.16 shows selected mean trajectories which begin during a QBO West phase and solar minimum. The relationship between  $Q_I$  and  $T'$  discussed in section 4.2 is shifted in Fig. 4.15 and 4.16 towards lower temperature anomalies of  $-0.1$  [st. dev.] (corresponding to  $-1$  K) and less. Moreover, by comparing the left with the right panels of Fig. 4.13-4.16 it is evident that the inclination of trajectories on the left panels determines the inclination of the approach to the mean climate state of the trajectories on the right panels<sup>22</sup>. Solar variability influences  $T'$  therefore in an important<sup>23</sup> but subtle way, which could be better understood by increasing the number of variables. The three-dimensional master equation shows that  $T'$  is increased

<sup>22</sup>Trajectories starting from many other initial conditions, which are not shown in Fig. 4.13-4.16, confirm this.

<sup>23</sup>The  $\pm 1$  [K] discussed above turns out to be a much stronger signal than the one shown, for instance, by Labitzke (2001, Fig. 3).

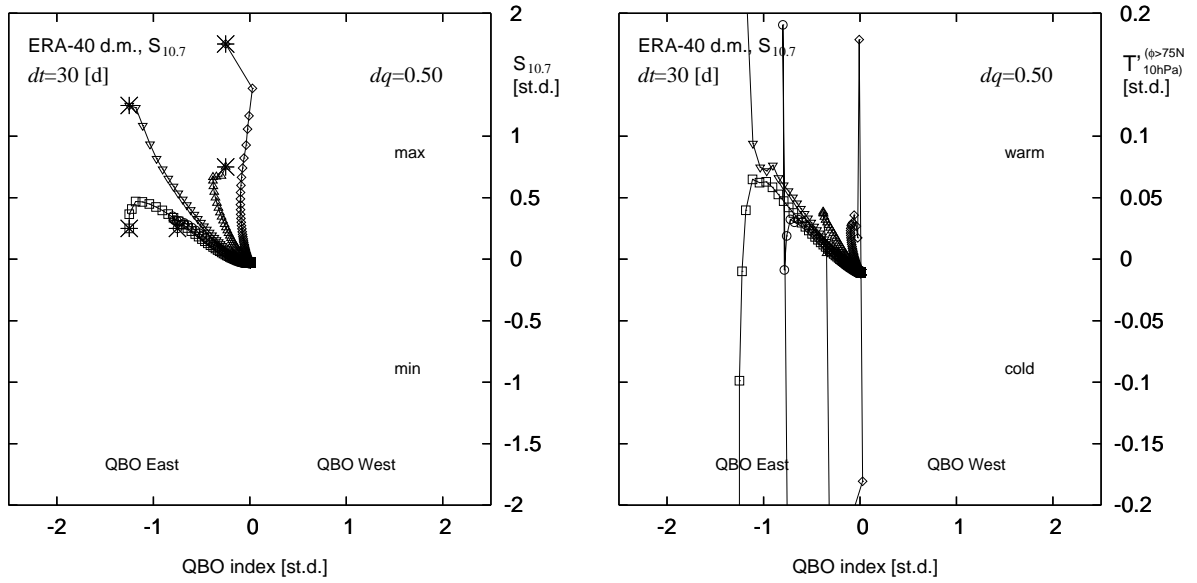


Figure 4.13: Selected mean trajectories delivered by a master equation for the variable set  $(Q_I, T', S_{10.7})$ , initialized with QBO East and high solar activity;  $dt = 30$  [days]; grid size  $dq = 0.50$  [st. dev.]; time series from ERA-40 daily means and the solar radio flux at 10.7 cm. Left [right] panel: projection on the  $(Q_I, S_{10.7})$  [ $(Q_I, T')$ ] phase plane. The stars  $*$  mark the sharp initial conditions; the symbols allow to distinguish between trajectories and give the mean position every 30 days.

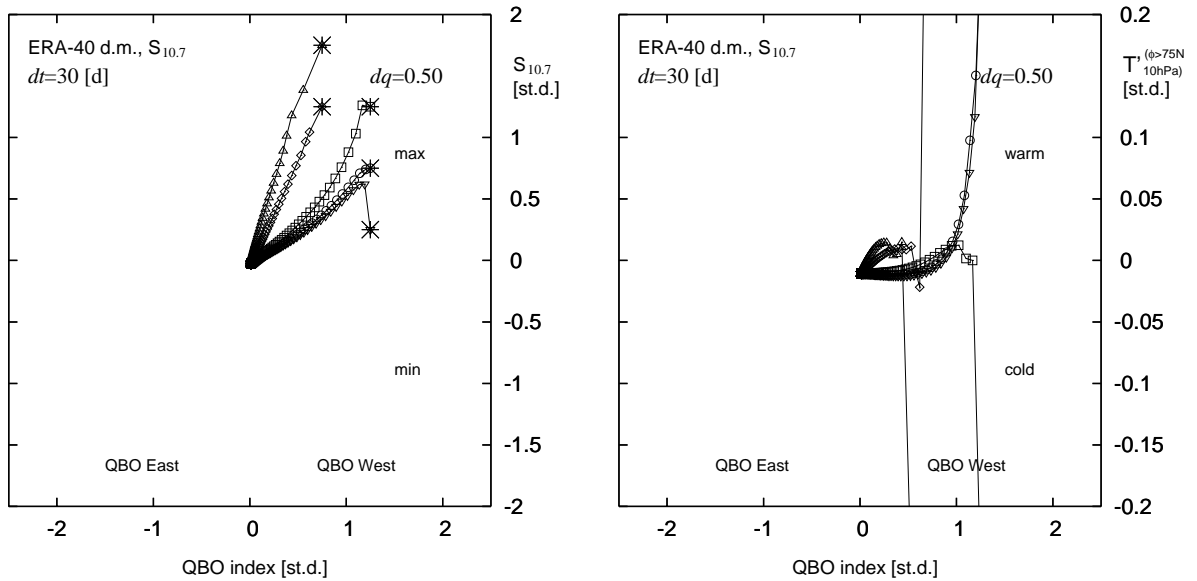


Figure 4.14: As in Fig. 4.13 but initialized with QBO West.

by high solar activity, and this, though to a different extent, during both phases of the QBO. This fosters the speculation that the 11-year SC enhances the Brewer-Dobson circulation during its higher phase. These results are somewhat different from the ones proposed by other studies (e.g. Labitzke, 2001, 2004a, Crooks and Gray, 2005, Salby and Callaghan, 2002), but further analysis of these differences is beyond the scope of this thesis.

The relationship between the QBO and  $T'$  seen in section 4.2 undergoes strong modifications as an effect of solar variability. Because of the limited fulfilment of the Markovian

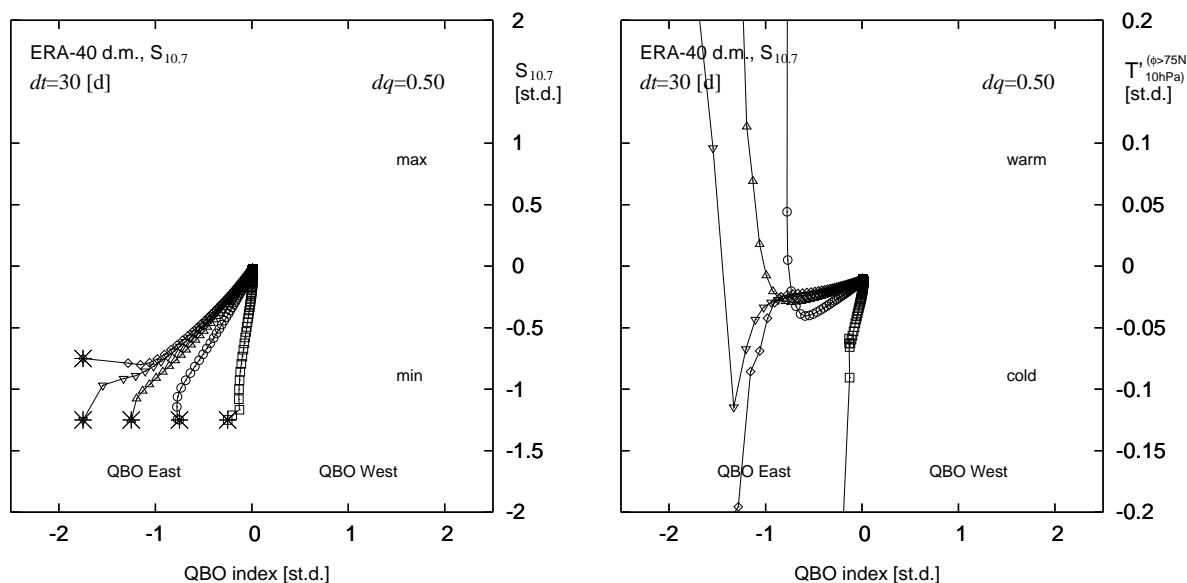


Figure 4.15: As Fig. 4.13 but initialized with low solar activity.

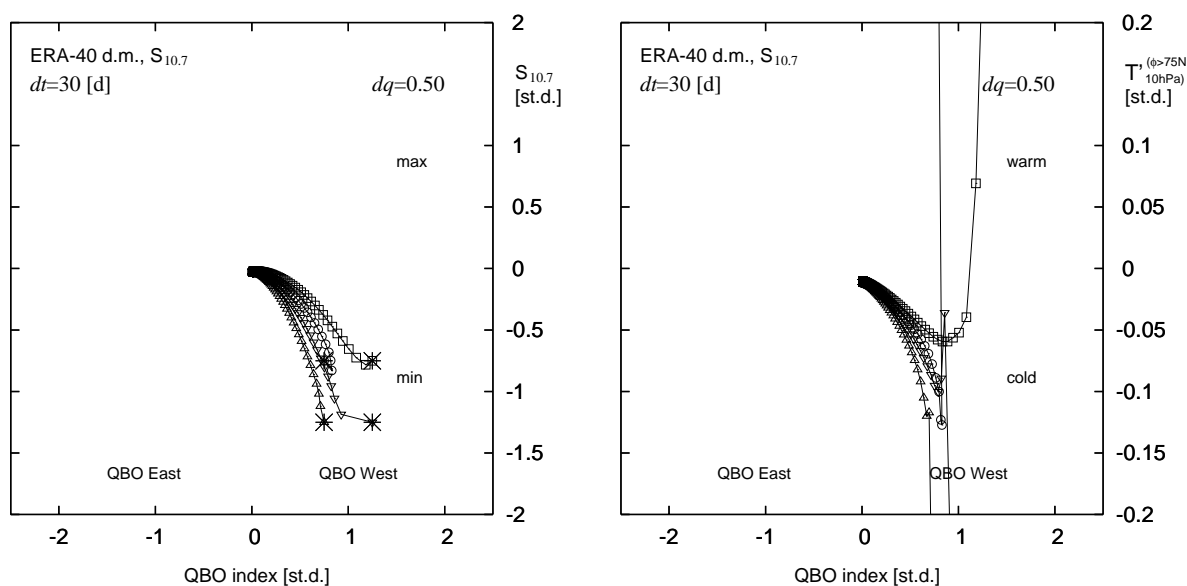


Figure 4.16: As Fig. 4.14 but initialized with low solar activity.

assumption, these interpretations are to some extent speculative. However, there is a strong indication of an influence of the 11-year SC and of the QBO on  $T'$  which is well established in the whole phase-space.

