A POD-Galerkin approach to the atmospheric dynamics of Mars

Oscar Martínez-Alvarado

Jesus College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2007

A POD-Galerkin approach to the atmospheric dynamics of Mars

Oscar Martínez-Alvarado

Jesus College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2007

The observation of less chaoticity and enhanced interannual periodicity of transient waves in the Martian atmosphere in comparison with that of the Earth suggests the hypothesis of a low-dimensional underlying atmospheric attractor. Grounded on this hypothesis, two questions can be asked: is there a small set of atmospheric modes, measured and classified by a suitable norm, capable of describing the atmosphere of Mars? If this set exists, are those atmospheric modes able to reproduce the dynamical behaviour of the atmosphere of Mars? The answer to these questions, constituting the central focus of this thesis, has led to the first application of POD-Galerkin methods to a state-of-the-art Mars general circulation model.

The proper orthogonal decomposition (POD) as a method for extracting coherent structures, called empirical orthogonal functions (EOFs), provided a means to answer the first question in the positive. An important amount of atmospheric total energy (TE) was found to be concentrated in a few EOFs (e.g., 90% TE in 20 EOFs). The most energetic EOFs were identified with atmospheric motions such as thermal tides and transient waves.

The Galerkin projection of the hydrostatic primitive equations onto the span of the EOFs provided a systematic method to establish physically plausible interactions between the most energetic EOFs. These interactions were complemented with closure schemes representing interactions with unresolved modes. This requirement proved to be essential in order to obtain bounded behaviour.

In the diagnostic analysis, represented by the POD alone, increasing the number of EOFs directly leads to a better approximation of the atmospheric state. In contrast, the dynamic reconstruction of the atmospheric evolution does not depend only on the number of included EOFs. Other important factors to obtain realistic evolution are the inclusion of every mode involved in the description of a particular kind of motion (diurnal tide, semidiurnal tide or transients) and the retention of higher order modes that may interact strongly with the modes of interest. Once these conditions are satisfied the behaviour of the reduced models is greatly improved. Implications of these findings for future work are discussed.

Por el enorme soporte
que desde el comienzo
me han brindado,

A MIS PADRES

Acknowledgements

The list of people that have helped me to bring this work to an end is certainly long. In the first place I must thank my supervisors Dr Irene Moroz at OCIAM and Prof Peter Read at AOPP for so many discussions and useful comments that gave this thesis its final form.

All my gratitude to Dr. Andy White in the MetOffice, Dr. Frank Kwasniok in the University of Exeter, Dr. Ulrich Achatz of the Leibniz-Institute für Atmosphärenphysik, Univeristät Rostock for helpful discussions and comments on my work.

Special thanks to the Applied Dynamical Systems at OCIAM and the Geophysical Fluid Dynamics at AOPP for all the group meetings and seminars during which I had the opportunity of exchanging points of view. In particular, I would like to thank Dr. Luca Montabone for so many fruitful discussions that led me to learn more about our neighbouring red planet and appreciate all the interesting physical processes taking place there.

Life would not be the same without friends. This fits specially well in the case of a DPhil student. Special thanks to Lucjan, Benjaporn, Pavel, Eleni, Christina, Alfonso who kindly provided the "little help from my friends". A special place is due to Angela Amphawan for her support and enjoyable company that made the last time of producing this thesis a beautiful time to be remembered.

Finally, I would like to thank the Government of Mexico, in particular the National Council of Science and Technology (CONACYT), for the support provided through the scholarship No. 179183, which allowed me to start the work that led to this thesis. I am also grateful to Jesus College for providing funding for conference attendance and a writing-up grant that was extremely helpful during the last stage of this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "So the Colleges of Cartographers evolved a Map of the Empire that was of the same Scale as the Empire and that coincided with it point for point."
>
> Jorge Luis Borges (On exactitude in science [19])

1.1 Mars

Mars is the fourth planet in the Solar System and the last of the inner or terrestrial planets. It is probably the planet that shares most similarities with the Earth as can be seen in Table 1.1 where a comparison between Venus, the Earth and Mars is shown. Mars is a rocky planet surrounded by a layer of gas that constitutes its atmosphere. A Martian day, or sol, is about 40 minutes longer than a day on the Earth while a Martian year is slightly less than two terrestrial years. The present obliquity of Mars[1] of 25.2 ° allows the existence of seasons just like on our planet.

Topographically the Earth and Mars are less similar. Despite being around half the size of the Earth, with a planetary radius of 3396.0 km, the surface of Mars is covered by topographic features of an amazing height such as Olympus Mons, the largest volcano in the Solar System, reaching heights of more than 20 km with respect to surrounding plains. There are also very deep depressions of which Hellas Planitia, with a depth of around −7 km, is the best example. Furthermore, there is a difference of about 5 km on average in elevation between the northern and southern hemispheres [113].

[1] The obliquity of Mars is in a chaotic state and can vary from 0 ° to 60 ° [60].

Table 1.1: Comparison between Venus, the Earth and Mars.

	Venus [51, 129]	Earth [91]	Mars [91]
Orbital radius (10^{11} m)	1.08	1.50	2.28
Orbital eccentricity	0.0068	0.017	0.093
Planetary obliquity	177.3 °	23.5 °	25.2 °
Rotation rate ($10^{-5}s^{-1}$)	-0.02992	7.294	7.088
Solar day (Earth days)	116.7	1.003	1.027
Year length (Earth days)	224.70	365.24	686.93
Equatorial radius (10^6 m)	6.052	6.378	3.396
Surface gravity (m s^{-2})	8.87	9.81	3.72
Surface pressure	9 MPa	101300 Pa	600 Pa
Equilibrium temperature, T_e (K)	227	256	210
Gas constant, R (m^2 s^{-2} K^{-1})	192	287	192
Scale height, $H = RT_e/g$ (km)	4.9	7.5	10.8
Atmospheric constituents	CO_2 (0.96)	N_2 (0.77)	CO_2 (0.95)
(molar ratio)	N_2 (0.03)	O_2 (0.21)	N_2 (0.027)
	Others (0.01)	H_2O (0.01)	Ar (0.016)
		Ar (0.009)	O_2 (0.0013)

1.1.1 The Martian atmosphere

The Martian atmosphere is thinner than the atmosphere of the Earth. The mean surface pressure is 600 Pa compared to 101.3 kPa on the Earth (Table 1.1). Its main constituent is carbon dioxide (CO_2) in amounts up to 95% in molar ratio. Other atmospheric components on Mars are molecular nitrogen (N_2, 2.7%), argon (Ar, 1.6%) and molecular oxygen (O_2, 0.13%) [91].

The scale height H is the height at which the atmospheric pressure has dropped by a factor of e with respect to the surface pressure in an isothermal atmosphere. It also gives the height below which 63% of the atmosphereic mass is concentrated, considering again isothermal conditions. The terrestrial scale height is 7.5 km at $T = 227$ K which is comparable to the highest peak on the Earth, Mount Everest (higher than 8 km above sea level). The scale height on Mars is 10.8 km at $T = 210$ K. This figure gives a clear indication of the significant influence of topography over the Martian atmospheric motion. Comparing it with the height of Olympus Mons, the depth of Hellas Planitia or the average difference between southern and northern hemisphere, we expect all these features to have a remarkable influence over the Martian atmospheric circulation. For example, it has been found that the north-south topographic asymmetry has an important impact over the Hadley cell structure so that the southern summer cell becomes predominant [94]

Oceans on the Earth act as energy reservoirs that tend to damp the atmospheric response to radiative forcing [38]. The lack of oceans on Mars makes this planet more susceptible to stronger temperature variations and to a more rapid response to solar radiative forcing [91]. On the other hand, dust is an atmospheric component that plays a more important role on Mars than on the Earth. The dust content in the atmosphere has a considerable influence over its optical properties since the absorption of solar heating is enhanced when the atmospheric dust load is heavier. One of the most striking features of the Martian atmosphere is its capacity to develop and sustain dust storms that virtually cover the whole planet. This process has been subject to extensive studies involving both observations (see, for example, [114, 22]) and data assimilation (see, for example, [63, 77, 79]). Modelling of the processes that start, sustain and make a dust storm to decay have also been studied using Mars general circulation models (GCMs) [81, 82, 119].

Another feature that has no parallel on the Earth is the fact that part of the atmosphere sublimates and condenses during the warmest and coldest season on the respective hemisphere. As a result, the surface pressure experiences a significant variation (of up to 30% peak-to-peak) during a Martian year [48]. In addition, there are two other effects that lead to a strong variation in surface pressure not only with time but also with latitude. The first one is due to the latitudinal variation in altitude. The second one arises because of the strong zonal winds found on Mars [48].

Since the atmosphere, either on Mars or the Earth, is less dense than the ocean, the gravitational influence of the Sun or the moon on the atmosphere is less important. This gives ocean tides their importance on the Earth. In contrast, an important component of Martian atmospheric dynamics are thermal tides. Thermal tides are planetary waves that are harmonics of the solar day directly forced by solar heating [91]. The most important thermal tides on Mars are the migrating, sun-synchronous, diurnal and semidiurnal tides [35]. The semidiurnal tide, especially, exhibits a strong sensitivity to vertical solar heating. Since dust in the atmosphere enhances its radiative absorption properties, the semidiurnal tide magnitude also shows a strong correlation with atmospheric dust content [63]. The interaction of the diurnal tides with the large Martian topography gives rise to non-migrating (longitude-dependent) tides [131]. We shall revisit and elaborate on this topic in Section 6.1.1.

Like the terrestrial atmosphere, the atmosphere on Mars largely contributes to the transfer of energy between the equator and the poles. Therefore, its role is central at reducing the temperature difference between these two regions. Like on the Earth, this energy transport is carried out by waves generated via barotropic and baroclinic

instabilities. Barotropic instability is associated with the reduction of horizontal shear. This instability has also been related to the circumpolar jet in the middle Martian atmosphere [31]. Baroclinic instability develops in the presence of latitudinal temperature gradients and is responsible for transporting energy polewards.

The first two successful theoretical descriptions of this instability are due to Charney [24] and Eady [32]. The model proposed by Eady considers what is probably the simplest configuration of a continuous atmosphere capable of developing this kind of instability. It consists of a quasi-geostrophic channel with rigid lids at the domain's top and bottom rotating at a constant rate. Furthermore, this model assumes incompressibility and a particularly simple basic zonal flow given by $U = \Lambda z$. Under these assumptions the conditions for unstable wave-like solutions can be determined. If appropriate values are provided, this model predicts values that are in good agreement with those observed, for example, on the Earth [91]. Charney's model is a more sophisticated model, where the rigid-lid assumption is eliminated, the β-effect (variation of Coriolis parameter with latitude) is restored and the working fluid is assumed to be compressible. Although it leads to essentially similar conclusions, the variety of unstable structures predicted by Charney's model is richer and more detailed (see, for example, [87]).

Evidence for baroclinic activity on Mars has been found in surface pressure time series from the Viking Lander 2 mission [14, 15] and in Mars GCM simulations [16, 27]. Observations and GCM simulations have given indications of the regularity of baroclinic waves in the Martian atmosphere [14, 15, 16, 27]. Through the analysis of the meteorological information sent back to the Earth by the Viking Lander 2 during two consecutive Martian years, Barnes [14, 15] found that what appeared to be baroclinic activity in the northern hemisphere was characterised by rather similar parameters during spring and autumn in both years. This striking regularity of transient waves was also found in simulations carried out using the NASA Ames GCM [16].

By carrying out numerical experiments with the aid of the Oxford Mars GCM, Collins *et al.* [27] showed that, without the influence of the diurnal cycle, the Martian atmosphere probably would display a behaviour characterised by the presence of coherent baroclinic waves during late autumn, winter and early spring with periods of either 2.2 or 5.5 sols, approximately. Waves with these characteristic periods were observed during the simulation of eight consecutive years. This result was interpreted as multiple coexisting equilibria. When the diurnal cycle was included the baroclinic waves did not display a regular behaviour any more. Instead, they behaved in a

more irregular way which was, nevertheless, highly influenced by oscillations with periods of 5.6 and 2.6 sols. This result suggests that the action of the thermal tides, induced by the diurnal cycle, is to perturb the system and make it switch between two equilibrium states.

A related example within the field of terrestrial dynamic meteorology is the Charney-DeVore model [25]. This model is a severely truncated barotropic spectral model aimed at explaining the existence of highly recurrent anomalies of the Earth's atmospheric circulation [25]. Using this model, Charney and DeVore [25] have shown the possibility of not only two but multiple coexisting equilibria with the system exhibiting random transitions between them.

Additionally, the global predictability of the Martian atmosphere throughout the Martian year has been assessed by Newman *et al.* [83]. By the use of breeding vectors [120], they found a clear indication that indeed the atmosphere of Mars is more predictable than the atmosphere of the Earth, specially during the interval from mid-spring to mid-autumn in the northern hemisphere.

1.2 Reduced-order modelling

The findings described at the end of Section 1.1.1 constitute rather intriguing features of the Martian atmosphere considering the high complexity that is expected in such a system. The enhanced regularity observed in the atmosphere of Mars with respect to its terrestrial counterpart suggests the hypothesis of a low-dimensional underlying attractor of the Martian climate. This hypothesis constitutes the motivation for the present thesis.

The hypothesis of low-dimensional attractors is not exclusive of this case. A related problem is, for example, the search for inertial manifolds (IMs), which are smooth finite-dimensional invariant manifolds that exponentially attract every trajectory in dissipative systems governed by certain partial differential equations [73].

The actual problem of finding IMs, either analytically or numerically, does not represent an easy task even for systems, such as the Kuramoto-Sivashinsky equation [111], which are known to possess an IM [117]. A more suitable and practical method of attacking this question is to seek the representation of the governing equations of the system of interest (in our case, the Martian atmosphere) by lower-order dynamical models so that these retain the essential dynamics. This is precisely the aim of reduced-order modelling techniques. These techniques will be reviewed in this section before stating the objectives of this work in Section 1.3.

1.2.1 Techniques for order reduction

There are various methods available for reducing the dimension of a system. The most immediate approach is probably heuristic. It consists of proposing a priori solutions to the equations of motion on the grounds of symmetry and boundary conditions. These solutions usually takes the form of a truncated series in terms of general sets of orthogonal functions, such as Fourier modes or spherical harmonics. In certain applications, this approach has been quite successful, especially in theoretical analyses of a variety of systems. Perhaps the best-known example is the Lorenz system [68], which arises from the severe truncation of the equations of free convection after the solution has been expanded in terms of Fourier modes. Other examples can be found in [25] and [69], for example.

An approach with a different underlying philosophy is represented by what here we call rational reduction methods. The main characteristic of these methods is that the proposed solutions are somehow related to a particular system. Antoulas [8] proposes a classification of rational methods into two main groups, namely, Krylov and singular value decomposition (SVD) methods. Krylov methods make use of iterations for finding approximations to large-scale dynamical systems. The SVD methods are based on the decomposition of the state vector into a set of vectors that can be ordered in a certain sense. These methods include balanced truncation and Hankel approximations for linear systems, and the proper orthogonal decomposition (POD) methods and empirical grammians for nonlinear systems. In this thesis we are concerned with the POD and its combination with Galerkin projection to produce reduced dynamical versions of the original large-scale system. The interested reader is referred to [8] for a more complete account of the available reduction methods.

The POD [52, 65] is a statistical technique to extract features from a given dataset by searching for patterns that optimally represent the dataset with respect to quantities such as variance or energy. The output of the POD is a set of time-independent orthogonal functions called empirical orthogonal functions (EOFs). Each EOF is associated with a certain amount of variance or energy. The first suggestion to use the POD in the analysis of dynamical systems was due to Lumley [71]. One of the first phenomena to be studied by means of reduced models arising from the use of the POD was turbulence [12, 18].

Perhaps for nonlinear systems the most studied method is the POD in conjunction with Galerkin projections [8]. However, this is not the only available method and some other ideas have been proposed, such as a decomposition into principal interaction patterns (PIPs), which explicitly incorporate information about the sys-

tem's dynamics within the determination of the patterns [42, 56, 57]. Recently, a new algorithm has been proposed in which the PIPs are computed in a sequential fashion, giving them a natural order just as for the EOFs in the POD [59]. With such an algorithm it should be possible, at least in principle, to tackle the reduction of very high dimensional problems such as the one here under study.

There are other techniques that might be suitable for developing low-order models such as the independent component analysis (ICA) [28]. The ICA aims at answering questions such as, what signal comes from what source? This problem is known as the *cocktail-party problem* [96]. The solution relies on the assumption of statistical independence of the sources and, therefore, of the signals. The outcome is a set of modes and a mixing matrix that sets the relationship between the modes in order to reconstruct the original mixed signal. As mentioned before, the ICA modes are statistically independent but not necessarily orthogonal and there is no specific order. Furthermore, they are not clearly related to any physical quantity that can help to decide what modes to retain and what to rule out.

These features may constitute a disadvantage when trying to construct low-order models since a truncation using these modes would lack an appropriate parameter to measure the expected accuracy of the resulting model. Original algorithms assume linearity although some efforts have been made to extend the formalism to nonlinear systems. For a more complete description of the method and the algorithms involved the reader is referred to the book of Roberts and Everson [96]. So far these ideas have not yet been further investigated in the context of dynamical systems.

1.2.2 Reduced-order models in atmospheric physics

The POD were introduced in atmospheric sciences from the late 1940s as a means of atmospheric diagnosis [84, 85], while dynamical models based on the EOFs have been sought since the late 1950s [102]. However, these early studies faced problems apparently associated to grid-point evaluation of higher order derivatives and an insufficient number of EOFs [102]. The work by Schubert [102, 103] and Selten [104, 105, 106] gave a new impulse to the development of EOF-based low-order models using barotropic and quasi-geostrophic models as the underlying equations of motion.

In the meantime, PIP-based models [5, 58] were being investigated. These models were shown to be an improvement with respect to EOF-based models regarding the number of necessary modes and prediction power [29, 59]. However, EOF-based models still enjoy the advantage of simplicity both in interpretation and implementa-

tion. Achatz and Opsteegh [3, 4] have brought these models a step forward towards more realistic models by developing primitive-equation-based reduced models for the terrestrial atmosphere. The work of these authors constitutes the most direct basis for the present study regarding order reduction.

1.2.3 Reduced-order models and the atmosphere of Mars

Work on low-order models in the context of planetary atmospheres other than the terrestrial atmosphere is still sparse despite the potential significance this research would have for the study of extraterrestrial atmospheres and comparative planetology [74]. Whitehouse *et al.* [124] analysed the possibility of representing an atmospheric dataset, produced by a simplified Mars GCM, by a small number of modes. Their aims were achieved by means of a vertical decomposition, based on the quasi-geostrophic vertical structure equation, followed by the POD of the reduced dataset. By this diagnostic procedure, they achieved a reduction from $\mathcal{O}(10^4)$ degrees of freedom in the original model to $\mathcal{O}(50)$ degrees of freedom in the reduced dataset. In a second article, Whitehouse *et al.* [125] presented the construction of dynamical reduced-dimension models. They managed to construct successful models with 80 degrees of freedom in a combination of vertical and horizontal modes derived during the diagnostic analysis previously described.

In this thesis we closely followed the programme established by Whitehouse *et al.* [124, 125] in order to investigate the low-order properties of a more realistic representation of the atmosphere of Mars. There are three important differences between that work and the present study, especially regarding the Oxford Mars GCM [36] dataset. These are the inclusion of topography, the diurnal cycle and the complex parameterisation of physical processes. In this respect, the present work is closer to that by Martínez-Alvarado *et al.* [74], who carried out the decomposition of a dataset from the Oxford MGCM in vertical eigenmodes and horizontal EOFs under the quasi-geostrophic theory, also following the programme by Whitehouse *et al.* [124]. Regarding the construction of low-order models, the main difference between this thesis and previous works lies in the use of the primitive equations rather than the filtered quasi-geostrophic equations as relevant equations of motion (see Section 2.2). This approach has led to the use of a multi-variate three-dimensional version of the POD, which produces EOFs to represent simultaneously velocity components and a thermal field (to be specified later). This approach has been used in the terrestrial contexts by Achatz and Opsteegh [3, 4] and in acoustic studies by Rowley *et al.* [98].

1.2.4 Reduced-order models in other fields

Nowadays the conjunction of the rapid development of computational systems and the detailed knowledge of a great deal of the important physical and chemical processes that have an influence over atmospheric dynamics has enabled scientists to simulate atmospheric motion in great detail.

This situation is not only prevalent in atmospheric sciences. Physics, biology, chemistry and many engineering sciences have also benefited from the combination of computational power and enhanced physical knowledge (see, for example, [8]). With such powerful models we certainly gain many advantages: an accurate representation of physical reality. Paradoxically the high complexity of a model still has disadvantages: the model becomes almost as complex as the system under study just as the map of the Empire in Borges' short story "On Exactitude in Science" [19]. The complicated relationships between variables and parameters make the analysis of the model as difficult as that of the original system, and the time to obtain predictions becomes of the same order as the time to observe the system or require the aid of powerful supercomputers. When time and space are primary constraints in applications, such as in forecasting and control, this can be a serious disadvantage.

1.3 Objectives

In view of the earlier indications of regularity in the Martian atmosphere discussed previously (cf. Section 1.1.1) and the results obtained by Whitehouse *et al.* [124, 125] discussed in Section 1.2.2, two important questions arise. The first question is related to the diagnostic understanding of the atmosphere of Mars and can be stated as whether the state of the Martian atmosphere, during a given time interval, can be represented by a small number of significant atmospheric modes.

Notice that, given a set of modes of this sort we could not only describe the atmospheric state at a given instant; we would also be capable of describing its evolution by means of a time series built from the reduced-order instantaneous description of the atmosphere at subsequent times. However, this is different to having a dynamical model capable of predicting the future of the atmosphere by knowing its state at a certain initial time. Addressing this issue, the second question is directly related to the dynamical behaviour of the atmospheric modes referred to previously. Hence, this question can be stated as whether these atmospheric modes are sufficient to describe the dynamical behaviour of the atmosphere of Mars. This question can be viewed as

the prognostic counterpart of the first one.

Pursuing the answers to these two questions directly leads to the objectives of the present thesis:

(*i*) To find a small number of atmospheric modes to represent a dataset for the Martian atmosphere.

(*ii*) To determine the sufficiency of the atmospheric modes found in (*i*) to explain the dynamical behaviour of the Martian atmosphere.

The first objective can be linked to the POD since, as we have seen in Section 1.2.1, this technique *per se* seeks static modes (independent of time) that can be used to reproduce a given dataset optimally (in the sense that the number of necessary modes to represent a given amount of variance or energy is less than in any other linear decomposition). The second objective requires a means of establishing physical interactions among the retained atmospheric modes. This means is provided by the projection of the equations of motion onto the span, in phase space, of the chosen modes.

Throughout this thesis we shall see that, within certain limits, the answers to both questions are indeed positive. The total energy in the Martian atmosphere is mostly contained in a few atmospheric modes of the order $\mathcal{O}(100)$. Furthermore, making these modes interact in a physically sensible way can in fact reproduce, albeit coarsely, the actual behaviour of these modes in the atmosphere. However, this is not a simple matter and involves a great deal of consideration directly related to the modelling process itself, such as the numerical treatment of high frequency waves, which are irrelevant to the general circulation. Although a separate study would be needed to construct long-term reduced models to climate change studies, for example, the positive answers to the questions posed at the beginning of this section are encouraging for the development of short term models.

1.4 Structure of the thesis

To reach our objectives we shall analyse two datasets. The first one comprises the output of the simplified GCM. Whitehouse *et al.* [124, 125] analysed an SGCM dataset as well. However, the regimes analysed in those and the present works are different. While Whitehouse *et al.* [124, 125] studied a wave-3 regime, arising from a simulation of perennial southern winter, in this thesis we study a weak wave regime superimposed over an axisymmetric background flow. The second dataset is obtained

from the output of the Oxford Mars GCM, a state-of-the art general circulation model used in Oxford since the mid-90s for the study of the Martian atmosphere. Prior to the analysis of these datasets, we shall introduce two central topics that constitute the background of this work.

Atmospheric dynamics is the first of these central topics. It will be reviewed in Chapter 2. In this chapter we shall also review different Mars GCMs focusing our attention on those details that will be important later on in the present thesis. In particular, we will explain the reasons for working within the framework of the primitive equations rather than other simpler theories.

The second important topic that must be reviewed before fully entering into pursuing our objectives is the theory behind the POD to be discussed in Chapter 3. This chapter will not only deal with theoretical aspects but also with the numerical implementation of the method. Moreover, we shall introduce and explain crucial assumptions that will give shape to our reduced-dimension approach to the Martian atmospheric dynamics.

After these introductory chapters, we shall proceed to the analysis of our two available datasets. The output of the simplified GCM was the first dataset to be investigated. Despite the dissimilarities in regimes and methods of analysis (see Section 1.2.3), the work in [124, 125] and the present study are still comparable. This comparison, presented in Chapter 4, is used both as a validation for the methods used here and as a point of reference for the interpretation of the results regarding our ultimate objective, the MGCM dataset. In Chapter 5, the POD-Galerkin projection is employed for the construction of EOF-based reduced models of the SGCM. The corresponding results for the MGCM are presented in Chapters 6 and 7.

Chapter 8 is concerned with certain theoretical details that should be considered in the development of future reduced-order models based on the POD. Rather than the last word on these issues, Chapter 8 aims at pointing out these issues and showing preliminary results that suggest that pursuing further investigation in these directions is likely to be worthwhile.

Finally, in Chapter 9, a general summary of this work is presented together with the most important conclusions that can be drawn from the research involved. There is a considerable variety of potential paths that can be followed to continue the research in this area. This potential future work will also be discussed in Chapter 9 bringing this work to a provisional end.

Chapter 2

Atmospheric dynamical equations

This chapter is devoted to the introduction of the dynamic equations governing the evolution of a planetary atmosphere. The so-called primitive equations of dynamic meteorology provide us with a good approximation to the equations that govern atmospheric motion, despite incorporating several simplifications. These equations are the object of Section 2.1 where we shall also discuss the choice of appropriate vertical coordinates depending on the intended application. Further simplifications to the equations of motion lead to different approximations. One such approximation is the quasi-geostrophic approximation which will be reviewed in Section 2.2 where we shall also discuss the convenience of using this approximation in the Martian context.

General circulation models (GCMs) constitute the most advanced class of atmospheric models available today. For the Earth, there is a whole variety of different GCMs. For Mars, the number is smaller although it is growing. In Section 2.3, we shall give an overview of GCMs and the numerical methods currently employed for solving the relevant dynamic equations.

2.1 Primitive equations

2.1.1 Momentum

A planetary atmosphere is a fluid system whose behaviour is completely described by the Navier-Stokes equations in conjunction with the continuity and thermodynamic energy equation, which express mass and energy conservation laws. These are usually complemented by other relations, which describe the conservation of partial atmospheric constituents such as water vapour, ozone, carbon dioxide or dust. The Navier-Stokes equations applied to a rotating planet using a rigidly rotating coordi-

nate system (fixed to the planet) can be written as (see [130], for example)

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + \frac{1}{\rho} \nabla p + \mathbf{g} + 2\mathbf{\Omega} \times \mathbf{v} - \mathbf{F} = 0, \tag{2.1}$$

where $\mathbf{v} = (u, v, w)^T$ is velocity, ρ is density, p is pressure, $\mathbf{\Omega}$ is the planetary rotation rate and $\mathbf{F}$ represents frictional forces. The rotational effects appear in these equations in the fourth and fifth terms. The apparent gravity $\mathbf{g}$ results from the combination of true gravity $\mathbf{g}'$ (directed towards the planetary centre) and the centrifugal force $\Omega^2\mathbf{R}$, where $\mathbf{R}$ is a position vector with respect to the planetary axis of rotation. The term $2\mathbf{\Omega} \times \mathbf{v}$ is the negative of the Coriolis force.

These equations, written in scalar form in planetographic coordinates (longitude λ, latitude ϕ and altitude z) have a rather complicated form (see [130], for example)

$$\frac{du}{dt} + \frac{uw}{r} - \frac{uv}{r} \tan\phi + lw - fv + \frac{1}{\rho}\frac{\partial p}{\partial x} - F_\lambda = 0, \tag{2.2}$$

$$\frac{dv}{dt} + \frac{vw}{r} + \frac{u^2}{r} \tan\phi + fu + \frac{1}{\rho}\frac{\partial p}{\partial y} - F_\phi = 0, \tag{2.3}$$

$$\frac{dw}{dt} - \frac{1}{r}(u^2 + v^2) - lu + \frac{1}{\rho}\frac{\partial p}{\partial z} + g - F_z = 0, \tag{2.4}$$

with

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z},$$

where

$$\frac{\partial}{\partial x} = \frac{1}{r\cos\phi}\frac{\partial}{\partial \lambda}, \qquad \frac{\partial}{\partial y} = \frac{1}{r}\frac{\partial}{\partial \phi}, \qquad \frac{\partial}{\partial z} = \frac{\partial}{\partial r},$$

and

$$l = 2\Omega\cos\phi, \qquad f = 2\Omega\sin\phi.$$

A method for simplifying (2.2)-(2.4) is to neglect small terms determined by scale analysis. This technique consists of a systematic comparison between the different terms occurring in the equations of motion after considering typical values of the variables involved in each term [130]. This leads to the simplified system

$$\frac{du}{dt} - \frac{uv}{a} \tan\phi - fv + \frac{1}{\rho}\frac{\partial p}{\partial x} - F_\lambda = 0, \tag{2.5}$$

$$\frac{dv}{dt} + \frac{u^2}{a} \tan\phi + fu + \frac{1}{\rho}\frac{\partial p}{\partial y} - F_\phi = 0, \tag{2.6}$$

$$\frac{1}{\rho}\frac{\partial p}{\partial z} = -g, \tag{2.7}$$

where the variable r has been substituted by the constant planetary radius a. This substitution implies the assumption that the atmosphere is shallow ($\sim$ 80 km) in comparison with the radius of the planet ($\sim$ 3000 km, for Mars) and should be made extensive to the definitions

$$\frac{\partial}{\partial x} = \frac{1}{a\cos\phi}\frac{\partial}{\partial \lambda}, \qquad \frac{\partial}{\partial y} = \frac{1}{a}\frac{\partial}{\partial \phi}.$$

Equation (2.7) is called the hydrostatic equation. Although there are models that do not make use of the hydrostatic approximation, it provides a sufficient accuracy for synoptic-scale motion on Mars [91], and will, therefore, be used in this work.

Equations (2.5)-(2.6) are the horizontal component of the equations of motion and can be written in vector form as

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{k}\times\mathbf{v} + \frac{1}{\rho}\nabla p - \mathbf{F} = 0, \tag{2.8}$$

where the total time derivative is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla + w\frac{\partial}{\partial z}$$

and all the vector quantities (with the exception of $\mathbf{k}$) must be re-interpreted as horizontal vectors so that $\mathbf{v} = (u, v)^T$, $\mathbf{F} = (F_\lambda, F_\phi)^T$ and

$$\nabla = \frac{\mathbf{e}_\lambda}{a\cos\phi}\frac{\partial}{\partial\lambda} + \frac{\mathbf{e}_\phi}{a}\frac{\partial}{\partial\phi}.$$

The horizontal momentum equations are part of the hydrostatic primitive equations. In addition, there are three other equations that need to be considered when modelling a planetary atmosphere. These are (*i*) the continuity equation, (*ii*) the thermodynamic energy equation, and (*iii*) an equation of state.

2.1.2 Continuity

The continuity equation represents mass conservation and is given in a rigidly rotating coordinate system (fixed to the planet) by (see e.g. [45])

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) + \frac{\partial(\rho w)}{\partial z} = 0, \tag{2.9}$$

where

$$\nabla \cdot \mathbf{A} = \frac{1}{a\cos\phi}\left(\frac{\partial A_\lambda}{\partial \lambda} + \frac{\partial}{\partial \phi}(A_\phi \cos\phi)\right).$$

2.1.3 Thermodynamic energy

The usual form of the thermodynamic energy equation, representing the first law of thermodynamics, is given by (see e.g. [45])

$$c_v \frac{DT}{Dt} + p\frac{D\alpha}{Dt} = J, \tag{2.10}$$

where T is temperature, $\alpha = \rho^{-1}$ is specific volume, J is the rate of diabatic heating per unit mass.

An appropriate equation of state, relating density, temperature and pressure, is the ideal gas equation

$$p\alpha = \frac{p}{\rho} = RT, \tag{2.11}$$

where R is the gas constant per unit mass (see Table 1.1). With the aid of (2.11), the first law of thermodynamics (2.10) can be written in terms of the entropy s as

$$\frac{Ds}{Dt} = \frac{J}{T}. \tag{2.12}$$

where the definition of the entropy of an ideal gas

$$s = s_0 + c_p \ln T - R \ln p \tag{2.13}$$

has been used.

Atmospheric studies often make use of a quantity called *potential temperature* defined as

$$\theta = T\left(\frac{p_0}{p}\right)^\kappa, \tag{2.14}$$

where p_0 is a (constant) reference pressure and κ is the ratio of the gas constant to the specific heat at constant pressure c_p. The interpretation of the potential temperature can be done by taking its logarithm and multiplying by c_p to obtain

$$c_p \ln\theta = c_p \ln T - R\ln p - R\ln p_0. \tag{2.15}$$

Comparing (2.15) with (2.13) we find that $s = c_p \ln\theta$. This is equivalent to saying that a surface of constant θ is an isentropic surface, i.e. a parcel of fluid subject only

to reversible processes must remain on a surface of constant θ. Thus, the first law of thermodynamics (2.12) can also be written in terms of θ as

$$c_p \frac{D}{Dt}(\ln \theta) = \frac{J}{T}. \tag{2.16}$$

Equations (2.8), (2.7), (2.9) and (2.16) complemented by (2.11) are the hydrostatic primitive equations of dynamic meteorology and constitute a system of six equations for six variables (three velocity components, temperature, pressure and density) determining the state of a planetary atmosphere.

2.1.4 Vertical coordinates

So far, we have written the dynamical equations that determine the state of the atmosphere in a rigidly rotating coordinate system which implies the use of height as a vertical coordinate. However, height is not always the most suitable vertical coordinate to be used in atmospheric modelling since the atmospheric equations can take more convenient forms in different coordinates. The transformation of the gradient from a general vertical coordinate to any single-valued monotonic function η of height or pressure as vertical coordinate is given by the alternative relation

$$\nabla_\eta = \nabla_z + \nabla_\eta z \frac{\partial}{\partial z}, \tag{2.17}$$

where ∇_x indicate that the differentiation must be carried out holding the variable x constant.

Isobaric coordinates

In particular, if in (2.17) z is height and η is pressure, we obtain the transformation from rigidly rotating coordinates to isobaric, or pressure, coordinates. In these coordinates, the pressure gradient term in (2.8) takes the form

$$-\rho^{-1}\nabla_z p = -\nabla_p \Phi, \tag{2.18}$$

where Φ is the geopotential gz (recall that g is the apparent gravity including the actual gravity and the centrifugal force due to rotation). After substituting (2.18) into (2.8), the horizontal component of the equations of motion become

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{k} \times \mathbf{v} + \nabla_p \Phi - \mathbf{F} = 0. \tag{2.19}$$

The total time derivative must be re-defined as

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_p + \omega \frac{\partial}{\partial p},$$

where $\omega = Dp/Dt$ and ∇_p is the gradient operator at constant pressure. In the isobaric coordinates the continuity equation takes the form

$$\nabla_p \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0. \tag{2.20}$$

Notice that there is no time derivative in (2.20) and that there is no reference to the density field in (2.19) or (2.20). These are properties of the isobaric coordinate system that give certain advantages to it. For completeness, we also include isobaric versions of the hydrostatic equation

$$\frac{\partial \Phi}{\partial p} = -\frac{1}{\rho} = -\frac{RT}{p}, \tag{2.21}$$

where the equation of state for an ideal gas has been used, and the thermodynamic energy equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_p \right) T - S_0 \omega = \frac{J}{c_p}, \tag{2.22}$$

where

$$S_0 = -\frac{1}{\rho} \frac{\partial \ln \theta}{\partial p}$$

is the static stability for the isobaric system. Equations (2.19), (2.20), (2.21) and (2.22) constitute the dynamical equations in isobaric coordinates. We shall use these equations in Section 2.2 when we deal with the so-called quasi-geostrophic theory.

Sigma-coordinates

As described in Chapter 1, Mars is a planet with very large topographic variations. In such a case, a terrain-following coordinate system is generally more appropriate. In a terrain-following system, the ground always coincides with a coordinate surface and the specification of the boundary conditions at ground level becomes simpler than using height or pressure as vertical coordinates. Probably the first terrain-following coordinate system was the sigma-coordinate system [88], where the vertical coordinate is defined as

$$\sigma = \frac{p(\mathbf{x}, z, t)}{p^s(\mathbf{x}, t)}, \tag{2.23}$$

where p^s is surface pressure and $\mathbf{x}$ represents a suitable set of horizontal coordinates.

The advantages of sigma-coordinates near the ground can represent a problem at higher altitudes. For example, since these coordinates have the print of topography throughout, constant fields over isobaric surfaces might exhibit steep gradients over sigma surfaces [91]. Hybrid-coordinates [101], a combination of sigma and isobaric coordinates, have been used in atmospheric models to circumvent these problems [9, 108]. However, preliminary tests carried out using the Mars general circulation model at the Laboratoire de Météorologie Dynamique (LMD) have shown that a representation over sigma-coordinates is sufficiently accurate on Mars, at least for altitudes below 100 km [91].

In order to transform from rigidly rotating coordinates to sigma-coordinates, we substitute σ for η in (2.17) to obtain

$$\nabla_\sigma = \nabla_z - \rho \nabla_\sigma \Phi \frac{\partial}{\partial p}, \tag{2.24}$$

where the hydrostatic equation (2.7) and the chain rule have been used. Application of (2.24) on p yields the pressure gradient term

$$-\frac{1}{\rho}\nabla_z p = -\nabla_\sigma \Phi - \frac{\sigma}{\rho}\nabla p^s. \tag{2.25}$$

By substituting (2.25) in (2.8) we obtain the horizontal component of the equations of motion in sigma-coordinates

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{k}\times\mathbf{v} + \nabla\Phi - \frac{\sigma}{\rho}\nabla p^s = \mathbf{F}, \tag{2.26}$$

where the gradient operator must be computed holding σ constant and

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla + \dot{\sigma}\frac{\partial}{\partial\sigma}, \quad \text{with} \quad \dot{\sigma} = \frac{D\sigma}{Dt}.$$

The continuity equation in sigma coordinates is more easily derived taking (2.20) as a starting point. In this case we need the transformation of the gradient operator from isobaric to sigma coordinates which can be computed by substituting p for z and σ for η in (2.17)

$$\nabla_p = \nabla_\sigma - \sigma\nabla(\ln p^s)\frac{\partial}{\partial\sigma}. \tag{2.27}$$

Using (2.27) and the relation [88]

$$\omega = p^s\dot{\sigma} + \sigma\frac{Dp^s}{Dt}$$

we obtain the continuity equation in the desired form (see also [45])

$$\frac{\partial p^s}{\partial t} + \nabla\cdot(p^s\mathbf{v}) + \frac{\partial p^s\dot{\sigma}}{\partial\sigma} = 0. \tag{2.28}$$

The thermodynamic energy equation in flux form in sigma-coordinates can be obtained by multiplying (2.16) by p^s and using (2.28)

$$\frac{\partial}{\partial t}(p^s c_p\ln\theta) + \nabla\cdot(p^s\mathbf{v}c_p\ln\theta) + \frac{\partial}{\partial\sigma}(p^s\dot{\sigma}c_p\ln\theta) = p^s\frac{J}{T}, \tag{2.29}$$

Finally, the hydrostatic equation in sigma coordinates can be obtained by the chain rule so that

$$\frac{\partial\Phi}{\partial\sigma} = -\frac{p^s}{\rho}, \tag{2.30}$$

or, using (2.11),

$$\frac{\partial\Phi}{\partial\sigma} = -\frac{RT}{\sigma}. \tag{2.31}$$

2.1.5 Dimensionless primitive equations in sigma-coordinates

In order to obtain a non-dimensional version of equations (2.26), (2.28), (2.29) and (2.31), the planetary radius a and the reciprocal rotation rate Ω^{-1} were used as length and time scales, respectively. A complete list of non-dimensionalising factors is given in Table 2.1. Thus, the primitive equations in their non-dimensional version become

$$\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} + \dot{\sigma}\frac{\partial\mathbf{v}}{\partial\sigma} + f\mathbf{k}\times\mathbf{v} + \nabla\Phi - \sigma\frac{\partial\Phi}{\partial\sigma}\nabla\ln p^s = \mathbf{F}, \tag{2.32}$$

$$\frac{\partial}{\partial t}(p^s\ln\theta) + \nabla\cdot(p^s\mathbf{v}\ln\theta) + \frac{\partial}{\partial\sigma}(p^s\dot{\sigma}\ln\theta) = p^s\frac{J}{T}, \tag{2.33}$$

$$\frac{\partial p^s}{\partial t} + \nabla\cdot(p^s\mathbf{v}) + \frac{\partial p^s\dot{\sigma}}{\partial\sigma} = 0, \tag{2.34}$$

$$\frac{\partial\Phi}{\partial\sigma} = -\frac{T}{\sigma}. \tag{2.35}$$

This is the system of equations to be used as a starting point for the derivation of the equations to be used in this work. Further assumptions, regarding surface pressure, will be introduced in Section 3.3. These assumptions will lead to the equations

Table 2.1: Non-dimensionalising factors

Dimension	Factor	Value
Length	a	3.3960×10^6 m
Time	Ω^{-1}	1.4108×10^4 s
Velocity	$a\Omega$	240.7 m/s
Temperature	$(a\Omega)^2/R$	303.1 K
Geopotential	$(a\Omega)^2$	5.794×10^4 m^2/s^2
Pressure	p_0	500 Pa
Frictional force	$a\Omega^2$	1.7062×10^{-2} m/s^2
Diabatic heat	$a^2\Omega^3/\kappa$	17.614 J(kg $\cdot$ s)$^{-1}$

presented in Section 5.1, which will constitute the actual system to be used in this thesis.

2.2 Quasi-geostrophic theory

Further scale analysis of (2.5) and (2.6) shows that at mid-latitudes ($25° \leq \phi \leq 80°$) on the Earth, the Coriolis and pressure gradient terms are dominant (see [45] or [130], for example). This leads to the geostrophic equations

$$-fv \simeq -\frac{1}{\rho}\frac{\partial p}{\partial x}, \quad fu \simeq -\frac{1}{\rho}\frac{\partial p}{\partial y}. \tag{2.36}$$

Since there is no time derivative in (2.36), these equations are diagnostic rather than prognostic relations. In order to obtain a prognostic system, an appropriate approximation to the tendencies must be adopted so that the resulting system incorporates the simplification given by (2.36). The appropriate approximation is achieved by writing horizontal velocity as composed of two parts: the geostrophic and ageostrophic winds. Equation (2.36), written in isobaric coordinates in vector form, leads to the definition of the geostrophic wind

$$\mathbf{v}_g = \mathbf{k} \times \nabla \left(\frac{\Phi}{f_0} \right) = \mathbf{k} \times \nabla \psi. \tag{2.37}$$

Notice that the constant f_0 has been substituted for f so that the geostrophic wind is non-divergent. The ageostrophic part is simply defined as

$$\mathbf{v}_a = \mathbf{v} - \mathbf{v}_g, \tag{2.38}$$

where a requirement for the theory to be valid is to assume that [45]

$$|\mathbf{v}_a| \gg |\mathbf{v}_g|. \tag{2.39}$$

The resulting system is called quasi-geostrophic (QG) theory.

There are available versions of QG theory on the sphere (e.g. [38, 116]). However, at this point we are interested only in a qualitative discussion of the theory. Therefore, for the sake of simplicity, the following analysis will be described in a mid-latitude β-plane, where the Coriolis parameter is substituted by the approximation

$$f = f_0 + \beta y, \tag{2.40}$$

where $y = a\phi$. This approximation results from the expansion of $f = f(\phi)$ in a Taylor series around the mean latitude ϕ_0 of the latitudinal belt under consideration (e.g. $25° \leq \phi \leq 80°$) so that f_0 is the value of the Coriolis parameter at ϕ_0 and

$$\beta = \left(\frac{df}{dy}\right)_{\phi_0} = \frac{2\Omega \cos\phi_0}{a}. \tag{2.41}$$

The fundamental equations of QG theory, in isobaric coordinates, are the QG vorticity equation

$$\frac{\partial \zeta_g}{\partial t} + \mathbf{v}_g \cdot \nabla \eta_g = f_0 \frac{\partial \omega}{\partial p}, \tag{2.42}$$

and the QG thermodynamic energy equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_g \cdot \nabla\right)\left(-\frac{\partial \Phi}{\partial p}\right) - \overline{S}_0 \omega = \frac{\kappa J}{p}, \tag{2.43}$$

where $\zeta_g = \mathbf{k} \cdot \nabla \times \mathbf{v}_g = \nabla^2 \psi$, $\eta_g = \zeta_g + f$ is the absolute QG vorticity and

$$\overline{S}_0 = -\frac{RT_0}{p}\frac{d \ln \theta_0}{dp}$$

is the horizontal average of S_0 with $T_0 = T_0(p)$ representing a static standard profile and θ_0 being the potential temperature corresponding to T_0. This step is necessary to conserve integral properties in the model [130] and implies an underlying assumption in QG theory: the decomposition of the thermodynamic fields in standard static profiles, which depend only on pressure, and deviations from these profiles [54, 122].

Thus, for a thermodynamic variable A we have

$$A(\mathbf{x}, p, t) = A_0(p) + A'(\mathbf{x}, p, t). \tag{2.44}$$

After some manipulation (2.42) and (2.43) can be brought to the conservation equation form

$$\frac{\partial q}{\partial t} + \mathbf{v}_g \cdot \nabla q = 0, \tag{2.45}$$

where q is the QG potential vorticity defined as

$$q = \frac{1}{f_0}\nabla^2\Phi + f + \frac{\partial}{\partial p}\left(\frac{f_0}{\overline{S}_0}\frac{\partial \Phi}{\partial p}\right). \tag{2.46}$$

Assuming that $\omega = 0$, at a reference pressure $p = p_0$ and adiabatic motion ($J = 0$), we obtain the lower boundary condition [45]

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_g \cdot \nabla\right)\left(-\frac{\partial \Phi}{\partial p}\right) = 0 \quad \text{at} \quad p = p_0. \tag{2.47}$$

Equation (2.46), supplemented with suitable boundary conditions, can be used to find Φ corresponding to a given q. This procedure is called static PV inversion [41, 54]. A complete description of the geopotential evolution requires the inversion of (2.45), given (2.47) and Φ (obtained by static PV inversion), to compute $\partial q/\partial t$ (equivalent to computing $\partial \Phi/\partial t$). This procedure is called prognostic PV inversion [41, 54] and is necessary in the integration of any quasi-geostrophic model [54].

Condition (2.39) can be stated more precisely as [45]

$$\frac{|\mathbf{v}_a|}{|\mathbf{v}_g|} \sim \mathcal{O}(\mathrm{Ro}), \tag{2.48}$$

where Ro is the Rossby number defined as

$$\mathrm{Ro} = \frac{U}{f_0 L}, \tag{2.49}$$

where U and L are appropriate horizontal velocity and length scales. It is a dimensionless parameter which gives the ratio of inertial to rotational forces. Thus, the validity of this approximation depends on the smallness of the Rossby number. On the Earth, typical values are $U_\oplus = 10$ m/s and $L_\oplus = 10^6$ m (the subindex $\oplus$ refers to the Earth), respectively, while $f_0 \sim 10^{-4}$. Thus, on the Earth the Rossby number

takes the value

$$\mathrm{Ro}_\oplus \simeq 0.1.$$

On Mars, the winds are twice stronger while the planet is about half the size of the Earth. Therefore,

$$\mathrm{Ro}_M \simeq 4\mathrm{Ro}_\oplus \simeq 0.4,$$

where the subindex M refers to Mars. Note that these values are based on observations of synoptic-scale motions on both planets and not necessarily coincide with the scales in Section 2.1 where the choice was made to match the scaling in the MGCM. Although Ro_M is still a small number, this simple analysis provides a first indication that the geostrophic wind is not such a good approximation to mid-latitude horizontal winds on Mars as it is on the Earth.

A second, and more fundamental, argument against the use of QG theory on Mars is given by Keeley and White [54, 123] summarised here as follows. By means of (2.17), the term $\nabla_p\Phi$ in (2.37) can be transformed to different vertical coordinates. For example, in potential temperature and height coordinate system, it appears as

$$\nabla_p\Phi = \nabla_\theta\Phi = \frac{1}{\rho}\nabla_z p.$$

The transformation of $\nabla_p\Phi$ to sigma-coordinates leads to the additional term

$$\nabla_p\Phi = \nabla_\sigma\Phi + \alpha\nabla_\sigma p,$$

where the hydrostatic equation has been used. As before, let us write the thermodynamic variables as [54]

$$A(\mathbf{x},\eta,t) = A_0(\eta) + A'(\mathbf{x},\eta,t) \tag{2.50}$$

where A' represents a deviation from the static standard profile A_0. Then, the geostrophic wind can be defined in sigma-coordinates as [54]

$$\mathbf{v}_{g,\sigma} = \frac{1}{f_0}\mathbf{k}\times\nabla_\sigma\psi, \tag{2.51}$$

where $\psi = \Phi' + \alpha_0 p'$. Assuming a condition of no flow ($\Phi_p = 0$) in the presence of mountains would lead to [54]

$$\nabla_p\Phi'_p = \nabla_\sigma(\Phi'_\sigma + \alpha_0 p') + \alpha'\nabla_\sigma p' = 0. \tag{2.52}$$

Since a null deviation from the static standard profile in isobaric coordinates would

correspond to a non-zero deviation in sigma-coordinates, (2.52) can be interpreted as the impossibility of decomposing the thermodynamic variables in the form prescribed by (2.50) [54], Hence, a null geostrophic wind in isobaric coordinates would correspond to a non-zero geostrophic velocity in sigma-coordinates

$$\mathbf{v}_{g,\sigma} = -\frac{\alpha'}{f_0}\mathbf{k} \times \nabla_\sigma p'. \tag{2.53}$$

This result implies an ageostrophic wind equal in magnitude but in opposite sign to (2.53), which is contrary to the assumption (2.48) [54]. These disadvantages can still be overcome if topographic variations are reasonably small. Unfortunately, this is not the case in Mars.

Given the higher numerical value of the Rossby number on Mars than on the Earth and the presence of large Martian topographic features, the two arguments above tend to indicate that QG theory is not a good approximation to be applied on Mars. These arguments do not invalidate previous work on reduced atmospheric models for Mars based on QG theory [124, 125] as they were carried out using simplified models on a Mars-like planet without topography. However, a more realistic model cannot neglect these important Martian characteristic. Therefore, the subsequent development of this thesis will be based on the primitive equations (2.32)-(2.35).

2.3 General circulation models

General circulation models (GCMs) are the most comprehensive models of the global circulation of a planetary atmosphere. In general, they are based on the primitive equations (2.32)-(2.35). Although the hydrostatic equation (2.35) is not essential, it is often included so that vertical sound waves, which are irrelevant for large-scale motion, are filtered out [130]. Still, the fastest waves allowed in a primitive equation model travel at the speed of sound given by $c_s = (\gamma RT)^{1/2}$, where γ is the ratio of specific heat capacities c_v/c_p. These waves are the so-called Lamb waves, which are characterised by a purely horizontal motion ($w = 0$) [130]. The maximum phase velocity allowed in a partial differential model has important consequences for any attempt at numerically integrating it as we shall see in the next section.

2.3.1 Numerical methods and GCMs

The integration of the primitive equations in a computer necessarily requires some kind of discretisation. The first attempt to simulate the general circulation by means of a numerical integration of the primitive equations is due to Richardson [92], who essentially used the equations in the form they are used today with height as vertical coordinate [92]. According to Lynch [72], his integration, carried out in 1922, failed because of the lack of balancing of the initial conditions to filter out high frequency gravity waves and the lack of consideration of the Courant-Friedrich-Lewy (CFL) condition for numerical convergence given by the inequality

$$\frac{U\Delta t}{\Delta x} \leq 1, \tag{2.54}$$

where Δx is the separation between grid points, Δt is the time step and U is the speed of the fastest wave allowed by the model. The CFL condition was discovered in 1928 [80, 130].

Most currently operational GCMs are based either on finite-differences [9] or spectral methods [47] to perform the discretisation in the horizontal direction. However these are not the only available discretisation methods. Two examples of alternative discretisation methods are the box [55] and finite-volume [6] methods. In the vertical direction, most models use finite-differences although there have been attempts at using spectral vertical approximations (see [47]).

Finite-difference models

The approach of finite-difference models to the solution of (2.32)-(2.35) consists of the division of physical space in a set of grid points covering the whole planetary surface. The atmospheric fields are approximated at these grid points and the derivatives involved in the equations are approximated by finite-differences [80].

A particularly relevant scheme for this thesis is the distribution of the discrete atmospheric fields over the physical grid points over a staggered Arakawa C-grid [9] (see also [100]). Figure 2.1 shows the distribution of variables over a C-grid. This distribution has shown to be advantageous at simulating geostrophic adjustment [9]. If vorticity and streamfunction are placed over potential-vorticity points, the scheme also caters for an accurate representation of quasi-geostrophic flow, conserving both kinetic energy and total enstrophy (defined as one half of the square vorticity) in the shallow-water approximation [9]. This distribution of variables is widely used by

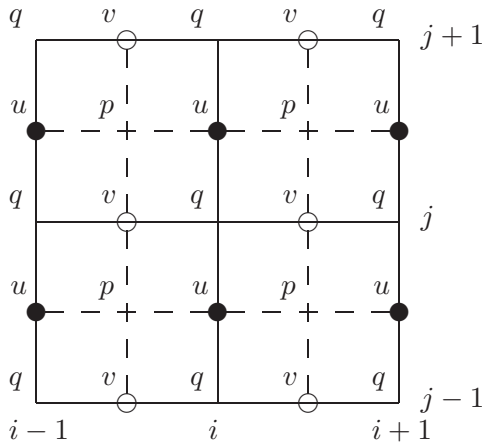


Figure 2.1: Arakawa C-grid. u and v denote zonal and meridional velocity, respectively and q is potential vorticity.

many finite-differences models [36, 109] and, following [3], this is the distribution used to derive the models in the present thesis.

Spectral models

Spherical harmonics constitute a set of orthonormal function over the sphere. They are defined as

$$Y_n^m(\lambda,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}\, P_n^m(\sin\phi)\, e^{im\lambda}, \tag{2.55}$$

where n represents total wavenumber, m is zonal wavenumber and P_n^m are the associated Legendre functions. In spectral models, atmospheric fields are decomposed in truncated series of spherical harmonics

$$X(\lambda,\phi,t) = \sum_{m=-M}^{M}\sum_{n=|m|}^{N_m} X_n^m(t)\, Y_n^m(\lambda,\phi), \tag{2.56}$$

where the actual limits depend on the truncation scheme to be used. There are two different truncation schemes: triangular ($N_m = M$) and rhomboidal ($N_m = |m| + M$). These are usually referred to by the letters T (for triangular) and R (for rhomboidal) followed by the maximum zonal wavenumber M allowed in the model. For example, a model at triangular truncation allowing a maximum zonal wavenumber 21 is referred to as T21. There is one alternative version in the case of triangular truncation: the jagged triangular truncation, where there are as many even-n as odd-n components

for each zonal wavenumber. The spectral coefficients are computed as

$$X_n^m(t) = \int_\Omega X(\lambda, \phi, t) Y_n^{m*}(\lambda, \phi) d\Omega, \tag{2.57}$$

where $d\Omega = \cos\phi \, d\lambda \, d\phi$. Generally, the prognostic variables in this kind of model are vorticity, divergence, absolute temperature and the logarithm of surface pressure because the resulting equations are more convenient for the application of the spectral scheme than the equivalent formulation in terms of the horizontal components of velocity [20].

Spectral models have various advantages over finite-difference models. Numerical dispersion and aliasing, which represent serious problems in finite-difference models, are not present in spectral models. Therefore, the evaluation of horizontal advection terms can be computed highly accurately [21]. Furthermore, spherical harmonics are defined over the sphere. Hence, unlike finite-difference models [9, 10], spectral models do not require special treatment at the poles [21]. However, these models were not popular until the development of the transform method [20, 86], which constituted a key element for their success. The transform method consists of the computation of nonlinear terms in physical space by transforming atmospheric fields from spectral space into grid points (at a convenient resolution so that aliasing is avoided) [20, 21, 47]. Once the nonlinear terms are calculated, the resulting tendencies are transformed back into spectral space.

Vertical coordinates

Although there were early models that used height as vertical coordinate [53], in general this coordinate is chosen to be a terrain-following coordinate such as sigma [88] or hybrid coordinates [101]. Independently of whether the horizontal discretisation scheme is finite-difference or spectral, the vertical coordinate is generally discretised using a finite-difference approach.

