POD-Galerkin Modelling of the Martian Atmosphere



Stephen Gerald Whitehouse

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 1999

POD-Galerkin Modelling of the Martian Atmosphere

Stephen Gerald Whitehouse, Pembroke College
Submitted for the degree of Doctor of Philosophy
Hilary Term, 1999

The aim of this thesis is to seek a low-dimensional description of baroclinic instability in general, and of the Martian atmosphere in particular, where both forcing and spatial resonance are relevant to the dynamics of the system being analysed.

The Proper Orthogonal Decomposition (POD) is used to determine a basis for the modal decomposition of climatic simulations of Mars, obtained by using two General Circulation Models (GCMs): (a) a simple GCM, which is an idealised model in which the meteorological primitive equations are solved on a sphere with simplified physical parameters and (b) the Martian GCM, a more realistic model in which a comprehensive range of the relevant Martian physical parameters and topography are represented. Results of these analyses are presented for a range of Martian seasons and climatic conditions. The effects of using different forms of energy norm in performing the analysis is considered, with the objective of providing analyses which represents the physically most significant components of the circulation, with optimal efficiency.

Reduced low-dimensional models that replicated the full simple GCM streamfunction simulations are formulated by projecting the spherical quasi-geostrophic equations onto the PODs of the large-scale calculations. The resulting models are analysed by using a combination of solution continuation and numerical integration methods. A thorough analysis of the models reveals that a 6-D POD model is capable of reproducing the amplitude, frequency and behaviour of the leading oscillatory structures of the simple GCM, to within a 1% error. Such an excellent reproduction of the original system is shown to be due to (1) an accurate vertical formulation scheme, (2) the use of the correct norm, (3) a sufficiently high level of truncation and (4) the fact that the original system is a steady wave flow.

The behaviour of the various regimes observed in the low-order models are comparable with observations from studies of large-scale waves and instabilities in planetary atmospheres, including a range of hydrodynamical experiments on baroclinic wave interactions of a stratified fluid in cylindrical containers.

Contents

1 Introduction **1**

1.1 The Planet Mars . . . 1
1.2 History of the POD . . . 4
1.3 EOF modelling of the atmosphere . . . 5
1.4 Proper Orthogonal Decomposition . . . 9
1.5 Fundamentals of POD . . . 10
1.5.1 An optimum model . . . 11
1.6 The governing equations . . . 12
1.6.1 Navier-Stokes on a sphere . . . 12
1.6.2 Selecting the governing equations . . . 12
1.7 Martian atmospheric behaviour and QG theory . . . 13
1.7.1 Advantages of the QG model . . . 14
1.8 Thesis aims . . . 14
1.9 Structure of thesis . . . 16

2 Vertical Structures **18**

2.1 Introduction . . . 18
2.2 The QG Vertical Structure equation . . . 18
2.2.1 Defining the Froude number . . . 20
2.2.2 Energy definitions . . . 20
2.2.3 Vertical modes . . . 21
2.3 The SGCM data . . . 24
2.3.1 Decomposition . . . 24
2.3.2 The numerical problem . . . 25
2.3.3 Calculation of the stratification parameter . . . 26
2.3.4 Eigenvalues and eigenmodes . . . 27
2.3.5 Energy . . . 27
2.4 The MGCM data . . . 30
2.4.1 Eigenvalues and eigenmodes . . . 31
2.4.2 Energy . . . 31
2.5 Summary and conclusions . . . 33

3 Mars **34**

3.1 Aim . . . 34
3.2 Martian atmospheric dynamics . . . 34

3.3 The SGCM . . . 36
3.3.1 The SGCM basic state . . . 39
3.3.2 The spectral grid . . . 41
3.3.3 Calculation of PODs in spectral space . . . 41
3.3.4 Constructing continuous vertical modes . . . 43
3.3.5 Projection of streamfunction or vorticity. . . . 43
3.3.6 The KE and TE norms . . . 45
3.3.7 Formation of the spectral energy matrix . . . 46
3.3.8 Symmetrising the eigenvector problem . . . 47
3.3.9 SGCM Eigenspectrum . . . 47
3.3.10 SGCM POD modes . . . 51
3.3.11 Power Spectra Analysis . . . 62
3.3.12 SGCM Principal Components and Frequency Spectra . . . 63
3.3.13 Energy in the SGCM eigenmodes . . . 70
3.4 The MGCM . . . 72
3.4.1 A winter run . . . 72
3.4.2 MGCM eigenspectrum . . . 74
3.4.3 MGCM POD modes . . . 77
3.4.4 MGCM Principal Components and Frequency Spectra . . . 86
3.4.5 Energy in the MGCM eigenmodes . . . 92
3.5 Conclusions . . . 94

4 POD-Galerkin Projection **96**
4.1 Spectral Galerkin methods . . . 96
4.2 Derivation of the Galerkin spherical QG model . . . 97
4.2.1 Defining the basic state . . . 99
4.2.2 PODs or spherical harmonics ? . . . 99
4.2.3 Boundary conditions . . . 100
4.2.4 Calculating the vertical derivative term . . . 100
4.2.5 Rescaling the spherical QG equations . . . 102
4.2.6 The Galerkin projection of the Martian equations . . . 102
4.2.7 Notation for inner products . . . 104
4.2.8 The matrix problem . . . 104
4.2.9 Forcing the low-dimensional model . . . 106
4.3 Overview of the general computational scheme . . . 107
4.3.1 Mak's 1985 low-dimensional model . . . 108
4.3.2 Fitting the POD modes to spherical harmonics . . . 110
4.3.3 Calculating the derivative fields . . . 111
4.3.4 Calculating the Interaction coefficients . . . 112
4.3.5 Test example: verification of computational scheme . . . 113
4.3.6 Analysing the Martian ODE set . . . 115
4.4 Conclusions . . . 115

5 2-v SGCM Galerkin models **117**
5.1 Background to baroclinic wave studies . . . 118
5.2 The advantages of using the total energy norm . . . 122
5.3 Linear Stability Method . . . 124
5.3.1 Earlier Results . . . 125
5.3.2 2-v linear stability results . . . 126
5.4 Robustness of the numerical nonlinear models . . . 130
5.5 Outline of modelling aims . . . 131
5.6 Assessment of bifurcation structures . . . 133
5.6.1 Comparisons between the models . . . 134
5.6.2 Model truncation . . . 136
5.6.3 Stability properties of the periodic orbits . . . 137
5.7 Single wave regimes . . . 139
5.7.1 Numerical integrations . . . 140
5.7.2 Power Spectra Results . . . 140
5.7.3 The system's attractor . . . 144
5.7.4 Phase space and attractor reconstructions . . . 145
5.7.5 Phase portraits . . . 146
5.8 Galerkin model comparisons . . . 148
5.8.1 20-D models . . . 148
5.8.2 10-D Galerkin models . . . 155
5.8.3 6-D models . . . 157
5.9 The mixed wave regime . . . 157
5.9.1 Phase Portraits . . . 157
5.9.2 Power spectra analysis . . . 159
5.10 Conclusions . . . 163

6 Extended 4-v Galerkin models **164**
6.1 Linear Stability method . . . 164
6.1.1 4-v linear stability results . . . 165
6.2 Assessment of bifurcation structures . . . 168
6.2.1 Comparisons between the 2-v and 4-v AUTO runs . . . 169
6.2.2 Stability properties of the periodic orbits . . . 169
6.3 Galerkin model comparisons . . . 170
6.3.1 20-D models . . . 170
6.3.2 10-D models . . . 175
6.3.3 6-D models . . . 177
6.4 The mixed wave regime . . . 177
6.4.1 Power spectra results . . . 177
6.4.2 Phase Portraits . . . 180
6.4.3 Transient behaviour . . . 180
6.4.4 The attractor . . . 183
6.4.5 Success in reproducing quasi-periodic behaviour . . . 184
6.5 The dynamical role of the double Hopf bifurcation . . . 186

6.5.1 A superposition of two waves 187
6.5.2 Unfolding the double Hopf bifurcation 188
6.5.3 The SGCM bifurcation set 191
6.6 Conclusions 194

7 Summary and Conclusions **196**

A Variational Principle for the POD **209**

B Floquet Theory **211**

Symbols and Acronyms

a	Radius of planet
a_i	i-th Galerkin wave amplitude coefficient
APE	Available potential energy
AV	Amplitude vacillation
b_i	i-th Galerkin mean-flow correction amplitude coefficient
BOD	Biorthogonal Decomposition
bc	Baroclinic component
β	A measure of the planetary vorticity gradient
bt	Barotropic component
CORR	Correlation
CO_2	Carbon dioxide
$\det(\mathbf{C})$	Determinant of matrix $\mathbf{C}$
∇_h^2	Horizontal Laplacian
DHB	Double Hopf bifurcation point
ENS	Enstrophy
EOF	Empirical Orthogonal Function
f	Coriolis parameter
f_0	Synoptic scale of motion for Coriolis parameter
F	Froude number
F_c	Critical Froude number
4-v	4 vertical modes retained in the modelling scheme
FFT	Fast Fourier Transform
g	Gravitational force
GCM	General Circulation Model
Γ	Dissipation parameter
HB	Hopf bifurcation point
h_i	Vertical pressure length scale for i-th mode
H_i	i-th vertical mode
IC	Initial condition
J	Jacobian operator
$\mathbf{J}$	Jacobian matrix
KE	Kinetic energy
KL	Karhunen-Loève
λ_i	i-th eigenvalue
L_s	Celestial orbit of the sun in Mars centred coordinates
m	Zonal wavenumber
$\mathbf{M}$	Energy matrix
MFC	Mean-flow correction
Mixed mode	Superimposed principal and subsidiary waves
MGCM	Martian General Circulation Model
μ	Sine (latitude)

n	Total wavenumber
N^2	Brunt-Väisälä frequency
NH	Northern hemisphere
NN	Highest total wavenumber retained in spectral series
ODE	Ordinary Differential Equation
Ω	Rotation rate
p	Pressure
p_*	Surface pressure
$P_{mn}(\mu)$	Legendre polynomial of the first kind (of degree l and order m)
PC	Principal Component
PDE	Partial Differential Equation
PE	Potential energy
Φ_i	i-th POD mode
ϕ	Longitudinal coordinate
PIP	Principal Interaction Pattern
ψ	Streamfunction
$\hat{\psi}(\theta)$	Streamfunction basic state component
$\bar{\psi}(\theta, t)$	Streamfunction mean-flow correction component
$\tilde{\psi}(\theta, \phi, t)$	Streamfunction wave component
$\bar{\psi}_m$	Amplitude of m-th baroclinic mode
POD	Proper Orthogonal Decomposition
Principal wave	Waves which arise from Hopf bifurcation HB_1
q	Quasi Geostrophic Potential Vorticity
QG	Quasi Geostrophic
QGPV	Quasi Geostrophic Potential Vorticity
r	Ekman dissipation parameter
R	Gas constant for dry air
$\mathbf{R}$	Correlation matrix
Rd	Rossby radius of deformation
ρ	Density
RHS	Right hand side
RMS	Root mean square
S	Stratification parameter
SC	Point of stability exchange
SGCM	Simple General Circulation Model
SH	Southern hemisphere
σ	Density in isentropic coordinates
sol	Martian day
Subcritical Hopf	Hopf which produces unstable periodic orbits
Subsidiary wave	Waves which arise from Hopf bifurcation HB_2
Supercritical Hopf	Hopf which produces stable periodic orbits
SV	Structural vacillation
TE	Total energy
T	Temperature
T21	Triangular truncation at total wavenumber 21
τ_D	Radiative relaxation timescale
τ_E	Frictional dissipation timescale
θ	Latitudinal coordinate
θ_T	Horizontal equilibrium of potential temperature
2-v	2 vertical modes retained in the modelling scheme
$\mathbf{V}$	State vector
V_{mni}	(m,n)-th position on the spectral grid of the i-th POD mode
z	Vertical coordinate
ζ	Absolute vorticity

Chapter 1

Introduction

1.1 The Planet Mars

Mars is the fourth planet from the Sun and is commonly referred to as the Red Planet, since the rocks, soil and sky have a red or pink tone. The distinctive red colour has been observed by astronomers throughout history. The planet was given its name by the Romans in honour of their god of war, while similar names were given by other civilisations, e.g. the ancient Egyptians named the planet Her Descher meaning the red one.

Before space exploration, extraterrestrial life was thought to inhabit Mars. Star observers thought they saw straight lines which crossed the planet's surface which led to the popular belief that irrigation canals had been constructed by intelligent beings. This fascination with the red planet culminated in 1938, when Orson Welles broadcasted a radio drama based on the science fiction classic War of the Worlds by H.G. Wells, where panic ensued as many believed that Martians were invading Earth.

Science for many years has been enchanted by the thought that life, at one time, existed on Mars. These beliefs were somewhat substantiated when early scientists observed apparent seasonal colour changes on the planet's surface. This phenomenon led to speculation that conditions might support a Martian vegetation during the warmer months and cause plant life to become dormant during colder periods. There is evidence that in the past a denser Martian atmosphere may have allowed water to flow on the planet. Physical features closely resembling shorelines, gorges, riverbeds and islands suggest that great rivers once marked the planet.

The Martian atmosphere

Mars and the Earth are classified as 'terrestrial planets', and are relatively small and dense, possessing rocky crusts and mantles. Their atmospheres contain mainly the heavier gases such as carbon dioxide and nitrogen[1]. Significantly, Mars is remarkably similar to Earth in many aspects. Notably, the inclination of the Martian polar axis to its orbital plane (see Table 3.1) suggests that Mars has seasons like Earth, and since both are rapidly rotating planets, rotating about the same speed, then the Coriolis forces will be of a similar magnitude. Such a similarity

[1]These contrast with the much larger 'gas giant' planets, Jupiter, Saturn, Uranus and Neptune, which are less dense and have no obvious solid surface (see James (1994)).

creates an immediate interest, especially since there is evidence (see below) that large-scale motions of the Martian atmosphere can be modelled by low-dimensional models.

The atmosphere of Mars is quite different from that of Earth. It is composed primarily of carbon dioxide with small amounts of other gases, the six most common components being Carbon Dioxide (CO_2) 95.32%, Nitrogen (N_2) 2.7%, Argon (Ar) 1.6%, Oxygen (O_2) 0.13%, Water (H_2O): 0.03%, Neon (Ne) 0.00025%. Martian air contains only about 1/1,000 as much water as our air, but even this small amount can condense out, and is capable of forming clouds that rise high in the atmosphere or swirl around the slopes of towering volcanoes.

Recent spacecraft observations, those made by Viking, have provided evidence for the existence of transient baroclinic eddies in the atmosphere of Mars. Such waves have been thought to exist since the work of Hess (1950), and their occurrence was confirmed by the work of Barnes (1980, 1981). These waves originate as a result of baroclinic instability of the type investigated by Charney (1947) and Eady (1949). In 1976 NASA landed two Viking landers on the surface of Mars, in the Northern Hemisphere, which measured wind direction and speed, temperature and pressure for many years. Barnes showed that measurements of pressure, temperature, wind speed and wind direction (measured from Viking Lander 2 situated at 48^oN, 226^oW on the planet) showed oscillations with phase relationships which were consistent with those found in terrestrial baroclinic waves, and those predicted by earlier analytic work by Leovy (1969) and Gadian (1978), as well as numerical studies by Leovy and Mintz (1969). The Viking lander observations also showed that the midlatitudes of Mars contains cyclones and anticyclones which are similar to those observed in the midlatitudes of Earth and are a result of baroclinic instability in the midlatitude temperature gradient.

Barnes (1981) showed that the oscillations of meteorological variables (in the frequency range relevant for Martian baroclinic waves) were often regular and repeatable, and that, by assuming the waves were geostrophic and sinusoidal, found that the flow was dominated by wavenumbers 3 and 4 which had phase speeds between 15-16 ms^{-1}. Barnes et al. (1993) studied the baroclinic wave activity (using the NASA Ames Mars GCM) and found that zonal wavenumbers 1-4 with periods 2-10 days existed, and that such oscillations were regular in time. It has since been shown that the large amplitude seasonal pressure cycle is due mainly to the exchange of CO_2 between the surface and the atmosphere, as well as changes in the large scale circulation patterns (see Hourdin et al. (1993)). In contrast, the shorter time scale oscillations are due to baroclinic waves (Barnes (1980, 1981)) and thermal tides (Zurek (1988)).

Although it is not known why Martian baroclinic eddies exhibit such regular behaviour, analysing the nonlinear behaviour of baroclinic waves on Mars might aid us in forming a better understanding of the nature of nonlinear baroclinic waves on Earth.

Derivation of a low-dimensional Martian atmospheric model

Our aim is to seek a low-dimensional description of a baroclinic atmosphere, with particular reference to baroclinic instability in general, where both forcing and spatial resonance are relevant to the dynamics of the system being analysed. It is important to choose a system

which has these dynamical features, and is capable of being reduced to a low-order system, but can still retain the original system's qualitative behaviour. An appropriate model has been developed by Collins and James (1995). Here, a simple General Circulation Model (SGCM) is used (an idealised model in which the primitive equations are solved on a sphere), which uses simplified parameterisations of heating and friction, and the parameter range is such that the waves which are found are comparable with those of an idealised Martian atmosphere.

Since the SGCM data is thought to exhibit low-dimensional behaviour (Collins and James (1995)), it is our aim to produce a Galerkin model which is capable of reproducing realistic dynamical flows. It is our aim to derive low-dimensional models using an optimal set of basis functions obtained from observations of the flow structures in the SGCM streamfunction data. These low-dimensional models will have a small number of degrees of freedom and so contain a small number of modes, and it is hoped that they will be able to describe systems (in some sense) of a higher dimension.

Reducing the system to its essential degrees of freedom may provide useful insight into the underlying physical processes which occur in the original SGCM. Hopefully, these models will reveal the dominant interactions which occur between the numerous components, and their contributions to atmospheric variability on various timescales. Dynamical systems theory[2] will be used to investigate the qualitative behaviour of the reduced models. The steady solutions of the governing equations will be of special importance as the stable states are those which will be observed in the original SGCM. Bifurcation theory will monitor the solutions of the nonlinear systems, and a thorough examination of the stability properties will be made.

Baroclinic instability

Most of the instabilities in meteorology are associated with wave propagation. The wave instabilities which are important for synoptic-scale meteorology are generally related to zonally asymmetric perturbations to a zonally symmetric basic flow field (Holton (1979)). Baroclinic instability is associated with the dynamics of atmospheres and oceans, and results in many cases from a rotating stratified fluid with vertically sheared basic currents which implies horizontal temperature gradients (Gill (1982)). The presence of horizontal temperature gradients implies the existence of available potential energy (APE) in the basic flow, and this is the energy source for baroclinic instability.

A rapidly rotating fluid with horizontal and vertical temperature gradients, such as the atmosphere, has APE which can be converted into kinetic energy (KE) in the form of a disturbance. For the APE to be an energy source for an instability, whilst releasing the energy the fluid must go through a process of rearranging itself into a state of minimum energy so that the isopycnals (lines of constant density) are as close to the horizontal as possible. The PE which is trapped in the sloping density surfaces can be released in the form of fluid motions (a release of KE) known as baroclinic waves (the motion is often referred to as 'sloping convection').

[2]Lorenz (1963a) showed that dynamical systems theory could be applied in modelling the 'state' of the atmosphere, defined to be an infinite dimensional phase-space in which the values of the state variables give the coordinates at each point of physical space.

Such waves are important in the atmosphere for transporting heat and angular momentum, being particularly evident in the midlatitudes of the Earth and Mars where observed cyclones, anticyclones and frontal systems are examples of large scale baroclinic instability.

The term baroclinic describes either a flow with vertical structure or a flow that transforms PE into KE (but were shown by Gill (1982) to be functionally identical). Evidence of transient baroclinic waves in the Martian atmosphere goes right back to the work done by Hess (1950). Much theoretical work has been done since on baroclinic instability, and a thorough discussion of the mathematical theory is found, for example, in Pedlosky (1979). Recent experimental as well as theoretical work is reviewed by Klein (1990). (A brief review is found in § 5.1).

1.2 History of the POD

The Proper Orthogonal Decomposition (POD) method is used to determine a basis for modal decomposition from an ensemble of signals. Such a method was introduced by Lumley (1967, 1981) for turbulent flows, but has its origin in the Karhunen-Loève (KL) decomposition demonstrated by Loève (1963). This technique yields a decomposition of the data that is optimal, in a well-defined sense (maximises the representation of the variance) and is a classical statistical technique for analysing spatiotemporal phenomena. Given an ensemble of patterns, the technique yields an orthogonal basis for the representation of the ensemble, as well as a measure of the relative contribution of each basis function to the total 'energy' (mean square fluctuation) of the ensemble. This basis is optimal in the sense that a truncated series representation of the data in it has a smaller mean square error than a representation by any other basis of the same dimension. These properties make the KL basis a natural one to consider when performing model reduction or data analysis (see Graham and Kevrekidis (1996)).

An early use of the KL analysis for the study of a spatiotemporal system was due to Lorenz (1956), who advocated its use in meteorology. Lorenz (1963a) introduced the idea of Empirical Orthogonal Functions[3] (EOFs) to the field of meteorology where the argument was used that the atmosphere continuously generates 'coherent' structures which are better represented by EOFs than spherical harmonics (see also Selten (1995a)). Lumley (1970) proposed that the decomposition would yield 'coherent structures' that dominate a turbulent flow. Coherent structures are also often found in laboratory experiments (see Cantwell (1981) for a review), and an overview of the theory of EOFs and its application to meteorology and oceanography is presented by Preisendorfer (1988). However, the method in its original formulation was often difficult to deal with numerically until Sirovich (1987) proposed the 'method of snapshots' to reduce the amount of computation necessary for the analysis of large data sets.

In the past several years, some of the applications of this technique have included construction of low-dimensional models of turbulent boundary layer flows (Aubry et al. (1988,1991)), time-dependent flows in complex geometries (Deane et al. (1991)), data analysis of turbulent Rayleigh-Bènard convection simulations (Sirovich and Park (1990), Park and Sirovich (1990)),

[3]EOF analysis is where experimental time series is used to construct an orthonormal basis. It has been used extensively as a tool in analysing both observations and numerical simulations of the global atmosphere.

simulations of the Ginzburg Landau equations (Sirovich (1989)), and temperature patterns in heterogeneous catalytic reaction-diffusion experiments (Graham et al. (1993a, b)), Chen and Chang (1993)).

The review of Berkooz et al. (1993) contains a presentation of the mathematical properties of the POD method, as well as a discussion of applications to turbulent flows. It is shown with mathematical rigour how to generate an approximate dynamical system by implementing Galerkin projections based upon POD modes, and the conditions for the preservation of symmetry by linear and nonlinear Galerkin projections. The importance of symmetry preservation is emphasised by King and Stewart (1992), who argued that not only is it important to reconstruct the attractor of the original system, but also to preserve its symmetries. The effect of solution symmetries on the POD procedure as well as some issues of long-term dynamics of truncated POD models is discussed by Aubry et al. (1993). Berkooz and Titi (1993) examined Galerkin projections of certain partial differential equations which retained the symmetries of the original nontruncated equations and concluded that it is important that the approximate dynamical systems retain the symmetries of the original system if it is to reflect the correct qualitative behaviour.

The POD method has also been used for modal reduction, in conjunction with an artificial neural network, to construct empirical dynamical systems from experimental data in an isothermal heterogeneous catalytic reactor-diffusion system (Krischer et al. (1993)). Broomhead and King (1986) have proposed a modification of the technique for cases in which only a scalar time series is available. In addition, arguments have been made for analysing the time derivative of the spatiotemporal field (Sirovich et al. (1990)), or for analysing combinations of the field and spatial derivatives of various orders (Kirby (1992)). The former has been proposed as an optimal basis for an 'inertial manifold' of a system, while the latter may be useful when a basis that efficiently captures gradients is desired.

An excellent discussion of the POD method and the related Bi-orthogonal decomposition (BOD) is made by Aubry (1991), who explained that BOD is in a sense the symmetric version of the POD, and thus permits the simultaneous investigation of the spatial and temporal evolution of the flow. She also discusses how it provides a framework for a spatiotemporal nonlinear dynamics tool and a means to study fluid behaviour. The advantage of such a decomposition is that it decomposes the flow into orthogonal spatial modes and orthogonal temporal modes, so that the temporal and spatial phase spaces are disjoint, and where there is a one-to-one correspondence between both spaces. This leads to a spatiotemporal theory of nonlinear dynamics.

1.3 EOF modelling of the atmosphere

Flows observed in the atmosphere and other hydrodynamical systems reveal a hierarchy of structures which differ in complexity, ranging from laminar to turbulent motions. Despite the intricate natures of these flows there have been many attempts to describe the behaviour by nonlinear models which only depend upon a few variables (such as Charney and DeVore

(1979)). The hypothesis which has long been tested is that the basic structures of observed flows depend only upon a finite (and possibly small) number of parameters that control it, i.e. there only exists a few degrees of freedom as it evolves in phase space (low dimensional).

There have been recent analyses of hydrodynamic experiments, made to obtain dimensionality estimates from observations. There is evidence that a relatively small number of degrees of freedom characterises observed turbulent flows. Guckenheimer and Buzyna (1983) analysed the transition to geostrophic turbulence in a rotating annulus and revealed a low dimensionality for the vacillation regime turning into a turbulent one. Brandstäter et al. (1983) presented empirical evidence for turbulent flows (which evolved in a low dimensional subspace of the whole space) when analysing modulated wavy vortices (Couette-Taylor flow).

Sellers (1957) was the first to formulate a barotropic model based on atmospheric EOFs. He constructed an empirical spectral model using 9 EOFs of 500 mb of height anomaly, from a 64 point grid covering Northern America, and reduced thus barotropic vorticity equation to a set of 9 coupled, nonlinear ordinary differential equations (ODEs). However, comparisons with the standard grid-point model showed the forecast skill of the EOF model to be poorer. Sellers pointed to difficulties in the grid point evaluation of the higher-order derivatives as a possible reason for the poor results and recommended that using spherical harmonics gave a better model for this problem.

Models of the atmosphere have been formulated in terms of EOFs by Rinne and Karhila (1975), where a spectral barotropic model given in horizontal EOFs is described. They produced a similar model to Sellers (1957), but included 57 EOFs over a (30×36) grid in the Northern Hemisphere. It was shown that the forecast skill was at least as good as the standard finite difference model. However, the model exhibited spurious oscillations in the energy invariant which were thought to be due to the EOF truncation level.

Schubert (1985) observed the wintertime interseasonal variability of the Earth's Northern Hemisphere circulation. EOFs were calculated from 10-day averaged 500-mb anomalies of the Northern Hemisphere winter. Here Schubert used a barotropic model which was projected onto a number of EOFs (of the streamfunction anomalies) which led to a system of quadratically nonlinear equations involving the EOF coefficients. The projection was made onto a small number of EOFs where all the EOF calculations were done in the wave domain. He stressed that this was a different procedure to the usual modelling of barotropic dynamics as the spectral modes (usually spherical harmonics) were replaced by functions which had been determined empirically from the data. The advantage of implementing such a method was that it allowed the explicit monitoring of arbitrary spatial modes with a dynamical framework. The first 40 EOFs accounted for 97.2 % of the variance and were substituted into the barotropic model. It was concluded that the EOF barotropic model produced results which were as good as a standard barotropic model which had many more components.