A final note is devoted to the master equation, which has proven in section 4.2 and in the current section 4.3 to be a very powerful tool. The evolution of the PDF shown in chapter 3 and/or in Fig. 4.6, and that of the mean trajectories seen in Fig. 4.7 and Fig. 4.13-4.16 describe quite complex behaviours. Such a rich description of dynamics in phase space may not be expected, for instance, from a regression model, which is commonly used in climate research (e.g. Baldwin et al., 2003). The linear regression model for a three-component

system is

$$\begin{aligned} q_1(t_n+dt) &= (1 + \alpha_1) q_1(t_n) + \beta_{12} q_2(t_n) + \beta_{13} q_3(t_n) + \zeta_1 \\ q_2(t_n+dt) &= \beta_{21} q_1(t_n) + (1 + \alpha_2) q_2(t_n) + \beta_{23} q_3(t_n) + \zeta_2 \\ q_3(t_n+dt) &= \beta_{31} q_1(t_n) + \beta_{32} q_2(t_n) + (1 + \alpha_3) q_3(t_n) + \zeta_3 \end{aligned} \quad (4.1)$$

where  $q_v$  are the variables, the constant coefficients  $\alpha_v$ ,  $\beta_{vw}$  can be estimated from data (e.g. von Storch and Zwiers, 1999, Egger, 2001), and  $\zeta_v$  are noise terms. Regression models were compared, for instance, by Egger (2001) to master equations and by Egger and Jönsson (2002) to the Fokker-Plank equation. The eigenvalues of a regression model describe the relations between the variables. Two cases are possible with three dimensions:

- the three eigenvalues are all real and represent damping rates towards the mean climate state, or
- there is a real eigenvalue describing a damping rate and two complex conjugate ones describing a damped rotation in phase space.

The features described by the master equation, including the various trajectory evolutions showing at times sharp bends, cannot be obtained neither through a linear regression model nor by interpreting correlation functions. Moreover, master equations are more general than Fokker-Plank equations. For a comparison between Fokker-Plank and master equations refer to Zwanzig (2001) and Egger (2002).

## 4.4 The Arctic Oscillation in the stratosphere and in the troposphere

In recent years there has been increasing evidence that the stratosphere plays an active role in the tropospheric circulations (e.g. Haynes, 2005), and that stratospheric climate variations may directly affect surface climate (e.g. Baldwin and Dunkerton, 1999, Thompson et al., 2002). Moreover, these links may help improve the skill of extended-range tropospheric forecasts (Thompson et al., 2002, Baldwin et al., 2003). Stratospheric modes of oscillation are indeed relevant for the troposphere (recall also the discussion in section 1.2). The troposphere in turn also influences the stratosphere. Planetary waves propagate from the troposphere into the stratosphere influencing the residual mean circulation through wave mean-flow interaction (e.g. Hauck and Wirth, 2001, Haynes, 2005). The question whether the stratosphere or the troposphere plays the greater role is debated (e.g. Haynes et al., 1996, Egger, 1996, Haynes et al., 1996, Wallace, 2000, Haynes, 2005). In Baldwin et al. (2003) it is pointed out that upward propagating waves lead to downward phase propagation of wind anomalies. Black (2002) and Ambaum and Hoskins (2002) introduced potential-vorticity considerations and suggested mechanisms of interaction between annular mode anomalies in the stratosphere and in the troposphere. Wallace (2000) argued the use of two paradigms, the NAO and the NAM, envisioning different phenomena and roles, and suggested rules to come to a consensus on which paradigm be more appropriate.

In this section, a master equation will be derived from a time series of the AO indices at stratospheric and tropospheric pressure levels defined in section 4.1. These indices were

defined as the normalised and de-seasonalised zonally averaged geopotential height difference between  $45^\circ\text{N}$  and  $67.5^\circ\text{N}$  (i.e. the geostrophic zonal wind in this latitude channel) at 10 ( $A_{10}$ ), 100 ( $A_{100}$ ), and 850 ( $A_{850}$ ) hPa. A similar approach was first used by Lorenz (1951), who defined  $U_{55}$  as the zonally averaged sea-level pressure difference between  $45^\circ\text{N}$  and  $65^\circ\text{N}$ <sup>24</sup>. Some recent studies referred to the AO/NAM as the leading Empirical Orthogonal Function (EOF) of slowly varying, wintertime, hemispheric geopotential at various pressure levels (e.g. Baldwin and Dunkerton, 1999, Baldwin et al., 2003). Wallace (2000) compared several NAM and NAO indices, and  $U_{55}$  was highly correlated with the principal component of the first EOF of ground pressure in the northern hemisphere ( $r > 0.9$ ). Finally, it should be remembered that in this thesis we consider de-seasonalised time series for the complete period covered by ERA-40.

The time series of the variable set ( $A_{10}$ ,  $A_{100}$ ,  $A_{850}$ ) obtained from ERA-40 daily means is shown in Fig. 4.17, 4.18, and 4.19. The first few lines of the times series of these three variables are reported for completeness in Table 4.3, where the first two lines represent the mean and standard deviation of the variables. A first glance at Fig. 4.17-4.19 clearly shows

# - mean =	3178.52550	3326.75105	939.06387
# / stdev=	5644.61584	1683.50820	514.34517
1957-09-01	0.15469	-0.31789	-0.43319
1957-09-02	0.13626	-0.29332	-0.55702
1957-09-03	0.13148	-0.18553	-0.83382

Table 4.3: The first few lines of a time series for the variable set ( $A_{10}$ ,  $A_{100}$ ,  $A_{850}$ ) obtained from the ERA-40 daily means as explained in section 4.1. The first two lines, which don't yet belong to the time series, represent the mean and the standard deviation of  $A_{10}$ ,  $A_{100}$ , and  $A_{850}$  in [ $m^2 s^{-2}$ ].

that time scales in the middle stratosphere ( $A_{10}$ ) are bigger than in the lower stratosphere ( $A_{100}$ ), which are again bigger than those in the lower troposphere ( $A_{850}$ , see also Fig. 4.21). Moreover, there are several cases where persistent AO anomalies are seen at the different pressure levels, but it is hard to identify whether these anomalies propagate preferentially in one direction. Furthermore, there is increasing noise at the lower heights (i.e. at the higher pressure levels). Already at 100 hPa it is hard to identify which parts of the anomalies come from the different seasons. Finally, the time series of  $A_{10}$  partly resembles that of  $T'$  in Fig. 4.3, the two indices being correlated with  $r = -0.6$ .

At this stage, a choice has to be made on how to partition the phase space spanned by the three variables  $A_{10}$ ,  $A_{100}$ , and  $A_{850}$ . A grid size  $dq = 0.50$ , as was used in section 4.2 and section 4.3, would lead here to quite high noise-to-signal ratios while estimating the transition coefficients ( $0.4 < R_w < 0.9$ , depending on the time step  $dt$ ). Therefore, a grid size  $dq = 1.00$  is chosen in order to increase the statistical significance of the estimates. The next choice to be made regards the time step  $dt$ . The time scales seen in the time series and in the correlation functions (Fig. 4.21) suggest that processes have no memory beyond a few weeks time, which would suggest  $dt = 7$  [days]. However, in order to avoid a rapid smearing of the PDF, we show results obtained with a time step  $dt = 14$ . The noise-to-signal ratio  $R_w$  obtained under these conditions is an acceptable  $R_w = 0.32$ .

Fig. 4.20 (a) shows a projection on the ( $A_{10}$ ,  $A_{100}$ ) phase plane of this time series' distribution of observed states. Figure 4.20 (b) shows isolines of this time series' observed state

<sup>24</sup>As the horizontal resolution of the ERA-40 daily means used here is  $7.5^\circ$ ,  $45^\circ\text{N}$  and  $67.5^\circ\text{N}$  are taken