As in the case of horizontal finite-difference schemes, a staggered grid is also useful in the vertical direction for achieving energy and angular-momentum conserving schemes [9, 108]. Figure 2.2 shows the energy-conserving distribution in a K-layer discretisation in sigma coordinates suggested by Arakawa and Lamb [9]. The layer boundary levels are labelled with half numbers. Sigma-velocity is computed on these levels. The rest of the variables are placed over full levels. This is the vertical distribution of variables used in this thesis.

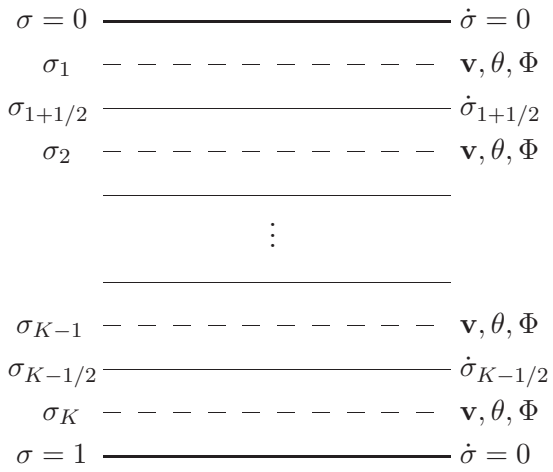


Figure 2.2: Vertical distribution of variables and boundary conditions.

2.3.2 Mars GCMs

Terrestrial GCMs are numerous among which, mentioning just a few, we can find the National Center for Atmospheric Research (NCAR) global GCM [53], the Geophysical Fluid Dynamics Laboratory (GFDL) global spectral model [39], the UK Meteorological Office atmospheric GCM [112] and the various medium-range forecast models at the European Centre for Medium-Range Weather Forecast (ECMWF) [109]. For a more detailed review on terrestrial GCMs and their history the interested reader is referred to [33].

In the case of Mars, there are at least four operational Mars GCMs (MGCMs). Most of them have been adapted from terrestrial GCMs with the possible exception of the Oxford MGCM (see Section 2.4). In the following lines, we give a summary of the development of these MGCMs. A much more complete history of atmospheric modelling for Mars is given in [91].

The NASA Ames MGCM [61] is a modified version of the Mintz-Arakawa model for the Earth. As such, it started as a two-level finite-difference model in sigma-coordinates [49]. Despite the limited available information at that time, regarding, for example, atmospheric composition and topography on Mars, the NASA Ames MGCM proved to be invaluable in the study of Mars (e.g. [61, 89]).

The GFDL MGCM [127] was adapted from the GFDL SKYHI terrestrial MGCM [34] using the box method [55] on a non-staggered horizontal grid. In the vertical, the model uses hybrid coordinates with up to 40 layers covering around 88 km height from the planet's surface. This model has been used in the study of thermal tides on Mars (e.g. [127, 128]) as well as the Martian polar warming [126].

The LMD MGCM has been adapted from the LMD terrestrial GCM [48]. It is a finite-difference model in sigma-coordinates with 15 layers in its initial version [48]. This model is currently in parallel development together with the Oxford MGCM (to be reviewed in Section 2.4). These models have been used for many different applications including atmospheric waves analysis [27], atmospheric dust [83] and the development of a climate database for Mars [64].

More recently, the Mars Weather Research and Forecasting (WRF) model [93, 118] has been under development. This is part of a broader project seeking the construction of a general planetary atmosphere model with views to be equally applicable to Venus, Mars and Titan (the last one being a satellite rather than a planet). Intended applications of this model regarding Mars include dust and water and CO_2 cycle studies and climate change [118],

2.4 Oxford Mars GCM

The Oxford MGCM is a spectral model based on the model of Hoskins and Simmons [47]. Its immediate precursor was the simplified GCM (SGCM) [26]. This was a spectral model at T21 resolution with 10 uniformly spaced levels in sigma-coordinates [26]. It did not include topography and the physical parameterisation package consisted of Rayleigh friction and Newtonian cooling for heating. The model parameters and equilibrium temperature profile were tuned to resemble Mars [26].

The SGCM was used in the study of baroclinic waves on Mars so that it was possible to produce a regime diagram with friction and radiation timescales as independent parameters [26]. In this thesis, the SGCM was used as a test model. In fact, this model has already been analysed from the point of view of the proper orthogonal decomposition and Galerkin-projected reduced-order models by Whitehouse *et al.* [124, 125]. As pointed out in Section 2.2, the difference between those studies and this work is that in this thesis we follow a primitive-equation approach whereas Whitehouse *et al.* [124, 125] constructed low-order models taking a QG approach in combination with a vertical structure decomposition. Chapter 4 in this thesis is entirely devoted to the analysis of the SGCM. A more complete description of the model will be given there.

The next step towards the current release of the Oxford MGCM was supplementing the initial SGCM with a physical parameterisation package. This was based on the physical package in the LMD MGCM [48]. As mentioned in Section 2.3.2, since 1995 the Oxford MGCM has been in constant development together with the LMD

MGCM giving rise to an European MGCM with two different, complementary versions. In its current release, the Oxford MGCM is written so that the resolution in the horizontal as well as in the vertical is variable. A typical resolution is jagged T31 with 25 unevenly-spaced sigma-levels. The uneven spacing in the vertical allows for an enhanced resolution near the surface [36]. The uppermost full level is located at $\sigma = 0.000056$ approximately corresponding to 97.8 km considering a constant scale height $H = 10$ km. This vertical extension allows the model to capture the fully developed Martian Hadley cell, which can reach heights of around 80 km [126].

2.4.1 Topography

The huge topography on Mars plays an essential role on the evolution of the general circulation. For example, coupled with the diurnal tide, it excites non-migrating tides [131]. In particular, a westward travelling diurnal Kelvin wave and an eastward travelling wavenumber-3 diurnal wave arise from this interaction. The diurnal Kelvin wave seems to have an influence on processes such as the development of dust storms [79]. Therefore, it is important to have an accurate representation of the Martian orography. The Oxford MGCM uses the highly accurate Mars topography dataset as measured by the Mars Orbiter Laser Altimeter (MOLA) aboard the Mars Global Surveyor (MGS) [113].

2.4.2 Parameterisation of physical processes

The physical processes included in the model can be divided into four classes: radiative heat transfer, surface processes, sub-grid dynamics and carbon dioxide mass exchange. The radiative response of the atmosphere of Mars is primarily determined by carbon dioxide, dust, water vapour and CO_2 and water ice. Radiative effects from water vapour and ice are neglected in the Oxford MGCM although efforts to include these effects are currently ongoing work [36]. The parameterisation of CO_2 ice clouds is made by adjusting the surface emissivity under condensation conditions.

The radiative effects explicitly parameterised in the Oxford MGCM are those due to gaseous CO_2 and dust. The effect of gaseous CO_2 is parameterised considering separately the thermal infrared, dominated by the strong 15 μm band, and the near-infrared (4.3 μm band). In the upper atmosphere (above 70 km), departures from local thermodynamic equilibrium become important. These have recently been incorporated into the Oxford MGCM through a simple model though work is being done towards improvement [36, 91].

The effect of dust on the radiative properties of the atmosphere is treated differently depending on whether the MGCM is run as a free-standing model or as part of a data assimilation process (see Section 2.4.3 below). In the first case, the dust optical depth is prescribed as a smooth version of the observations from the Thermal Emission Spectrometer (TES) aboard the MGS. In the second case, the dust optical depth is periodically updated in accordance with these observations. In both cases, the shape of the dust distribution is prescribed.

Surface temperature is determined by direct solar insolation, thermal radiation from the atmosphere and the surface itself and turbulent fluxes as well as thermal conduction in the soil below the surface. This is simulated in the Oxford MGCM by an 11-layer model governed by a conduction equation for each vertical column. These processes depend on the surface distribution of albedo and thermal inertia as measured by spacecrafts.

Subgrid dynamics include turbulent processes in the planetary boundary layer (PBL), convection and parameterisation of gravity wave and small-scale topography drag. Turbulent mixing of horizontal velocity components and potential temperature in the PBL is governed by a turbulent closure scheme whose mixing coefficients are computed with the aid of an additional dynamic equation for turbulent kinetic energy. Convection is modelled by means of a standard energy-conserving rapid mixing in the presence of unstable temperature profiles. Gravity wave drag, mainly in the middle and upper atmosphere, and subgrid-scale topography drag, near the surface, are modelled by a joint scheme.

As explained in Section 2.4.3, the surface pressure on Mars displays large interannual variations mainly due to the exchange of carbon dioxide between the atmosphere and the polar ice caps during winter and summer. The CO_2 mass exchange scheme in the MGCM considers sublimation and condensation during local summer and winter, respectively. The interested reader is referred to [36, 91] where a detailed account of physical processes parameterisations and further references are given.

2.4.3 Applications

The Oxford MGCM has been used in a variety of applications. As mentioned before, together with the LMD MGCM, it was used to compile a climate database for Mars [64]. It has also been used in dust cycle studies [81, 82] as well as on predictability studies by means of breeding vectors [83]. More recently, this model has been used to perform data assimilation of temperature and dust measurements from the MGS-TES

instrument [77, 78]. In turn, the data assimilation into the Oxford MGCM has been applied to the analysis of thermal tides on Mars [63] and to investigate the interannual variability of dust storms on Mars [77, 79].

Chapter 3

Proper Orthogonal Decomposition

The proper orthogonal decomposition (POD) is a method used to extract those structures that contain most of the energy (in a general sense) in a given system. The appropriate definition of a generalised energy, such as variance, kinetic energy or total energy, will depend on the system under study. Variance can be useful when an arbitrary signal is being analysed whereas kinetic energy or total energy may be more relevant for physical systems. The use of POD in the context of turbulence was first proposed in 1967 by Lumley [70] and, although the acceptance of this technique as a useful tool was not immediate [13], it is now widely used in a variety of applications from which fluid dynamics (e.g. [37, 97, 98, 99]) and atmospheric physics (e.g. [3, 4, 74, 104, 105, 106, 107, 124, 125]) are only two examples.

3.1 Theoretical framework

The discussion in this section follows the work of Berkooz *et al.* [18] and Holmes *et al.* [44]. Let us suppose that we are interested in describing a system S whose state is defined by $\psi \in \mathcal{H}$. The set $\mathcal{H}$ is a Hilbert space, whose dimension may be either finite or infinite, with an inner product (χ, ψ), where χ is also a member of $\mathcal{H}$, and a norm given by $||\psi|| = (\psi, \psi)^{1/2}$. To keep our argument as general as possible we assume that $\mathcal{H}$ is defined over $\mathbb{C}$ and the inner product is of the form

$$(\chi, \psi) = \int_V \chi_i^*(x) E_{ij} \psi_j(x) dV, \tag{3.1}$$

where χ_i^* is the complex conjugate of χ_i. The integral in (3.1) is defined over the domain V where χ and ψ are also defined. V is not restricted to be a three-dimensional physical space, but may be any system of generalised coordinates. Accordingly $x \in V$

denotes those generalised coordinates collectively. The sub-indices i and j run from 1 to d, and Einstein's summation convention is assumed. Therefore, this definition is suitable for a space $\mathcal{H} = \mathcal{H}_d \otimes \mathcal{H}_\infty$, given as a tensor product of a vector space $\mathcal{H}_d$ with finite dimension d and an infinite-dimensional function space $\mathcal{H}_\infty$ [75]. The energy tensor E_{ij} defines the geometry of the space and, therefore, the relations between the components of ψ and χ. It is a metric tensor which depends on the system and the quantity to be maximised. Examples of this quantity can be variance, kinetic energy or potential energy and the choice depends on the intended application. Note that this inner product has the property

$$(\chi, \psi) = (\psi, \chi)^*, \tag{3.2}$$

the proof arising immediately from the direct application of (3.1).

There are two important special cases that may be recovered from (3.1):

(*i*) If $d = 1$ and $E_{11} = 1$, then we obtain the L^2 inner product for scalar functions.

(*ii*) If $d > 1$ and $E_{ij} = \delta_{ij}$, where δ_{ij} is the Kronecker delta, then the L^2 inner product for vector functions arises.

The first case is important for incompressible fluids studies since it is proportional to kinetic energy.

Let us denote by $\mathcal{K} = \{\psi^{(k)}\}_{k=1}^{M} \subset \mathcal{H}$ a set of realisations of S, an ensemble, containing a sufficiently large number M of elements. The actual number depends on the application and the minimum number of dimensions to be retained. There is no upper limit and, in fact, to include as many ensemble members as possible is advisable as this would tend to provide a better phase space sample. $\mathcal{K}$ might have been obtained from either experiments or numerical simulations. Assuming S is ergodic and M contains a sufficiently large number of elements, we can consider $\mathcal{K}$ as a representative subset of the phase space region $\mathcal{Q} \subset \mathcal{H}$ where the evolution of S takes place. We seek [44]

$$\phi \in \mathcal{H} \ : \ \max_{\phi} \frac{\langle |(\phi, \psi)|^2 \rangle}{||\phi||^2}, \tag{3.3}$$

where $|\cdot|$ represents modulus, $\langle \cdot \rangle$ indicates an ensemble average over $\mathcal{K}$ and $||\phi||$ is the norm of ϕ. The statement (3.3) expresses the requirement that the average projection of $\mathcal{K}$ onto ϕ is maximal. An alternative interpretation of (3.3) is that it requires the average angle between $\mathcal{K}$ and ϕ to be minimal.

Rewriting (3.3) as a variational problem it is possible to find a solution [44].

Defining a functional $J[\phi]$ as

$$J[\phi] = \langle |(\phi, \psi)|^2 \rangle + \lambda(||\phi||^2 - 1), \tag{3.4}$$

and looking for an extremal of (3.4) using variations of the form $\phi + \delta\xi$, where $\delta\xi \in \mathbb{R}$, the following condition arises [44]:

$$2\Re\{\langle(\xi, \psi)(\psi, \phi)\rangle - \lambda(\xi, \phi)\} = 0, \tag{3.5}$$

where the symbol $\Re\{\cdot\}$ indicates the real part of the argument, and the property (3.2) was used. Using (3.1) and since $\delta\xi$ is an arbitrary variation, (3.5) yields

$$\int_V \langle \psi_i(x)\psi_k^*(x')\rangle E_{kj}\phi_j(x')dV' = \lambda\phi_i(x). \tag{3.6}$$

This is an homogeneous Fredholm equation of the second kind for ϕ whose kernel

$$K_{ij}(x, x') = \langle \psi_i(x)\psi_k^*(x')\rangle E_{kj} \tag{3.7}$$

is not Hermitian since $K_{ij}(x, x') \neq K_{ji}^*(x', x)$. However, it can be brought to a Hermitian form by means of the linear transformation $U : \mathcal{H}_D \to \mathcal{H}_H$, defined by $\zeta_i = U_{ij}\psi_j$, where U_{ij} is given by the relation $E_{ij} = U_{ik}U_{kj}$ and $\zeta \in \mathcal{H}_H$ [11].

Using this transformation, (3.6) becomes

$$\int_V K_{ij}^H(x, x')\vartheta_j(x')dV' = \lambda\vartheta_i(x), \tag{3.8}$$

where $\vartheta_i = U_{ij}\phi_j$ and the Hermitianised kernel is given by

$$K_{ij}^H(x, x') = \langle \zeta_i(x)\zeta_j^*(x')\rangle. \tag{3.9}$$

According to Hilbert-Schmidt theory if the kernel is continuous, there is an infinite number of eigenfunctions $\vartheta^{(j)}$ and associated eigenvalues $\lambda^{(j)} \in \mathbb{R}$, where $\lambda^{(1)} \geq \lambda^{(2)} \geq \ldots \geq 0$ and $j = 1, \ldots, \infty$ [11]. Furthermore, the eigenfunctions $\vartheta^{(j)}$ are orthogonal under the L^2-norm, so that

$$(\vartheta^{(i)}, \vartheta^{(j)})_{L^2} = \int_V \vartheta_k^{(i)*}(x)\vartheta_k^{(j)}(x)dV = ||\vartheta^{(i)}||^2\delta_{ij}. \tag{3.10}$$

From here it is easy to show, by substituting the definition of ϑ back in (3.10), that the eigenfunctions of the original kernel are orthonormal under the inner product

(3.1):

$$(\phi^{(i)}, \phi^{(j)}) = \delta_{ij}, \tag{3.11}$$

where the eigenfunctions have been normalised. To derive (3.11) we used the fact that E_{ij} is a symmetric metric tensor.

We define Λ as the set of strictly positive eigenvalues of (3.6) and Φ as the set of associated eigenvectors. We neglect those eigenvalues whose value is zero because they contain no energy in average and are, therefore, not important for the evolution of the system. Λ and Φ are generally called the sets of empirical eigenvalues and empirical orthogonal functions (EOFs), respectively. Additionally we define the Hermitianised EOFs as $\Phi^H = U(\Phi)$ which may be used to prove any other property of Φ using the well-established theory for Hermitian kernels [44]. In particular, it is possible to show that $\text{span}(\Phi) = \text{span}(\mathcal{K})$. Therefore, Φ may be considered as a basis for $\mathcal{Q}$ rather than for $\mathcal{H}$. The expansion of the elements of $\mathcal{K}$ in terms of the EOFs,

$$\psi = \sum_{j=1}^{\infty} a_j \phi^{(j)}, \tag{3.12}$$

is indeed what is called the POD of $\psi^{(k)}$. We will use this name in general to refer to the decomposition of the elements of $\mathcal{Q}$ so that we can approximate the state vector by means of the finite sum

$$\psi^N = \sum_{j=1}^{N} a_j \phi^{(j)}. \tag{3.13}$$

This operation can be viewed as the projection of the infinite-dimensional $\mathcal{H}$ onto an space of finite dimension N. The energy contained by the approximate state vector ψ^N is, on average, [44]

$$\langle E^N \rangle = \sum_{j=1}^{N} \lambda^{(j)}. \tag{3.14}$$

The upper index N in the last two expressions indicates the order of the approximation. It is possible to show that a decomposition in terms of EOFs converges more rapidly than any other linear expansion [18, 44]. This energy content is, therefore, the largest possible. The only part that remains to be determined are the coefficients a_j. By taking the inner product of $\phi^{(k)}$ and either (3.12) or (3.13) we obtain

$$a_k = (\phi^{(k)}, \psi). \tag{3.15}$$

These time-dependent coefficients are called the principal components (PCs) of ψ and

contain all the information about the evolution of the system S.

Although the POD is a linear procedure, the method does not make any assumptions on the linear properties of the system to be represented [18]. In this sense, decomposing a function in terms of EOFs is as general as decomposing in terms of Fourier modes or Chebyshev polynomials. The advantage over these two possible bases (among others) is the information about the specific system that is contained by Φ. Moreover, in the case when more than one scalar variable is necessary to describe the system, Φ contains information on the interactions between them. This is a fundamental difference with respect to ordinary expansions in terms of more general functions and gives the POD an additional advantage: it contains information about the correlated behaviour between the different state variables. Due to this correlation, if the system under study is nonlinear, the EOFs will contain information about the nonlinear character of the system.

3.2 Numerical implementation

So far we have been concerned with the general case of vector continuous functions. In practice, however, one always deals with functions evaluated only at a finite number of spatial points or expanded as a finite series in terms of a general basis such as Fourier modes, Legendre associated functions or Chebyshev polynomials. Therefore, it is important to study the changes that this introduces into the picture above.

The generality of the formalism allows us to apply it to the new case although not before taking into account certain considerations. In order to see this, suppose that the components ψ_i are discretised on a set of N physical points,

$$\psi_i^j = \psi_i(x_j), \qquad j = 1, \ldots, N. \tag{3.16}$$

Alternatively, they might be given as a finite series in terms of a certain orthonormal basis $\{\eta^{(j)}\}_{j=1}^N$,

$$\psi_i = \sum_{j=1}^{N} \psi_i^j \eta^{(j)}. \tag{3.17}$$

In both cases the effective result is the change of the infinite dimensional space $\mathcal{H}_\infty$ to the N-dimensional one $\mathcal{H}_N$. Thus, the resulting space has $D \times N$ dimensions.

The integral in (3.1) may be approximated by a finite sum of the form

$$\int_V \sum_{i,j} \chi_i^*(x) E_{ij} \psi_j(x) dV \simeq \sum_{k,l} \sum_{i,j} \chi_i^{k*} E_{ij}^{kl} \psi_j^l, \tag{3.18}$$

where the definition of E_{ij}^{kl} depends on the discretisation scheme. If (3.16) is used, then $E_{ij}^{kl} = E_{ij} \Delta V^{kl}$, where ΔV^{kl} is a diagonal matrix of discrete volumes; on the other hand, if ψ_i is approximated by (3.17), then $E_{ij}^{kl} = E_{ij} \delta^{kl}$. The upper indices indicate operations in the space $\mathcal{H}_N$ in both definitions. Exchanging each pair of indices $\{ {}_i^k \}$ with a single index $\{\nu\}$ we recover the whole formalism applied to purely discrete systems. However, it is sometimes useful to keep the treatment in terms of tensor products of vector spaces since it is easier to keep clear the meaning of each entry in the state vector.

3.2.1 Sirovich's snapshots method

Let us move now onto the numerical implementation of the POD. To keep everything as simple as possible we work first in the Hermitianised space $\mathcal{H}^H$ and then extend those results to the original space $\mathcal{H}$. Since $\mathcal{K}$ is a set of realisations of S at M different instants where M is a sufficiently large number, it is possible to approximate the kernel K_{ij}^H as

$$K_{ij}^H(x, x') = \frac{1}{M} \sum_{k=1}^{M} \zeta_i^{(k)}(x) \zeta_j^{(k)*}(x'), \tag{3.19}$$

and, since $\mathrm{span}(\Phi) = \mathrm{span}(\mathcal{K})$, it is possible to write the EOFs as

$$\vartheta^{(j)}(x) = \sum_{k=1}^{M} b_k^{(j)} \zeta^{(k)}(x), \tag{3.20}$$

assuming that the realisations $\{\zeta^{(k)}\}_{k=1}^M$, and therefore $\{\psi^{(k)}\}_{k=1}^M$, are linearly independent. Substituting (3.19) and (3.20) into (3.8) gives

$$C_{lk} b_k^{(m)} = \lambda^{(m)} b_l^{(m)}, \tag{3.21}$$

where

$$C_{lk} = \frac{1}{M} (\zeta^{(l)}, \zeta^{(k)})_{L^2}. \tag{3.22}$$

This is the method of snapshots introduced by Sirovich in 1987 [110]. It enables us to transform an eigenvalue problem which in principle would involve a matrix of

order $D \times N = O(10^5)$ (after discretising or expanding the continuous fields) into a new eigenvalue problem which involves a matrix of order $M = O(10^3)$. This is advantageous since (as can be seen from the numbers given) typically $M \ll D \times N$. If this last relation did not hold, the direct method [110], consisting of the direct numerical evaluation of (3.6) or (3.8), should be chosen.

Pre-multiplying (3.20) by U^{-1}, the inverse of U, and using the definition of ζ we obtain

$$\phi^{(j)}(x) = \sum_{k=1}^{M} b_k^{(j)} \psi^{(k)}(x). \tag{3.23}$$

Using the definition of ζ in (3.22), we also obtain

$$C_{lk} = \frac{1}{M}(\psi^{(l)}, \psi^{(k)}). \tag{3.24}$$

Relations (3.23) and (3.24) show that the coefficients $b_k^{(j)}$ are the same in both the original and the Hermitianised spaces. Therefore, it is equivalent to work either in the original space under the norm induced by (3.1) or in the Hermitianised space under the L^2-norm .

We have now a general theoretical framework that can be applied to the case of an arbitrary energy tensor. In the following section we will address the important question of the appropriate form of this tensor when the formalism is employed to analyse a system such as the case of interest here, a planetary atmosphere.

3.3 Energy and metrics

The choice of an appropriate inner product and a norm is of central importance to the POD because it determines the quantity to be maximised. For some problems the choice is almost obvious. In the case of isothermal incompressible flow, for example, the inner product is defined as

$$(\mathbf{v}_1, \mathbf{v}_2) = \int_\Omega \mathbf{v}_1 \cdot \mathbf{v}_2 \, dV \tag{3.25}$$

The norm induced by (3.25) is the variance which is directly proportional to kinetic energy. Its physical interpretation is, therefore, straightforward.

In contrast, the appropriate choice of an inner product when the variables describing the system have different dynamical character is still not clear and remains an open question. In this case the inner product definition depends on the relevant

variables and the phenomena to be studied. This is the situation, for example, in studies of compressible flow, where the velocity is insufficient to completely describe the system, and additional thermal variables, such as temperature or entropy, are required.

In the case of compressible flow, for instance, one can use a state space formed by velocity $\mathbf{v}$ and sound velocity c_s, the latter being related to the temperature in an ideal gas by $c_s = (\gamma RT)^{1/2}$, where γ is the specific heats ratio c_p/c_v . Rowley *et al.* [98] have suggested as a suitable alternative inner product

$$(q_1, q_2)_\alpha = \int_\Omega \mathbf{v}_1 \cdot \mathbf{v}_2 + \frac{2\alpha}{\gamma(\gamma-1)}\, c_{s1} c_{s2}\, dV, \tag{3.26}$$

where the state vectors are defined to be $q = (\mathbf{v}, c_s)$. The constant α can be either 1, in which case the induced norm is proportional to the total enthalpy, or γ, in which case the norm is proportional to the total energy. In proposing this inner product, an underlying assumption of constant density has been made. This approximation is valid when the fluid's velocity is small in comparison with the sound velocity and temperature gradients are small [98].

In the case of atmospheric dynamics, the total energy (TE) norm has been shown to be an appropriate choice [3, 124]. The total atmospheric energy in sigma-coordinates (assuming hydrostatic stability) is given by

$$E = \frac{1}{g} \int_V p^s \left(\frac{\mathbf{v} \cdot \mathbf{v}}{2} + c_p T \right) dV, \tag{3.27}$$

where $dV = a^2 \cos\phi \, d\phi \, d\lambda \, d\sigma$ is a volume element in spherical coordinates with σ acting as vertical coordinate. Since we want to identify TE with the norm it is necessary to express it as a quadratic form of the state variables. This aim can be achieved following the suggestion of Achatz and Opsteegh [3] with a slight modification to make the vector state dimensionally consistent. We therefore introduce the variable $\tau = \sqrt{R\theta}$ and, using the appropriate non-dimensionalising factors from table 2.1, write the non-dimensional energy as

$$E = \int_V p^s \left(\frac{\mathbf{v} \cdot \mathbf{v}}{2} + \frac{P\tau^2}{\kappa} \right) dV, \tag{3.28}$$

where $P = (p/p_0)^\kappa$ and $dV = \cos\phi \, d\phi \, d\lambda \, d\sigma$. The energy scale is $E_0 = (a^2\Omega)^2 p_0/g = 89.82$ EJ (1 EJ $= 10^{18}$ J.). Defining the state vector as $\psi = (\mathbf{v}, \tau)^T$, equation (3.28)

can be identified with a squared norm

$$E = ||\psi||^2 = (\psi, \psi). \tag{3.29}$$

This leads directly to the definition of the inner product as

$$(\psi_1, \psi_2) = \int_V p^s \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{2} + \frac{P\tau_1\tau_2}{\kappa} \right) dV, \tag{3.30}$$

where the surface pressure p^s is assumed to be time-independent. This is the additional assumption mentioned at the end of Section 2.1. It leads to the equations of motion presented in Section 5.1. With this definition of the inner product, the energy tensor $\mathbf{E}$ becomes

$$\mathbf{E}(\lambda, \phi, \sigma) = \frac{1}{2\kappa} \begin{pmatrix} \kappa p^s(\lambda, \phi) & 0 \\ 0 & 2p^s(\lambda, \phi)P(\lambda, \phi, \sigma) \end{pmatrix}. \tag{3.31}$$

3.3.1 Further remarks

There are two points that deserve further comment. The first one concerns the definition of TE in atmospheric dynamics. It can be argued that this definition should be based on available potential energy (APE) rather that total potential energy (TPE) since this is the only fraction of potential energy available to be transformed into kinetic energy [66, 67]. However, APE is only well-defined over isentropic or barotropic surfaces [66] while neither of these constitute a terrain-following coordinate system. On the other hand, the choice of TPE allows the use of terrain-following sigma coordinates without major modifications. Still, we will see in the following chapters that the thermal field splits into two easily distinguishable fields. The first, explained by the first (and most energetic) EOF, corresponds to the background potential energy unavailable to be transformed into kinetic energy. The second, comparable with APE, corresponds to higher order EOFs. We will see that indeed this part clearly interacts with velocity fields originating an energy exchange between them.

The second point that deserves further comment is the assumption of time independence of surface pressure. This assumption is required to give a proper definition to (3.30) since a time-dependent inner product would avoid the EOFs being global atemporal system representatives. However, this assumption can be a very strong constraint in the case of the Martian atmosphere where the surface pressure can present variations up to 30% peak-to-peak throughout a Martian year [91]. There are three main reasons for this variation: the condensation-sublimation of CO_2 in the

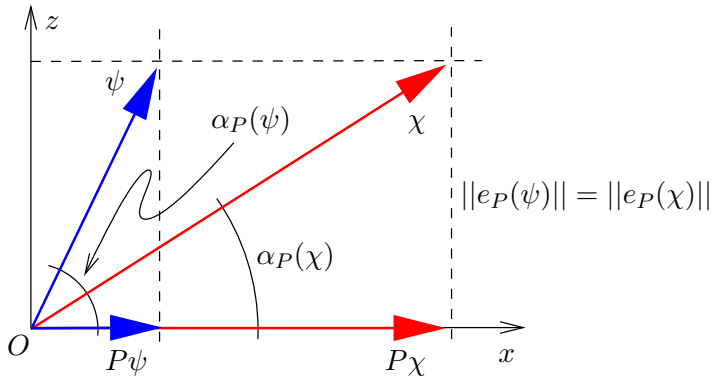


Figure 3.1: Subspace approximation error and angles between ψ and χ and their projections in an orthogonal space.

polar caps, the large latitudinal variations of topography and the dynamical effect of geostrophic balance [48]. The condensation-sublimation process might be taken into account by including the contribution of the ice caps to internal energy to the TE expression. However, this is left as a next step for developing long-term atmospheric models. Instead, in order to avoid a possible inconsistency coming from the assumption of time-independent surface pressure, this study is restricted to a seasonal analysis. The considered period is the transition from autumn to winter in the northern hemisphere since it is during this interval when the baroclinic activity is enhanced.

3.4 Accuracy

The projection error is given by [46]

$$e_P(\psi) = P\psi - \psi, \tag{3.32}$$

where P is the projection operator onto the span of the leading EOFs. From its definition it can be interpreted as the distance from a vector to its projection. It has been used to evaluate the accuracy of a projection (see, for example, [46, 90]). However, when the phase space is spanned by an orthogonal basis, that measurement may be misleading since two vectors can have the same projection error. For example, figure 3.1 shows two (two-dimensional) vectors ψ and χ in the (x, z) space and their (one-dimensional) projections onto the x subspace.

In this case, a better way to evaluate the accuracy of the projection is to look at a relative distance, namely the quotient of the projection to the norm of the original

vector. In fact, the mathematical formulation of the POD problem (3.3) is given in these terms. A more elegant way of expressing this is by a proximity angle defined to be the angle between the original vector and its projection

$$\alpha_P(\psi) = \text{ang}(\psi, P\psi) = \cos^{-1}\left(\frac{(\psi, P\psi)}{||\psi||\,||P\psi||}\right). \tag{3.33}$$

Since the basis is orthogonal, $P\psi$ and $e_P(\psi)$ are orthogonal, which can be verified by expanding in term of EOFs

$$P\psi = \sum_{k=1}^{N} a_k \phi^{(k)}, \quad e_P(\psi) = \sum_{k=N+1}^{\infty} a_k \phi^{(k)},$$

so that the inner product becomes

$$(P\psi, e_P(\psi)) = \sum_{k=1}^{N}\sum_{l=N+1}^{\infty} a_k^* a_l \delta^{(kl)} \equiv 0, \tag{3.34}$$

which provides the desired result. This implies a Pythagorean relation between the original vector, its projection and the error between them. Indeed, using (3.32) and (3.34), we have

$$||\psi||^2 = (P\psi + e_P(\psi), P\psi + e_P(\psi)) = ||P\psi||^2 + ||e_P(\psi)||^2. \tag{3.35}$$

In turn, in order to establish a relation between the norm of the projection error and the proximity angle we can write

$$||e_P(\psi)||^2 = (\psi - P\psi, e_P(\psi)) = ||\psi||^2 - ||\psi||\,||P\psi||\cos\alpha_P(\psi),$$

where (3.32) and (3.34) have been used. Solving for $\cos\alpha_P(\psi)$ and using (3.35) we finally obtain

$$\cos\alpha_P(\psi) = \left(1 - \frac{||e_P(\psi)||^2}{||\psi||^2}\right)^{1/2}, \tag{3.36}$$

which shows that $\cos\alpha_P(\psi)$ solely depends on the relative error. Furthermore, if the relative error is small so that the angle is also close to zero we have

$$\cos\alpha_P(\psi) = 1 - \frac{1}{2}\alpha_P^2(\psi) + \ldots = 1 - \frac{1}{2}\frac{||e_P(\psi)||^2}{||\psi||^2} - \ldots$$

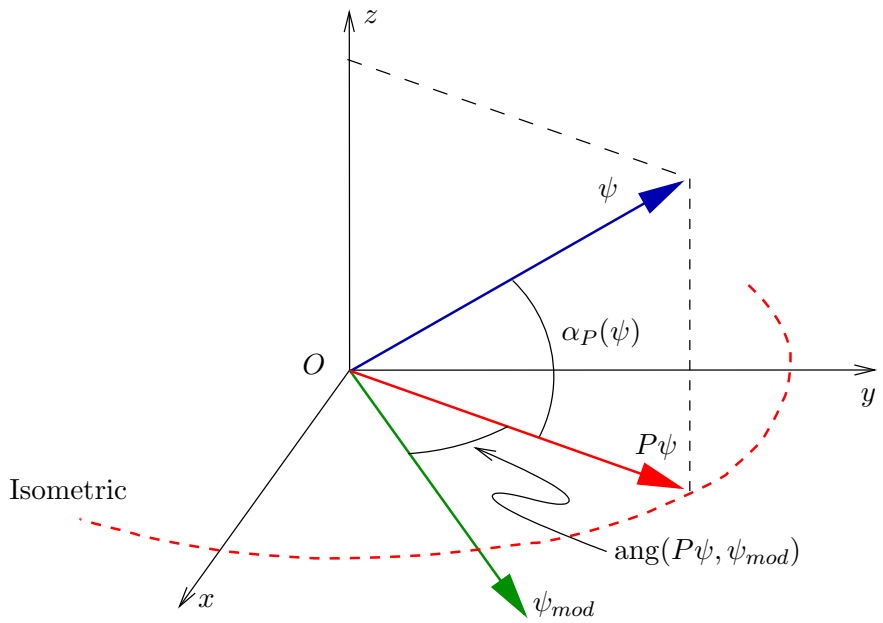


Figure 3.2: A vector ψ in a three-dimensional orthogonal phase space, its projection $P\psi$ on the x-y subspace and a predicted vector ψ_{mod} from the integration of a hypothetical reduced model acting on the x-y subspace. The isometric is the locus of vectors of equal norm in the reduced space.

which immediately leads to the conclusion that

$$\alpha_P(\psi) \sim \frac{||e_P(\psi)||}{||\psi||}.$$

We can now define the projection angle as the angle between the original phase space $\mathcal{Q}$ and the projection hyperplane $\mathscr{P}$ given by the ensemble average of (3.33)

$$\alpha_{\mathscr{P}}(\mathcal{Q}) = \text{ang}(\mathcal{Q}, \mathscr{P}) = \langle \alpha_P(\psi) \rangle.$$

This quantity provides a global measure useful at assessing the level of approximation reached by a given dimension reduction: the smaller the projection angle, the better the approximation. In fact, this is just another way of mathematically stating the POD problem. This approach has also been used in other contexts such as document comparison and information filtering (see [17], for example).

When comparing two vectors that lie within the same reduced space, a single parameter is not enough to describe the error between them because of the lack of a Pythagorean relation between the involved vectors. This situation, depicted in figure 3.2, can be encountered when comparing the integration of a reduced model

ψ_{mod} with the projection of the corresponding *true trajectory*[1] ψ. In this case, in addition to the angle between vectors, the ratio between norms

$$r(\psi_{mod}, P\psi) = \frac{||\psi_{mod}||}{||P\psi||} \tag{3.37}$$

can be used in order to compare the sizes of both vector. These two measures provide a full comparative picture with information on direction and amplitude errors.

The present chapter has been completely devoted to the theoretical and numerical ideas behind the proper orthogonal decomposition. In the next two chapters we shall apply these ideas to the particular case of atmospheric dynamics on Mars in the two different models presented in Chapter 2. First, we shall discuss the SGCM to develop a framework from which to analyse the more complex MGCM.

[1]The true trajectory would be the trajectory produced by a full model (a GCM) in the full phase space starting from initial condition whose projection on the reduced space would produce the initial conditions for the low-order model.

Chapter 4

POD of the simplified GCM

As described in Section 2.4, the Oxford Mars GCM [36] is a state-of-the-art atmospheric model which includes the parameterisation of a number of relevant physical processes. All these parameterisations contribute to the forcing of the system in a way that is not always easy to discern. Explicitly including all those physical processes into a reduced-order model would be contrary to the spirit of such a model.

In this work, the aim of a reduced-order model is to retain the most important features of the system with the smallest number of modes. The simplified GCM (SGCM) was investigated as a means for validating a reduced model. Newtonian cooling and Rayleigh friction are sufficiently simple schemes to be implemented straightforwardly in a dynamical reduced-order model. These are precisely the friction and radiative schemes incorporated into the SGCM. Hence, this model naturally arises as a good benchmark to evaluate the performance of a dynamical reduced model. Furthermore, the SGCM was previously studied by Whitehouse et al. [124, 125] using the POD-Galerkin method and QG theory. This gives a possibility of comparison and validation for our results. This chapter will be focused on the application of the POD on an SGCM time series and the analysis of the so-obtained EOFs.

4.1 SGCM

The original dataset from the SGCM [26, 27] consists of a simulation in perennial winter in the northern hemisphere carried out for 200 sols at a rate of $s = 12$ samples/sol constituting an ensemble of 2400 elements. Both the SGCM and the MGCM produce time series of the spectral coefficients of absolute vorticity, divergence, temperature and (logarithm of) surface pressure. The resolution of the original model is T21 (jagged triangular truncation with maximum total wavenumber $n_{\max} = 21$)

Table 4.1: Sigma-levels and equivalent altitudes in the SGCM dataset. The equivalent altitudes shown are computed as $z = -H \ln \sigma$, where $H = 10$ km (see text).

Level	σ	z (km)	Level	σ	z (km)
1	0.05	29.957	6	0.55	5.978
2	0.15	18.971	7	0.65	4.308
3	0.25	13.863	8	0.75	2.877
4	0.35	10.498	9	0.85	1.625
5	0.45	7.985	10	0.95	0.513

over ten equally-spaced sigma-levels, shown in Table 4.1. The equivalent altitudes are computed as $z = -H \ln \sigma$, where $H = 10$ km is a constant scale height. Recalling that $\sigma = p/p^s$, where p is pressure and p^s is surface pressure, they must be interpreted as heights with respect to the planetary surface. This convention will be used throughout the thesis unless specified otherwise.

The thermal forcing in the SGCM is given by Newtonian cooling while the frictional forces are represented by Rayleigh friction. Under the Newtonian cooling assumption, the volumetric diabatic heat rate is given in non-dimensional form by $Q = (T_* - T)/\tau_r$, where the system is relaxed towards the axisymmetric state $T_* = T_*(\phi, \sigma)$ shown in figure 4.1 with a constant relaxation time of $\tau_r = 2.0$ sols. Using τ as the relevant variable, the volumetric heat rate takes the form $Q = (\sigma p^s)^\kappa(\tau_*^2 - \tau^2)/\tau_r$ in dimensionless form. The Rayleigh friction model assumes the friction to be given by $\mathbf{F} = -r\mathbf{v}$, where $r = r(\sigma)$ is the Rayleigh friction factor. In the model, it is considered to act only in the lowest layer ($\sigma = 0.95$), where it has a value $r(0.95) = 1/3$ sol^{-1}, which corresponds to a timescale of $\tau_D = 3.0$ sols.

4.1.1 SGCM dataset climatology

Using friction and radiative relaxation times as parameters, Collins and James [26] were able to identify different regimes in the perennial southern hemisphere winter of the SGCM. In particular, they found what they called weak waves (small-amplitude disturbances over a zonally symmetric flow) and waves with zonal wavenumber three [26]. The wavenumber-3 wave was later the subject of a POD analysis carried out by Whitehouse *et al.* [124]. We shall discuss this work later in this chapter.

In accordance with their regime diagram for the SGCM [26], the values of τ_r and τ_D used in this work (2.0 and 3.0 sols, respectively) would place the model in a wave 3 regime. Figure 4.2 shows that this is not the case. In it we can see four

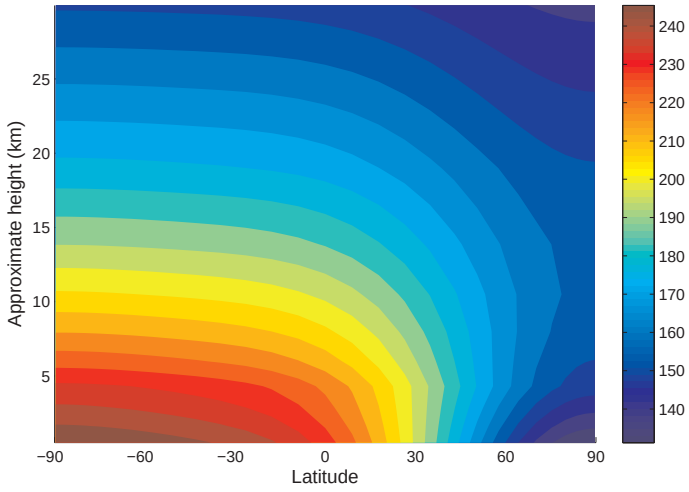
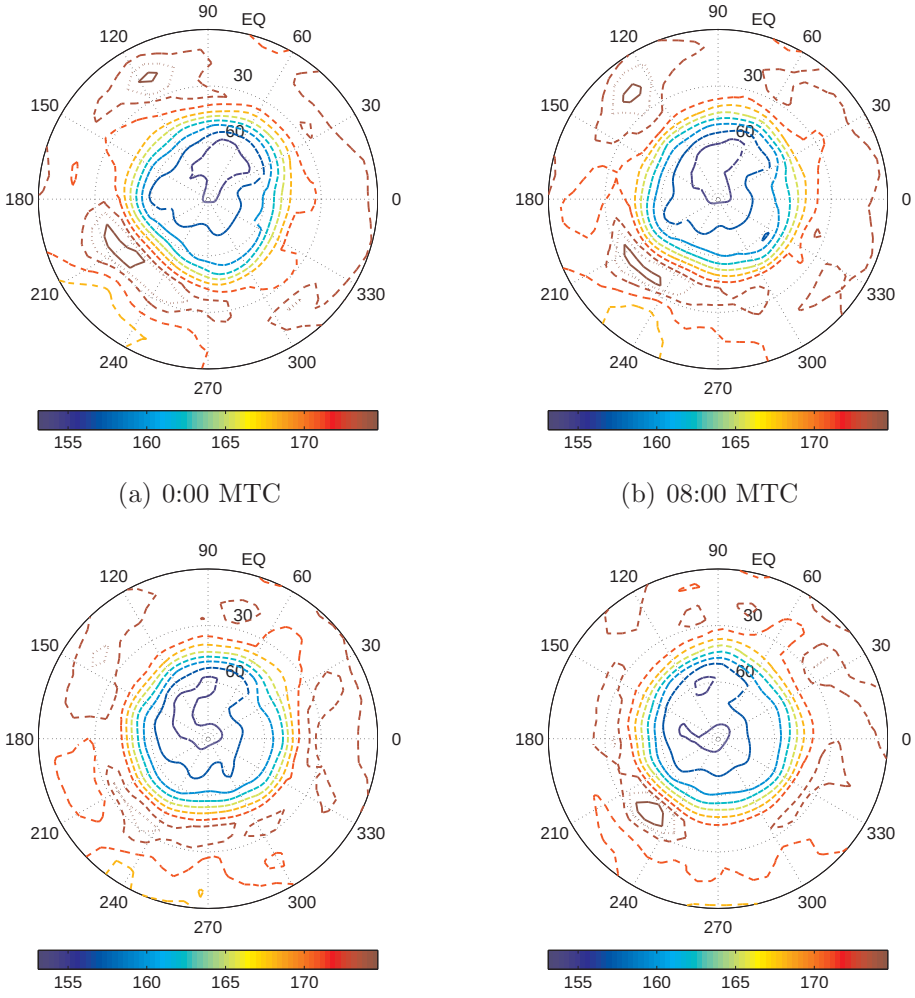


Figure 4.1: Latitude-height section showing the axisymmetric equilibrium temperature profile, measure in K, to which the SGCM was relaxed.

temperature snapshots at an approximate height $z = 18.97$ km above the planet's surface. The patterns correspond to a weak wave rather than a wave 3, specially at mid-latitudes. In the tropical region, the patterns display a more complex wave pattern. The difference in regime is not surprising if we bear in mind that this chapter is concerned with a perennial northern hemisphere winter. Moreover, we used a different equilibrium temperature profile to the one used in [26] and it has been shown that different equilibrium temperature profiles are likely to yield different regimes [26]. Considering that topography is not included, we can see this situation as having similar planets rotating in opposite directions.

Figure 4.3 shows the temporal and zonal average profiles of temperature, potential temperature and zonal velocity in the original SGCM dataset. As expected, the temperature profile (figure 4.3(a)) resembles the equilibrium temperature to which the atmosphere is being relaxed (cf. figure 4.1) with characteristics typical of winter solstice: warmer temperatures in the southern hemisphere with small temperature gradients and strong temperature gradients towards the north pole. A key difference between Mars and the Earth is the atmospheric temperature distribution in the summer hemisphere. As shown in figure 4.3(a), maximal temperature is reached near the summer pole. As a result, a temperature gradient develops between high latitudes in the southern hemisphere and the north pole. This characteristic temperature distribution is mainly due to the lack of clouds allowing solar heating to reach the surface [40]. Another important factor is the lack of oceans which induces a lower surface thermal inertia [40, 91].



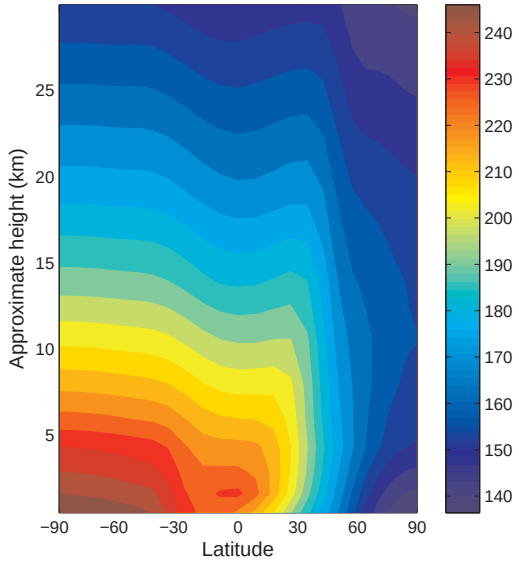
(a) 0:00 MTC

(b) 08:00 MTC

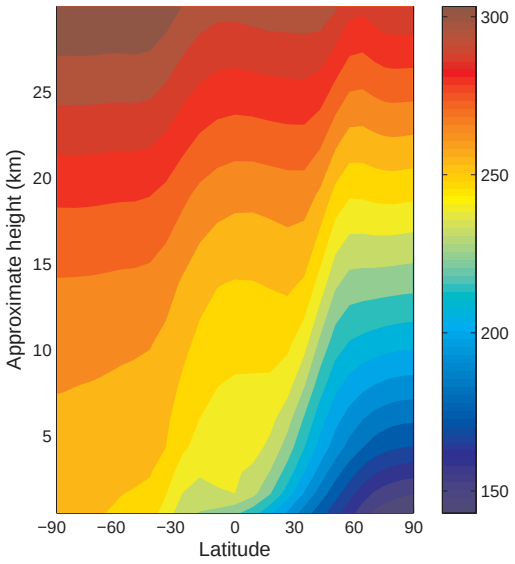
(c) 16:00 MTC

(d) 24:00 MTC

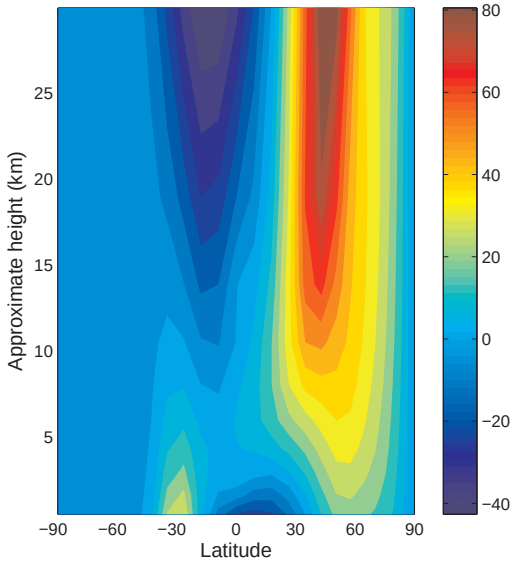
Figure 4.2: Temperature, measured in K, in the original dataset at four different times (indicated in each frame) during one sol at $\sigma = 0.15$ (approximate height $z \simeq 18.97$ km). MTC stands for Coordinated Mars Time, the Martian equivalent to the terrestrial Coordinated Universal Time (UTC) [7].



(a) Temperature



(b) Potential temperature



(c) Zonal velocity

Figure 4.3: (a) Temporal and zonal average of temperature, (b) potential temperature and (c) zonal velocity in the SGCM dataset, measured in K and m/s, respectively. Positive horizontal velocity directed eastwards.

Potential temperature (figure 4.3(b)) shows that the atmosphere's static stratification is stable everywhere, while zonal velocity (figure 4.3(c)) displays westerly jets in the middle atmosphere above the northern hemisphere as a consequence of the strong horizontal temperature gradients encountered in that region and it is at least partially explained by the thermal wind relation (see [130], for example)

$$\frac{\partial \mathbf{v}}{\partial p} = -\frac{R}{fp}\mathbf{k} \times \nabla_p T, \tag{4.1}$$

where geostrophic balance has been assumed. Hence, the validity of (4.1) is given by the smallness of the Rossby number (recall that the SGCM does not include topography and, therefore, the restrictions to the validity of QG theory due to this factor are not valid in this context). Equation (4.1) expresses that the zonal wind is expected to increases with height in the presence of a horizontal thermal gradient.

4.1.2 From spectral space to longitude-latitude grid

The data from the SGCM were transformed into time series of zonal and latitudinal velocity and square root of potential temperature on the 10 sigma-levels used in the SGCM. Each layer was discretised in a 64×32 regular longitude-latitude grid, corresponding to squares of angular length $5.625°$ in either direction and equivalent to a maximum arc length of $s = 333$ km.

In [9] Arakawa and Lamb discussed five different distributions of the atmospheric variables over a finite-difference grid. In that work they focused on the accuracy of these schemes when simulating geostrophic adjustment. They found that the so-called staggered Arakawa C-grid was the best choice under these criteria. This distribution was chosen in this thesis, following the work by Achatz and Opsteegh [3]. Thus, the variables were placed on the longitude-latitude grid according to the distribution shown in figure 4.4.

The variation in the global average surface pressure in the MGCM dataset is of only 0.03% (peak-to-peak) justifying the assumption of time-independent surface pressure. This assumption imposes conditions on the velocity components. Under these circumstances, sigma-velocity is

$$\dot{\sigma} = -\frac{1}{p^s}\int_0^\sigma \nabla \cdot (p^s \mathbf{v})\, d\sigma_1.$$

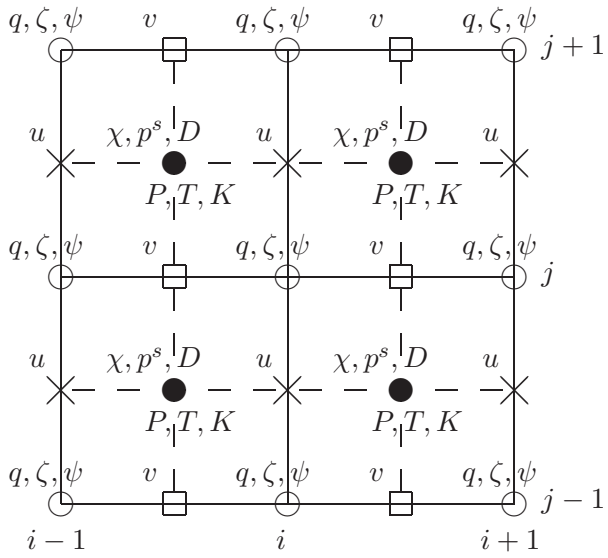


Figure 4.4: Horizontal distribution of variables.

With the boundary conditions $\dot{\sigma} = 0$ at $\sigma = 0$ and $\sigma = 1$, we conclude that

$$\nabla \cdot (p^s [\![\mathbf{v}]\!]) = 0, \tag{4.2}$$

where $[\![\cdot]\!] = \int_0^1 \cdot \, d\sigma$ so that $[\![\mathbf{v}]\!]$ gives the barotropic velocity.

Since surface pressure is time-dependent in both the SGCM and MGCM, (4.2) is not satisfied in general. To make the data consistent with a time-independent surface pressure, a pre-processing of both GCM time series was required. The procedure is now described.

The streamfunction ψ and velocity potential χ are computed by solving the Poisson equations $\nabla^2 \psi = \zeta - f$, where f is the Coriolis parameter, and $\nabla^2 \chi = D$ in spectral space to take advantage of the properties of the spherical harmonics, in particular, of their being eigenfunctions of the Laplacian in spherical coordinates. In terms of stream function and velocity potential, velocity is given by

$$\mathbf{v} = \nabla \chi + \mathbf{k} \times \nabla \psi. \tag{4.3}$$

Since we are interested in the product $p^s [\![\mathbf{v}]\!]$, we integrate (4.3) from top ($\sigma = 0$) to bottom ($\sigma = 1$) and multiply it by p^s to obtain, after some algebra,

$$p^s [\![\mathbf{v}]\!] = p^s \nabla [\![\chi]\!] - \mathbf{k} \times [\![\psi]\!] \nabla p^s + \mathbf{k} \times \nabla (p^s [\![\psi]\!]). \tag{4.4}$$

By choosing $\nabla[\![\chi]\!] = \mathbf{k} \times [\![\psi]\!] \nabla(\ln p^s)$, the velocity can be computed using the formula

$$\mathbf{v} = \frac{1}{p^s}\mathbf{k} \times \nabla(p^s[\![\psi]\!]) + \nabla(\chi - [\![\chi]\!]_\circ) + \mathbf{k} \times \nabla(\psi - [\![\psi]\!]), \tag{4.5}$$

where $[\![\chi]\!]_\circ$ indicates the barotropic velocity potential in the original time series. The surface pressure in (4.5) is an ensemble average of the corresponding time series in the original dataset. This average surface pressure is also used in the definition of the energy matrix. The first term in the right hand side of (4.5) is the non-divergent barotropic velocity. The last two terms together represent the baroclinic velocity. These procedure is equivalent to filtering out external Kelvin waves [3].

The planetographic poles are points of special interest since they are singularities in spherical coordinates and velocity cannot be defined there. Following Achatz and Opsteegh [3] the poles are represented by a set of ζ-points, where vorticity and potential vorticity have a single value, to avoid defining velocity components there.

The inner product (3.30) and, therefore, the energy matrix (3.31) were discretised in accordance with the discretisation of the state variables. For simplicity, the energy matrix can be written in three parts corresponding to the three different state variables, namely zonal velocity, latitudinal velocity and square root of potential temperature. These parts are

$$E_1 = \operatorname{diag}\left\{\left(\frac{1}{2}\Delta V \overline{p^s}^{\lambda}\right)_\alpha\right\}_{\alpha=1}^{I(J-1)K}, \tag{4.6}$$

$$E_2 = \operatorname{diag}\left\{\left(\frac{1}{2}\Delta V \overline{p^s}^{\phi}\right)_\beta\right\}_{\beta=1}^{I(J-2)K}, \tag{4.7}$$

$$E_3 = \operatorname{diag}\left\{\left(\frac{P}{\kappa}\Delta V p^s\right)_\gamma\right\}_{\gamma=1}^{I(J-1)K}, \tag{4.8}$$

where $\overline{(\cdot)}^{\lambda}$ and $\overline{(\cdot)}^{\phi}$ indicate the arithmetic mean along λ and ϕ, respectively, and $\Delta V_{j,k} = (\cos\phi \Delta\lambda \Delta\phi)_j \Delta\sigma_k$. The indices α, β and γ represent the triplets $\alpha = (i, j + \frac{1}{2}, k)$, $\beta = (i + \frac{1}{2}, j, k)$, $\gamma = (i + \frac{1}{2}, j + \frac{1}{2}, k)$, running across every possible value of i, j and k in a definite order. The upper limit is defined by the number of grid points for each variable.