Schubert (1986) extended the EOF modelling to baroclinic EOFs which were calculated from observations in different frequency bands. He employed a two-layer quasi-geostrophic (QG)

model to study the energy conversion between various modes of variability over different time scales. Particular difficulties were found in estimating the higher order derivative terms in the model and resulted in significant differences in the results as compared with other similar studies. However, both of Schubert's studies (1985,1986) gave encouraging results for the EOF modelling approach, but in neither of them were the differential equations integrated to determine whether a model (which was based on a relatively small number of EOFs) was capable of producing realistic dynamical behaviour.

Ideas in a similar context have been used by Mundt and Hart (1994) where reduced low-dimensional models were formulated by projecting the governing equations (a standard QG f-plane dynamical model of baroclinic instability) onto the EOFs of the large-scale circulations yielding a set of ODEs for the amplitude of the nonlinear eigenfunctions. They studied the number of EOFs required to reconstruct flows with 99% of the variance retained. The streamfunction was decomposed into the baroclinic and barotropic wavy and zonal mean parts, allowing them to consider what effects certain symmetries had upon the kind of bifurcations which might occur. This, they stressed was an important consideration since the breaking of symmetries correlates with the onset of more complex (possibly chaotic) motion in the flow. It was shown that as the number of active unstable modes increased, the ability of the EOF low-dimensional system to efficiently model the full spatio-temporal dynamics diminished. They concluded by remarking that their EOF-based models behaved well when modelling a transition region from a simple (i.e. steady) flow to a chaotic or weakly-turbulent behaviour. Such models were also found to be capable of extracting mode-mixing behaviour which occurred in quasi-periodic regimes.

Selten (1993) used a two-level QG hemispheric model which had been formulated in spherical harmonics and used EOFs to describe the evolution of the circulation. He argued whether it would be be advantageous to describe the dynamics in terms of EOFs rather than decompose the flow fields into Fourier components (and thus describe the interactions among the structures by many interaction coefficients). It was also shown that EOFs are not uniquely defined but depend for instance on the inner product used in their formulation. A discussion was made on whether the basis should use the total state vector or describe deviations from the mean state (as standard EOFs do). The EOFs calculated used both a kinetic energy (KE) and total energy (TE) norm, and the evolution equations for the amplitudes of the EOFs were derived by the Galerkin projection of the model equations onto the EOF basis. Selten investigated how many EOFs must be retained to describe the attractor of the complete model. It was shown that the global structure of the system can only be described in a truncated TE model since the KE model failed to simulate the energy conversion processes which occurred between the modes (which is fundamental to baroclinic flow processes).

Selten (1995) used a T21[4] spectral barotropic model and showed that it could be accurately modelled by using 20 EOFs, which were computed in spectral space. It was shown that the construction of an EOF model requires particular attention, such as its integral constraints. Two different EOF models were evaluated at a truncation limit of 20, 40 and 80 EOFs. Both

[4]T21 refers to a triangular truncation at total wavenumber 21. For details refer to § 3.3.2.

the ability to produce accurate deterministic forecasts and the ability to simulate the T21 climate was investigated. It was found that the quality monotonically improved when more EOFs were retained. The proposition was made that the mean effect of the neglected interactions is to damp the resolved modes and that it can be adequately described by a linear damping. With the closure assumption it was found that the T21 climatology and variability were well reproduced by the EOF model with 20 EOFs.

Selten (1997) addressed whether EOFs are more efficient in describing large-scale atmospheric dynamics compared to spherical harmonics, and whether using these EOFs improves the forecast skill and the climatology of the model. The experiments were performed using a three-level QG, hemispheric spectral T21 model. He argued that it was important to observe the vertical structures of the EOFs. By calculating the pattern correlation between the different levels it was shown that there existed an equivalent barotropic structure, and so the norm used favoured these structures[5]. He argued that the TE norm would have been used if he had found more pronounced baroclinicity in the leading EOFs.

Selten (1997) concluded that an EOF basis is more efficient in describing large-scale atmospheric dynamics compared to spherical harmonics. This was proved by comparing the predictive skill of EOF truncated models and a T17 version of the T21 model. It was found that with less than half the number of phase space dimensions the EOF truncated model had the same predictive skill as T17[6]. However, he disappointingly concluded that even though his work for a barotropic model (Selten (1995)) showed that the EOF-based models were superior to traditional spectral models, this was not the case for the baroclinic model. The predictions from the EOF-baroclinic model were found to be relatively poor. He argued that the reasons for the poor skill may have been due to the coarse vertical resolution, by using a constant potential vorticity forcing term, the representation of the orographic forcing and the QG assumption.

In Achatz and Schmitz (1997) EOFs and Principal Interaction Patterns[7] (PIPs) were used to test the possibilities of energy-conserving closures[8] in the reduction of multiple baroclinic wave cycles in a two-layer QG model on the sphere. The paper addressed two questions: (1) how to obtain the best closed reduced model for the full system and (2) which is the best method (EOFs or PIPs) for basic patterns in obtaining low-order models which have energy-conserving closure. They showed that both the time and zonal climate mean (and the typical frequency behaviour of the turbulent energy) were much better reproduced with than without closure. As for the reductive capabilities, it was found that a closed model with 17 EOFs (7 zonally symmetric and 10 in wave subspace) was able to simulate the most important aspects of the original time series. It was also shown that the same number of retained PIPs were at least as good as using EOFs (in a closed model) in the simulation of important time and zonal means.

[5]Selten chose weights such that all three levels had the same amount of variance.

[6]However, the EOF truncated model is computationally more expensive since an equivalent of the Fast transform for EOFs does not exist.

[7]PIP analyses look for an optimal subspace of the full state space for representing the dynamics of the system on the basis of using as few degrees of freedom as possible. PIP analysis is a generalisation of the optimisation procedure described by EOFs in that now not only the interaction coefficients but also the reduced subspace (used for the projection) is an object of optimisation.

[8]The parameterisation of the influence of the unresolved modes onto the selected patterns is called 'closure' in this context.

1.4 Proper Orthogonal Decomposition

In the POD-Galerkin scheme, the objective is to compute an orthonormal basis which will minimise, on average, the least squares error between the truncated representation and the true solution to the equations. Given a time series of such a solution $\mathbf{v}(\mathbf{x}, t)$, it can be shown (see Appendix A) that a suitable basis is obtained by calculating the eigenvectors (called the POD modes) of the time averaged autocorrelation function:

$$\int_{\Omega} \mathbf{R}(\mathbf{x}, \mathbf{x}')\phi(\mathbf{x}')d\mathbf{x}' = \lambda\phi(\mathbf{x}), \tag{1.1}$$

where

$$\mathbf{R}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{v}(\mathbf{x}, t)\mathbf{v}(\mathbf{x}', t)\rangle, \tag{1.2}$$

and the angled brackets indicate a time average. As we are considering fields over a sphere, the coordinate vector $\mathbf{x} = (\theta, \phi)$ represents the position on the sphere where θ denotes the latitudinal and ϕ the longitudinal directions. The integral equation (1.1) must be discretised, which gives a matrix eigenvalue problem. The correlation function and the eigenfunction are respectively discretised as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}(\mathbf{x}_1, \mathbf{x}'_1) & \cdots & \mathbf{R}(\mathbf{x}_1, \mathbf{x}'_n) \\ \vdots & & \vdots \\ \mathbf{R}(\mathbf{x}_n, \mathbf{x}'_1) & \cdots & \mathbf{R}(\mathbf{x}_n, \mathbf{x}'_n) \end{pmatrix}, \qquad \mathbf{\Phi}^{(m)} = \begin{pmatrix} \phi_{\mathbf{1}}{}^{(m)}(\mathbf{x}_1) \\ \vdots \\ \phi_{\mathbf{1}}{}^{(m)}(\mathbf{x}_n) \end{pmatrix}, \tag{1.3}$$

where n is the number of spatial points. This reduces the integral equation (1.1) to the matrix equation:

$$\mathbf{R}\mathbf{\Phi}^{(m)} = \lambda_m \mathbf{\Phi}^{(m)}. \tag{1.4}$$

The eigenvalues $\{\lambda_i\}$ are measures of the variance of the POD modes, and the *relative eigenvalue* is defined as $p_k = \frac{\lambda_k}{\sum_{i=1}^{n} \lambda_i}$ which gives the *energy* content of the k-th mode.

Since the Fredholm integral equation in (1.1) is on a sphere, the necessary Fredholm integral equation to be calculated in spherical coordinates is:

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} R(\theta, \phi, \theta', \phi')\Phi(\theta', \phi')r^2 \sin(\theta')d\theta' d\phi' = \lambda\Phi(\theta, \phi). \tag{1.5}$$

As the kernel is not symmetric, we introduce $\Psi(\theta, \phi) = \sqrt{\sin\theta}\ \Phi(\theta, \phi)$, substituting this into (1.5) to obtain

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \left[R(\theta, \phi, \theta', \phi')\sqrt{\sin\theta \sin\theta'}\right] \Psi(\theta', \phi')r^2 d\theta' d\phi' = \lambda\Psi(\theta, \phi), \tag{1.6}$$

where we see from the above that the matrix bracket [ ] is the new symmetrised kernel.

Once we have obtained the eigenvalues and vectors by solving (1.6) we substitute back for the eigenfunction Φ. This method illustrates that the matrix can always be symmetrised so that NAG routines can be used to reduce the matrix to a tridiagonal form. Such routines run in a time proportional to n^3 (where n is the number of spatial points of the grid). Therefore,

the computational time depends upon the resolution required, which often depends upon the problem being considered. The matrix was integrated using the Simpson's and Trapezoidal methods.

1.5 Fundamentals of POD

The POD method is similar to EOF analysis and Berkooz (1991) provides an excellent discussion of POD analysis applicable to problems in turbulence and this is summarised below.

Assume that the fields are scalar, and the functions are complex-valued and square-integrable on an interval Ω of the real line, defined as a family of L^2. Then by definition $(f,g) = \int_\omega f(x)g^*(x)dx$ is the inner product with norm $\|f\| = (f,f)^{\frac{1}{2}}$. An important property of the POD method is that it maximises the (ensemble) mean square of the projection (defined by the L^2 canonical scalar product) of the field onto the eigenfunction $\phi^{(1)}$ and repeats this maximisation process so that it converges optimally fast in the quadratic mean. Therefore, by the calculus of variations the function ϕ must satisfy the condition:

$$\max_{\psi(x)\in L^2(x)} \left\langle \frac{(u,\psi)}{(\psi,\psi)} \right\rangle = \left\langle \frac{(u,\phi)}{(\phi,\phi)} \right\rangle. \tag{1.7}$$

Variance is to be maximised (see Chapter 3 for a discussion on kinetic and total energy optimisation), and so a necessary condition is determined by implementing a variational principle where ϕ is the optimal function to be obtained, giving

$$\frac{d}{d\delta} \frac{< |(u,\phi+\delta\psi)|^2 >}{(\phi+\delta\psi,\phi+\delta\psi)}|_{\delta=0} = 0. \tag{1.8}$$

By differentiating the above function, implementing the condition $\delta = 0$, and then equating the numerator to zero it has been shown that, after rearranging the integrals and using Fubini's Theorem, a sufficient condition is that ϕ is the eigenfunction of the two-point correlation tensor:

$$\int_\Omega < u(x)u^*(x') > \phi(x')dx' = \lambda\phi(x'), \tag{1.9}$$

where

$$\mathbf{R}(x,x') =< u(x)u^*(x') >, \tag{1.10}$$

and <> denotes the averaging operation in time. Since $\mathbf{R}$ is a symmetric kernel, Hilbert-Schmidt theory can be used (provided Ω is bounded) which states that there exists an infinite number of eigenfunctions $\{\phi_k\}$ of $\mathbf{R}$, which can be normalised so that $\phi_k = 1$ for all k. The eigenvalues are ordered so that $\lambda_i \geq \lambda_{i+1}$ where $\lambda_i \geq 0$ due to the non-negative definiteness of the correlation matrix $\mathbf{R}(x,x')$. There also exists a diagonal decomposition of the correlation tensor:

$$\mathbf{R}(x,x') = \sum_k \lambda_k \phi_k(x)\phi_k^*(x'), \tag{1.11}$$

where the largest eigenvalue corresponds to the maximum of (1.9).

It is shown that almost every member (in a measure sense) of the ensemble may be reproduced by a modal expansion in terms of the eigenfunctions:

$$u(x) = \sum_k a_k \phi_k(x), \tag{1.12}$$

where

$$< a_k a_{k'}^* >= \delta_{kk'} \lambda_k. \tag{1.13}$$

Berkooz (1991) shows that if the POD spectrum decays fast enough, then all the likely realisations in the ensemble can be found in a relatively small set of bounded subspace. We note that the definition of the dimension (according to Aubry et al. (1991)) is the number of nonzero eigenvalues in the POD decomposition, but this is also the dimension of the smallest linear subspace which contains the dynamics. Such a definition does not take into account the existence of noise and Broomhead and King (1985) showed mathematically that if white noise (which is uncorrelated with the given field) is added to the time series then all the eigenvalues will be non-zero. However, it is important to determine whether the use of such optimal bases affect the bifurcation structure of the original system and also whether the original pattern (or flow) with all its associated dynamics can be reproduced by using a finite number of the modes.

Although PODs are of a statistical nature they are the optimum set of basis functions for a given truncation, i.e. no other basis set can explain more of the variance averaged over the dimensional space. The eigenvalues define the variance which the eigenvectors contribute to the time evolution.

The immediate advantage with implementing POD is its linearity, but such simplicity can lead to mathematical restriction as there is no requirement that the problem to which the method is applied be linear. However, Penland and Ghil (1993) argued that since PODs are determined by the fully nonlinear dynamics, a POD model automatically includes some of the nonlinear effects.

1.5.1 An optimum model

A property of the Galerkin system which has used orthogonal basis functions is that the total variance of the signal equals the sum of the squared moduli of the POD coefficients (which is a consequence of Parseval's theorem). When the POD method is implemented, Loève (1955) showed that the set of basis functions obtained contain the variance in the lower-order functions. The POD method is being used so that optimum low-order models can be derived. It is hoped that the fields under consideration will have eigenvalues λ_n which decrease **rapidly** with increasing n, resulting in the majority of the variance of the system being represented by a limited number of functions. If the eigenvalues λ_n decrease more slowly, then the variance will be distributed across a much larger set of functions.

1.6 The governing equations

We will now discuss the Navier-Stokes and QG equations on a sphere and highlight the relationship between them. We will also examine the problems associated with both sets of equations. Such a discussion allows the conclusion to be made that (for the purpose of this project) the spherical QG equations were best suited to be used as the governing equations, and the reasons behind this decision will be discussed.

1.6.1 Navier-Stokes on a sphere

Consider (θ, ϕ, z) to be the coordinate axes, where ϕ is longitude, θ latitude, z the vertical distance above the surface of the planet and $\mathbf{U} = (u, v, w)$ to be the relative velocity. The distance to the centre of the planet is r where $r = a + z$ (a is the radius of the planet). Since $z \ll a$ for the regions of the atmosphere which are meteorologically relevant, we take $r = a$ as a first approximation. If the Coriolis force $2\mathbf{\Omega}$ is expressed in components and the friction is taken to be $\mathbf{F}_r = (F_x, F_y, F_z)$, then the equations of motion on a sphere are :

$$\frac{du}{dt} - \frac{uv\tan\theta}{a} + \frac{uw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\Omega v\sin\theta - 2\Omega w\cos\theta + F_x, \tag{1.14}$$

$$\frac{dv}{dt} + \frac{u^2\tan\theta}{a} + \frac{uw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\Omega u\sin\theta + F_y, \tag{1.15}$$

$$\frac{dw}{dt} - \frac{u^2+v^2}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + 2\Omega u\cos\theta + F_z. \tag{1.16}$$

From James (1994), the pressure derivatives in spherical coordinates become $\frac{\partial p}{\partial x} \rightarrow \frac{1}{a\cos\theta}\frac{\partial p}{\partial \phi}$ and $\frac{\partial p}{\partial y} \rightarrow \frac{1}{a}\frac{\partial p}{\partial \theta}$. The terms proportional to $\frac{1}{a}$ on the left hand sides of (1.14) - (1.16) are the curvature terms. In this process r has been replaced by a which means that the Coriolis terms which are proportional to $\cos\theta$ in (1.14) and (1.16) must be neglected if the equations are to satisfy angular momentum conservation. The resulting set of equations are called the *primitive equations*.

1.6.2 Selecting the governing equations

In order to select the set of governing equations which best suits our requirements, we want the equations to describe the flow, but not be excessively computationally intensive.

From the governing physical laws - the equations of momentum, continuity and thermodynamic balance, the primitive equations can be derived which involve absolute vorticity ξ, divergence D, temperature T, the surface pressure p and the geopotential ϕ, suitably nondimensionalised using the length, time, temperature and pressure scales of the motion.

The primitive equations describe the necessary fields on a sphere (as well as the thermal tide), but it was soon realised that the amount of work involved in solving this system would be unreasonable. The NAG routines used in the fortran programs (to determine the eigenvalues and vectors of the autocorrelation matrix) cannot be used because of the excessive size of the correlation matrix. This is due to the many variables such as vorticity, divergence, temperature

and surface pressure, which would all have to be expanded in POD modes, over the various levels and correlated with each other. For the QG equations, the fields can be expressed in terms of the streamfunction which allows the correlation matrix to be of a manageable size. However, this is not possible for the primitive equations as a similar convenient relationship does not exist between the fields.

If the primitive equations are used as the governing equations, then the matrices become too large for simple algorithms, and the retrieval of the eigenvalues and eigenvectors could only be achieved by using sophisticated and expensive algorithms such as an Iterative Lanzcos algorithm[9] (ILA). According to Buizza and Palmer (1995), the ILA was first proposed in 1950 by Lanczos and was intended to compute a few of the eigenvalues and the corresponding eigenvectors of a large symmetric matrix. Golub and Van Loan (1983) explains that the algorithm is advantageous when about 1 % of the spectrum of a large symmetric $n \times n$ matrix is required. The method is applied to large, sparse, symmetric eigenvalue-problems, and the details of its implementation can be found from Buizza et al. (1993, 1995).

Although the highly truncated system which we intend to obtain satisfies this criterion, with such an implementation arises the immediate problem of coding the adjoint operator (so we have the adjoint model). According to Buizza and Palmer (1995) the adjoint operator must be coded into the fortran statements which describe the evolution operator. This coding would be a task in itself, and care would have to be taken in using the algorithm.

Another (simpler) theory involves the formulation of the QG equations on a sphere. We will now briefly discuss QG theory and why the spherical QG equations were chosen as the governing set of equations for the reduced low-order model, paying particular attention to how relevant the QG approximation is to modelling Martian atmospheric behaviour.

1.7 Martian atmospheric behaviour and QG theory

Like Earth, Mars is a rapidly rotating planet. To describe it, the Navier-Stokes equations in a frame of reference rotating with constant angular velocity Ω can be reduced to the primitive equations of dynamical meteorology (as was shown in § 1.6.1). Maximum wind speed on Mars reaches $O \sim 100$ ms^{-1} at a height of 30-50 km above the surface. This suggests that the Rossby number ϵ, given by $\frac{U}{fL}$, where f is the Coriolis parameter and L is a length scale, is less than unity for very large scales > 1000 km. This is important as it suggests the QG approximation holds for atmospheric motions with frequencies $< \Omega$, the angular frequency of the planet (see Andrews et al. (1987)).

It is intended to model large-scale motion (*geostrophic* motion) by characterising the motion by single scales for velocity and length which result in the details of the interaction of the large-scale motion with motions of smaller scales to be ignored. The governing equations used will retain the full Coriolis parameter ($f = 2\Omega \sin\theta$, where Ω is the speed of the rotating planet

[9]This is discussed by Strang (1986) and details can be found in the book by Golub and Van Loan (1983), where it discusses the problems associated with systems which have large numbers of degrees of freedom, resulting in the failure of conventional matrix algorithms to solve the eigenvalue problems.

and θ the latitude of the fluid element in question), and the parameter f is the local component of the planetary vorticity normal to the planet's surface. By including f it allows the consideration of motions on a sphere, where the particular class of motion will be large-scale, essentially geostrophic motion.

If L is the horizontal scale and a the sphere's radius (and L is large enough so that the Rossby number is small), then particular attention can focus on the situation where L is considerably smaller than a. This causes the geostrophic approximation to fail in the equatorial regions where the Coriolis force on the horizontal currents is extremely small. The restriction to small $\frac{L}{a}$ ($\epsilon = \frac{U}{2\Omega L} = O(\frac{L}{a}) \ll 1$) allows considerable simplification of the Navier-Stokes equations on a sphere and permits the derivation of the spherical QG equations on a sphere, which will be used as the governing set of equations for our studies (e.g. see Holton (1979), Gill (1982) or James (1994) for derivations).

1.7.1 Advantages of the QG model

The introduction of the geostrophic and hydrostatic assumptions result in the QG model. Such a model can be used as a numerical forecaster since the dynamical equations have been dramatically simplified, and both the sound and gravity waves have also been filtered out. The principal reason for filtering out sound and gravity waves is that for the unsimplified equations of motion the meteorologically important motions can be lost due to the noise generated by these waves unless care is taken with the initial conditions.

If a hydrostatic condition is assumed then this prevents sound waves propagating in the vertical direction. Nevertheless, even though vertical sound waves cannot occur there is still the possibility of acoustic or 'Lamb' waves propagating in the horizontal directions. These oscillations display the characteristic that their maximum amplitudes occur at the lower boundary, and away from this boundary the amplitude of the oscillations decay. By implementing the condition $w = \frac{dp}{dt} = 0$ at the lower boundary these waves can be eliminated.

Even though acoustic waves have now been removed, there is still the possibility of internal gravity waves occurring. It is necessary that there is a divergent and time varying horizontal velocity field for the existence of such waves and so by neglecting the local rate of change of the horizontal divergence, they can also be removed. As the local derivative $\nabla \cdot \mathbf{U}$ does not explicitly occur in the momentum equation, the equations can be rewritten so that this term naturally occurs (the details are given by Holton (1979)).

The filtered models so far discussed are various approximations of the QG system and have the advantage of being 80-90 % accurate (in terms of order Rossby number) in the midlatitudes.

1.8 Thesis aims

PODs will be used to determine a basis for the modal decomposition of climate simulations of Mars obtained using two General Circulation Models: (a) an idealised model (SGCM) in which

the meteorological primitive equations are solved on a sphere with simplified physical parameters (Collins and James (1995)) and (b) a more realistic model, namely the Oxford Martian GCM (MGCM) in which a comprehensive range of the relevant Martian physical processes and topography are represented (Collins et al. (1996), Hourdin et al. (1993)). The results of these analyses will be presented for a range of Martian seasons and climatic conditions, and discussions will be made of possible physical interpretations of the identified modes. We will also consider the effects of using different forms of energy norm in performing the analysis, with the objective of providing analyses which represent the physically most significant components of the circulation with optimal efficiency.

It should be realised that a similar spectral basis function expansion would previously have been assumed to be a Fourier series, but these are not necessarily optimal in representing the dynamical variables. Fourier expansions, moreover are global, i.e. all the fields interact with one another on all scales. However, the need for optimality is essential, as it is hoped that the dynamical structure of the system can be represented in a low level of truncation, implying that a low-order model is attainable. By such a process it is hoped that we will understand the most important aspects of the atmosphere's circulation and how the physical processes affect the equation sets.

It is then intended to derive low-dimensional systems of ODEs which will describe the qualitative behaviour of the Martian fields by implementing a spectral method (Hussaini and Zang (1987)), where the dynamical variables of the selected governing equation will be expanded in a series of orthogonal basis functions. By projecting the PODs onto the truncated series (by the implementation of the Galerkin method) a set of evolution equations will be produced for the time-varying coefficients of the series, extracting the large-scale motions from the data.

The aim then is to reproduce the time-series of the SGCM's principal components (PCs), using such a minimal equation set, and to determine whether the dynamics of the SGCM is actually low-dimensional. If the system is in fact low-dimensional, then a thorough examination of the ODEs may provide us with an understanding of some of the underlying dynamics of the SGCM. By such an analysis it might be possible to identify the important energy exchanges which occur between the physical modes, besides specifying which modes are influential in determining a particular flow regime.

An understanding of the SGCM is initially required in order that a similar study of the full MGCM can be made. Attempting to derive a low-order model of the MGCM poses numerous problems since it is a far more comprehensive model, which includes modelled physical processes of the Martian atmosphere as well as surface topography. However, investigating the behaviour of a full Martian atmospheric model will be of great interest since Collins et al. (1996) analysed surface pressure data from the Viking Lander mission and found that the dominant four EOFs captured 57 % of the variance of the system, and so suggesting low-dimensional behaviour on Mars. They also observed similar behaviour in the MGCM and demonstrated that for the simulated Viking Lander 2 pressure field the leading four EOFs captured either 50 % or 60 % of the total variance (depending upon whether the diurnal cycle

was included or not). There is then considerable evidence that a low-dimensional EOF model for the MGCM is also attainable.

In summary, it is our aim to study the possibility of using PODs as a basis to obtain a reasonable low-order SGCM, i.e. to determine whether a model based on a relatively small number of POD modes can reproduce the global structure of the atmospheric attractors.

1.9 Structure of thesis

Chapter 2 examines the properties of the vertical structure equation (applicable to any planetary atmosphere) which arises, for example, from the linearised, adiabatic QG potential vorticity equation. It is shown that the vertical equation is an immediate consequence of two contrasting mathematical approaches. The first uses the standard solution method by assuming the streamfunction is separable (and so the horizontal and vertical variables are separated), whereas the alternative approach is based on a variational principle, which has the advantage of providing additional insight into the behaviour of the vertical modes, particularly their energy content.

Chapter 3 introduces the data sets and considers how the POD method is applied to the SGCM and MGCM streamfunction data fields. It will be shown that the PODs are not uniquely defined, but depend (for instance) upon the choice of inner product. Discussions are made between the eigenvectors which describe deviations from the mean state and those which describe the total state vector. The efficiency of the POD algorithm is examined in terms of its representation of the streamfunction fields. Particular attention focuses upon the various structures contained in the PODs and the PCs, and whether any physical interpretations can be made of the identified modes. Considerations are made of the effects of using different forms of energy norm in performing the analysis.

Chapter 4 discusses the general numerical scheme used in the construction of low-order models. It will be shown that reduced, projected and truncated models have been obtained by an approximation of the system's state vector, and by the use of the spherical QG equations (which are taken to be the governing equations of the reduced system), and by the implementation of a numerical Galerkin scheme. Particular attention focuses on the computation of the derivative fields and the interaction coefficients.

Chapters 5 and 6 will consider a variety of models derived from the SGCM data, and whether any of the generated low-dimensional ODE sets can reproduce observed dynamics in the original data. Early work focuses on linear stability analyses. A thorough analysis of the ODEs is achieved by the use of AUTO (Doedel (1981)), which allows the determination of their bifurcation and stability properties. Comparisons are made between the original and reconstructed attractors and it is investigated whether the low-dimensional models are picking out the most important dynamical behaviour of the original system.

Finally, the conclusion considers to what extent the low-order POD models have been successful

in reproducing the dynamics of the SGCM streamfunction flow, and what has been learnt from this exercise. Possible modifications and improvements are considered as well a proposal for further studies.