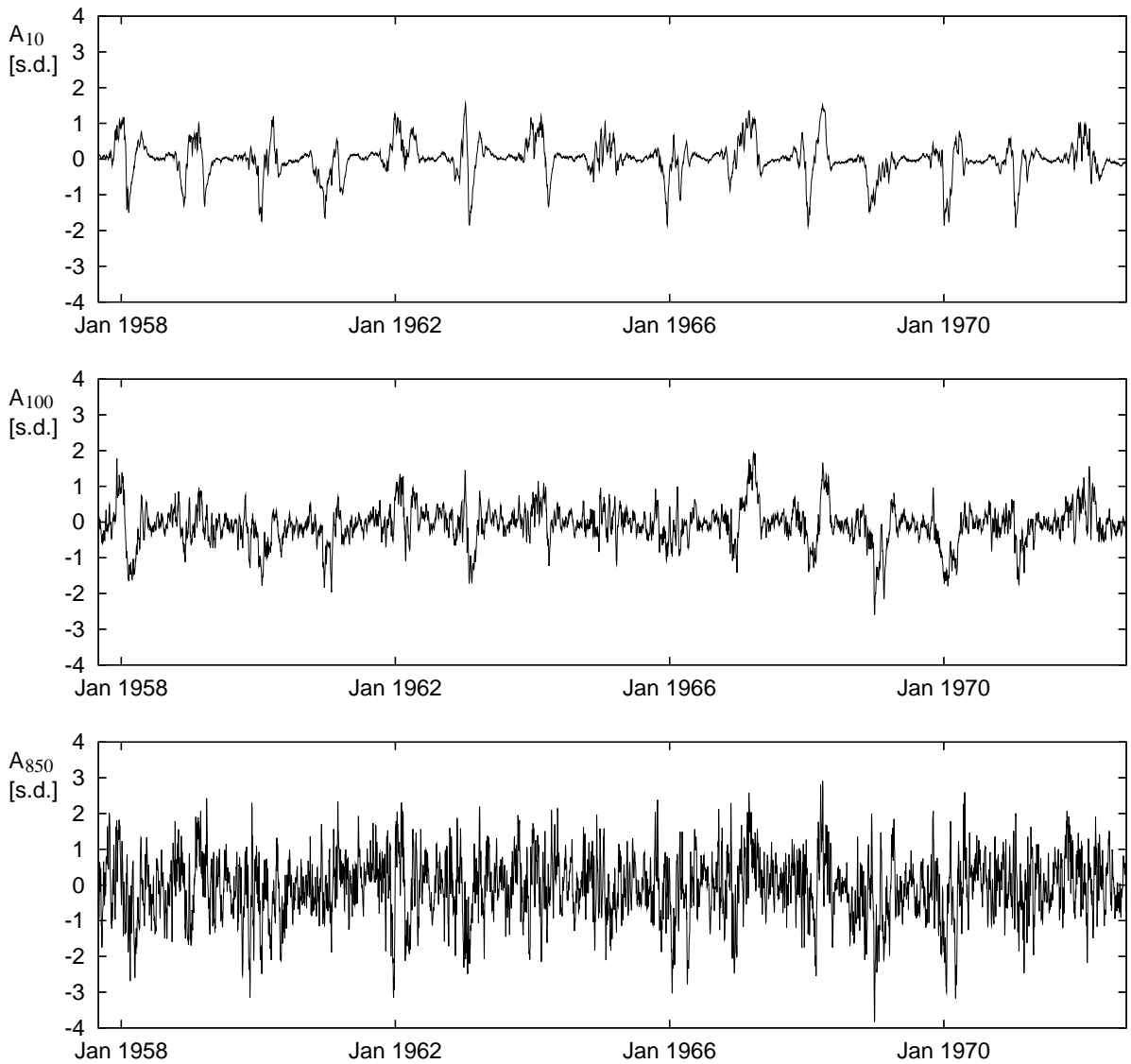


Figure 4.17: Time series of the normalised and de-seasonalised AO index in 10, 100, and 850 hPa, obtained from the ERA-40 daily means for the period September 1957 - August 1972. Top panel:  $A_{10}$ ; middle panel:  $A_{100}$ ; bottom panel:  $A_{850}$ . Positive [negative] anomalies indicate westerly [easterly] anomalies of the geostrophic zonal wind in middle latitudes in the northern hemisphere.

density  $\rho$  (integrated along the  $A_{850}$ -axis) for a phase space partition  $dq = 1.0$  [st. dev.]. The same applies to Fig. 4.20 (d) and (e) where values are projected on the  $(A_{10}, A_{850})$  phase plane, and consequently  $\rho$  is integrated along the  $A_{100}$ -axis. By observing Fig. 4.20 (a) and (d), one notes that the distribution of observed states shows a lower variability at lower pressure levels<sup>25</sup>. The long-term, stationary PDF prediction from a master equation with time step  $dt = 14$  [days] for an arbitrary initial condition is shown in Fig. 4.20 (c, f) and is almost not discernible from Fig. 4.20 (b, e). This is a first sign that the master equation correctly describes the observed system. Moreover, Fig. 4.20 indicates that  $A_{10}$  is more closely linked to  $A_{100}$  than to  $A_{850}$ . Analogous projections on the  $(A_{100}, A_{850})$  phase plane

as reference latitudes.

<sup>25</sup>At lower pressure levels, the seasonal cycle dominates the variability.



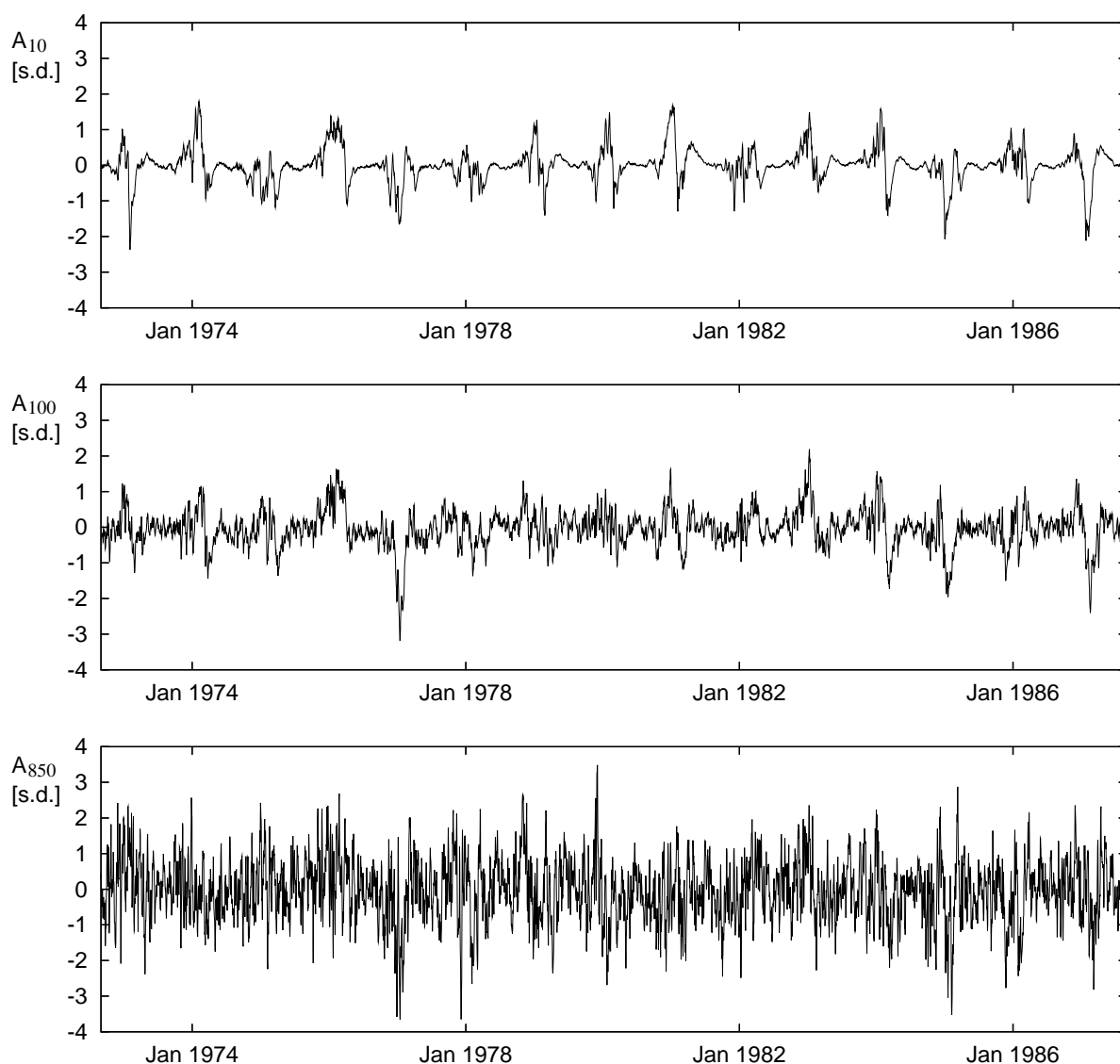


Figure 4.18: Time series of the normalised and de-seasonalised AO index obtained from the ERA-40 daily means for the period September 1972 - August 1987. Top panel:  $A_{10}$ ; middle panel:  $A_{100}$ ; bottom panel:  $A_{850}$ .

indicate that  $A_{100}$  is more strongly linked to  $A_{10}$  than it is to  $A_{850}$  (not shown).

Before discussing dynamics in phase space, let us consider Fig. 4.21, where correlation functions as seen in the data and as delivered by the master equation are shown as a function of time lag. The master equation captures the decay of the auto-correlation functions (top panels) and the fact that this decay is more pronounced at higher pressure levels like 850 hPa. This is a clear evidence that time scales in the middle stratosphere are longer than those in the lowermost stratosphere, which are in turn longer than those in the troposphere. The cross-correlation functions are also well reproduced by the master equation. All in all, the Markovian approximation is satisfactorily fulfilled. Furthermore, the higher values of  $r_{100,10}$  and  $r_{850,10}$  for negative lags are probably related to a downward propagation of AO anomalies. Only  $r_{850,100}$  shows relatively high values for positive lags. Ambaum and Hoskins (2002) discussed correlation functions for a NAO index and an index