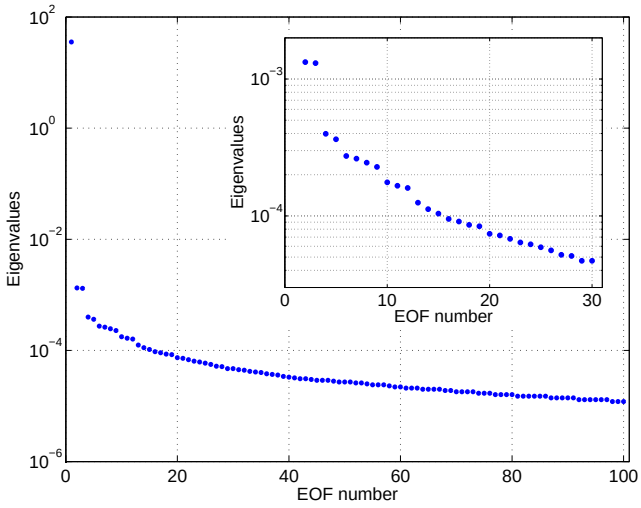


Figure 4.5: Eigenvalues obtained from the POD of the SGCM dataset.

4.2 Eigenvalues and energy contents

The empirical eigenvalues corresponding to the SGCM dataset are shown in figure 4.5. By definition, the first mode explains most of the total atmospheric energy. This energy is almost constant throughout the analysed time interval fluctuating around the time average of the total energy in the ensemble. In comparison, higher order modes are more dynamically active than the first mode. As we shall see, these facts allow the fields to be decomposed into a background state, based on the first EOF, and a perturbation or eddy state defined by the sum of higher order modes.

From EOF2 onwards the eigenvalues may be sorted in groups of two, three or more modes separated by gaps of decreasing size as the EOF number increases. This degeneracy indicates the presence of travelling waves whose Fourier-like components are associated with the same eigenvalue. There are five leading groups. The first one is constituted by EOF2 and EOF3. The first and most important gap is located between EOF3 and EOF4 always bearing in mind that the difference between the first mode and the rest is the most significant. The next group is the pair EOF4-5 and then a quartet composed of EOF6-9. However, looking at the eigenspectrum alone might be misleading. It would seem, for instance, that the next group is a trio composed of EOF10-12. Instead, we shall see in the subsequent analysis and discussion that this is indeed a pair comprising EOF10-11, while EOF12 is part of a quartet together with EOF13-15. From EOF16 onwards the spectrum appears to be continuous, fading into a long tail.

Figure 4.6 shows the normalised cumulative total eddy energy (i.e. excluding

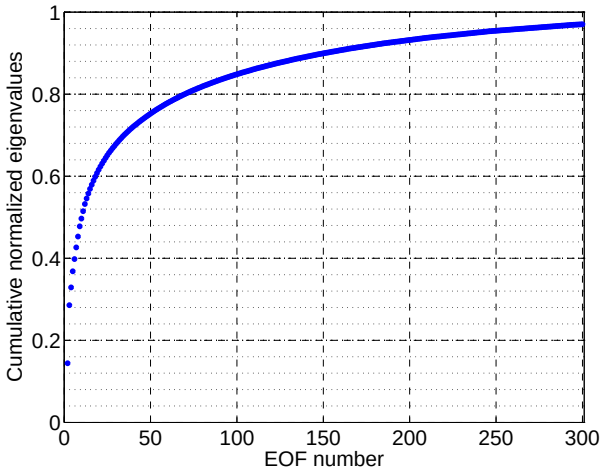


Figure 4.6: Cumulative energy in the SGCM dataset normalised with respect to the total energy in 500 EOFs (without considering energy in the first EOF).

the energy in the background state) throughout the POD modes with respect to the average energy in 500 EOFs. It is useful in determining which modes to retain in a reduced-order model. Thus, for instance, if we intend to explain 60% of atmospheric total eddy energy, it is necessary to retain 19 modes, whereas we should retain 70 modes in order to explain 80% TE in the eddy fields. However, an appropriate truncation cannot be based purely on an average energy content to be explained. The classification of the EOFs as constituents of groups makes it important to include every mode in an EOF group to enable a reduced-order model to represent the waves described by that group.

Another important aspect to be considered when deciding the number of modes to be retained is the influence of higher order modes on the dynamics of lower order modes. For example, it has been shown that the performance of EOF-based reduced models of the Charney-DeVore model [25] crucially depend on the number of retained EOFs [29]. The original Charney-DeVore model exhibits regime transitions if the parameters are carefully chosen [30]. However, reduced-models based on three and five EOFs were unable to reproduce regime transitions, whereas a 4-EOF model was able to display this behaviour. These results clearly show that the performance of a reduced-model is not only determined by the number of included modes but also by the interaction between them. Similar results, where the behaviour of reduced-models drastically changed by the inclusion or exclusion of certain modes, were found in the present work. These will be discussed in more detail in Section 5.3 after we deal with reduced-order models for the SGCM.

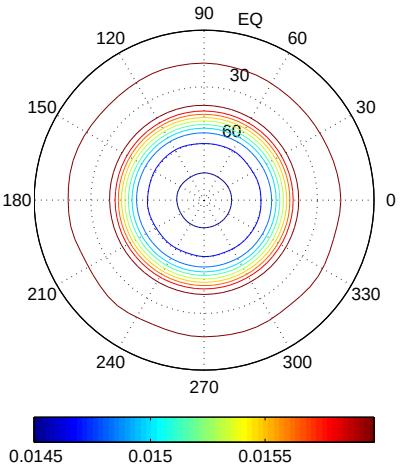


Figure 4.7: Temperature, in arbitrary units, in the first EOF for the SGCM at an approximate height $z \simeq 18.97$ km in the northern hemisphere.

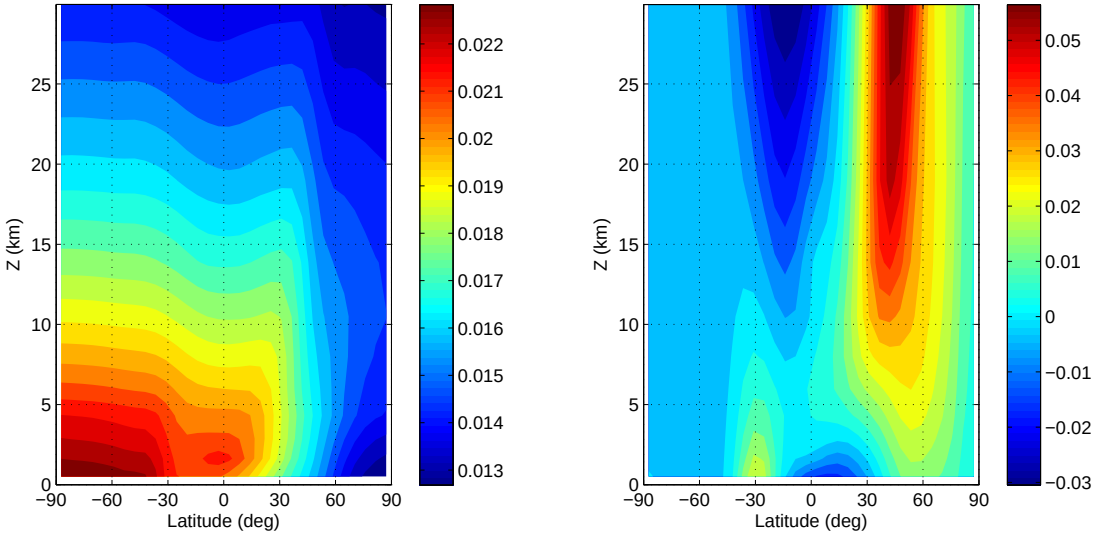
4.3 The first EOF

By definition, the first EOF is the most energetic mode. Its structure is highly axially symmetric as shown in figure 4.7. This was expected as the longitude-dependent influence provided by topography is not included in the SGCM. A latitude-height section of the zonal average of temperature and zonal wind for this mode is shown in figure 4.8.

The temperature profile has steep horizontal gradients in the northern hemisphere. The vertical structure of the zonal wind displays a jet flowing eastwards around 45°N and an easterly jet around 15°S, both at an approximate altitude of 30 km. Given these features, the section clearly resembles the zonal-mean structure of the atmosphere during winter in the northern hemisphere [62]. Indeed, a comparison between EOF1 and the temporal average state (figure 4.3) suggests that they are closely related.

This result was expected since the temporal average fundamentally contains the immense amount of total potential energy not available to be transformed into kinetic energy through atmospheric processes [66, 67]. Furthermore, the principal component associated with EOF1 (PC1) is nearly constant with a maximum absolute variation of 0.001, which is small compared with the variation in PC2 or PC3 whose maximum absolute variation is 0.140.

The relation between the ensemble average and EOF1 is made clear by assuming



(a) (b)

Figure 4.8: (a) Zonal average of temperature and (b) zonal velocity in the first SGCM EOF (both in arbitrary units). Positive horizontal velocity directed eastwards.

EOF1 to be proportional to the ensemble average

$$\phi^{(1)} = K\langle\psi\rangle, \tag{4.9}$$

where K is a scalar constant. By (3.23), EOF1 can be written as

$$\phi^{(1)} = \sum_{k=1}^{M} b_k^{(1)}\psi^{(k)}. \tag{4.10}$$

On the other hand, the ensemble average can be computed by

$$\langle\psi\rangle = \frac{1}{M}\sum_{k=1}^{M}\psi^{(k)}. \tag{4.11}$$

Substituting (4.10) and (4.11) into (4.9), it is easy to conclude that the coefficients $b_k^{(1)}$, for $k = 1, \ldots, M$, are constant and equal to K/M.

According to the Sirovich snapshots method [110], the coefficients $b_k^{(1)}$ for $k = 1, \ldots, M$, are the components of the first eigenvector of the matrix (3.24). This

implies that

$$\sum_{l=1}^{M} C_{kl} = C_0. \tag{4.12}$$

In other words, the row sum of the elements of C_{kl} is equal to a single constant C_0. Writing the state vector as the sum of the mean state $\Psi = \langle \psi^{(k)} \rangle$ and a perturbation to the mean $\psi'^{(k)} = \psi^{(k)} - \Psi$, the elements of the matrix C_{kl} can be written as

$$C_{kl} = \frac{1}{M} \left\{ (\Psi, \Psi) + (\Psi, \psi'^{(k)} + \psi'^{(l)}) + (\psi'^{(k)}, \psi'^{(l)}) \right\}.$$

Since

$$\sum_{k=1}^{M} \psi'^{(k)} = 0,$$

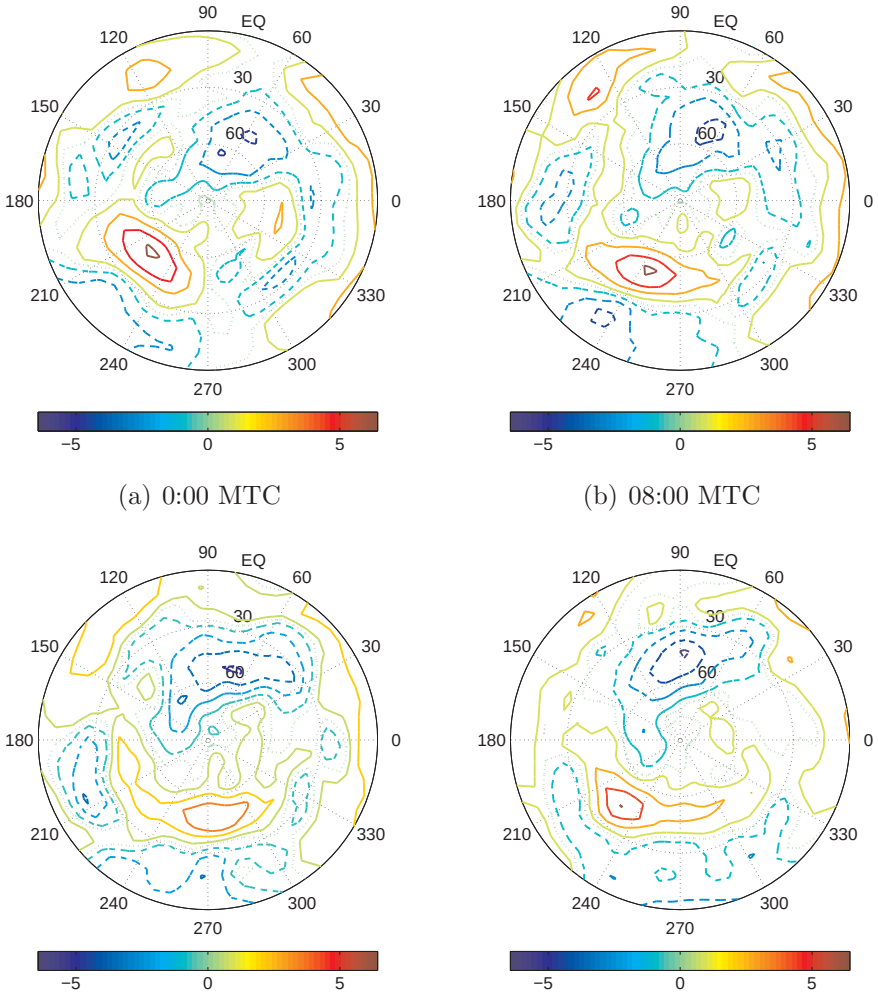
the sum (4.12) is

$$\sum_{l=1}^{M} C_{kl} = (\Psi, \Psi) + (\Psi, \psi'^{(k)}). \tag{4.13}$$

Therefore, the approximate constancy of the sum (4.13) depends on the relative size of the perturbation field product $(\Psi, \psi'^{(k)})$ with respect to the mean field product (Ψ, Ψ). In the case of the SGCM, the ensemble average $\left\langle \frac{(\Psi, \psi'^{(k)})}{(\Psi, \Psi)} \right\rangle = 0$ with standard deviation $s = 3.0117 \times 10^{-5}$. This shows that the time average is virtually proportional to EOF1. We shall see in Section 6.3 that a similar picture arises in the analysis of the Oxford MGCM so that EOF1 represents a time average state there as well.

4.4 Background state and eddy fields

We are interested in nonlinear atmospheric waves, their creation and evolution, describing them by reduced-order models. These models are intended to specifically explain the dynamics of the Martian atmosphere. In theoretical atmospheric studies it is customary to split the fields into a background flow and perturbations in the form of eddy components. The eddy components are simply defined as the difference between the instantaneous system state and the background state. The background state, on the other hand, has been subject to different definitions. In theoretical atmospheric wave studies, the background flow is traditionally defined as an axially symmetric zonal-mean flow (see [45], for example). In POD-Galerkin modelling, on the other hand, it is usually defined as an ensemble or time average state (e.g. [2, 3]). If the ensemble or time average state is kept for the computation of EOFs, the POD provides an alternate definition that for our purposes is more suitable. The



(a) 0:00 MTC

(b) 08:00 MTC

(c) 16:00 MTC

(d) 24:00 MTC

Figure 4.9: Temperature eddy component, measured in K, in the original dataset at four different times (indicated in each frame) during one sol at $\sigma = 0.15$ (approximate height $z \simeq 18.97$ km).

background state can be defined as

$$\overline{\psi} = a_1 \phi^{(1)}, \tag{4.14}$$

where $\phi^{(1)}$ is EOF1 and a_1 is the first principal component given by $a_1 = (\phi^{(1)}, \psi)$. Given the attributes of the SGCM EOF1 described in the preceding paragraphs, the structure of the background state is very close to the ensemble average, which in turn is highly zonally symmetric due to the flatness of the modelled planet. This is in close correspondence with both definitions of the background state as either a zonal-mean flow or as an ensemble or time average. This situation is similar in the case of the Oxford MGCM as we shall see in Section 6.3. Therefore, the background state defined by (4.14) may be viewed as a generalisation of the concept of background state from the highly idealised SGCM to the highly realistic MGCM. However, this would not be true in the case where the temporal average was subtracted from the source dataset from the beginning.

The eddy field ψ' is defined as the remainder after subtracting the background state from a complete field, i.e.

$$\psi' = \psi - \bar{\psi}, \tag{4.15}$$

where ψ' strongly depends on both space and time. The evolution of temperature eddies during one sol is shown in figure 4.9. It is apparent that it is composed of a combination of different waves. The POD provides a way to resolve the waves embedded into atmospheric motion. In this respect, the definition of the background state by (4.14) makes the interpretation of higher order EOFs transparent.

4.4.1 Higher order EOFs

In Section 4.2 we saw that the EOFs can be organised in groups of various sizes according to their associated empirical eigenvalues. In this section, we shall discuss the structure of the modes that compose each EOF group and analyse the time evolution of their amplitudes by means of Fourier analysis. In this way we shall be able to discern the frequencies that characterise their evolution.

The first EOF group is formed by EOF2 and EOF3. Figures 4.10 and 4.11 show these two modes in the northern hemisphere at $\sigma = 0.55$ ($z \simeq 5.98$ km above the surface) and $\sigma = 0.15$ ($z \simeq 18.97$ km above the surface), respectively. Although the structure of the modes varies with height, the modes remain structurally similar to each other throughout the vertical. At $\sigma = 0.55$ (figure 4.10) the modes consist of

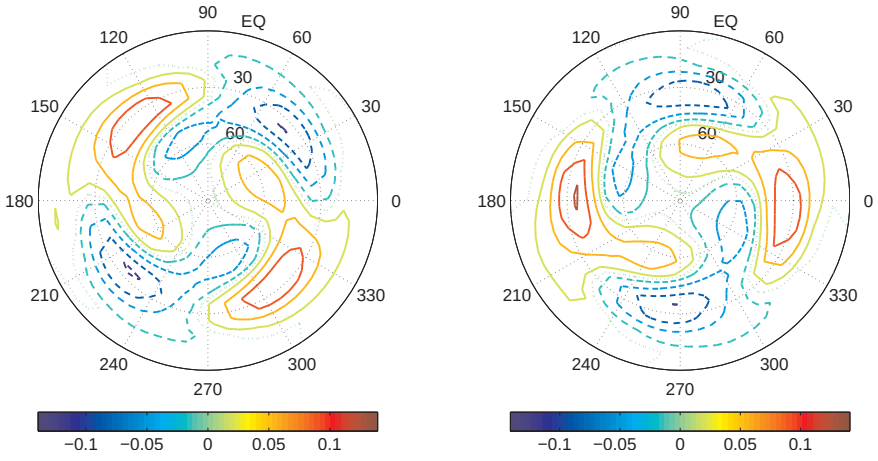


Figure 4.10: Square root of potential temperature, in arbitrary units, in the SGCM eigenfunction (a) EOF2 and (b) EOF3 at $\sigma = 0.55$ (approximate height $z \simeq 5.98$ km) in the northern hemisphere.

a primary longitudinal wavenumber two wave located around 40°N and a secondary longitudinal wavenumber two wave around 50°N of minor intensity. The equator does not display strong activity at this height. In contrast, the structure of the modes is more complicated at $\sigma = 0.15$ (figure 4.11). On that sigma-level there are two waves of longitudinal wavenumber two with similar intensity located around 10°N and 45°N spanning the whole northern hemisphere. The longitude difference between EOF2 and EOF3 is approximately 45°.

The corresponding PC2 and PC3 are shown in figure 4.12(a). They are two quasi-sinusoidal signals of comparable amplitude and frequency but different phase. The power vs period plots of these two principal components show that the dominant period in both signals is $\mathcal{P} = 5.88$ sols (figure 4.12). Additionally, the phase difference is $\Delta\phi = 1.567$ rad, equivalent to $\Delta t = \mathcal{P}\Delta\phi/2\pi = 1.467$ sols. Thus, these modes describe a pair of waves travelling eastwards with an approximate period of 5.88 sols.

The Fourier spectra can be used as a means for classifying the POD modes into groups. In general, these groups coincide with the EOF groups constructed by considering the associated eigenvalues relative sizes (figures 4.13 and 4.14). EOF12 constitutes a notable exception. According to the eigenspectrum it must be grouped together with the modes ten and eleven forming a trio. However, Fourier analysis shows that EOF10 and EOF11 form a couple with two period bands around 5 and 15 sols, respectively (figure 4.13(g)-(h)). EOF12 is more appropriately grouped together with EOF13, EOF14 and EOF15 whose Fourier spectra, with period bands

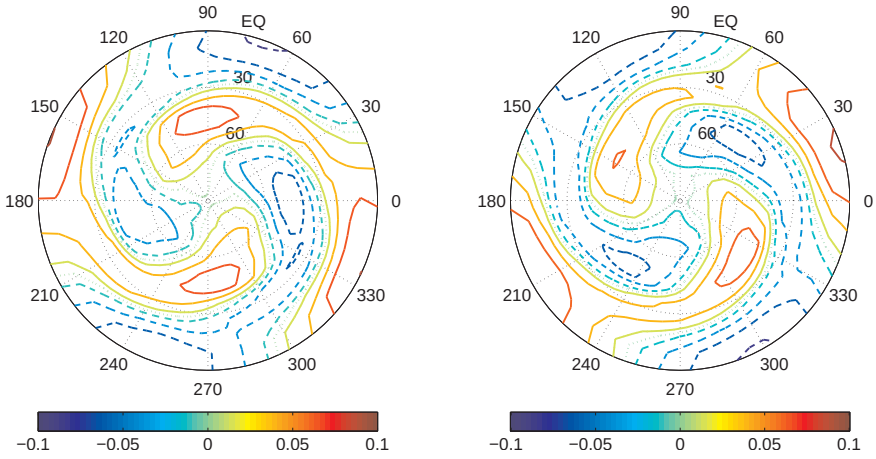


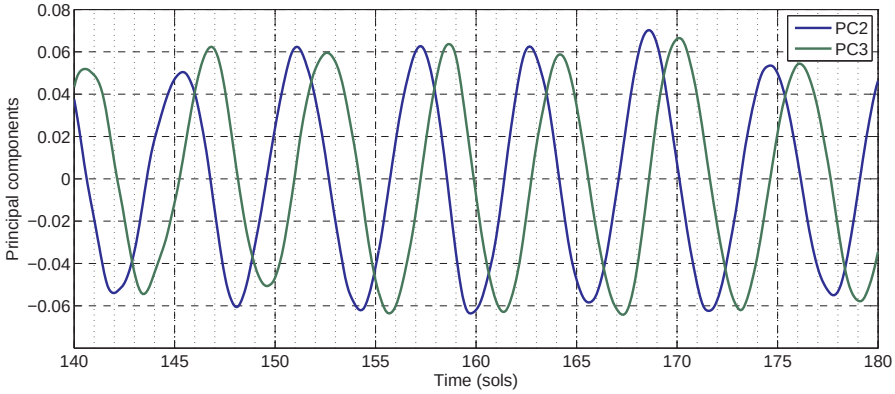
Figure 4.11: Square root of potential temperature, in arbitrary units, in the SGCM eigenfunctions (a) EOF2 and (b) EOF3 at $\sigma = 0.15$ (approximate height $z \simeq 18.97$ km) in the northern hemisphere.

around 12.5 sols, look alike (figure 4.14(a)-(d)). The structure of these modes at an approximate height $z \cong 5.98$ km above the surface is shown in figure 4.19.

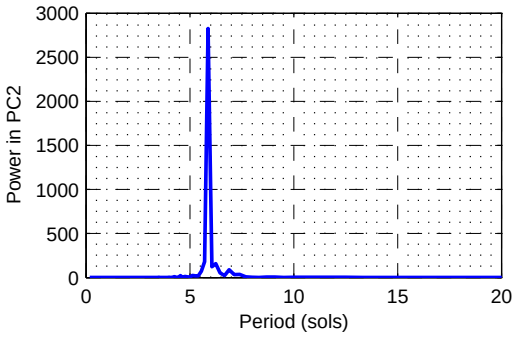
Each group displays distinctive characteristics. The pair EOF4-5 is constituted by a zonal wavenumber-2 wave between 30 and 60°N (figure 4.15(a)-(b)). The power vs period plots of the corresponding principal components (PC4 and PC5) show peaks at $\mathcal{P}_1 = 5.5$ sols and, close to each other, $\mathcal{P}_2 = 6.0$ and $\mathcal{P}_3 = 6.5$ sols. These two close periods are responsible for the beating displayed by both time series (figure 4.15(c)). Indeed, half the difference of the corresponding frequencies, i.e $\mathcal{P}_2^{-1}$ and $\mathcal{P}_2^{-1}$ corresponds to a period of an envelope of $\mathcal{P}_e = 156$ sols (cf. figure 4.15(c)). The beating observed in PC4 and PC5 combined gives rise, after summing up the product of PCs and corresponding EOFs, to a wave that displays amplitude vacillation.

The quartet EOF6-9 has a zonal wavenumber-1 wave at $\sigma = 0.15$ (figure 4.16). Their behaviour in time (given by PC6-9) is governed by longer periods ($\mathcal{P} > 11$) (figure 4.13(c)-(f)). Unfortunately, the resolution of the power vs period plots becomes poorer in that region. The first peaks in those plots are located at $\mathcal{P} = 12.5$ and $\mathcal{P} = 13.25$ sols. The behaviour, as observed in the time domain, seems more complex than the previous examples although it is still possible to observe signs of beating and, therefore, amplitude vacillation (figure 4.17). The apparent correlation between PC6-9 confirms that the corresponding EOFs constitute a single EOF group.

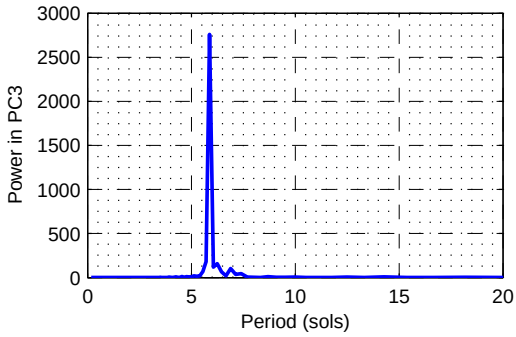
It has been shown that the axial symmetry of the thermal forcing term imposes conditions onto the form of the expected EOFs so that every EOF belongs to the



(a) Time series



(b) PC2

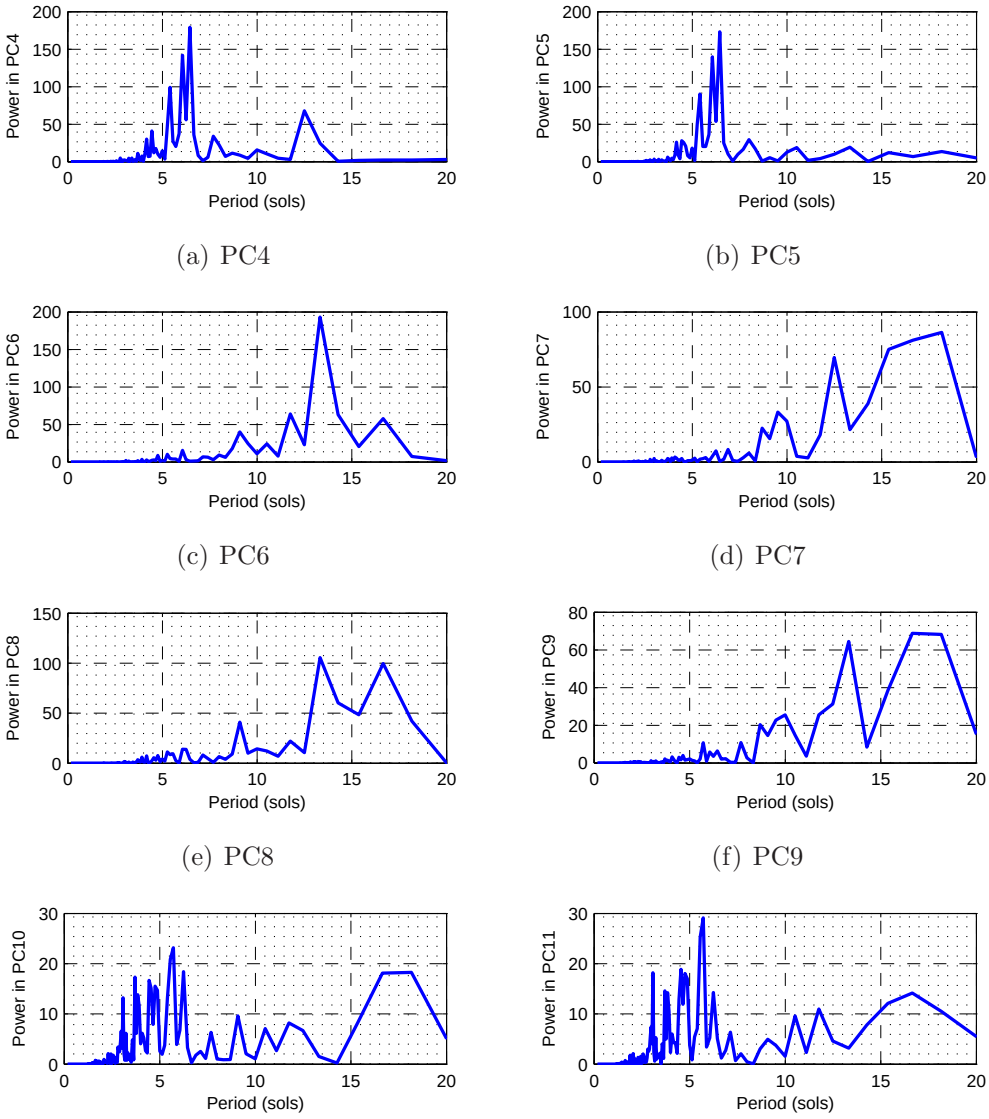


(c) PC3

Figure 4.12: (a) PC2 and PC3 during a 40-sol time interval. Power vs period plots corresponding to the original SGCM dataset corresponding to (b) EOF2 and (c) EOF3.

subspace of one single zonal wavenumber [5]. This fact was used to assess the statistical convergence of the obtained EOFs. What we found was that the first 9 EOFs are indeed contained in a single zonal wavenumber subspace (apart from small deviations). For higher order EOFs the statistical convergence decreases when the spread of the modes between two or more zonal wave numbers is more noticeable. This is mainly due to the small number of ensemble members. We still describe some of the higher order modes since they are still characteristic to the SGCM dataset.

The structure of the pair EOF10-11, at an approximate height $z \simeq 5.98$ km, consists of a zonal wavenumber-3 wave around 50 °N and weaker zonal wavenumber-1 waves around 35 °N and 70 °N (figure 4.18). In contrast, the components of the quartet EOF12-15 exhibits zonal wavenumber-two waves around 50°N, which supports the conclusion, drawn from Fourier analysis, that these EOFs constitute a separate group.



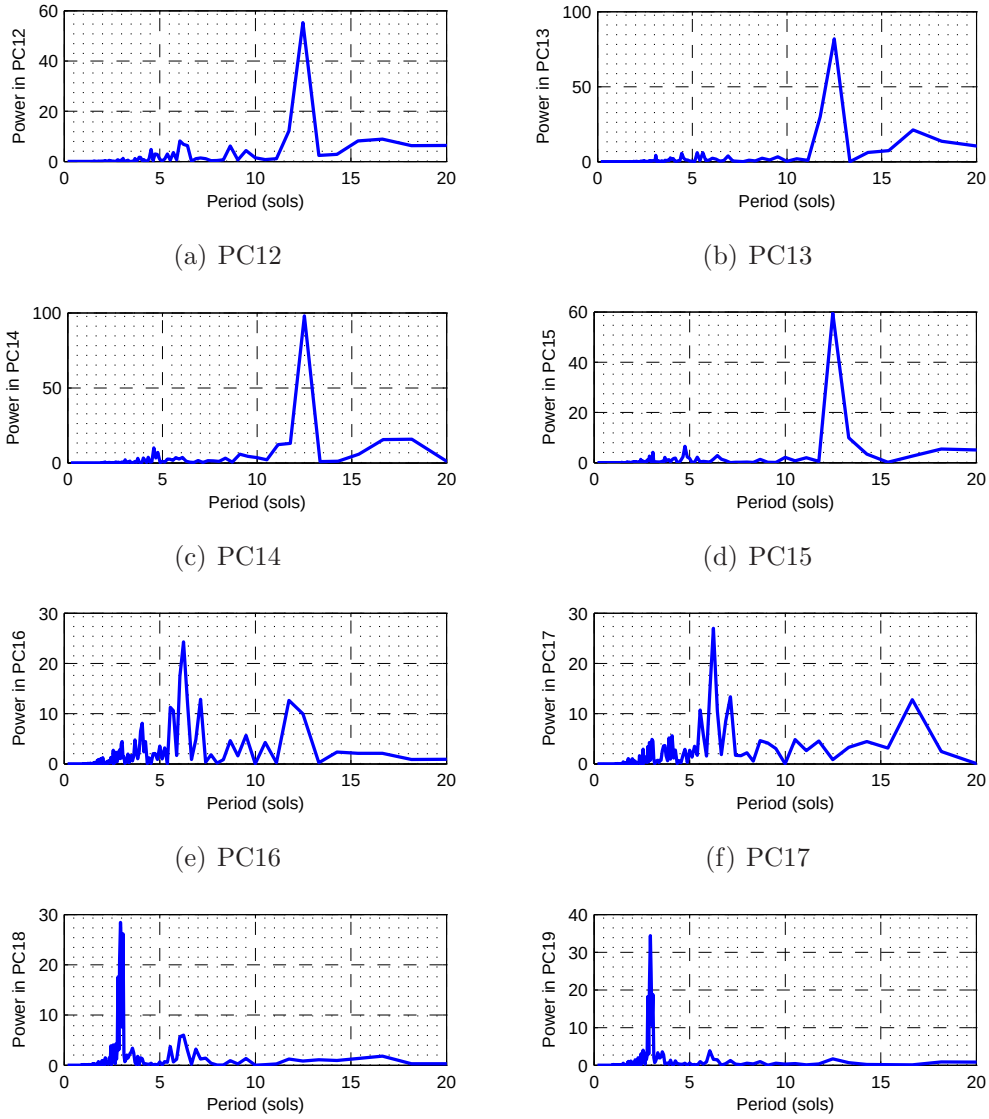
(a) PC4 (b) PC5

(c) PC6 (d) PC7

(e) PC8 (f) PC9

(g) PC10 (h) PC11

Figure 4.13: Power vs period plots for the original SGCM dataset corresponding to (a)–(h) EOF4–EOF11.



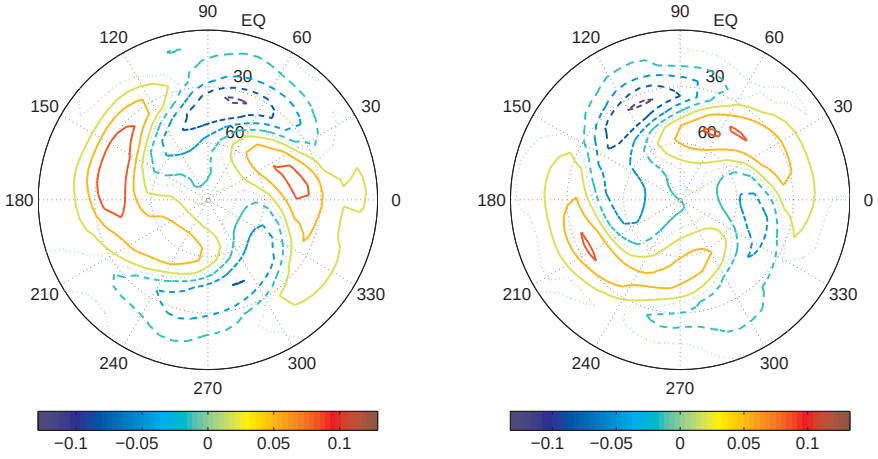
(a) PC12 (b) PC13

(c) PC14 (d) PC15

(e) PC16 (f) PC17

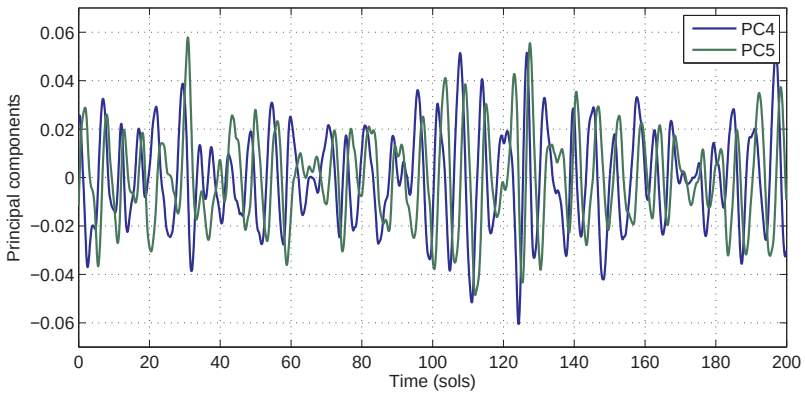
(g) PC18 (h) PC19

Figure 4.14: Power vs period plots for the original SGCM dataset corresponding to (a)–(h) EOF12–EOF19.



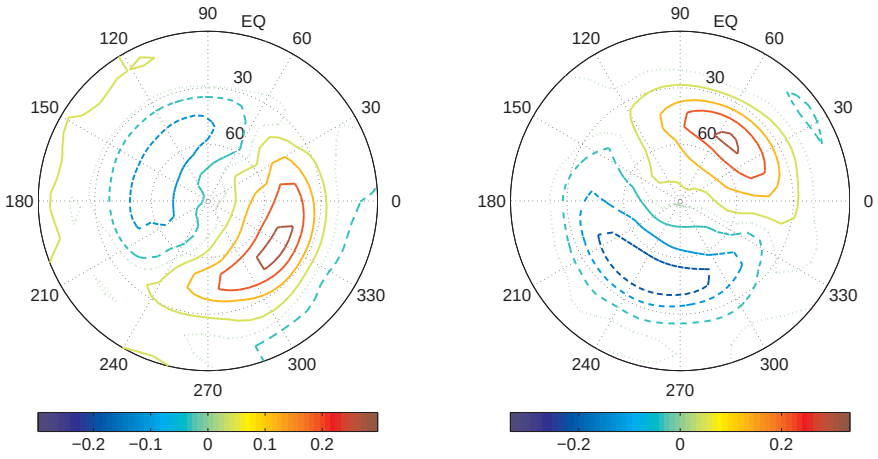
(a) EOF4

(b) EOF5

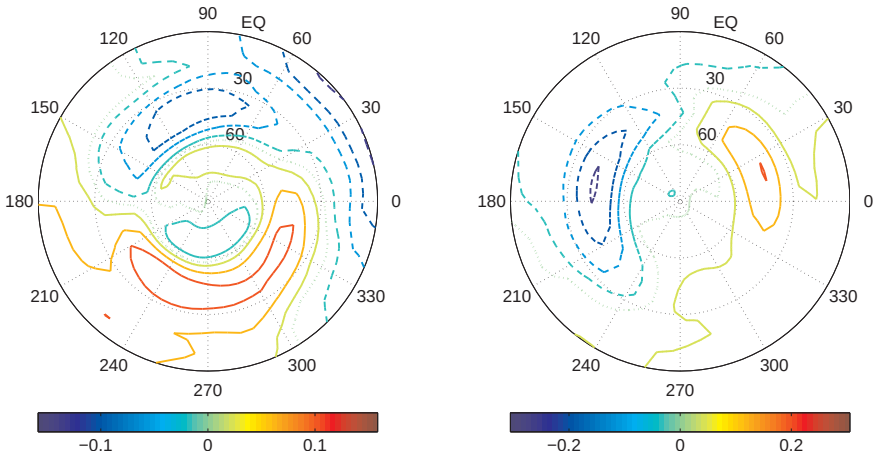


(c) Principal components

Figure 4.15: Square root of potential temperature, in arbitrary units, in the SGCM eigenfunction (a) EOF4 and (b) EOF5 at $\sigma = 0.55$ in the northern hemisphere; (c) time series corresponding to PC4 and PC5.



(a) EOF6 (b) EOF7



(c) EOF8 (d) EOF9

Figure 4.16: Square root of potential temperature, in arbitrary units, in (a) EOF6, (b) EOF7 (c) EOF8 and (d) EOF9 at $\sigma = 0.15$ in the northern hemisphere.

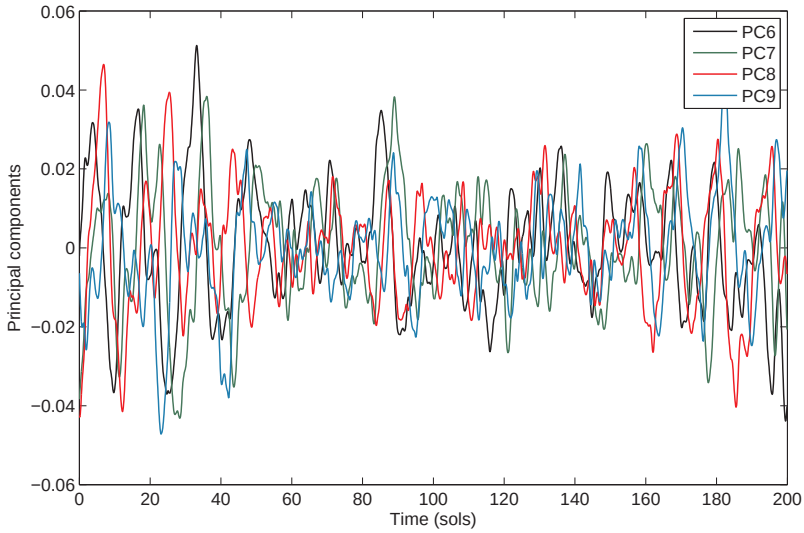


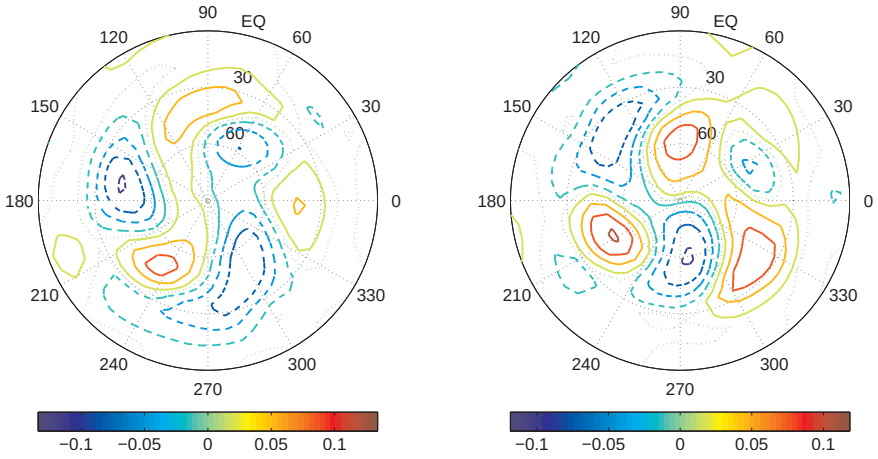
Figure 4.17: PC6 to PC9 during a 200-sol time interval.

Harmonic analysis also neatly separates the next four modes into two different pairs. The frequency spectrum shows a period band around a dominant peak of 6 sols for EOF16 and EOF17 (figure 4.14(e)-(f)) whereas it exhibits a narrow peak around 3 sols for EOF18 and EOF19 (figure 4.14(g)-(h)). As before, the mode structure supports this classification. Figure 4.20 shows the four modes at an approximate height $z \simeq 18.97$ km. EOF16 and 17 consist of a zonal wavenumber-3 wave with a strong wavenumber-1 component around 45°N (figure 4.20(a)-(b)) whereas the couple EOF18-19 is composed of a zonal wavenumber-4 wave around 20°N and a mixture of wavenumber-1 and wavenumber-3 waves around 45°N (figure 4.20(c)-(d)). The classification of modes beyond EOF19 is not as clear as it is for the first 19 modes.

4.4.2 Comparison with previous work

The performance of the POD on the SGCM, despite its inherent simplicity, establishes a benchmark to analyse the results of the POD of more general models. Before moving on, we would like to discuss the present results in the view of previous work on reduced-models of the SGCM, in particular, the work by Whitehouse *et al.* [124]. They analysed a wave 3 during perennial southern winter described previously by Collins and James [26]. The analysis was carried out using a combination of vertical decomposition (using the QG vertical structure equation (see, for example, [87])) and POD with different norms (variance, kinetic energy and total energy) within the framework of QG theory.

Despite not being directly comparable, the results by Whitehouse *et al.* [124]



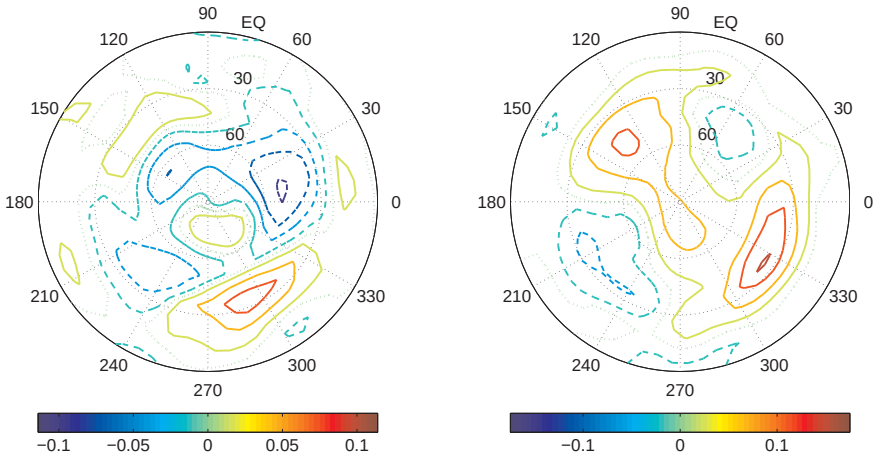
(a) EOF10 (b) EOF11

Figure 4.18: Square root of potential temperature, in arbitrary units, in the SGCM at $\sigma = 0.55$ for (a) EOF10 and (b) EOF11.

and those presented here bear important similarities. Though more slowly in the present case, the distribution of energy among the modes decays in a similar way. The slower decay is probably due to the fact that we have not employed QG theory (compare figure 4.6 and figure 9(a) in [124]). Therefore, the dataset in this thesis has not been drastically filtered, retaining motions on a wider scale range. Still, the reduction of degrees of freedom from $\mathcal{O}(10^4)$ in the primitive-equation SGCM to $\mathcal{O}(10^2)$ is comparable to that in [124].

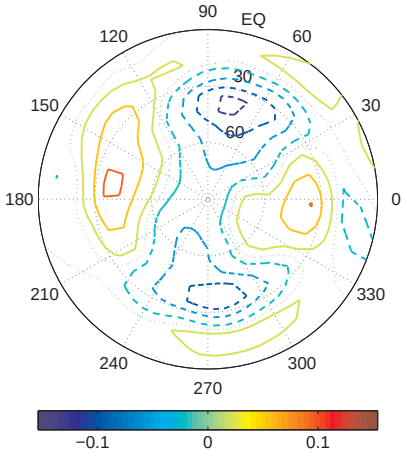
The EOF structure and resulting principal components are comparable as well, always considering the differences between the regimes under analysis. In both cases, the EOFs can be classified in conjugate pairs or groups. Whitehouse *et al.* [124] found that the most energetic group corresponds to a zonal wavenumber-3 travelling wave clearly corresponding to the original flow. In the present work, the POD was able to identify underlying wave motion not evident at first sight. Thus, the most energetic mode (apart from the zonally symmetric background flow) corresponds to a zonal wavenumber-2 wave.

The most important differences between both sets of EOFs is the location of wave motion. In [124], the wave activity is located in the southern hemisphere, whereas in the present work it is located in the northern hemisphere. Again, this is not surprising since the original time series correspond to southern and northern winter, respectively. However, these results point out the diversity of modes that might be needed to generate reduced-models incorporating the seasonal cycle.

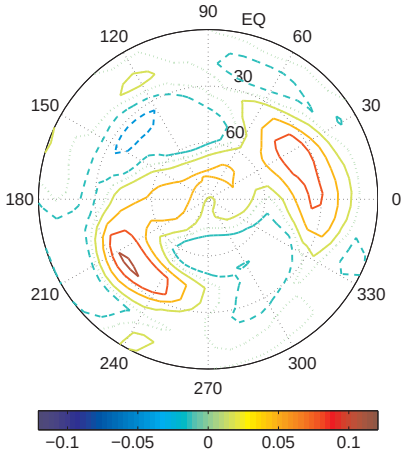


(a) EOF12

(b) EOF13

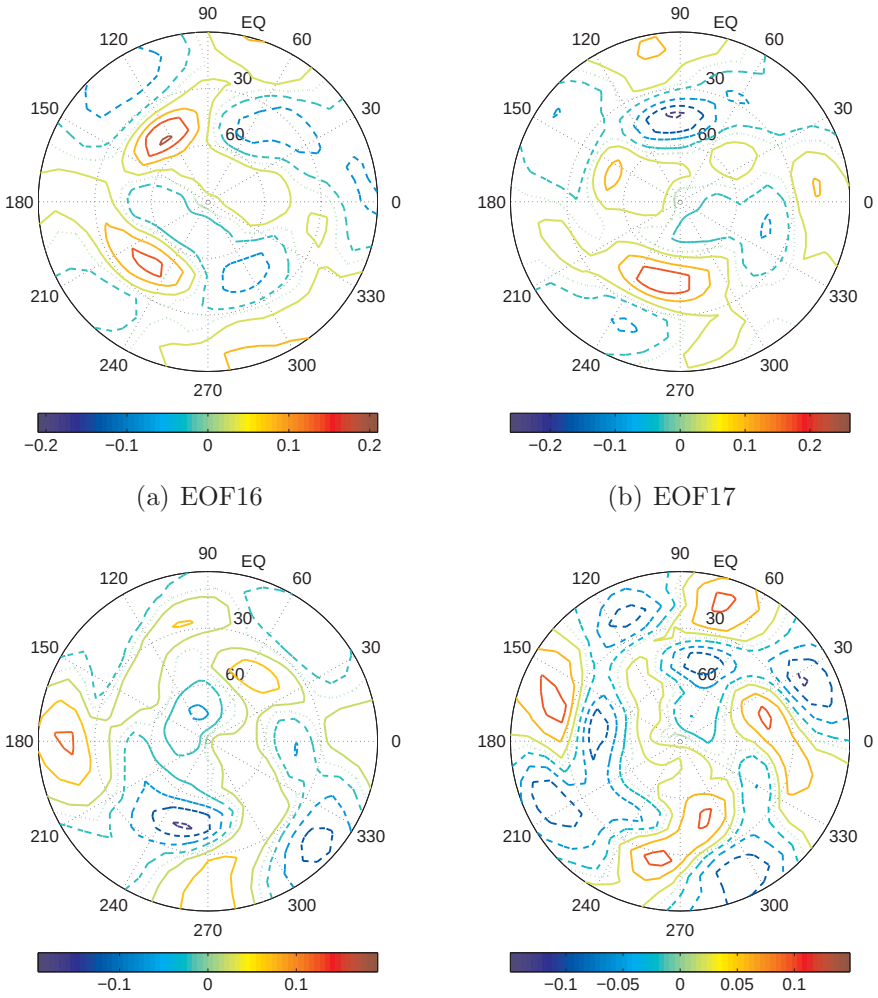


(c) EOF14



(d) EOF15

Figure 4.19: Square root of potential temperature, in arbitrary units, in the SGCM at $\sigma = 0.55$ for (a)–(d) EOF12–EOF15.



(a) EOF16 (b) EOF17

(c) EOF18 (d) EOF19

Figure 4.20: Square root of potential temperature, in arbitrary units, in the SGCM at $\sigma = 0.15$ (approximate height $z \simeq 18.97$ km) for (a)–(d) EOF16–EOF19.

4.5 Summary

In this chapter we have performed a three-dimensional multivariate POD on a dataset produced by the SGCM for perennial northern winter. The resulting EOFs were classified according to their spatial structure and the Fourier spectra of the corresponding PCs. By doing so, we have been able to extract different modes with physical meaning. First, the most energetic EOF1 was clearly identified with the background zonal flow. Hence, the rest of the EOFs were identified as components of the eddy field. These were also classified in groups representing different waves corresponding to different combinations of zonal wave numbers and periods.

The EOF basis for the SGCM will used in Chapter 5 to construct Galerkin-projected low-order models that capture the essential dynamics of the SGCM during the period under consideration. In that chapter, the EOFs and EOF groups shall be discussed in greater detail within a framework of prognostic character rather than the diagnostic point of view adopted here.

Chapter 5

Reduced-order models of the SGCM

The POD generates an optimal basis in the sense that fewer modes are required to explain the same amount of energy in comparison with any other linear decomposition. We have seen in the previous sections that atmospheric TE in the SGCM can be explain up to a significant percentage by a small number of POD modes. The first twenty modes, for example, are able to explain as much as 90% of TE in the eddy field of the MGCM (cf. figure 6.5). We shall see that the outlined feasibility of previous chapters indeed lead to actual models capable of representing the evolution of the most energetic modes in Martian atmospheric dynamics.

We begin this section by describing the process of constructing low-order models. This description will include the parameterisation of unresolved modes and an appropriate way of forcing these models. Based on this theoretical framework, we shall construct and analyse various models for the SGCM. Furthermore, we shall employ the same ideas and methods to construct low order dynamical models of the MGCM data in Chapter 7. Our analysis of reduced-order models includes a comparison against the original datasets.

5.1 POD-Galerkin models

Let us consider the dimensionless hydrostatic primitive equations in sigma-coordinates in the form given by Achatz and Opsteegh [3]. This system already incorporates the time-independent surface pressure assumption introduced in Section 3.3. The

momentum, thermodynamic energy and continuity equations can be written as

$$\frac{\partial \mathbf{v}}{\partial t} + \zeta \mathbf{k} \times \mathbf{v} + \nabla(K + \Phi) + P\tau^2 \nabla \ln p^s + \dot{\sigma}\frac{\partial \mathbf{v}}{\partial \sigma} = \mathbf{F}, \tag{5.1}$$

$$\frac{\partial}{\partial t}(p^s \tau) + \nabla \cdot (p^s \mathbf{v}\tau) + \frac{\partial}{\partial \sigma}(p^s \dot{\sigma}\tau) = \frac{p^s J}{2P\tau}, \tag{5.2}$$

$$\nabla \cdot (p^s \mathbf{v}) + \frac{\partial}{\partial \sigma}(p^s \dot{\sigma}) = 0. \tag{5.3}$$

The dimensionless hydrostatic equation can be written in the following two forms that are useful for discretisation [9]

$$\delta(\Phi\sigma) = -(P\tau^2 - \Phi)\delta\sigma, \tag{5.4}$$

$$\kappa\delta\Phi = -\tau^2 \delta P. \tag{5.5}$$

In these equations the prognostic variables are $\mathbf{v}$, horizontal velocity, and τ the square root of potential temperature; $\zeta = f + \xi$ is absolute vorticity, where $f = 2\sin\phi$ is the Coriolis factor and $\xi = \mathbf{k} \cdot \nabla \times \mathbf{v}$ is the vertical projection of relative vorticity; K is kinetic energy, Φ is geopotential height, p^s is surface pressure, $\dot{\sigma} = d\sigma/dt$ is sigma-velocity and $P = (\sigma p^s)^\kappa$ where $\kappa = c_p/R$ is the ratio of specific heat at constant pressure to the atmospheric gas constant.

The forcing is given by friction and diffusion, $\mathbf{F}$, and diabatic heating per unit mass, J. The explicit form of these terms depends ultimately on the modeller's choice. In general, their complexity will be in accordance with the complexity of the discretisation scheme of the equations of motion, the dynamical core, and the intended application. In a GCM, for instance, these terms will include parameterisation schemes for modelling radiative heating, surface interactions and sub-grid scale effects as well as dissipation to remove unwanted energy accumulation.

The continuity equation (5.3) already incorporates the assumption of having time-independent surface pressure. Written in this form, this equation constitutes an auxiliary relation to compute sigma-velocity as

$$\dot{\sigma} = -\frac{1}{p^s}\int_0^\sigma \nabla \cdot (p^s \mathbf{v}) d\sigma_1 = \frac{1}{p^s}\int_\sigma^1 \nabla \cdot (p^s \mathbf{v}) d\sigma_1, \tag{5.6}$$

since $\dot{\sigma}$ is assumed to be zero at the ground and at infinity. In turn, the hydrostatic equation (5.4) and (5.5) is an auxiliary relation to compute geopotential height.