Chapter 2

Vertical Structures

2.1 Introduction

It is our aim to obtain a set of normal modes which model the vertical activity of the original system. If, in the POD-Galerkin modelling process, a poor vertical resolution scheme is used then important information concerning physical processes and energy exchanges may not be captured in the reduced model.

The separable solution approach was used by Flierl (1978) in an oceanographic context, who showed that linear eigenmodes in a system with flat horizontal boundaries are integrated functions of N^2 (the Brunt-Väisälä frequency) which determine the evolution of the geostrophic streamfunction. In a system with homogeneous boundary conditions, a simple Sturm-Liouville problem has to be solved in order to obtain the normal modes. In practice with such problems, it is typically found that the barotropic[1] and the first baroclinic modes are the most significant (and consistent with the two modes generally represented by two-layer or two-level quasi-geostrophic models).

It is suggested that a clear distinction should be made between kinetic energy (KE) and available potential energy[2] (APE) in selecting vertical modes in an approximate model, especially since the barotropic mode contains only KE.

2.2 The QG Vertical Structure equation

The vertical structure equation which we will now introduce arises from the primitive equations. However, for the purpose of our studies it is convenient to consider the adiabatic, frictionless Quasi-Geostrophic Potential Vorticity (QGPV) equation which, in isobaric coordinates, can be written (see Pedlosky (1979)) as

$$\frac{D_g q}{Dt} = (\frac{\partial}{\partial t} + \mathbf{V}_g \cdot \nabla_h)(\frac{1}{f_0}\nabla_h^2 \Phi + f_0 \frac{\partial}{\partial p}(\frac{1}{S}\frac{\partial \Phi}{\partial p}) + f) = 0, \tag{2.1}$$

[1] The term barotropic means either vertically homogeneous, or a process that converts KE of a background flow directly into KE of a disturbances (Holton (1979)).

[2] According to Pearce (1978) the available potential energy of the atmosphere is defined as the 'property of the atmosphere for which the source is energy generated by the distribution of the heat sources and sinks, ie. their deviations from the mean heating rate, and the sink its conversion to kinetic energy.'

where $\frac{D_g}{Dt} = \frac{\partial}{\partial t} + \mathbf{V}_g \cdot \nabla_h$ is the derivative following the geostrophic flow; q is the QGPV defined by

$$q = (\frac{1}{f_0}\nabla_h^2\Phi + f_0\frac{\partial}{\partial p}(\frac{1}{S}\frac{\partial \Phi}{\partial p}) + f); \tag{2.2}$$

the horizontal Laplacian is defined as

$$\nabla_h^2 = \frac{1}{a^2}\left(\frac{\partial^2}{\partial\theta^2} + \frac{1}{\cos^2\theta}\frac{\partial^2}{\partial\phi^2}\right), \tag{2.3}$$

where a is the radius of the planet, θ is the latitudinal and ϕ the longitudinal direction. The geostrophic component of velocity is $\mathbf{V}_g = \frac{\mathbf{k}\wedge\nabla_h\mathbf{\Phi}}{f}$ where $\mathbf{k}$ is the unit vector in the vertical direction; the isobaric distribution of the geopotential is $\Phi(\theta, \phi, p, t)$; pressure is p; the geostrophic vorticity term is $\xi_g = \nabla_h^2\left(\frac{\Phi}{f_0}\right)$; the Coriolis parameter is $f = 2\Omega\sin\theta$; the stratification parameter is

$$S = \frac{N^2}{g^2\rho^2}, \tag{2.4}$$

where N is the Brunt-Väisälä frequency, g is the gravitational force and ρ is the density; f_0 is a synoptic scale of motion for the Coriolis parameter and Ω is the rotation rate.

Writing the geopotential in separable form[3] as $\Phi = \sum_i \tilde{\Phi}_i(\theta, \phi, t)H_i(p)$ and substituting into equation (2.1), we find that the vertical structure equation to be solved is:

$${f_0}^2\frac{d}{dp}(\frac{1}{S_i}\frac{dH_i}{dp}) + \lambda_i H_i = 0, \tag{2.5}$$

with homogeneous boundary conditions

$$\frac{dH_1}{dp}|_{p_1} = \frac{dH_1}{dp}|_{p_n} = 0, \tag{2.6}$$

where the λ_i's in (2.5) are a discrete countable set of eigenvalues, S_i is the stratification parameter for the i-th mode and p_1, p_n are the pressures at the upper and lower boundaries respectively. The H_i's form a complete orthogonal set and can be orthonormalised by setting

$$\frac{1}{P}\int_{p_1}^{p_n} H_i(p)H_j(p)dp = \delta_{ij}, \tag{2.7}$$

where P is the pressure difference between these two boundaries.

The derivation of equation (2.5) is similar to the equation derived in a different context by Flierl (1978), except that here equation (2.5) is written in isobaric coordinates.

[3]Tanaka (1985) and Tanaka et al. (1986) used a similar separable solution in order that a diagnostic energetic scheme could be developed to investigate the energy flows among three-dimensional normal modes (Hough modes in the horizontal) in the atmosphere. It was concluded that their studies of three-dimensional normal mode energetics was a useful diagnostic tool to supplement the standard spectral energetics analysis of the general circulation.

2.2.1 Defining the Froude number

The Rossby radius of deformation (Rd) is the characteristic length scale of disturbances in the midlatitudes (James (1994)) and can be interpreted as the horizontal length scale over which the height field adjusts whilst approaching geostrophic equilibrium. In the separation of variables which leads to (2.5), $\frac{1}{\sqrt{\lambda_i}}$ is a similar horizontal length scale which is specific to each of the vertical and horizontal modes $\{H_i\}$, so that the relevant Rossby Radius of deformation (Rd_i) for each mode can simply be defined as

$$Rd_i = \frac{1}{\sqrt{\lambda_i}}. \tag{2.8}$$

Since by definition the Froude number[4] (F) satisfies

$$F = \frac{a^2}{Rd^2}, \tag{2.9}$$

it is then immediate from (2.8) that

$$F = \lambda \times a^2. \tag{2.10}$$

This implies that (2.5) provides a spectrum of Froude numbers $\{F_i\}$, where each Froude number (F_i) in turn corresponds to a different vertical scaling. Therefore, $a^2\lambda_0$ in (2.5) represents the Froude number of the barotropic mode, $a^2\lambda_1$ the Froude number of the first baroclinic mode, and so on. The global Froude number for the SGCM is taken to be $F = a^2\lambda_1$ since the vertical length scale of the first baroclinic mode is the depth of the planetary model (for further details see Table 2.1).

2.2.2 Energy definitions

We now derive vertical normal modes which are solutions to the linearised vorticity equation by assuming a separable form for the streamfunction:

$$\psi(\theta, \phi, p, t) = \sum_{l=0}^{\infty} \tilde{\psi}_l(\theta, \phi, t) H_l(p), \tag{2.11}$$

which allows the definition of the total KE of the system on a sphere of radius a to be defined as:

$$\begin{aligned} KE &= \frac{1}{2} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{p_1}^{p_n} (\nabla_h \psi \cdot \nabla_h \psi)\, a^2 \sin\theta \; dp d\phi d\theta \\ &= \frac{1}{2} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \sum_{l=0}^{\infty} \left| \nabla_h (\tilde{\psi}_l) \right|^2 a^2 \sin\theta \; d\phi d\theta, \end{aligned} \tag{2.12}$$

[4]The Froude number is proportional to (inertial force)/(gravitational force) and is used in momentum transfer in general and open channel flow and wave and surface behaviour calculations in particular.

where $\int_{p_1}^{p_n} H_l H_m dp = \delta_{lm}$ is assumed and p_1, p_n are the pressures at the upper and lower boundaries respectively. Taking $H_0(p)$ to be the barotropic mode and $H_k(p)$ the k-th baroclinic mode (where $k \neq 0$), by using the same definition of APE[5] as Selten (1993), the contribution to APE from the k-th baroclinic mode is:

$$\begin{aligned} APE_k &= \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\int_{p_1}^{p_n}\left(\frac{{f_0}^2}{S}\left(\frac{dH_k}{dp}\right)^2\right)_k \tilde{\psi}_k^2 a^2 \sin\theta \; dpd\phi d\theta \\ &= \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\left(\frac{{f_0}^2}{S}\left(\frac{dH_k}{dp}\right)\right)_k \frac{dH_k}{dp}]_{p_1}^{p_n}\tilde{\psi}_k^2 a^2 \sin\theta \; d\phi d\theta \\ &\quad - \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\int_{p_1}^{p_n}\frac{d}{dp}\left(\frac{{f_0}^2}{S}\left(\frac{dH_k}{dp}\right)\right)_k H_k\tilde{\psi}_k^2 a^2 \sin\theta \; dpd\phi d\theta. \end{aligned} \tag{2.13}$$

By assuming $\frac{dH_k}{dp}|_{p_1} = \frac{dH_k}{dp}|_{p_n} = 0$ in (2.13) and by using the identity given in (2.5), we obtain:

$$APE_k = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\int_{p_1}^{p_n}\lambda_k\Big(H_k\tilde{\psi}_k\Big)^2 a^2 \sin\theta \; dpd\phi d\theta, \tag{2.14}$$

where the subscript k refers to the k-th mode.

Pedlosky (1979)

Alternatively, following Pedlosky (1979) the energy of the system ε can be written as:

$$\varepsilon = \frac{\rho}{2}({u_g}^2 + {v_g}^2 + \frac{{\theta^2}_T}{S}), \tag{2.15}$$

where the geostrophic component of velocity is $\mathbf{V}_g$, and so the APE is directly proportional to the departure of the potential temperature surface (θ) from the horizontal equilibrium configuration (θ_T). Thus

$$\frac{APE}{KE} = \frac{{\theta^2}_T}{({u_g}^2 + {v_g}^2)S} = O(\frac{L}{L_D})^2, \tag{2.16}$$

where the static stability profile $S(p)$ can be seen to be a measure of the stratification ($S = (\frac{L_D}{L})$), where L is the horizontal length scale and L_D is the Rossby radius of deformation. We note that when the scale of motion is $L = O(L_D)$ then $APE = O(KE)$, and if

$$\begin{aligned} L \ll L_D \Rightarrow KE \gg APE, \\ L \gg L_D \Rightarrow APE \gg KE, \end{aligned} \tag{2.17}$$

which are consistent with the energy definitions given in equations (2.12) and (2.14).

2.2.3 Vertical modes

The vertical structure equation (2.5) must be solved in order to produce normal modes which determine the evolution of the geopotential Φ. Alternatively, we now show that the equation

[5]The definition used for APE is proportional to the global mean temperature variance which can be partitioned into zonal and eddy components.

can be derived from a variational principle by minimising APE in the vertical direction. As before, we assume there exists a separable form for the geopotential $\Phi = \tilde{\Phi}(\theta, \phi, t)H(p)$ so that the APE integrand is proportional to $\frac{{f_0}^2}{S}(\frac{dH}{dp})^2$, as given in equation (2.13). We may now pose the variational problem to be that of seeking stationary solutions of the APE functional:

$$I[H] = \int_{p_1}^{p_n} \frac{{f_0}^2}{S}\left(\frac{dH}{dp}\right)^2 dp, \tag{2.18}$$

subject to the constraint:

$$J[H] = \int_{p_1}^{p_n} H^2 dp = 1, \tag{2.19}$$

since the vertical modes are normalised, with homogeneous boundary conditions $\frac{dH}{dp}|_{p_1} = \frac{dH}{dp}|_{p_n} = 0$, and so assuming the atmosphere is bounded at some upper boundary height. These functions are used since the original SGCM has an artificial 'sponge' level at the top of the model (in order to reduce the problems of spurious energy reflection from the top boundary) and so the vertical modes are consistent with this condition.

Using the standard Calculus of Variations approach (for example see Mikhlin (1964)) we obtain the Euler-Lagrange equation:

$$f_0^2 \frac{d}{dp}\left(\frac{1}{S}\frac{dH}{dp}\right) = -\gamma H, \tag{2.20}$$

where γ is the Lagrange multiplier to be determined. Note that (2.20) is identical to equation (2.5) (ie. where $\lambda = \gamma$) and is equivalent to

$$\mathcal{L}H = \lambda H, \tag{2.21}$$

where $\mathcal{L} = -\frac{d}{dp}\left[\frac{f_0^2}{S(p)}\frac{d}{dp}\right]$. Since f_0 is independent of p we have an eigenvalue equation for $\mathcal{L}$, where H is the eigenfunction corresponding to the eigenvalue λ.

By defining $F = \frac{f_0^2}{S}(\frac{dH}{dp})^2$ and

$$E = \frac{\partial^2 F}{\partial H'^2} - \lambda\frac{\partial^2 H^2}{\partial H'^2}, \tag{2.22}$$

where $H' = \frac{dH}{dp}$, then $E = \frac{2f^2}{S} > 0$ implies that $I[H]$ in (2.18) has a minimum within the range of integration. This is a mathematically significant result, though it may not be immediately obvious why it is desirable from a practical viewpoint to produce vertical eigenfunctions which minimise APE.

If total energy (TE) in the system is invariant, by minimising APE in effect maximises KE. This is of importance since any buoyancy-driven process arises from the conversion of APE of the basic state into KE of eddy motion. The first eigenmode which is derived from equation (2.20) is the barotropic mode which, by definition, contains no APE but instead captures almost all of the KE of the system. The remaining modes will then tend to contribute progressively less

to the total KE of the system. The additional benefit of using barotropic and baroclinic modes is the use of simple expressions for KE and APE as are defined in equations (2.12) and (2.14).

From Sturm-Liouville theory (for example see Mikhlin (1964)), since $\frac{{f_0}^2}{S} \geq 0$ and $\int_{p_1}^{p_n} (\frac{{f_0}^2}{S})^{-1} dp$ is bounded we have:

$$|H|_e{}^2 = \int_{p_1}^{p_n} \frac{f_0^2}{S} \left(\frac{dH}{dp}\right)^2 dp = \lambda_n, \tag{2.23}$$

where $|H|_e$ is the energy norm of function H. This is identical to the vertical component of the APE norm used in (2.13), where the energy norm $[\ \ ]_e$ satisfies:

$$[H_n, H_m]_e = \int_{p_1}^{p_n} \frac{f_0^2}{S} \left(\frac{dH_n}{dp}\right) \left(\frac{dH_m}{dp}\right) dp = \lambda_n \delta_{nm}, \tag{2.24}$$

illustrating that these functions are orthogonal in the energy norm. The set $\{H_n(p)\}_{n=1}^{\infty}$ is complete both in terms of energy and in the sense of convergence in the mean.

If we now consider the equation $\frac{d}{dp}\frac{{f_0}^2}{S}(\frac{dH}{dp}) = -\lambda H$, then if both sides are multiplied by H and integrated by parts, then

$$\begin{aligned} \int_{p_1}^{p_n} H \frac{d}{dp}\frac{{f_0}^2}{S}(\frac{dH}{dp})dp &= -\lambda \int_{p_1}^{p_n} H^2 dp \\ \Rightarrow H^2 \frac{{f_0}^2}{S}\frac{dH}{dp}|_{p_1}^{p_n} - \int_{p_1}^{p_n} (\frac{dH}{dp})\frac{{f_0}^2}{S}(\frac{dH}{dp})dp &= -\lambda \int_{p_1}^{p_n} H^2 dp, \end{aligned} \tag{2.25}$$

and by using the homogeneous boundary conditions $\frac{dH}{dp}|_{p_1} = 0$, $\frac{dH}{dp}|_{p_n} = 0$, we have

$$\lambda = \frac{\int_{p_1}^{p_n} \frac{{f_0}^2}{S}(\frac{dH}{dp})^2 dp}{\int_{p_1}^{p_n} H^2 dp}, \tag{2.26}$$

so all non-zero λ's are positive. We note that $\lambda = 0$ is satisfied if $H = 1$ (the purely barotropic mode).

Alternatively, as we are solving an atmospheric problem, and as there does not exist a lid at the top of the atmosphere in reality (Flierl's work was based on an oceanic model), other boundary conditions could be used in calculating the normal modes in order that they were consistent with the real atmosphere. This is certainly an important issue since it was shown by Lindzen et al. (1968) that bounded atmospheres reproduce oscillations of the infinite atmosphere with an accuracy which depends upon height. More particularly, they found that the bounded atmosphere models produced spurious free oscillations and concluded that bounded atmospheres do not properly respond to oscillations which propagate vertically.

So we could use functions which are obtained from only using the lower homogeneous boundary condition $\frac{dH}{dp}|_{p_1} = 0$, and not prescribing the upper boundary condition (except that $|H|^2$ remains finite as $z \to \infty$, where z is the vertical coordinate). In order that nontrivial baroclinic

modes can be examined it is necessary to reduce (2.5) to standard form by the transformation:

$$H(p) = \frac{S^{\frac{1}{2}}}{f} G(p), \tag{2.27}$$

then G satisfies:

$$\frac{d^2G}{dp^2} + \left[\frac{\lambda S}{f^2} - S^{\frac{1}{2}} \frac{d^2}{dp^2} \left(\frac{1}{S^{\frac{1}{2}}} \right) \right] G = 0, \tag{2.28}$$

and using the final condition $\frac{dH}{dp} = 0$ at $p = p_1$ the lower boundary condition is:

$$\frac{1}{S} \frac{dG}{dp} + \frac{f}{S^{\frac{1}{2}}} \frac{d}{dp} \left(\frac{S^{\frac{1}{2}}}{f} \right) = 0. \tag{2.29}$$

If S is finite and nonzero at infinity then the condition $|H|^2$ remaining finite as $z \to \infty$ merely requires G to be finite as $z \to \infty$.

However, for the purpose of this thesis, the vertical structure equation (2.5) with homogeneous boundary conditions is used in order that the obtained vertical modes are consistent with the bounded atmospheric model.

2.3 The SGCM data

We now apply equation (2.5) to a simplified GCM (SGCM) so that we can examine how APE is distributed among the modes and whether our mathematical findings aid us in understanding how this is achieved.

In the present example we have used data from a fully nonlinear GCM based on the meteorological primitive equations on a sphere. This model uses simplified parameterisations of the dominant physical processes which occur under conditions analogous to the Martian atmosphere. The simplified parameterisations of heating and friction have been used to examine baroclinic waves in the parameter range appropriate to the Martian atmosphere, during a continual seasonal numerical simulation of the southern hemisphere winter ($L_s = 90^o$)[6]. The model experiment is described in detail by Collins and James (1995) and will be thoroughly discussed in Chapter 3.

2.3.1 Decomposition

By assuming the decomposition of the streamfunction field ψ is in separable form we may calculate $\overline{\psi}_m$ which is the amplitude of baroclinic mode m for each profile (as a function of θ, ϕ or a profile location). Denoting ψ_j as the streamfunction for level j, then

$$\sum_{m=1}^{N} \overline{\psi}_m \mathbf{H}_m(\sigma_j) = \psi_j, \tag{2.30}$$

[6] L_s is the celestial orbit of the sun in Mars centred coordinates. $L_s = 0^\circ$ corresponds to northern spring equinox, $L_s = 90^\circ$ to northern summer solstice, $L_s = 180^\circ$ to northern autumn equinox and $L_s = 270^\circ$ to northern winter solstice.

where $\mathbf{H}_i(\sigma_j)$ is the i-th vertical mode in layer σ_j and N is the number of vertical levels and hence the total number of normal modes[7].

Now $\frac{1}{L}\int_{\sigma_1}^{\sigma_n} \mathbf{H}_i(\sigma)\mathbf{H}_j(\sigma)d\sigma = \delta_{ij}$ (where L is the sigma value difference between the upper and lower boundaries), then

$$\frac{1}{P}\sum_{j=1}^{N} \delta\sigma_j \mathbf{H}_i(\sigma_j)\mathbf{H}_k(\sigma_j) = \delta_{ik}, \tag{2.31}$$

where $P = \sum_{i=1}^{N} \sigma_i$. The difference in σ between two consecutive sigma surfaces is fixed (for details see Table 3.2) so $\delta\sigma_j$ is constant, which gives the following outer product identity

$$\mathbf{H}_k\mathbf{H}_k^T = N\mathbf{I}, \tag{2.32}$$

where $\mathbf{I}$ is the identity matrix, implying that

$$\overline{\psi}_m = \frac{1}{N}\sum_{j=1}^{N} < \mathbf{H}_m(\sigma_j), \psi_j > . \tag{2.33}$$

Such an expression for the amplitude of each of the baroclinic waves allows the calculation of the KE and APE in each of the modes by using equations (2.12) and (2.14).

2.3.2 The numerical problem

The streamfunction ψ was assumed to have a separable decomposition of the form:

$$\psi(\theta, \phi, \sigma, t) = \sum_{k=0}^{N} \tilde{\psi}_k(\theta, \phi, t)\mathbf{H}_k(\sigma), \tag{2.34}$$

where $\tilde{\psi}_k$ is the horizontally varying part, $\mathbf{H}_0(\sigma)$ is the barotropic mode and $\mathbf{H}_k(\sigma)$ the k-th baroclinic mode in order that the data can be decomposed into horizontal and vertical components.

Equation (2.5) is solved numerically in a finite-difference form (given the appropriate profile of the stratification parameter $S(\sigma)$ derived from the SGCM) in the following way:

$$\frac{\partial}{\partial\sigma}\frac{{f_0}^2}{S_i}\frac{\partial}{\partial\sigma}H_i \sim \frac{{f_0}^2}{S_{i+1}(\delta\sigma_{i+1})^2}(H_{i+1} - H_i) - \frac{{f_0}^2}{S_i(\delta\sigma_i)^2}(H_i - H_{i-1}). \tag{2.35}$$

By allowing $\delta\sigma_i = K$ for some constant $K \in \mathbb{R}$, results in the vertical modal problem being solved over equally spaced sigma levels, so:

$$\frac{{f_0}^2}{S_i K^2}H_{i-1} - \frac{{f_0}^2}{K^2}(\frac{1}{S_i} + \frac{1}{S_{i+1}})H_i + \frac{{f_0}^2}{S_{i+1}K^2}H_{i+1} + \lambda H_i = 0. \tag{2.36}$$

This leads to an eigenvalue problem $\mathbf{C}\mathbf{H}_k = \lambda_k\mathbf{H}_k$, where $\mathbf{C}$ is a symmetric, tridiagonal

[7]The spectral primitive equation model has 10 equally spaced σ levels (where σ=pressure/surface pressure). It is therefore convenient to calculate the normal modes on these surfaces. For details refer to Table 3.2.

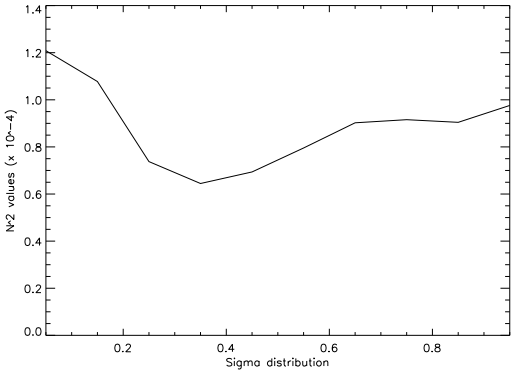
matrix and vector $\mathbf{H}_k$=$(H_k(\sigma_1), ..., H_k(\sigma_N))$, where N was taken here to be 10 (as the SGCM experiments of Collins and James (1995) were carried out using 10 vertical levels equally spaced in σ). A symmetric matrix results from the use of a constant difference in σ between levels and thereby yields orthogonal eigenfunctions. The boundary conditions $\frac{d\mathbf{H}}{d\sigma} = 0$ at $\sigma = \sigma_1$ and $\sigma = \sigma_N$ were chosen (and were implicitly satisfied in (2.35) by specifying S_1=S_{N+1}=∞). The numerical code in (2.36) was verified by setting the stratification parameter $S(\sigma)$ to be constant in the vertical structure equation (2.5), and as expected, the obtained H_i's were sine and cosine functions.

2.3.3 Calculation of the stratification parameter

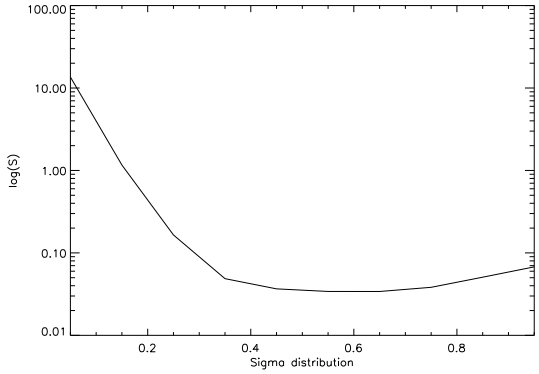
In order to calculate the values of the stratification parameter $S(p)$ in equation (2.36) the N^2 profile (the buoyancy or Brunt-V*äisälä* frequency) was calculated. As the vertical problem is solved in an isobaric coordinate system, the following definition of N^2 is used:

$$N^2 = -\frac{g^2\rho}{\theta_T}(\frac{\partial\theta_T}{\partial p}) = -\frac{g^2 P}{RT\theta_T}(\frac{\partial\theta_T}{\partial p}), \tag{2.37}$$

where g is the gravitational force; ρ is the density; θ_T is the potential temperature surface from equilibrium; p is pressure; T is temperature and R is the gas constant for dry air. This then allows the stratification parameter to be computed[8] by using the identity $S = \frac{N^2}{g^2\rho^2}$.



(a) The Brunt-V*äisälä* frequency (N^2 profile) for the SGCM at latitude 52.61^o in the SH (representing a typical profile). Note that N^2 is almost constant (to $\pm$ 20% or so).



(b) The stratification parameter (S) at latitude 52.61^o in the SH obtained by using $S = N^2/g^2\rho^2$. Note the steep gradient near the top of the model (i.e. for $\sigma < 0.4$).

Figure 2.1: The form of (a) the SGCM N^2 profile and (b) the stratification parameter (S) at latitude 52.61^o in the SH. The N^2 profile is almost constant but the corresponding S profile rapidly increases for $\sigma < 0.4$. The reason for this strong baroclinic structure near the upper limits of the model is that $S \sim O(\frac{1}{\rho^2})$ (as N^2 is practically constant) and between sigma surfaces $\sigma = 0.25$ and 0.05 the vertical height of the model quickly increases from 13.86 km to 29.95 km (from ground level), resulting in a fast decrease in the density (ρ) values.

[8]The following parameters and constants are relevant to the SGCM: $p = p_0 + p'$ is pressure where $p_0 = 610KP_a$ is the basic term and p' is the departure; $T = T_0 + T'$ where T_0=200^oK; θ_T is the potential temperature surface from equilibrium; $R = 191.2JKg^{-1}K^{-1}$ is the gas constant for dry air; $g = 3.72\text{ms}^{-2}$ is the gravitational force; ρ is the density; $\Omega = 7.08822 \times 10^{-5}\text{s}^{-1}$ is the rotational rate and $a = 3.394 \times 10^6$m is the radius of the planet.

The SGCM N^2 profile and the stratification parameter

The N^2 profile at latitude 52.61^o in the southern hemisphere (SH) (representing a typical profile) is shown in Figure 2.1(a) and the corresponding stratification parameter profile in Figure 2.1(b). Near the top of the model (i.e. for $\sigma < 0.4$) the S values rapidly increase, depicting a strong baroclinic structure near the upper limits of the model (an artifact of the sponge layer). Because the streamfunction has a dominant steady wave 3 in the SH latitude band $[40, 60]^o$ (see Figures 3.3 and 3.6), it was decided that the stratification parameter would be averaged over this interval because of the significant atmospheric activity that existed there. Otherwise, if averaging took place over the whole sphere it would be likely that the important vertical structures would be lost. The consequence of such a derivation is that the normal modes are optimal over this chosen band.