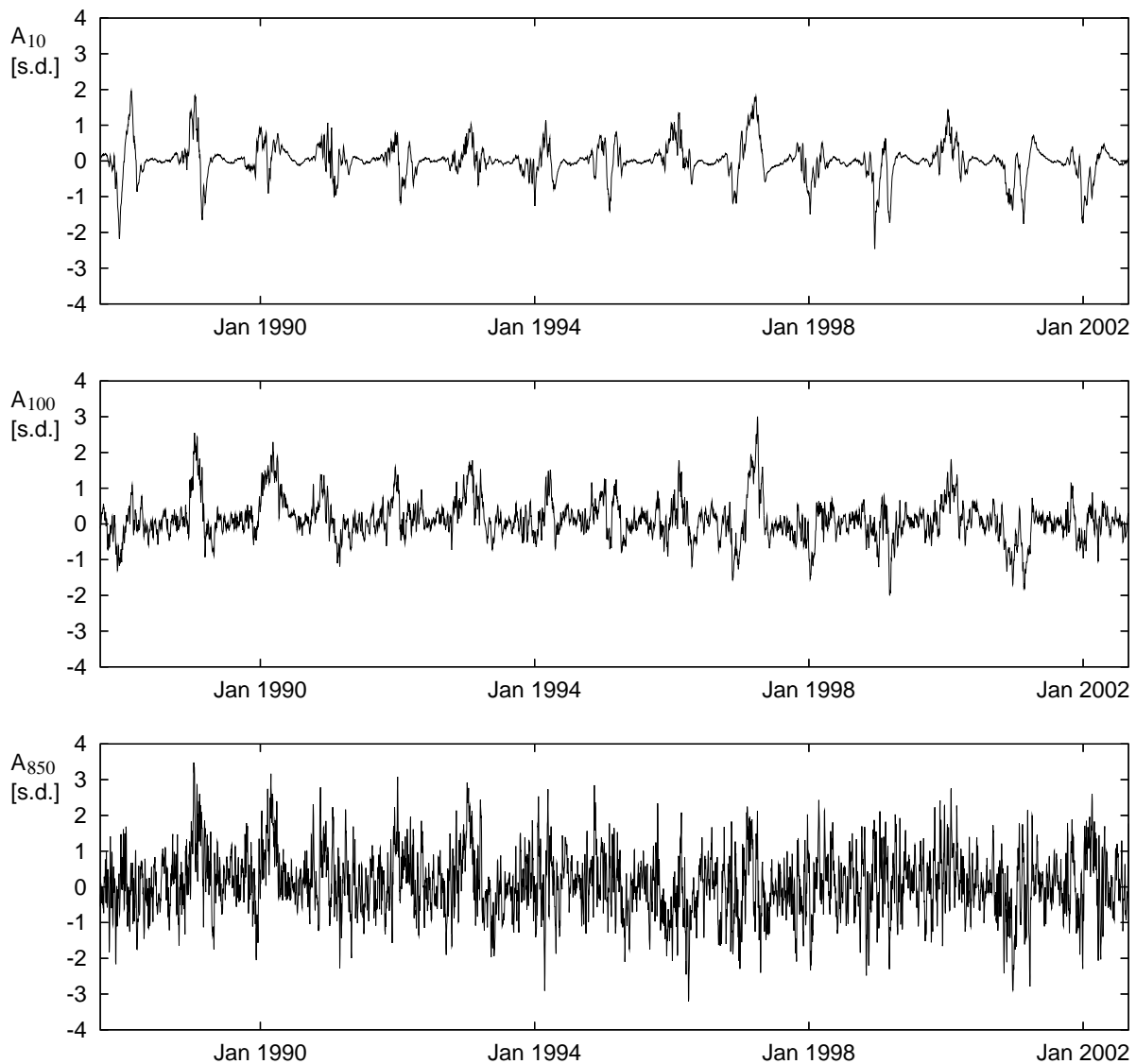


Figure 4.19: Time series of the normalised and de-seasonalised AO index obtained from the ERA-40 daily means for the period September 1987 - August 2002. Top panel:  $A_{10}$ ; middle panel:  $A_{100}$ ; bottom panel:  $A_{850}$ .

measuring the strength of the stratospheric polar vortex estimated from data for the winter months. They found a small peak in the cross-correlations for the NAO index leading the stratospheric index by 4 days. It is, however, difficult to identify clear time-scales on the base of the results in Fig. 4.21 as the correlations don't show other peaks besides lag zero.

To better evaluate dynamics in phase space we discuss PDF evolutions as predicted by the master equation. Master equation runs are started from all the grid boxes where the state vector  $(A_{10}, A_{100}, A_{850})$  resides 30 or more times within the time series; the transition coefficients for the first time step(s) will therefore have smaller confidence intervals. Projections on the three phase planes of the mean trajectories are shown in Fig. 4.22, 4.23, and 4.24 for phase planes  $(A_{10}, A_{100})$ ,  $(A_{10}, A_{850})$ , and  $(A_{100}, A_{850})$  respectively. Let us first consider Fig. 4.22, which shows projections of the mean trajectories from these master equation runs on the  $(A_{10}, A_{100})$  phase-plane. Use is made of different symbols which give

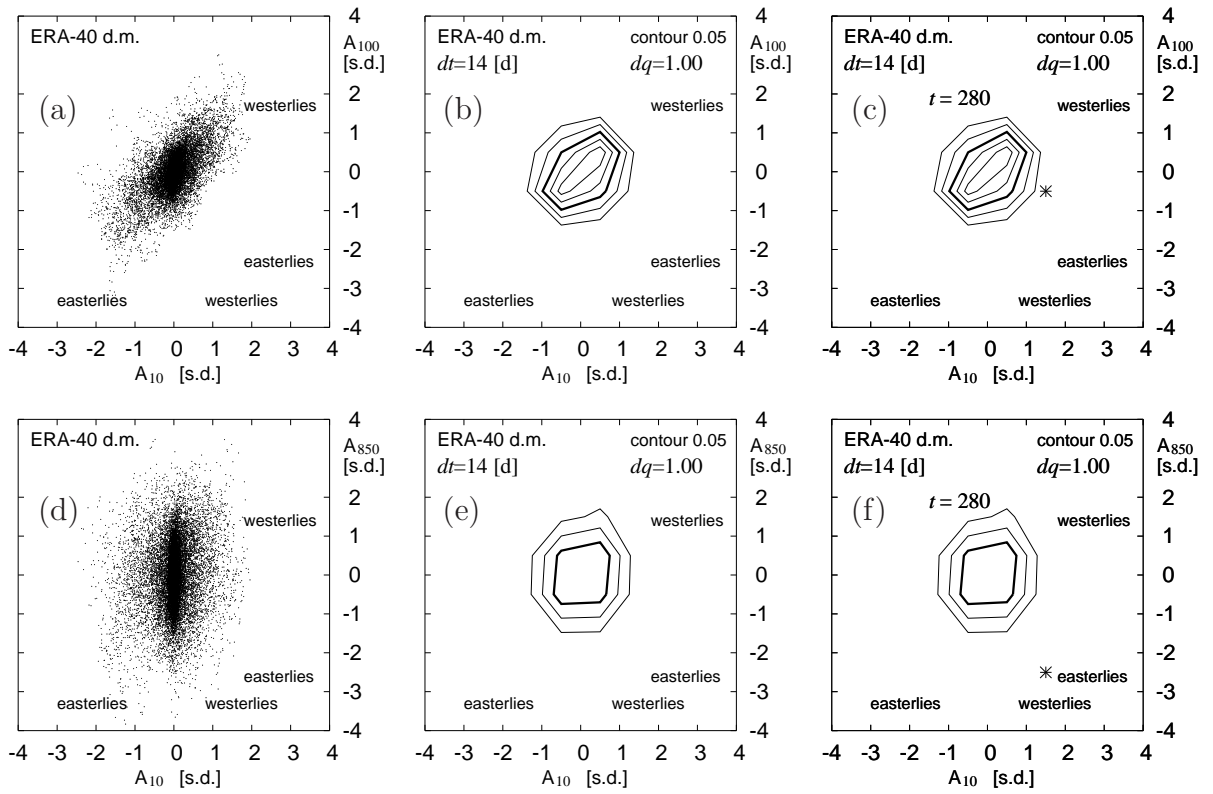


Figure 4.20: (a) [(d)] Daily mean values from the time series of the variable set  $(A_{10}, A_{100}, A_{850})$  are projected on the  $(A_{10}, A_{100})$  [ $(A_{10}, A_{850})$ ] phase plane. After discretising the phase space with  $dq = 1.0$  [st. dev.] and integrating along the normal axis, one obtains the observed state density  $\rho$  whose isolines are shown in (b) [(e)], where the time step  $dt = 14$  [days]. (c) [(f)] The stationary PDF (at  $t = 280$ ) obtained from a master equation for the time series of the three variables. The sharp initial condition is marked by a star  $*$ ; time step  $dt = 14$  [days]; grid size  $dq = 1.00$  [st. dev.]; the PDF has been integrated along the normal axis. The isoline contour interval in (b), (c), (e), and (f) is 0.05, and every third isoline is bold.

the mean position every 14 days in order to better recognise the trajectories. The right panels in Figs. 4.22-4.24 zoom in by a factor 10 towards the origin. If  $A_{10}$  and  $A_{100}$  were just slightly related to another, and showed approximately the same time scales, one would expect the trajectories in Fig. 4.22 to begin and continue along a straight line towards the mean climate state (i.e. the origin). As  $A_{10}$  has a longer memory than  $A_{100}$ , one would expect two curves recalling, say, parabolas with vertical axes, and vertexes at the origin. What we see, however, is a clear cyclonic rounding of the contracting motion. Moreover, in both panels, and even more clearly in the right one of Fig. 4.22, we note that trajectories starting close to the  $A_{10}$ -axis exhibit a strong cyclonic rounding, whereas trajectories starting close to the  $A_{100}$ -axis exhibit a less pronounced one. This means that anomalies of  $A_{10}$  clearly propagate downwards to 100 hPa with a time scale of 14 (or more) days, i.e. one (at times two) point(s) on the trajectories<sup>26</sup>. It should be reminded that we are dealing with a three-dimensional discretised phase space<sup>27</sup>, and that no runs are started in

<sup>26</sup>This time scale was obtained also from numerical runs with  $dt = 7$  [days] and  $dq = 0.5$  [st. dev.], which delivered similar correlation functions.

<sup>27</sup>For this reason, different trajectories appear to originate from the same point in a projected view. Moreover, interesting initial conditions on the axes are not possible as these are not the centres of any grid boxes.