5.1.1 Building EOF-based reduced models

A reduced-order model is obtained by projecting (5.1) and (5.2) onto the most energetic EOFs. Since each EOF comprises three scalar components, namely the two horizontal velocity components and τ, a more suitable form of the equations of motion is

$$\frac{\partial \psi}{\partial t} = \Psi_t(\psi, t), \tag{5.7}$$

where $\psi = (\mathbf{v}, \tau)^T$ and Ψ_t is a nonlinear operator given by

$$\Psi_t = \begin{pmatrix} -\zeta \mathbf{k} \times \mathbf{v} - \nabla(K + \Phi) - P\tau^2 \nabla \ln p^s + \dot{\sigma} \frac{\partial \mathbf{v}}{\partial \sigma} + \mathbf{F} \\ -\frac{1}{p^s} \nabla \cdot (p^s \mathbf{v} \tau) - \frac{\partial}{\partial \sigma}(\dot{\sigma} \tau) + \frac{J}{2P\tau} \end{pmatrix}. \tag{5.8}$$

In principle, the construction of a Galerkin reduced-order model comprises two steps:

(i) writing the equations of motion in terms of basis functions, and

(ii) taking the inner product of these equations with the basis functions.

These two steps reduce the original PDE system into a set of ODEs. In practice, it is more convenient to use a modified version of the transform method [20], used in spherical harmonics spectral models [20, 47] to compute nonlinear terms. Following this idea, we can transform from EOF coefficients into spatial grid points. Tendencies are then computed using a discrete version of (5.8). Finally, taking the inner product of tendencies and EOFs projects tendencies back into the reduced EOF space.

The discrete version of (5.8) was written following the model developed by Achatz and Opsteegh [3] with minor modifications. In particular, in that work the chosen state variables are the barotropic streamfunction Ψ that arises as a consequence of (4.2), namely $p^s [\![\mathbf{v}]\!] = \mathbf{k} \times \nabla \Psi$, the baroclinic velocity $\hat{\mathbf{v}} = \mathbf{v} - [\![\mathbf{v}]\!]$ and τ. Barotropic streamfunction tendencies are computed via the time derivative of relative barotropic vorticity and the relation [3]

$$(\Delta A [\![\zeta]\!])_{i,j} = (\mathrm{V}\Psi)_{i,j},$$

where $\Delta A = \cos\phi \Delta\lambda \Delta\phi$ is the discrete area differential, V is the discrete version of $\nabla \cdot \frac{\nabla}{p^s}$ and the sub-indices (i, j) label a particular grid point. Baroclinic velocity tendencies are computed by subtracting barotropic velocity tendencies from full velocity tendencies. With this choice, Achatz and Opsteegh [3] ensure that external Kelvin waves are filtered out.

In this study, external Kelvin waves are filtered out by subtracting the divergent part from full velocity tendencies. The barotropic streamfunction Ψ can be related to a certain barotropic vorticity

$$Z = \mathbf{k} \cdot \nabla \times (p^s [\![\mathbf{v}]\!]) = \nabla^2 \Psi. \tag{5.9}$$

Equation (5.9) can be easily solved for Ψ by using of spherical harmonics Y_n^m, which are eigenfunctions of the Laplacian in spherical coordinates. In spectral form (5.9) takes the form

$$\nabla^2 \left(\sum_{n=0}^{N} \sum_{m=-n}^{n} \Psi_n^m Y_n^m \right) = \sum_{n=0}^{N} \sum_{m=-n}^{n} Z_n^m Y_n^m, \tag{5.10}$$

where Ψ_n^m and Z_n^m are the spectral coefficients of Ψ and$Z$, respectively, and $N \leq I/2$ is the maximum total wavenumber allowed by the chosen truncation (where I is the number of longitudes in the physical grid). Since $I = 64$, the truncation used in this work is a jagged T21 in order to avoid aliasing [47]. Equating coefficients on both sides of (5.10) and recalling that $\nabla^2 Y_n^m = -n(n+1)Y_n^m$ we find that

$$\Psi_n^m = \frac{-Z_n^m}{n(n+1)}, \quad n \neq 0. \tag{5.11}$$

The coefficient corresponding to $n = 0$ represents a global mean and can be chosen to be $\Psi_0^0 = 0$ without loss of generality.

Once the barotropic streamfunction has been computed, the filtered velocity can be recovered by

$$\mathbf{v} = \mathbf{v}^o - [\![\mathbf{v}^o]\!] + \frac{1}{p^s} \mathbf{k} \times \nabla \Psi, \tag{5.12}$$

where $\mathbf{v}^o$ is the velocity before filtering. Since the operators $\frac{\partial}{\partial t}$ and ∇ act on different domains and p^s is time-independent, (5.12) can be rewritten for tendencies simply by substituting $\mathbf{v}$, $\mathbf{v}^o$, $[\![\mathbf{v}^o]\!]$ and Ψ with their corresponding time derivatives. This version of (5.12) produces the desired result, i.e. the momentum equations with a non-divergent barotropic component.

Another difference between the model in [3] and the one here lies in the number of sigma layers: [3] used 3, we used 10. These layers were equally spaced for the SGCM and unevenly spaced for the MGCM, distributed in accordance with the description in Chapter 6.

The construction of reduced models based on EOFs should consider the energy partition structure over the POD modes not only for deciding the level of energy to be kept but also for including all the modes that belong to a degenerate energy

level. Therefore, the way in which the modes are included into progressively complex reduced models depends not only on the desired energy level but on the atmospheric structure the modeller wishes to reproduce. The first modes in the SGCM and MGCM provide us with a good example. The SGCM eigenspectrum (figure 4.5) shows that in order for a model to simulate travelling waves, it is necessary to include at least three modes: the background state (EOF1) and the first degenerate pair (EOF2 and EOF3). On the other hand, an MGCM-based reduced model (figure 6.4) requires the inclusion of at least four modes: the background state (EOF1), the second single mode (EOF2) and the first degenerate pair (EOF3 and EOF4).

5.1.2 Parameterisation of unresolved modes

Before constructing a reduced model, one further aspect must be considered: the dynamical impact of unresolved modes. There are different proposals for modelling this influence [2, 3, 59, 107]. The models developed in this thesis make use of an empirical approach [2, 3] that determines a set of closure parameters by minimising the relative error between the tendencies predicted by the reduced model $\dot{a}_k$ and those computed from the original time series $\dot{a}_k^{ts}$, [2, 3]

$$e_r = \frac{\left\langle \sum_{k=1}^{N} (\dot{a}_k^{ts} - \dot{a}_k)^2 \right\rangle}{\left\langle \sum_{k=1}^{N} (\dot{a}_k^{ts})^2 \right\rangle}, \tag{5.13}$$

where N is the number of modes included in the reduced model and $\langle \cdot \rangle$ indicates time (or ensemble) average.

The form of the parameters to be determined can be deduced through the following argument. Let a_k^r and a_k^n denote the resolved and unresolved modes, respectively. A full model, incorporating the action of resolved and unresolved modes on resolved modes up to quadratic nonlinearities, would take the form

$$\dot{a}_k^r = \alpha_k + \beta_{kl} a_l^r + \gamma_{klm} a_l^r a_m^r + (\delta_{kl} a_l^n + \epsilon_{klm} a_l^n a_m^n) + \zeta_{klm} a_l^n a_m^r, \tag{5.14}$$

where the tensors δ_{kl} and ϵ_{klm} represent the linear and quadratic contribution of the unresolved modes and ζ_{klm} is the interaction between resolved and unresolved modes. Since the latter are not known, (5.14) can be rewritten in a more compact form (after dropping the upper index r) as

$$\dot{a}_k = \alpha_k + \beta_{kl} a_l + \gamma_{klm} a_l a_m + \Delta_k + Z_{kl} a_l, \tag{5.15}$$

where $\Delta_k = \delta_{kl}a_l^n + \epsilon_{klm}a_l^n a_m^n$ and $Z_{kl} = \zeta_{klm}a_l^n$. Thus, Δ_k introduces N scalar parameters whereas Z_{kl} introduces N^2 additional ones.

5.1.3 Forcing

Finally, the form of the forcing to be used must be specified. The SGCM was forced by Newtonian cooling to a prescribed zonally symmetric temperature field (cf. figure 4.1) and Rayleigh friction of the lower-most level to emulate the planetary boundary layer [124, 125]. The reduced models described here are forced in the same manner although a linear Newtonian relaxation forcing in a temperature equation leads to a quadratic forcing in an equation for τ. Thus,

$$\mathbf{F}(\sigma) = \begin{cases} -r\mathbf{v}(\sigma), & \sigma = \sigma_K \\ 0, & \text{otherwise} \end{cases}, \tag{5.16}$$

$$J = \frac{P\left(\tau_*^2 - \tau^2\right)}{\tau_R}, \tag{5.17}$$

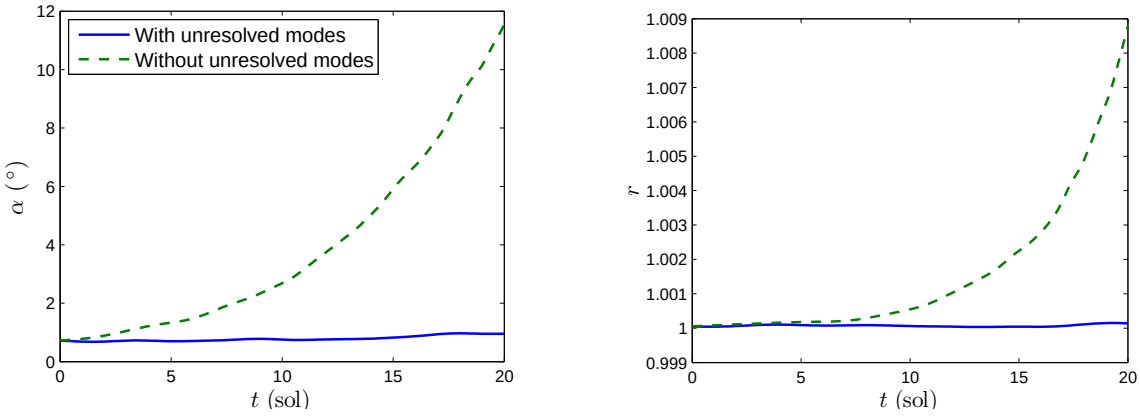
where σ_K is the lower-most sigma layer, τ_* is the restoration state to which the model is relaxed and, with the same values as in the SGCM, $1/r = 3$ sol and $\tau_R = 2$ sol are friction and radiative time constants, respectively.

5.2 POD-Galerkin models of the SGCM

5.2.1 Unresolved modes

We begin the examination of SGCM reduced models by analysing the impact of unresolved modes on an 11-EOF model. By including this number of EOFs, we ensure that the model comprises the first five SGCM EOF groups and 52% of TE in the eddy field. The original dataset constitute a set of realisations of the system. This set may be considered as a representative subset of the region in phase space where the evolution of the system takes place. A reduced model is expected to evolve in the same region in phase space as the original system. This property was used for analysing the influence of unresolved modes on the behaviour of a reduced-model. To evaluate this property, we computed the evolution of the system around the background state of (4.14), namely $\overline{\psi} = a_1\phi^{(1)}$, assuming it to be a good representative of the original system's presence in phase space.

Figure 5.1 shows the evolution of two quantities: the angle α between an 11-EOF model and the background state, and the 11-EOF model's solution norm r (nor-



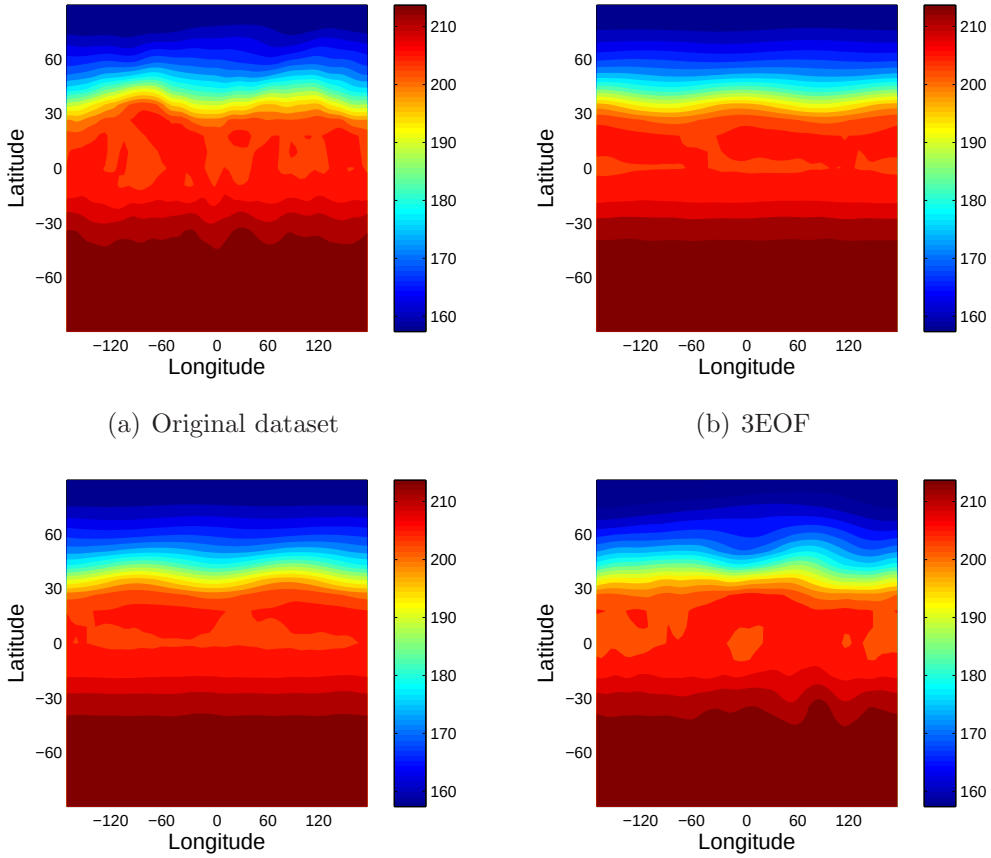
(a) (b)

Figure 5.1: (a) Angles between the average state, defined by (4.14) and an 11-EOF model solution, and (b) 11-EOF model solution norm (relative to the average state). In both frames, the 11-EOF model was integrated with and without parameterisation of unresolved modes.

malised with respect to the background state). We computed these two quantities for the cases when the reduced model runs with and without parameterisation of unresolved modes. The integration time interval was 20 sols and the initial conditions were set as the projection of the first state in the original dataset. Figure 5.1 makes it clear that even for this short period the behaviour of the inclusion of a parameterisation scheme for unresolved modes is better and, therefore, the importance of having an unresolved-mode parameterisation scheme. The state predicted by this model oscillates at around 1 ° about the average state. In contrast, the model without unresolved mode parameterisation drifts away from the average state. As a result, the angle between the average state and the solution produced by this model increases monotonically with time. Similarly, the norm of the solution from the parameterised model remains close to 1 whereas the norm of the solution from the model without parameterisation, drifts away from unity after a short period of approximately 5 sols when it remains close to the background state.

Table 5.1: Comparison between three different EOF models. Angle between original time series and model integration.

	3-EOF	5-EOF	11-EOF
α	0.9881 °	0.2602 °	2.0144 °



(a) Original dataset

(b) 3EOF

(c) 5EOF

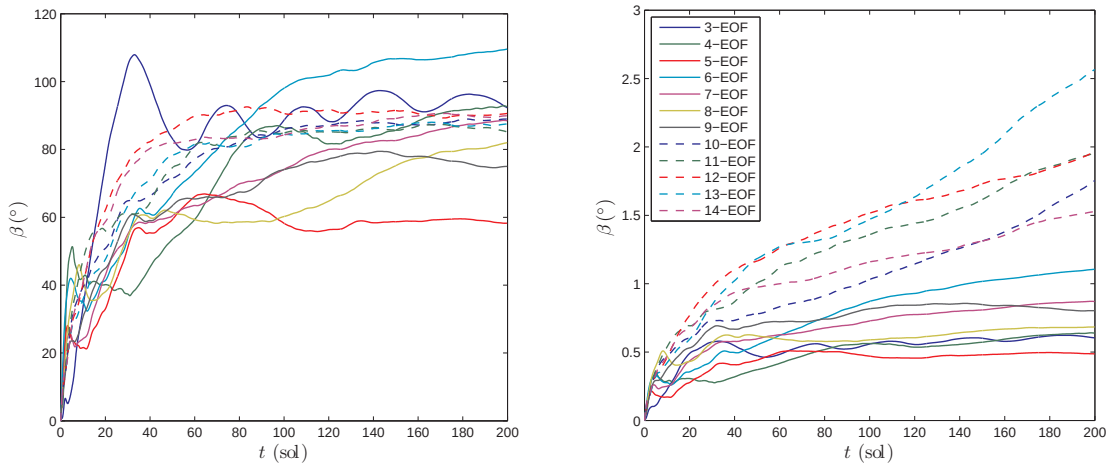
(d) 11EOF

Figure 5.2: Temperature, measured in K, at $\sigma = 0.45$ ($z \simeq 7.99$ km with respect to the surface). (a) State at $t = 100$ sol in the original dataset in comparison with a 100-sol integration of reduced models including (b) 3 EOFs, (c) 5 EOFs and (d) 11 EOFs.

5.2.2 Truncation order

In order to compare the behaviour of models at different truncation orders, three different models were integrated retaining three, five and eleven modes starting from the same initial conditions. As before, the initial conditions were set to be the projection onto the corresponding reduced space of the first state in the original dataset. Figure 5.2 shows temperature in the final states predicted by these models at $t = 100$ sols in comparison with the corresponding state in the original dataset.

Although we would have expected the 11-EOF model to show the best performance, it was, in fact, the 5-EOF model the one that produced better results. The temperature waves in the northern hemisphere (around 30 °N) seem to be around 90 ° out of phase in the 3-EOF model (figure 5.2(b)). The 5-EOF model, on the other



(a) With no background state

(b) With background state

Figure 5.3: Mean angular deviation, defined by (5.18) for different SGCM reduced models (a) with no background state and (b) including the background state.

hand, is able to reproduce these waves whereas the 11-EOF model exhibits stronger waves than the original dataset at both the northern and southern hemispheres.

These results can be confirmed by a quantitative method by looking at the angles between the original dataset and the reduced models. These angles are computed after projecting the original dataset onto the reduced space where the model lives. This is done to avoid spurious deviations as a result of superfluous modes that are not represented by the model. As pointed out in Section 3.4, the complete comparison of vectors lying on the reduced space requires the specification of a second measure that can be their relative norm. However, by simply observing the angle we can draw conclusions. The results shown in Table 5.1 confirm that the best results are produced by the 5-EOF model.

This leads to the question about how many modes are to be included in order to obtain the best approximation to the original dataset. In order to answer this question we can compare the evolution of the predicted solution with that of the original dataset by defining the mean angular deviation as

$$\beta(t) = \frac{1}{t}\int_0^t \cos^{-1}\left(\frac{\mathbf{a}^{ts}(t_1)\cdot\mathbf{a}(t_1)}{||\mathbf{a}^{ts}(t_1)||\,||\mathbf{a}(t_1)||}\right) d\,t_1, \tag{5.18}$$

where $\mathbf{a}$ and $\mathbf{a}^{ts}$ are N-dimensional vectors of coefficients arising from the integration of a reduced model and the projection of the original time series, respectively.

Figure 5.3 shows the mean angular deviation as a function of time for different

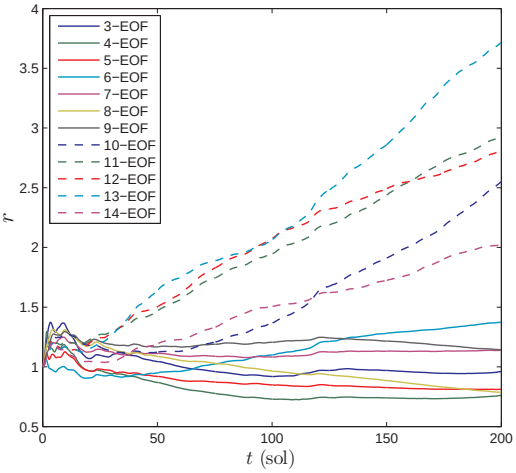


Figure 5.4: Temporal mean of the relative norm as a function of time for reduced models at different truncation orders after 200-sol integrations. The background state (EOF1) has been removed.

truncation orders from two different points. The first frame (figure 5.3(a)) shows the projection of these results onto the EOF1-subspace in order to eliminate the influence of this mode over the results. After a rapid growth of angular deviation, the mean angle difference tend to settle after 60 sols. The settlement value depends on the number of retained EOFs with a minimum of 60 ° for a 5-EOF model and a maximum around 110 ° for a 6-EOF model. So far, every model seems to behave in a similar way.

The second frame (figure 5.3(b)) shows the results as viewed from the origin of the reduced phase space (i.e. including the background state). In this figure the dissimilarities between models are more noticeable. The smallest angular deviation is produced by the 5-EOF model. Models with more than 10 EOFs display a rapid growth in angular deviation with time. This is an indication of energy accumulation in higher order modes. This phenomenon has been observed in previous studies [5] and has been removed by including additional damping to dissipate that energy [2, 3, 58, 59]. This involves the optimisation of a number of parameters that depends on the chosen diffusion scheme. Since we are interested in developing as simple a model as possible, this step was not followed so that we are able to identify the optimal reduced EOF set. However, energy accumulation is unavoidable without the implementation of additional dissipation schemes [59]. Therefore, this step must be considered essential in the development of future operational models, specially if these are intended for long-term simulations and climate studies.

For the sake of completeness, we include information on the relative norm of

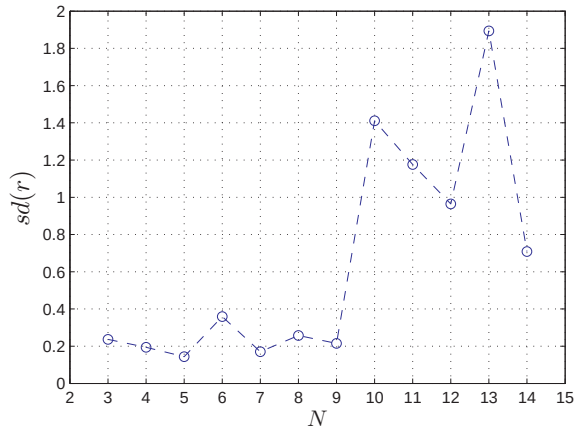


Figure 5.5: Temporal standard deviation of the relative norm for different SGCM-based reduced models after 200 sols. The background state (EOF1) has been removed.

reduced model with respect to the original time series in figures 5.4 and 5.5, which show the temporal mean (as a function of time) and the standard deviation (for the complete time interval) for the same reduced models as in figure 5.3. In this case, PC1 must be removed since it is much larger than the rest and results produced while keeping it would be strongly dominated by its influence. Figure 5.4 is in agreement with figure 5.3(b) so that the relative norm of models with $N < 10$ remains close to the expected value of 1, whereas the relative norm of models with $N \geq 10$ rapidly drifts away from such a value.

The standard deviation (figure 5.5) behaves in a similar manner. It remains small for models with $N < 9$, suddenly increasing for models with ten or more modes. From figures 5.3(b) and 5.4, we already know that models beyond $N = 10$ are drifting apart from the original orbit. Thus, it is not surprising that for $N \geq 10$ the standard deviation becomes large. Again, these results support the hypothesis of energy being accumulated at higher order modes. Moreover, these results may also be related to the statistical convergence of the EOFs. As mentioned in Section 4.4.1, the first nine modes belong to the subspace of a single zonal wavenumber, which indicates that these modes have already converged. Higher order modes spread over different zonal wave numbers, which shows that convergence has not been attained yet. Linking these results to those presented in this section, we find a clear correlation between statistical convergence and appropriate behaviour in reduced models.

5.2.3 Evolution of EOF groups

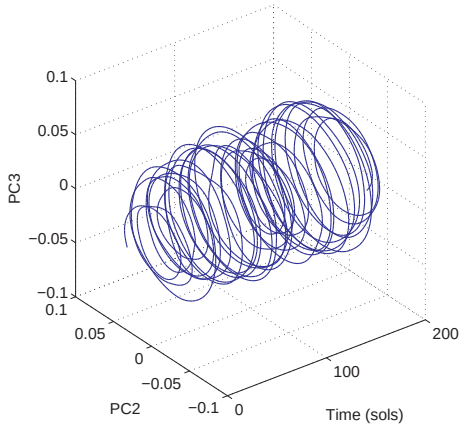
In order to analyse the response of the reduced models to different truncations we can observe the predicted evolution of EOF groups. Figure 5.6 shows a comparison between the evolution of EOF2–3 in the SGCM dataset (figure 5.6(a)) and various reduced models. In Section 4.4.1 we found that these groups represent waves with a period of 5.88 sols. Additionally, figure 5.6(a) shows that EOF2–3 exhibits a weak amplitude vacillation. A 3-EOF model fails at predicting this behaviour. Instead, it only shows a steady increase in amplitude with a period of $\mathcal{P} = 6.25$ sols (figure 5.6(b)). A 4-EOF model (not shown) reproduces the correct period but the amplitude decays steadily. It is until 5 EOFs are included that amplitude vacillation is predicted by the reduced model (figure 5.6(c)). However, the displayed behaviour appears to be more regular than in the original time series. The 9-EOF model seems to produce the best results although the predicted amplitude is smaller than the original (figure 5.6(f)).

Figure 5.7 shows the results corresponding to the pair EOF4–5. In this case, the period predicted by the model coincides with that of the SGCM dataset ($\mathcal{P} = 5.5$ sols). As in the case of EOF2–3, a 5-EOF model predicts enhanced regularity and a very weak amplitude vacillation. The models that produced the best results are the 7-EOF (figure 5.7(d)) and 9-EOF (figure 5.7(f)) models, although in both cases the amplitude is amplified with respect to the original.

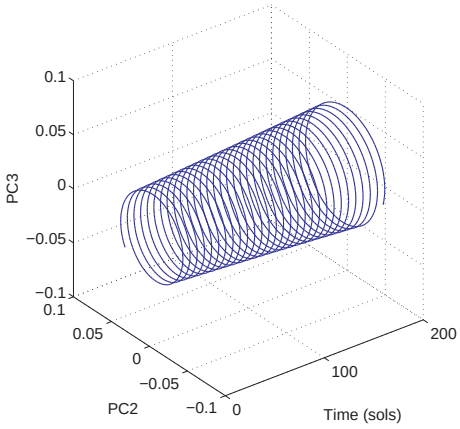
The last group included in models with $N < 10$ is the quartet EOF6–9. The evolution of this group (in the space PC6–PC7) is shown in figure 5.8. The only reduced model that completely contains this quartet is the 9-EOF model (figure 5.8(d)). For the sake of comparison, we also include the evolution predicted by the 7-EOF (figure 5.8(b))and 8-EOF (figure 5.8(c)) models. In this case, the best results are produced by the 9-EOF model specially at the beginning of the integration. As time goes on, the predicted amplitude becomes larger than the amplitude of the corresponding modes in the original dataset. At this point, we could conjecture that these result are due to the lack of resolution of modes beyond EOF9. As a result, this last group would not be allowed to interact directly with higher modes that might modulate its behaviour.

5.3 Discussion

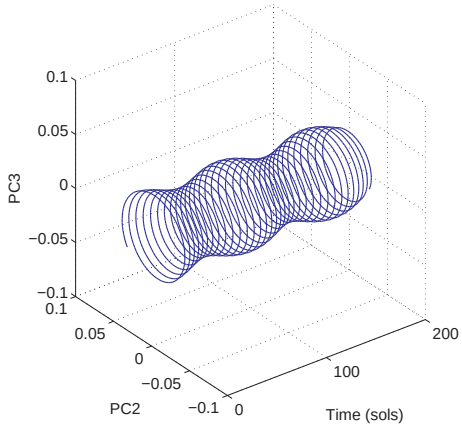
Let us conclude this section by revisiting the discussion started at the end of Section 4.2, We pointed out in that section the importance of choosing the number of



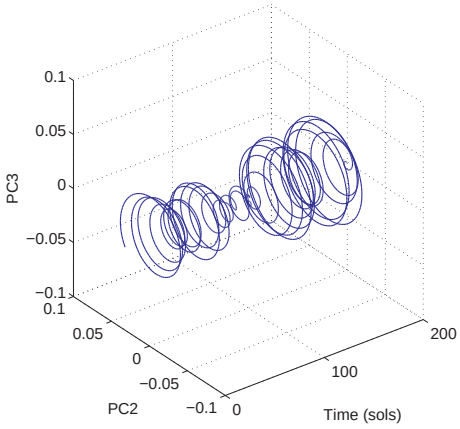
(a) SGCM dataset



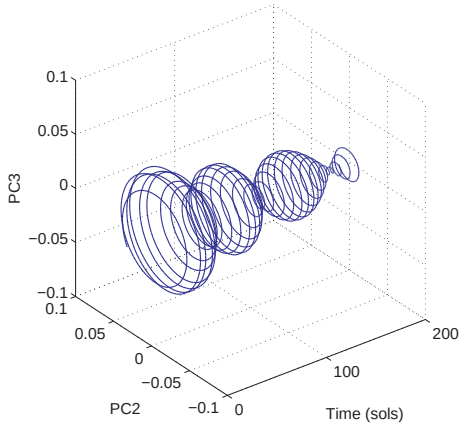
(b) 3-EOF



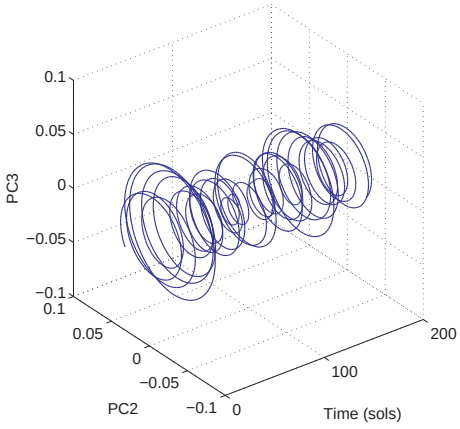
(c) 5-EOF



(d) 7-EOF

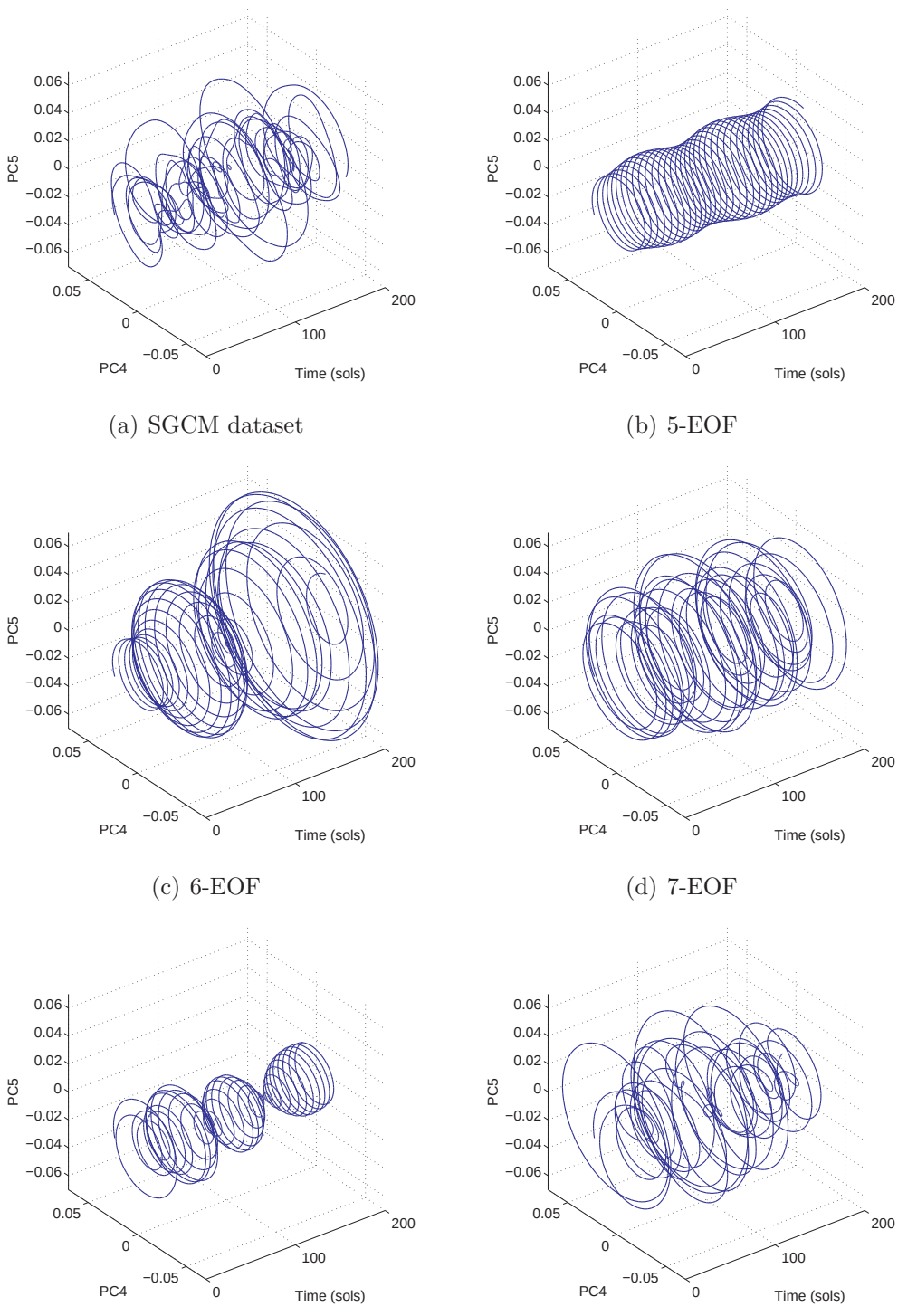


(e) 8-EOF



(f) 9-EOF

Figure 5.6: Evolution of the pair EOF2-3 as viewed from the (a) original SGCM dataset, and (b) 3-EOF, (c) 5-EOF, (d) 7-EOF, (e) 8-EOF and (f) 9-EOF models.



(a) SGCM dataset

(b) 5-EOF

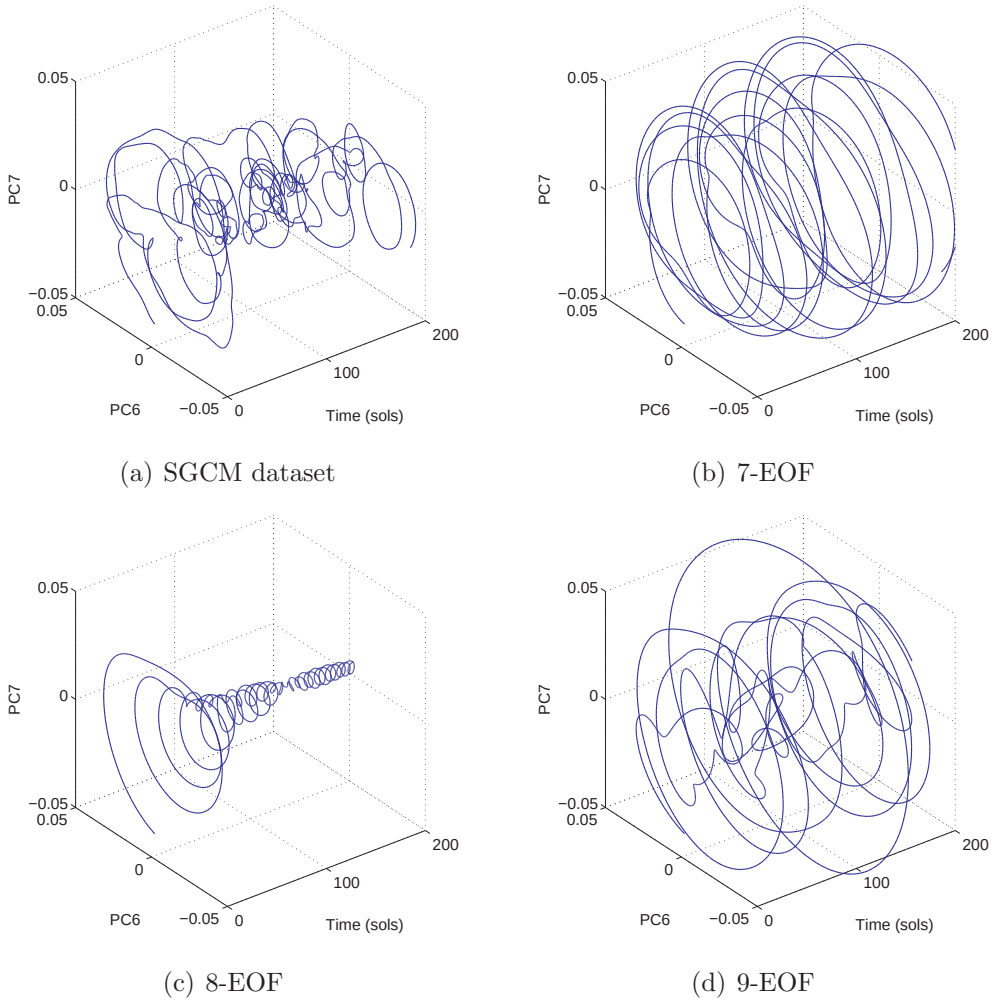
(c) 6-EOF

(d) 7-EOF

(e) 8-EOF

(f) 9-EOF

Figure 5.7: Evolution of the pair EOF4-5 as viewed from the (a) original SGCM dataset, and (b) 5-EOF, (c) 6-EOF, (d) 7-EOF, (e) 8-EOF and (f) 9-EOF models.



(a) SGCM dataset (b) 7-EOF

(c) 8-EOF (d) 9-EOF

Figure 5.8: Evolution of the pair EOF6-7 as viewed from the (a) original SGCM dataset, and (b) 7-EOF, (c) 8-EOF and (d) 9-EOF models.

modes to be retained not only by the level of energy that these represent but also by considering both the completeness of EOF groups, so that the waves represented by them are properly reproduced, and the interaction between higher order modes and the most energetic EOFs.

In Section 5.2 we have found neat examples of the relevance of these requirements. For instance, three EOFs represent 29.0% of TE in the eddy field (cf. figure 4.6) and contains the two modes of the pair EOF2-3. Still, the behaviour of a 3-EOF model is far from the expected orbit. Instead, the relative norm of the solution oscillates around unity with a quadratic mean amplitude of 5.3×10^{-4}. Including one mode more, to explain 33.5% TE in the eddy field, brings the model integration closer to the original dataset by predicting the correct period of PC2 and PC3. Therefore, we can conclude that the interaction of EOFs 1 to 3 with at least one mode of higher

order is essential to reproduce dynamics close to the original dataset.

Including four EOFs is, nevertheless, not enough to include the pair EOF4-5. Five EOFs explain 37.5% TE in the eddy field. Moreover, the behaviour of the resulting reduced model is closer to the expected dynamics. Thus, the 5-EOF model corresponds to the first minimum in figures 5.5 and with the lowest angular deviation for the time interval considered (figure 5.3). Furthermore, this model predicts not only the correct period but also shows amplitude vacillation although this is weak when compared with the original time series. Truncation order $N = 8$ corresponds to the next minimum in the relative norm standard deviation plot (see figure 5.5). This truncation order also corresponds to low angular deviation in figure 5.3. However, the model that seems to produced better results, specially when looking at the evolution of the time series, is the 9-EOF model. This might be explained by the fact that by including EOF9 another EOF group is fully represented by the resulting model.

Although separated by totally different approaches, these results are consistent with those by Whitehouse *et al.* [125] since the need to include a sufficient number of modes was also found in this work. However, as we have explained in this section this criterion seems to be insufficient to have a more accurate model. Indeed we can say that our results are more closely related to those of Crommelin and Majda [29] with respect to the somewhat inconsistent behaviour of the models as the number of modes is increased. In the next chapter we will find a further, more detailed example in the context of the MGCM.

5.4 Summary

The EOFs derived in Chapter 4 were used here as a basis on which to project (in the sense of Galerkin) the hydrostatic primitive equations in order to construct low-order models of the SGCM. The lowest order models ($N < 15$) were integrated, using as initial conditions, the projection of the original dataset on the corresponding reduced space. The results were compared with the original dataset to find that despite the importance of including sufficient modes to explain a certain amount of energy, there are other factors that are significant as well. These factors are EOF convergence, EOF groups completeness so that the waves represented by each group is properly represented and, perhaps more importantly, the influence of higher order EOFs over the dynamics of the low order modes of interest.

Chapter 6

POD of the Oxford MGCM

The Oxford MGCM was comprehensively described in Section 2.4. In Chapter 6, we shall analyse a dataset produced by this model from the point of view of the POD. After describing this dataset and its climatology, we shall present the eigenvalues and distribution of energy throughout the modes. Then, we shall relate the EOFs to the atmospheric motions present in the MGCM dataset. This will include atmospheric tides and transient waves as the most important components of the general circulation.

6.1 Oxford MGCM

As for the SGCM, the output of the MGCM consists of a set of time series for absolute vorticity, divergence, temperature and (logarithm of) surface pressure in terms of spherical harmonics coefficients at a jagged triangular resolution T31. The model uses a finite difference scheme in the vertical direction with 25 unevenly spaced sigma-levels. Table 6.1 shows the location of these levels and their equivalent altitude z in km above the planetary surface.

In order to reduce the amount of information to be handled by the POD, only 10 levels were retained. These were the levels labelled with an odd number from 3 to 21 in the GCM (see highlighted levels in Table 6.1). Thus, the highest level is situated at approximately 73.0 km above the planet's surface whereas the lowest level is about 174 m above surface level. The horizontal resolution was the same as for the SGCM, i.e. $I = 64$ longitudes and $J = 32$ latitudes for vorticity, potential vorticity and streamfunction. This corresponds to a direct transformation from spherical harmonics to physical grid without oversampling.

The MGCM output over the chosen sigma-levels was pre-processed in the same way as that coming from the SGCM in order to apply the algorithms developed for

Table 6.1: Sigma-levels and equivalent altitudes in the original dataset. The sigma-levels actually used to perform the POD are highlighted.

Level	σ	z (km)	Level	σ	z (km)
1	0.000056	97.902	14	0.455263	7.869
2	0.000250	82.940	15	0.600911	5.093
3	0.000678	72.964	16	0.731006	3.133
4	0.001542	64.747	17	0.831173	1.849
5	0.003199	57.449	18	0.899629	1.058
6	0.006290	50.688	19	0.942591	0.591
7	0.011955	44.266	20	0.968097	0.324
8	0.022181	38.085	21	0.982737	0.174
9	0.040323	32.108	22	0.990977	0.091
10	0.071666	26.357	23	0.995562	0.045
11	0.123568	20.910	24	0.998098	0.019
12	0.203945	15.899	25	0.999500	0.005
13	0.316690	11.498			

the SGCM. As before, the external Kelvin waves were filtered out using the procedure outlined in Section 4.1.2. The pre-processing produced time series of horizontal velocities, square root of potential temperature and an ensemble-average profile of surface pressure to be used as time-independent surface pressure during the analysed period.

6.1.1 Oxford MGCM dataset climatology

The MGCM dataset covers a 90-sol interval from $L_s = 267.7\,^\circ$ (sol 511) to $L_s = 323.2\,^\circ$ (sol 601). Here L_s is called areocentric longitude and gives the position of the planet in its orbit around the sun. $L_s = 0, 90, 180, 270\,^\circ$ correspond to the northern spring equinox, summer solstice, autumn equinox and winter solstice, respectively. Thus, the MGCM dataset corresponds to the autumn-winter transition in the northern hemisphere. The simulation was carried out with background MGS dust loading, i.e. an imposed dust loading that produces temperature profiles close to those observed by the MGS spacecraft [91]. The time series was sampled every two hours to make an ensemble of 1080 elements. The rather limited size of the MGCM dataset restricts the EOF convergence. Since forcing in the MGCM is not axially symmetric, in this case the well-converged EOFs are not restricted to belong to a single zonal wavenumber subspace. Based on the results in Chapter 4, we expect to achieve convergence for $N \sim \mathcal{O}(10)$ EOFs.

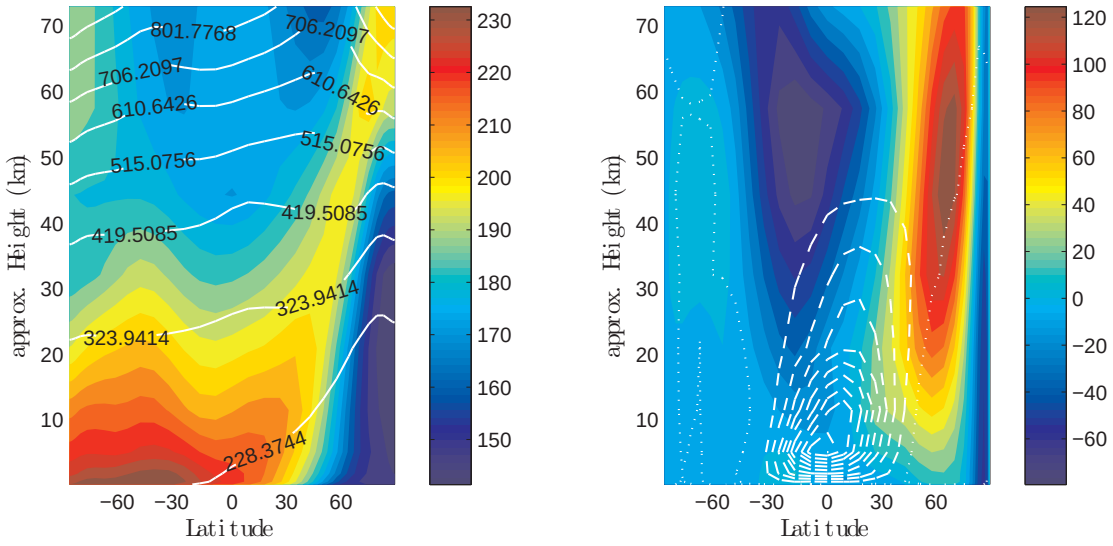


Figure 6.1: (a) Temporal and zonal average of temperature and potential temperature (white contours), and (b) zonal wind and meridional mass transport streamfunction (white contours): solid, dashed and dotted lines represent positive, negative and null contours, respectively. The separation between contours is $\Delta\chi = 129.2$ kg/s^3.

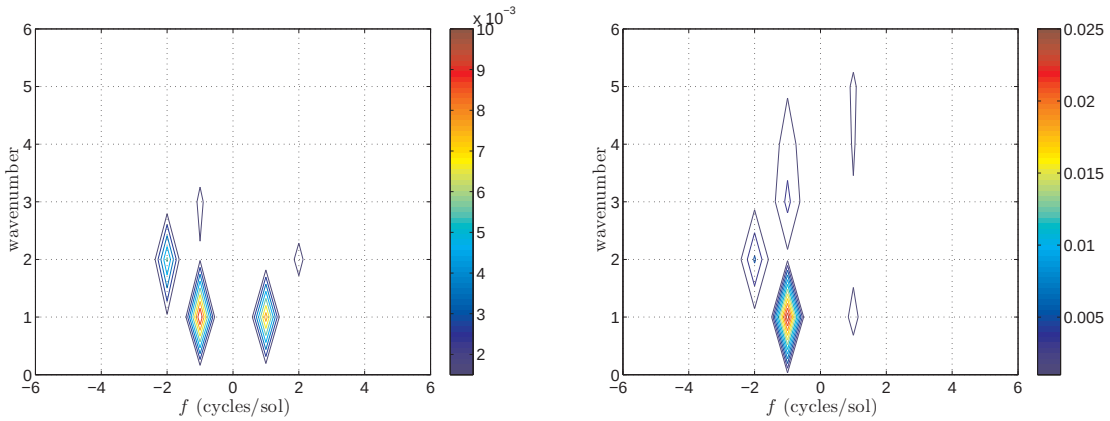
Zonal mean

Figure 6.1 shows the temporal and zonal average of temperature, potential temperature and zonal wind in the MGCM dataset. Figure 6.1(b) also shows the meridional mass transport streamfunction χ defined, in sigma-coordinates, as (see e.g. [38])

$$\overline{p^s v} = -\frac{1}{\cos\phi}\frac{\partial \chi}{\partial \sigma}, \quad \overline{p^s \dot{\sigma}} = \frac{\partial \chi}{\partial y}, \tag{6.1}$$

where $\overline{(\cdot)}$ and $dy = \cos\phi\, d\phi$ indicates zonal average, so that positive streamfunction indicates counter-clockwise circulation.

Figure 6.1 presents an atmospheric structure similar to that presented in figure 4.3 for the SGCM dataset. As in that case, the atmosphere is stably stratified everywhere. Moreover, the steep horizontal temperature gradient between the equator and the north pole gives rise to a very strong winter westerly jet in the middle atmosphere above the northern hemisphere, reaching speeds of 140 m/s in average. This jet is accompanied by a weaker easterly jet above the equator. An important difference with respect to the SGCM dataset appears in the temperature profile (figure 6.1(a)): the so-called winter polar warming that takes place in the middle atmosphere between



(a) (b)

Figure 6.2: Thermal tides in the MGCM dataset as seen in (a) surface pressure (cf. figure 5.9 in [91]), and (b) temperature at $\sigma = 0.943$ (approximate height $z = 591$m above the planetary surface). Both figures correspond to an equatorial longitude circle. Negative frequency indicates westward travelling waves.

50 and 80 km [91]. This phenomenon is related to the Hadley cell, which in Mars extends up to 80 km above the surface and spans almost the whole planet [126, 36]. Notice again the strong asymmetry of the Martian solstice general circulation, which contrasts with the nearly symmetric circulation found on the Earth. As explained in Section 4.1.1, this is mainly due to the shorter radiative time scale as a result of the absence of oceans [40].

Atmospheric tides

There are two important differences between the SGCM and the MGCM datasets: the presence of diurnal and seasonal cycles and the inclusion of topography. For this study, only the diurnal cycle and topography are relevant since the seasonal cycle is not being considered due to the assumption of time-independent surface pressure.

The presence of the diurnal cycle gives rise to atmospheric tides. These are atmospheric motions whose characteristic periods are harmonics of the diurnal cycle. They can be classified as sun-synchronous (migrating) and longitude-dependent (non-migrating) tides. The most significant migrating tides are the diurnal and semidiurnal tides (see, e.g. [127]).

The diurnal tide is a wave travelling westwards, characterised by a wavenumber one and a period of 1 sol. The semidiurnal tide is also a westward travelling wave, though it is characterised by a wavenumber two and a period of half a sol. While

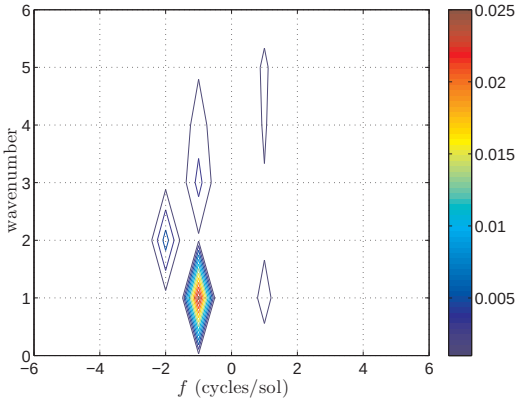


Figure 6.3: Thermal tides in the MGCM dataset as seen in temperature at the equator and at $\sigma = 0.943$ (approximate height $z = 591$m above the planetary surface) after the external Kelvin waves have been removed. Negative frequency indicates westward travelling waves.

the diurnal tide is mainly the response of the atmosphere to surface heating, the semidiurnal tide reflects the vertical radiative response of the atmosphere. Hence, the activity of the semidiurnal tide has a central influence in the atmospheric response to phenomena such as dust storms, when the vertical radiative properties are changed due to the enhance presence of dust loadings [63].

Among the non-migrating tides, the diurnal Kelvin wave is one of the most important. This wave is characterised by a wavenumber one and a period of 1 sol but travelling eastwards (see, e.g. [131, 35, 63]). It arises from the interaction between the diurnal tide and topography dominated by a strong wavenumber-2 pattern [131]. Hence, it can only be simulated in a model which includes appropriate topographic information (see [35, 91]). The interaction between the diurnal tide and topography also gives rise to a westward propagating wavenumber-3 wave with a period of 1 sol.

Figure 6.2 is a graphical representation of atmospheric tides present in the MGCM dataset as seen by two different fields. Figure 6.2(a) corresponds to surface pressure (cf. figure 5.9 in [91]); figure 6.2(b) presents the same information as seen by the temperature at an approximate height of 591 m above the planetary surface. These figures show the tidal amplitude obtained by the double Fourier transform (in time and longitude) of a surface pressure (or temperature) time series at the equator [91, 63].

The strongest signal in both fields corresponds to the diurnal cycle. The semidiurnal tide and the diurnal Kelvin wave are present in both fields as well as the westward propagating diurnal wavenumber-3 wave arising from the interaction between diurnal

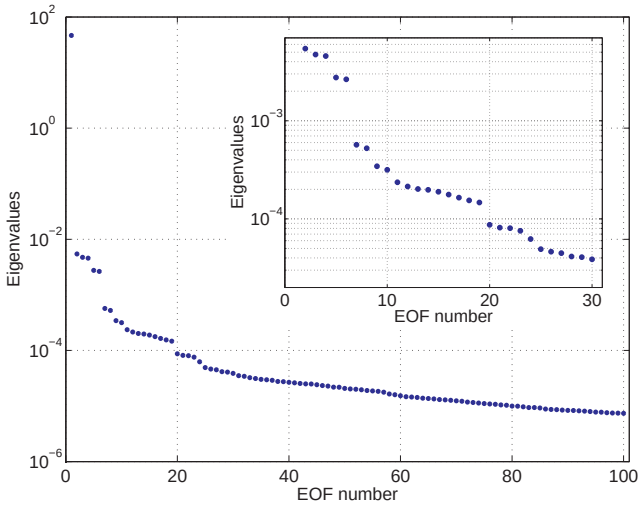


Figure 6.4: Eigenvalues obtained from the POD of the Oxford Mars GCM.

tide and topography. However, these waves have different weight on each field. In the surface pressure signal, the diurnal Kelvin wave is the second strongest followed by the semidiurnal tide. In contrast, in the temperature field at the level of observation, the semidiurnal mode is the second most significant followed by the westward propagating diurnal wavenumber-3 wave. This is simply an indication of vertical tidal structure.

Figure 6.3 is equivalent to figure 6.2(b) after external Kelvin waves have been filtered. Apart from slight differences in the amplitudes these two figures are virtually the same. These figures show that the action of filtering external Kelvin waves as a consequence of the time-independent surface pressure assumption, does not affect the atmospheric structure as far as the thermal tides are concerned.

6.2 Eigenspectrum

To begin with the POD analysis of the Oxford MGCM, we look at the empirical eigenspectrum. Figure 6.4 shows the Oxford MGCM empirical eigenvalues for the autumn-winter transition in the northern hemisphere. As expected, the first mode contains by far most of the atmospheric TE on average, the energy content decreasing as the EOF index increases. It is interesting to compare this figure with its equivalent for the SGCM (figure 4.5). The first noticeable feature is that, unlike the SGCM dynamics, the MGCM eigenspectrum splits into more well-defined groups of modes of comparable energy content. Using the diagram in figure 6.4 and Fourier analysis, it

Table 6.2: MGCM leading EOFs groups and characteristic periods. When doublets and triplets in a pair or trio have different second periods, these appear in parentheses.

Group	Type	EOFs	Leading period (sol)	Second period (sol)
1	single	1	seasonal	(very weak) diurnal
2	single	2	seasonal	(weak) diurnal
3	pair	3, 4	diurnal	
4	pair	5, 6	3.6	3.2
5	pair	7, 8	4.5	4.1
6	pair	9, 10	semidiurnal	
7	pair	11, 12	1.73	(10.0,1.83)
8	trio	13 to 15	1.63	(1.47,1.73,1,66)
9	pair	16, 17	10.0	(1.63,1.01)
10	pair	18, 19	1.01	0.99
11	single	20	12.8	8.2
12	pair	21, 22	1.80	1.70
13	pair	23, 24	4.7	(5.6,2.9)
14	single	25	7.5	10.0
15	pair	26, 27	1.12	1.38
16	continuous	≥ 28	various	various

is possible to characterise and classify these groups according to energy and dominant periods present in their corresponding principal components (PCs) time series. There are at least sixteen groups as shown in Table 6.2. These can be formed by either one, two or three modes. We can refer to the components of these groups as singlets, doublets and triplets, respectively. The continuous tail starts with EOF28.

Singlets can be interpreted as standing waves. The first and second singlets carry information about the seasonal trend. Since the models that will be constructed do not consider the seasonal cycle, these features are not expected to be resolved by the low-order models. When the seasonal trend is removed, the first and second singlets reveal the influence of the diurnal cycle. The ratios $\mathrm{PC1}_{rms}/\lambda_1^{1/2} = 1.35 \times 10^{-4}$ and $\mathrm{PC2}_{rms}/\lambda_2^{1/2} = 1.23 \times 10^{-1}$, where X_{rms} is the quadratic mean of X and λ_k is the k-th empirical eigenvalue, show the relative significance of the diurnal cycle to each of these modes. This influence is more noticeable in PC2 than in PC1.