2.3.4 Eigenvalues and eigenmodes

The λ_k's can be arranged to form an increasing sequence which are illustrated in Figure 2.2(a) and listed in Table 2.1. For any $k > 0$, $\mathbf{H}_k$ has one more zero in the interval (σ_1, σ_{10}) than does $\mathbf{H}_{k-1}$. This may be seen in Figure 2.2(b) where the purely barotropic mode (a) and the nine baroclinic modes are shown in (b) to (j).

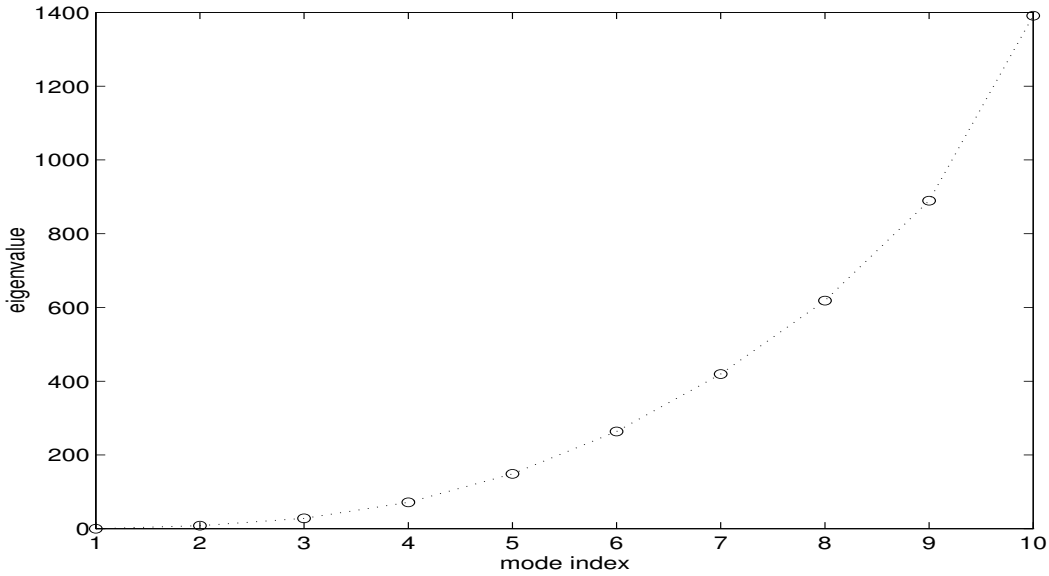
Mode	Eigenvalue	Rossby radius of deformation	Froude number
Barotropic	$\lambda_0 = 0$	$Rd_0 = \infty$	0.0
Baroclinic 1	$\lambda_1 = 7.06 \times 10^{-7}$ km^{-2}	$Rd_1 = 1190.1$ km	8.14
Baroclinic 2	$\lambda_2 = 2.44 \times 10^{-6}$ km^{-2}	$Rd_2 = 640.2$ km	28.1
Baroclinic 3	$\lambda_3 = 6.21 \times 10^{-6}$ km^{-2}	$Rd_3 = 401.3$ km	71.5
Baroclinic 4	$\lambda_4 = 1.29 \times 10^{-5}$ km^{-2}	$Rd_4 = 278.4$ km	148.8
Baroclinic 5	$\lambda_5 = 2.29 \times 10^{-5}$ km^{-2}	$Rd_5 = 208.9$ km	263.7
Baroclinic 6	$\lambda_6 = 3.64 \times 10^{-5}$ km^{-2}	$Rd_6 = 165.7$ km	419.4
Baroclinic 7	$\lambda_7 = 5.37 \times 10^{-5}$ km^{-2}	$Rd_7 = 136.5$ km	618.4
Baroclinic 8	$\lambda_8 = 7.72 \times 10^{-5}$ km^{-2}	$Rd_8 = 113.8$ km	889.4
Baroclinic 9	$\lambda_9 = 1.21 \times 10^{-4}$ km^{-2}	$Rd_9 = 90.9$ km	1390.9

Table 2.1: Table of Froude numbers of the associated vertical eigenmodes for the SGCM case.

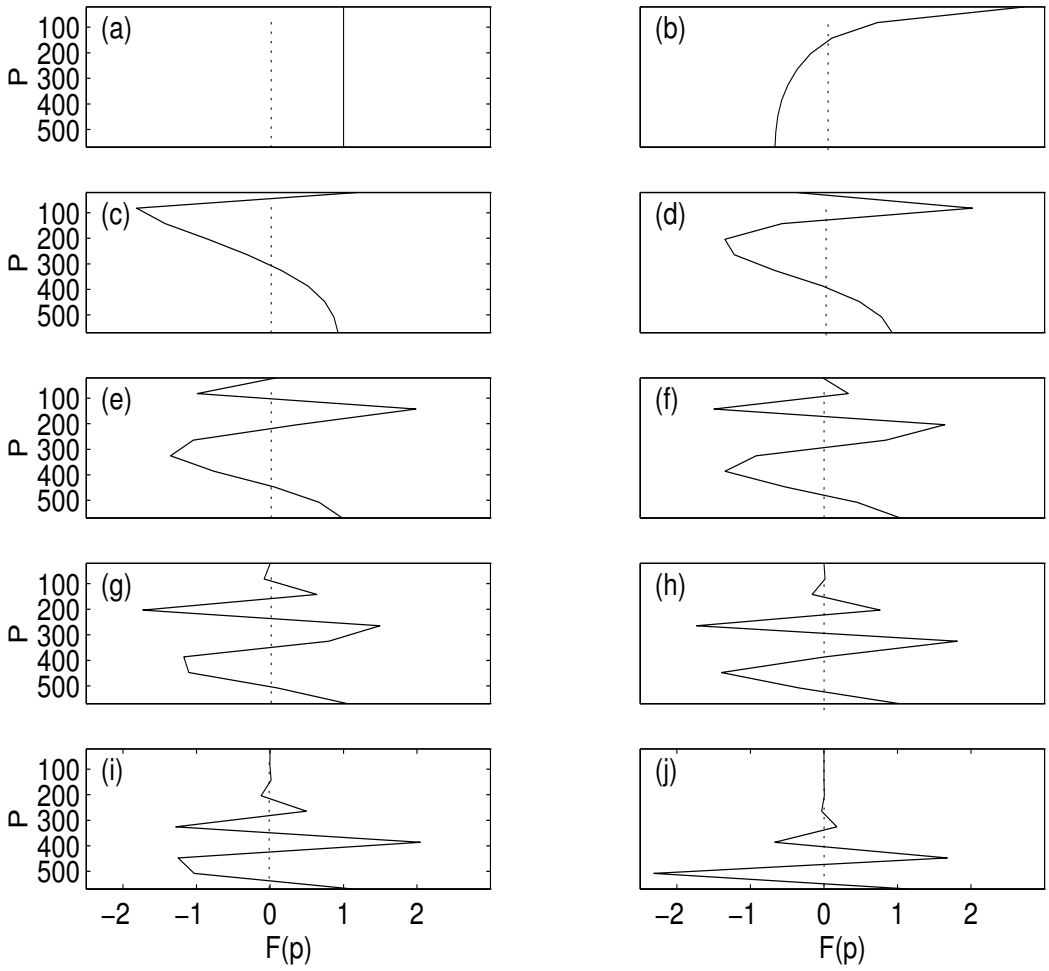
It is noticed from Figure 2.2(b) that the purely barotropic mode takes the value 1 for all levels while the first two baroclinic modes have a dominant baroclinic structure near the top of the model, which is strongly affected by the dissipative 'sponge' layer in the original GCM as well as the rapid decrease in ρ at high altitudes. Fluctuations in the vertical structures become progressively more dominant for the higher order modes, particularly near the ground level. Such disturbances seem to resemble a travelling wave moving from the mid-altitudes to the ground.

2.3.5 Energy

Although the normal mode problem discussed above minimises APE, with respect to KE it can be seen that Figure 2.3 shows that the purely barotropic function represents 89% of the total KE, whereas the first and second baroclinic functions supplement KE by just 6.1% and



(a) Froude numbers calculated from the vertical structure equation.



(b) Vertical normal modes from the vertical equation.

Figure 2.2: Froude numbers (a) and eigenmodes (b) of the vertical structure equation from a buoyancy frequency profile obtained from the SGCM streamfunction data.

2.5% respectively. Each of the remaining modes contribute less to the total KE than its previous mode, except for baroclinic modes 8 and 9 which contain 0.19 and 0.20% of KE respectively.

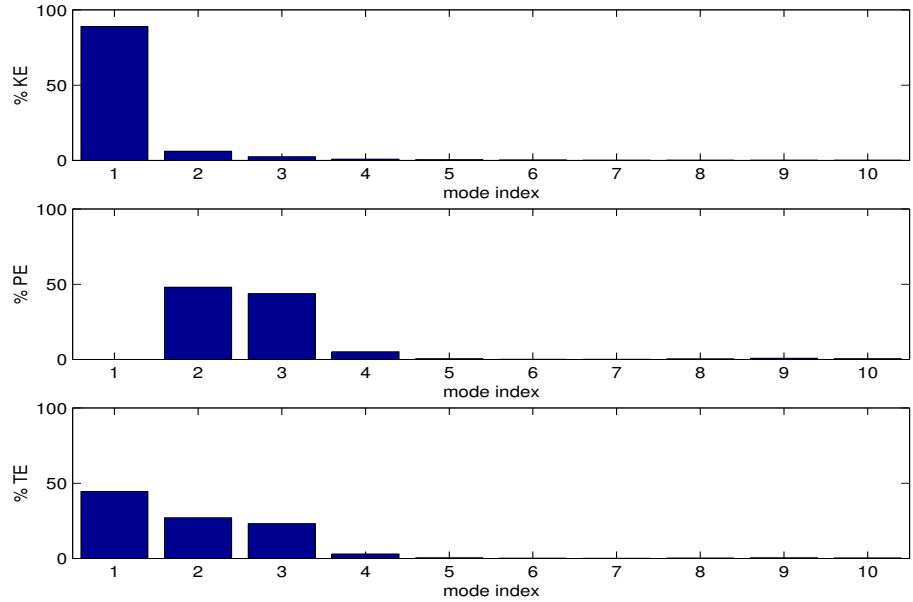
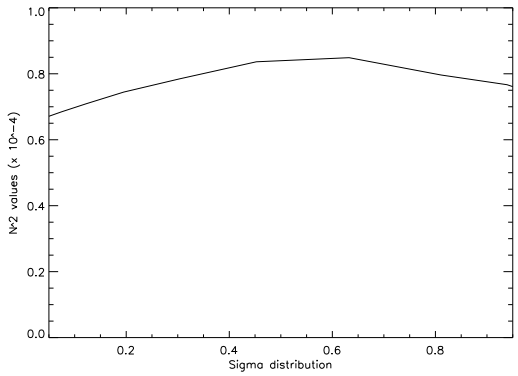


Figure 2.3: Contributions to KE and PE from the barotropic and baroclinic modes.

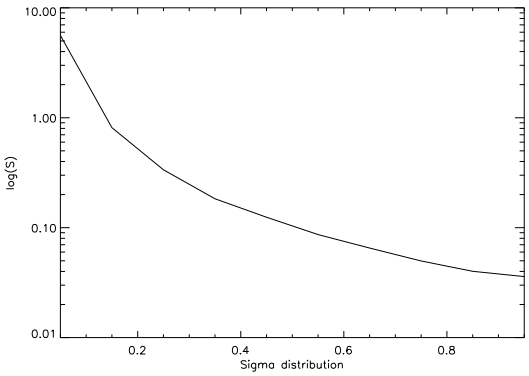
From the increasing eigenvalues given in Table 2.1 it is evident that, for example, were the atmosphere to achieve an equipartitioning in energy with respect to this basis, thus placing an equal excitation in each of the vertical modes, then the amount of APE in each mode would entirely depend upon the magnitude of its associated eigenvalue λ, which would produce an increasing APE contribution function. But from Figure 2.3 this is clearly not the case, and the hypothesis of an energy equipartition is not supported. In fact, it can be seen that the model places the majority of the APE of the system in the first two or three baroclinic modes, accounting for 97.1% of the APE in modes 1 to 3. An appropriate interpretation of this result is that the model evidently seeks a state of minimum APE, and this minimisation is optimised by placing most of its APE in those modes which are least efficient at representing APE. This accounts for the majority of APE appearing in the modes which are effectively the most baroclinically relaxed.

By definition the barotropic mode accommodates no APE, but in the experiment analysed here the first, second and third baroclinic eigenfunctions respectively represent 48.1, 43.9 and 5.1% of the total APE, while the remaining modes contribute proportionately less until baroclinic modes 8 and 9 are reached (which contain 0.89 and 0.64% of the total APE respectively).

If baroclinic modes 8 and 9 are neglected, the contribution of each successive mode to TE is also a decreasing function. It was thought initially that the high order modal contributions were due to the excitation of small scale horizontal structures, particularly as the vertical scaling of the higher order modes reduces significantly. If large horizontal scale behaviour only was being modelled then such contributions could be neglected. To understand the reason for the significant contributions to TE by the final two baroclinic modes, the T21 spectral data



(a) The Brunt-V*äisälä* frequency (N^2 profile) for the MGCM at latitude 52.61^o in the NH (representing a typical profile). Note that N^2 is almost constant (to $\pm$ 10% or so) and virtually symmetric about σ =0.5.



(b) The stratification parameter (S) at latitude 52.61^o in the NH obtained by using $S = N^2/g^2\rho^2$. Note the steep gradient near the top of the model (i.e. for $\sigma < 0.3$).

Figure 2.4: The form of (a) the MGCM N^2 profile and (b) the stratification parameter (S) at latitude 52.61^o in the NH. The N^2 profile is almost constant but the corresponding S profile rapidly increases for $\sigma < 0.3$. The reason for this strong baroclinic structure near the upper limits of the model is that $S \sim O(\frac{1}{\rho^2})$ (as N^2 is practically constant) and between sigma surfaces $\sigma = 0.25$ and 0.05 the vertical height of the model quickly increases from 13.86 km to 29.95 km (from ground level), resulting in a fast decrease in the density (ρ) values.

was reduced to a T5 truncation in order to reduce the contribution of small-scale processes. Despite this truncation significant KE and APE contributions were still found in the baroclinic modes 8 and 9. From these analyses it can be concluded that such contributions are mainly artifacts of the vertical truncation level as well as consequences of the severe reduction in the vertical length scales, all resulting in numerical inaccuracies.

2.4 The MGCM data

Our aim now is to apply the vertical structure equation (2.5) to a more realistic Martian GCM (MGCM), and determine whether similar observations (to the SGCM) are found, i.e. whether the atmospheric model places the majority of APE in the leading order minimum APE modes. The MGCM uses a spectral primitive equation model which has a T21 truncation and 10 equally spaced σ levels in the vertical. The difference is that this version of the GCM includes such parameters as radiative heating and friction, as described by Hourdin et al (1993). A northern winter season ($L_s = 239.6^o$ to 300.42^o) is chosen so that the seasonal conditions would be comparable to those of the perpetual southern winter run of the SGCM ($L_s = 90^o$). This model will be explained more comprehensively in Chapter 3.

The MGCM N^2 profile and the stratification parameter

The N^2 profile at latitude 52.61^o in the northern hemisphere (NH) (representing a typical profile) is shown in Figure 2.4(a) and the corresponding stratification parameter profile in Figure 2.4(b). The N^2 values are almost symmetric about $\sigma = 0.5$ and reasonably uniform, but the S values rapidly increase (for $\sigma < 0.3$), again depicting a strong baroclinic structure near

the upper limits of the atmosphere (as was observed for the SGCM in Figure 2.1(a)).

Typical model streamfunction data from the MGCM is illustrated in Figure 3.24, where a stationary or slowly moving wave can be seen at about $\pm 50^o$ latitude, together with a rapidly moving diurnal tide which is seen as a travelling wave. Because the streamfunction has a dominant steady wave in the NH latitude band $[40, 60]^o$, it was decided that the stratification parameter would be averaged over this interval. The consequence of such a derivation is that the normal modes are optimal over this chosen band.

2.4.1 Eigenvalues and eigenmodes

The λ_k's from the MGCM data form an increasing sequence shown in Table 2.2 and illustrated in Figure 2.5(a). The leading eigenvalues in Table 2.2 are greater in magnitude than those which were shown in Table 2.1 for the SGCM data.

Mode	Eigenvalue	Rossby radius of deformation	Froude number
Barotropic	$\lambda_0 = 0$	$Rd_0 = \infty$	0.0
Baroclinic 1	$\lambda_1 = 9.64 \times 10^{-7}$ km^{-2}	$Rd_1 = 1018.5$ km	11.1
Baroclinic 2	$\lambda_2 = 3.76 \times 10^{-6}$ km^{-2}	$Rd_2 = 515.7$ km	43.3
Baroclinic 3	$\lambda_3 = 7.84 \times 10^{-6}$ km^{-2}	$Rd_3 = 357.1$ km	90.3
Baroclinic 4	$\lambda_4 = 1.24 \times 10^{-5}$ km^{-2}	$Rd_4 = 283.9$ km	142.6
Baroclinic 5	$\lambda_5 = 2.02 \times 10^{-5}$ km^{-2}	$Rd_5 = 222.5$ km	232.5
Baroclinic 6	$\lambda_6 = 3.12 \times 10^{-5}$ km^{-2}	$Rd_6 = 179.0$ km	359.7
Baroclinic 7	$\lambda_7 = 4.77 \times 10^{-5}$ km^{-2}	$Rd_7 = 144.8$ km	549.6
Baroclinic 8	$\lambda_8 = 7.07 \times 10^{-5}$ km^{-2}	$Rd_8 = 118.9$ km	814.3
Baroclinic 9	$\lambda_9 = 1.01 \times 10^{-4}$ km^{-2}	$Rd_9 = 99.5$ km	1167.9

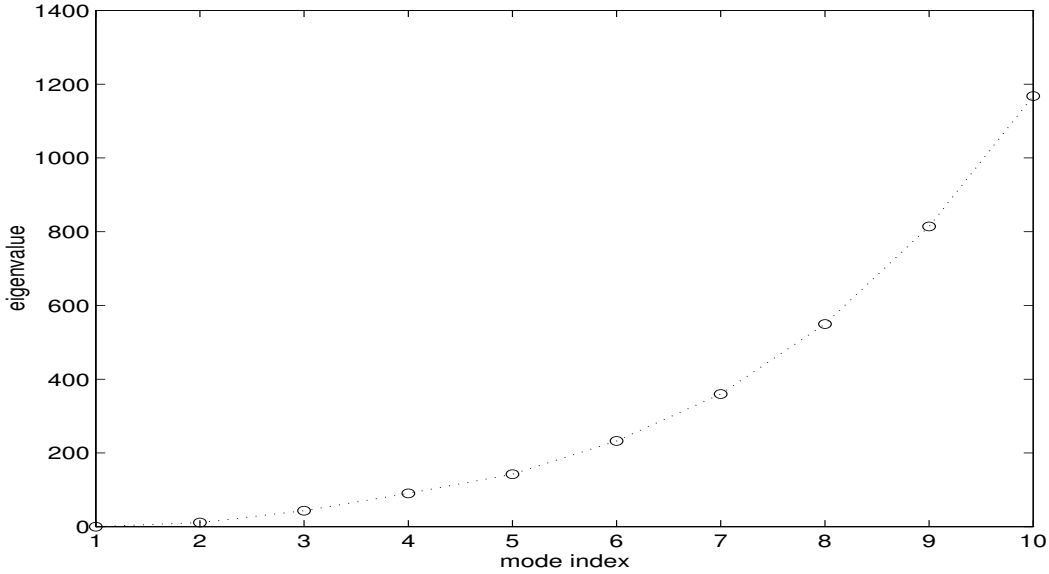
Table 2.2: Table of Froude numbers of the associated vertical eigenmodes for the MGCM case.

The modes are very similar to those seen from the SGCM data set. Again, the form of the two leading baroclinic modes have a dominant baroclinic structure near the top of the model, having been caused by the dissipative layer and the rapid decrease in ρ.

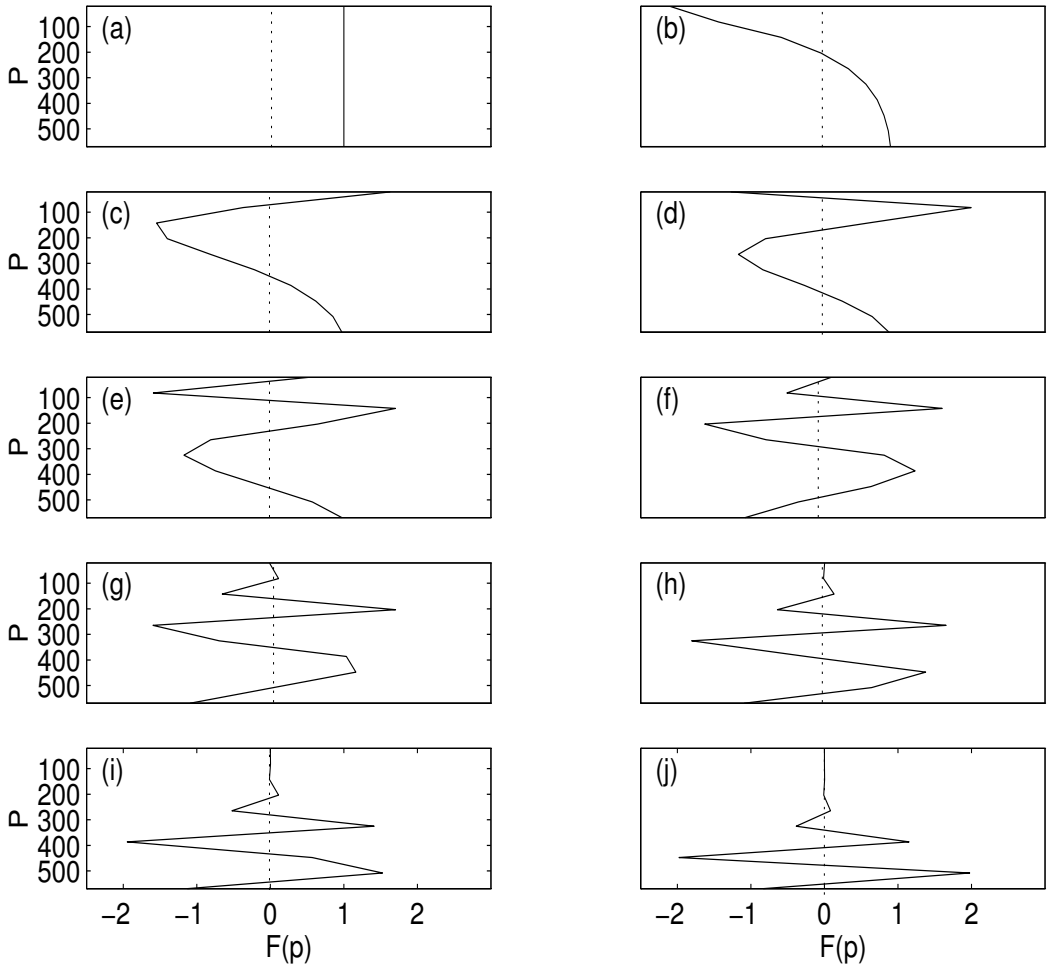
2.4.2 Energy

It is evident from Figure 2.6 that the purely barotropic mode contains 80% of the total KE and the leading two baroclinic modes provide an additional 6.2% and 5.1% to total KE respectively. Unlike the SGCM, each mode contributes less to KE than its previous mode and this is also evident in the amount of TE contained in each of the modes, where it is seen that each mode contains less of the TE than the preceding mode.

Again, there is no evidence of equipartitioning in energy amongst the vertical modes, but the model does arrange the majority of the APE of the system so that it appears in the leading baroclinic modes. Since the modes are consistent with a variational principle, the model is seeking a state of minimum APE and distributing the majority of APE in the most baroclinically relaxed modes.



(a) Froude numbers calculated from the vertical structure equation.



(b) Vertical normal modes from the vertical equation.

Figure 2.5: Froude numbers (a) and eigenmodes (b) of the vertical structure equation from a buoyancy frequency profile obtained from the MGCM streamfunction data.

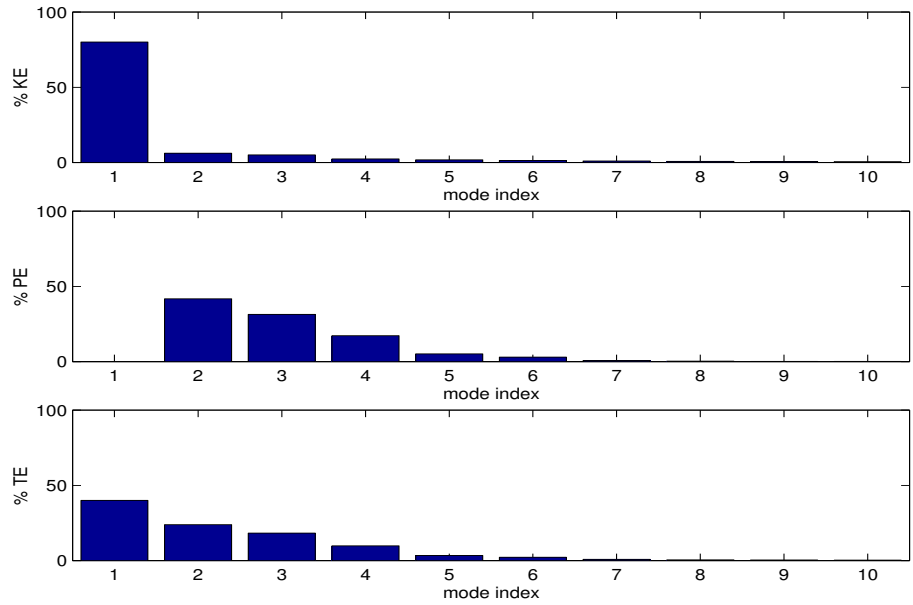


Figure 2.6: Contributions to KE and PE from the barotropic and baroclinic modes.

The first, second and third baroclinic modes account for 41.7, 31.4 and 17.3% of APE while the remaining six modes contribute an additional 4.5% to APE.

2.5 Summary and conclusions

This Chapter relies on the work done by Whitehouse et al. (1999) and shows that the linearised vertical structure equations of a form equivalent to that of Flierl (1978) are consistent with an extremum principle. It has been shown here that the baroclinic modes arising from this principle actually minimise APE in the vertical and so could be referred to as baroclinically 'relaxed' modes.

Since atmospheric circulations are widely interpreted as arising from the release of APE produced by the action of differential solar heating with a presence of the planetary gravitational field, the APE minimisation principle makes it clear why a decomposition of real atmospheric flows in eigenmodes derived from Flierl-type vertical structure equations are so efficient in representing atmospheric dynamical structures.

In our examples it was shown that the atmospheric states of our Martian GCMs were effectively seeking states of minimum APE by placing the majority of the APE in the leading order minimum APE modes. As no particular atmospheric properties have been assumed, these results can be taken to apply to any rapidly rotating baroclinic atmosphere, where we conclude that similar processes will occur.