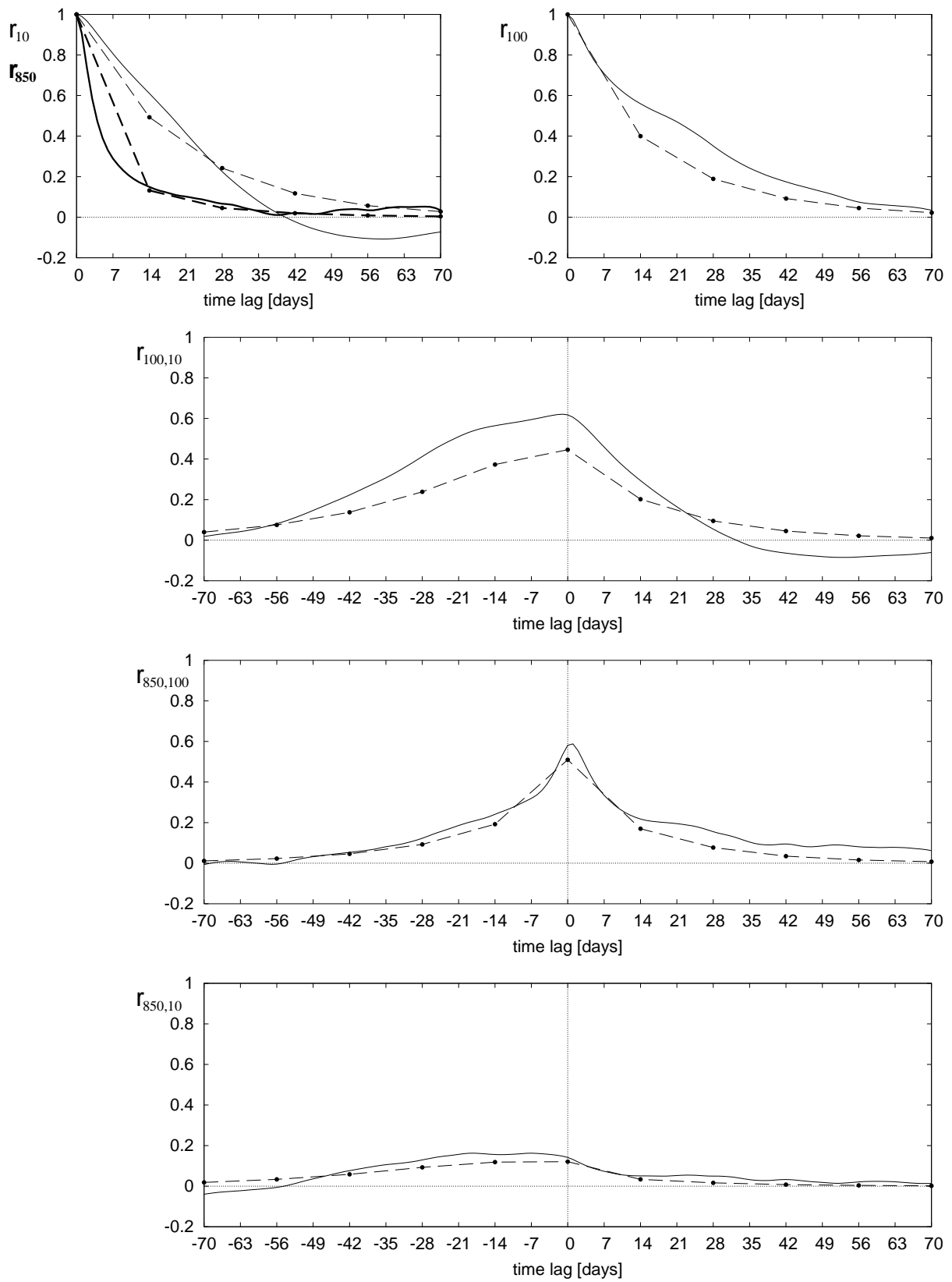


Figure 4.21: Correlation functions as observed (*solid*), and as reproduced by the master equation (*dashed*) derived as explained in the text from a time series of the variable set  $(A_{10}, A_{100}, A_{850})$ ; time step  $dt = 14$ ; phase space grid size:  $dq = 1.00$  [st. dev.]. The *bold* lines in the top left panel refer to values of  $r_{850}$ .

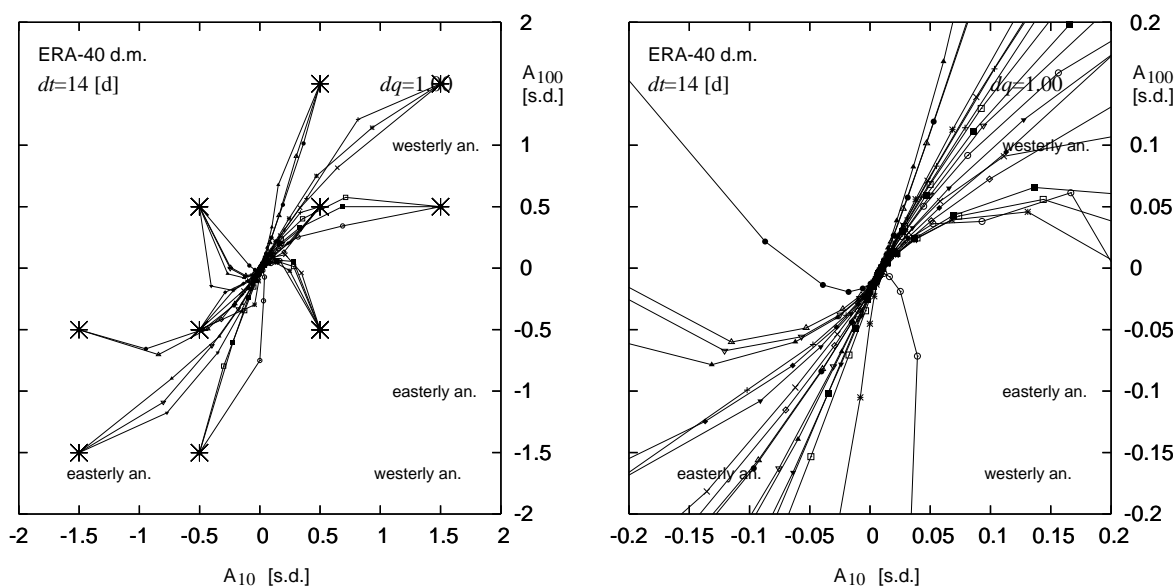


Figure 4.22: Selected mean trajectories delivered by a master equation for the variable set  $(A_{10}, A_{100}, A_{850})$ ; time step  $dt = 14$  [days]; grid size  $dq = 1.00$  [st. dev.]; the time series was obtained from ERA-40 as explained in section 4.1. The trajectories have been projected on the  $(A_{10}, A_{100})$  phase plane. The stars  $*$  mark the sharp initial conditions; the different symbols allow to distinguish between trajectories and give the mean position every 14 days.

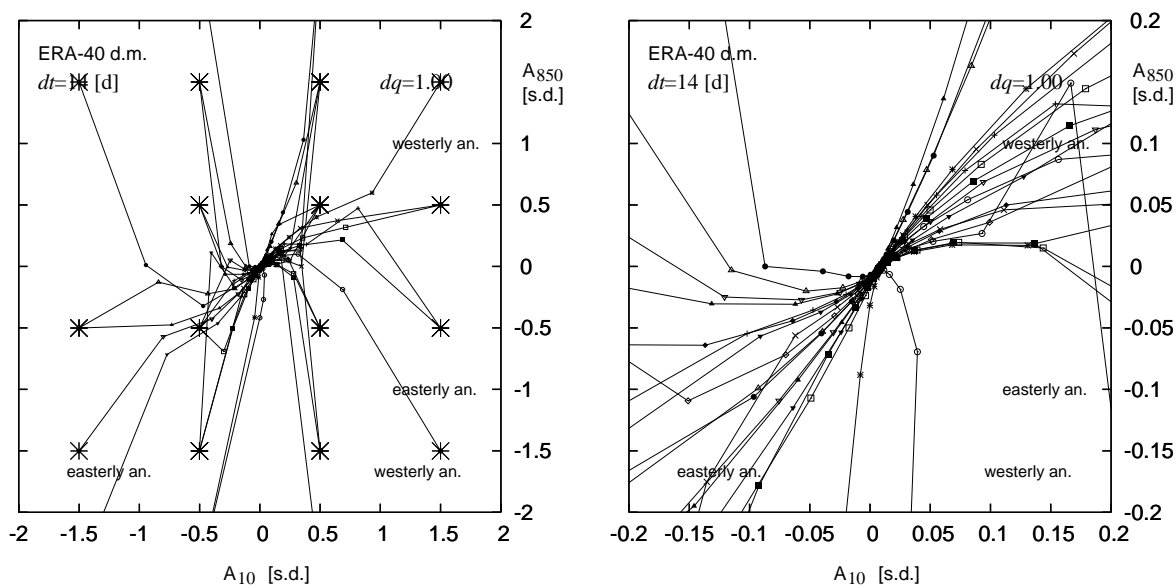


Figure 4.23: As in Fig. 4.22 but with trajectories projected on the  $(A_{10}, A_{850})$  phase plane.

grid boxes that have been seldom visited in the time series.

The same trajectories seen in the preceding paragraph are now projected onto the  $(A_{10}, A_{850})$  phase plane, which are shown in Fig. 4.23. Here the cyclonic rounding of trajectories is less pronounced than in the other projection, and can be better appreciated in the right panel of Fig. 4.23. This is probably an effect of the increasing distance between these layers, as 10 hPa correspond to a height above sea level of approximately 30 km, and 850 hPa to 1.5 km. The typical time scale for the propagation of an AO anomaly from

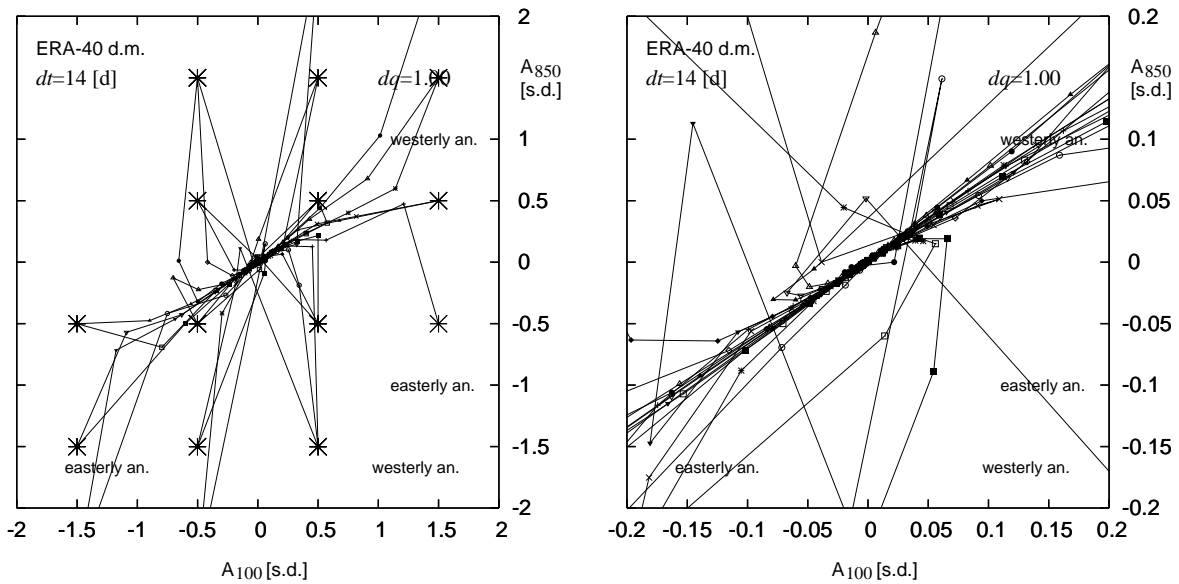


Figure 4.24: As in Fig. 4.22 but with trajectories projected on the  $(A_{100}, A_{850})$  phase plane.