As in the SGCM case, doublets represent travelling waves. There are two special cases which are characterised by a single period. These are pairs 3 and 6, characterised by a diurnal and semidiurnal period, respectively, which seems to indicate a relation between these groups and the diurnal and semidiurnal tides. Other pairs can be

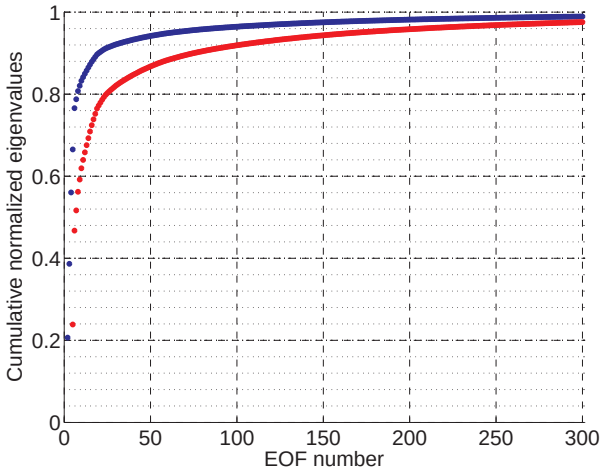


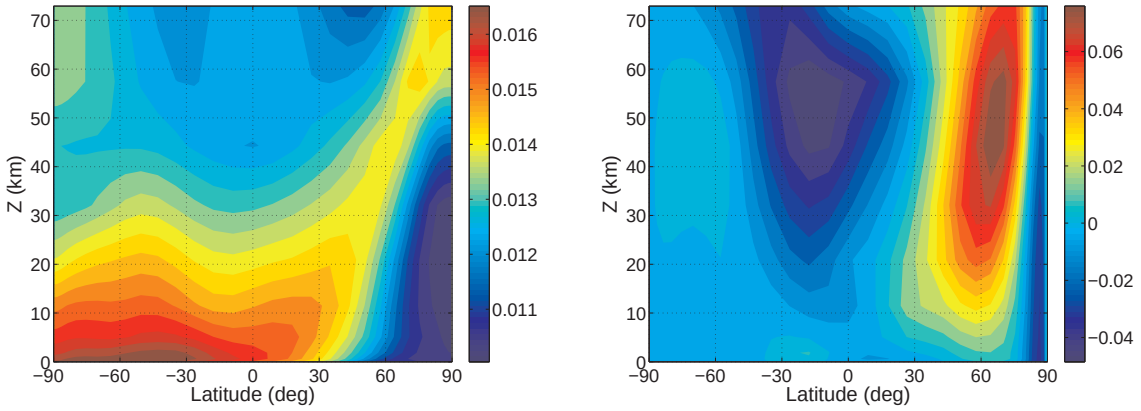
Figure 6.5: Cumulative energy in the Oxford Mars GCM normalised without considering background energy (in blue), and without considering background and diurnal heating energy (in red) with respect to TE in 1079 and 1076 EOFs, respectively.

classified by their secondary period. The second period can be either the same for the pair or different for each member.

6.2.1 Energy accumulation over the EOFs

An interesting point of comparison between the SGCM and the MGCM is given by the cumulative energy diagrams. Figure 6.5 shows the normalised cumulative eddy TE for the MGCM for two cases: (*i*) after deletion of background energy and (*ii*) after deletion of background and diurnal heating energy (associated with EOF3–4, see Table 6.2 and Section 6.4). These cases were normalised with respect to average eddy TE in 1079 and 1076 EOFs, respectively.

The first apparent characteristic is the rapid growth in the first modes reaching 80% with only 7 EOFs and 90% in 20 EOFs when no background energy is considered. This represents a marked contrast with the SGCM case where it was necessary to retain at least 151 modes to explain 90% of the eddy TE in that dataset (see figure 4.6). Even in the absence of diurnal modes (i.e. EOF3–4), it is only necessary to retain 75 modes to explain 90% eddy TE. This fact might indicate that the region of phase space where the MGCM evolves can be represented by fewer modes than the phase space of the SGCM. This statement, might be true also for Mars itself. This feature is due to the presence of the diurnal cycle which tends to accumulate the energy in the first modes corresponding to diurnal motions.



(a) (b)

Figure 6.6: (a) Zonal average of temperature and (b) zonal velocity in the MGCM EOF1 (both in arbitrary units). Positive horizontal velocity directed eastwards.

6.3 EOF1 and EOF2

As in the SGCM, EOF1 (figure 6.6) represents a background state, closely related to the time average of the dataset. This can be seen by comparing figures 6.6 and 6.1. The horizontal structure of EOF1 shows a clear topographic influence throughout the vertical (figure 6.7).

EOF2 is a correction to the mean term. It also shows influence from topography, specially in the southern hemisphere where two anti-cyclonic structures are located in the regions of the Argyre Planitia ($\sim 45\,^{\circ}$W, $50\,^{\circ}$S) and Hellas Planitia ($\sim 60\,^{\circ}$E, $40\,^{\circ}$S) (figure 6.8(a)). However, this influence is limited to the lower atmosphere ($z < 20$ km). Above this level a zonally symmetric component becomes more significant (figure 6.8(b)).

6.4 Higher order EOFs and atmospheric motion

In this section, we shall identify some of the most important atmospheric motions to the MGCM EOFs. The motions to be considered are atmospheric tides and transient waves.

6.4.1 Diurnal tide

Group 3 is formed by EOF3 and EOF4, which share a similar structure. The corresponding coefficients also share a similar behaviour clearly showing the influence

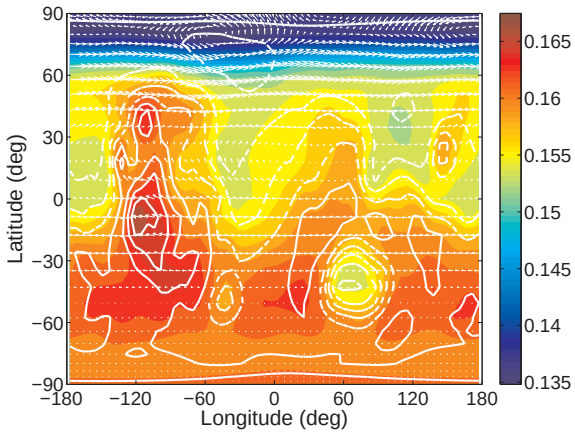


Figure 6.7: Longitude-latitude map at $\sigma = 0.0403$ ($z =\simeq 32.11$ km) showing square root of potential temperature in the MGCM EOF1. The arrows represent horizontal velocity and solid and dashed white lines represent positive and negative topography, respectively.

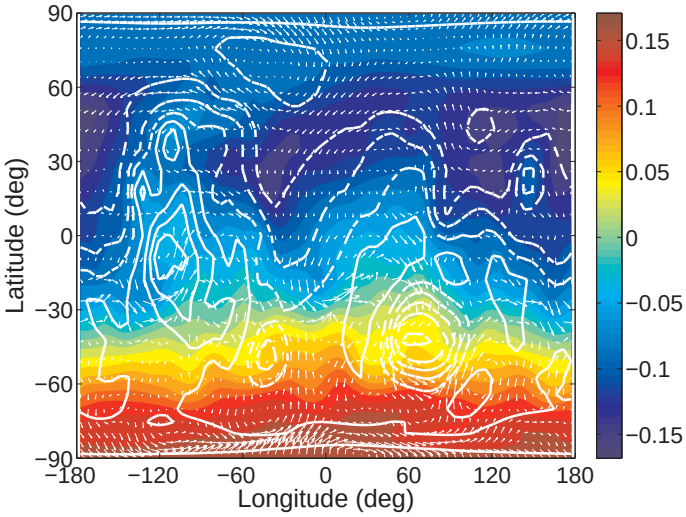
of the diurnal cycle. For example, we focus our attention on EOF3 (figure 6.9). It is a zonal wavenumber-one wave, stronger in the southern hemisphere as expected, since the original dataset corresponds to the transition from spring to summer in that hemisphere [127]. The vertical structure of EOF3 has the expected behaviour as well (cf. [127, 91]). The maximum amplitude is located near the surface with a vertically propagating wavenumber-one wave above the equator. Away from the tropics the wavenumber-one wave is vertically trapped in the lower atmosphere.

Despite the striking structural resemblance between EOF3–4 and the diurnal tide, we must recall at this point that the POD is not explicitly designed to extract atmospheric tides. Instead, the tidal modes are expected to appear distributed over more than one EOF. This distribution was computed as the normalised time-average TE in every POD mode after projecting the diurnal tide over the first 30 EOFs. Thus,

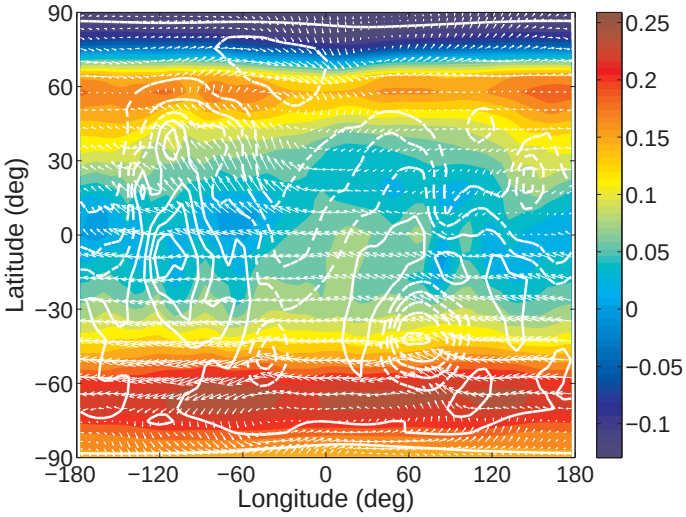
$$E_k^d = \frac{1}{Z} \sum_{j=1}^{M} \left[a_k^d(t_j) \right]^2 , \tag{6.2}$$

where $a_k^d(t_j) = (\phi^{(k)}, \psi^d(t_j))$ is the projection of the diurnal tide component $\psi^d(t_j)$ over the k-th EOF at the discrete time $t_j = j\Delta t$ with $\Delta t = 1/12$ sol, M is the number of snapshots and $Z = \sum_{k=1}^{N} \sum_{j=1}^{M} \left[a_k^d(t_j) \right]^2$, is a normalising factor with $N = 30$ as the number of retained EOFs.

The tides were computed using an extended version of the method outlined by Lewis and Barker in [63], which consists of a double Fourier transform in longitude

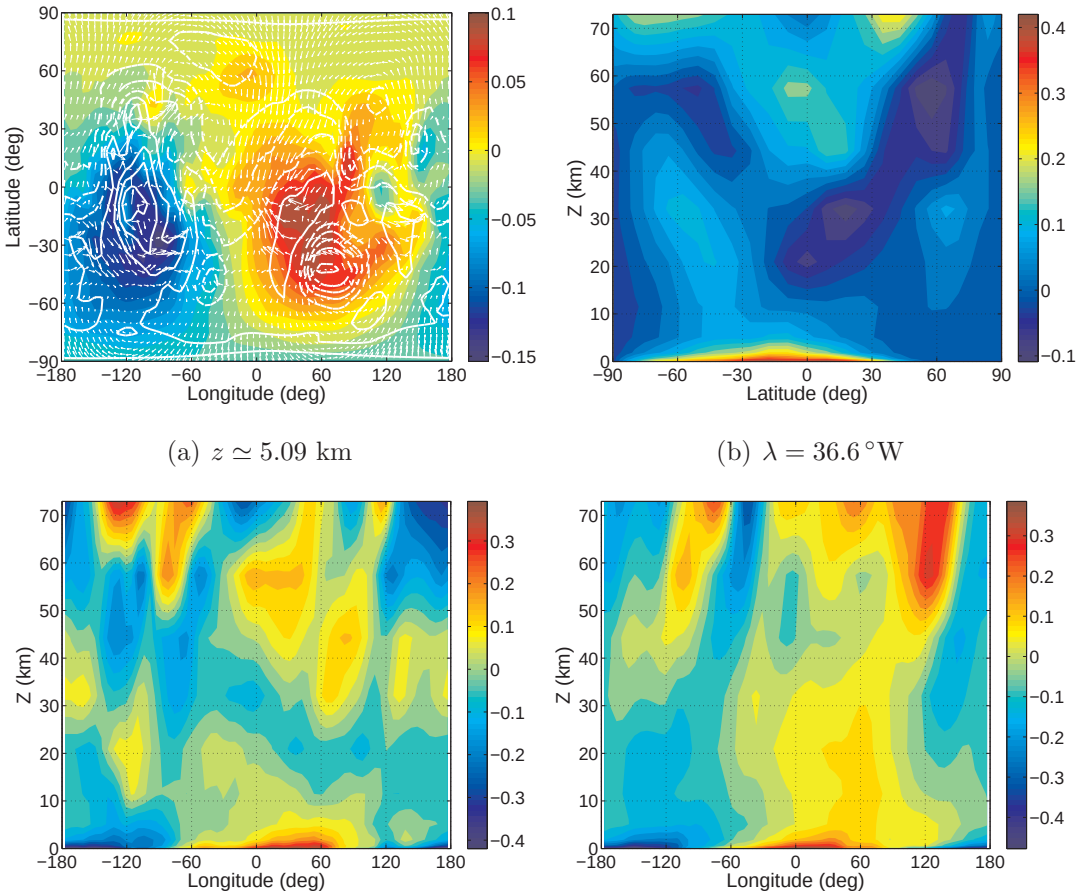


(a)



(b)

Figure 6.8: Longitude-latitude map at (a) $\sigma = 0.8312$ ($z \simeq 1.85$ km) and (b) $\sigma = 0.0403$ ($z =\simeq 32.11$ km) showing the square root of potential temperature in the MGCM EOF2. The arrows represent horizontal velocity and solid and dashed white lines represent positive and negative topography, respectively.



(a) $z \simeq 5.09$ km

(b) $\lambda = 36.6\,^{\circ}$W

(c) Equator

(d) $\phi = 34.9\,^{\circ}$S

Figure 6.9: Square root of potential temperature in MGCM EOF3: (a) Longitude-latitude map at $\sigma = 0.6$ ($z \simeq 5.09$ km above the planetary surface) (the arrows represent horizontal velocity); (b) latitude-altitude section at $\lambda = 36.6°$W, and longitude-altitude sections (c) at the equator and (d) $\phi = 34.9°$S.

and time along every latitudinal circle and height. Keeping only the frequency n and wavenumber s of interest, an inverse Fourier transform was applied to recover the tides over physical grid points. This procedure was applied to two velocity components and temperature. The tidal component in τ was computed as

$$\tau' = -\mathfrak{T} + \mathfrak{T}\left(1 - \frac{T'}{P\mathfrak{T}^2}\right)^{1/2} \simeq \frac{T'}{2P\mathfrak{T}}, \tag{6.3}$$

where T' is the temperature tidal component under analysis and $\mathfrak{T}$ is the τ component in the background state given by (4.14).

The decomposition of the diurnal tide ($n = -1$ sol^{-1} and $s = 1$) over the EOFs is shown in figure 6.10. As expected, the pair EOF3–4 contains 96.6% of TE whereas

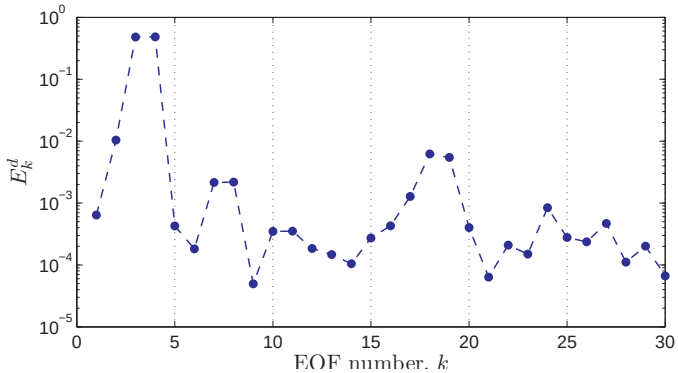


Figure 6.10: Diurnal tide decomposition over 30 EOFs.

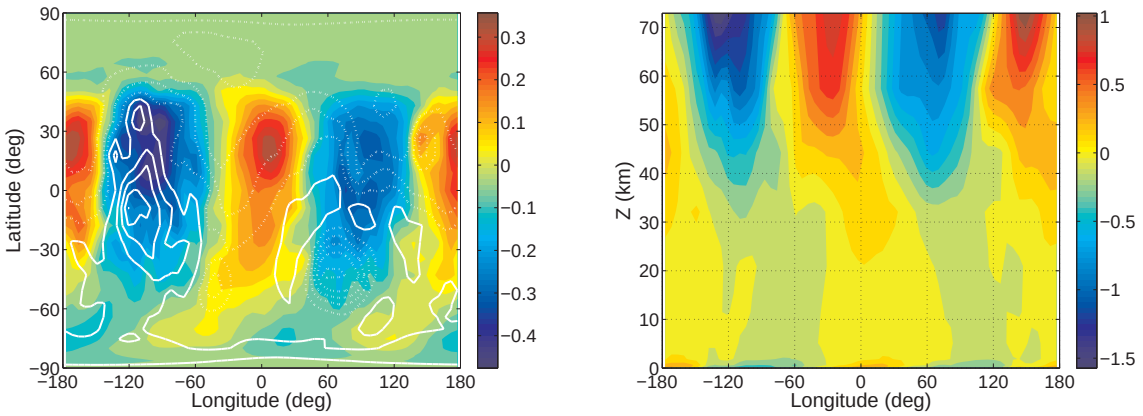
the next most energetic group, EOF18–EOF19, contains 1.2%. In conclusion, the pair EOF3–4 is a good approximation to the diurnal tide in the reduced space. Alternatively, we can say that the diurnal tide is the most energetic atmospheric mode after the background state, in the dataset under analysis.

6.4.2 Semidiurnal tide

EOF9–EOF10 are zonal wavenumber-two waves (figure 6.11(a)). Like the semidiurnal tide, which has a vertical wavelength > 100 km [132, 91], the pair EOF9–10 have a long vertical wavelength (figure 6.11(b)). However, the semidiurnal tide ($n = -2$ sol^{-1} and $s = 2$) is not well represented by only one EOF pair. Performing the same analysis as for the diurnal tide case, we find that Group 6 indeed explains 82% of TE in the semidiurnal tide (figure 6.12). The second most energetic pair is the pair EOF5–EOF6, which contains 14.9% of TE, which is still significant especially since EOF5–6 are more energetic than EOF9–10 when the full atmospheric motion is considered.

Indeed, the MGCM EOF5 power spectrum shows the signature of semidiurnal motion (figure 6.13). However, this signature is weak compared with the already reported maxima in these time series, namely 3.6 and 3.2 sols (cf. table 6.2). A similar case is found when computing the distribution of the diurnal Kelvin wave, shown in figure 6.14(a). These two cases reveal the nature of the EOFs as combinations of waves that can be further separated by means of appropriate techniques. This topic will be revisited in Section 6.4.4 where transient baroclinic waves are to be projected onto the EOF basis.

As an additional comment to the decomposition of the diurnal and semidiurnal tide, we must mention, for example, the case of pair EOF9-10 whose components



(a) $z \simeq 174$ m (b) $\phi = 17.8°$N

Figure 6.11: Square root of potential temperature, in arbitrary units, in the MGCM EOF9: (a) longitude-latitude map at the lower-most sigma-level and (b) longitude-altitude section at 17.8°N.

appear to contain different amounts of TE in the diurnal tide decomposition (figure 6.10). This feature indicates that a travelling wave of zonal wavenumber two, with the characteristics given by EOF9-10, is not a component of the diurnal tide. Therefore, these EOFs behave independently like two standing waves which, nevertheless, modulate the amplitude of the tide. Similarly, a zonal wavenumber-1 wave with the characteristics given by the pair EOF4-3 is not a component of the semidiurnal tide. As a result, EOF3 and EOF4 exhibit different TE content in the semidiurnal tide decomposition (figure 6.12).

6.4.3 Diurnal Kelvin and wavenumber-3 waves

As discussed in Section 6.1.1, the interaction between the diurnal tide and the topography, dominated by a wavenumber-2 pattern, gives rise to two waves. The diurnal Kelvin wave is a wavenumber-1 wave which propagates eastwards with period of 1 sol. The second wave that arises from the interaction diurnal tide-topography is a wavenumber-3 wave which propagates westwards with period of 1 sol.

The decomposition of the diurnal Kelvin wave ($n = 1$ sol^{-1} and $s = 1$) is shown in figure 6.14(a). The best representation of this wave is given by group 5 (EOF7-EOF8). However, this group only explains 60% of TE in this wave. The rest is distributed over various modes without clear localised energy peaks.

The diurnal wavenumber-3 wave ($n = -1$ sol^{-1} and $s = 3$), on the other hand, is distributed among a larger number of EOFs that include the pairs EOF3–4 and

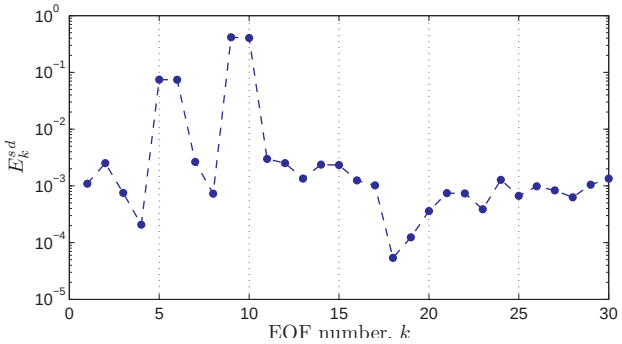


Figure 6.12: Semidiurnal tide decomposition over 30 EOFs.

EOF18–19, already identified with a period of 1 sol. There are other individual EOFs such as EOF12, 13 and 15, that contribute to the representation of this wave. The additional comments given at the end of Section 6.4.2 are equally valid for these individual EOFs. The incomplete appearance of pair EOF11-12 and trio EOF13-15 is due to the fact that the travelling waves represented by these groups are not integral part of the diurnal wavenumber-3 wave: only individual components appear in its decomposition modulating its intensity.

6.4.4 Transient waves

Generally, transient waves arise as a consequence of a combination of barotropic and baroclinic instabilities and, as for the Earth, they are responsible for the energy transport polewards from the equator. These waves are not characterised by a definite wavenumber or period. Hence, the separation of these waves by means of double Fourier transform in space and time is not as clear as in the case of tidal components.

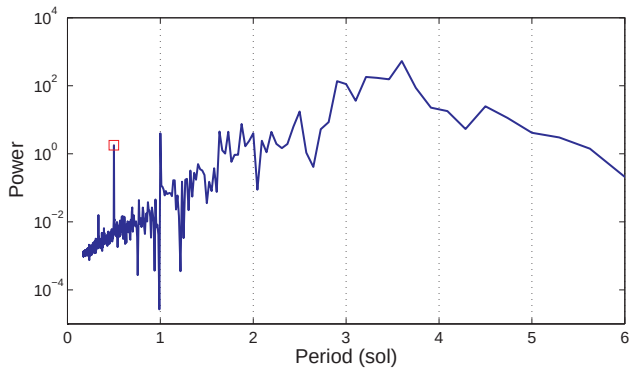
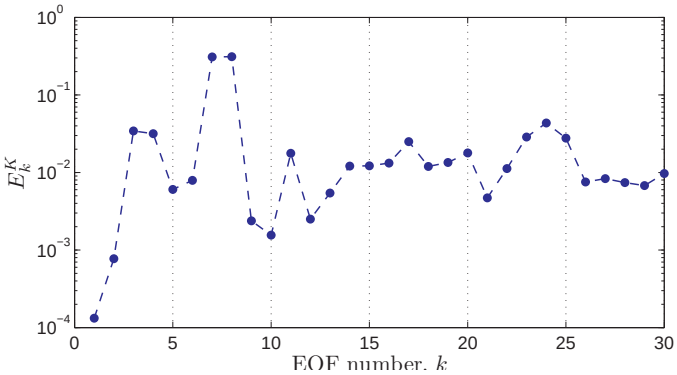
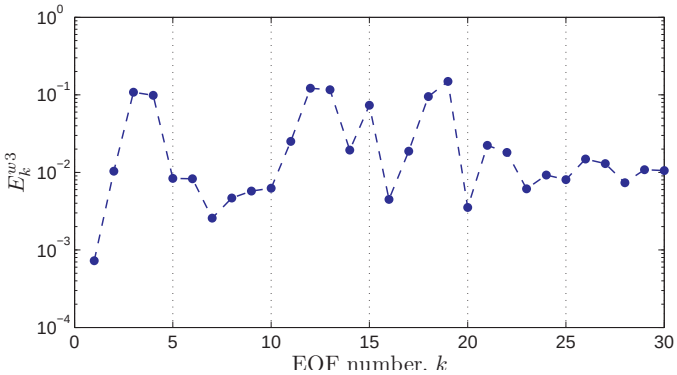


Figure 6.13: MGCM EOF5 power spectrum and the contribution of the semidiurnal tide (red square at 0.5 sol).



(a) Diurnal Kelvin wave



(b) Diurnal wavenumber-3 wave

Figure 6.14: Distribution over 30 EOFs of (a) the diurnal Kelvin wave and (b) westward propagating diurnal wavenumber-3 wave.

Figure 6.15 shows an attempt to extract transient waves from the MGCM dataset by subtracting the tides from the complete fields. From this figure we can see that the most important contribution comes from EOF5-6. EOF7–8, 11–12, 13–15 for a set of secondary contributory EOFs to the transient-wave representation.

EOF5–6, with a leading period of 3.6 sol and a second common period of 3.2 sol is an example of an EOF group with a single secondary period. This group is constituted by disturbances that resemble baroclinic waves. These waves were expected since the season under analysis correspond to an active period of baroclinic activity (typically from late autumn to early spring in each hemisphere [91]) Figure 6.16 shows contours of sigma-velocity and superimposed vectors of horizontal velocity. Recalling that $w \propto -\dot{\sigma}$, it can be seen that air masses moving northwards go upwards whereas air masses moving towards the equator go downwards. This structure is maintained from the surface to an altitude of about 40 km. However, the thermal wave extends to higher altitudes (figure 6.17). As we have seen in Section 6.4.2, the structure that

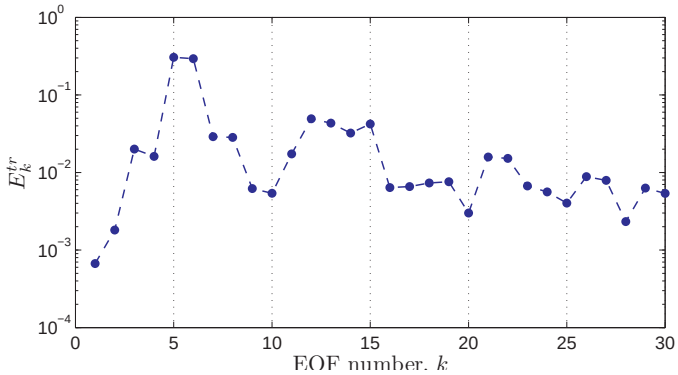


Figure 6.15: Distribution of transient waves over 30 EOFs.

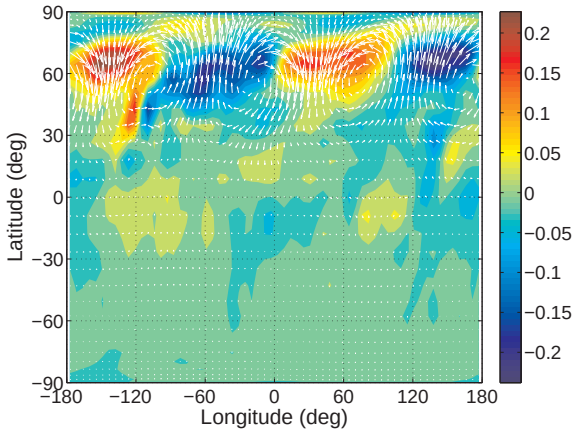
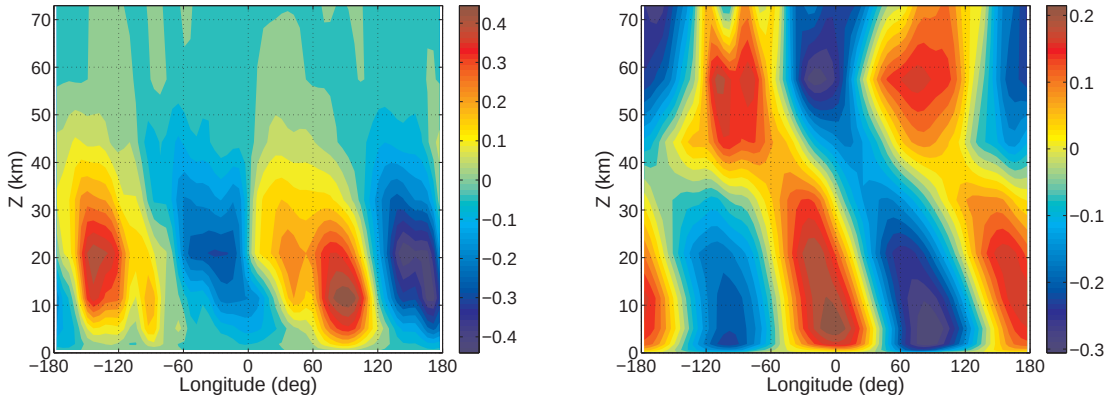


Figure 6.16: MGCM EOF5 sigma-velocity contours on a longitude-latitude map at $z \simeq 32.1$ km. The arrows represent horizontal velocity.

rises into higher altitudes is in fact part of the semidiurnal wave that appears as a secondary component of EOF5-6.

Figure 6.18 shows a longitude-latitude map and a longitude-vertical section of EOF7 as a representative of the pair EOF7–8. This pair consists of a wavenumber-1 travelling wave with characteristics of a baroclinic wave located around 70°N . In fact, the structure of this wave is similar to that of the pair EOF3–4. The sloping convective structure is maintained up to an altitude of 30 km whereas the thermal wave reaches heights up to 60 km although it reaches higher altitudes at higher latitudes (almost 70 km at around 80°N).

The pair EOF11-12 is an example of a pair in which each component has its own secondary frequency. Both of them exhibit sloping convection near the surface (figure 6.19). However, these two modes have very different structures. The surface wave in EOF11 seems to have a dominant zonal wavenumber two whereas the wave in EOF12 has a dominant zonal wavenumber three extending up to around 40 km.



(a) Sigma-velocity

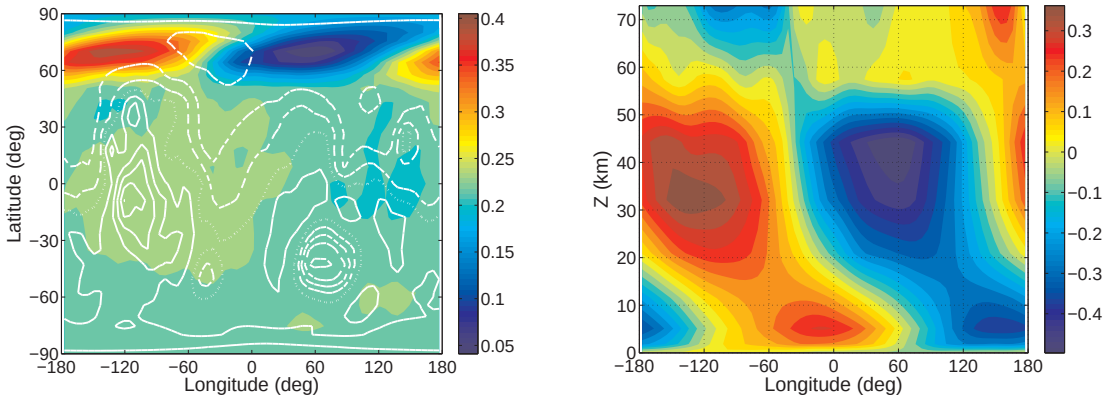
(b) Square root of potential temperature

Figure 6.17: Longitude-altitude sections at $\phi = 64.3\,^{\circ}$N showing (a) sigma-velocity and (b) square root of potential temperature in MGCM EOF5.

Figure 6.20 shows these two modes at two different heights. From this point of view, the wave in EOF11 appears as a wave with dominant zonal wavenumber one restricted to higher latitudes in the northern hemisphere. In turn, EOF12 appears to have the structure of a zonal wavenumber-3 wave near the surface, although this is lost at around $z \simeq 30$ km, constraining baroclinic wave-like structures to the lower atmosphere, where horizontal temperature gradients are stronger.

Each component in EOF13-15 has its own secondary period. Figure 6.21 shows longitude-latitude maps of τ in these three modes at two different altitudes. Near the surface, all of them exhibit a zonal wavenumber-3 wave structure. These waves can be associated with sloping convection and, therefore, these modes can be identified with baroclinic waves. However, around $z \simeq 30$ km this structure is lost in EOF14 and 15 to give place to zonal wavenumber-one waves in both modes. EOF13 reaches around 40 km keeping the wavenumber-3 structure.

As we can see from these examples, EOF groups with more than one secondary period have a very complex structure and rather than interpreting them as individual travelling waves they must be thought of mixtures of waves that modulate the general circulation. In fact, this same statement is strictly valid for any other EOF group but, as we have seen before, when all the components in a group are defined by the same periods (leading and secondary), its characterisation becomes simpler.



(a) $z \simeq 32.1$ km (b) $\phi = 70.2\,^{\circ}\mathrm{N}$

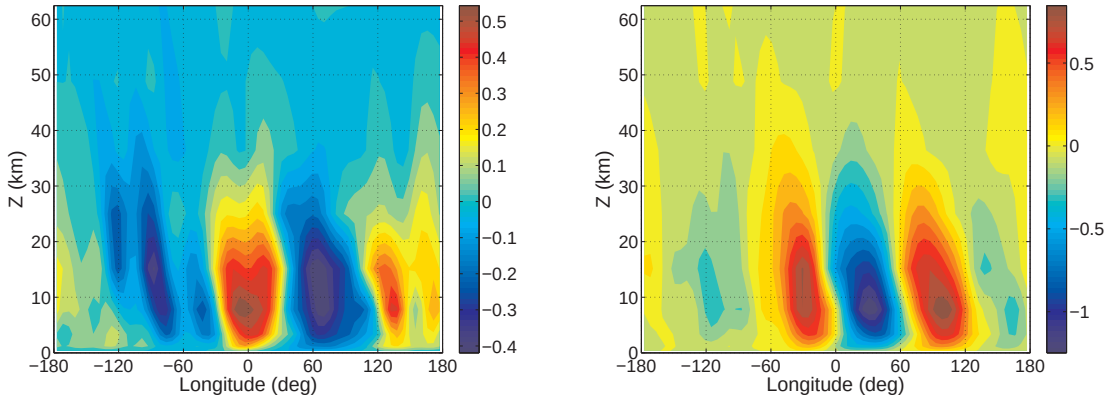
Figure 6.18: (a) Longitude-latitude map at $z \simeq 32.1$ km, and (b) longitude-altitude section at $\phi = 70.2\,^{\circ}\mathrm{N}$ of the square root of potential temperature in MGCM EOF7.

6.5 Further remarks

The various atmospheric motions present in the MGCM dataset are much richer than their counterparts for the SGCM. This variety is fundamentally due to the inclusion of topography and the diurnal cycle. Accordingly, the EOFs extracted from the POD of the MGCM dataset exhibit an enhanced richness in structure when compared with the SGCM (cf. Section 4.4).

Through the identification of the first EOFs with the atmospheric modes represented by them, we have reached a clearer physical interpretation of the EOFs not only as the most energetic atmospheric modes but also as actual representatives of those atmospheric motions that have been identified as primary driving components of the general circulation on that planet, namely thermal tides and transient waves (see e.g. [91]).

This work is comparable, though not directly, to the work by Martínez-Alvarado *et al.* [74]. Following the analysis by Whitehouse *et al.* [124] on the SGCM, Martínez-Alvarado *et al.* [74] carried out a decomposition of the same MGCM dataset in vertical and horizontal modes, using the QG vertical structure equation, respectively. The norm employed in the POD was a TE-norm, though it was defined as the sum of KE and APE, rather than total PE, as it is in the present study. Despite these differences in the method, the results are similar, especially when considering the structure of the first modes. In [74], EOF1 and EOF2 are characterised by a diurnal period corresponding to EOF3 and EOF4 in the present thesis. Furthermore, EOF3 and EOF4 mainly correspond to transient waves as EOF5 and EOF6 do in this work.



(a) EOF11 (b) EOF12

Figure 6.19: Sigma-velocity longitude-altitude sections at $\phi = 70.2\,^{\circ}\mathrm{N}$ in MGCM (a) EOF11 and (b) EOF12.

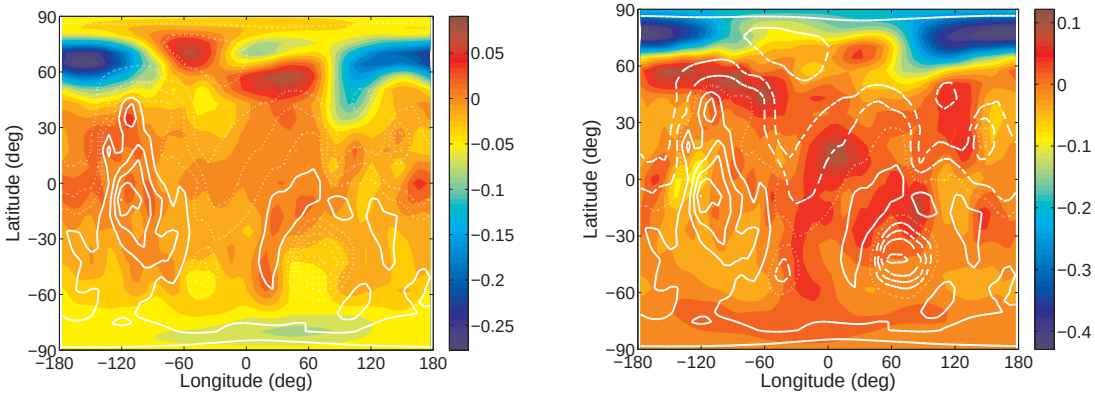
The apparent shift in order is simply due to the fact that in [74] the fields were separated in zonal (axisymmetric) and eddy parts.

It is appropriate here to point out that these modes are not expected to be invariant with respect to the dataset used in their derivation. Instead, a different set of EOFs is likely to be extracted from a different dataset corresponding not only to different seasons but also to the same period over different conditions, for example, under different dust loadings or at different sampling rate. In fact, changing the sampling rate would change the Nyquist frequency, i.e. the maximum resolvable frequency to avoid aliasing. As a result, the range of frequencies captured by the EOFs. Depending on the length of the time series and the actual value of the Nyquist frequency, the resolution of the PCs evolution might also be affected.

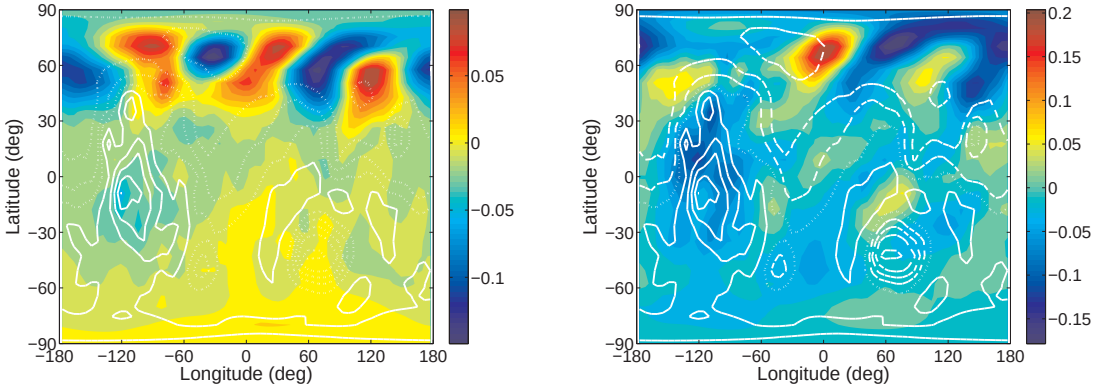
This constraint can possibly be relaxed by including as many and as varied snapshots in the original dataset from where the EOFs are to be obtained. However, a dynamical model derived from a given set of EOFs and, therefore from a given dataset, can be expected to reproduce not only that particular dataset but intermediate states as there is no restriction on the coefficients (PCs). This effectively means that, for example, the intermediate states between two given ensemble elements can be interpolated as to obtain an accurate approximation.

Another operation that could lead to the production of a different set of EOFs can be, for example, data assimilation. Nevertheless, this might be advantageous if the analyses rather than the background states[1] are considered the elements of the

[1]In the data assimilation jargon, the background state is the initial conditions first guess whereas the analysis is the best initial obtained after data assimilation.



(a) EOF11



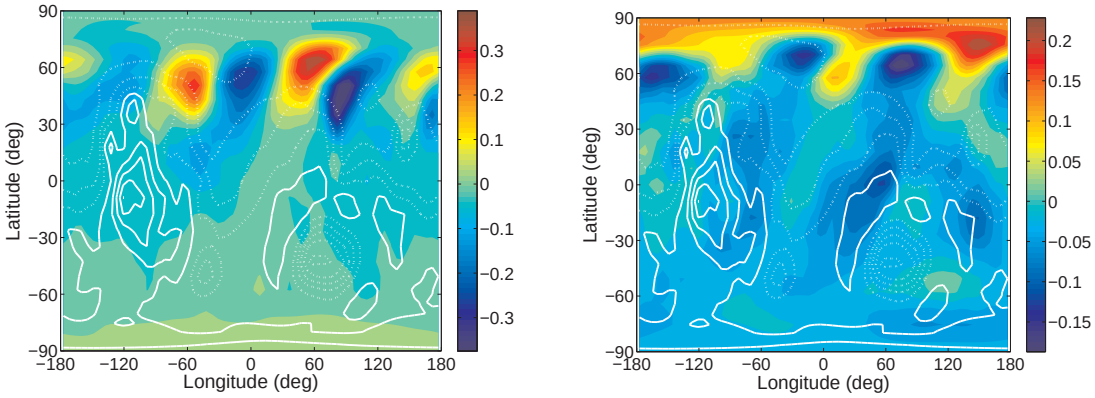
(b) EOF12

Figure 6.20: Longitude-latitude maps of square root of potential temperature, in arbitrary units, at $\sigma = 0.3167$ ($z \simeq 11.5$ km) (left column) and $\sigma = 0.0403$ ($z \simeq 32.1$ km) (right column) for MGCM (a) EOF11 and (b) EOF12.

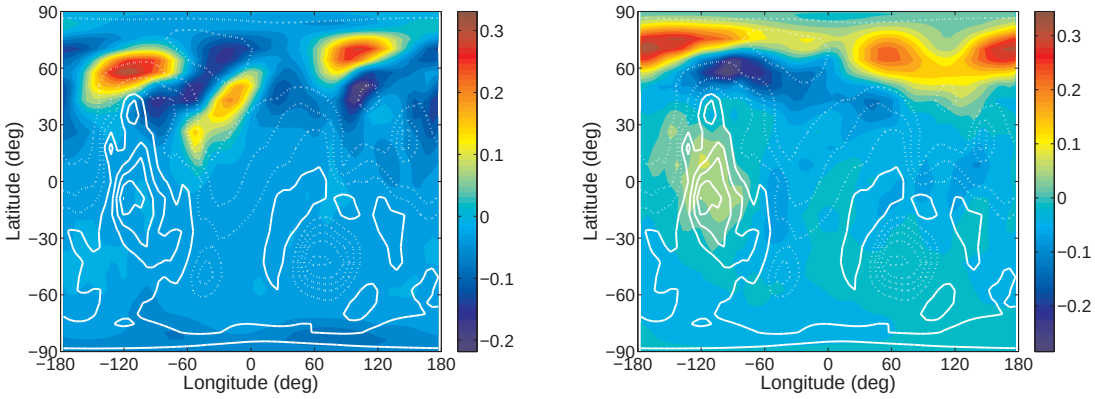
source ensemble on which to perform the POD. The reason for this would be that the analyses are expected to lie or, at least, be close to the slow manifold.

6.6 Summary

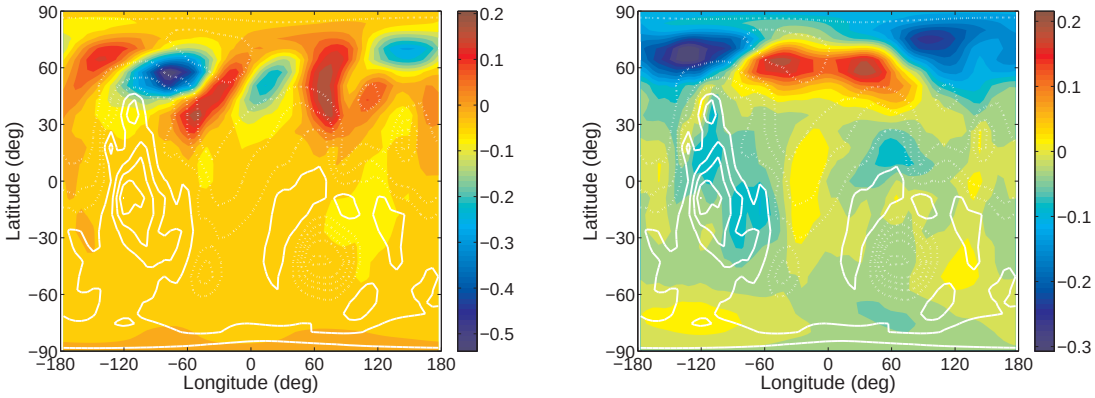
In this chapter we have extracted the most energetic atmospheric modes in the MGCM dataset by means of the POD. As for the SGCM dataset, we have also found the eigenvalues (related to TE content) of the MGCM. The cumulative energy in this model appears to grow more rapidly than its counterpart for the SGCM. This can be explained by the strong energy content posed on the first EOFs, directly forced by the diurnal cycle. This result seems further to support the hypothesis that motivated this work: that the atmospheric dynamics of Mars lies essentially on a phase space



(a) EOF13



(b) EOF14



(c) EOF15

Figure 6.21: Contours of square root of potential temperature in MGCM (a) EOF13, (b) EOF14 and (c) EOF15 at $\sigma = 0.8312$ ($z \simeq 1.9$ km) (left column) and $\sigma = 0.0403$ ($z \simeq 32.1$ km) (right column).

of low-dimension, especially when compared to the Earth.

The EOFs produced by the POD over the MGCM dataset were classified by their energy contents, three-dimensional structure and Fourier spectra. This procedure allowed their identification with important atmospheric motions such as tides and transient baroclinic waves. The EOFs and their classification will be used in Chapter 7 in order to produce low-order models of the MGCM in the same manner as we did in Chapter 5 regarding the SGCM.

Chapter 7

Reduced-order models of the MGCM

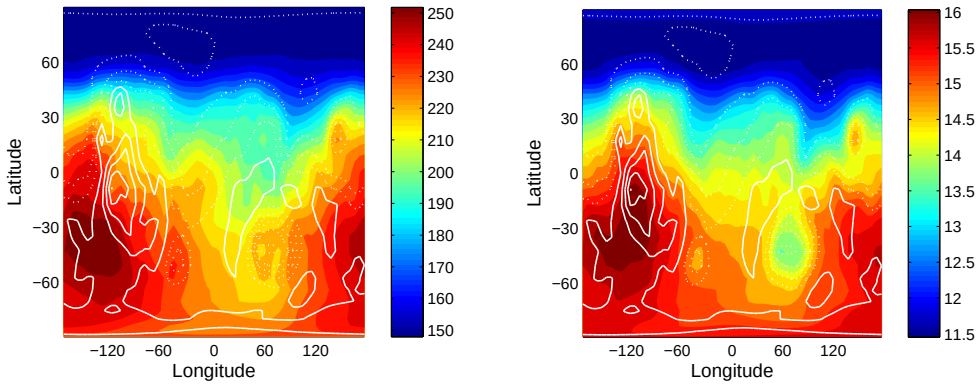
In Chapter 7, we shall project the primitive equation onto the span of the leading EOFs, derived in Chapter 6, in order to construct low-order models of the MGCM. The development of low-order models of the MGCM presents different challenges from those found in the SGCM case. First, the forcing in the source model is a rather complicated mixture of parameterisations to emulate relevant physical processes. Secondly, the large scale Martian topographic structures gives rise to important characteristics of the general circulation (cf. Section 6.1.1).

After introducing an appropriate forcing for the MGCM reduced models, we shall assess their behaviour by direct comparison with the original MGCM dataset. By analysing a variety of reduced models we shall find indications of essential interactions between modes. This will be further confirmed by carrying out numerical experiments where the influence of the diurnal cycle is investigated. These experiments will lead to significant conclusions regarding the nature of certain interactions between tidal and baroclinic modes.

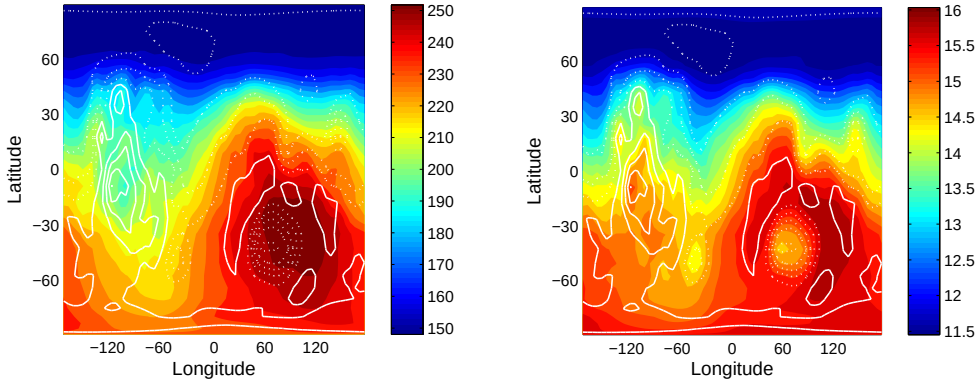
7.1 MGCM vs SGCM: a more realistic Mars

7.1.1 Forcing and physical parameterisations

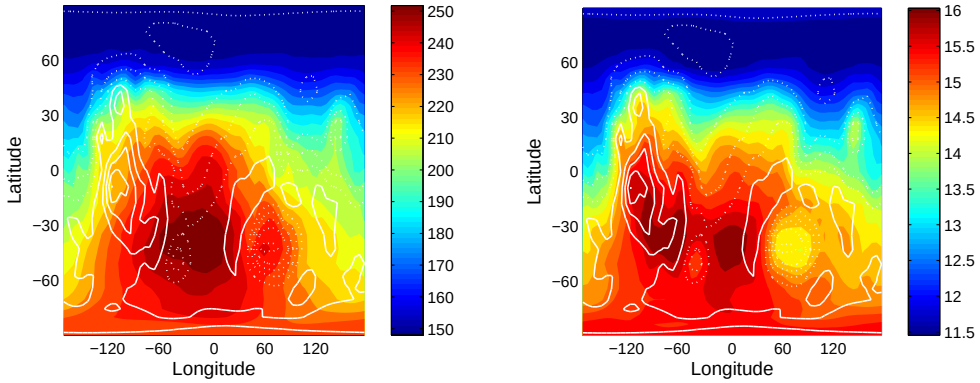
Considering that the physical parameterisation represents 90% of the computing time used in the integration of a GCM, including such a complex set of implicit parameters would go against the spirit of a low-order model, at least, at this stage. For this reason, Newtonian cooling and Rayleigh friction were used as a first approach.



(a) 00:00 MTC



(b) 08:00 MTC



(c) 16:00 MTC

Figure 7.1: Restoration temperature T_* (left column) and corresponding restoration state τ_* (right column) for different times of the day at an approximate height of 174 m. Solid and dashed lines represent positive and negative topography, respectively. (MTC = Mars Time Coordinate)

The diurnal cycle was emulated by relaxing the atmosphere towards a periodic state with period 1 sol. This was constructed by averaging temperature in the original time series for each available sol fraction (recall that the original time series is sampled every 2 Martian hours, i.e. there are 12 samples/sol). Mathematically it can be written as

$$T_*(\mathbf{x}, t_j) = \frac{n}{M} \sum_{d=1}^{M/n} T(\mathbf{x}, d + t_j), \tag{7.1}$$

where M is the number of snapshots in the time series, n is the number of snapshots in 1 sol, $\mathbf{x}$ represents spatial coordinates and $t_j = j/n$ for $j = 0, 1, 2, \ldots, n-1$ is sol fraction. In this case $n = 12$, so that the restoration temperature was sampled every two Martian hours. The actual relaxation state used in the model was

$$\tau_*(\mathbf{x}, t) = (\sigma p^s(\mathbf{x}))^{-\kappa/2} \sqrt{T_*(\mathbf{x}, t \,(\mathrm{mod} 1))}, \tag{7.2}$$

so that (5.16) and (5.17) were used to compute forcing using the same relaxation times as for the SGCM case. These relaxation times lie in the interval thought to be appropriate for Mars [26, 50], though a more extensive analysis would be required to determined the sensitivity of reduced models to these parameters.

The modulo operator (mod) in (7.2) is necessary to make τ_* periodic with the required period of 1. Note that in this case, the restoration temperature is not zonally symmetric but fully dependent on the three spatial variables. This is necessary to get realistic models of Mars not only because of the diurnal cycle signal but also because of the huge Martian topography that generates a considerable longitudinal temperature difference that remains at any time between high volcanoes and deep basins as shown in figure 7.1.

7.1.2 Topography

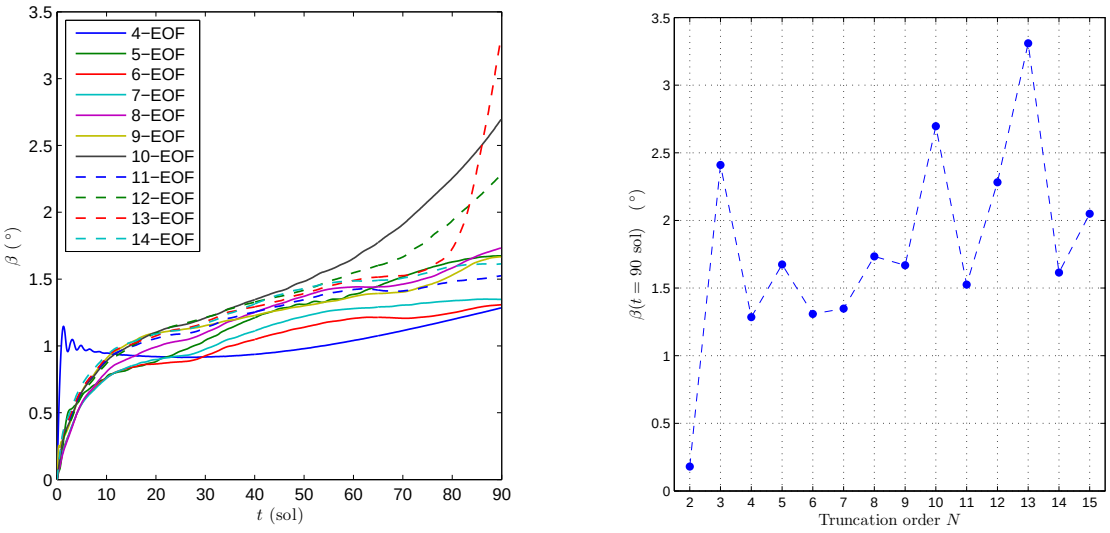
Topography enters the equations of motion through the surface geopotential Φ_s in the pressure gradient term in (5.1). However, as was pointed out by Achatz and Opsteegh [3], filtering external Kelvin waves out, through the process described in 5.1, would leave Φ_s out from the equations (since Φ_s only appears under the gradient operation and the filtering process involves taking the curl of the horizontal component of the equations of motion). Therefore, the topography must be included indirectly by means of the time-independent surface pressure, as it is done in [3]. The time-independent surface pressure is taken as the average of surface pressure throughout the MGCM dataset, containing the topographic signal.

7.2 POD-Galerkin models of the MGCM

7.2.1 Angles and relative norms

We begin the study of POD-Galerkin models of the MGCM by looking at the mean angular deviation and the relative norm. The mean angular deviation is presented without projecting it onto the EOF1 subspace since, like in the MGCM case, this would only show that a nearly constant phase difference between the full model and its reduced counterpart is present for every reduced model. Figure 7.2(a) shows the comparison of eleven reduced models corresponding to different truncation orders. It is equivalent to figure 5.3 for the SGCM, where a similar behaviour can be observed. There are some models that remain close to the original time series for longer time intervals. As before, there is no clear tendency as the number of retained EOFs increases. Instead, after a maximum in angular deviation, a minimum may follow and after a large angular deviation we may find a sudden fall. Perhaps the best example of this phenomenon is found between $N = 13$ and $N = 14$ where, after reaching more than $3\,^\circ$, the deviation decreases to $\beta(t = 90 \text{ sol}) \simeq 1.5\,^\circ$. This is shown in figure 7.2(b) where β is plotted as a function of truncation order at $t = 90$ sol. According to this figure, the best models appear to be those based on 4, 6, 7, 11 and 14 EOFs. Notice that the 2-EOF model yields the smallest angular deviation. Nevertheless, this is a trivial model unable to reproduce any travelling wave motion, a characteristic that can be made extensive to the 3-EOF model.