Chapter 3

Mars

3.1 Aim

The SGCM is an idealised model of atmospheric circulation in a generic planetary atmosphere, in which the meteorological primitive equations are solved on a sphere with simplified physical parameterisations and is described in detail by Collins and James (1995). The simplified parameterisations of heating and friction have been used to examine baroclinic waves in the parameter range appropriate to the Martian atmosphere, during a continual season numerical simulation of the southern hemisphere winter ($L_s = 90^o$). The MGCM is a more sophisticated model, which includes comprehensive parameterisations of radiative heating and friction besides modelling of boundary layer processes etc. during a northern winter run (from $L_s = 239.6^o$ to 300.42^o), and is described by Collins et. al (1996).

The question of whether we should use standard EOFs (which describe deviations from the mean state) or whether the basis should describe the total state vector (and so include norms which maximise energy extractions) will be addressed. POD modes will be optimised to describe the variance, kinetic energy (KE) and total energy (TE=APE+KE, where APE is the available potential energy) of the system. The standard correlation norm will be compared with the two energy norms which have distinct physical interpretations.

Within this Chapter we will be addressing the question of what norm should be used to monitor energy conversion processes within the data, what emphasis each norm places on certain spatial and temporal structures within the data and how many patterns are required to reproduce the most 'significant' dynamics of the model.

3.2 Martian atmospheric dynamics

Mars is one of the terrestrial planets, having a relatively thin atmosphere in terms of physical depth, and covering a solid surface. Interestingly, many of the basic planetary characteristics relevant to Martian atmospheric dynamics are similar to those on Earth (see Table 3.1). The inclination of the Martian polar axis to the ecliptic plane implies that Mars has seasons like Earth, and both are rapidly rotating planets, rotating about the same speed so that Coriolis forces will have similar magnitudes. Much of the water on Mars is believed to be in the per-

mafrost which is beneath the surface, leading to the low observed atmospheric concentrations (Farmer (1979)). The Martian atmosphere does not contain enough water vapour to permit large amounts of latent heat release in convective or frontal clouds, however, absorption of short wave radiation by suspended dust can produce significant heating rates (Gierash and Goody (1972), Pollock et al. (1979)). This results in a fundamental difference in how the two planetary circulations are driven, as there is no large amount of latent heat release in the Martian tropics to drive the zonal-mean circulation (as there is in the Earth's troposphere). Instead, the Martian circulation is primarily driven directly by heating of the surface and dust suspended in the atmosphere. It has been suggested that the Martian weather systems are much more regular and coherent than corresponding weather systems on Earth. It is because the Martian atmosphere has certain characteristics which are very different to those of Earth, that a presence of regular travelling baroclinic waves occur.

Physical Parameters	**Earth**	**Mars**
Radius	6378 km	3397 km
Surface gravity	9.8 ms^{-2}	3.72 ms^{-2}
Mean distance from sun	1.49 x 10^8 km	2.28 x 10^8 km
Inclination of Equator to orbit	23.45^o	25.2^o
Rotation Period	24 h (1 day)	24 h 39.5 m (1 sol)
Length of year	365.25 days	669 sols
Surface pressure	1000 mb	5-10 mb
Surface Temperature	250-300 K	150-300 K
Scale Height	8 km	10 km

Table 3.1: Characteristics of Earth and Mars.

Even though there is a basic similarity in the values of relevant parameters for large-scale motions for Mars and the Earth, Martian meteorology has some unique aspects (Leovy (1969, 1979)). Interestingly, many of these aspects are particularly relevant to baroclinic instability, such as the very rapid radiative relaxation, with damping time scales on the order of several days (Goody and Belton (1967)); very large topography, where a large zonally symmetric component exists in the northern hemisphere (Kliore et al. (1973), Linda et al. (1979)) and considerable changes in the zonal-mean circulation due to global dust storms (Haberle et al. (1982)).

The winter circulation of the Martian atmosphere is dominated by the presence of heat transporting baroclinic transients. During the summer months there occur only small fluctuations of pressure, and travelling waves are usually absent in either hemisphere. This contrasts with the autumn and winter months when large amounts of carbon dioxide (CO_2) evaporate and condense out of the atmosphere onto the southern cap during the southern winter (and northern hemisphere summer). The condensation of CO_2 at the winter pole and its evaporation at the summer pole causes the pressure to rise and allows this period to be more predictable. This is a fundamental factor in the circulation of the Martian atmosphere (for further details see the extensive work of Kieffer et al. (1992)).

3.3 The SGCM

In the first part of our study we used data from a fully nonlinear GCM based on the meteorological primitive equations on a sphere and is similar to the model used by James and Gray (1986) and James and James (1992). The SGCM uses simplified parameterisations of the physical processes which occur at sub-gridscale, under conditions analogous to the Martian atmosphere and is described in detail by Collins and James (1995). The model solves the adiabatic primitive equations of meteorology, as described by Hoskins and Simmons (1975), in spherical coordinates, by using a spectral representation in the horizontal directions and finite-difference σ-coordinates in the vertical (where $\sigma = p/p_*$, p=pressure and p_*=surface pressure), having been used by Joshi (1994) and by Collins and James (1995) in idealised studies of the Martian atmosphere. The spectral primitive equation model employs a triangular truncation at total wavenumber 21 (denoted T21) and has 10 equally spaced σ levels in the vertical over a pressure range of 0 to 600 Pa (shown in Table 3.2). Surface drag is represented by Rayleigh friction with a time-scale of 3 days (denoted τ_D); heating by Newtonian relaxation toward a zonally symmetric equilibrium temperature distribution with a time-scale of 2 days (denoted τ_E). For simplicity topography was excluded in the model in order that the baroclinic waves could be isolated and would not suffer from orographic modulation.

Level	$\sigma = p/p_*$	p/Pa	Height/km, $Z = -Hln(\sigma)$, $H \sim 10km$
10	0.95	579.5	0.51
9	0.85	518.5	1.63
8	0.75	457.5	2.83
7	0.65	396.5	4.3
6	0.55	335.5	5.98
5	0.45	274.5	7.99
4	0.35	213.5	10.49
3	0.25	152.5	13.86
2	0.15	91.5	18.97
1	0.05	30.5	29.95

Table 3.2: Data level information for the SGCM and MGCM.

Figure 3.1 shows the range of experiments performed using the SGCM by Collins and James (1995). It can be seen that each square marks an experiment in the model's (τ_D, τ_E) parameter space, where a solid square represents a weak regime flow and an open square a wave 3 regime. The data which will be analysed in this thesis is obtained from a wave 3 integration run, corresponding to the position (3,2) in parameter space. According to Figure 3.1, a concentration of regular wave 3 flows are observed (for frictional timescales $\tau_D \in$ [3,..6]), surrounded by the more irregular weak wave integrations, which suggests that dissipative mechanisms are important in the maintenance of the large amplitude wave 3 flows.

In Figure 3.2(a) the time-averaged zonal winds with respect to latitude and height (km) for the SGCM are shown. It can be seen that there exists a zonal westerly jet in the southern hemisphere (SH) which ranges from 10-70 ms^{-1}. Temperature is illustrated in Figure 3.2(b) where the largest valued temperature profiles (of 160-200 K) occur in the northern hemisphere (NH), between 0-10 km from the ground. It can be seen from the SH stereographic projection

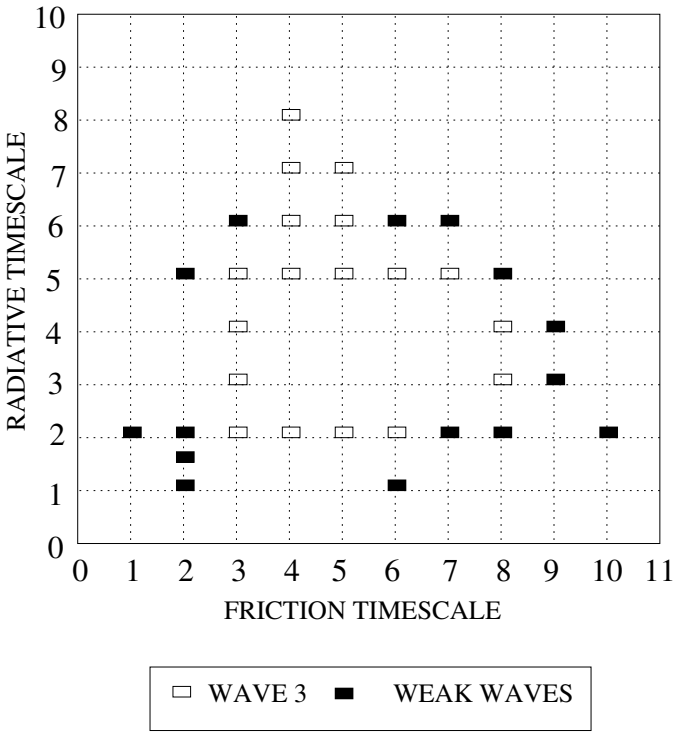


Figure 3.1: A regime diagram of the SGCM with squares marking all the experiments carried out by Collins and James (1995). Open squares correspond to the wave 3 regime, and solid squares to the weak wave regime. The abscissa is the frictional timescale τ_D and the ordinate is the radiative timescale τ_E both in Martian days.

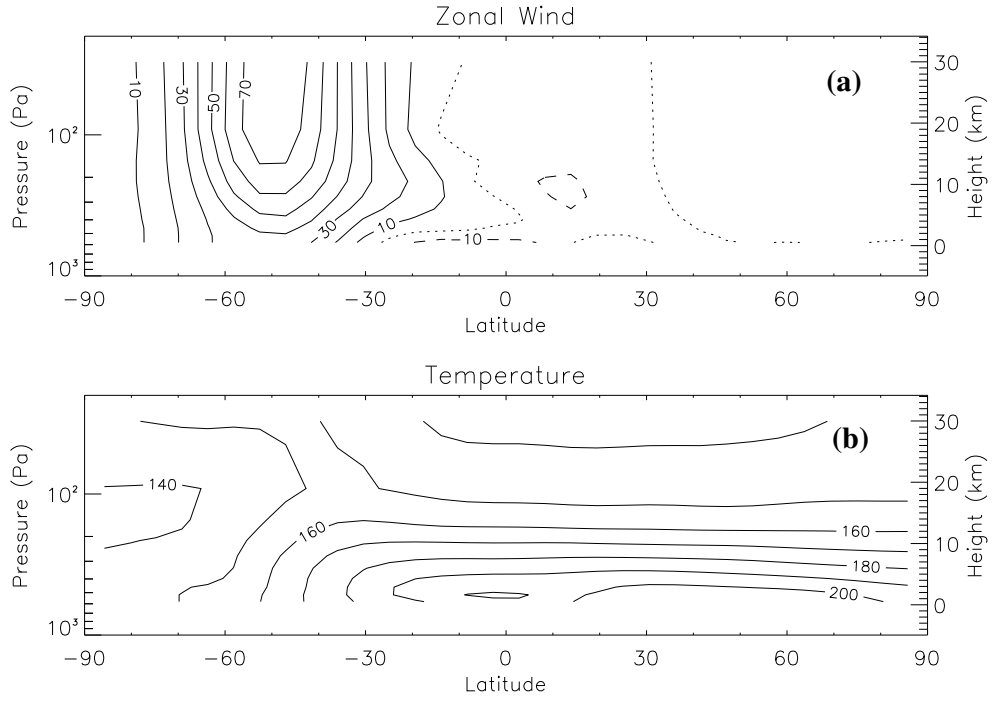


Figure 3.2: Pressure-latitude maps of the (a) SGCM time-averaged zonal wind and (b) temperature (on $\sigma = 0.45$) over 300 time levels during the southern winter solstice ($L_s = 90°$).

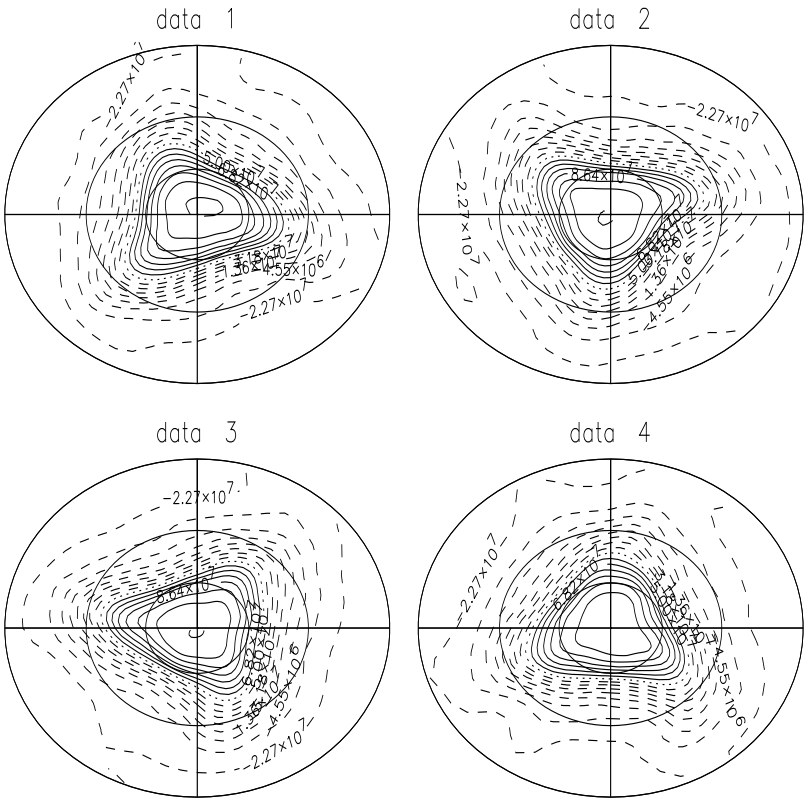


Figure 3.3: Southern hemisphere stereographic projection maps of the SGCM modelled streamfunction data during the southern winter solstice ($L_s = 90°$). The steady wave 3 streamfunction is shown at the first 4 days on $\sigma = 0.45$.

plots of the SGCM modelled streamfunction data in Figure 3.3 that a steady zonal wavenumber 3 structure is depicted in the SH.

3.3.1 The SGCM basic state

In Figure 3.4 the two definitions of the barotropic and the first three baroclinic components (defined in § 2.3.1) of the SGCM basic state are shown. In Figure 3.4(a) the basic-state is defined as the time average of the zonally averaged flow and in Figure 3.4(b) is defined as the steady state of the 2-D version (zonally symmetric) of the SGCM (by suppressing the waves) after 400 sols (i.e. 400 Martian days), to as to ensure that the state has equilibrated. The basic-state defined here, and in the remaining part of the thesis, is taken to be that which is shown in Figure 3.4(b), i.e. the zonally symmetric version of the SGCM in order that our used definition is consistent with that used by Collins (1993). (For further details on the used basic state see § 4.2.1 and § 5.3.2).

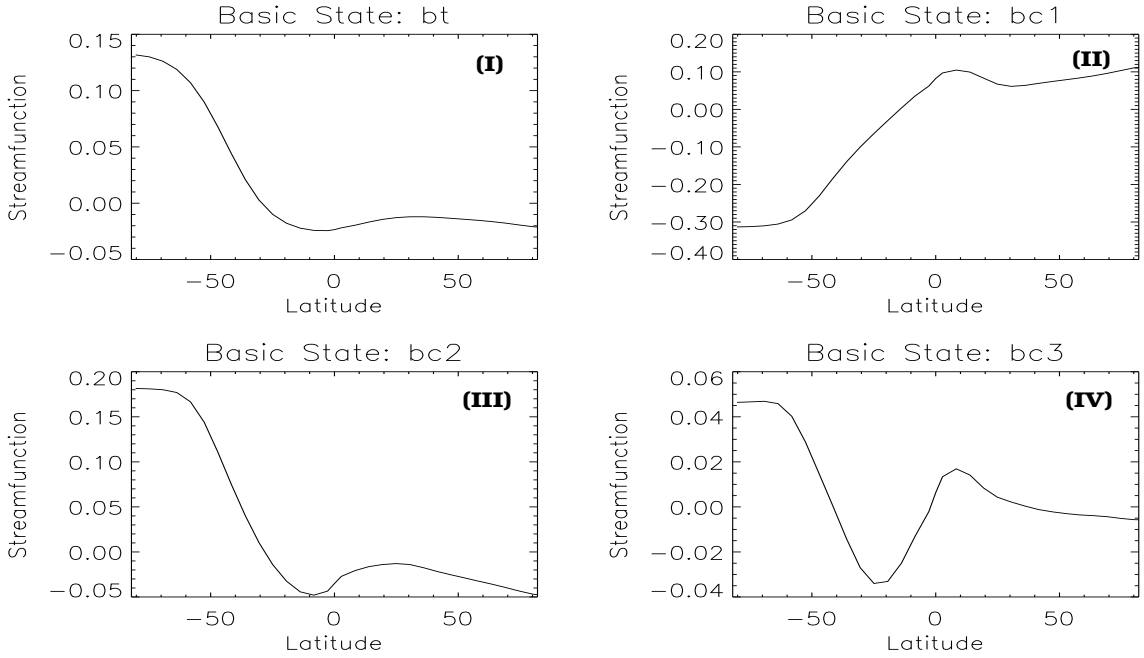
It can be seen in Figure 3.4 that the differences between the above two definitions for the basic-state are negligible. The top left frame (block I) in Figure 3.4 shows the barotropic basic state component (for the definitions of the barotropic and baroclinic components refer to § 2.3.1) of the SGCM data. This profile can be interpreted as an equal weighted average over the vertical levels of the model. The largest profile exists in the SH which grows rapidly between 30-80^o in the SH and then levels off at the most southern latitudes.

Baroclinic components 2 and 3 (blocks III and IV respectively) of the SGCM basic state are depicted in the bottom two plots of Figure 3.4. These both have dominant southern profiles, rapidly increasing between 30-80^o in the SH.

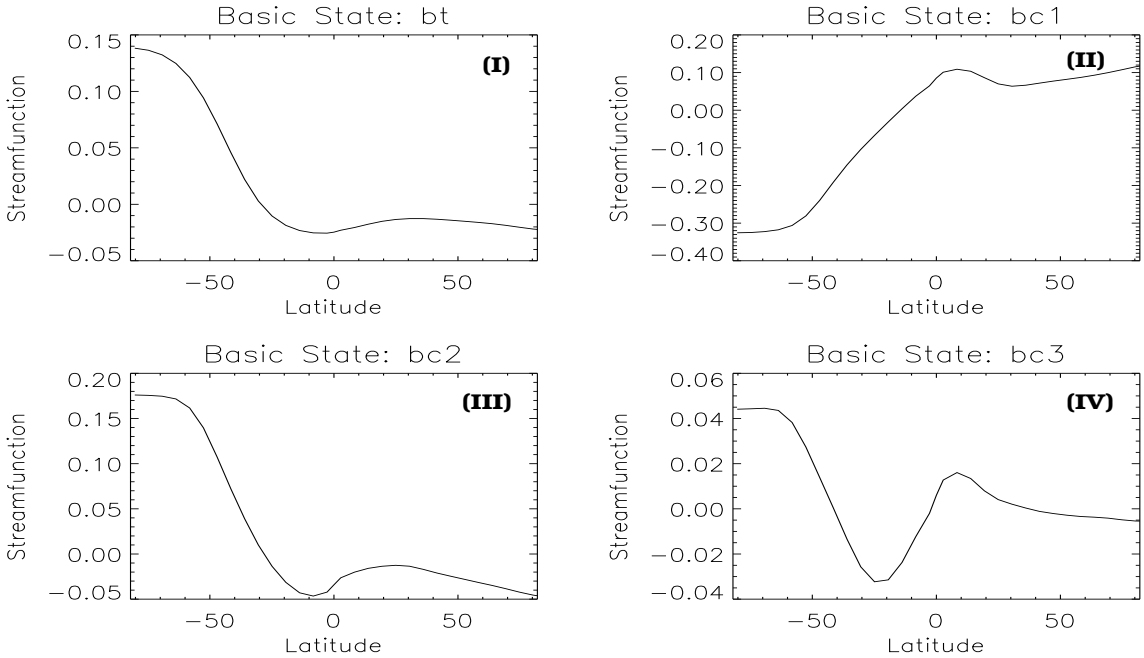
The third baroclinic structure is shown in the bottom right quadrant of Figure 3.4 and is by far the smallest of the baroclinic components, being almost a third the size of the others.

Whilst the barotropic and baroclinic components 2 and 3 of the basic state are very similar in shape the first baroclinic mode (block II) is almost a mirror-image of them about the zero amplitude line. The remaining six baroclinic components of the basic-state are not shown here, but it was found that the barotropic and first two baroclinic components were dominant in amplitude and energy (see § 2.3.5).

It is now clear from the SGCM data set that it is necessary to retain at least the first two baroclinic modes (besides the barotropic component) in the used vertical scheme for the low-dimensional model. The reason is that in both cases the barotropic and the two leading baroclinic modes dominate in terms of their amplitude. In comparison, it was found that the higher order baroclinic modes were insignificant in size and energy (see § 2.3.5).



(a) A latitudinal cross-sectional profile of the basic state, defined here as the time average of the zonally averaged flow.



(b) A latitudinal cross-sectional profile of the basic state, defined here as the steady state of a 2-D version run of the SGCM (by suppressing the waves) after 400 sols, so as to ensure that the state had equilibrated.

Figure 3.4: Basic state SGCM streamfunction components. The profiles are of (I) the barotropic, (II) the first, (III) the second and finally (IV) the third baroclinic components of the SGCM basic-state from a southern winter solstice ($L_s = 90^o$) continual run. The basic state in this thesis is defined as the zonally symmetric version of the SGCM (shown in (b)).

3.3.2 The spectral grid

The SGCM streamfunction data lies on a jagged T21 spectral grid and so it is necessary to explain the ordering of the spectral coefficients on such a grid.

The order of data storage is by level, beginning at the uppermost level (i.e. at σ=0.05). For each level there exist odd and even complex coefficients (an odd for a coefficient of a cosine and an even for a sine) and there are equal number of odd and even coefficients at each level. Taking n to be the total wavenumber and m the zonal wavenumber (and so n - $|m|$ is the number of zeros between the north and south poles), then the ordering of the coefficients is similar to that shown in Figure 3.5 for the T5 and T4 truncations, where the coefficients are read in order of increasing n with increasing m, beginning with the even coefficients, then the odd (for full details see Blackburn (1985)).

Taking A_n^m to be a spectral coefficient at position (n, m) on a spectral grid, then the coefficients are read in the following order:

$$\boxed{\text{EVEN}}\ A_0^0, A_2^0, A_4^0, ..., A_1^1, A_3^1, ..., A_2^2, A_4^2, ..., A_{NN-1}^{NN-1},$$
$$\boxed{\text{ODD}}\ A_1^0, A_3^0, A_5^0, ..., A_2^1, A_4^1, ..., A_3^2, A_5^2, ..., A_{NN-1}^{NN-1},$$

where NN is the highest total wavenumber retained in the spectral series.

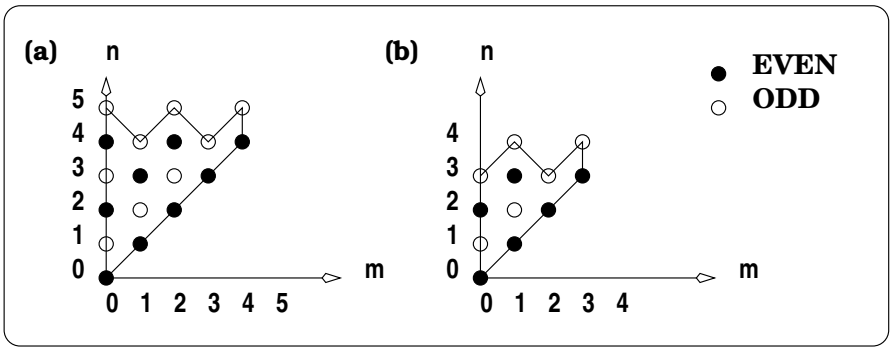


Figure 3.5: The spectral grids for a jagged (a) T5 and (b) T4 triangular truncations. Coefficients are read in order of increasing n with increasing m (where n is the total wavenumber and m the zonal wavenumber) for the even then the odd coefficients (which are depicted as black and empty circles respectively).

3.3.3 Calculation of PODs in spectral space

POD modes are usually calculated in physical grid space, but because of the calculations of Schubert (1985) and the discussions of Selten (1995), the spatial eigenmodes in this thesis will be calculated in spectral space. Whether the calculations are made in grid or spectral space is irrelevant since both methods are equivalent (and connected via a linear transformation).

Since the data lies on a T21 spectral grid it was also felt that the PODs should also have a T21 resolution i.e. retaining all wavenumbers up to wavenumber 21. If the PODs were cal-

culated in grid space then a covariance matrix of size (64 × 32,64 × 32) would be required if a T21 spectral resolution is to be attained. However, a T21 data set can be fully resolved by 242 real odd and even spectral data points which implies that only a spectral covariance matrix of size (484,484) would be necessary for the same resolution. As the numerical eigenvector retrieval scheme (using F02ABF from the NAG routine library which calculates all the eigenvalues and eigenvectors of a real symmetric matrix) operates at a speed of l^3 (where l is the length of the matrix), then the spectral problem runs approximately $(4.23)^3$ times faster.

If the covariance function of the streamfunction in the continuous case is:

$$C(\theta,\phi,\theta',\phi') = r^2\sqrt{\sin\theta\sin\theta'}\left(\overline{\psi(\theta,\phi)\psi(\theta',\phi')}\right), \tag{3.1}$$

where the overbar indicates a time average, then the following eigenvalue problem holds on a sphere:

$$\frac{1}{2\pi^2}\int_0^{2\pi}\int_0^{\pi} C(\theta,\phi,\theta',\phi')\Upsilon_i(\theta',\phi')d\theta'd\phi' = \lambda_i\Upsilon_i(\theta,\phi), \tag{3.2}$$

where $\Phi_i(\theta,\phi) = \frac{1}{\sqrt{\sin\theta}}\Upsilon_i(\theta,\phi)$ is the i-th POD mode.

Expanding both the streamfunction $\psi(\theta,\phi,t)$ and the spatial eigenmodes $\Phi_i(\theta,\phi)$ as a jagged T21 spectral truncation, we have:

$$\begin{aligned}\psi(\theta,\phi,t) &= \sum_{m=0}^{20}\sum_{n=m}^{N_m}\psi_{mn}P_{mn}e^{im\phi},\\ \Phi_i(\theta,\phi) &= \sum_{m=0}^{20}\sum_{n=m}^{N_m}V_{mni}P_{mn}e^{im\phi},\end{aligned} \tag{3.3}$$

where ψ_{mn} and V_{mni} are the spectral coefficients, $P_{mn}(\mu)$ are the associated Legendre polynomials of the first kind (of degree l and order m and where μ is the sine of the geographic latitude) which satisfy the normalisation condition

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} {P_l}^m(\sin\theta){P_l}^m(\sin\theta)\cos\theta d\theta = 1, \tag{3.4}$$

and the summation N_m is defined as:

$$N_m \;=\; \begin{cases} 21 & m \quad \text{even} \\ 20 & m \quad \text{odd.} \end{cases} \tag{3.5}$$

Substituting (3.3) into (3.2) gives:

$$\sum_{m=0}^{20}\sum_{n=m}^{N_m}\overline{\psi_{mn}{\psi^*}_{m'n'}}V_{m'n'i} = \lambda_i V_{mni}, \tag{3.6}$$

where * denotes complex conjugation. We notice that as

$$\psi_{-mn} = (-1)^m\psi^*_{mn}, \tag{3.7}$$

then the standard correlation matrix of (3.2) then satisfies the following relationship:

$$\frac{1}{2\pi^2}\int_0^{2\pi}\int_0^{\pi}\psi(\theta,\phi)\psi(\theta,\phi)d\theta d\phi = \sum_{m=0}^{20}\sum_{n=m}^{N_m}\psi_{mn}^2. \tag{3.8}$$

Therefore (3.6) can be rewritten as

$$\sum_{m=0}^{20}\sum_{n=m}^{N_m}\overline{\psi_{m'n'}\psi_{mn}}V_{mni} = \lambda_i V_{m'n'i}, \tag{3.9}$$

where V_{mni} is the spectral coefficient of the the i-th POD mode V_i at the spectral grid point (m,n).