10 hPa to 850 hPa is approximately 4 weeks<sup>28</sup> (two points in Fig. 4.23). As 850 hPa corresponds to the top of the planetary boundary layer over land in mid-latitudes, the rounding of the trajectories shown in Fig. 4.23 is a clear indication that anomalies of  $A_{10}$  propagate deep into the troposphere. These results are consistent with other studies (e.g. Baldwin and Dunkerton, 1999, Black, 2002, Baldwin et al., 2003). The time scales suggested so far by the master equation are higher than those suggested by other studies; for instance, Baldwin and Dunkerton (1999) suggested a time scale of about three weeks for the propagation of AO anomalies from 10 hPa to the surface. Longer time scales may improve the potential for weather prediction (e.g. Thompson et al., 2002, Baldwin et al., 2003).

These first results would somehow suggest that tropospheric AO anomalies hardly influence the stratosphere. Figure 4.24, where the well acquainted trajectories are projected on the  $(A_{100}, A_{850})$  phase plane, induces a further discussion. Dynamics appear more vivid when projected on this phase plane, suggesting a strong interaction. Moreover, a clear cyclonic behaviour does not appear. The impact of large tropospheric AO anomalies to the lower stratosphere is mostly confined to the first time step. In no way, however, do we see an anticyclonic rounding in the approach to the mean climate state of the trajectories. These results support suppositions that the troposphere affects the stratosphere on shorter time scales (e.g. Ambaum and Hoskins, 2002), and that dynamical mechanisms accounting for the flux, at higher frequencies, of waves from the troposphere to the stratosphere (Baldwin et al., 2003, Haynes, 2005) play an important role<sup>29</sup>.

All in all, the master equation again proves to be a very powerful tool and provides a contribution to a puzzling issue by clearly displaying dynamics in phase space. The master equation shows that anomalies of the AO propagate from the middle stratosphere

<sup>28</sup>Again, this time scale is also confirmed by numerical runs with  $dt = 7$  [days] and  $dq = 0.5$  [st. dev].

<sup>29</sup>It should be emphasised that downward propagating zonal wind anomalies do not mean that the lower layers of the atmosphere do not influence the upper ones. Upward propagating waves are, for instance, responsible for the downward propagating zonal wind anomalies observed within the QBO.

downward into the lower stratosphere and further deep into the troposphere, and suggests time scales for this phenomenon.

# Chapter 5

## Conclusions

Data-based master equations have been developed and their numerical properties have been studied. Master equations have been subsequently implemented to study various modes of stratospheric climate variability with focus on the northern hemisphere.

Correlation functions as estimated from data and as delivered by the master equation were compared in order to verify the fulfilment of the Markovian assumption inherent in the master equation. The significance problem was addressed in terms of the ratio between the confidence interval of the transition coefficients and the coefficients themselves.

A numerical study was conducted on the base of time series generated by running the three-component Lorenz convection model extended with white Gaussian noise. Thus time series of the desired characteristics could be generated easily. Moreover, by knowing the model equations it was possible to determine the evolution of an ensemble of initial conditions, which served as a reference to compare with the probability density function (PDF) forecasts delivered by the master equations.

An adequate grid size to discretise the phase space can be defined by carefully considering the analysed phenomena and the characteristics of the underlying time series. The grid size is a compromise between resolution in phase space on one side, and available data (i.e. statistical significance) and computer resources on the other. All in all, the grid size should not be too fine.

The strong impact of the time series length on the success of the master equation was assessed from various perspectives. The statistical significance of the transition coefficients improved with growing time series length. The PDFs predicted by master equations derived from time series of different lengths could be compared to the distribution of ensembles of initial conditions calculated by numerically integrating the model equations. It turned out that for a given phase space partition and time resolution, there is an optimal time series length beyond which the forecast skill hardly improves. A new approach considering the transition coefficient matrix  $\mathbf{T}$  as a multi-dimensional vector, showed that  $\mathbf{T}$  converges with time series length going to infinity. The numerical study showed that good PDF forecasts can be obtained already with moderate time series lengths.

The influence of time resolution on the master equation was assessed, assuming that the time resolution remains less than half of the period of the shortest time scale to be resolved. Surprisingly, master equations derived from time series with *coarser* time resolutions led to better forecast skills and delivered better correlation functions.

The dimension of the master equation also plays a crucial role. A master equation



derived for time series of the X and Z-components of the Lorenz model with stochastic forcing led to poor predictions and did not fulfil the Markovian assumption. An extension to further variables may therefore help improve the quality of predictions and reduce violations of the Markovian assumption.

The next part of this thesis was to derive master equations from time series of stratospheric climate indices to explore various modes of stratospheric climate variability. These master equations confirmed the results from the previous numerical study discussed in the preceding paragraphs. The stratosphere is chosen here since it shows various variability modes on the longer time scale, and some of these, like the quasi-biennial oscillation of equatorial zonal wind (QBO), are not correctly described by global circulation models; moreover, there is growing evidence that these modes are relevant for the surface climate. These climate indices were chosen to best represent various modes of stratospheric climate variability. The time series were derived from ERA-40 daily means and from daily observations of the solar radio flux at a wavelength of 10.7 cm. Each time series was normalised with the standard deviation of that particular variable; the mean and the seasonal cycle were subtracted. No filtering of the data took place. Averaging was done exclusively on the 6-hourly values in order to obtain daily means. On the basis of the results from the numerical study based on the Lorenz model with stochastic forcing, the transition coefficients were estimated from the daily time series on the basis of time steps of up to 30 days.

A three-dimensional master equation was first derived for a time series of indices of the QBO at 30 hPa (denoted by  $Q_I$ ) and 20 hPa in the equatorial stratosphere, and of the temperature anomaly index for the arctic middle stratosphere  $T'$ . No clear effect of the QBO on the temperature in the arctic stratosphere could be seen in the data. The relationship suggested by [Holton and Tan \(1980\)](#) on the base of monthly means for the winter months could therefore not be immediately confirmed. It should be reminded that we observed the distribution in the  $(Q_I, T')$  phase plane of daily mean data from the whole period covered by ERA-40, i.e. without a seasonal preference. However, the evolution of the PDF in the phase space of these variables showed that the QBO has an impact on the temperature anomaly of the middle stratosphere in the arctic region. It turned out that during the QBO East phase the arctic stratosphere is 2 K warmer than during the QBO West phase. A moderate cooling takes place during the transition to QBO West and during the West phase itself. The correlation functions of the climate variables derived by the master equation well reproduced the correlations observed in the data for surprisingly long time scales, confirming that the Markovian assumption held. This was not the case when a two-dimensional master equation was derived from a time series of  $Q_I$  and  $T'$ . However, the latter results for the two-dimensional master equation helped interpret the role of the 11-year solar cycle (SC) on climate in the stratosphere.

A master equation was then derived from a time series of the QBO index  $Q_I$ , of the index for temperature anomaly in the arctic middle stratosphere, and of the index for the solar flux at a wavelength of 10.7 cm. Solar variability affects the arctic stratosphere. The relationship between the QBO and the temperature in the arctic stratosphere discussed in the last paragraph is shifted towards warmer [colder] temperatures of +1 [-1] K during phases of high [low] solar activity. However, the correlation functions delivered by the master equation suggest that a five-dimensional master equation is needed to better capture the features of this interaction.

Finally, the dynamical coupling of the stratosphere with the troposphere was investi-

gated in terms of the Arctic Oscillation (AO). A master equation was derived from time series of AO indices in the middle and lower stratosphere and in the lower troposphere. In the stratosphere, the AO is linked to the temperature anomaly in the arctic region. In the troposphere, the AO is a mode of variability which has a considerable impact on surface weather and climate in the mid- and high latitudes. The master equation showed that AO anomalies tend to propagate downwards from the middle stratosphere to the lower stratosphere in approximately two weeks, and then deeply into the troposphere with a time scale of four weeks. The impact of large tropospheric AO anomalies is confined to the lower stratosphere.

This thesis has shown that the use of master equations may develop understanding of connections between various phenomena. The results presented here concerning the stratosphere increase the knowledge on stratospheric modes of climate variability and may help to improve climate and weather prediction. An improved representation of the stratospheric modes would have a high potential for climate and weather predictions in the troposphere. Moreover, the climate over mid- and high latitudes is extremely sensitive to the phase of the AO. Thus the potential for improving climate predictions is of great interest for Europe. The evaluation of modelled anthropogenic climate change may also profit from an improved understanding of stratospheric processes.