Figure 7.3 shows results regarding the relative norm after projecting onto the EOF1 subspace. The time mean relative norm deviation after 90 sols is shown in figure 7.3(a) (recall that a relative norm of unity is what we are aiming at). The corresponding standard deviation is shown in figure 7.3(b). The agreement between these two figures is fairly satisfactory. The large deviations in the relative norm time mean and the standard deviation at $N = 3$ is not surprising since this model is unable to capture any wave motion potentially present in the atmosphere. Apart from this case, in both figures the models based on 10 and 13 EOFs exhibit the worse behaviour. This fact is clearer in the plot of the time mean relative norm deviation. In the MGCM case, we can draw no conclusion regarding the statistical convergence of the modes since there is no clear separation between low-order converged and higher order non-converged modes as we found in the MGCM. Although some disagreement may arise as what model shows the best behaviour, we notice that figure 7.3(a) closely resembles figure 7.2(b) (for $N > 3$). In particular, we notice that there is no clear tendency in the accuracy of the reduced models as more EOFs are retained. Similar



(a) (b)

Figure 7.2: Mean angular deviation, defined by (5.18), in MGCM reduced models for different truncation orders (a) as a function of time and (b) at $t = 90$ sol.

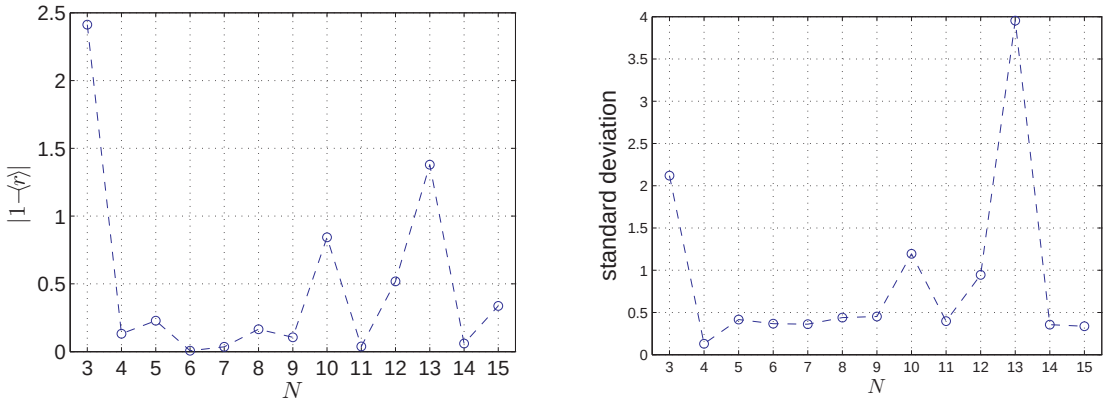
information can be extracted from figure 7.3(b).

7.2.2 A physical point of view

Let us look at these results from a more physical point of view. The frames in figure 7.4 show zonal mean temperature, zonal wind and mean meridional streamfunction averaged over the whole time interval under analysis for five different low-order models in comparison with the original dataset. The results depicted in figure 7.4 correspond to those models that exhibited the best behaviour in terms of mean angular deviation, namely 4-, 6-, 7-, 11- and 14-EOF models (cf. figure 7.2(b)).

Despite subtle discrepancies, all of them look structurally similar. This structure corresponds to what is expected for early northern winter: a steep horizontal temperature gradient near the surface in the winter hemisphere, a middle-atmosphere winter westerly jet, a weaker easterly jet above the tropics and winter polar warming (cf. 6.1.1).

The differences between original and reduced model depend on the truncation order. For instance, 4-EOF model predicts stronger winds and hot temperatures going up to higher altitudes than the source model. In contrast, higher truncation models show weaker easterly jets. The predicted meridional circulation is weaker as well. As a result, the predicted Hadley cells appear to extend to lower altitudes while



(a) (b)

Figure 7.3: Relative norm for different MGCM-based reduced models: (a) deviation of temporal mean from unity, and (b) standard deviation. The background state (EOF1) has been removed.

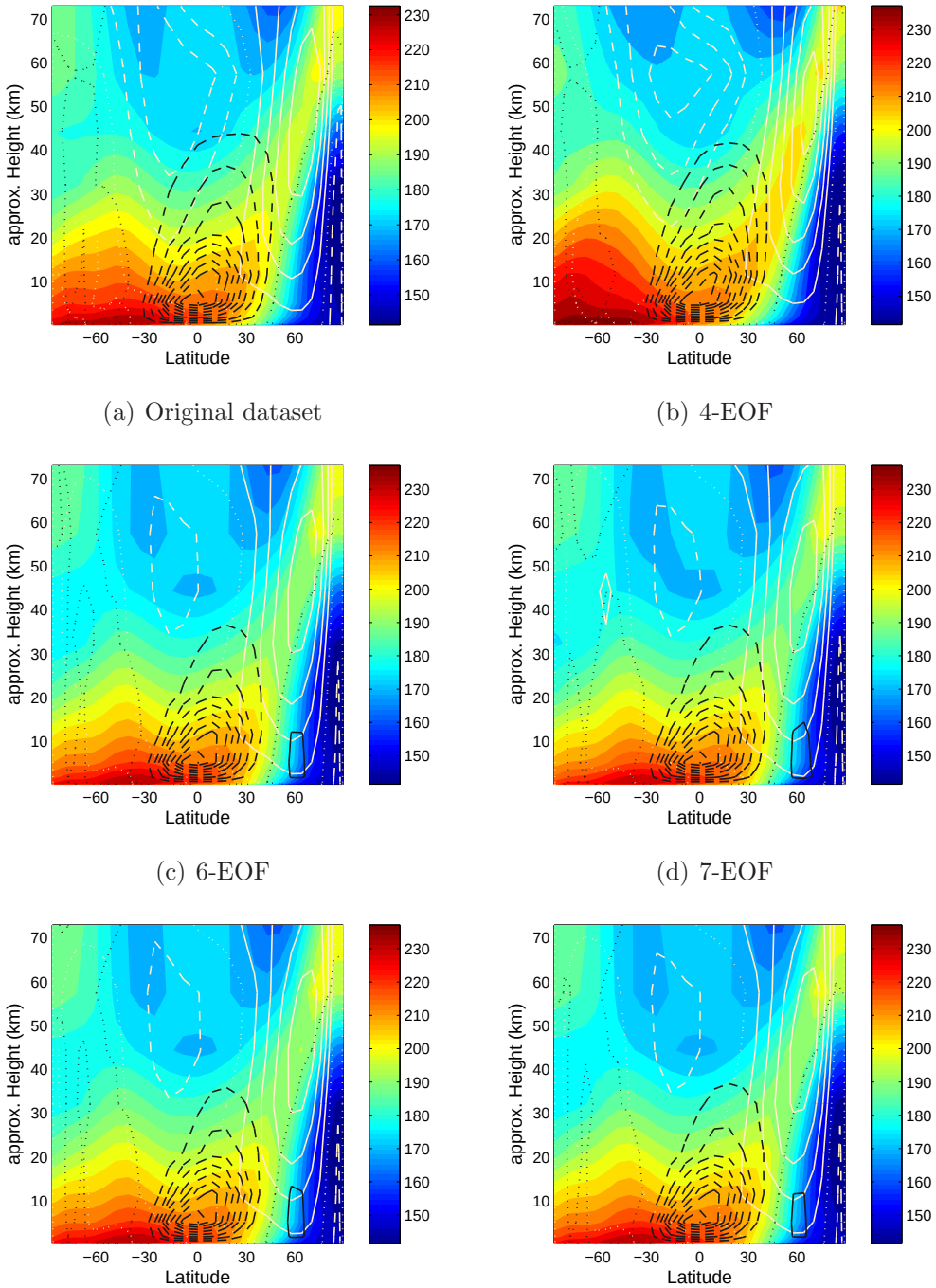
the westerly winter jet shows about the same strength as in the MGCM dataset.

7.3 Evolution of EOF groups

Despite the slight discrepancies between models, the five models described in the previous section seem to behave very well in terms of zonal and time averages. However, when looking at the individual evolution of EOF coefficients, more fundamental differences are to be found. Figures 7.5–7.8 show the evolution of various EOF groups as they appear in the MGCM dataset and as predicted by different reduced models for integrations of 90 sols (the length of the MGCM dataset).

The first group to be analysed in this manner is the pair EOF3–4, dominated by a period of $\mathcal{P} = 1$ sol. This pair represents the diurnal tide and a certain proportion of the diurnal Kelvin and wavenumber-3 waves (see Section 6.4). In the original dataset, EOF3–4 displays an almost constant amplitude which slowly decays as a result of the seasonal cycle (figure 7.5(a)). Despite including the two components of this group, namely EOF3 and EOF4, a 4-EOF model is definitely bad at reproducing its behaviour (figure 7.5(b)). The amplitude predicted by this model rapidly decays to a nearly constant value which represents around 7.5% of the observed value.

The 10-, 12- and 13-EOF models (not shown) start oscillating around zero but later on drift away. This is the reason of the large angular deviations of these models, shown in figure 7.2. The 15-EOF model shows a distinctive behaviour as it grows steadily so that its amplitude becomes twice as large as the amplitude of the original



(a) Original dataset

(b) 4-EOF

(c) 6-EOF

(d) 7-EOF

(e) 11-EOF

(f) 14-EOF

Figure 7.4: Time-averaged zonal mean temperature (colour), zonal wind (white) and mean meridional transport streamfunction (black). Solid lines represent positive values whereas dashed lines represent negative values. Dotted lines represent null contours. The gaps between contours are $\Delta u = 27$ m/s and $\Delta\chi = 129$ kg/s^3.

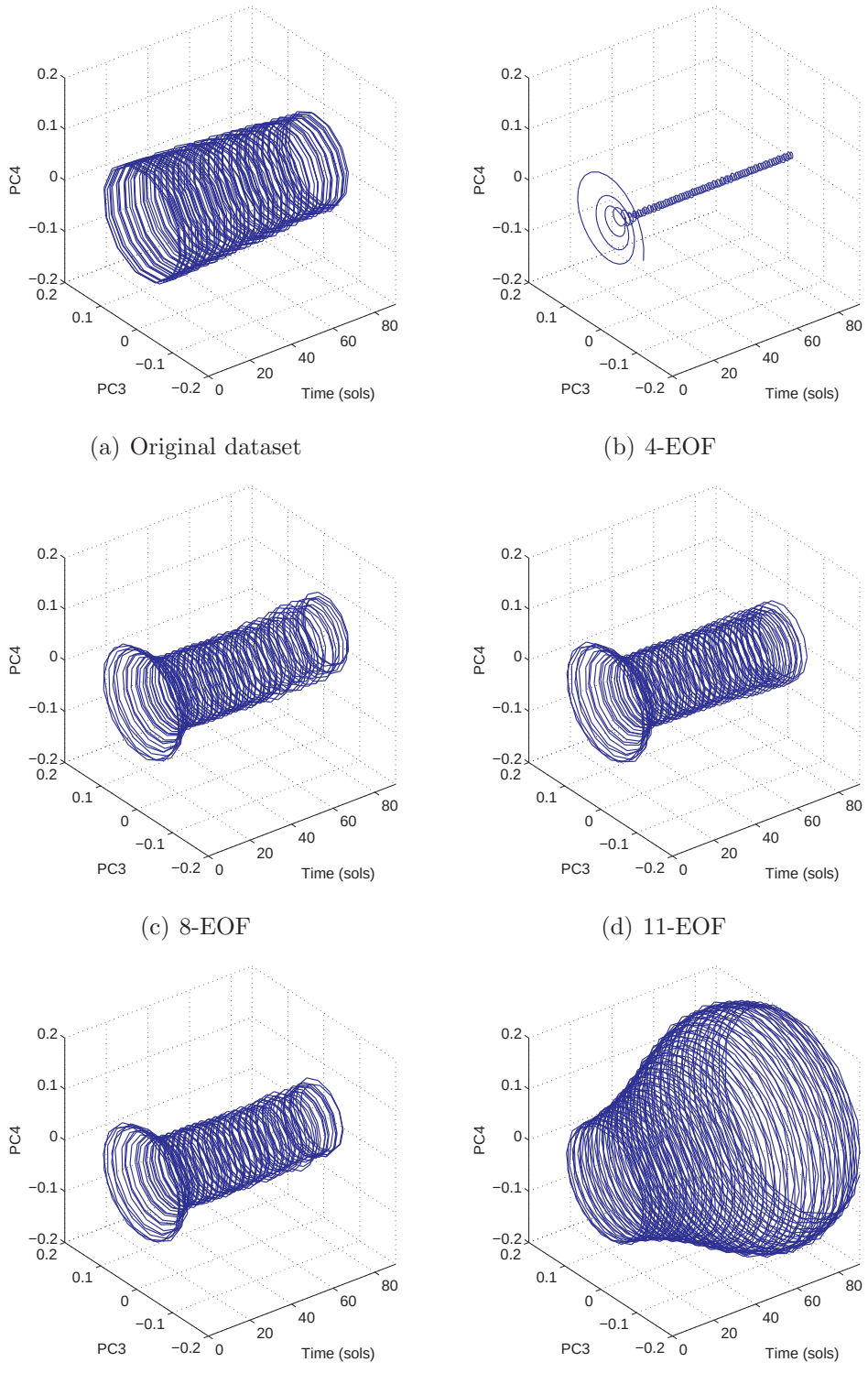
time series in 70 sols. Afterwards, the amplitude decays so that at the end of the test interval (90 sols), it is 50% larger than the original (figure 7.5(f)). The rest of the models display a rapid decay for about 15 sols that leads to a nearly constant amplitude 25% smaller than the original (see, for example, figure 7.5(c)-(e)), which still can be considered as a good approximation.

The pair EOF5–6 is the most energetic representative of transient waves and explains 14.9% of TE in the semidiurnal tide (cf. Section 6.4). This pair has a dominant period of $\mathcal{P}_1 = 3.6$ and a secondary period of $\mathcal{P}_2 = 3.2$ which results in a beat-like behaviour, as can be seen in figure 7.6(a). This behaviour is not fully reproduced by any of the studied models. However, the power spectrum closely resemble the original for models with $N > 5$ (see, Section 7.3.1 below). Thus, the dissimilarities lie on the predicted amplitudes, which in most reduced models tend to be larger than expected (see, for example, figure 7.6(c)-(e)). The 7- and 15-EOF models are exceptions in this respect since both show amplitudes in the same order of magnitude as the original. In contrast, the 12- and 13-EOF models (not shown) predict amplitudes 300 and 1000% larger than the MGCM dataset.

A similar case is found in the pair EOF7–8. It also represents transient waves and, together with EOF3–4, part of the diurnal Kelvin wave (cf. Section 6.4). This pair, with principal and secondary periods of $\mathcal{P}_1 = 4.5$ and $\mathcal{P}_2 = 4.1$, respectively, also shows a beat-like behaviour (figure 7.7(a)). As in the EOF5–6 case, most reduced models overestimate the amplitude of EOF7–8. In fact, only the 11- and 15-EOF models predict the correct amplitude and principal period in average (figure 7.7). However, even these two models fail at predicting beating.

Finally, simulations of the pair EOF9–10, representative of the semidiurnal tide are shown in figure 7.8. In this case, every analysed model exhibits a rapid amplitude shrinking so that in less than 5 sols the amplitude is 30% the size of the amplitude in the MGCM. The only exception is the 15-EOF model which, after a rapid decay in the first five sols, grows as to reach a size in the order of the original (figure 7.8(d)).

A possible way to improve the behaviour of the MGCM reduced models is the use of a time-dependent parameterisation of unresolved modes [3]. This idea was tried out by using periodic parameterisation schemes with period of 1 sol. However, the smallness of the MGCM dataset did not allow a smooth scheme. As a result, the irregularity introduced in the system was so large that it prevented the attainment of better results with respect to those just presented. Most of the models produced unbounded solutions. Those that were successfully integrated showed enhanced irregularity which, in comparison to the version with time-independent parameterisation,



(a) Original dataset

(b) 4-EOF

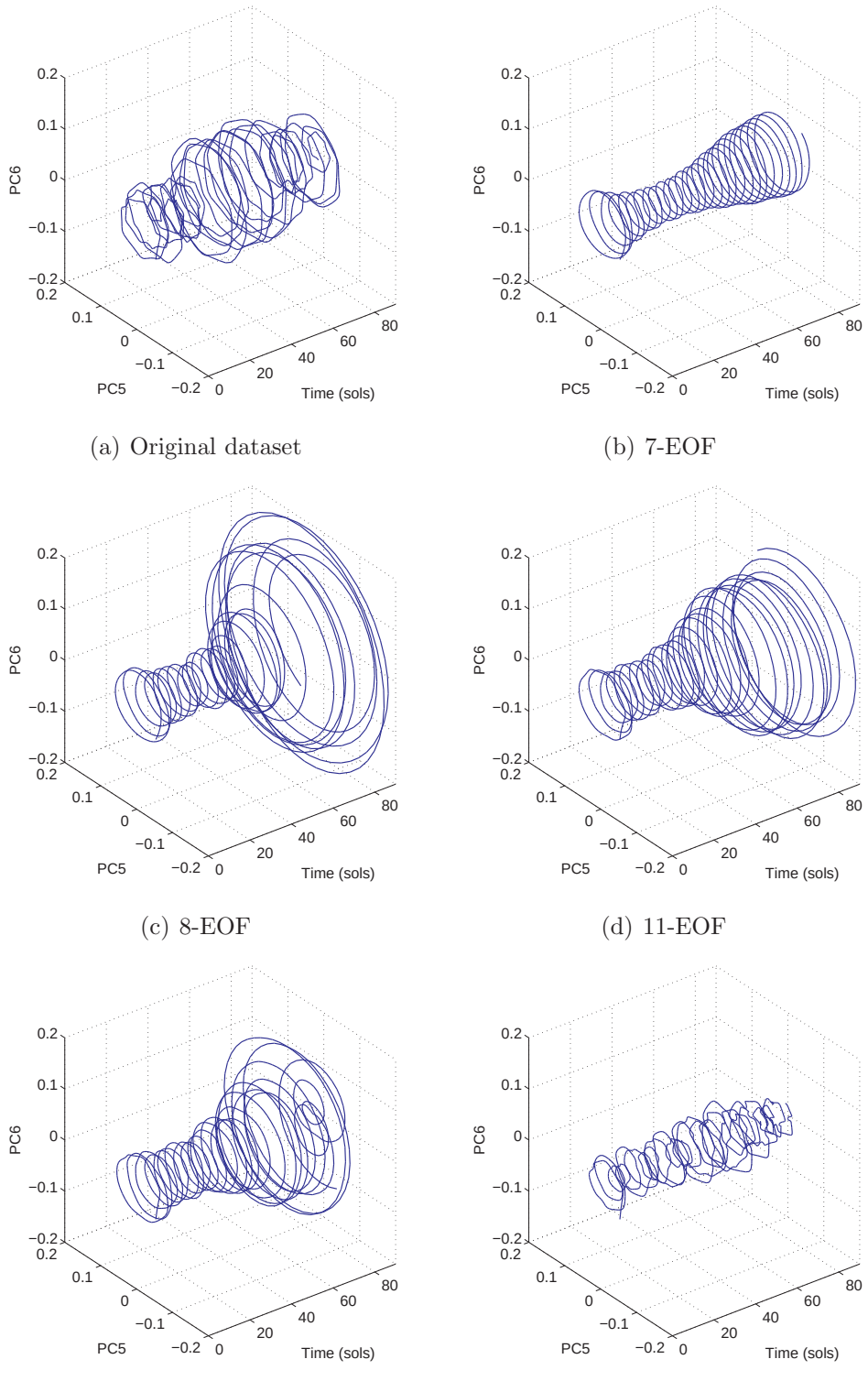
(c) 8-EOF

(d) 11-EOF

(e) 14-EOF

(f) 15-EOF

Figure 7.5: Evolution of the pair EOF3–4 as viewed from the (a) original MGCM dataset, and (b) 4-EOF, (c) 8-EOF, (d) 11-EOF, (e) 14-EOF and (f) 15-EOF models.



(a) Original dataset

(b) 7-EOF

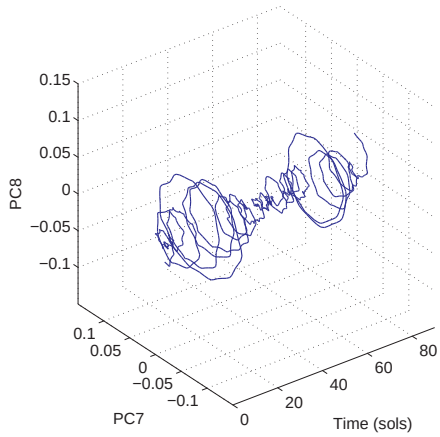
(c) 8-EOF

(d) 11-EOF

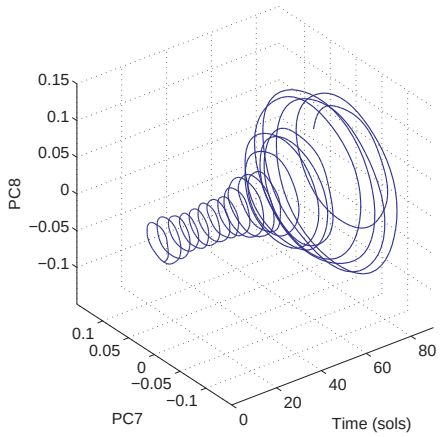
(e) 14-EOF

(f) 15-EOF

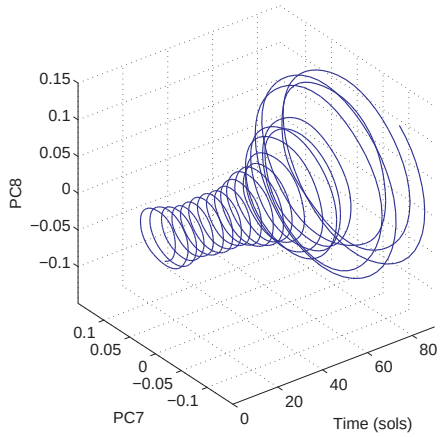
Figure 7.6: Evolution of the pair EOF5–6 as viewed from the (a) original MGCM dataset, and (b) 7-EOF, (c) 8-EOF, (d) 11-EOF, (e) 14-EOF and (f) 15-EOF models.



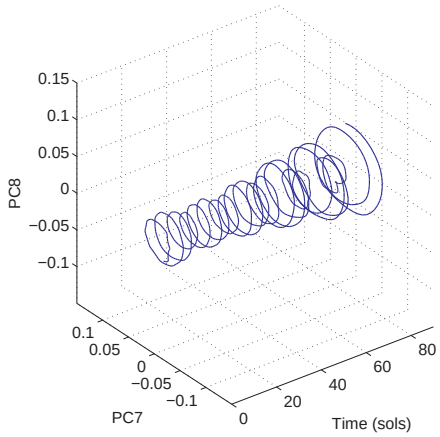
(a) Original dataset



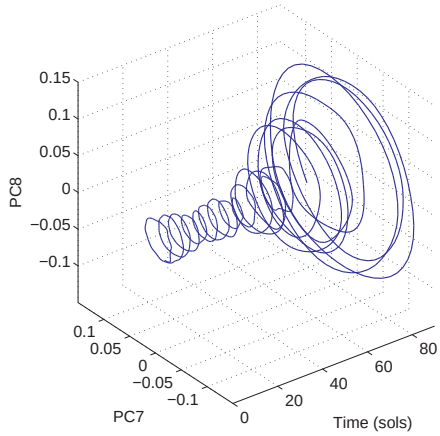
(b) 8-EOF



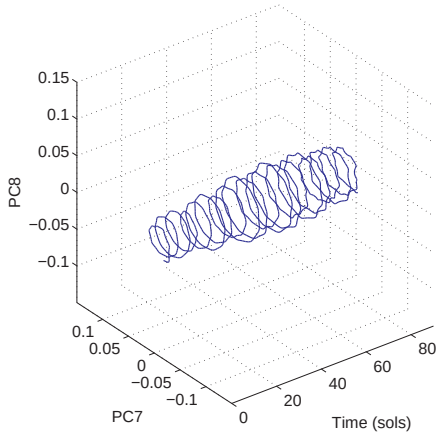
(c) 9-EOF



(d) 11-EOF

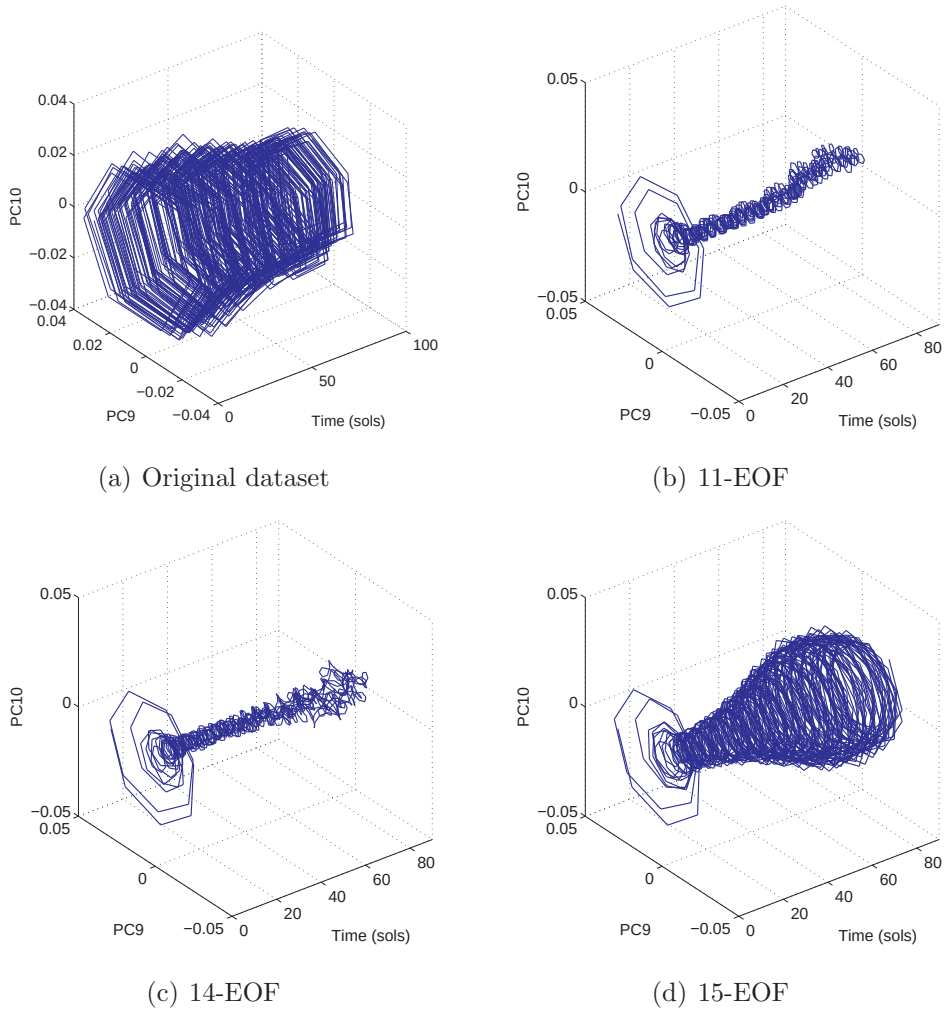


(e) 14-EOF



(f) 15-EOF

Figure 7.7: Evolution of the pair EOF7–8 as viewed from the (a) original MGCM dataset, and (b) 8-EOF, (c) 9-EOF, (d) 11-EOF, (e) 14-EOF and (f) 15-EOF models.



(a) Original dataset

(b) 11-EOF

(c) 14-EOF

(d) 15-EOF

Figure 7.8: Evolution of the pair EOF9–10 as viewed from the (a) original MGCM dataset, and (b) 11-EOF, (c) 14-EOF and (d) 15-EOF models.

could not be considered as an improvement.

7.3.1 Fourier analysis

Figure 7.9(a) shows the Fourier transform amplitude of the principal component corresponding to EOF3 as it appears in the original dataset and after simulations using 3-, 4- and 5-EOF reduced models. As we have seen in Section 6.4, EOF3 and 4 constitute the most representative pair for the diurnal tide. A 3-EOF model, which considers EOF4 as an unresolved mode, produces a bad approximation to the Fourier spectrum of the original time series, overestimating amplitudes at every frequency. More importantly, this model fails to predict a peak at 1 cycle/sol, characteristic frequency of the diurnal tide.

A 4-EOF model displays a peak at that frequency (1 cycle/sol). However, it also fails at predicting the first peak at 0.28 cycles/sol (corresponding to a period of 3.57 sol). Instead, the coefficient predicted by the reduced model exhibits a peak at a slightly higher frequency of 0.44 cycles/sol (corresponding to a period of 2.27 sol). The same behaviour is found in PC4 when four EOFs are included as shown in figure 7.9(b).

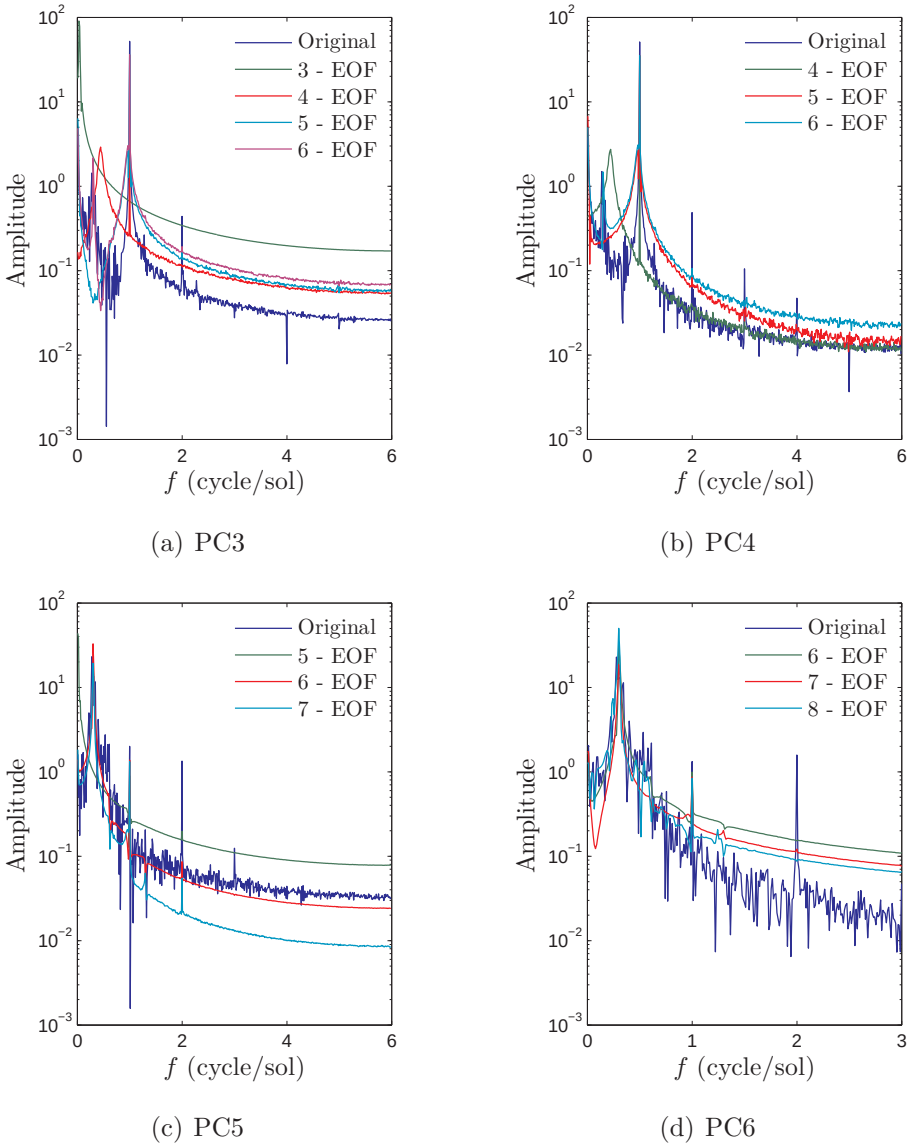
A 5-EOF model produces a better approximation to the Fourier spectrum of these to coefficients (PC3 and PC4). However, it fails at producing the correct spectrum for PC5 (figure 7.9(c)). Finally, the inclusion of six EOFs predicts the correct spectrum for the three PCs under examination (PC3, PC4 and PC5), producing a good approximation to PC6 Fourier spectrum as well (figure 7.9(d)).

A similar explanation can be given for the case depicted in figure 7.2(b) where, after a very high angular deviation between the original time series and that predicted by a 13-EOF model, there is a minimum corresponding to a truncation order of 14. The frames in figure 7.10 show a comparison between the original Fourier spectra of PC11-14 and those obtained using reduced-order models at different truncation orders. These frames make clear that, although the inclusion of 11 and 12 EOF yields a good approximation to the original spectrum, a 13-EOF model fails at predicting a realistic spectrum overestimating amplitudes at every frequency. Including fourteen EOFs produces a good approximation again inducing the minimum in angular deviation observed in figure 7.2(b).

In summary, both the evolution of the modes in phase space and Fourier analysis show that the modelling of a given mode (an EOF group) using a reduced-order model depends on the inclusion not only of all the components in that particular group (for example, EOF3 and 4, in the case of the diurnal tide) but also of all the modes that have a significant influence on the group of interest. Moreover, it was shown that the increase in truncation order does not necessarily imply a better approximation to the original system's behaviour and care must be taken when choosing the number of modes in the construction of a reduced-model.

7.4 Tidal perturbations and chaos

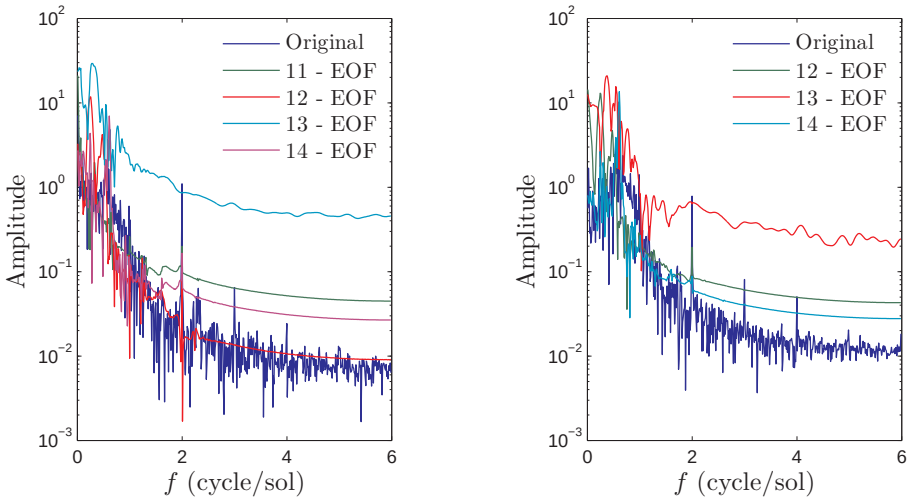
The title of this section was taken from [91], where we find the description of two related numerical experiments carried by means of the Oxford MGCM. These experiments were briefly described in Section 1.1.1. They were related to the meteorological information retrieved by the two Viking Landers (VL) sent on mission to Mars in 1975,



(a) PC3 (b) PC4

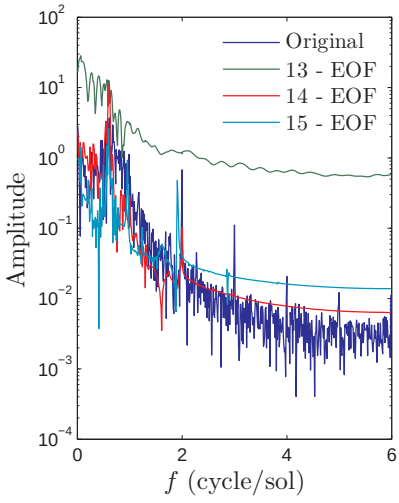
(c) PC5 (d) PC6

Figure 7.9: Fourier transform amplitude of the principal components (a) PC3, (b) PC4, (c) PC5 and (d) PC6 in the original MGCM dataset in comparison with simulations using different reduced-models.

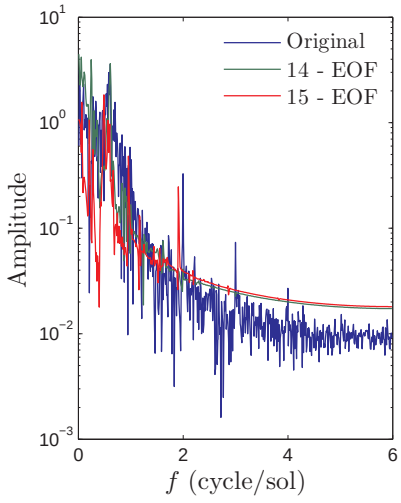


(a) PC11

(b) PC12



(c) PC13



(d) PC14

Figure 7.10: Fourier spectrum of the principal components (a) PC11, (b) PC12, (c) PC13 (d) PC14 in the original MGCM dataset in comparison with simulations using different reduced-models.

landing on Mars in 1976. The measurements of surface pressure showed for the first time the annual oscillation of this variable mainly due to the mass exchange between the polar caps and the atmosphere during the sublimation and condensation seasons. These measurements also led to the first real evidence of baroclinic activity on Mars [91]. The landing sites of these two spacecrafts were located at 22.5°N for VL1 and 48°N for VL2. This information was useful to infer that the detected wave activity took place at mid-latitudes during winter in the northern hemisphere [14].

The numerical experiments using the MGCM consisted of two 9-Martian year simulations [27]. Both simulations were carried out under identical conditions, the only difference being the presence of the diurnal cycle. Thus, in the first experiment no diurnal cycle was included whereas, in the second, this cycle was put back into the model [27].

Both simulations were able to reproduce broadly the main features of the surface pressure cycle (see figures 5 and 8 in [27]). Thus, the differences between these simulations laid in the sort of waves predicted by each model. In particular, the model with diurnal cycle showed the predominance of the diurnal and semidiurnal thermal tides during summer and early autumn in the northern hemisphere. The model with no diurnal cycle did not show oscillations during those seasons. This result was expected since the diurnal cycle was not present to force the activity of thermal tides.

The most striking findings appeared from late autumn to early spring. During this interval both models produced baroclinic waves in accordance with what had been observed in Mars [14, 15] and in previous MGCM simulations [16]. However, the waves displayed very distinctive features in each simulation. The simulations with no diurnal cycle produced two kinds of wave, a high-frequency wave with an approximate period of 2.2 sols and a low-frequency wave with an approximate period of 5.5 sols. High-frequency waves were associated with zonal wavenumber 2 whereas low-frequency waves were related to zonal wavenumber 1. There were two out of eight Martian years in the simulation that exhibited high-frequency waves during the whole winter season. The rest displayed high-frequency waves at the beginning, switching later into low-frequency motion.

A different picture was obtained from the simulation with diurnal cycle. In this case, the output from the simulation exhibited a more irregular signal with similar characteristics to the measurements from VL2. This signal was apparently still dominated by two periodic states with wave numbers 1 and 2 characterised by low and high frequencies, respectively. However, in this case the system did not remain in

either of these states. Instead, it seemed to switch randomly between them. Since the only difference between these two simulations was the presence of the diurnal cycle, these results suggest that this cycle acts like a perturbation making the system continually switch between two coexisting oscillatory equilibria.

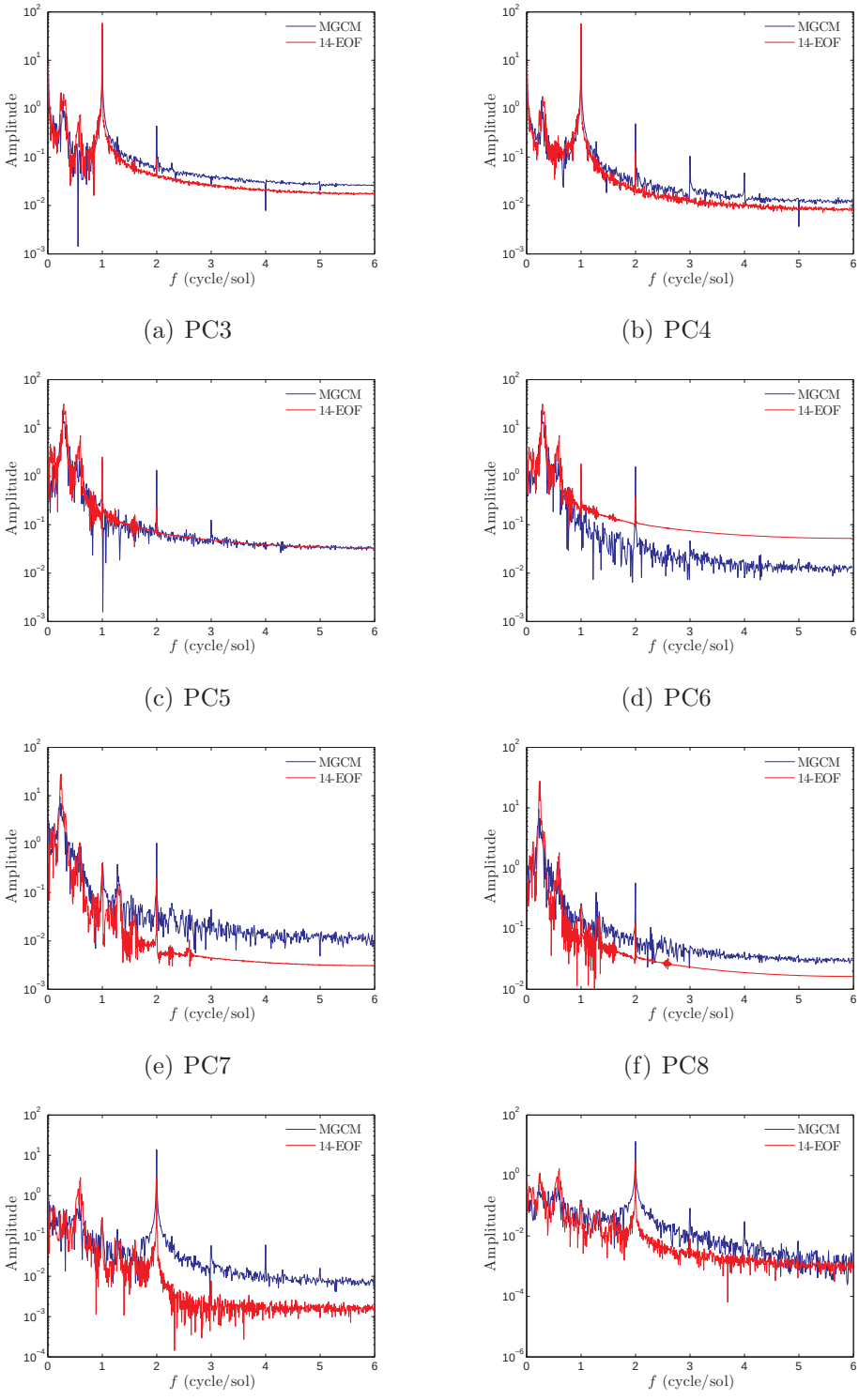
The second objective of this thesis was to investigate the possibility of dynamical interactions between the EOFs so that the large-scale atmospheric motion is reproduced up to, at least, certain level of approximation. One consequence of this statement is the possibility of analysing the interaction between atmospheric modes by means of reduced-order models. The MGCM numerical experiments just described provide an interesting framework to test this possibility.

7.4.1 14-EOF model

We have chosen the 14-EOF model because it includes the atmospheric motions of interest. In particular, it includes the diurnal (EOF3–4) and semidiurnal (EOF9–10 and EOF5–6) tides, the diurnal Kelvin wave (EOF3–4 and EOF 7–8), the diurnal wavenumber-3 wave (EOF3–4). Moreover, baroclinic modes with the appropriate zonal wavenumber are also present in this model. This waves are represented by the pair EOF5–6, which produces a zonal wavenumber-2 northern wave, and EOF7–8, which yields a zonal wavenumber-1 wave.

Another important characteristic of this model is the fact that it shows good approximation with respect to the original MGCM dataset regarding angular deviation and relative norms (cf. Section 7.2.1). When looking at the detailed behaviour of individual PCs, the evolution produced by this model is not worse than other models. The evolution of PC2–3, strongly dominated by the diurnal tide, is fairly well reproduced. Although the rest of the PCs display inaccurate amplitudes (see figures 7.6-7.8), the power spectra for most PCs are approximately correct. This is shown in figure 7.11, where we can see that the spectra of PC3 to PC10 are well reproduced apart from discrepancies at frequencies higher than 2 cycles/sol. The power in these frequencies is always underestimated although there are two notable exceptions: in PC6 it is overestimated while in PC5 the power spectrum is very accurately reproduced. The power at high frequencies in higher order modes is overestimated by this model as shown in figure 7.10.

The spectra of PC9 and P10 show higher power at frequencies between $f = 1.5$ and $f = 2$ cycles per sol. This is similar to what was observed in PC3 and PC4 in models based in less than five EOFs (see Section 7.3.1). Hence, it can be caused



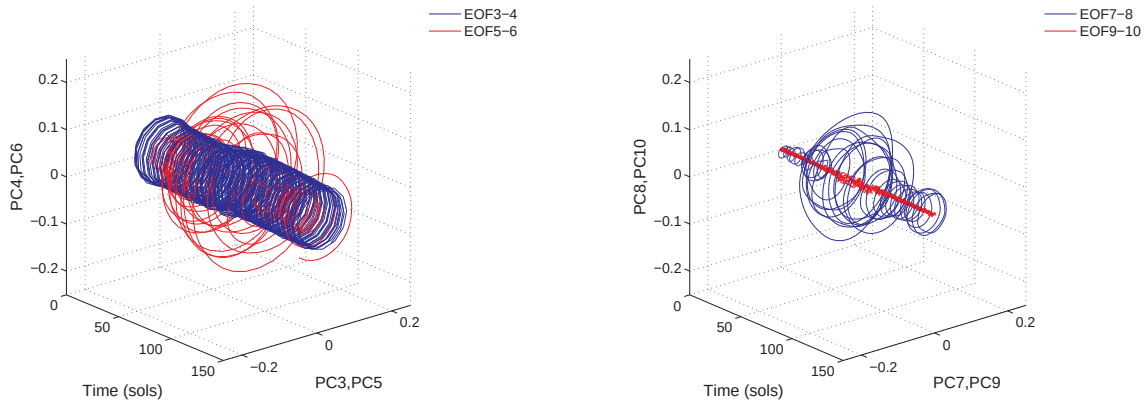
(a) PC3 (b) PC4

(c) PC5 (d) PC6

(e) PC7 (f) PC8

(g) PC9 (h) PC10

Figure 7.11: Fourier spectrum of the principal components (a) PC3, (b) PC4, (c) PC5 (d) PC6, (e) PC7, (f) PC8, (g) PC9 and (d) PC10 in the original MGCM dataset (in blue) in comparison with simulations using the 14-EOF models (in red).



(a) EOF3–4 and EOF5–6

(b) EOF7–8 and EOF9–10

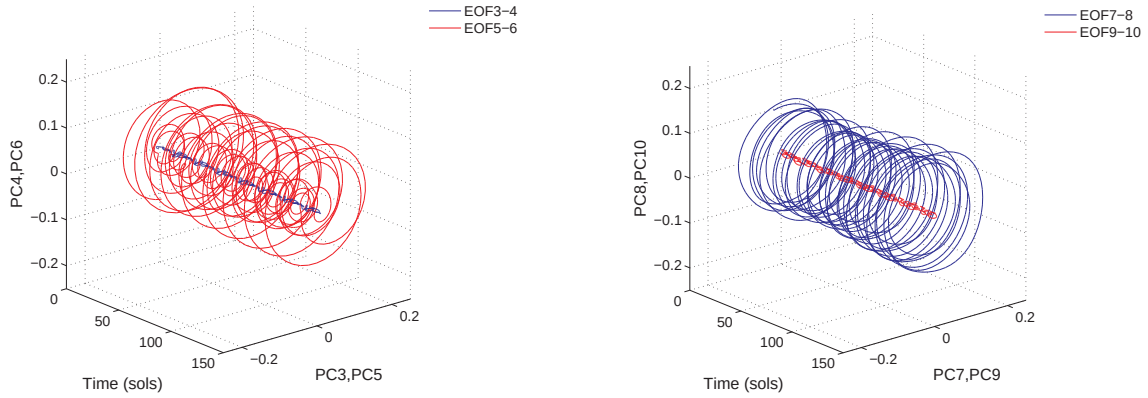
Figure 7.12: Conjugate evolution of the pairs (a) EOF3–4 (blue) and EOF5–6 (red), and (b) EOF7–8 (blue) and EOF9–10 (red) in the 14-EOF model with diurnal cycle.

by the lack of interaction with higher order modes that would need to be explicitly resolved.

7.4.2 Simulation with diurnal cycle

To perform the simulation with diurnal cycle, the 14-EOF model was integrated during 160 sols. the initial conditions were taken to be the projection of the twentieth element in the MGCM dataset onto the 14-EOF space. The first ten sols have been discarded to eliminate a short transient period. The radiative forcing in this model consisted of Newtonian cooling relaxed to a periodic relaxation state with period $\mathcal{P} = 1$ sol and friction in the PBL was modelled by means of Rayleigh friction as described in Section 7.1. The relaxation times were the same as those in that section as well.

The results of this simulation are shown in figure 7.12. Figure 7.12(a) shows the evolution of the diurnal mode and the pair EOF5–6. It is evident in this figure the predominance of the diurnal cycle although the baroclinic pair becomes stronger during certain periods. Figure 7.12(b) shows the semidiurnal mode and the pair EOF7–8 in the second. In this case, the semidiurnal mode is weaker than the baroclinic mode EOF7–8 throughout the interval under analysis. At this point we must recall that the amplitude of the semidiurnal mode is underestimated by the 14-EOF model. However, its magnitude was still expected to be lower than that of the baroclinic mode EOF7–8 since the latter is associated with more energetic EOFs.



(a) EOF3–4 and EOF5–6

(b) EOF7–8 and EOF9–10

Figure 7.13: Conjugate evolution of (a) EOF3–4 (blue) and EOF5–6 (red), and (b) EOF7–8 (blue) and EOF9–10 (red) in the 14-EOF model with no diurnal cycle.

7.4.3 Simulation with no diurnal cycle

The relaxation state used in the simulation with no diurnal cycle was derived from the diurnal average of (7.1). Therefore, this state was time-independent. However, the longitude dependence remained in order to include the signal imposed by the topography. The Newtonian cooling scheme was parameterised by the same value for the relaxation time as for the case including diurnal cycle. As in that case, Rayleigh friction was used in order to model frictional forces in the PBL with the same relaxation times. The model was initialised to identical initial conditions as for the case with diurnal cycle. However, it was necessary to discard a much longer interval of 450 sols in order to eliminate transients. This is not surprising if we consider that the projection of any element of the MGCM dataset is not expected to lie on the climate attractor under the new conditions imposed by the lack of the diurnal cycle.

Figure 7.13 shows the results of this simulation. It shows the same information as in figure 7.12 but for the case with no diurnal cycle. The difference between these two figures is evident. As expected, the diurnal mode (EOF3–4) decays becoming a weak signal in comparison with the baroclinic mode EOF5–6, while this pair displays a much more regular behaviour (figure 7.13(a)). A similar situation is found in figure 7.13(b), where the baroclinic mode EOF7–8 has become more regular. The semidiurnal mode displays an amplitude of comparable size to the case with diurnal cycle. However, the ratio of between the amplitudes of EOF7–8 and the semidiurnal mode has increased in average.

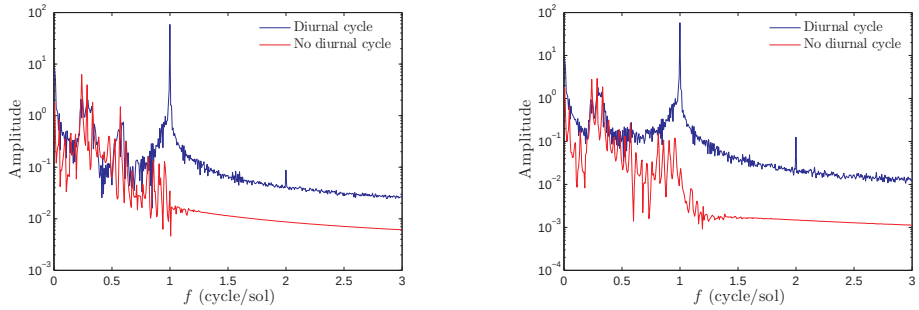
7.4.4 Influence of the diurnal cycle

Figure 7.14 is a comparison between the Fourier spectrum produced by both simulations. The first difference is the absence of strong diurnal signals in the simulation with no diurnal cycle. Again, this result is not surprising since there the diurnal cycle is not exciting these modes. A more interesting result is the quasi-invariance that the spectra show at low frequencies. Indeed, every mode contains the same power at low frequencies in both simulations although the frequency where the spectra diverge varies from mode to mode. For example, the spectra corresponding to PC3 (and PC4) remains close up to a frequency of $f = 0.5$ cycle/sol whereas the spectra corresponding to PC5 (and PC6) remain similar up to $f = 1$ cycle/sol.

At high frequencies, the spectra displayed by the simulation lacking diurnal forcing seem to contain less energy than their counterparts with diurnal cycle. This suggests that this cycle is not only responsible for the excitation of tidal modes but also of every mode with a period shorter than 1 sol. There are two exceptions given by PC7 and PC8, which show higher power at high frequencies.

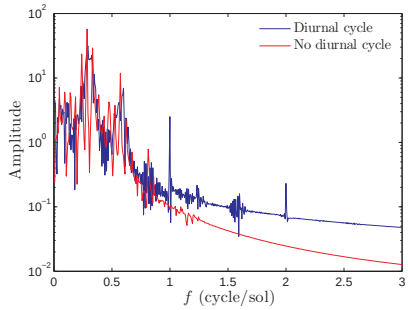
The effect of the lack of tides is the isolated excitation of low frequencies in the diurnal pair EOF3–4 (figures 7.14(a) and (b)). This excitation is due to the presence of baroclinic waves. This is in agreement with the results in Section 7.3.1, where we found that including EOF5–6 in a reduced model was necessary to obtain the correct behaviour of PC3 and PC4 at low frequencies. Here, it is made apparent by the fact that both coefficients oscillate in phase after the transient period has decayed as can be seen in figure 7.15(a). We can also see in figure 7.15(a) the remarkable degree of periodicity that these two modes have reached. We stress that this does not mean that the diurnal tide is being excited by the presence of baroclinic waves. Instead, we must recall that each EOF is a mixture of waves even though some of them are identifiable to a large extent with classical atmospheric motions such as thermal tides or transient waves.

The semidiurnal pair is probably being excited by the baroclinic modes as well. However, in this case it is not the primary but the secondary frequency in EOF3-4 (or EOF7-8) which is displayed by this mode (figures 7.14(g) and (h)). This gives rise to a high frequency oscillation with a period of $\mathcal{P} = 1.7$ sols approximately. As in the previous case, both the baroclinic pair EOF7–8 and the semidiurnal mode display a highly periodic behaviour as can be seen in figure 7.15(b). An apparent difference between EOF5–6 and EOF7–8 is that the former exhibits amplitude vacillation whereas the latter shows a nearly constant amplitude though still modulated by a signal of longer period.