So in summary, by formulating the POD modes in spectral space a high resolution could be maintained relatively economically in terms of CPU time.

3.3.4 Constructing continuous vertical modes

The state vector was defined as $\mathbf{V}(\theta,\phi,\sigma) = \left(\tilde{\psi}_0,\tilde{\psi}_1,\tilde{\psi}_2,\tilde{\psi}_3\right)$ where each of the baroclinic components $\tilde{\psi}_k$ are functions of latitude and longitude and are defined as:

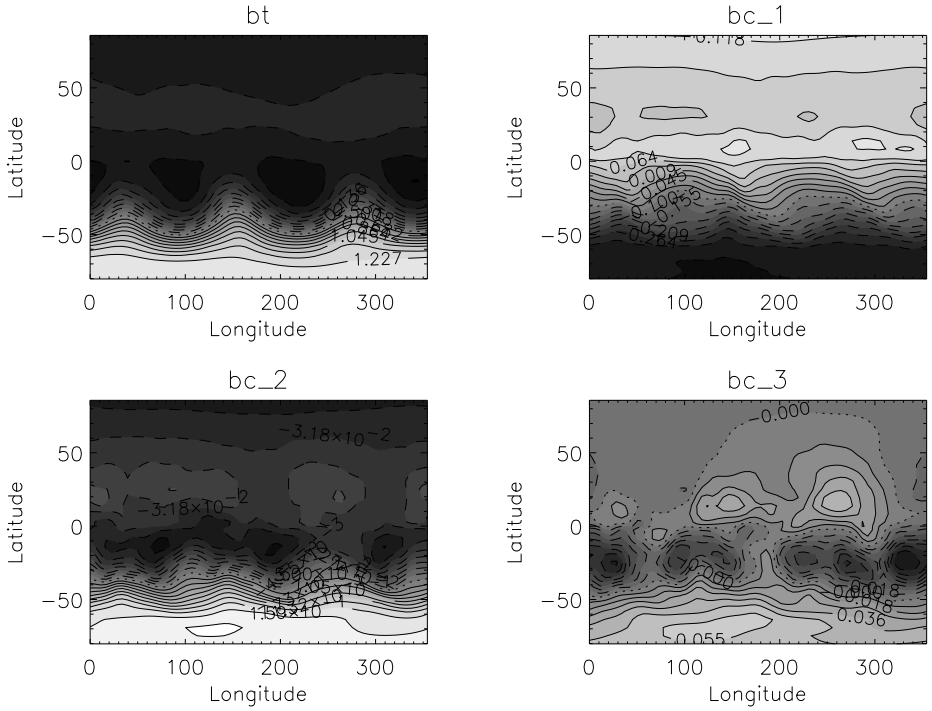
$$\tilde{\psi}_k(\theta,\phi,\sigma) = \frac{1}{N}\sum_{j=1}^{N}\langle\mathbf{H}_k(\sigma_j),\psi_j(\theta,\phi)\rangle\,, \tag{3.10}$$

where $\mathbf{H}_i(\sigma_j)$ is the i-th baroclinic mode in layer σ_j and N is the number of vertical levels (and the number of normal modes). For further details on the decomposition of (3.10) refer to § 2.3.1.

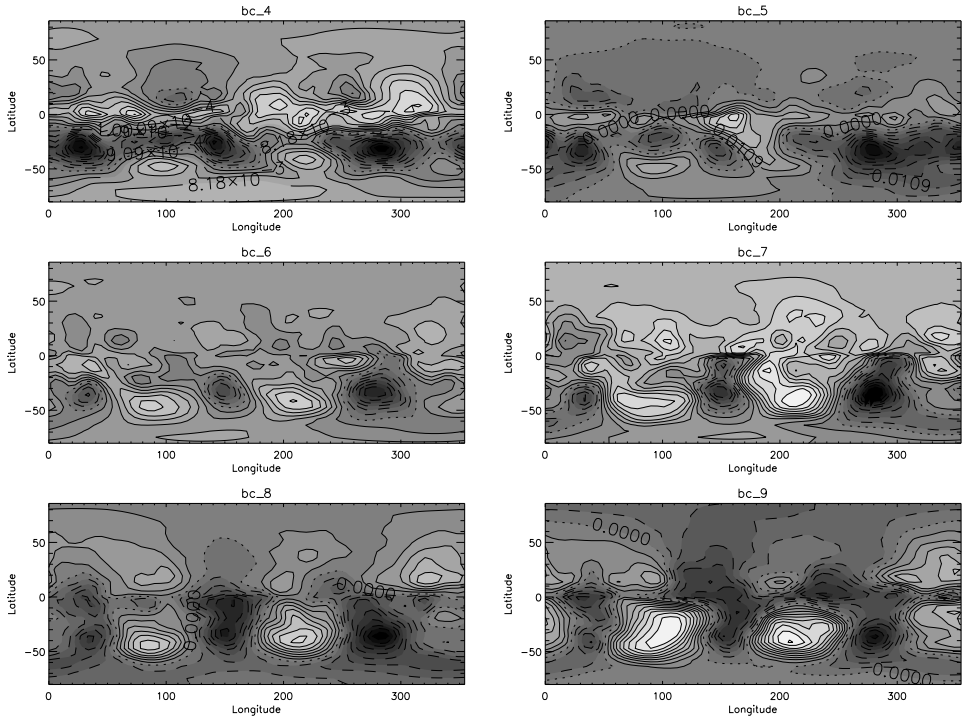
Figure 3.6 shows the decomposition of the SGCM streamfunction data into its purely barotropic and nine baroclinic components. Figure 3.6(a) shows the instantaneous barotropic and first three baroclinic components of the streamfunction data during the first day. The barotropic component can be viewed as an equal-weighted averaged flow over the vertical levels, which reveals a steady wave flow between $40-60^o$ in the SH. The first and second baroclinic components contain modulated wave 3 profiles in the SH, whereas the third component possesses a baroclinic wavenumber 3 structure, centred at a latitude of 30^o in the SH, surrounded by a more complicated flow. Figure 3.6(b) shows the instantaneous baroclinic components 4 to 9 of the streamfunction data, where baroclinic modes 4 to 7 contain wavenumber 3 structures, centred at a latitude of 50^o in the SH and baroclinic modes 8 and 9 have wavenumber 2 structures at a latitude of 45^o in the SH (where negative contour values are shaded).

3.3.5 Projection of streamfunction or vorticity.

Selten (1995) raised the question of whether to expand in terms of the POD modes of the streamfunction or vorticity. If vorticity is chosen, then the POD modes need to be optimised to describe the vorticity instead of the streamfunction, and the inner product $<\psi,\psi>=\int_\Gamma\nabla^2\psi\cdot\nabla^2\psi\;d\Omega$ defines enstrophy. Such an inner product is usually used to emphasise small-scale rather than the squared norm (ψ_i,ψ_j), but as our model is intended to describe large-scale



(a) Decomposing the instantaneous streamfunction field into its purely barotropic (bt) and its first three baroclinic components (bc_1, bc_2 and bc_3) during the first day.



(b) Decomposing the instantaneous streamfunction field into its baroclinic components 4 to 9 (bc_4 to bc_9) during the first day.

Figure 3.6: Decomposition of the SGCM streamfunction data into its purely barotropic and nine baroclinic components. The figure illustrates the instantaneous fields of (a) the purely barotropic and first three baroclinic components and (b) baroclinic components 4 to 9 during the first day.

circulation, this basis will not be used.

If the streamfunction is expanded in a truncated set of PODs, it is important to note that the corresponding vorticity and planetary vorticity field is not described exactly by the same set of PODs. By projecting the streamfunction onto a truncated series of PODs then it is not possible to derive a truncated POD model which conserves both KE and enstrophy (ENS) simultaneously, which is explained by Schubert (1985). As KE is restricted to larger scales this conservation law is considered more important. Furthermore, if forcing and dissipative terms are added then neither KE nor ENS are conserved.

3.3.6 The KE and TE norms

The standard correlation norm would give:

$$\mathbf{R\Phi} = \lambda\mathbf{\Phi}, \tag{3.11}$$

with the covariance matrix $\mathbf{R}$ defined by:

$$R_{ij} = \overline{(V_i - \bar{V}_i)(V_j - \bar{V}_j)}, \tag{3.12}$$

where the overbar denotes the time mean, so that the POD modes describe deviations from the mean state. But following Selten (1993), the same POD method can be applied to construct a set of basis patterns which describe the total state vector. Here the covariance matrix $\mathbf{R}$ of equation (3.11) is replaced by matrix $\mathbf{C}$ such that:

$$C_{ij} = \overline{V_i V_j}, \tag{3.13}$$

so that the mean state is retained in the POD expansion to allow for dynamical interactions between the mean background flow and the anomalous POD modes. The POD modes found by using matrix $\mathbf{R}$ are optimised to describe the variance whereas those obtained by using matrix $\mathbf{C}$ are optimised to describe the energy.

The inner product which defines the energy matrix $\mathbf{M}$ must satisfy the condition:

$$< \mathbf{V}, \mathbf{V^T} >= \mathbf{V}^T\mathbf{M}\mathbf{V}, \tag{3.14}$$

then the eigenvalue problem becomes:

$$< \mathbf{V}, \mathbf{V}^T > \mathbf{M}^T\mathbf{U} = \lambda\mathbf{U}. \tag{3.15}$$

Noting that $\mathbf{M}^T = \mathbf{M}$ (as the energy matrix $\mathbf{M}$ is diagonal) then

$$\mathbf{C}\mathbf{M_p}\mathbf{U} = \lambda\mathbf{U}, \tag{3.16}$$

where the diagonal energy matrix $\mathbf{M}_p$ depends on the product:

$$< \psi, \psi >_p = \frac{1}{2}\int_0^{2\pi}\int_0^{\pi} (\nabla^p\psi \cdot \nabla^p\psi)\, a^2 \sin\theta d\theta d\phi, \quad p = 0, 1, 2, ... \tag{3.17}$$

so the basis patterns **U** describe the total state vector.

We note that $< \psi, \psi >_0$ is the correlation norm of the streamfunction. As Selten (1995) further remarks as p increases there is much more emphasis towards the smaller scales.

3.3.7 Formation of the spectral energy matrix

The energy matrices used in this thesis optimise KE or TE. Recalling the equality given in (2.12) for KE:

$$\begin{aligned} \text{KE} &= \frac{1}{2}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\sum_{k}\left[\nabla_h\tilde{\psi}_k\cdot\nabla_h\tilde{\psi}_k\right]a^2\sin\theta d\theta d\phi \\ &= -\frac{1}{2}\sum_{k}\int_{0}^{\pi}\int_{0}^{2\pi}\left[\tilde{\psi}_k\cdot\nabla_h^2\tilde{\psi}_k\right]a^2\sin\theta d\theta d\phi \\ &= \sum_{mnk} n(n+1)\tilde{\psi}^2_{mnk}, \end{aligned} \tag{3.18}$$

and (2.14) for TE:

$$\begin{aligned} \text{TE} &= \text{KE} + \frac{1}{2}\sum_{k}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\lambda_k\tilde{\psi}_k^2\, a^2\sin\theta d\theta d\phi \\ &= \text{KE} + \sum_{mnk}\lambda_k\tilde{\psi}^2_{mnk} \\ &= \sum_{mnk}\left(n(n+1)+\lambda_k\right)\tilde{\psi}^2_{mnk}, \end{aligned} \tag{3.19}$$

where $\tilde{\psi}_k$ is the k-th baroclinic component in grid space and $\tilde{\psi}_{mnk}$ is the coefficient at point (m,n) on the spectral grid of the k-th baroclinic component. Since these expressions must satisfy the inner product condition of the energy matrix given in (3.14), then it is clear that the spectral energy matrices take the form of the diagonal matrix shown in Figure 3.7(a). Here, the D_i^E and D_i^O blocks are square matrices whose diagonal entries depend upon which norm is being used. The D_i^E blocks are for the even spectral entries of the i-th baroclinic component (i=0 is the barotropic and i=(1,2,3) are the first, second and third baroclinic components respectively) and the D_i^O blocks for the odd entries.

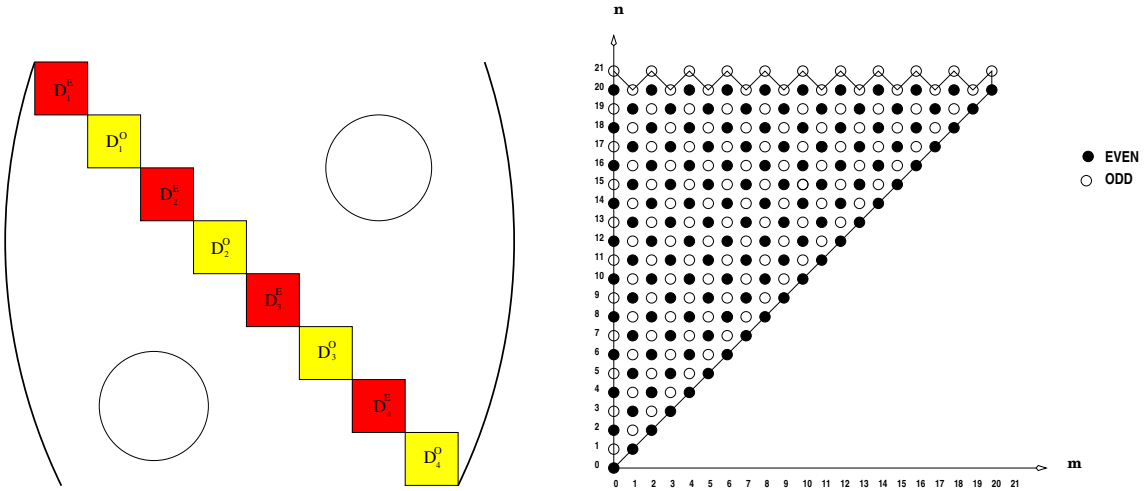
The diagonal blocks of the even spectral entries for the KE and TE matrices (which are shown in Figure 3.7(a)) are as follows:

$$\begin{aligned} \boxed{\textbf{KE blocks}}\quad \mathbf{D}_i^E &= \mathbf{diag}\left[n_1(n_1+1),\ n_2(n_2+1),\ldots,n_s(n_s+1)\right], \\ \boxed{\textbf{TE blocks}}\quad \mathbf{D}_i^E &= \mathbf{diag}\left[n_1(n_1+1)+\bar{\lambda}_i,\ n_2(n_2+1)+\bar{\lambda}_i,\ldots,n_s(n_s+1)+\bar{\lambda}_i\right], \end{aligned}$$

denoting n_i as the total wavenumber (the n value) of the i-th even spectral coefficient and where the subscript s is defined as

$$s = \frac{n(m+1)+\max(m,n)+1}{2}, \tag{3.20}$$

reading the coefficients in order of increasing n with increasing m as shown in Figure 3.7(b).



(a) The form of the square energy matrices. The red blocks (D_i^E) represent the diagonal matrices for the even spectral entries and the yellow (D_i^O) for the odd.

(b) The spectral grid for a jagged T21 triangular truncation. Coefficients are read in order of increasing n with increasing m. The jagged T21 truncation has 121 odd and even complex data points.

Figure 3.7: The form of the diagonal energy matrix is shown in (a) and the jagged T21 triangular truncation grid is shown in (b). The diagonal entries of matrices D_i^E and D_i^O depend upon the ordering of the n values (the total wavenumbers) in the T21 spectral grid.

Here $\bar{\lambda}_i$ corresponds to the i-th Froude number obtained from the vertical structure equation in (2.5). Similar blocks are constructed for the odd spectral entries D_i^O.

3.3.8 Symmetrising the eigenvector problem

When implementing the diagonal matrix the correlation matrix is no longer symmetric. It is necessary to diagonalise matrix **CM** in (3.16) to form a symmetric matrix $\mathbf{D} = \mathbf{M}^{\frac{1}{2}}\mathbf{C}\mathbf{M}^{\frac{1}{2}}$ (where $\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}} = \mathbf{M}$) which has the same eigenvalues as **CM**. So we now solve

$$\mathbf{D}\mathbf{E}' = \lambda\mathbf{E}', \tag{3.21}$$

where the eigenvectors $\mathbf{E}'$ of **D** are related to eigenvectors **E** of **CM** by

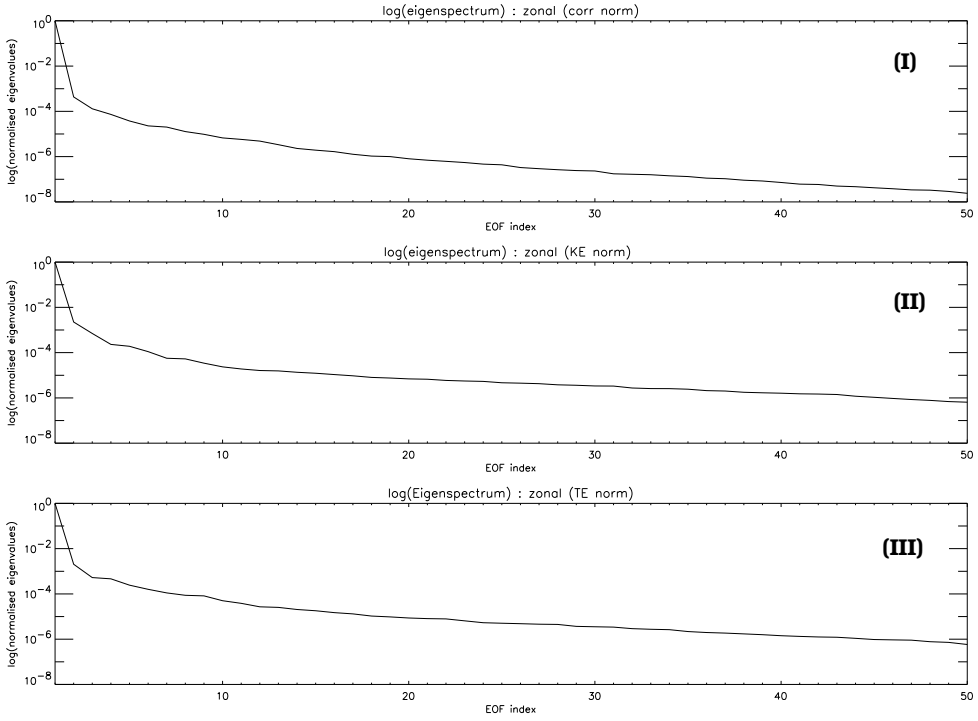
$$\mathbf{E}' = \mathbf{M}^{\frac{1}{2}}\mathbf{E}, \tag{3.22}$$

since $\mathbf{M}^{\frac{1}{2}}\mathbf{C}\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{U} = \lambda\mathbf{M}^{\frac{1}{2}}\mathbf{U}$. Once the energy POD modes **E** are found, we use $\mathbf{E} = \mathbf{M}^{-\frac{1}{2}}\mathbf{E}'$ to return to Euclidean spectral space, then the Fast Fourier Transform (FFT) to transform the spatial modes to physical grid space and so allow direct comparisons to be made between the qualitative structures of each of the modes obtained from the various norms.

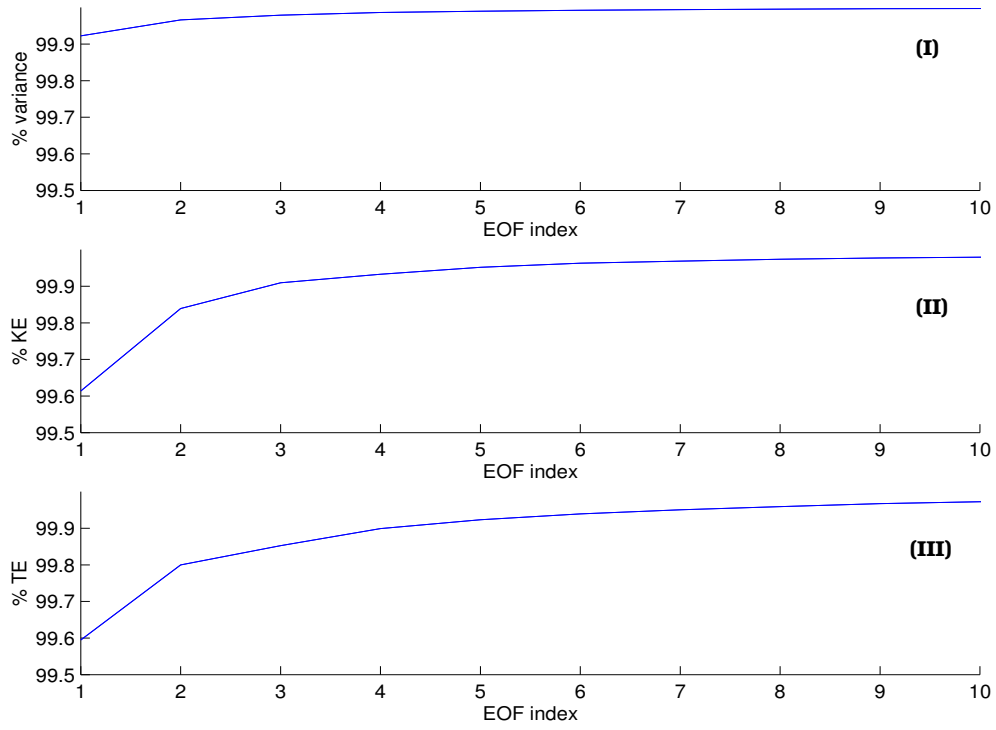
3.3.9 SGCM Eigenspectrum

Figures 3.8 and 3.9 respectively illustrate the eigenspectra[1] of the zonal and the departure fields using the correlation, KE and TE norms, where each logarithmic profile of the zonal and

[1]The eigenvalues are normalised so that their sums equal 1.

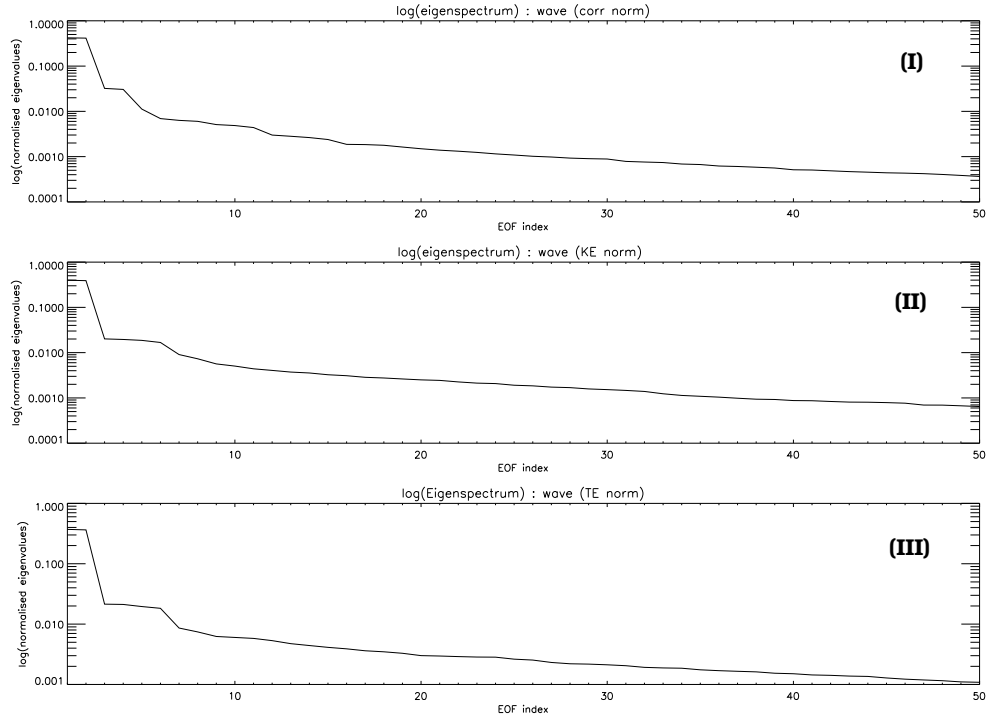


(a) The logarithmic eigenspectra zonal profiles using the (I) correlation, (II) KE and (III) TE norms.

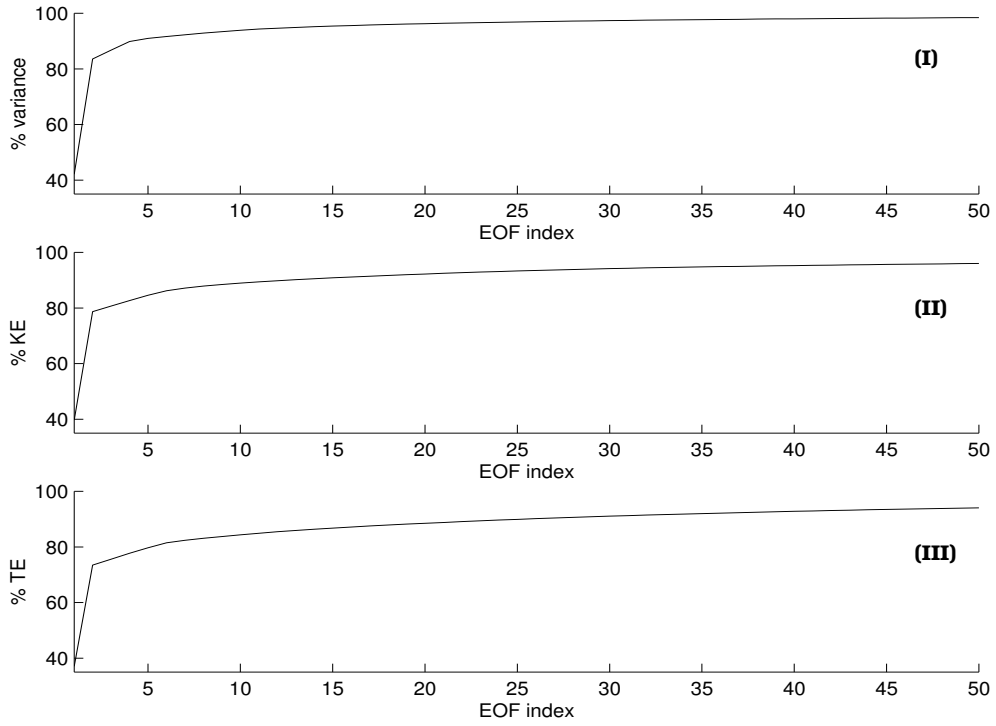


(b) The cumulative (I) variance, (II) KE and (III) TE capture of the zonal field using the (I) correlation, (II) KE and (III) TE norms respectively.

Figure 3.8: The (a) logarithmic and (b) cumulative eigenspectra of the SGCM zonal flow using the (I) correlation, (II) KE and (III) TE norms.