# Appendix A

## List of acronyms

AO	Arctic Oscillation
ECMWF	European Centre for Medium-range Weather Forecast
ENSO	el niño/southern oscillation
EOF	Empirical Orthogonal Function
ERA-15	Reanalysis project of the ECMWF for the period 1979-1993
ERA-40	Reanalysis project of the ECMWF for the period 1957-2002
EUV	extreme ultraviolet solar radiation
LMSF	Lorenz model extended with a stochastic forcing
NAM	northern annular mode
NAO	North Atlantic oscillation
PDF	probability density function
PV	potential vorticity
QBO	quasi-biennial oscillation of equatorial zonal wind in the stratosphere
SC	(11-year) solar cycle
UTC	universal coordinate time

# Appendix B

## List of symbols

$A_{10}$	Arctic Oscillation index for the middle stratosphere (10 hPa)
$A_{100}$	Arctic Oscillation index for the lowermost stratosphere (100 hPa)
$A_{850}$	Arctic Oscillation index for the lower troposphere (850 hPa)
$\alpha$	factor of the white Gaussian noise in the Lorenz model equations (3.1)
$B$	numbers of boxes as in (3.2)
$b$	parameter in the Lorenz model equations (3.1)
$C_{q_k q_l}(\tau)$	covariance function for time lag $\tau$ as in section 2.5
$c_{q_k q_l}(\tau)$	contribution to the covariance function for time lag $\tau$ from a single master equation run
$C_w$	convergence coefficient as in (3.4)
$c_i^{i'}$	90% confidence interval for the transition coefficient $T_i^{i'}$
$\mathcal{D}_i^{i'}$	transition rate from grid box ( $i$ ) to grid box ( $i'$ ) for a master equation in operator form
$\mathcal{D}$	transition rate matrix for a master equation in operator form
$dq$	phase space grid size
$dq_v$	grid size along the $v^{\text{th}}$ axis
$\Delta t$	length of a time series
$dt$	time step of a master equation (eventually time resolution of a time series)
$dt_{RK}$	time step for the Runge-Kutta method of the fourth order
$\delta_{ii'}$	Kronecker symbol: $\delta_{ii'} = 1$ if $i = i'$ , $\delta_{ii'} = 0$ if $i \neq i'$
$\delta(\tau)$	Dirac function of $\tau$
$f_i(t_n)$	probability density in grid box ( $i$ ) at time $t_n = t_0 + n dt$
$\mathbf{f}(t_n)$	probability density matrix at time $t_n$
$\bar{\mathbf{f}}$	equilibrium probability density distribution
$\phi$	latitude angle, $\phi = 0$ at the equator
$\mathbf{g}$	conditional probability matrix
$\gamma_j$	value delivered by a random number generator with Gaussian deviate at its $j^{\text{th}}$ numerical generation
$i, i'$	grid box indices
$\mathbf{i}$	grid box index in a multi-dimensional phase space
$i^{\text{max}}$	number of interval in a one-dimensional partition
$i_v$	grid box index along the $v^{\text{th}}$ axis
$\lambda_2$	second biggest eigenvalue of matrix $\mathbf{U}$

$M_i^{i'}$	number of transitions from grid box ( $i$ ) to grid box ( $i'$ ) observed in a time series
N	North
$N$	number of observations ( $N + 1$ , see <a href="#">subsection 2.5.1</a> )
$N_i$	number of events when an observation falls in grid box ( $i$ )
$n$	time index
$Q_I$	quasi-biennial oscillation index
$q, q_v$	variable in phase space
$q^{min} [q^{max}]$	minimum [maximum] value of variable $q$ in the partitioned phase space region
$\tilde{q}$	coordinate value of the centre of a grid box
$\hat{\mathbf{q}}$	mean position of a PDF cloud
$\hat{q}_v$	mean value of the $v^{\text{th}}$ variable of a PDF cloud
$R_w$	noise to signal ratio for a given estimate of transition matrix $\mathbf{T}$
$r$	parameter in the Lorenz model equations <a href="#">(3.1)</a>
$r$	correlation coefficient
$r_{q_k q_l}(\tau)$	correlation function for time lag $\tau$ as in <a href="#">(2.11)</a>
$\rho$	observed state density
$\rho_i(l)$	observed state density in grid box $i$ for a long time series as in <a href="#">(3.4)</a>
$S$	forecast skill as in <a href="#">(3.3)</a>
S	factor of the time series step as in <a href="#">section 4.1</a>
$s$	time series index as in <a href="#">section 4.1</a>
$S_{10.7}$	solar flux index
$\sigma$	standard deviation of a PDF cloud, parameter in the Lorenz model equations <a href="#">(3.1)</a>
$T_i^{i'}$	transition rate from grid box ( $i$ ) to grid box ( $i'$ )
$T_i^{i'}$	transition coefficient (probability) from grid box ( $i$ ) to grid box ( $i'$ )
$T_i^{i'}(l) [T_i^{i'}(s)]$	transition coefficient (probability) from grid box ( $i$ ) to grid box ( $i'$ ) for a long [short] time series as in <a href="#">(3.4)</a>
$\mathbf{T}$	transition matrix
$T'$	temperature anomaly index for the arctic middle stratosphere
$t_r$	time scale of the residence in an attractor wing
$t_w$	time scale of the evolution around the centre of an attractor wing
$\tau$	time lag
$\mathbf{U}$	transition matrix for a master equation in matrix form
$U_{e20}$	zonally averaged zonal wind in 20 hPa above the equator
$V$	number of variables
$v_{max}$	maximum value among all velocity components in phase space
X	X-component of the Lorenz model equations <a href="#">(3.1)</a>
$\xi$	white (uncorrelated) noise
$\xi_j$	value of $\xi$ at its $j^{\text{th}}$ numerical generation
Y	Y-component of the Lorenz model equations <a href="#">(3.1)</a>
Z	Z-component of the Lorenz model equations <a href="#">(3.1)</a>

# Bibliography

- Ambaum, M. H. P. and Hoskins, B. J. (2002). The NAO Troposphere-Stratosphere Connection. *Journal of Climate*, 15:1969–1978. [48](#), [51](#), [56](#)
- Andrews, D. G., Holton, J. R., and Leovy, C. B. (1987). *Middle Atmosphere Dynamics*. Academic Press. [33](#)
- Baldwin, M. P. and Dunkerton, T. J. (1999). Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research*, 104:20937–30946. [3](#), [31](#), [48](#), [49](#), [56](#)
- Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M. (2001). The Quasi-biennial oscillation. *Reviews of Geophysics*, 39(2):179–229. [30](#), [33](#), [36](#)
- Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., and O’Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, 301:636–640. [3](#), [36](#), [47](#), [48](#), [49](#), [56](#)
- Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, 32:12–40. [10](#)
- Black, R. X. (2002). Stratospheric Forcing of Surface Climate in the Arctic Oscillation. *Journal of Climate*, 15:268–277. [48](#), [56](#)
- Crommelin, D. T. (2004). Observed Nondiffusive Dynamics in Large-Scale Atmospheric Flow. *Journal of the Atmospheric Sciences*, 61:2384–2396. [2](#), [7](#), [8](#), [9](#), [11](#)
- Crooks, S. A. and Gray, L. J. (2005). Characterization of the 11-Year Solar Signal Using a Multiple Regression Analysis of the ERA-40 Dataset. *Journal of Climate*, 18:996–1015. [46](#)
- Ditlevsen, P. D. (1999). Observation of  $\alpha$ -stable noise induced millennial climate changes from an ice-core record. *Geophysical Research Letters*, 26:1441–1444. [1](#)
- Egger, J. (1996). Comments on ”on the ’downward control of extratropical diabatic circulations by eddy-induced mean zonal forces”. *Journal of the Atmospheric Sciences*, 53:2103–2107. [3](#), [48](#)
- Egger, J. (2001). Master equations for climatic parameter sets. *Climate Dynamics*, 17:169–177. [2](#), [8](#), [9](#), [10](#), [11](#), [24](#), [31](#), [48](#)

- Egger, J. (2002). Master Equations for Himalayan Valley Winds. *Stochastics and Dynamics*, 2(3):381–394. 2, 8, 10, 11, 48
- Egger, J. and Jönsson, T. (2002). Dynamic models for Icelandic meteorological data sets. *Tellus*, 54A:1–13. 2, 48
- Fraedrich, K. (1988). El Niño/Southern Oscillation Predictability. *Monthly Weather Review*, 116:1001–1012. 2
- Gambolati, G. (1994). *Lezioni di Metodi Numerici per Ingegneria e Scienze Applicate*. Edizioni Libreria Cortina, Padova. 16
- Gardiner, C. W. (1983). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer. 1, 20
- Gradišek, J., Siegert, S., Friedrich, R., and Grabec, I. (2000). Analysis of time series from stochastic processes. *Physical Review E*, 62(3):3146–3155. 1
- Gray, L. J., Crooks, S., Pascoe, C., Sparrow, S., and Palmer, M. (2004). Solar and QBO Influences on the Timing of Stratospheric Sudden Warmings. *Journal of the Atmospheric Sciences*, 61:2777–2796. 3, 41
- Hauck, C. and Wirth, V. (2001). Diagnosing the impact of stratospheric planetary wave breaking in a linear model. *Journal of the Atmospheric Sciences*, 58:1357–1370. 48
- Haynes, P. (2005). Stratospheric Dynamics. *Annual Review of Fluid Mechanics*, 37:263–293. 48, 56
- Haynes, P. H., Marks, C. J., McIntyre, M. E., Shepherd, T. G., and Shine, K. P. (1991). On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *Journal of the Atmospheric Sciences*, 48:651–678. 3
- Haynes, P. H., McIntyre, M. E., and Shepherd, T. G. (1996). Reply to the comments on "on the 'downward control of extratropical diabatic circulations by eddy-induced mean zonal forces". *Journal of the Atmospheric Sciences*, 53:2103–2107. 3, 48
- Holton, J. R. (1992). *An Introduction to Dynamic Meteorology*. Academic Press, third edition. 33
- Holton, J. R. and Tan, H. C. (1980). The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *Journal of the Atmospheric Sciences*, 37:2200–2208. 3, 31, 33, 36, 39, 59
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. (2001). *Climate Change 2001: The Scientific Basis*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK. 3, 4
- Kaplan, D. and Glass, L. (1995). *Understanding Nonlinear Dynamics*. Springer. 13, 14, 27, 44
- Kodera, K. and Kuroda, Y. (2002). Dynamical response to the solar cycle. *Journal of Geophysical Research*, 107. Doi:10.1029/2002JD002224. 3, 41