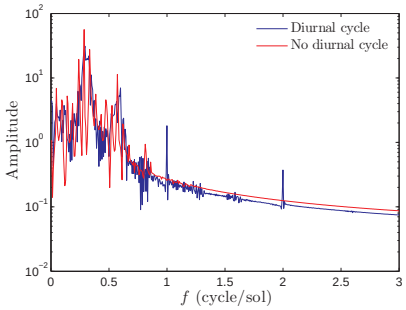


(a) PC3

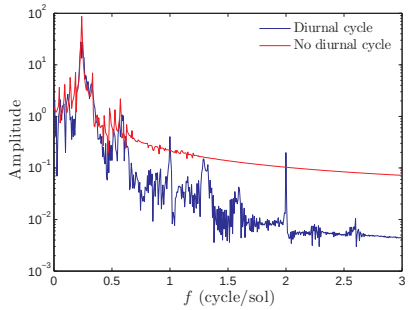
(b) PC4



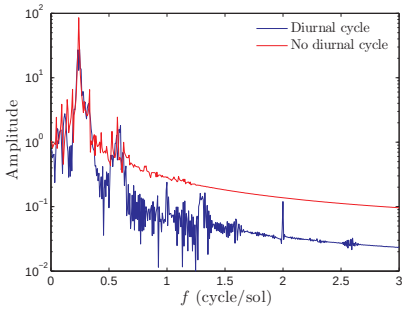
(c) PC5



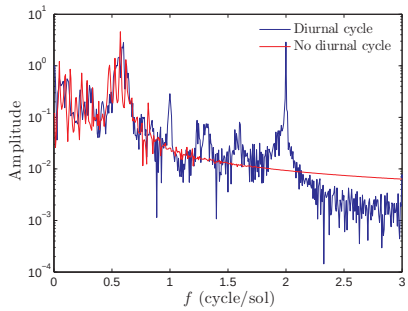
(d) PC6



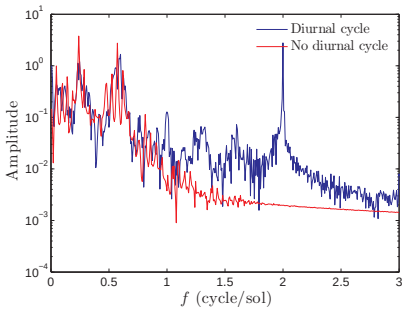
(e) PC7



(f) PC8

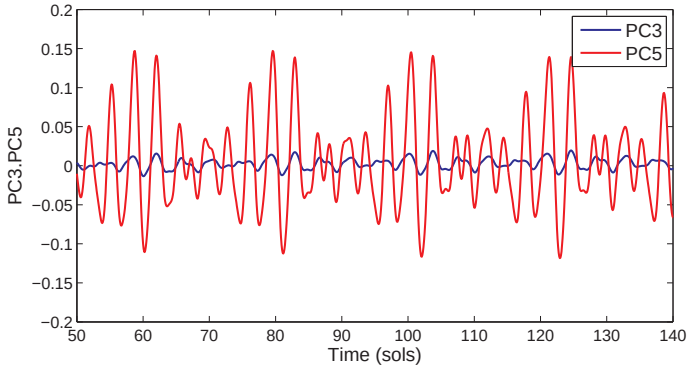


(g) PC9

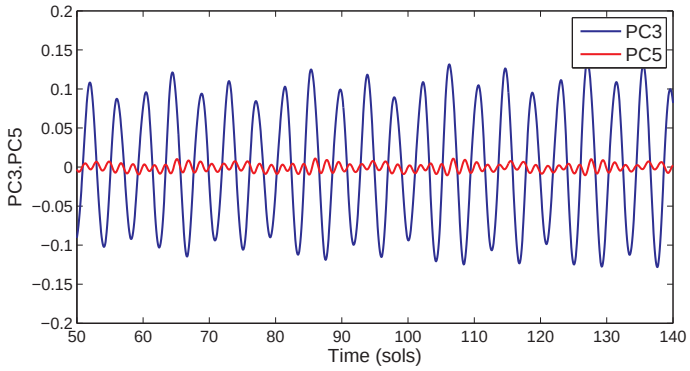


(h) PC10

Figure 7.14: Fourier spectrum of the principal components, indicated in each frame, in the simulations including (in blue) and not including (in red) diurnal cycle in the 14-EOF model.



(a) PC3 and PC5

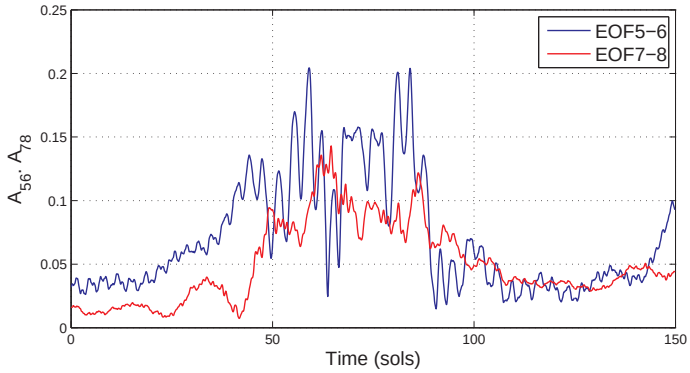


(b) PC7 and PC9

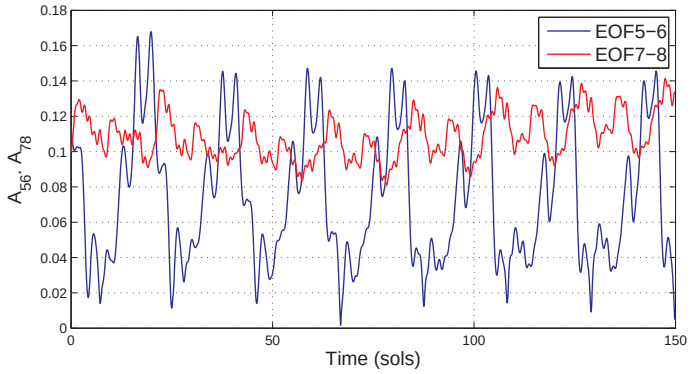
Figure 7.15: Time series corresponding to (a) PC3 (blue) and PC5 (red), and (b) PC7 (blue) and PC9 (red), as simulated by the 14-EOF model with no diurnal cycle.

At this point, it is appropriate to define quantitatively the amplitude of a pair of EOFs by the formula $A_{kl}(t) = (a_k^2(t) + a_l^2(t))^{1/2}$, where a_k is the k-th PC. Figure 7.16 shows the amplitudes of EOF5–6 and EOF7–8 as predicted by both simulations. The time series from the simulation with diurnal cycle appears to be more irregular while the amplitudes of both pairs are of similar magnitude throughout the period under analysis (figure 7.16(a)). In contrast, the simulation with no diurnal cycle yields highly regular, nearly periodic amplitudes for both EOF pairs in concordance with the results in figure 7.15. Furthermore, the predominance is shared by both pairs so that the system flips between zonal wavenumber-1 and -2 baroclinic waves. The times of predominance are different for each baroclinic pair so that EOF5–6 is predominant during 25% of the analysed interval whereas the rest is dominated by EOF7–8.

This result seems to be in agreement with the findings by Collins *et al.* [27], where six out of eight years exhibited transitions from wavenumber 2 into wavenumber 1. Moreover, the experiments with the 14-EOF model seem to support the hypothesis



(a) With diurnal cycle



(b) With no diurnal cycle

Figure 7.16: Amplitudes of EOF5-6 and EOF7–8 as simulated by the 14-EOF model (a) with diurnal cycle, and (b) with no diurnal cycle.

that the influence of the diurnal cycle over the Martian atmospheric general circulation is to act as a destabilizing agent over transient waves so that they are unable to settle into one single mode, switching instead between two different regimes.

These are certainly preliminary results that leave open a number of questions. For example, the reasons for having certain years with a predominant wavenumber-2 wave cannot be explained in the light of these results as described in [27]. Another unanswered questions would be why the system tends to start oscillating with a high frequency wavenumber-2 wave to then switch into the low frequency mode. These questions would require further investigation. In this respect, the advantage of having an available reduced model is the separation of tidal modes from baroclinic waves. Despite this separation not being perfect, it allows a clearer study of the interaction between these two kinds of motion. On the other hand, this sort of separation cannot be achieved with a full MGCM since tides and transient waves are intimately intermingled in this case. In other words, it is not possible to suppress

the presence of these waves at will. An EOF-based reduced model can help in this purpose.

7.5 Concluding remarks

We finish this section with some concluding remarks not only to the section itself but also to the discussion started at the end of Section 4.2 and continued at Section 5.3. Regarding the models' performance as the number of included EOFs increases, we have found the same non-monotonic behaviour in the MGCM reduced-models as we found in the SGCM models. Moreover, these results are in general agreement with the findings of Crommelin and Majda when constructing low-order models for the Charney-DeVore model [29]. The conclusion here is certainly the same: the performance of a reduced-dimension model depends not only on the number of modes included; it also depends on the completeness of the groups that represent the waves of interest. Furthermore, it is important to retain certain higher order modes whose interaction with the modes of interest might be significant.

As an example, let us consider the diurnal tide represented by EOF3–4 (figure 7.9). The correct representation of this tide requires, as expected, having EOF3 and EOF4 among the resolved modes. However, it is necessary to include EOF5–6 as well in order to obtain a better approximation to the Fourier spectra observed in the MGCM dataset. Recalling that EOF5–6 primarily represents a transient baroclinic wave and part of the semidiurnal tide, we can interpret this result as a strong coupling between the diurnal tide, a transient baroclinic wave of period around 3.2 and 3.6, and, possibly, the semidiurnal tide. This conclusion is even more evident from the experiments described in Section 7.4 where the diurnal cycle was switched on and off.

As a final comment we would like to point out that there are always discrepancies at high frequencies, which are difficult to model at any truncation (figures 7.9 and 7.10). An improved model would require the implementation of numerical dissipation to remove or add energy from these frequencies as required. The suggestion would be first to include the appropriate number of modes so that the approximation to the original Fourier spectrum is sufficiently accurate (based on a criterion set by the modeller) and secondly to parameterise numerical dissipation over high frequencies in order to achieve better resemblance between full and reduced models.

7.6 Summary

In this chapter we have applied the POD and Galerkin projection to construct low-order models of the Oxford MGCM. The EOF classification carried out in Chapter 6 allowed a neat interpretation of the results after projecting the equation of motion onto the the most energetic EOFs' subspace and constructing low-order models. In particular, by analysing the evolution of various EOF groups (representing tides and transient waves) and the Fourier spectra of the amplitudes predicted by reduced order models, we have been able to identify, for example, significant interactions between the diurnal tide and transient baroclinic waves and possibly the semidiurnal tide.

This was achieved by means of a direct comparison between the output of reduced models and the MGCM dataset, and by carrying out numerical experiments aimed at investigating the influence of the diurnal cycle over the dynamics of the Martian atmosphere. From these experiments it was possible to draw some conclusion that are in agreement with results from the MGCM. In particular, we showed evidence that supports the conjecture that the diurnal cycle tends to destabilize the baroclinic activity during northern winter, which would otherwise be highly regular, therefore introducing larger variability into the system's dynamics.

Chapters 6 and 7 have covered the central topic of this thesis. In Chapter 8, we shall discuss some theoretical issues that could ultimately lead to further development of low-order models based on the POD. This would consist of the search for separate bases for different fields by means of suitable transformations of the EOFs.

Chapter 8

Investigating new reduced spaces

In the previous four chapters, we have shown realisations of low-order models using three-component three-dimensional EOFs. However, there are certain theoretical difficulties that arise from the finite nature of the original dataset. A potential problem is that of the independent representability of different components. For example, the correct representation of the advective term in the horizontal components of the equations of motion requires the expansion of the horizontal velocity $\mathbf{v}$ alone. In other words, we need to represent a vector of the form $\psi = (\mathbf{v}, 0)^T$. Since $\tau = 0$, the spectral coefficients a_k are simply given by $(\phi^{(k)}, \psi) = \frac{1}{2}\int_V p^s \phi_{\mathbf{v}} \cdot \mathbf{v}\, dV \neq 0$. Substituting back into the reconstruction equation, we obtain

$$\psi_r = \sum_{k=1}^{N} a_k \begin{pmatrix} \phi_{\mathbf{v}}^{(k)} \\ \phi_{\tau}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_r \\ \tau_r \end{pmatrix}, \tag{8.1}$$

where $\mathbf{v}_r, \tau_r \neq 0$, i.e. the vector reconstruction gave origin to a spurious τ component. This effect is due to the incompleteness of Φ and the finite truncation in (8.1). A similar problem was pointed out in reference to the turbulent flow analysed in [12] by Moffat [76] and Holmes [43].

One way of solving this problem arises by looking for alternative separate bases to represent each component. The first alternative basis, here called transformed empirical functions (TEFs) will arise from the unitary transformation from the original EOFs into new sets of orthogonal functions. The most important property of the transformations is to decouple the state vector components. As a result, the new functions have the ability to approximate the prognostic equations independently from each other. In order to gain better insight into the action of the proposed transformations and the new orthogonal functions, they will be applied to a low dimensional

system and, with these results as a guideline, the same transformation will be applied to the dynamical system of interest, namely the Oxford MGCM.

A different alternative basis arises from considering the time series for each component and carrying out a POD analysis for each one, independently. By this procedure, however, the vector character of the system's state and, consequently, the correlation between its different components are lost. A different solution can be obtained by looking for a transformation of Φ that conserves its properties giving independence to its components at the same time.

8.1 Transformed empirical functions

Assuming that the norm induced by (3.1) is interpreted as the system's TE and that the matrix E_{ij} is at least block-wise diagonal, it is possible to split it into two parts and define two partial inner products. The K-product,

$$(\chi, \psi)_K = \int_V \chi_p^*(x)\, E_{pq}\, \psi_q(x)\, dV, \tag{8.2}$$

with $p, q = 1, \ldots, d_1$, related to kinetic energy (KE), induces the KE norm $||\psi||_K = (\psi, \psi)_K^{1/2}$. The P-product,

$$(\chi, \psi)_P = \int_V \chi_r^*(x)\, E_{rs}\, \psi_s(x)\, dV, \tag{8.3}$$

with $r, s = d_1 + 1, \ldots, d_1 + d_2$, related to potential energy (PE), induces the PE norm $||\psi||_P = (\psi, \psi)_P^{1/2}$.

This procedure also separates the components of a vector into two subsets so that the first d_1 components of ψ, denoted by $\mathbf{v}$, are associated with KE while the rest, denoted by τ, is associated with PE. Thus, a vector will have the general form $\psi = (\mathbf{v}, \tau)^T$.

Let us define the matrices $E^{(ij)} \equiv (\phi^{(i)}, \phi^{(j)})$, $K^{(ij)} \equiv (\phi^{(i)}, \phi^{(j)})_K$ and $P^{(ij)} \equiv (\phi^{(i)}, \phi^{(j)})_P$ so that $E^{(ij)} = K^{(ij)} + P^{(ij)}$. Since $\{\phi^{(k)}\}_{k=1}^{\infty}$ is orthonormal under (3.1), we have that $E^{(ij)} = \delta^{(ij)}$, the Kronecker delta. Therefore, although in general $K^{(ij)}$ and $P^{(ij)}$ are not diagonal, we have that

$$K^{(ij)} = \delta^{(ij)} - P^{(ij)}. \tag{8.4}$$

Let $U^{(kl)}$ be a unitary matrix so that $\mathscr{K}^{(ij)} = U^{(ki)*} K^{(kl)} U^{(lj)}$ is diagonal. Then, it can

be proved by direct application of (8.4) that $\mathscr{P}^{(ij)} = U^{(ki)*} P^{(kl)} U^{(lj)}$ is also diagonal.

8.1.1 Definition and properties

Let us define the set of transformed empirical functions (TEF) as $\Theta = \{\theta^{(l)} : \theta^{(l)} = \phi^{(k)} U^{(kl)}\}$. This definition enables us to write

$$(\theta^{(i)}, \theta^{(j)})_K = \mathscr{K}^{(ij)}, \tag{8.5}$$

which shows that the subset $\{\theta_1^{(l)}\}$ associated with KE is orthogonal under the KE norm. Similarly,

$$(\theta^{(i)}, \theta^{(j)})_P = \mathscr{P}^{(ij)}, \tag{8.6}$$

so that the subset $\{\theta_2^{(l)}\}$ associated with PE is orthogonal under the PE-norm. Furthermore, Θ is orthogonal under the TE norm, inheriting this property from the original set. These properties enable the elements of Θ to represent independently the components of a vector so that its components are no longer proportional to each other. This is important, for example, when the dynamical equations contain terms that act exclusively on one subspace but not on the other. It must be mentioned that this transformation was independently suggested by Waleffe [121] and used by Smith *et al.* [115] in turbulence studies.

Given these three different orthogonality properties, there are two ways of decomposing a vector. The first makes use of orthogonality under the TE norm,

$$\psi^N(x,t) = \sum_{k=1}^{N} b_k(t)\theta^{(k)}(x) \tag{8.7}$$

where, as customary,

$$b_k(t) = (\theta^{(k)}(x), \psi(x,t)). \tag{8.8}$$

The second decomposition relies on the simultaneous orthogonality of the components θ_1 and θ_2 under the KE and PE norms, respectively. In this case we can write

$$\begin{aligned} \psi^N(x,t) &= \sum_{k=1}^{N} (\mu_k(t)\ \nu_k(t)) \cdot \theta^{(k)}(x) \\ &= \sum_{k=1}^{N} \begin{pmatrix} \mu_k(t)\theta_1^{(k)}(x) \\ \nu_k(t)\theta_2^{(k)}(x) \end{pmatrix}, \end{aligned} \tag{8.9}$$

where

$$\mu_k(t) = (\theta^{(k)}(x), \psi(x,t))_K / K_k, \quad \text{and} \quad \nu_k(t) = (\theta^{(k)}(x), \psi(x,t))_P / P_k, \tag{8.10}$$

and $K_k \equiv ||\theta^{(k)}||_K^2$ while $P_k \equiv ||\theta^{(k)}||_P^2$. Under the first decomposition, span(Θ) is equal to the span of the original set of eigenmodes. The second decomposition enables us to represent vectors in the form $\psi = (\mathbf{v}, 0)^T$ and $\psi = (0, \tau)^T$ as well as vectors where the vector components are not proportional in the way prescribed by the EOFs. The first description is, thus, a special case of the second where $\mu_k = \nu_k$.

Since the elements of Θ are linear combinations of the EOFs, they are no longer eigenfunctions of (3.6). Moreover, the whole structure of the TEFs is more complex than that of the EOFs (see Section 8.1.4). Therefore, it is not possible to characterise them by a definite amount of average TE. However, given a certain number of EOFs, the modified set will be characterised by the same amount of average TE explained by the EOFs. This can be shown by expanding ψ in terms of the empirical eigenvectors,

$$\psi = \sum_{k=1} b_k \theta^{(k)} = \sum_{kl} b_k \phi^{(l)} U^{(lk)} = \sum_{l} a_l \phi^{(l)}. \tag{8.11}$$

Therefore, the coefficients in both expansions are related by

$$a_l = U^{(lk)} b_k. \tag{8.12}$$

Energy in the TEFs

The average energy in an order-N approximation to ψ is given by

$$\langle E_N^{\text{EOF}} \rangle = \sum_{k=1}^{N} \langle a_k(t) a_k^*(t) \rangle = \sum_{k=1}^{N} \langle b_k(t) b_k^*(t) \rangle, \tag{8.13}$$

where the last equality was derived using (8.12) and $U^{(kl)} U^{(km)*} = \delta_{lm}$ (since U^{kl} is unitary). As a result, the average TE in Θ is the same as in Φ and is given by (cf. (3.14))

$$\langle E_N^c \rangle = \sum_{k=1}^{N} \langle b_k(t) b_k^*(t) \rangle = \sum_{k=1}^{N} \lambda_k, \tag{8.14}$$

where λ_k is the eigenvalue associated to the k-th EOF so that $\lambda_k \geq \lambda_{k+1}$. The superscript c indicates that E_N was computed with θ as a basis for ψ (with coupled

components). In the general case, when $\mu_k \neq \nu_k$, the average energy is given by

$$\langle E_N^d \rangle = \sum_{k=1}^{N} \left(K_k \langle \mu_k(t) \mu_k^*(t) \rangle + P_k \langle \nu_k(t) \nu_k^*(t) \rangle \right). \tag{8.15}$$

The superscript d indicates that E_N was computed accordingly to (8.9) (i.e. considering $\mathbf{v}$ and τ as decoupled components).

A question arises at this point: which basis, EOFs or TEFs, explains more TE on average? In order to answer this question we use (8.8) and (8.10) to obtain

$$b_k = K_k \mu_k + P_k \nu_k. \tag{8.16}$$

We see from (8.16) that b_k is indeed an average of μ_k and ν_k weighted by K_k and P_k, respectively. Subtracting the energy contribution from every component at any given time using (8.4) yields,

$$b_k b_k^* - (K_k \mu_k \mu_k^* + P_k \nu_k \nu_k^*) = -P_k K_k |\mu_k - \nu_k|^2 \leq 0. \tag{8.17}$$

From here it follows immediately that

$$\langle E_N^d \rangle \geq \langle E_N^c \rangle, \tag{8.18}$$

which demonstrate that, on average, a decomposition in terms of $\{\theta_2^{(k)}, \theta_2^{(l)}\}$, where $k, l = 1, \ldots, N$, explains more TE in the original ensemble than an EOF decomposition with the same number of modes. Furthermore, the difference between $\langle E_N^d \rangle$ and $\langle E_N^{\mathrm{EOF}} \rangle = \langle E_N^c \rangle$ is given by

$$\langle E_N^d \rangle - \langle E_N^{\mathrm{EOF}} \rangle = \sum_{k=1}^{N} P_k K_k \left\langle |\mu_k - \nu_k|^2 \right\rangle. \tag{8.19}$$

In other words the advantage in using TEFs over EOFs, in terms of average TE, is directly related to the magnitude of the component coefficient difference $|\mu_k - \nu_k|$. Notice that the EOFs are still optimal representatives of the original dataset explaining more TE than any other linear decomposition in the full phase space. The TEFs are able to explain more TE because they split the phase space into two independent subspaces defining a different kind of decomposition.

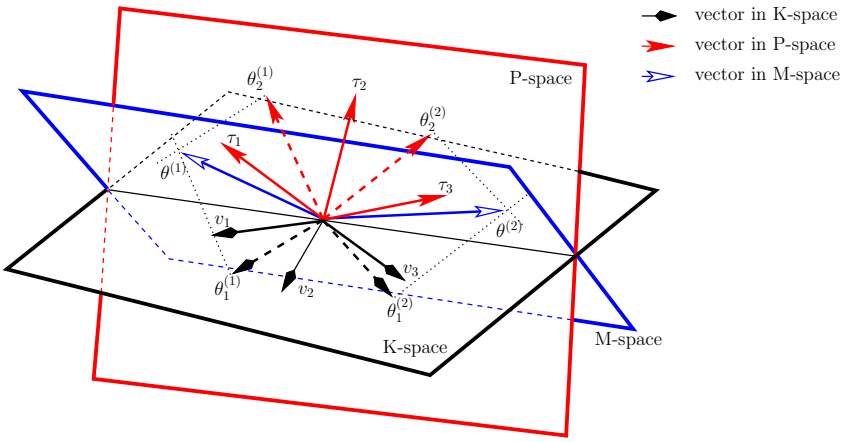


Figure 8.1: The whole phase space and different subspaces. $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ represents the K-space. (τ_1, τ_2, τ_3) represent the P-space. $\{\theta^{(1)}, \theta^{(2)}\}$ spans the M-space. The projections of these vectors onto the K- and P-spaces are shown as dashed arrows.

Interpretation of the matrix $U^{(kl)}$

In order to understand the effect of $U^{(kl)}$ let us represent the phase space, that in principle could be infinite dimensional, by a six-dimensional space (figure 8.1). The K-space is defined by the KE-related components $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. and is therefore measured by the KE norm. The P-space, on the other hand, is defined by the PE-related components (τ_1, τ_2, τ_3) and is measured by the PE norm.

Let us assume, that as a result of the POD, it has been possible to identify the maximum energy subspace, given by the hyperplane spanned by the TEFs $\{\theta^{(1)}, \theta^{(2)}\}$ and labelled as M-space in figure 8.1. The projection of these functions onto the K-space produces the components $\theta_1^{(1)}$ and $\theta_1^{(2)}$. They are mutually orthogonal in the K-space and span a hyperplane there. An important difference between the EOFs and TEFs is the non-orthogonality of the components of the EOFs within their respective spaces, K and P. Another difference, perhaps more fundamental, is that in general every EOF has a projection onto both the K- and P-space regardless of the possibility of having subspaces with different dimensions. This fact may cause ambiguity at defining a basis for each subspace. Each TEF, on the contrary, only has a projection onto the K- and P-spaces if it is non-orthogonal to them. Therefore, the dimension of the reduced hyperplane spanned by the components $\theta_1^{(1)}$ and $\theta_1^{(2)}$ is given by the number of TEFs that are non-orthogonal to the K-space. Similarly, projecting $\theta^{(1)}$ and $\theta^{(2)}$ onto the P-space produces an orthogonal basis spanning a hyperplane in the P-space, whose dimension is the number of TEFs that are non-orthogonal to the P-space.

8.1.2 Further decoupling

In general, this procedure can be carried on iteratively in order to decouple as many components as necessary. This can be easily achieved by normalising the elements in the new bases. For instance, let us suppose that we are required to decouple the components of $\mathbf{v}$. Then, we just need to normalise the set $\{\theta_1^{(l)}\}$ so that the assumptions in Section 8.1 are satisfied. Therefore we can look for a new orthogonal transformation to make the component bases independent.

In principle, this procedure would lead to additional phase space coverage by a reduced basis. Nevertheless, the further decoupling procedure just outlined is not available for the reduction of a primitive-based-equation model under the present assumptions. The reason for this is the time-independence of surface pressure. In order to see this let us assume the existence of independent bases for the velocity components u and v so that they can be written as

$$u = \sum_j \tilde{u}_j \theta_1^j, \quad v = \sum_j \tilde{v}_j \theta_2^j, \tag{8.20}$$

where $\tilde{u}_j$ and $\tilde{v}_j$ are the expansion coefficients and $\{\theta_1^j\}_j$ and $\{\theta_2^j\}_j$ are the assumed bases for u and v, respectively. The index j runs over a suitably defined set of integers. The sigma-velocity equation (5.6) in spherical coordinates is

$$\dot{\sigma} = -\frac{1}{p^s} \int_0^\sigma \frac{d\sigma_1}{\cos\phi} \left\{ \frac{\partial}{\partial\lambda}(p^s u) + \frac{\partial}{\partial\phi}(p^s v \cos\phi) \right\}. \tag{8.21}$$

Substituting (8.20) into (8.21) and evaluating at the surface where $\dot{\sigma} = 0$ yields

$$\sum_j \left\{ \tilde{u}_j \frac{\partial}{\partial\lambda} \left(p^s [\![\theta_1^j]\!] \right) + \tilde{v}_j \frac{\partial}{\partial\phi} \left(p^s [\![\theta_2^j]\!] \cos\phi \right) \right\} = 0. \tag{8.22}$$

Each term in this equation must vanish for the boundary conditions to be satisfied independently of the truncation order. Thus,

$$\tilde{v}_j = -\tilde{u}_j \frac{\frac{\partial}{\partial\lambda} \left(p^s [\![\theta_1^j]\!] \right)}{\frac{\partial}{\partial\phi} \left(p^s [\![\theta_2^j]\!] \cos\phi \right)}, \tag{8.23}$$

which leads to a contradiction since we assumed $\tilde{u}$ and $\tilde{v}$ to be mutually independent.

Moreover, since the EOFs automatically satisfy the boundary conditions we can write

$$\nabla \cdot \left(p^s [\![\theta^j]\!]\right) \cos\phi = \frac{\partial}{\partial \lambda}\left(p^s [\![\theta_1^j]\!]\right) + \frac{\partial}{\partial \phi}\left(p^s [\![\theta_2^j]\!] \cos\phi\right) = 0, \tag{8.24}$$

which, with (8.23), implies that $\tilde{v}_j = \tilde{u}_j$, and the velocity components must be simultaneously represented by a coupled basis.

By using a different set of state variable, this problem can be circumvented. One such alternative is constituted by the barotropic streamfunction, the two baroclinic components of horizontal velocity and the square root of potential temperature [3, 4]. With this choice of variables, the coupled part of horizontal velocity automatically remains coupled while the baroclinic components can be decoupled. The price to be paid for this capacity is the addition of the barotropic streamfunction as part of the state variables.

8.1.3 Finite-dimensional systems

The ultimate objective of the present thesis is the application of dimension-reduction techniques to atmospheric dynamics on Mars governed by infinite dimensional partial differential equations. However, we start the study of practical applications of TEFs by the analysis of finite-dimensional systems, which are not less important considering that any high resolution numerical model is necessarily of finite dimension. In this section, we present the projection of a general finite-dimensional nonlinear system with at most quadratic nonlinearities onto a TEF space and show the relation between EOF- and TEF-based reduced models.

Consider a dynamical system in $\mathbb{R}^n$ with quadratic nonlinearities, where $n < \infty$. The general form of such a system is

$$\begin{aligned} \dot{\mathbf{v}}_i &= A_i^1 + B_{ij}^1 \mathbf{v}_j + C_{ij}^1 \tau_j + D_{ijk}^1 \mathbf{v}_j \mathbf{v}_k + \\ &+ F_{ijk}^1 \mathbf{v}_j \tau_k + G_{ijk}^1 \tau_j \tau_k, \qquad (8.25) \\ \dot{\tau}_i &= A_i^2 + B_{ij}^2 \mathbf{v}_j + C_{ij}^2 \tau_j + D_{ijk}^2 \mathbf{v}_j \mathbf{v}_k \\ &+ F_{ijk}^2 \mathbf{v}_j \tau_k + G_{ijk}^2 \tau_j \tau_k. \qquad (8.26) \end{aligned}$$

Here $\mathbf{v} \in \mathbb{R}^{d_1}$ and $\tau \in \mathbb{R}^{d_2}$ represent vectors in the K- and P-space, respectively, and superscript '1' and '2' refer to the K- and P- spaces, respectively. As before, the total state vector can then be written as $\psi = (\mathbf{v}, \tau)^T$. In terms of the total state vector,

(8.25) and (8.26) can be combined into a single equation of the form

$$\dot{\psi}_i = \alpha_i + \beta_{ij}\psi_j + \gamma_{ijk}\psi_j\psi_k, \tag{8.27}$$

where

$$\alpha = \begin{pmatrix} A^1 \\ A^2 \end{pmatrix},$$

$$\beta = \begin{pmatrix} B^1 & C^1 \\ B^2 & C^2 \end{pmatrix}$$

and

$$\gamma_i = \begin{pmatrix} D_i^1 & F_i^1 \\ 0 & G_i^1 \end{pmatrix}, \quad i = 1, \ldots, d_1,$$

$$\gamma_{d_1+i} = \begin{pmatrix} D_i^2 & F_i^2 \\ 0 & G_i^2 \end{pmatrix}, \quad i = 1, \ldots, d_2.$$

In this case, the empirical eigenvectors are the solutions of

$$CE\phi^{(k)} = \lambda_k \phi^{(k)}$$

where $C = \langle \psi \psi^T \rangle$ is the correlation matrix. As discussed above, there are two alternative ways of expanding a vector in term of the TEFs, either

$$\psi = \sum_{k=1}^{N} a_k \phi^{(k)} \tag{8.28}$$

or

$$\mathbf{v} = \sum_{k=1}^{N} \mu_k \theta_1^{(k)}, \quad \tau = \sum_{k=1}^{N} \nu_k \theta_2^{(k)}. \tag{8.29}$$

In both cases, we deal with an order N approximation. However, (8.29) implies twice the number of degrees of freedom as (8.28).

Following the classical projection algorithm, consisting of taking the appropriate inner product, after substituting (8.28) into (8.27), we obtain the classical result [44]

$$\dot{a}_i = \hat{\alpha}_i + \hat{\beta}_{ij} a_j + \hat{\gamma}_{ijk} a_j a_k, \tag{8.30}$$

where

$$\begin{aligned}\hat{\alpha}_i &= \phi_l^{(i)} E_{lm} \alpha_m, \\ \hat{\beta}_{ij} &= \phi_l^{(i)} E_{lm} \beta_{mn} \phi_n^{(j)}, \\ \hat{\gamma}_{ijk} &= \phi_l^{(i)} E_{lm} \gamma_{mnp} \phi_n^{(j)} \phi_p^{(k)}.\end{aligned}$$

The transformed equations in the extended empirical subspace are obtained by substituting (8.29) into (8.25) and (8.26) and taking the corresponding inner product:

$$\begin{aligned}\dot{\mu}_i &= \hat{A}_i^1 + \hat{B}_{ij}^1 \mu_j + \hat{C}_{ij}^1 \nu_j + \hat{D}_{ijk}^1 \mu_j \mu_k + \\ &+ \hat{F}_{ijk}^1 \mu_j \nu_k + \hat{G}_{ijk}^1 \nu_j \nu_k, \qquad (8.31) \\ \dot{\nu}_i &= \hat{A}_i^2 + \hat{B}_{ij}^2 \mu_j + \hat{C}_{ij}^2 \nu_j + \hat{D}_{ijk}^2 \mu_j \mu_k + \\ &+ \hat{F}_{ijk}^2 \mu_j \nu_k + \hat{G}_{ijk}^2 \nu_j \nu_k. \qquad (8.32)\end{aligned}$$

Fixed points

Fixed points are important entities in phase space because of their influence on the dynamics of systems. We end this section with a discussion on fixed points in reduced-order models. We shall show that in general the projection of a system onto a reduced space gives rise to artificial fixed points that do not correspond to fixed points in the original system.

As pointed out by Moffat [76], fixed points in a reduced model are expected to correspond to fixed points in the original system. We start the analysis of this issue with the EOF reduction (8.30). The fixed points of this system are given by

$$\dot{a}_i = \phi_l^{(i)} E_{lm} \dot{\psi}_m = 0, \qquad (8.33)$$

where $\dot{\psi}_m$ is given by (8.27). According to the projection theorem for Hilbert spaces (e.g. [75]), although the projections of different full system's fixed points may coincide, every fixed point of the full system has a corresponding fixed point in the reduced model. However, the converse depends on the kernel or null space of the matrix $\phi_l^{(i)} E_{lm}$.

One way of finding fixed points is to solve the nonlinear equation

$$\alpha_i + \beta_{ij} \bar{\psi}_j + \gamma_{ijk} \bar{\psi}_j \bar{\psi}_k = \xi_i, \qquad (8.34)$$

where $\xi_i \in \mathrm{Ker}[\phi_l^{(k)} E_{lm}] \subset \mathbb{R}^d$. The solution defines a potential fixed point curve

Table 8.1: Five-dimensional model constants.

	$j=1$	$j=2$
Δ_j	0.0835	0.0301
β_j	0.0095	0.0032
ν_j	0.5236	-0.1752
γ_j	0.2793	-1.7707

in phase space. The intersection of this curve with $\text{span}\{\phi^{(k)}\}_{k=1}^{N}$ determines the location of the fixed points.

In a completely analogous way, an equilibrium point in a TEF reduction (8.32) may be identified by finding the intersection of $\text{span}\{\theta_a^{(k)}\}_{k=1}^{N}$ and a potential fixed point curve given by the solution to

$$\begin{aligned} A_i^a + B_{ij}^a \bar{\mathbf{v}}_j + C_{ij}^a \bar{\tau}_j + D_{ijk}^a \bar{\mathbf{v}}_j \bar{\mathbf{v}}_k \\ + F_{ijk}^a \bar{\mathbf{v}}_j \bar{\tau}_k + G_{ijk}^a \bar{\tau}_j \bar{\tau}_k = f_{ai}, \end{aligned} \tag{8.35}$$

where $f_{ai} \in \text{Ker}[(\mathscr{X}^{-1})^{(ij)} \theta_{ap}^{(j)} E_{apq}] \in \mathbb{R}^{d_a}$ with $\mathscr{X} = \mathscr{K}$ and $a = 1$ for the K-space, and $\mathscr{X} = \mathscr{P}$ and $a = 2$ for the P-space.

A simpler way of locating the fixed point is to directly solve the nonlinear system of algebraic equations given by (8.30) (for an EOF model) and (8.31)-(8.32) (for a TEF model) equated to zero. Nevertheless, the former technique has the potential advantage of containing information about the reduced space. Both methods might be used in conjunction to investigate the structure of the reduced space.

Once the equilibrium points have been found the only way so far to decide whether they actually occur in the original system or they are artificially created during the dimension reduction is by analysing the full-dimensional system. This is not a major disadvantage if we consider that the original ensemble was obtained by numerical integration of the full-dimensional system and therefore it is known and available. Furthermore, the dimension reduction still represents an advantage as it is in general easier to find fixed points in a reduced model than in a full dimensional one if the reduction is significant.

To have a one-to-one correspondence between the fixed points of the complete system and those in the reduced system the only solution to (8.33) should be the trivial solution. However, since $\phi_l^{(k)} E_{lm}$ is an N by d matrix ($N < d$), the latter condition does not hold in general and artificial fixed points may be created in any reduced model. This fact may have important consequences over the faithfulness of

the reduced model to resemble the dynamics of the original one. Moreover, since artificial fixed points arise as a consequence of having unresolved modes, the parameterisation of unresolved modes (closure schemes) can be interpreted as a controlled way of making the reduced system to avoid approaching artificial fixed points.

A five-dimensional example

As an example of the new method we used a system of five dimensions originally obtained as a severely truncated two-layer quasi-geostrophic model of baroclinic instability [69];

$$\begin{aligned}
\dot{\mathbf{v}}_1 &= -\Delta_1 \mathbf{v}_1 + \beta_1 \mathbf{v}_2 - (\nu_1 + \gamma_1 \tau_3)\tau_2, \\
\dot{\mathbf{v}}_2 &= -\beta_1 \mathbf{v}_1 - \Delta_1 \mathbf{v}_2 + (\nu_1 + \gamma_1 \tau_3)\tau_3, \\
\dot{\tau}_1 &= -\Delta_2 \tau_1 + \beta_2 \tau_2 - (\nu_2 + \gamma_2 \tau_3)\mathbf{v}_2, \\
\dot{\tau}_2 &= -\beta_2 \tau_1 - \Delta_2 \tau_2 + (\nu_2 + \gamma_2 \tau_3)\mathbf{v}_1, \\
\dot{\tau}_3 &= -\bar{\Delta}\tau_3 + \bar{\gamma}(\mathbf{v}_1 \tau_2 - \mathbf{v}_2 \tau_1),
\end{aligned} \tag{8.36}$$

where Δ_j, β_j, ν_j, γ_j with $j = 1, 2$ and $\bar{\Delta} = 0.0279$, $\bar{\gamma} = 3.8554$ are all real-valued parameters chosen so that the system was in a chaotic regime. The values of the remaining parameters are given in Table 8.1. For details on the derivation of the original model the interested reader is referred to [69].

We are not interested here in the physical meaning of these equations. Therefore, the system will be considered only from an abstract point of view. We just point out that $\mathbf{v}$ is associated with baroclinic components whereas τ is related to barotropic ones [69]. Thus, in this case it would be perhaps more sensible to talk of baroclinic/barotropic energy rather than kinetic/potential energy. The energy matrices are $E_1 = I_2$ and $E_2 = I_3$ so that the full-dimensional phase space is indeed measured by the variance norm.

Table 8.2: Empirical eigenvalues. (TE = Total energy; CTE = Cumulative total energy)

k	λ_k	% TE	% CTE
1	0.00800	67.00	67.00
2	0.00240	19.65	86.65
3	0.00150	12.26	98.91
4	0.00008	0.66	99.57
5	0.00005	0.43	100.0

Table 8.3: Transformed basis energy contents.

	$\theta^{(1)}$	$\theta^{(2)}$
$\mathbf{v}_1$	0.0339	-0.7117
$\mathbf{v}_2$	0.0354	0.6812
τ_1	0.0071	0.1102
τ_2	-0.0052	0.1315
τ_3	-0.9988	0.0001
K (%)	0.24	97.06
P (%)	99.76	2.94

The original dataset consisted of a single integration of (8.36) for a time interval $t = 5000$ with a time step $\Delta t = 1/3$ or $M = 15000$ snapshots. $\mathbf{v}(0) = (0.00, 0.01)^T$ and $\tau(0) = (0.02, 0.01, -0.08)^T$ were used as initial conditions. Since the time interval is finite, the exact numerical values of the EOF will, in general, depend on the original dataset. However, the qualitative behaviour is general and does not depend on the original time series.

Since each TEF implies two degrees of freedom, retaining one EOF would lead to two degrees of freedom in the extended space. On the other hand, since the original system is five-dimensional, retaining more than two EOFs would lead to extended spaces of more degrees of freedom than the original. As a result, the only sensible number of EOFs to be retained in order to calculate TEFs is two. Consequently we analysed a model projected onto two EOFs and the corresponding 2-TEF reduced model. Of course it can be argued that both reductions are not comparable because the TEF model implies four degrees of freedom while the corresponding 2-EOF model implies only two. under these circumstances, we also performed the analysis of a model projected onto four EOFs with four degrees of freedom.

Table 8.2 shows the empirical eigenvalues for the full-space POD. It can be seen that an EOF-based reduced space retaining two modes explains 86.65% of the original system average TE whereas a four-dimensional reduced space contains 99.57%. Despite the large average TE content in the latter, we shall see that the performance of the 4-EOF model is less faithful than that of the 2-TEF model.

Table 8.3 shows the TEF components in the original space. Since the K-space is two-dimensional, it is totally spanned by $\theta_1^{(1)}$ and $\theta_1^{(2)}$. In contrast, $\theta_2^{(1)}$ and $\theta_2^{(2)}$ reduce the three-dimensional P-space into a two-dimensional space almost perpendicular to the τ_1-τ_2 plane. The percentage of KE and PE in each TEFs is also displayed in Table 8.3. Note that the first TEF is characterised by a large amount of PE whereas the second contains almost all the KE. This is an observation that can also be made

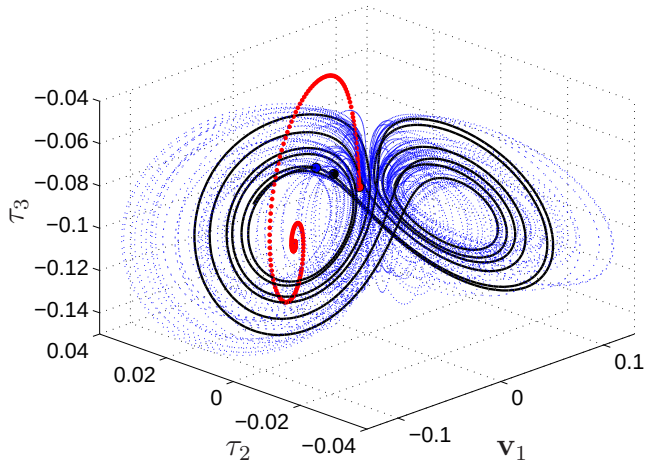
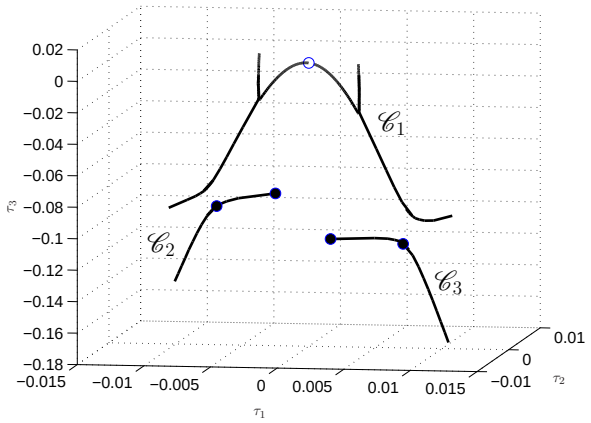


Figure 8.2: The subspace $(\mathbf{v}_1, \tau_2, \tau_3)$ showing: (a) the original five-dimensional system (dotted); (b) the projection onto two EOFs (bold dotted), and (c) the projection onto two TEFs (solid).

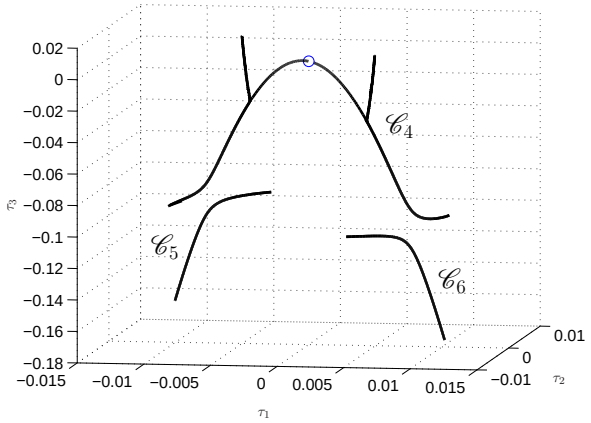
in large dimensional systems. This fact could be considered as a means of classifying the TEFs. However, it is not possible to reduce the space further (by selecting the most energetic modes, for example) since the TEFs are derived from a set of EOFs chosen previously and this set determines the amount of average TE explained by the truncation.

In order to verify the theoretical remarks presented in Section 8.1.1, the original ensemble was projected onto span$\{\theta^{(1)}, \theta^{(1)}\}$ according to both (8.7) and (8.9). As expected, the ensemble average of TE using (8.7) gave $\langle E_2^c \rangle = \lambda_1 + \lambda_2 = 0.0104$, whereas (8.9) yielded $\langle E_2^d \rangle = 0.0119 \geq \langle E_2^c \rangle$. It is interesting to note that in this case $\langle E_2^d \rangle = \sum_{k=1}^{4} \lambda_k = \langle E_4^c \rangle$. In fact, every numerical experiment starting from different initial conditions and carried out during different integration intervals produced similar results with $\langle E_2^d \rangle$ close to the sum of the first four empirical eigenvalues. This observation might be an indication of the relation between a N-TEF subspace and a $2N$-EOF subspace.

The 2-EOF model produced a generic orbit as shown in figure 8.2 (bold dotted line). As a reference, the orbit of the full-dimensional system used as original dataset is also shown (dotted line). Since the reduced space is two-dimensional it was not expected to retain the chaotic character of the full-dimensional system. Instead, two stable fixed points were created. Each stable fixed point has its own basin of attraction. which, given a set of initial conditions, determines the final state of the reduced system.



(a) 4-EOF model



(b) 2-TEF model

Figure 8.3: Potential fixed point curves in (a) a 4-EOF model and (b) a 2-TEF model for the five-dimensional system (8.36) in the P-space. The circles represent actual (white) and artificial (black) fixed points.

The 4-EOF model exhibited artificial fixed points as well. As discussed in Section 8.1.3, the fixed points in a reduced model are given by intersections between the reduced space and the potential fixed point curves given for an EOF model by (8.34). The curves and fixed points in the 4-EOF model are shown in figure 8.3(a). The white dot symbolises the only actual fixed point occurring in the original model at the origin. The black dots mark four artificial fixed points produced by the dimension reduction. As can be seen, Figure 8.3 reveals a fundamental difference between reduced spaces of different nature in this particular example. The 4-EOF reduced space structure consists of the intersection between the original five-dimensional space and the four-dimensional 4-EOF span. Its structure is complex as is evident from the position of the four additional intersections corresponding to artificial fixed points. In comparison, the 2-TEF reduction implies a reduction of both the K- and P-space into a two-dimensional space each. Hence, the 2-TEF reduced P-space is just a plane to which $\mathscr{C}_4$ is transversal. As a result, the 2-TEF model only exhibit one fixed point corresponding to the actual one at the origin (figure 8.3(b)).

The existence of artificial fixed points in the 4-EOF has significant influence on the behaviour of the reduced model. A generic orbit is shown in figure 8.4 as a solid line. At the beginning, it resembles the original five-dimensional system (dotted line). However, it eventually fell into the neighbourhood of an artificial fixed point. In contrast, the 2-TEF reduction with its four degrees of freedom exhibited a more faithful behaviour. It displayed chaos as can be seen from the generic orbit shown in solid line in the same figure. Furthermore, the phase space regions visited by the complete and the reduced systems seem to coincide (bearing in mind that the reduced space in the P-space is two dimensional).

The analysis of the average TE (with respect to the integration time) in the reduced models produced interesting results. Using (8.13) to compute the average TE in the 2-EOF model (substituting ensemble average by time average) yielded $\langle E_2^{\mathrm{EOF}} \rangle = 0.0152$, a value that is much bigger than the value computed from the projection of the original dataset ($\langle E_2^c \rangle = 0.0104$). The time-averaged TE in the 2-TEF model, on the other hand, remained equal to its initial value $\langle E_2^d \rangle = 0.0119$. This fact reinforces the claim that the 2-TEF model is more faithful than the 2-EOF model. In the case of the 4-EOF model, the application of (8.13) to compute the time average total energy yielded $\langle E_2^{\mathrm{EOF}} \rangle = 0.0145$, a value that is greater than the expected $\langle E_4^c \rangle = \sum_{k=1}^{4} \lambda_k = 0.0119$.

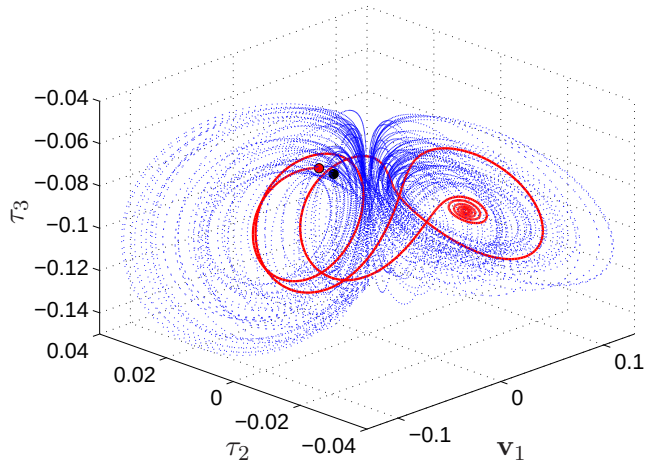


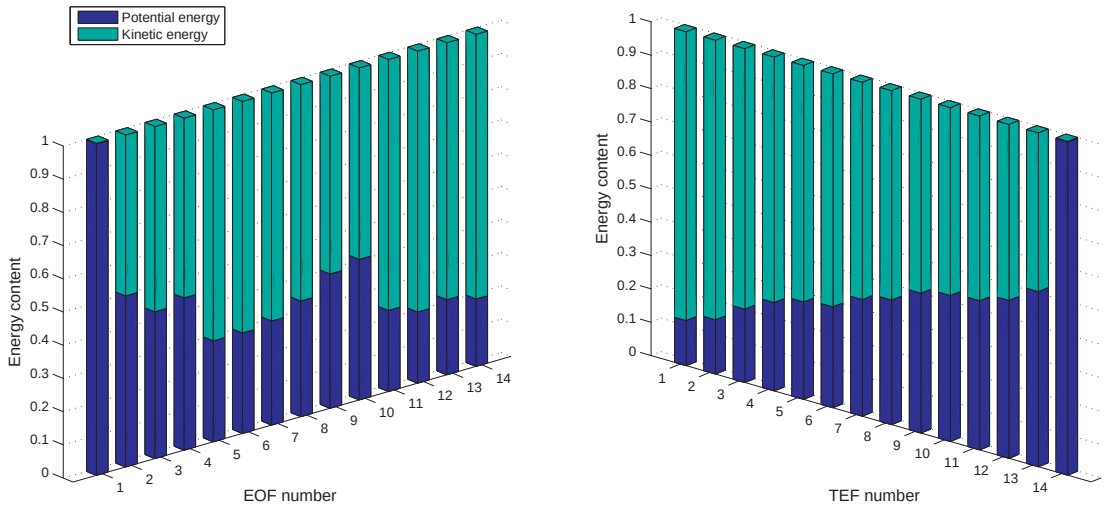
Figure 8.4: The subspace $(\mathbf{v}_1, \tau_2, \tau_3)$ showing: (a) the original five-dimensional system (dotted), and (b) the projection onto four EOFs (solid).

8.1.4 Infinite-dimensional space

On the grounds of the theoretical and numerical evidence just provided, we return to the study of the Mars GCM and analyse the infinite dimensional case within this context. We consider the maximum TE subspace given by the span of the first 14 EOFs, which represent 86.5% of average TE (cf. Section 6) and which has proved to produce well-behaved models for various truncation orders (cf. Chapter 7).

Figure 8.5 shows the distribution of energy throughout the EOFs and corresponding TEFs. As expected, these two distributions are different. Perhaps, the most important difference between these distributions is the lack of order, regarding PE (or KE) content, in the EOFs. However, there is a striking similarity between EOF1 and TEF14[1]. In fact, the difference in PE content can be found in the second decimal digit (99.77% for EOF1 versus 99.79% for TEF14) while the angle between them is only 0.8330 °. Therefore, we can identify TEF14 with the TEF representation of the background state composed by a large PE which is not transformed into KE. This result is not peculiar to the truncation order presented here and can be found, in general, at any truncation order. Thus, the claim in Section 3.3.1 regarding the use of TPE rather than APE in the definition of an appropriate inner product is further supported by these findings. This is an added advantage that has arisen from using the TE-norm and the full system's state without neglecting interaction between mean

[1]Recall that the TEF order is arbitrary. It can be done according to ascending or descending PE content.



(a) EOFs (b) TEFs

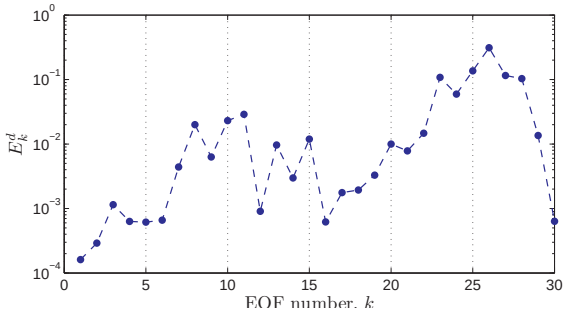
Figure 8.5: Energy distribution over (a) 14 EOFs and (b) the corresponding TEFs.

state and anomalies.

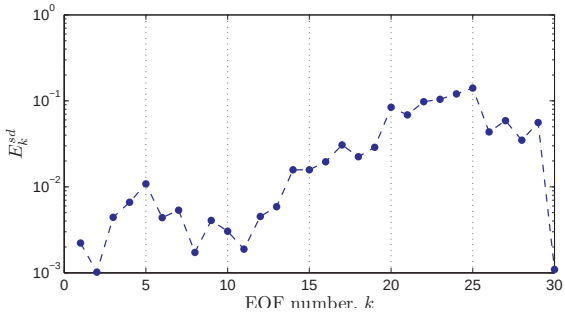
Figure 8.6 shows the decomposition of the diurnal and semidiurnal tides and the diurnal Kelvin wave over 30 EOF. This number was chosen to have a direct comparison with the results in Section 6. After comparing the figures in that section with those presented here, one immediately notices the lack of a pattern in the TEFs tidal decompositions. For example, there is no group of modes clearly representing either tide. Instead, the tides are distributed over the whole set with a slight shift towards those modes with higher PE content. Moreover, while in the EOF case two pairs (EOF3-4 and EOF18-19) were sufficient for explaining 98% of diurnal TE, the same amount of energy is distributed among 16 TEFs in the present case. Thus, the decomposition of tides reveals every TEF as a mixture of atmospheric wave without any clear physical interpretation. This constitutes one major drawback with using TEFs. However, the physical interpretation of the results can be recovered in a simple way by means of the transformation equations (8.16) and (8.12).

8.2 Ab-initio decoupled EOFs

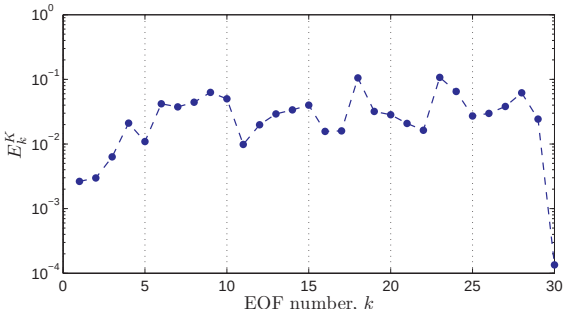
Achatz [1] has suggested the potential convenience of computing ab-initio decoupled EOFs (D-EOFs) as a means of avoiding the huge computation involved in extracting full-space EOFs. This comment has motivated the following investigation.



(a) Diurnal tide



(b) Semidiurnal tide



(c) Diurnal Kelvin wave

Figure 8.6: Tidal decomposition over 30 TEFs: (a) diurnal tide, (b) semidiurnal tide and (c) diurnal Kelvin wave.

8.2.1 Definition

The theory in Chapter 3 is perfectly valid in this case although the original dataset is divided into K-space and P-space from the beginning. The relevant inner products are those defined by (8.2) and (8.3), respectively, and the D-EOFs satisfy the eigenvalue equations

$$\int_V \langle \psi_l(x)\psi_p^*(x')\rangle \, E_{pq} \, \varphi_q(x') \, dV' = \lambda_K \varphi_l, \tag{8.37}$$

with $l, p, q = 1, \ldots, d_1$ and

$$\int_V \langle \psi_m(x)\psi_r^*(x')\rangle \, E_{rs} \, \varphi_s(x') \, dV' = \lambda_P \varphi_m, \tag{8.38}$$

with $m, r, s = d_1 + 1, \ldots, d_1 + d_2$, where $k = 1, \ldots, \infty$ and d, d_1 and d_2 are the dimensions of the full-, K- and P-spaces, respectively. For the sake of simplicity let us denote a KE-related vector by a subindex 1 and a PE-related vector by a subindex 2 as we did in Section 8.1.1. Thus, we can denote the new bases as $\Phi_1 = \{\varphi_1^{(k)}\}_{k=1}^{\infty}$ and $\Phi_2 = \{\varphi_2^{(k)}\}_{k=1}^{\infty}$, whose elements are orthonormal with respect to the appropriate norm.

The components of a state vector can be decomposed in these new bases as

$$\psi_1^N(x,t) = \sum_{k=1}^{N_1} c_k(t) \, \varphi_1^{(k)}(x) \tag{8.39}$$

$$\psi_2^N(x,t) = \sum_{k=1}^{N_2} d_k(t) \, \varphi_2^{(k)}(x), \tag{8.40}$$

where

$$c_k(t) = (\varphi_1^{(k)}(x), \psi(x,t))_K, \quad \text{and} \quad d_k(t) = (\varphi_2^{(k)}(x), \psi(x,t))_P. \tag{8.41}$$

Notice that in this case the vector components are completely independent of each other and, therefore, the truncation orders N_1 and N_2 are not necessarily the same.