(a) The logarithmic eigenspectra departure profiles using the (I) correlation, (II) KE and (III) TE norms.



(b) The cumulative (I) variance, (II) KE and (III) TE capture of the departure field using the (I) correlation, (II) KE and (III) TE norms respectively.

Figure 3.9: The (a) logarithmic and (b) cumulative eigenspectra of the SGCM departure flow using the (I) correlation, (II) KE and (III) TE norms.

departure fields displays a rapid decrease in the magnitude of the eigenvalues.

Figure 3.8(b) shows that the first eigenvalue of the zonal field dominates the flow and captures 99.92% of the variance (in I), 99.61% of KE (in II) and 99.60% of TE (in III) if the correlation, KE or TE norms respectively are used. The fall-off rate of the correlation zonal eigenspectrum is faster than that observed in the energy eigenspectra, confirmed by the behaviour of the logarithmic spectra profiles in Figure 3.8(a) which are illustrated in (I), (II) and (III).

Figure 3.9(b) shows that the two leading eigenvalues of the departure flow represent 84% of the variance (in I), 79% of KE (in II) and 73% of TE (in III) of the wave flow if the correlation, KE or TE norms respectively are used. The logarithmic departure eigenspectra profiles illustrated in Figure 3.9(a) reveal that the first four wave modes appear in pairs. The dominance of the leading four eigenmodes suggests that the original SGCM system could be governed by two wave pairs (i.e. a complex bimodal system).

Norm	90 %	95 %	4 modes	10 modes	50 modes
Correlation	5 modes	14 modes	89.83 %	93.88 %	98.44 %
KE	13 modes	38 modes	82.68 %	88.94 %	96.01 %
TE	26 modes	56 modes	77.72 %	84.34 %	94.07 %

Table 3.3: POD modal truncation against variance/energy capture in the departure flow.

Table 3.3 shows that the subspaces spanned by 10 correlation, KE and TE eigenvectors respectively contain 93.88 %, 88.94 % and 84.34 % of the variance, KE and TE of the the wave flow. Only 5 correlation modes are required to capture 90 % of the variance whereas 13 KE or 26 TE eigenvectors are needed if 90 % of the KE or 90 % of the TE respectively is to be retained. If an additional 5 % of the variance, KE or TE of the system is sought then an extra 9, 25 or 30 correlation, KE and TE eigenvectors are required. This illustrates that the correlation eigenspectrum has a faster fall-off rate and this can be seen in the logarithmic eigenspectra in Figure 3.9(a).

A sudden decrease from large to small eigenvalues with increasing EOF index offers a natural criterion for choosing how many patterns are needed to be retained in the description. It can be seen in Figure 3.9(a), for example, that a sudden decrease by a factor of 100 exists beyond wave pattern 2. Therefore, only two wave modes would be sufficient to describe the model dynamics to the first order (for further details see Kantz and Schreiber (1997)). This may be true for short-term predictions, but in order to reasonably reproduce the original system many more components may need to be included.

The results indicate that a small number of PODs ($\approx$ 10-20) are required to meet the variance threshold (say $>$ 90 %). It may be possible to construct a low-order dynamical system (of dimension 10-20) that mimics the spatial and temporal behaviour of the entire SGCM flow field, as well as possibly providing useful insight into the number of important degrees of freedom in the SGCM.

3.3.10 SGCM POD modes

Figures 3.10 - 3.11 illustrate the horizontal structures of the barotropic components of the various spatial correlation modes of the departure flow. Here, modes 1 and 2 represent a steady wavenumber 3 in the SH, which are $\frac{\pi}{2}$ out of phase. Modes 3 and 4 correspond to a steady wavenumber 2 and mode 5 depicts a mean zonal flow. It is also evident that a wavenumber 6 structure exists in modes 6, 8 and 10 (see Figure 3.11).

Figure 3.10 illustrates that modes 1 and 2, 3 and 4 form conjugate pairs. However, the choice of norm affects the ordering of the modes and this is illustrated in Table 3.4 which shows the zonal wavenumber of the barotropic components of each of the POD modes for each of the norms, where it can be seen that the first pair of spatial modes form a conjugate pair, although the ordering of the remaining modes is norm dependent.

MODES	CORR	KE	TE
1	3	3	3
2	3	3	3
3	2	2	2
4	2	6	6
5	zonal flow	2	2
6	6	6	6
7	4	4	4
8	6	4	4
9	4	3	4
10	4	3	4

Table 3.4: Zonal wavenumbers associated with each of the POD modes.

Figures 3.12 - 3.17 illustrate the barotropic and baroclinic components of the SGCM departure POD modes. We will now compare the various patterns produced by the correlation, KE and TE norms, examining the barotropic and baroclinic components of each of the modes. For convenience, only the odd departure eigenmodes will be discussed here (the positive contour lines are shaded in red, the negative contours in blue and the zero contour lines in green).

Modes 1 and 2 form a conjugate pair for each of the norms and thus denote a travelling wavenumber 3 in the longitudinal direction (i.e. a wave which has a varying phase surface in time). The structure of mode 1 in each case is unaffected by the choice of norm (seen in Figures 3.12, 3.14 and 3.16), containing a strong wavenumber 3 component centred at a latitude of about 45^o in the SH. The barotropic components are almost sinusoidal in longitude. The first baroclinic components comprise two parallel wavenumber 3 structures, centred at latitudes of 45^o and 15^o in the SH, displaced by 50^o in longitude. There exists additional wavenumber 3 patterns centred at 20^o in the NH. The second baroclinic components also contains similar parallel wavenumber 3 patterns, centred at latitudes of 45^o and 25^o in the SH, displaced by 25^o in longitude with a strong wavenumber 3 structure at 20^o in the NH. The third baroclinic modes are almost symmetric about the equator, with tilted structures in the NH. A sinusoidal wavenumber 3 pattern exists at 20^o in the NH, besides a further two wavenumber 3 patterns at latitudes of 15^o and 50^o in the SH, where the latter wave pattern is displaced by 50^o in longitude.

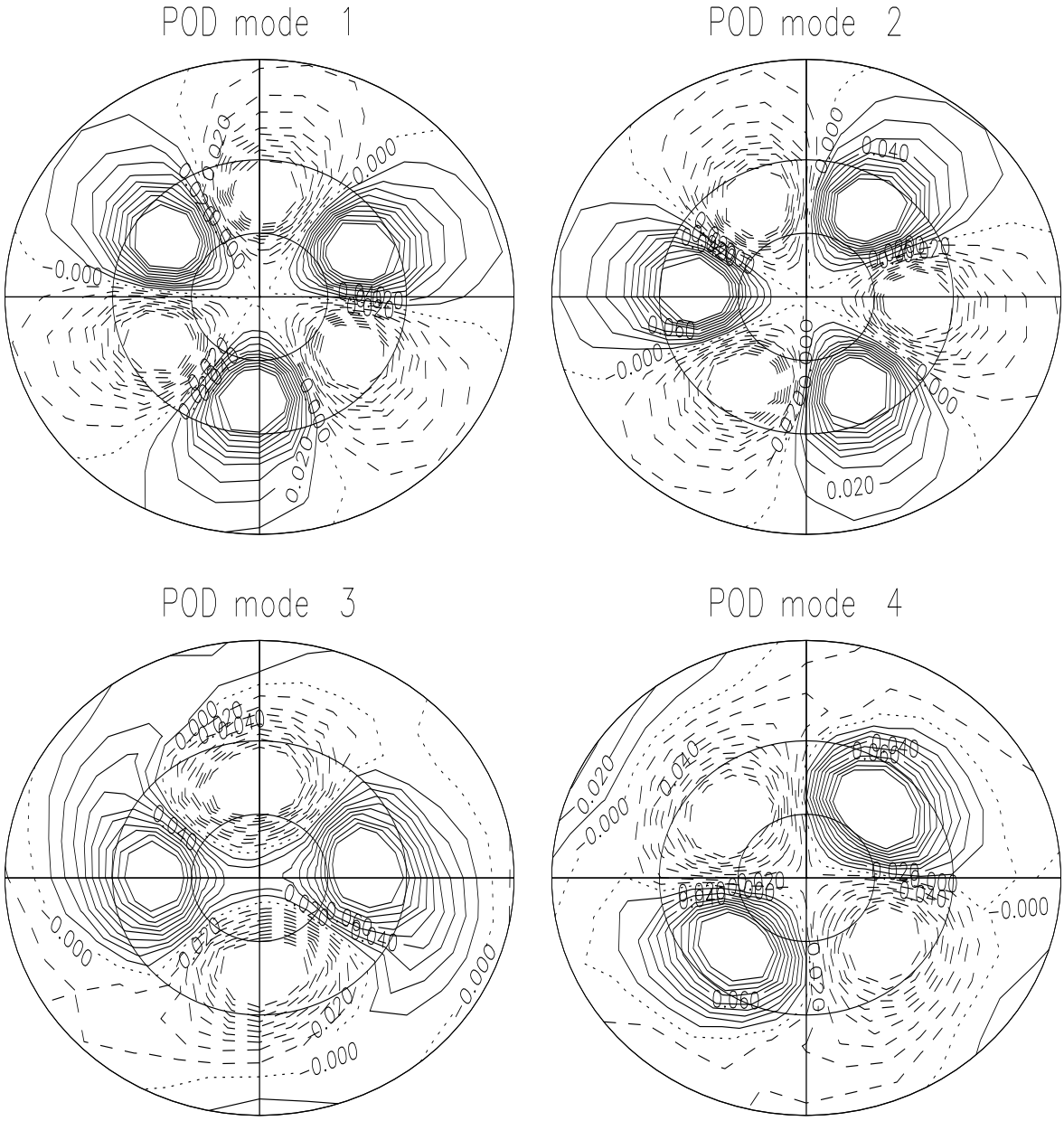


Figure 3.10: Southern hemisphere stereographic projection plots of the first four SGCM POD modes of the barotropic component obtained using the correlation norm.

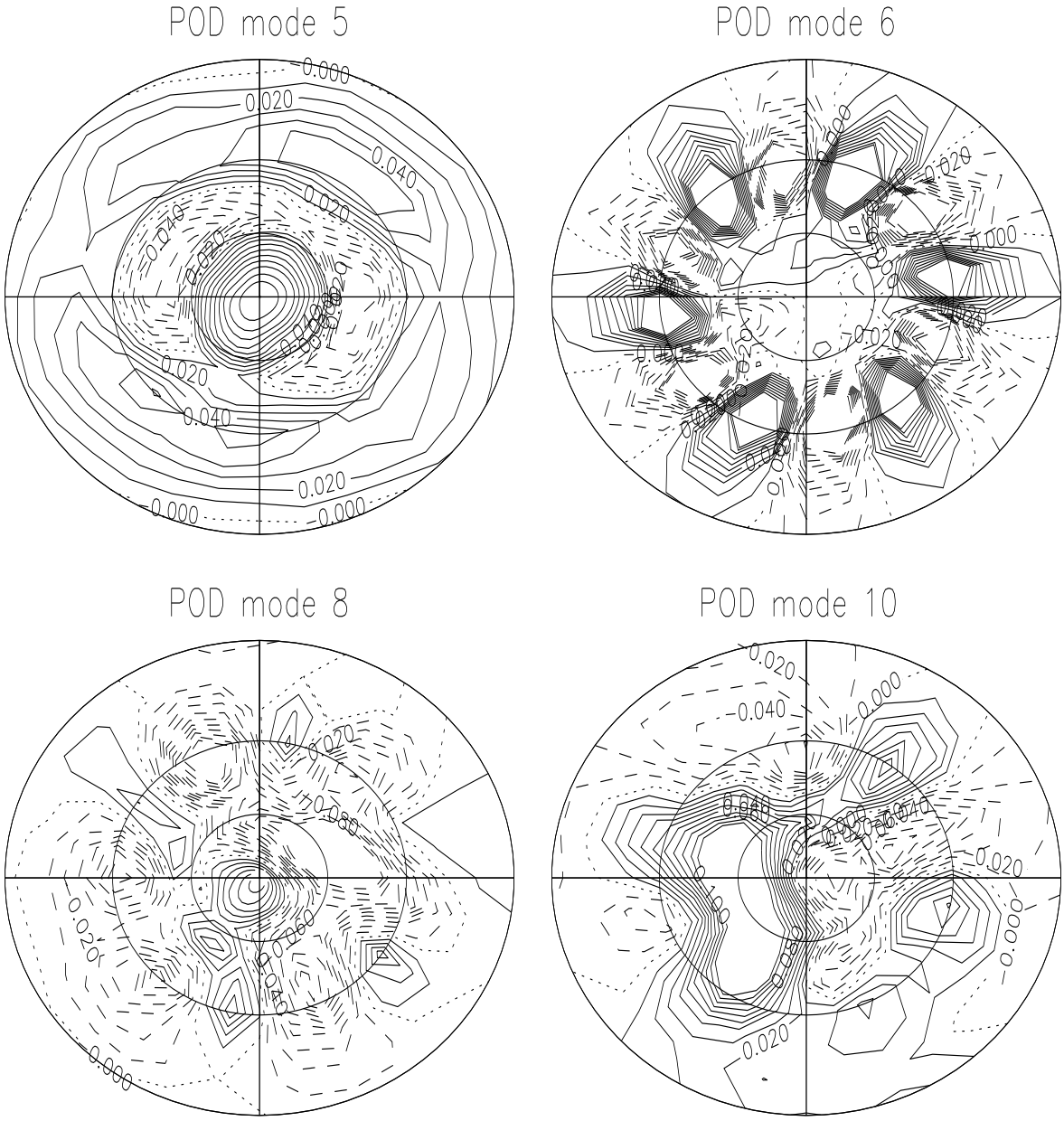


Figure 3.11: Southern hemisphere stereographic projection plots of SGCM POD modes 5, 6, 8 and 10 of the barotropic component obtained using the correlation norm.

Mode 3 (Figures 3.12, 3.14 and 3.16) contains a strong wavenumber 2 pattern in each of the vertical components. If the correlation or the KE norm is used then the resulting barotropic component appears to be almost sinusoidal in form (Figures 3.12 and 3.14), but somewhat distorted if the TE norm is selected (Figure 3.16). Similarly, the first and second baroclinic components of the correlation and KE modes contain a latitudinally stretched wavenumber 2 pattern, together with a small wave number 3 structure centred at a latitude of 20^o in the NH, which is displaced by 100^o in longitude. However, the TE norm mode contains severely modulated wavenumber 2 structures, centred at a latitude of 25^o in the SH. If the third baroclinic components are examined it is observed that the correlation and KE norms produce very similar structures which are almost symmetric about the equator. Both wave patterns are centred at latitudes $\pm 45^o$, with a thin band of distortion about the equator (almost a condensed wavenumber 3 pattern on the 0^o latitude line). However, the TE norm produces a sinusoidal wavenumber 6 pattern at a latitude of 25^o in the SH. Clearly, correlation modes 3 and 4 form a conjugate pair whereas energy modes 3 and 5 form their respective pairings. Since correlation mode 5 (in Figure 3.13) describes a departure from the mean flow, this suggests that the choice of norm may significantly effect the individual ordering of the modes.

Mode 7 contains a wave number 4 structure (Figures 3.13, 3.15 and 3.17). The energy norms retrieve an almost sinusoidal pattern with dominant coupled features, whereas the correlation norm creates several complicated structures. The barotropic component (for each of the norms) has a wavenumber 4 structure, centred at a latitude of 35^o in the SH, where the energy norms retrieve a more regular spatial pattern. The first barotropic components contain two parallel wavenumber 4 structures, centred at latitudes of 45^o and 15^o in the SH, displaced by 25^o in the longitude, where the energy norm modes contain more regular structures. The second baroclinic components contain two parallel wave number 4 structures, centred at latitudes of 50^o and 25^o in the SH, with a large wavenumber 1 structure in the NH, centred at a latitude of 30^o. The third baroclinic component of the KE and TE mode 7 also contain wavenumber 4 structures, centred at latitudes of 45^o and 15^o in the SH, displaced by 25^o in longitude, whereas the same component of correlation mode 7 contains a wavenumber 6 structure centred at a latitude of 25^o in the SH. It should be noted that both energy modes 7 and 8 contain dominant zonal wavenumber 4 structures whereas the correlation norm modes 7 and 8 respectively contain zonal wavenumber 4 and 6 structures.

The correlation norm generates wavenumber 4 patterns in the barotropic and first baroclinic components of mode 9 (Figure 3.13), though the second and third baroclinic components contain wave number 6 structures centred at a latitude of 25^o in the SH. The barotropic component of KE mode 9 (Figure 3.15) contains two parallel wavenumber 3 structures centred at latitudes of 50^o and 25^o in the SH, displaced by 50^o in longitude. However, the TE norm extracts a wavenumber 4 structure centred at a latitude of 45^o in the SH. The first baroclinic components of the KE and TE mode 9 are tilted wavenumber 3 patterns centred at a latitude of 35^o in the SH, with an additional wavenumber 3 structure centred at a latitude of 20^o in the NH, displaced by 25^o in longitude. The energy norms produce wavenumber 3 patterns in the second baroclinic component, centred at latitudes of 45^o in the SH and 0^o (the equator),

displaced by 15^o in longitude. The third baroclinic component of the energy modes contain three wavenumber 3 patterns centred at latitudes of 50, 25^o in the SH and 20^o in the NH.

Summary

Modal similarities are a feature of the lower-order modes, where the correlation and energy norms create modes which are almost identical in amplitude and form. Structural differences are visible if the higher-order modes are compared. It has been shown that the energy norms tend to extract symmetrical patterns and often produce small-scale spatial structures.

It has been shown that the use of the different norms affect the ordering of the POD modes. It is clear that modes 1 and 2 (for each of the norms) form a complex conjugate pair and thus depict a travelling wavenumber 3 structure, whereas modes 3 and 4 for the correlation norm and modes 3 and 5 for the energy norms together depict a travelling wavenumber 2 structure. Hence two distinct travelling waves can be observed from the leading order POD modes. Clearly, the use of the various norms has affected the ordering of the spatial eigenmodes, and so direct comparisons between modes at the same truncation level is misleading. An understanding of how many modes form complex conjugate pairs (for a given truncation level) is crucial so that it is known how many travelling waves are retained in a given truncation level.

It is difficult to determine how the higher order spatial POD modes evolve in time, so to provide us with additional information regarding the behaviour of the spatial patterns we will study their temporal behaviour by analysing the principal components of the flow.

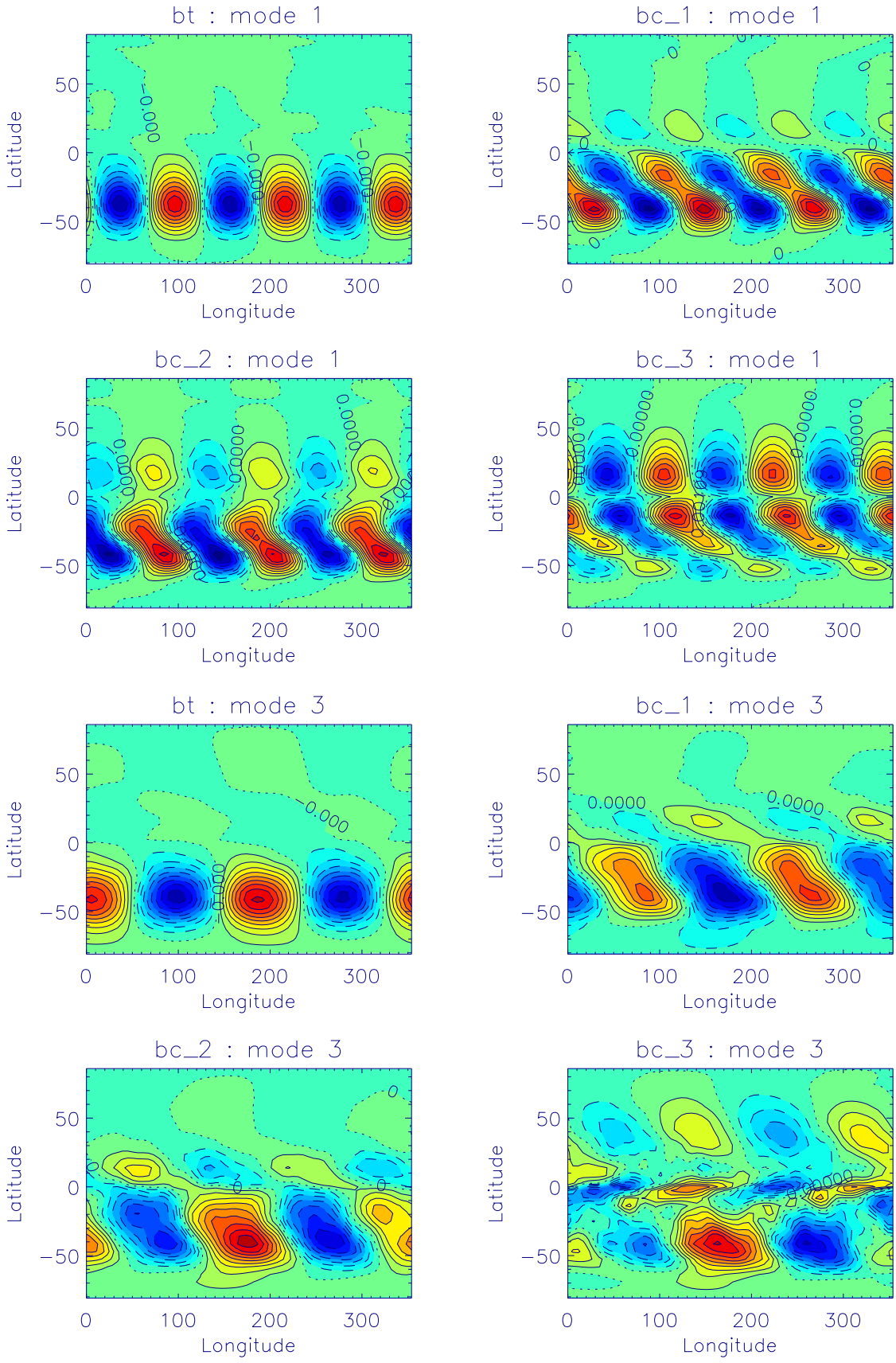


Figure 3.12: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the correlation norm.

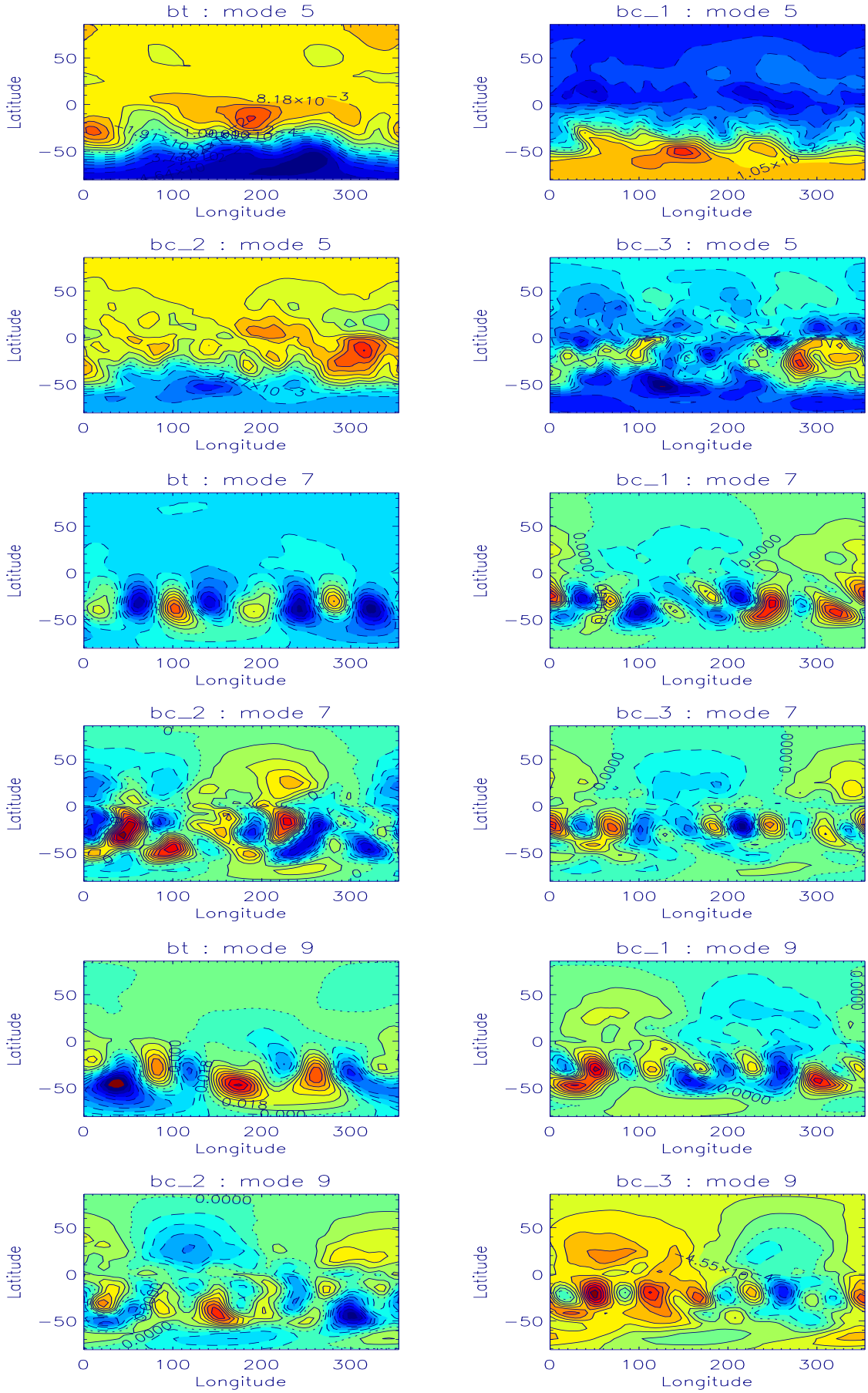


Figure 3.13: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the correlation norm.