- Labitzke, K. (1987). Sunspots, the QBO, and the Stratospheric Temperature in the North Polar Region. *Geophysical Research Letters*, 14(5):535–537. 30, 31
- Labitzke, K. (2001). The global signal of the 11-year sunspot cycle in the stratosphere: Differences between solar maxima and minima. *Meteorologische Zeitschrift*, 10(2):83–90. 30, 41, 45, 46
- Labitzke, K. G. (1998). *Die Stratosphäre, Phänomene, Geschichte, Relevanz*. Springer-Verlag, Berlin Heidelberg. 3, 4, 30, 41
- Labitzke, K. G. (2004a). On the signal of the 11-year sunspot cycle in the stratosphere and its modulation by the quasi-biennial oscillation. *Journal of Atmospheric and Solar-Terrestrial Physics*, 66:1151–1157. 3, 31, 41, 46
- Labitzke, K. G. (2004b). On the signal of the 11-year sunspot cycle in the stratosphere over the Antarctic and its modulation by the Quasi-Biennial Oscillation (QBO). *Meteorologische Zeitschrift*, 13(4):263–270. 3
- Labitzke, K. G. and van Loon, H. (1999). *The Stratosphere, Phenomena, History, and Relevance*. Springer. 3, 4, 41
- Levy, P. (1948). *Processus Stochastiques et Mouvement Brownian*. Gauthier - Villars, Paris. 14
- Lorenz, E. N. (1951). Seasonal and irregular variations of the northern hemisphere sea-level pressure profile. *Journal of Meteorology*, 8:52–59. 49
- Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20:130–141. 10, 13, 14
- Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T. (1996). The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122:73–119. 14
- Nicolis, C. (1990). Chaotic dynamics, Markov processes and climate predictability. *Tellus*, 42A:401–412. 2, 10
- Nicolis, C., Ebeling, W., and Baraldi, C. (1997). Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. *Tellus*, 49A:108–118. 2, 8, 10
- Palmer, T. N. (1993). Extended Range Atmospheric Prediction and the Lorenz Model. *Bulletin of the American Meteorological Society*, 74:49–65. 14, 15, 19
- Pasmanter, R. A. and Timmermann, A. (2002). Cyclic Markov chains with an application to an intermediate ENSO model. *Nonlinear Processes in Geophysics*, 9:1–14. 2, 7, 8, 9, 10, 28, 31
- Pichler, H. (1997). *Dynamik der Atmosphäre*. Spektrum Akademischer Verlag, Mannheim, Germany, third edition. 14, 15



- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1999). *Numerical Recipes in Fortran 77. The Art of Scientific Computing. Volume 1 of Fortran Numerical Recipes*. Cambridge University Press, second edition. 14
- Qian, B., Corte-Real, J., and Xu, H. (2000). Is the North Atlantic Oscillation the most important atmospheric pattern for precipitation over Europe? *Journal of Geophysical Research*, 105:11901–11910. 4
- Rind, D. (2002). The sun’s role in climate variations. *Science*, 296:673–677. 4, 41
- Salby, M. and Callaghan, P. (2000). Connection between the Solar Cycle and the QBO: The Missing Link. *Journal of Climate*, 13:328–338. 3, 41
- Salby, M. and Callaghan, P. (2002). Evidence of the Solar Cycle in the General Circulation of the Stratosphere. *Journal of Climate*, 17:34–46. 46
- Schönwiese, C. D. (1985). *Praktische Statistik für Meteorologen und Geowissenschaftler*. Gebrüder Borntraeger, Berlin - Stuttgart, second edition. 11, 36
- Siegert, S., Friedrich, R., and Peinke, J. (1998). Analysis of data sets of stochastic systems. *Physics Letters A*, 243:275–280. 1
- Sparrow, C. (1982). *The Lorenz equations: bifurcations, chaos, and strange attractors*. Springer. 14
- Spekat, A., Heller-Schulze, B., and Lutz, M. (1983). Über Großwetter und Markov-Ketten. *Meteorologische Rundschau*, 36:243–248. 2, 7, 8, 9, 10
- Stolarski, R. S., Bloomfield, P., and McPeters, R. D. (1991). Total Ozone Trends deduced from Nimbus 7 TOMS data. *Geophysical Research Letters*, 18:1015–1018. 3, 33
- Thompson, D. W. J., Baldwin, M. P., and Wallace, J. M. (2002). Stratospheric Connection to Northern Hemisphere Wintertime Weather: Implications for Predictions. *Journal of Climate*, 15:1421–1428. 3, 4, 39, 48, 56
- Thompson, D. W. J. and Wallace, J. M. (2001). Regional Climate Impacts of the Northern Hemisphere Annular Mode. *Science*, 293:85–89. 3
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., and collaborators (2005). The ERA-40 Reanalysis. *Quarterly Journal of the Royal Meteorological Society*. Accepted. 4, 29
- van Loon, H. and Labitzke, K. (1998). The Global Range of the Stratospheric Decadal Wave. Part I: Its Association with the Sunspot Cycle in Summer and in the Annual Mean, and with the Troposphere. *Journal of Climate*, 11:1529–1537. 3, 41
- Vautard, R., Mo, K. C., and Ghil, M. (1990). Statistical Significance Test for Transition Matrices of Atmospheric Markov Chains. *Journal of the Atmospheric Sciences*, 47(15):1926–1931. 8
- von Storch, H. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press. 1, 11, 15, 48

- Wallace, J. M. (2000). North Atlantic Oscillation/annular mode: Two paradigms-one phenomenon. *Quarterly Journal of the Royal Meteorological Society*, 126A:791–805. [3](#), [48](#), [49](#)
- Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics*. Oxford University Press, New York. [1](#), [5](#), [6](#), [7](#), [48](#)

# Acknowledgements

This doctorate in meteorology is the fulfilment of a dream I would not have dared think of during school time. Prof. J. Egger has given me a great chance I will always be grateful for. Prof. J. Egger has allowed freedom of action while keeping an eye to be sure that the requirements for this scientific work be fulfilled within a reasonable time.

I wish to thank Prof. K. Elsässer, Dr. W. Ulrich, PD Dr. M. Dameris, Dr. M. Hornsteiner, Dr. R. Goler and many colleagues at the Meteorological Institute of the University of Munich who spent time in fruitful discussions. I heartily thank my fiancée Johanna, my family, Liviana, Franco, and Matteo, Agnes, Stefano, P. Dominik, and all those who kept up my morale during the most challenging phases of this thesis.

Financial support was received from the German Ministry of Education and Research (*Bundesministerium für Bildung und Forschung*) and the German Space Agency (*Deutsches Zentrum für Luft- und Raumfahrt*) as part of the German Climate Research Program (*Deutsches Klimaforschungsprogramm*). This doctoral thesis was part of the project CLIMESTO (diagnosis of stratospheric CLimate variability with MEchanistic and STOchastic models, Contract No. 01 LD 0033), a joint research program with the Institute for atmospheric Physics (IPA) at the University of Mainz, Germany. My sincere thanks are extended to the European Centre for Medium-Range Weather Forecasts for developing the ERA-40 Reanalysis project, and the IPA for help in acquiring the used data set. I am also grateful to the staff of the Dominion Radio Astrophysical Observatory, Canada, for making the time series of the solar flux available.

# Curriculum Vitae of the author

## Personal information

First name: Mauro  
Surname: Dall'Amico  
Nationality: Italian  
Sex: male  
Birth date and place: 1973-10-30, Thiene, Venetia, Italy  
Current address: Auenstr. 8, D-86938 Schondorf a. A., Bavaria, Germany  
Home address: Via Braglio 27, I-36030 Mossòn di Caltrano, Venetia, Italy  
Current affiliation: Meteorologisches Institut der Universität München  
Theresienstr. 37, D-80333 Munich, Germany  
Email: [mauro@meteo.physik.uni-muenchen.de](mailto:mauro@meteo.physik.uni-muenchen.de)

## Education

### *Secondary school:*

Institution, years: Ponoka Composite High School, Ponoka, Canada, Sep 1990 to Jul 1991  
Certificate: Alberta High School Diploma Exams in Social studies, Mathematics, Biology, July 1991 (*Exchange year with AFS*)

Institution, years: Liceo Ginnasio Statale F. Corradini, Thiene, Sep 1987 to Jul 1992  
Certificate: **Maturità Scientifica**, July 1992 (*Italian grade 13 secondary school diploma with scientific major*)

### *University studies:*

Institution, years: Università degli studi di Padova, Padua, Italy, Oct 1992 to Apr 1998  
Certificate: **Laurea in Ingegneria Meccanica**, April 1998, 110 points out of 110 (*Former Italian Doctor's diploma, Mechanical Engineering, regular period of study: 11 semesters after a grade 13 secondary school diploma*)

Thesis, supervisor, dept.: Studio di fattibilità di una vela solare a nastro, Prof. F. Angrilli, CISAS (*Feasibility Study of a tape-shaped Solar Sail, Interdepartmental Centre for Studies and Activities for Space, <http://cisas.unipd.it>*)

Institution, years: Ludwig-Maximilians-Universität München, Munich, Germany, since Jan 2002; Doctoral candidate in Meteorology

Thesis, supervisor, dept.: Data-based Master Equations for the Stratosphere, Prof. J. Egger, MIM (*Meteorological Institute of the University of Munich, [www.meteo.physik.uni-muenchen.de](http://www.meteo.physik.uni-muenchen.de)*)

## Work experience

- Employer, years: Dr. R. Frau, Thiene, Summer 1998; Applied Thermodynamics, Engineer
- Employer, years: Aereonautica Militare, Vicenza, Italy, Aug 1998 to Jun 1999; AGI (*Italian Air Force, military service, assistant for disabled veterans*)
- Employer, years: Tecnomare SpA, Venice, Italy, Summer 1999; Oceanography, Engineer
- Employer, years: Coca-Cola Greater Europe, Brussels, Belgium, Jan 2000 to Dec 2001; Food and Beverage Industry, Laboratory Engineer; laboratory testing, quality assurance, supplier approval procedures, development of new testing methods
- Employer, years: Meteorologisches Institut der Universität München, Munich, Jan 2002 to Feb 2005; Climate Research, Scientist; employed in the project KLIMESTO within the German Climate Research Program ([www.deklim.de](http://www.deklim.de))
- Employer, years: Institut für Italienische Philologie der Universität München, since Apr 2003; Lecturer of Italian language

## Language skills

- Italian native language
- German fluent
- English fluent (TOEFL, 663 points out of 677, June 1998)
- French orally fluent, written: working knowledge

## Computer skills

- Operating systems: Linux, Unix, Ms-Windows, DOS
- Programming languages: Fortran, scripting languages, C++ (basics), Visual Basic (basics)
- Meteorological Data: GrADS
- Software applications: Ms-Office,  $\text{\LaTeX}$ , graphics packages, Matlab (basics), Autocad (basics)

## Personal interests

- Languages and Culture, Politics, Nature, Astronomy;
- Alpine skiing, Ski touring, Swimming, Skating, Mountaineering.





---

# **Data-based Master Equations for the Stratosphere**

**Mauro Dall'Amico**

---

This Ph.D. thesis is available on the physics faculty dissertations web page of the Munich University Library:

<http://edoc.ub.uni-muenchen.de/view/subjects/fak17.html>