8.2.2 Ab-initio decoupled eigenvalues for the MGCM

Figure 8.7 shows the eigenvalues corresponding to each subspace (K- and P-spaces). It also shows the eigenvalues obtained from the full-space POD (cf. figure 6.4). As can be seen from this figure, at $N = 1$ the energy is mostly PE in accordance with the large PE content in EOF1 or the last TEF. PE decays more rapidly than KE as

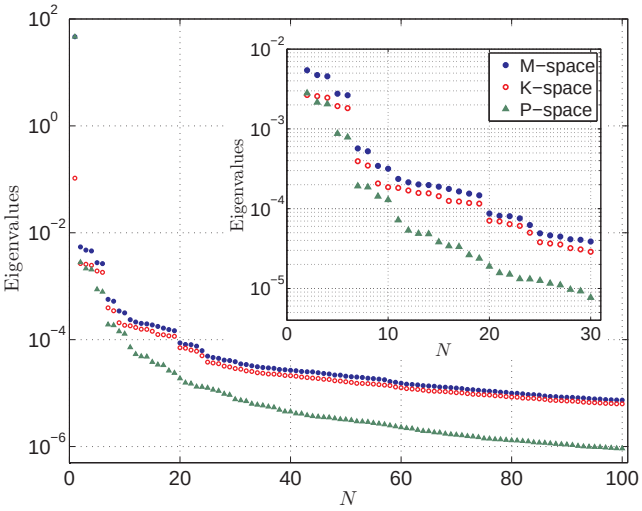


Figure 8.7: Eigenvalues from PODs carried over full space (M-space) (cf. figure 6.4) and separate subspaces (K- and P-spaces).

N increases so that, at $N = 3$, KE content in a given K-EOF is larger than PE in the corresponding P-EOF. After $N = 5$, the energy distribution in the K-space very closely resembles that in the full-space while the energy in the P-space is already very low. This can be interpreted as an indication that the horizontal velocity components are more dynamically important for the atmospheric evolution on Mars. The EOFs groups are still apparent in both eigenspectra and remain the same as in the full-space POD case.

8.2.3 Comparison with TEFs and full-state EOFs

In order to understand the relation between the D-EOFs and the TEFs, let us decomposed the latter in terms of the former to obtain

$$\theta_1^{(k)} \simeq \sum_j c_j^{(k)} \varphi_1^{(j)} \tag{8.42}$$

$$\theta_2^{(k)} \simeq \sum_j d_j^{(k)} \varphi_2^{(j)}, \tag{8.43}$$

where

$$c_j^{(k)} = (\varphi_1^{(j)}, \theta_1^{(k)})_K, \quad \text{and} \quad d_j^{(k)} = (\varphi_2^{(j)}, \theta_2^{(k)})_P. \tag{8.44}$$

Taking the K-product of $\theta_1^{(l)}$ and (8.42) we have

$$(\theta_1^{(l)}, \theta_1^{(k)})_K = \mathscr{K}^{(kl)} \simeq \sum_j c_j^{(k)}\, c_j^{(l)*}. \tag{8.45}$$

Similarly, taking the P-product of $\theta_2^{(l)}$ and (8.43) yields

$$(\theta_2^{(l)}, \theta_2^{(k)})_P = \mathscr{P}^{(kl)} \simeq \sum_j d_j^{(k)}\, d_j^{(l)*}. \tag{8.46}$$

Using (8.9) and the expansions (8.39) and (8.40), we find, after assuming $N_1 = N_2 = N$, the relation between the TEF coefficients and the D-EOFs:

$$c_j \simeq c_j^{(k)} \mu_k \quad \text{and} \quad d_j \simeq d_j^{(k)} \nu_k. \tag{8.47}$$

Now, the average TE in an order N approximation to ψ using D-EOFs is given by

$$\langle E_N^{\mathrm{dEOF}} \rangle = \sum_{k=1}^{N} \left(\langle c_k(t) c_k^*(t) \rangle + \langle d_k(t) d_k^*(t) \rangle \right). \tag{8.48}$$

Subtracting the K-component in (8.15) from the corresponding component in (8.48) we obtain

$$\sum_k \left(\langle c_k\, c_k^* \rangle - K_k \langle \mu_k\, \mu_k^* \rangle \right) \simeq \sum_{j,l} \mathscr{K}^{(jl)} \langle \mu_j\, \mu_l^* \rangle - \sum_k K_k \langle \mu_k\, \mu_k^* \rangle = 0, \tag{8.49}$$

where (8.45) and (8.47) have been used. If the equality were satisfied, (8.49) would show that D-EOFs and TEFs are expected to explain almost the same amount of KE. However, the equality is only valid when $N \to \infty$ but, in this case, this analysis would not make any sense as any given decomposition would be equivalent to any other. A similar conclusion can be drawn regarding PE by means of an analogous procedure. As an immediate consequence, D-EOFs would explain more TE than full-space EOFs at the same truncation order. From the point of view of average energy content, given the result expressed by (8.49), TEFs and D-EOFs would yield equivalent decompositions. However, we shall see that, in general, there are structural differences between TEFs and D-EOFs.

A more accurate and practical way of comparing different decompositions is to

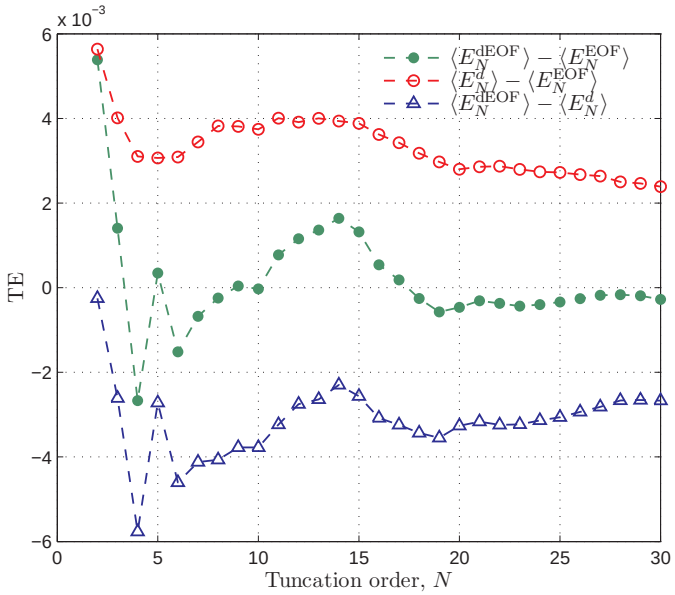


Figure 8.8: Differences in TE between EOFs and D-EOFs (green dots), TEFs and EOFs (red open circles) and TEFs and D-EOFs (blue triangles) as functions of truncation order. All quantities are normalised with respect to the eddy average TE in a full-space EOF decomposition with truncation order N.

use (3.14) which is also valid in the case of D-EOFs in the form

$$\langle E^{\mathrm{dEOF}}_{N_1,N_2}\rangle = \sum_{k=1}^{N_1} \lambda_K^{(k)} + \sum_{k=1}^{N_2} \lambda_P^{(k)}. \tag{8.50}$$

If $N_1 = N_2 = N$

$$\langle E^{\mathrm{dEOF}}_{N}\rangle = \sum_{k=1}^{N} \left(\lambda_K^{(k)} + \lambda_P^{(k)}\right). \tag{8.51}$$

This expression is available for comparison with the average energy in an full-state EOF-based decomposition, namely

$$\langle E^{\mathrm{dEOF}}_{N}\rangle - \langle E^{\mathrm{EOF}}_{N}\rangle = \sum_{k=1}^{N} \left(\lambda_K^{(k)} + \lambda_P^{(k)} - \lambda^{(k)}\right). \tag{8.52}$$

Subtracting (8.19) from (8.52) gives the difference between a TEF and a decoupled EOF decomposition

$$\langle E^{\mathrm{dEOF}}_{N}\rangle - \langle E^{d}_{N}\rangle = \sum_{k=1}^{N} \left[\lambda_K^{(k)} + \lambda_P^{(k)} - \left(\lambda^{(k)} + P_k K_k \left\langle |\mu_k - \nu_k|^2 \right\rangle\right)\right]. \tag{8.53}$$

Figure 8.8 shows the results from these comparisons. The difference between the energy explained by D-EOFs and full-state EOFs is positive only for

$$N \in \{11, 12, 13, 14, 15, 16, 17\},$$

having a maximum at $N = 14$. TEF models for $N < 4$ are not being considered since they do not include any travelling wave EOF group and, therefore, are unable to represent such waves. In contrast, TEFs always capture energy more efficiently than EOFs, reaching a plateau at $N = 8$ which extends until $N = 15$. The comparison between D-EOFs and TEFs is, not surprisingly, always negative for the decoupled EOFs.

The quantities shown in figure 8.8 have been normalised with respect to the eddy average energy that would be expected using full-space EOFs. This is given by (3.14) with j running from 2, rather than 1, to N. Thus, we can see that in general the improvement of using different bases is relatively weak. The actual improvement of using one or another basis will be evaluated in the next chapter where we will be concerned with the construction and evaluation of reduced models.

8.3 Reduced-order dynamical models

8.3.1 Reduced-order models over TEFs

Building low-order models over TEFs follows, in essence, the same method as for building models over EOFs. For the sake of simplicity, it is convenient to normalise TEFs so that the KE and PE fractions (given by K_k and P_k, respectively) are not carried over during projections. The normalised versions of the TEFs are given by

$$\vartheta_1^{(k)} = \frac{\theta_1^{(k)}}{K_k^{1/2}} \quad \text{and} \quad \vartheta_2^{(k)} = \frac{\theta_2^{(k)}}{P_k^{1/2}}. \tag{8.54}$$

With these definitions, the decomposition of the state fields is (cf. (8.9))

$$\mathbf{v}(x,t) = \sum_{k=1}^{N} v_k(t)\vartheta_1^{(k)}(x) \tag{8.55}$$

$$\tau(x,t) = \sum_{k=1}^{N} \tau_k(t)\vartheta_2^{(k)}(x), \tag{8.56}$$

Table 8.4: Relative norm temporal mean for TEF-based reduced models at different truncation orders after 90 sols.

N	Rel. Norm Mean	Rel. Norm Std. Dev.	Angular Mean	Angular Std. Dev.
4	1.0372	0.0262	7.6490	3.3633
6	1.0560	0.0713	5.8155	4.8300
8	1.0016	0.0013	0.9855	0.1604

where (cf. (8.10))

$$v_k(t) = (\vartheta_1^{(k)}(x), \psi(x,t))_K, \quad \text{and} \quad \tau_k(t) = (\vartheta_2^{(k)}(x), \psi(x,t))_P, \tag{8.57}$$

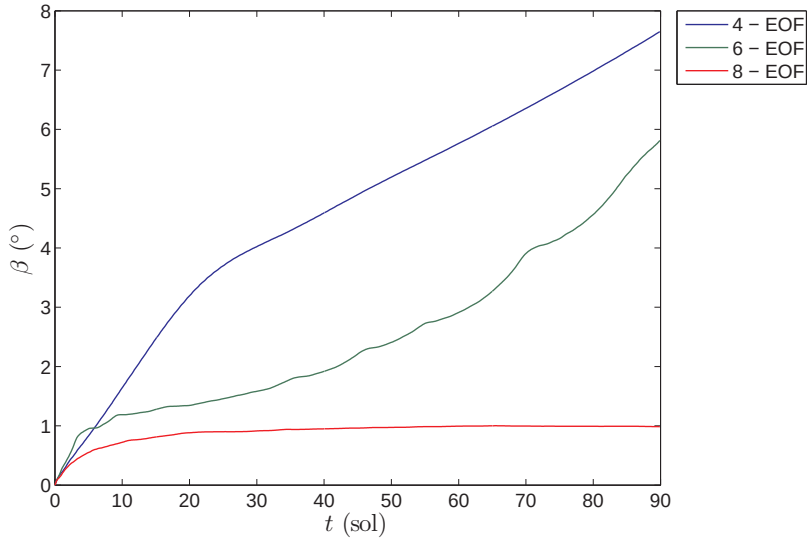
The equations to be projected are (5.1) and (5.2). The unresolved modes can be parameterised using the same approach as described in Section 5.1.1 although the theoretical form of the operators to be determined would be more complex. This, however, does not imply any further computing complications apart from the increased size of these operators for a given number of fundamental EOFs. Thus, a TEF model incorporating parameterisation of unresolved modes is given by (5.15) (rewritten here for clarity)

$$\dot{a}_k = \alpha_k + \beta_{kl} a_l + \gamma_{klm} a_l a_m + \Delta_k + Z_{kl} a_l, \tag{8.58}$$

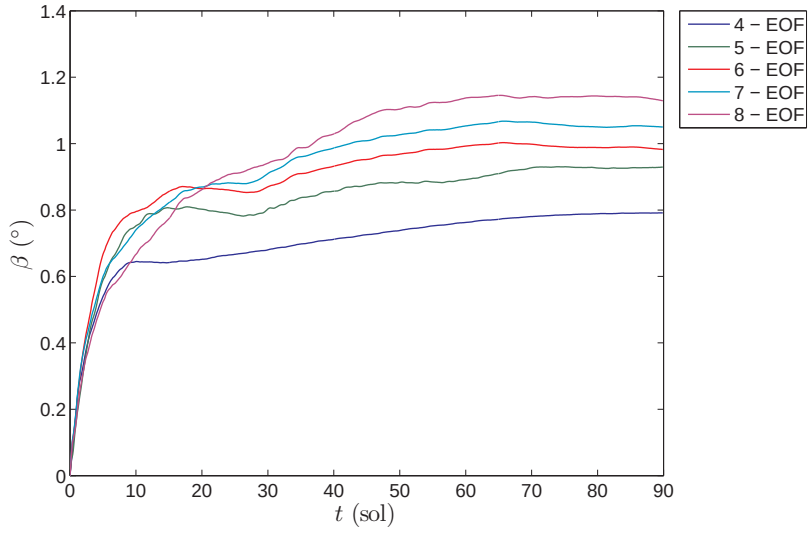
with the vector a_k being a concatenation of v_k and τ_k, namely $a_k = (v_k, \tau_k)^T$, and the sub-indices k, l, m running from 1 to N for the the K-modes and from $N+1$ to $2N$ for the P-modes. In this case, Δ_k and Z_{kl} introduce $2N$ and $(2N)^2$ scalar parameters, respectively. The rapid increase in the number of parameters to be determined due to unresolved modes is a considerable disadvantage to the use of TEFs in comparison with EOFs.

8.3.2 Reduced-order models over D-EOFs

The method for constructing low-order models over D-EOFs is in all respects the same as that described in Section 8.3.1, employed for building models over normalised TEFs. This implies the use of (8.55) and (8.57), with ϑ being substituted by φ, to represent the prognostic fields. Similarly, the form of the resulting D-EOFs models would be given by (8.58).



(a) TEFs



(b) D-EOFs

Figure 8.9: Mean angular deviation, defined by (5.18), in MGCM reduced models for different truncation orders as a function of time using (a) TEFs and (b) D-EOFs as basis functions.

Table 8.5: Relative norm temporal mean for D-EOF-based reduced models at different truncation orders after 90 sols.

N	Rel. Norm Mean	Rel. Norm Std. Dev.	Angular Mean	Angular Std. Dev.
4	0.9973	0.0018	0.7907	0.1146
5	1.0025	0.0017	0.9282	0.1833
6	1.0017	0.0012	0.9798	0.1662
7	1.0007	0.0005	1.0485	0.1891
8	0.9985	0.0013	1.1287	0.2578

8.3.3 Model inter-comparison

Figure 8.9 shows the angular deviations in TEF- (figure 8.9(a)) and D-EOF-based models (figure 8.9(b)). These figures only show the models for $N \geq 4$ (which include travelling waves) that have produced bounded solutions below $\beta = 8\,^\circ$ during the test time interval (90 sols). No model for $N > 8$ has complied with these requirements. These results have two surprising characteristics. Firstly, the TEF-based models, despite explaining more TE in average than D-EOF (cf. Section 8.2.3), have displayed worse performance. In fact, as can be seen from figure 8.9(a), just a few TEF-model integrations remained below $\beta = 8\,^\circ$. Still, the 8-TEF model seems to produce better results than its EOF-based counterpart (see figure 7.2). A model including the strongest part of the semidiurnal tide (EOF9–10) was unsuccessful, although it is still explained by EOF5–6 to a certain extent (14.9 % of TE in that tide, cf. Section 6.4.2).

A second notable feature of the alternatively based models is that every D-EOF-based model exhibited better behaviour than the corresponding EOF-based models despite explaining less or an equal amount of TE in average than the EOFs (cf. Section 8.2.3). This can be seen by comparing figures 8.9(b) and 7.2, where every EOF-based model is shown to have a mean angular deviation above $1\,^\circ$. In fact, looking at tables 8.4 and 8.5, we can see that every evaluation variable (mean and standard deviation of relative norm and angle), concerning the 8-TEF model and every D-EOF-model, describes a remarkably good behaviour of the corresponding models.

8.4 Summary and conclusions

After discussing a potential theoretical problem that arises with the construction of low-order models based on multi-dimensional, multi-variate EOFs, we have shown

two possible ways to overcome it. Both options are grounded on the idea of having an independent basis for each variable.

As a first alternative, we investigated a unitary transformation of the multi-variate three-dimensional EOFs obtained in Chapter 6. Based on this idea, a five-dimensional test model was reduced with remarkably good results: a 2-TEF model showed to be a better approximation to the original (chaotic) dynamics than 2- and 4-EOF models. A second alternative resulted from a simpler procedure consisting of performing the POD over each state variable separately.

The results obtained from the construction of reduced-order models over these two alternative bases are contradictory. On the one hand, one 8-TEF and every D-EOF-based reduced model exhibited better behaviour in the sense that their output seems to be closer in phase space to the MGCM dataset than the solutions provided by EOF-based models. On the other hand, the integration of these models seems to be subject to larger instabilities that did not allow us to go beyond $N = 8$. Every model after this limit produced unbounded solutions after finite time. These contradictory conclusions, in the light of the preliminary results presented, show the potential value of exploring these alternatives but also give evidence of the need of further research in this direction.

In particular, there is one important ingredient which has not been added to any of the reduced models presented in this thesis and which seems to be particularly important in the case of D-EOF- and TEF-based models. This ingredient is additional dissipation and is necessary even in full GCMs as a means of removing energy at small scales (or high wavenumber, in the case of spectral models) [59, 91]. As explained at the end of Section 5.2.2, it is also essential for achieving long-term realistic models. Two proposed methods that have proved successful adding extra dissipation are Newtonian damping [2, 3] and additional horizontal diffusion [59]. The latter is perhaps the preferred method in the case of full GCMs [36].

Chapter 9

Conclusions

9.1 Summary

Motivated by the observation of enhanced regularity of certain transient components of the Martian atmosphere in comparison with the terrestrial atmosphere, and encouraged by previous work on simplified Martian atmospheric low-order models, this thesis has focused on two different aims:

(*i*) to evaluate the possibility of realistically representing the dynamical behaviour of the Martian atmosphere by a reduced number of atmospheric modes, and

(*ii*) to construct low-order dynamical models of this system based on the modes determined in (*i*).

The dataset which constituted the main target of this study was a sample output of the Oxford Mars general circulation model (MGCM) [36].

In the pursuit of these aims, two complementary techniques were used. The POD, based on a total energy norm, was employed to find an optimal set of basis functions to represent the original dataset as a decomposition of a small number of atmospheric modes. Additionally, a combination of the POD and Galerkin projection was adopted in order to construct low-order models of the Martian atmosphere. Thus, the primitive equations were projected onto the span of the empirical orthogonal functions (EOFs) obtained from the POD. The hydrostatic primitive equations were favored over the simpler quasi-geostrophic theory due to two reasons: first, the order of the planetary Rossby number (being around 0.4) is not as small as is necessary to ensure a good approximation of the actual dynamics by QG theory and secondly, the large-scale Martian topographic features required the use of terrain-following coordinates.

Before pursuing results for the principal problem of interest (the MGCM dataset), these two methods were applied onto a simpler, precursory version of the MGCM, the so-called simplified GCM (SGCM) [26]. The principal differences between the SGCM and the MGCM is that the former does not include topography and the parameterisation of forcing consists solely of Newtonian cooling and Rayleigh friction, whereas the latter includes topographic information and a state-of-the-art parameterisation of physical processes, which represents forcing and dissipation in the model.

The POD applied to the SGCM dataset showed that the total energy (TE) was concentrated in a relatively low number of EOFs. As a consequence of the lack of topography, the first and most energetic mode corresponded to a highly axisymmetric zonal flow. The next most energetic group of modes corresponded to baroclinic waves. This result was expected given the lack of diurnal cycle forcing other kind of atmospheric motions such as thermal tides. These findings were in concordance with previous studies based on the same simplified model [124]. The SGCM EOFs were classified in groups according to their spatial structure and temporal properties of the corresponding principal components (PCs). By projecting the hydrostatic primitive equations onto the span of the most energetic EOFs, we were able to produce a variety of reduced-order dynamical models. The projection of the equations of motion onto the span of the most energetic EOFs provided a means of finding physically plausible interactions between the resolved modes.

The results for the SGCM were highly satisfactory, providing confidence in the analysis of the more realistic MGCM dataset. As before, the POD showed the feasibility of reducing the order of the MGCM dataset by means of an EOF approximation. In this case, the most energetic mode, EOF1, was again related to a nearly zonal background flow, in this case, modulated by topography. Moreover, PC1 exhibited a slow decay corresponding to the seasonal cycle. In contrast with the SGCM case, the next most energetic modes represented the diurnal tide as a result of the strong diurnal forcing imposed on this model. As in the SGCM case, the MGCM EOFs were classified taking into consideration their spatial structure and the frequencies embedded in the evolution of the corresponding PCs. This classification was complemented with an analysis of the projection of thermal tides (diurnal, semidiurnal, diurnal Kelvin and westward propagating wavenumber-3 diurnal wave) and transient baroclinic waves. Thanks to this classification, it was possible to provide a physical interpretation of the MGCM EOFs beyond their energy content.

Reduced-order dynamical models for the MGCM were also constructed by projecting the hydrostatic primitive equations onto the subspace spanned by the most

energetic EOFs. In addition to the analysis established for the study of the SGCM low-order models, the MGCM models were analysed by observing the change in the models' frequency response after retaining different EOFs. This analysis led to the identification of interactions between transient waves and the semidiurnal tide with other waves such as the diurnal tide, the diurnal Kelvin wave and the wavenumber-3 diurnal wave characterised by a period of 1 sol.

Numerical experiments using a 14-EOF model were carried out in order to investigate the relevance of the diurnal cycle in the evolution of baroclinic waves. These experiments shed further light into the question of the interaction between different kinds of atmospheric motion.

Finally, in an attempt to achieve a better approximation to the tendency space, which is not necessarily the same as the state space, alternative reduced spaces were investigated. These alternative spaces were obtained by either decoupling the components of multivariate three-dimensional EOFs or directly applying the POD over independent components. Although, in theory, the use of these bases should involve advantages, such as a more accurate representation of state variables, in practice, the results seem to indicate that these models are at most as good as the simpler EOF models.

9.2 Conclusions

Given the results just summarised, we can now give answers to the questions posed in the introduction. To the question as whether there is a small number of significant atmospheric modes being excited in a realistic simulation of the Martian atmosphere, the answer is yes, there is at least one set of functions that can be used to describe the state of the atmosphere by retaining just a small number of members in the set. This set is constituted by the EOFs. There may be alternative sets equally useful as an answer to our question. Moreover, there may be different EOF sets depending on the definition of the norm employed to measure the phase space. In this thesis, each EOF is associated with a certain amount of average TE and each can also be associated with certain atmospheric motions known to be important for the general circulation of the Martian atmosphere.

To the question as whether the atmospheric modes, determined by the answer to the previous query, are sufficient to describe the dynamical behaviour of the atmosphere of Mars, the answer is not as rigorous as in the previous case. We have certainly found that the EOFs can interact dynamically to approximately resemble

the actual evolution of the corresponding PCs within the MGCM. This resemblance was assessed by the reduced models' ability to evolve near the region in phase space where the original system evolves and their capability for reproducing the most important frequencies in the corresponding PCs' Fourier spectra. These are global methods appropriate for nonlinear systems and, from this point of view, the results were satisfactory. However, the detailed reproduction of the original orbit was not achieved. This is not entirely surprising as we are dealing with a nonlinear system likely to be sensitive to initial conditions. Moreover, the projection of the initial state onto the EOF reduced-space necessarily implies a perturbation to the initial conditions and, as a result, a different orbit in phase space.

The accuracy of the reduced models depends not only on the number of EOFs being resolved in the reduced models but also on the completeness of EOF groups representing a certain type of atmospheric wave. For example, the pair EOF3–4 is a good representative of the diurnal tide and is also part of the decomposition of other atmospheric motions such as transient waves. To achieve travelling wave behaviour it is necessary to include both EOFs. Otherwise, the most that can be attained is a standing wave. This is analogous to having real and imaginary parts in a Fourier series.

Another important ingredient for the success of a reduced model was found to be the interaction of resolved modes with certain higher order modes. As a result, it is necessary to include these as explicitly resolved modes in order to obtain the correct modulation of the evolution of the modes of interest. Suppose, for example, that we are interested in reproducing the evolution of the pair EOF3–4 and that we include only the first four EOFs. As we have just discussed, including both modes enables the model to represent travelling waves. If the forcing is periodic with a period of 1 sol the model's response will exhibit this period. However, when compared with the Fourier spectrum of the original time series of, say, PC3, we will notice that the response at low frequencies is not well reproduced. It is necessary to include at least EOF5 in order to obtain a better approximation to the Fourier spectrum of the corresponding original time series. We have used the lowest pair capable of representing travelling waves as an example. However, the same sort of response is obtained when higher order pairs are considered. These results are consistent with the work of Crommelin and Majda [29], who found that increasing the number of modes in a reduced model does not necessarily imply a more accurate approximation.

The numerical experiments using the 14-EOF model described in Section 7.4 brought in further information about the way in which different modes interact with

each other. In particular, these experiments gave further indications about the role that the diurnal cycle plays in the development and evolution of baroclinic waves during winter in the northern hemisphere on Mars. Previous work using the MGCM [27, 91] pointed towards the hypothesis that the baroclinic activity on Mars was organised by two coexisting equilibrium states related to two different kinds of baroclinic waves. These two equilibrium states would be destabilised by. the presence of the diurnal cycle. In this thesis, the POD of the MGCM dataset showed that there were two important kinds of baroclinic waves embedded within the original dataset. Moreover, the experiments with the 14-EOF model showed that the interaction between these baroclinic waves and the diurnal (and possibly the semidiurnal) tide leads to an irregular behaviour which would not be present otherwise.

These results put the present work in the context of certain terrestrial models where the existence of multiple equilibrium states has also been hypothesised. In fact, Charney and DeVore [25] where the first to put forwards such a hypothesis. It led to the Charney-DeVore model [25] which was used precisely with the purpose of searching for multiple steady or oscillatory equilibrium states. As mention in Section 1.2.1, this is a heuristically derived low-order model consisting of a severe truncation of a spectral barotropic channel model. Their results were convincing; however, the lack of realism was still a drawback.

EOF models have also been used in research related to this hypothesis [4]. By using a 30-EOF model based on primitive equations, Achatz and Opsteegh [4] were able to reproduce the climate statistics of the terrestrial GCM used as a reference and source of data. However, they were unable to identify the hypothetical steady states responsible for organising the climate dynamics in such a way. Extrapolating the results in Section 8.1.3 to the infinite dimensional case of the terrestrial atmosphere, we would have expected more rather than less fixed points in the reduced models. The conclusion would be in agreement with that in [3] in the questioning to the quasi-stationarity hypothesis.

In the case of Mars, on the other hand, the hypothesis of two coexisting oscillatory equilibrium states seems to have more observational and numerical evidence supporting it. In that sense, we had the advantage of an apparently simpler system since it is mainly organised by only two kinds of motion, and clear suggestions coming from observations and previous studies with the MGCM [27] pointing towards the diurnal cycle as the source of instability.

Despite the good results obtained, the dynamical models thus derived must be considered in their right dimension as models at the starting point of development.

Their predictive power is in agreement with the highly idealised conditions under which they operate. For example, the dissipative and radiative forcing used here consisted only of the rather simple Rayleigh friction and Newtonian cooling parameterisation schemes. Furthermore, they do not include, so far, any means of removing spurious energy accumulation at high frequencies, which cause the models to fail at long-term simulations. We should also consider that, despite being solved at a fairly high resolution (T31 with 25 vertical levels in this case), it still consists of a truncated spherical harmonic series. Hence, the MGCM is already an approximate, albeit accurate, representation of the dynamical behaviour of the atmosphere of Mars. Therefore, the models constructed here are reductions of an already reduced system.

A different question, more theoretical than practical, was raised in Chapter 8, where the problem of the independent representability of different components was introduced. It is true that, as Holmes *et al.* [44] have pointed out, a dynamical system does not necessarily wander around the whole phase space. However, it would be useful to make sure the dynamical system we are concerned with does not tend to visit the particular regions not covered by the span of the EOFs. The reduced models presented here for both low- and high-dimensional models seem to indicate that this might be the case making research in that direction worthwhile pursuing.

As a last remark, let us just comment that a significant difference between typical spectral models and those reliant on three-dimensional EOFs is that usually the spectral discretisation is done only horizontally. Trials with a discretisation of the vertical coordinates have not been as successful as the latter and there has been no further research in that direction. On the other hand, with the POD of the three-dimensional fields, we effectively obtain a set of functions that represents in three dimensions the structure of the fields. Hence, the resulting Galerkin models are, in effect, spectral in the three dimensions. Moreover, representing the horizontal velocity components and a thermal variable by a single set of EOFs would mean, effectively, that we are in the presence of a multi-variate-spectral model in three-dimensions.

9.3 Future work

9.3.1 Further basic investigation

Given the limitations pointed out in Section 9.2, the next step in future work to be suggested is easily identified. In the search for a long-standing reduced model of Martian atmospheric dynamics, the removal of spurious energy at high frequencies

becomes important. Possible alternatives to the solution of this problem have already been investigated [3, 59]. These alternatives could be Newtonian damping [3] or enhanced horizontal diffusion over anomalies [59]. The latter is analogous to the scheme generally employed in both finite-difference and spectral GCMs [36].

Another strong assumption made in the development of the models presented in this thesis was that of the time-independence of surface pressure. This is an open question which requires further attention within the interest of developing long-term models. A time-independent surface pressure is an assumption that can be regarded as valid only during a certain season or limited period of time as the pressure on Mars varies greatly from one time of the year to another. As suggested in Section 3.3.1, one possible way out of this problem is to consider the contribution of the carbon dioxide mass to the total energy since the sublimation and condensation of this atmospheric component is the main cause for the variation in surface pressure.

Forcing reduced models by means of realistic physical parameterisations is an important topic for both short- and long-term simulations as it has a direct impact on the predictive power of the model. It is, nevertheless, not an easy question. Physical parameterisation schemes constitute a significant sink of computational resources in state-of-the-art MGCMs [91] and, as pointed out in Section 7.1, this would be against the spirit of a reduced-order model. However, keeping just the highly simplified parameterisations used during this study would keep rather limited the predictive capacity and, thus, the usefulness of these models. Therefore, a compromise between complexity and simplicity must be sought.

As explained in Section 9.2, the theoretical question of the independent representability of components might be a worthwhile branch to be investigated. This research might also shed light onto questions related to the interaction between atmospheric fields. An example of this was already found when proving the impossibility of decoupling the bases for the horizontal components of velocity in the presence of a time-independent pressure (cf. Section 8.1.2). The problem of the independent representability of components might also be possible to connect it to other theoretical problems such as the topological, and geometrical, structure of the atmospheric attractor and the existence of the slow manifold.

Despite the vast physical knowledge already available, modelling is still a trial and error process, evidence that it is not only physics what matters for achieving a successful model. An example are the most advanced general circulation models available to date. In spite of the great amount of physics included in these models, there is still intense research towards the best way of parameterising these physical

processes in order to improve the predictive power of these models.

As we have just seen, reduced-order modelling is not excluded from this trial and error road to development. Reduced-order modelling is still young compared to other techniques such as finite differences or spectral methods. However, it is a promising path that should be further investigated.

9.3.2 Potential applications

Apart from the theoretical aspects that can be studied to reach a better understanding of the models themselves, a different sort of future work that can be scheduled is related to actual use of reduced models for scientific applications.

Analysis of modal interactions and meteorological phenomena

We have seen, that the POD alone can be a powerful tool to analyse the distribution of total energy over the atmosphere. This property can be exploited in the analysis of different datasets at different times during the year, in order to study inter-seasonal variability, or at the same time for different years, in order to study inter-annual variability. We have started, for example, the analysis of a planet-encircling dust storm on Mars by means of the POD with promising, though preliminary results.

We have also shown that the construction of low-order models can become a significant source of information regarding interactions between modes. The studies started with the results presented in Chapter refsec:mgcmmodels can be carried out forward to investigate, for example, the open questions regarding the nature of the transitions between baroclinic waves with no diurnal cycle. Considering that the seasonal variability in Mars is larger than on the Earth, these analyses can be extended to include during different seasons as well.

Data assimilation

Data assimilation is a field of great potential applications. It is concerned with the determination of the best initial conditions possible in order to achieve the best forecast and with the search for the best fit of a model's output to a given set of observations. Generally, data assimilation schemes usually require the integration of the model and its adjoint forwards and backwards in time, respectively. These operations usually involve the manipulation of enormous datasets with sizes of the order of magnitude of the original state vector (i.e. $\mathcal{O}(10^6)$) even employing simplified models. The compression that a reduced dynamical model could provide is likely to

be very useful in this context. Research in this direction has been started within idealised cases in oceanic studies [23, 95].

Other areas of potential application

As pointed out in the introduction, not only atmospheric sciences face the problem of high complexity in terms of the number of degrees of freedom to be manipulated. From this point of view, engineering sciences are also potential costumers benefiting from the use of low-order models. Hence, the development of reduced atmospheric models should, in principle, lead to the improvement of reduced models in other fields. Evidently, the reverse statement should also be true. One area that has shown a considerable interest in the potentialities of reduced-order models is that of control engineering, which deals with increasingly complex systems, arising from increasingly complex devices, that required to be solved in as short the time as possible.

Bibliography

[1] U. Achatz. Private communication. 2006.

[2] U. Achatz and G. Branstator. A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *J. Atmos. Sci.*, **56**:3140–3160, 1999.

[3] U. Achatz and J.D. Opsteegh. Primitive-equation-based low-order models with seasonal cycle. Part I: Model construction. *J. Atmos. Sci.*, **60**(3):465–477, 2003.

[4] U. Achatz and J.D. Opsteegh. Primitive-equation-based low-order models with seasonal cycle. Part II: Application to complexity and nonlinearity of large-scale atmospheric dynamics. *J. Atmos. Sci.*, **60**(3):478–490, 2003.

[5] U. Achatz and G. Schmitz. On the closure problem in the reduction of complex atmospheric models by PIPs and EOFs: A comparison for the case of a two-layer model with zonally symmetric forcing. *J. Atmos. Sci.*, **54**:2452–2474, 1997.

[6] A. Adcroft, J.M. Campin, C. Hill, and J. Marshall. Implementation of an atmosphere-ocean general circulation model on the expanded spherical cube. *Mon. Weather Rev.*, **132**(12):2845–2863, 2004.

[7] M. Allison and R. Schmunk. Technical notes on Mars solar time as adopted by the Mars24 Sunclock. Technical report, Goddard Institute for Space Studies, 2005.

[8] A. C. Antoulas. *Approximation of large-scale dynamical systems.* Number 6 in Advances in design and control. SIAM, 2005.

[9] A. Arakawa and V.R. Lamb. Computational design of the basic dynamical processes of the UCLA General Circulation Model. *Methods Comput. Phys.*, **17**:173–265, 1977.

[10] A. Arakawa and V.R. Lamb. A potential enstrophy and energy conserving scheme for the shallow water equations. **109**:18–36, 1981.

[11] G.B. Arfken and H.J. Weber. *Mathematical methods for physicists*. Academic Press, 4th edition, 1995.

[12] N. Aubry. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, **192**:115–173, 1988.

[13] N. Aubry. On the hidden beauty of the proper orthogonal decomposition. *Theoret. Comput. Fluid Dynamics*, **2**:339–352, 1991.

[14] J.R Barnes. Time spectral analysis of midlatitudes disturbances in the martian atmosphere. *J. Atmos Sci.*, **37**(9):2002–2015, 1980.

[15] J.R Barnes. Midlatitude disturbances in the martian atmosphere: A second mars year. *J. Atmos Sci.*, **38**(2):225–234, 1981.

[16] J.R. Barnes, J.B. Pollack, R.M. Haberle, and C.B. Leovy. Mars atmospheric dynamics as simulated by the NASA Ames general circulation model. *J. Geophys. Res.*, **98**(E2):3125–3148, 1993.

[17] J. Becker and D. Kuropka. Topic-based vector space model. In W. Abramowicz and G. Klein, editors, *Business information systems, Proceedings of BIS 2003*, pages 7–12, Colorado Springs, USA, 2003.

[18] G. Berkooz, P. Holmes, and J.L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.*, **25**:539–575, 1993.

[19] J.L. Borges. On rigor in science. In *A universal history of infamy*. Penguin Books, 1975. A translation of Historia universal de la infamia, Buenos Aires, 1954.

[20] W. Bourke. An efficient, one-level, primitive equation spectral model. *Mon. Weather Rev.*, **100**(9):683–689, 1972.

[21] W. Bourke. A multi-level spectral model. I. Formulation and hemispheric integrations. *Mon. Weather Rev.*, **102**:687–701, 1974.

[22] B.A. Cantor. MOC observations of the 2001 Mars planet-enrcircling dust storm. *Icarus*, **186**:60–96, 2007.

[23] Y. Cao, J. Zhu, I.M. Navon, and Z. Luo. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids*, **53**:1571–1583, 2006.

[24] J.G. Charney. The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.*, **4**:135–163, 1947.

[25] J.G. Charney and J.G. DeVore. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**:1205–1216, 1979.

[26] M. Collins and I.N. James. Regular baroclinic transient waves in a simplified global circulation model of the Martian atmosphere. *J. Geophys. Res. Planet*, **100**(7):14421–14432, 1995.

[27] M. Collins, S.R. Lewis, P.L. Read, and F. Hourdin. Baroclinic wave transitions in the martian atmosphere. *Icarus*, **120**:344–357, 1996.

[28] P. Comon. Independent component analysis: a new concept? *Signal processing*, **36**:287–314, 1994.

[29] D.T. Crommelin and A.J. Majda. Strategies for model reduction: comparing different optimal bases. *J. Atmos. Sci.*, **61**:2206–2217, 2004.

[30] D.T. Crommelin, J.D. Opsteegh, and F. Verhulst. A mechanism for atmospheric regime behavior. *J. Atmos. Sci.*, **61**:1406–1419, 2004.

[31] Michelangeli D.V, R.W. Zurek, and L.S. Elson. Barotropic instability of midlatitude zonal jets on Mars, Earth and Venus. *J. Atmos. Sci.*, **44**(15):2031–2041, 1987.

[32] E.T. Eady. Long waves and cyclone waves. *Tellus*, **1**(3):33–52, 1949.

[33] P.N. Edwards. A brief history of atmospheric general circulation modelling. In D.A. Randall, editor, *General circulation development, past presente and future: the proceedings of a symposium in honor of Akio Arakawa*, pages 67–90, New York, 2000. Academic Press.

[34] S.B. Fels, J.D. Mahlam, M.D. Schwarzkopf, and R.W. Sinclair. Stratospheric sensitivity to perturbations in ozone and carbon dioxide: radiative and dynamical response. *J. Atmos. Sci.*, **37**:2265–2297, 1980.

[35] J.M. Forbes. Tides in the middle and upper atmospheres of Mars and Venus. *Advances in Space Research*, **33**:125–131, 2004.

[36] F. Forget, F. Hourdin, R. Fournier, C. Hourdin, O. Talagrand, M. Collins, S.R. Lewis, and P.L. Read. Improved general circulation models of the martian atmosphere from the surface to above 80 km. *J. Geophys. Res. Planet*, **104**(10):24155–24176, 1999.

[37] J.B. Freund and T. Colonius. POD analysis of sound generation by a turbulent jet. *AIAA paper 2002-0072*, 2002.

[38] A.E. Gill. *Atmosphere-ocean dynamics*. Number 30 in International Geophysical Series. Academic Press, 1982.

[39] C.T. Gordon and W.F. Stern. A description of the GFDL global spectral model. *Mon. Weather Rev.*, **110**(7):625–644, 1982.

[40] J.M. Haberle, J.B. Pollack, J.R. Barnes, R.W. Zurek, C.B. Leovy, J.R. Murphy, H. Lee, and J. Schaeffer. Mars atmospheric dynamics as simulated by the NASA Ames General Circulation Model 1. The zonal-mean circulation. *J. Geophys. Res*, **98**(E2):3093–3123, 1993.

[41] G.J. Hakim, D. Keyser, and L.F. Bosart. The Ohio Valley wave-merger cyclogenesis event of 25-26 January 1978. Part II: Diagnosis using quasi-geostrophic potential vorticity inversion. *Mon. Weather Rev.*, **124**:2176–2205, 1996.

[42] K. Hasselmann. PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93**(D9):11015–11021, 1988.

[43] P. Holmes. On Moffat's paradox or can empirical projections approach turbulence? In J.L. Lumley, editor, *Whither turbulence? Turbulence at the crossroads*, pages 306–309, New York, 1990. Springer-Verlag.

[44] P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems, and symmetry*. Cambridge monographs on mechanics. Cambridge University Press, 1996.

[45] J.R. Holton. *An introduction to dynamic meteorology*, **88** of *International Geophysics Series*. Elsevier Academic Press, 4th edition, 2004.

[46] C. Homescu, L.R. Petzold, and R. Serban. Error estimation for reduced-order models of dynamical systems. *SIAM J. Numer Anal.*, **43**(4):1693–1714, 2005.

[47] B.J. Hoskins and A.J. Simmons. A multilayer spectral model and the semi-implicit method. *Quart. J. R. Met. Soc.*, **101**:637–655, 1975.

[48] F. Hourdin, P. Le Van, F. Forget, and O. Talagrand. Meteorological variability and the annual surface pressure cycle on Mars. *J. Atmos. Sci.*, **50**(21):3625–3640, 1993.

[49] D.R. Johnson and A. Arakawa. On the scientific contributions and insight of Professor Yale Mintz. *J. Climate*, **9**:3211–3224, 1996.

[50] M. Joshi, S.R. Lewis, P.L. Read, and D.C. Catling. Western boundary currents in the atmosphere of mars. *Nature*, **367**, 1994.

[51] L.W. Kamp, F.W. Taylor, and S.B. Calcutt. Structure of Venus's atmosphere from modelling of night-side infrared spectra. *Nature*, **336**:360–362, 1988.

[52] K. Karhunen. Zur spektral Theorie stochastischer Prozesse. *Ann. Acad. Sci. Fenicae*, **A1**:34, 1946.

[53] A. Kasahara and W. Washington. NCAR global general circulation model of the atmosphere. *Mon. Weather Rev.*, **95**(7):389–402, 1967.

[54] N. Keely and A. White. Remarks on the theory and practice of potential vorticity inversion. Technical Report 439, Meteorological Office, 2004.

[55] Y. Kurihara and Jr. J.L. Holloway. Numerical integration of a nine-level global primitive equations model formulated by the box method. *Mon. Weather Rev.*, **95**(8):509–530, 1967.

[56] F. Kwasniok. The reduction of complex dynamical systems using principal interaction patterns. *Physica D*, **92**:28–60, 1996.

[57] F. Kwasniok. Optimal galerkin approximations of partial differential equations using principal interaction patterns. *Phys. Rev. E*, **55**(5):5365–5375, 1997.

[58] F. Kwasniok. Empirical low-order models of barotropic flow. *J. Atmos. Sci.*, **61**:235–245, 2004.

[59] F. Kwasniok. Reduced atmospheric models using dynamically motivated basis functions. *J. Atmos. Sci.*, 2007.

[60] J. Laskar and P. Robutel. The chaotic obliquity of the planets. *Nature*, **361**:608–612, 1993.

[61] C. Leovy and Y. Mintz. Numerical simulations of the atmospheric circulation and climate of mars. *J. Atmos. Sci.*, **26**(6):1167–1190, 1969.

[62] S.R. Lewis. Modelling the Martian atmosphere. *Astronomy and Geophysics*, **44**(4):4.06–4.14, 2003.

[63] S.R Lewis and P.R. Barker. Atmospheric tides in a Mars general circulation model with data assimilation. *Advances in Space Research*, **36**:2162–2168, 2005.

[64] S.R. Lewis, M. Collins, P.L. Read, F. Forget, F. Hourdin, R. Fournier, C. Hourdin, O. Talagrand, and J.P. Hout. A climate database for Mars. *J. Geophys. Res.*, **104**(E10):24177–24194, 1999.

[65] M. Loève. Fonctions aleatoire de second ordre. *C. R. Acad. Sci. Paris*, 1945.

[66] E.N. Lorenz. Available potential energy and the maintenance of the general circulation. *Tellus*, **7**(2):157–167, 1955.

[67] E.N. Lorenz. Energy and numerical weather prediction. *Tellus*, **12**(4):364–373, 1960.

[68] E.N. Lorenz. Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**:130–141, 1963.

[69] A.F. Lovegrove, I.M. Moroz, and P.L. Read. Bifurcations and instabilities in rotating two-layer fluids: I. f-plane. *Nonlinear Proc. Geophys.*, **8**:21–36, 2001.

[70] J.L. Lumley. The structure of inhomogeneous turbulent flows. In A.M. Yaglom and V.I. Takerski, editors, *Atmospheric turbulence and radio wave propagation*, pages 166–178, Moscow, 1967. Nauka.

[71] J.L. Lumley. *Stochastic tools in turbulence*. Number 12 in Applied mathematics and mechanics. Academic Press, 1970.

[72] P. Lynch. Richardson's marvellous forecast. In M.A. Shapiro and S Grønås, editors, *The life cycles of extratropical cyclones*, pages 61–73, Boston, 1999. Amer. Met. Soc.

[73] J. Mallet-Paret and G.R. Sell. Inertial manifolds for reaction diffusion equations in higher space dimensions. *J. Am. Math. Soc.*, **1**(4):805–866, 1988.

[74] O. Martínez-Alvarado, I.M. Moroz, P.L. Read, and S.R. Lewis. Reduced order models of the Martian atmospheric dynamics. In *Proceedings of the 5th EUROMECH Nonlinear Dynamics Conference*. EUROMECH, 2005.

[75] A. Messiah. *Quantum mechanics*. Dover, New York, 1999.

[76] H.K. Moffat. Fixed points of turbulent dynamical systems and supression of nonlinearity. In J.L. Lumley, editor, *Whither turbulence? Turbulence at the crossroads*, pages 250–257, New York, 1990. Springer-Verlag.

[77] L. Montabone, S.R. Lewis, and P.L. Read. Interannual variability of Martian dust storms in assimilation of several years of Mars Global Surveyor observations. *Advances in Space Research*, **36**:2146–2155, 2005.

[78] L. Montabone, S.R. Lewis, P.L. Read, and D.P. Hinson. Validation of Martian meteorological data assimilation for MGS/TES using radio occultation measurements. *Icarus*, **185**:113–132, 2006.

[79] L. Montabone, O. Martinez-Alvarado, S.R. Lewis, P.L. Read, and M.D. Smith. Meteorology of the 2001 global dust storm on Mars in an assimilation of Thermal Emission Spectrometer data from Mars Global Surveyor. In *Seventh International Conference on Mars*, Pasadena, California, July 2007.

[80] K.W. Morton and D.F. Mayers. *Numerical solution of partial differential equations*. Cambridge University Press, 1994.

[81] C.E. Newman, S.R. Lewis, P.L. Read, and F. Forget. Modeling the Martian dust cycle 1. representations of dust transport processes. *J. Geophys. Res.*, **107**(E12, 5123), 2002.

[82] C.E. Newman, S.R. Lewis, P.L. Read, and F. Forget. Modeling the Martian dust cycle 2. multiannual radiatively active dust transport. *J. Geophys. Res.*, **107**(E12, 5124), 2002.

[83] C.E. Newman, P.L. Read, and S.R. Lewis. Investigating atmospheric predictability on Mars using breeding vectors in a general circulation model. *Q. J. R.Meteorol. Soc.*, **128**:1–999, 2002.

[84] G.R. North. Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, **41**(5):879–887, 1984.

[85] G.R. North, T.L. Bell, and R.F. Cahalan. Sampling errors in the estimation of empirical orthgonal functions. *Mon. Weather Rev.*, **110**:699–706, 1982.

[86] S.A. Orszag. Transform method for the calculation of vector-coupled sums: application to the spectral form of the vorticity equation. *J. Atmos. Sci.*, **27**:890–895, 1970.

[87] J. Pedlosky. *Geophysical fluid dynamics*. Springer-Verlag, 2nd edition, 1987.

[88] N.A. Phillips. A coordinate system having some special advantages for numerical forecasting. *J. Atmos. Sci.*, **14**:184–185, 1957.

[89] J.B. Pollac, C. B. Leovy, P.W. Greiman, and Y. Mintz. A Martian general circulation experiment with large topography. *J. Atmos. Sci.*, **38**(1):3–29, 1981.

[90] M. Rathinam and L.R. Petzold. A new look at proper orthogonal decomposition. *SIAM J. Numer. Anal.*, **41**(5):1893–1925, 2003.

[91] P. L. Read and S.R. Lewis. *The Martian climate revisited*. Springer-Praxis books in geophysical sciences. Springer-Praxis, 2004.

[92] L.F. Richardson. *Weather prediction by numerical process*. Cambridge University Press, Cambridge, 1922. Reprinted Dover, New York, 1965.

[93] M. I. Richardson, C. E. Newman, and A. D. Toigo. PlanetWRF - A flexible, multi-scale model for planetary atmospheres. In *Bulletin of the American Astronomical Society*, **38** of *Bulletin of the American Astronomical Society*, pages 625–+, September 2006.

[94] M.L. Richardson and R.J. Wilson. A topographically forced asymmetry in the Martian circulation and climate. *Nature*, **416**:298–301, 2002.

[95] C. Robert, S. Durbiano, E. Blayo, J. Verron, J. Blum, and F.-X Le Dimet. A reduced-order strategy for 4D-Var data assimilation. *J. Marine Systems*, **57**:70–82, 2005.

[96] S. Roberts and R. Everson. *Independent component analysis. Principles and practice*. Cambridge University Press, 2001.

[97] C.W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *Int. J. on Bifurcation and Chaos*, **15**(3):997–1013, 2005.

[98] C.W. Rowley, T. Colonius, and R.M. Murray. Dynamical models for control of cavity oscillations. *AIAA paper 2001–2126*, 2001.

[99] C.W. Rowley, T. Colonius, and R.M. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Phys. D*, **189**:115–129, 2004.

[100] R Sadourny. The dynamics of finite-difference models of the shallow-water equation. *J. Atmos. Sci.*, **32**:680–689, 1975.

[101] W.E. Sangster. A method of representing the horizontal pressure force without reduction of station pressures to sea level. *J. Meteor.*, **17**:166–176, 1960.

[102] S.D. Schubert. A statistical-dynamical study of empirically determined modes of atmospheric variability. *J. Atmos. Sci.*, **42**(1):3–17, 1985.

[103] S.D. Schubert. The structure, energetics and evolution of the dominant frequency-dependent three-dimensional atmospheric modes. *J. Atmos. Sci.*, **43**(12):1210–1237, 1986.

[104] F.M Selten. Toward an optimal description of atmospheric flow. *J. Atmos. Sci.*, **50**(6):861–877, 1993.

[105] F.M Selten. An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.*, **52**(7):915–936, 1995.

[106] F.M Selten. Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.*, **54**:2099–2114, 1997.

[107] F.M Selten. A statistical closure of a low-order barotropic model. *J. Atmos. Sci.*, **54**:1085–1093, 1997.

[108] A.J. Simmons and D.M. Burridge. An energy and angular-momentum conservig vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Weather Rev.*, **109**:758–766, 1981.

[109] A.J. Simmons, D.M. Burridge, M. Jarraud, C. Girard, and W. Wergen. The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.*, **40**:28–60, 1989.

[110] L. Sirovich. Turbulence and the dynamics of coherent structures. Part I: Coherent structures. *Quart. Appl. Math.*, **45**(3):561–571, 1987.

[111] G.I. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames –I. Derivation of basic equations. *Acta Astronautica*, **4**:1177–1206, 1977.

[112] A. Slingo, R.C. Wilderspin, and S.J. Brentnall. Simulation of the diurnal cycle of outgoing longwave radiation with an atmospheric GCM. *Mon. Weather Rev.*, **115**:1451–1457, 1987.

[113] D.E. Smith, M.T. Zuber, S.C. Solomon, R.J. Phillips, J.W. Head, J.B. Garvin, W.B. Banerdt, D.O. Muhleman, G.H. Pettengill, G.A. Neumann, F.G. Lemoine, J.B. Abshire, O. Aharonson, C.D. Brown, S.A. Hauck, A.B. Ivanov, P.J. McGovern, H.J. Zwally, and T.C. Duxbury. The global topography of Mars and implications for surface evolution. *Science*, **284**(5419):1495–1503, 1999.

[114] M.D. Smith, B.J. Conrath, J.C. Pearl, and P.R. Christensen. Thermal Emission Spectrometer observations of Martian planet-encircling dust storm 2001a. *Icarus*, **157**:259–263, 2002.

[115] T.R. Smith, J. Moehlis, and P. Holmes. Low-dimensional modelling of turbulence using the proper orthogonal decomposition: a tutorial. *Nonlinear Dyn.*, **41**:275–307, 2005.

[116] J. Thuburn. *Modelling of large-scale unstable waves in the middle atmosphere.* DPhil thesis, University of Oxford, 1988.

[117] E.S. Titi. On approximate inertial manifolds to the Navier-Stokes equations. *J. Math. Anal. Appl.*, **149**:540–557, 1990.

[118] A. Toigo, M.I. Richardson, and C.E. Newman. MarsWRF: a general purpose, local to global numerical model for the martian climate and atmosphere. In *Seventh International Conference on Mars*, Pasadena, California, July 2007.

[119] A.D. Toigo, M.I. Richardson, R.J. Wilson, H. Wang, and A.P. Ingersoll. A first look at dust lifting and dust storms near the south pole of Mars with a mesoscale model. *J. Geophys. Res.*, **107**(E7, 10.1029/2001JE001592), 2002.

[120] Z. Toth and E. Kalnay. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Astron. Soc.*, **74**(12):2317–2330, 1993.

[121] F. Waleffe. Transition in shear flows. Nonlinear normality versus non-normal linearity. *Phys. Fluids*, **7**(12):3060–3066, 1995.

[122] A.A. White. Modified quasi-geostrophic equations using geometric height as vertical coordinate. *Q. J. R. Met. Soc.*, **103**(437):383–396, 1977.

[123] A.A. White. Private communication. 2005.

[124] S.G. Whitehouse, S.R. Lewis, I.M. Moroz, and P.L. Read. A simplified model of the Martian atmosphere. Part 1: a diagnostic analysis. *Nonlinear Proc. Geophys.*, **12**:603–623, 2005.

[125] S.G. Whitehouse, S.R. Lewis, I.M. Moroz, and P.L. Read. A simplified model of the Martian atmosphere. Part 2: a POD-Galerkin analysis. *Nonlinear Proc. Geophys.*, **12**:625–642, 2005.

[126] R.J. Wilson. A general circulation model simulation of the Martian polar warming. *Geophys. Res. Lett.*, **24**(2):123–126, 1997.

[127] R.J. Wilson and K. Hamilton. Comprehensive model simulation of thermal tides in the Martian atmosphere. *J. Atmos. Sci.*, **53**(9):1290–1326, 1996.

[128] R.J. Wilson and M.I. Richardson. The Martian atmosphere during the Viking mission, I Infrared measurements of atmospheric temperatures revisited. *Icarus*, **145**:555–579, 2000.

[129] C.F. Yoder. Astrometric and geodetic properties of Earth and the Solar System. In T.J. Ahrens, editor, *Global Earth Physics*, **1**, pages 1–31, 1995.

[130] W. Zdunkowski and A. Bott. *Dynamics of the atmosphere. A course in theoretical meteorology*. Cambridge University Press, 2003.

[131] R.W. Zurek. Diurnal tide in the martian atmosphere. *J. Atmos. Sci.*, **33**:321–337, 1976.

[132] R.W. Zurek. Inference of dust opacities for the 1977 Martian great dust storms from Viking Lander 1 pressure data. *Icarus*, **48**:202–215, 1981.