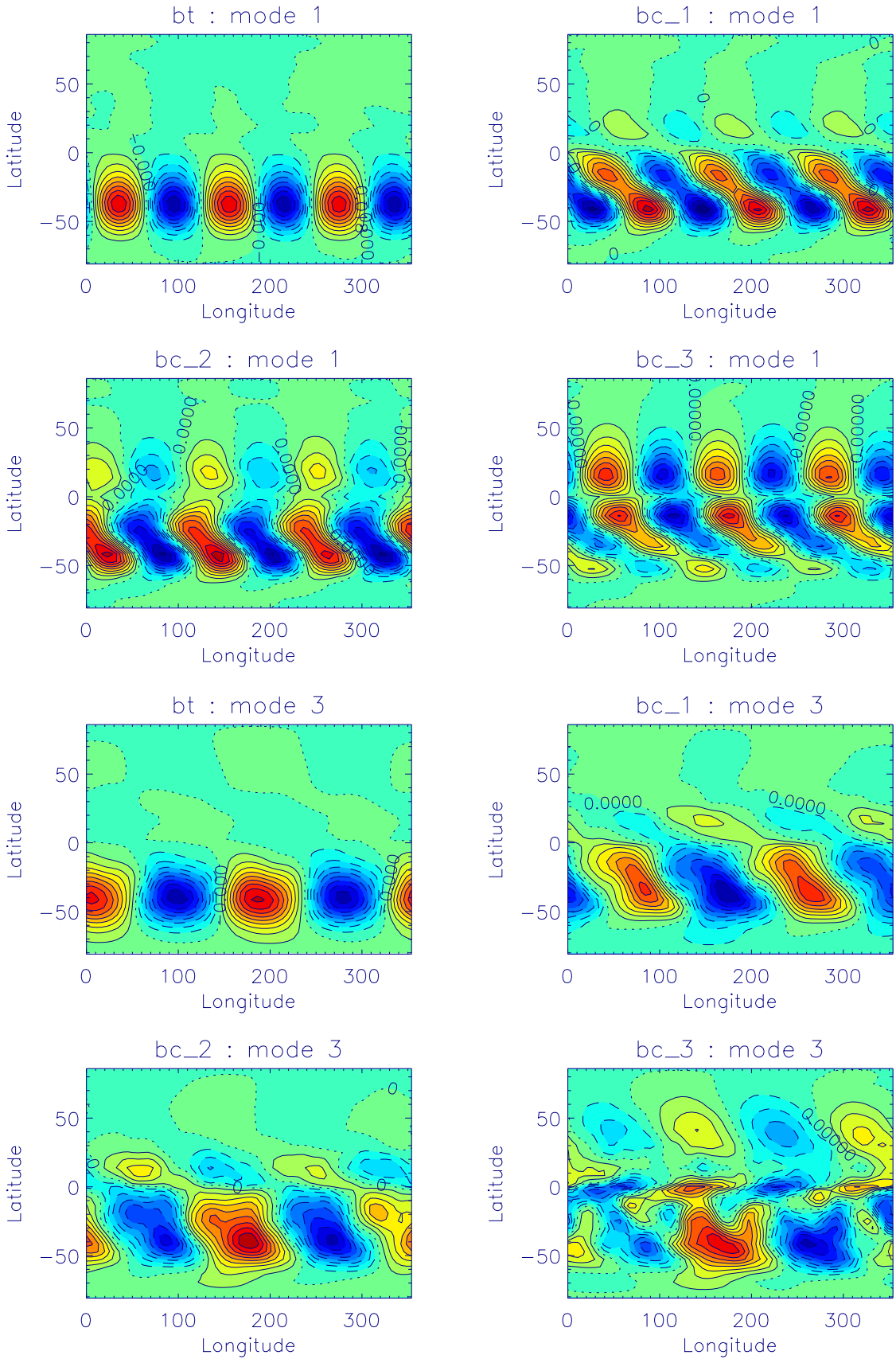


Figure 3.14: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the KE norm.

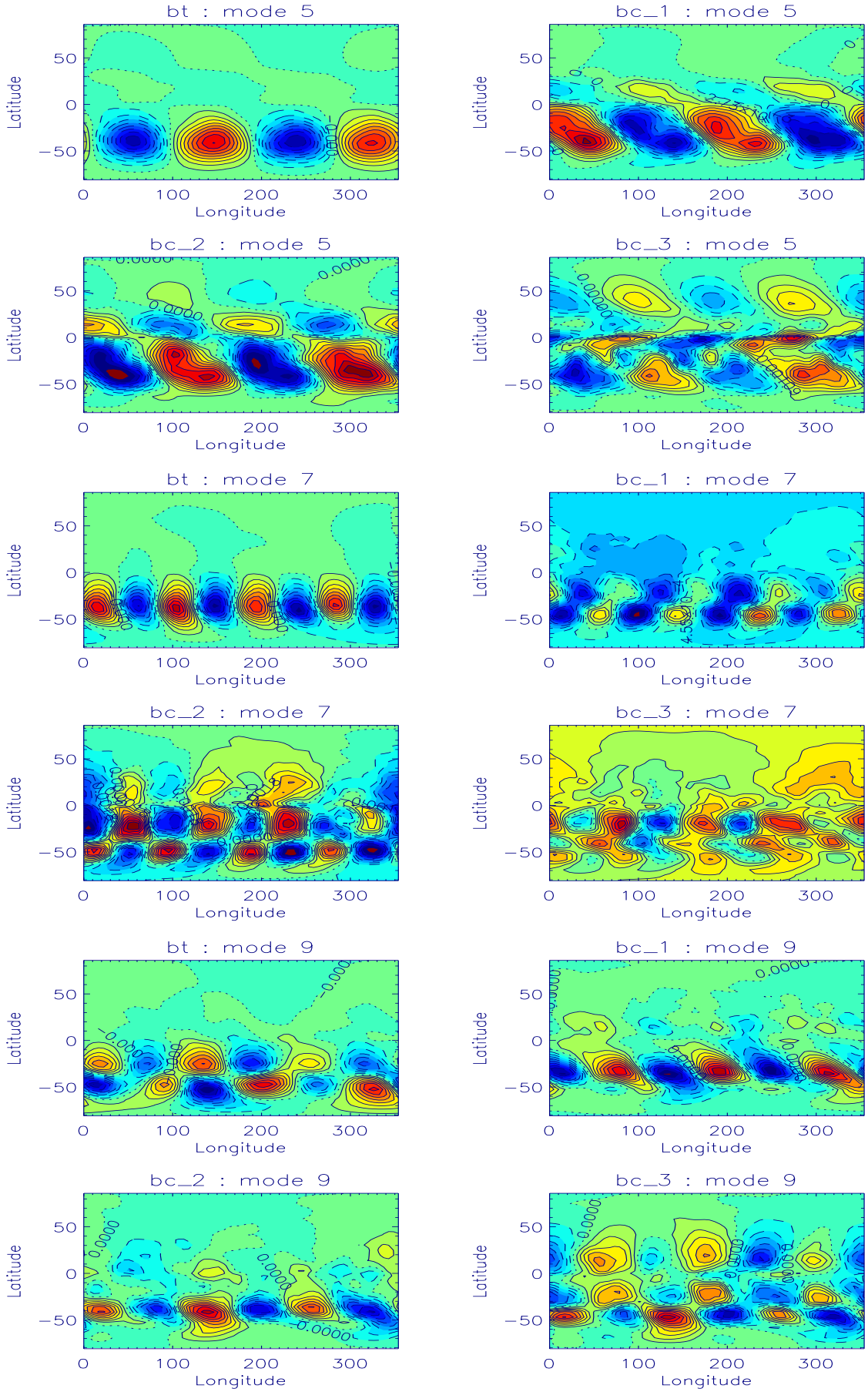


Figure 3.15: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the KE norm.

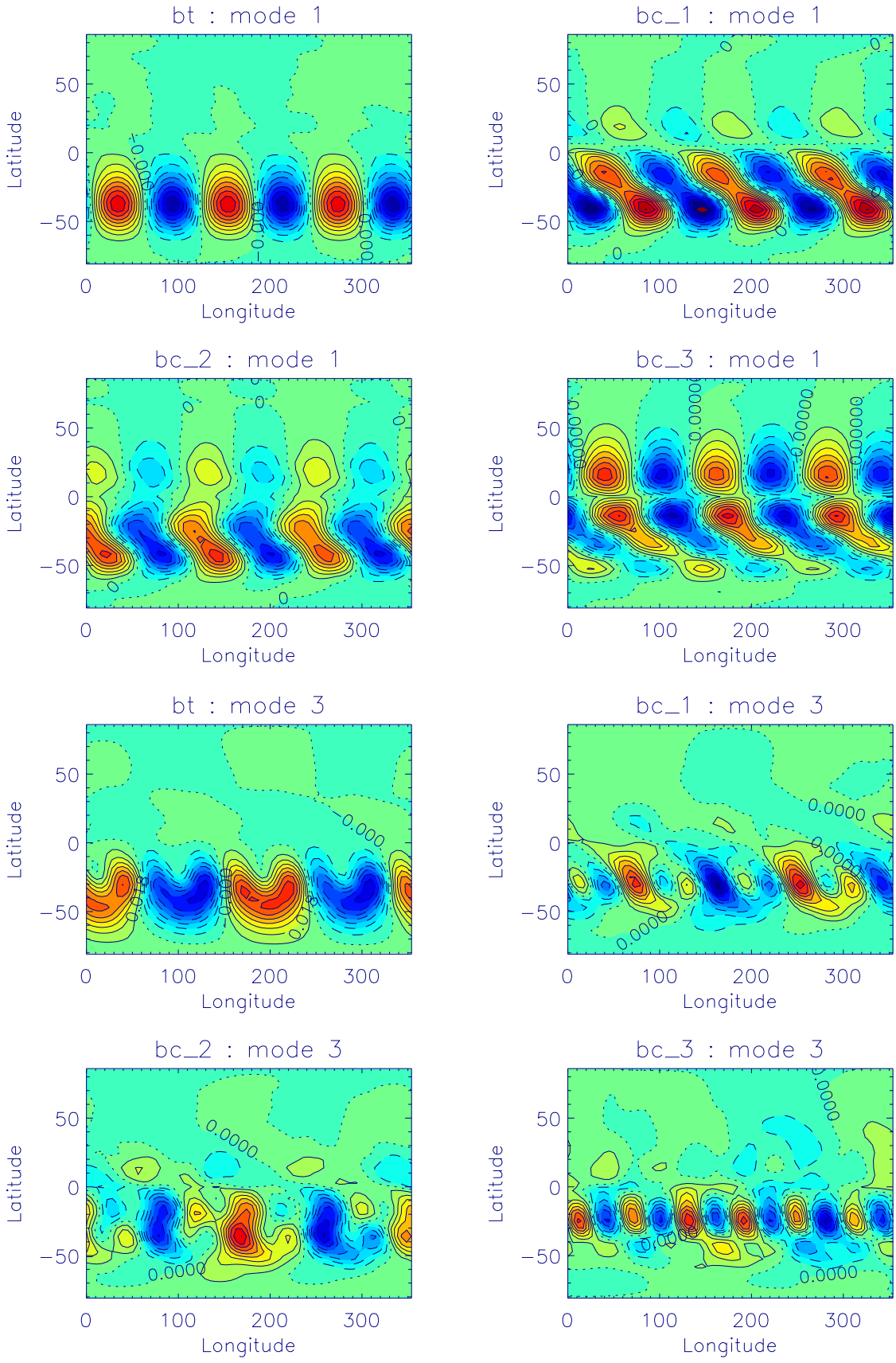


Figure 3.16: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the TE norm.

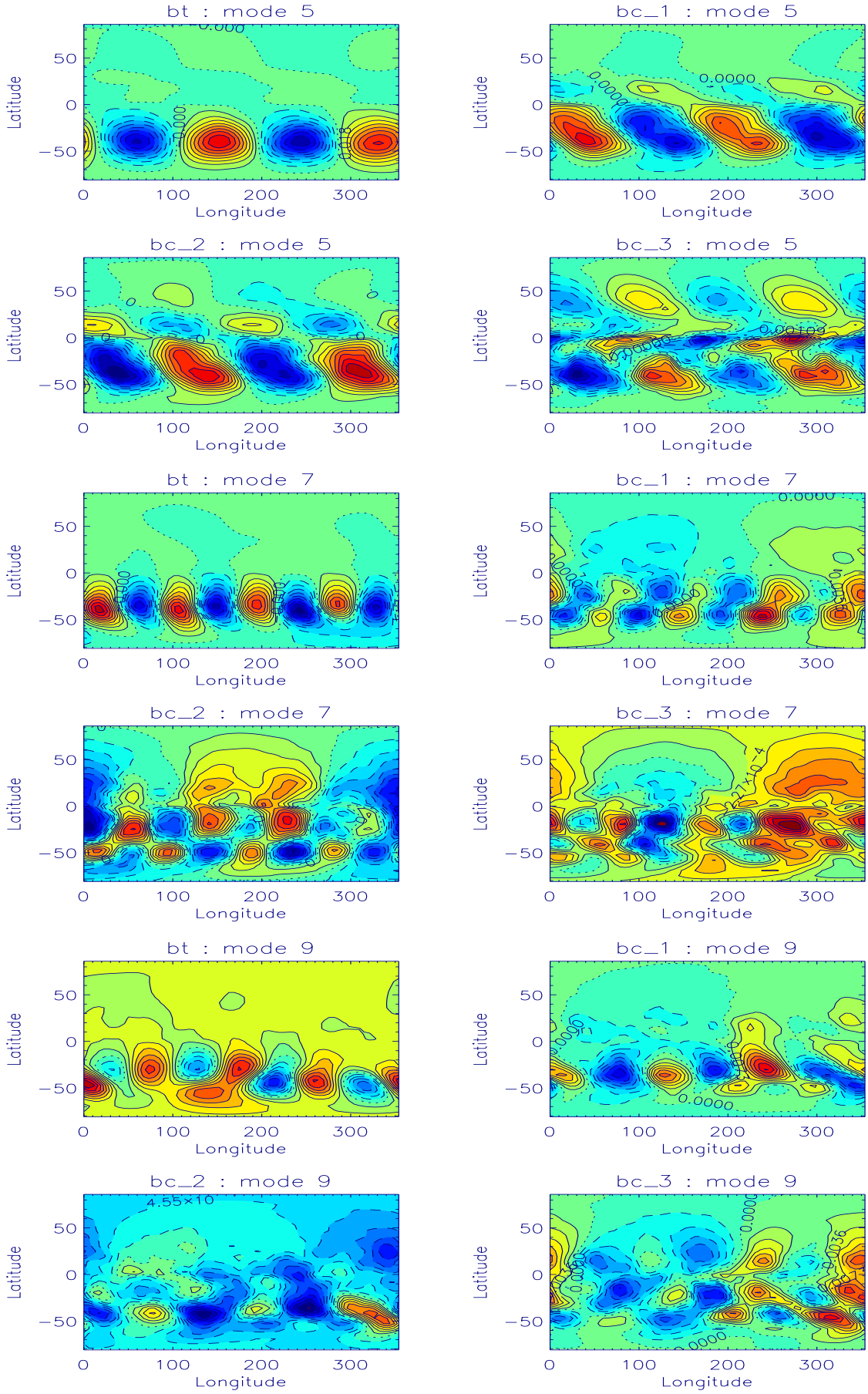


Figure 3.17: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the TE norm.

3.3.11 Power Spectra Analysis

One common way of picking out signals in numerical data is to find the Fourier transform or frequency spectrum of the data. The objective is typically to detect the components of the data with particular frequencies, or ranges of frequencies.

Instead of describing the statistical properties of a signal in real space its properties in Fourier space will be analysed. The Fourier transform establishes a one-to-one correspondence between the signal at certain times (time domain), how certain frequencies contribute to the signal, and how the phases of oscillations are related to the phases of other oscillations (frequency domain).

The Fourier transform of a function $s(t)$ is given by

$$\tilde{s}(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(t) e^{2\pi i f t} dt, \tag{3.23}$$

and that of a finite, discrete time series by

$$\tilde{s}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_n e^{2\pi i k n/N}. \tag{3.24}$$

Here, the frequencies in physical units are $f_k = k/N\delta t$, where $k = -N/2, ..., N/2$ and δt is the sampling interval. Both operations are invertible and are linear transformations.

The power spectrum is defined as the squared modulus of the continuous Fourier transform, $S^s(f) = |\tilde{s}(f)|^2$. It is the square of the amplitude, by which the frequency f contributes to the signal. If an estimate is to be made from a finite discrete series, we can use the periodogram, $S_k^{(s)} = |\tilde{s}_k|^2$.

The power spectrum is particularly useful for studying oscillations in a system. There will be sharper or broader peaks at the dominant frequencies and their integer multiples, or harmonics. Purely periodic or quasi-periodic signals will show sharp spectral lines. (For further details see Kantz and Schreiber (1997)).

The Fast Fourier Transform (FFT) was implemented by the use of MATLAB (which had the FFT algorithm as a signal processing function[2]). Using $\mathrm{fft}(\mathbf{x})$ to denote the discrete Fourier transform[3] of vector $\mathbf{x}$ of length N then

$$\mathrm{fft}(\mathbf{x}) = \sum_{j=1}^{N} \mathbf{x}(j) \omega_N^{(j-1)(k-1)}, \tag{3.25}$$

and the inverse transform function is

$$\mathrm{ifft}(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^{N} \mathrm{fft}(\mathbf{x}) \omega_N^{-(j-1)(k-1)}, \tag{3.26}$$

[2]The Fast Fourier Transform is used by the commands fft and the inverse ifft in Matlab.

[3]The FFT is computed with a radix-2 if the length of $\mathbf{x}$ is a power of 2, and with a mixed radix algorithm if the length is not.

where $\omega_N = e^{-2\pi i/N}$ is an N-th root of unity.

By looking at the absolute values of the Fourier data the frequencies of the original signal can be identified.

3.3.12 SGCM Principal Components and Frequency Spectra

If $\mathbf{x}_j$ is taken to be a window in the time series then it can be expanded as a linear combination of the eigenvectors such that:

$$\mathbf{x}_j = \sum_{i=1}^{m} a_{ij}\mathbf{V}_i, \tag{3.27}$$

where a_{ij} is the principal component (PC) of the i-th eigenvector at position j in the time series. The PCs represent the amplitude of the eigenvectors at a particular time. When the eigenvectors form an oscillating pair, then the square root of the sum of the squares of the two PCs (in the L^2 norm) represent the amplitude of the oscillation at any time.

Figure 3.18 shows the temporal behaviour of the departure spatial eigenmodes for the different norms over 300 Martian days (a Martian day is 24h 37 m). The structures are wave-like, with the two leading eigenvectors forming pairs that have almost the same shape but are displaced one quarter of a period out of phase. This indicates that the pair explain a recurrent oscillation pattern in the time series (see Vautard et al. (1992)). There is a strong periodic behaviour in PCs 1 and 2, where roughly 14 oscillations occur every 60 Martian days and oscillate between the nondimensionalised values [-4,4]. The form of the leading oscillations is virtually unaffected by the choice of norm and evidently contain periodic structures. The higher order modes appear to be more erratic, particularly PCs 6 to 10 where high frequency, irregular structures can be observed. It must be remembered that the ordering of the spatial POD modes is affected by the choice of norm, and so comparisons between the various principal components must take this into account.

Figure 3.19 shows the original SGCM PCs for the correlation, KE and TE norms during a 60 Martian day interval. The leading two PCs form a complex conjugate pair, and are only weakly affected by the choice of norm as they are almost identical in shape and frequency. The power peaks are seen to cluster about a frequency of 0.23-0.27 cycles per sol (in Figure 3.20(a)) which corresponds to the spatial wavenumber 3 structures oscillating as a travelling wave at a rate of 14 to 16 cycles every 60 days.

PCs 3 and 4 in Figure 3.19 are shown to be far more distorted and erratic than the two leading PCs. It is observed that correlation PCs 3 and 4, and energy PCs 4 and 5 respectively form complex conjugate pairs, and so respectively represent travelling waves. Therefore, the travelling wavenumber 3 structures oscillate at a rate of about 5 oscillations during every 60 days which gives a frequency of about 0.08 cycles/sol. This is consistent with the frequency values picked out by the power spectra analysis in Figure 3.20(a), where sharp peaks have clustered about the 0.08-0.1 cycles/sol frequency band, corresponding to 5-6 oscillations every 60 days.

The correlation and TE norms extract similar wave patterns in PC 4, and this can be seen in the clustering of peaks about 0.1 cycles/sol in Figure 3.20(a). This is interesting as the spatial patterns contained in POD mode 4 are very different, the correlation norm mode contains a wavenumber 2 structure and the TE norm mode possesses a wavenumber 6. However, the wavenumber 6 structure contained in KE mode 4 has a very different temporal behaviour to that produced by the correlation and TE norms, where its PC 4 contains a much higher frequency oscillatory structure, which is small in amplitude and slightly fluctuates about the zero amplitude line, and has a cluster of frequencies about 0.5 cycles/sol seen in Figure 3.20(a).

As already stated, the KE and TE PCs 3 and 5 form complex conjugate pairs respectively and thus describe a travelling wave. This is confirmed by comparing the power spectra of PC 3 in Figure 3.20(a) with that which is observed in PC 5, which is shown in Figure 3.20(b). The energy norms produce oscillatory structures which have a dominant frequency of about 0.1 cycles/sol, whereas the zonal spatial flow contained in the correlation POD mode 5 has a temporal behaviour which is highly erratic, and is small in amplitude, having almost a negligible frequency of about 0.006 cycles/sol.

In Figure 3.19 it can be seen that PCs 6 to 10 are far more complicated and irregular than the leading five PCs. This is why the power spectra in Figure 3.20(b) reveal a range of interacting frequencies in the form of a broadband power spectrum, indicative of a chaotic evolution (see Bergé et al. (1984), p. 118).

Clearly from the power spectra in Figure 3.20 it can be seen that the energy modes 4 and 6 are not complex conjugate pairs, where there exists a concentration of peaks about the 0.45-0.5 cycles/sol band. This reveals that the wavenumber 6 spatial structures which were observed in the energy POD modes 4 and 6 do not depict travelling waves, but are instead standing waves i.e. waves which have a constant phase with respect to Mars. It should also be noted that since information propagates with the group velocity and not the phase speeds, propagation can still occur in these stationary waves (for further details see Andrews et al. (1987)).

The temporal behaviour of PC 7 is almost norm independent, where the power spectra peaks cluster about the 0.35-0.5 cycles/sol range. Similarly, for the KE and TE PC 8 there is a concentration of peaks about the same frequency range, with an additional cluster of peaks about 0.25 cycles/sol. The correlation PC 8 behaves differently, with large peaks occurring between 0.45-0.5 cycles/sol. Since both the energy POD modes 7 and 8 contain a spatial wavenumber 4 structure, then the PCs reveal that they are not complex conjugate pairs as their power spectra profiles are different.

The energy norms produce similar temporal behaviour in PCs 9 and 10, where a concentration of peaks are focused about a frequency of 0.25 cycles/sol. The correlation norm produces a floor of frequencies which is indicative of a chaotic evolution.

Figure 3.21 shows the temporal behaviour of the mean-flow correction modes. It can be seen

that PC 1 is almost insensitive to the choice of norm, but PCs 2 to 10 containing strikingly different oscillatory behaviour. This suggests that the zonal modes are more sensitive to the choice of norm than the departure PCs. A possible reason for this might be due to the fact that almost all of the TE of the flow is contained in the zonal mean (in excess of 90 %), and since the APE in the zonal flow accounts for almost 50 % of the TE then the maximisation of KE and TE (by use of the energy norms) produces significant differences in the spatial and temporal patterns.

Summary

Figure 3.20 shows the power spectra for the leading SGCM PCs using the correlation, KE and TE norms. As the leading travelling waves oscillate at a frequency of about 0.2 cycles/sol (i.e. the spatial wavenumber 3 structure) and 0.9 cycles/sol (i.e. the spatial wavenumber 2 structure), it might be then reasonable to presume that a similar behaviour can be duplicated by using a low-dimensional model. The fact that the analysis picks out two frequencies of oscillation implies that there could be some degree of weak interaction occurring between the leading order spatial modes, suggesting that it is crucial to retain a 'sufficient' amount of information regarding the behaviour of the two travelling spatial wave patterns if a low-dimensional model is to be obtained.

It has been seen that irrespective of norm, almost the same amount of information is contained in the ten leading eigenmodes, but the use of the different norms affects the orderings of the spatial and temporal structures. Thus the ordering of the modes is not robust. By looking at the temporal behaviour of the spatial POD modes it has enabled us to identify the travelling and standing waves contained in the leading eigenmodes. The results presented here have shown that the frequency distribution of the highly regular baroclinic wave activity in the idealised Martian atmosphere is nearly bimodal, i.e. governed by two leading travelling waves.

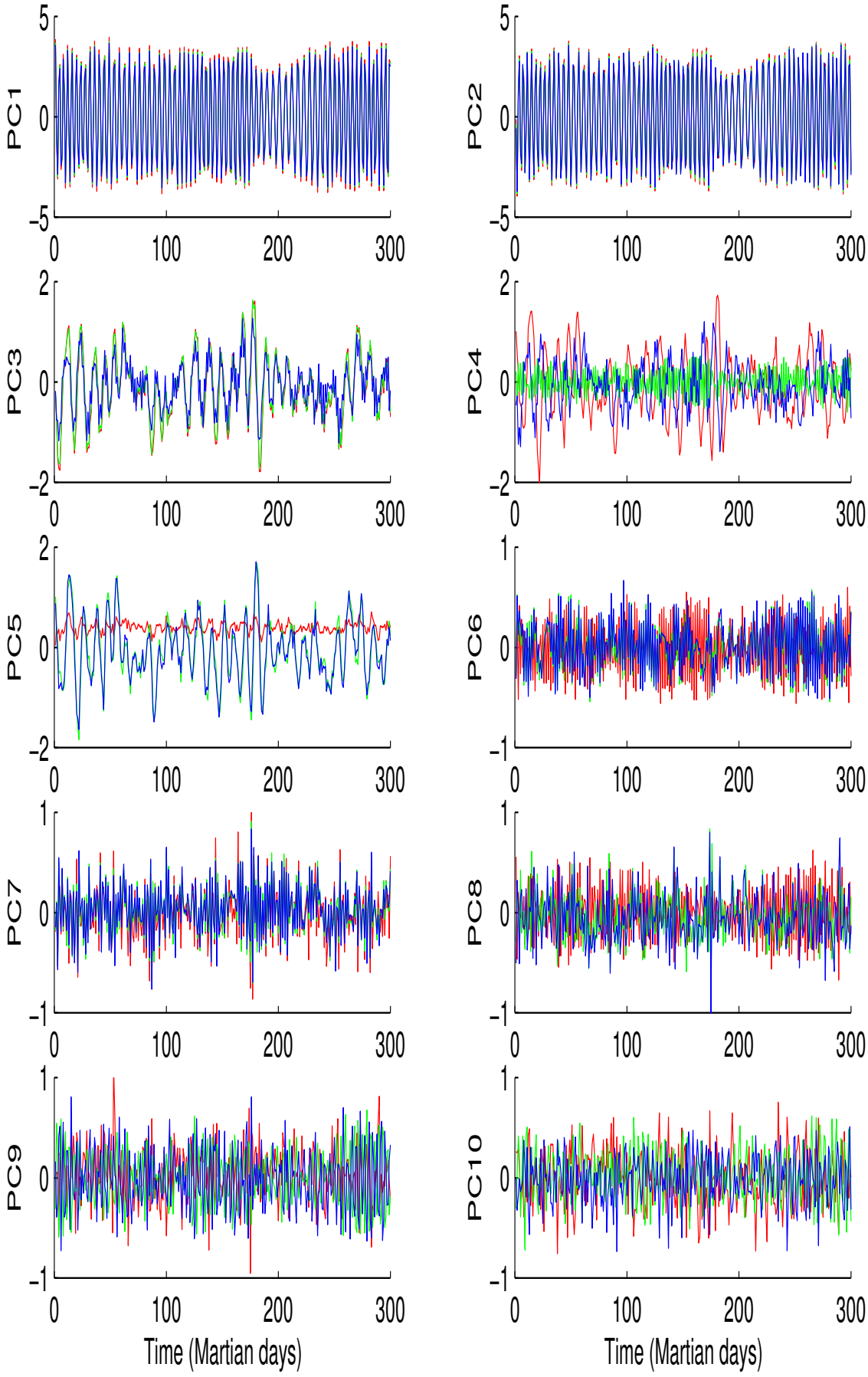


Figure 3.18: Original SGCM departure principal components (PCs) using the correlation (red), the KE (green) and the TE norms (blue) shown during an interval of 300 Martian days.

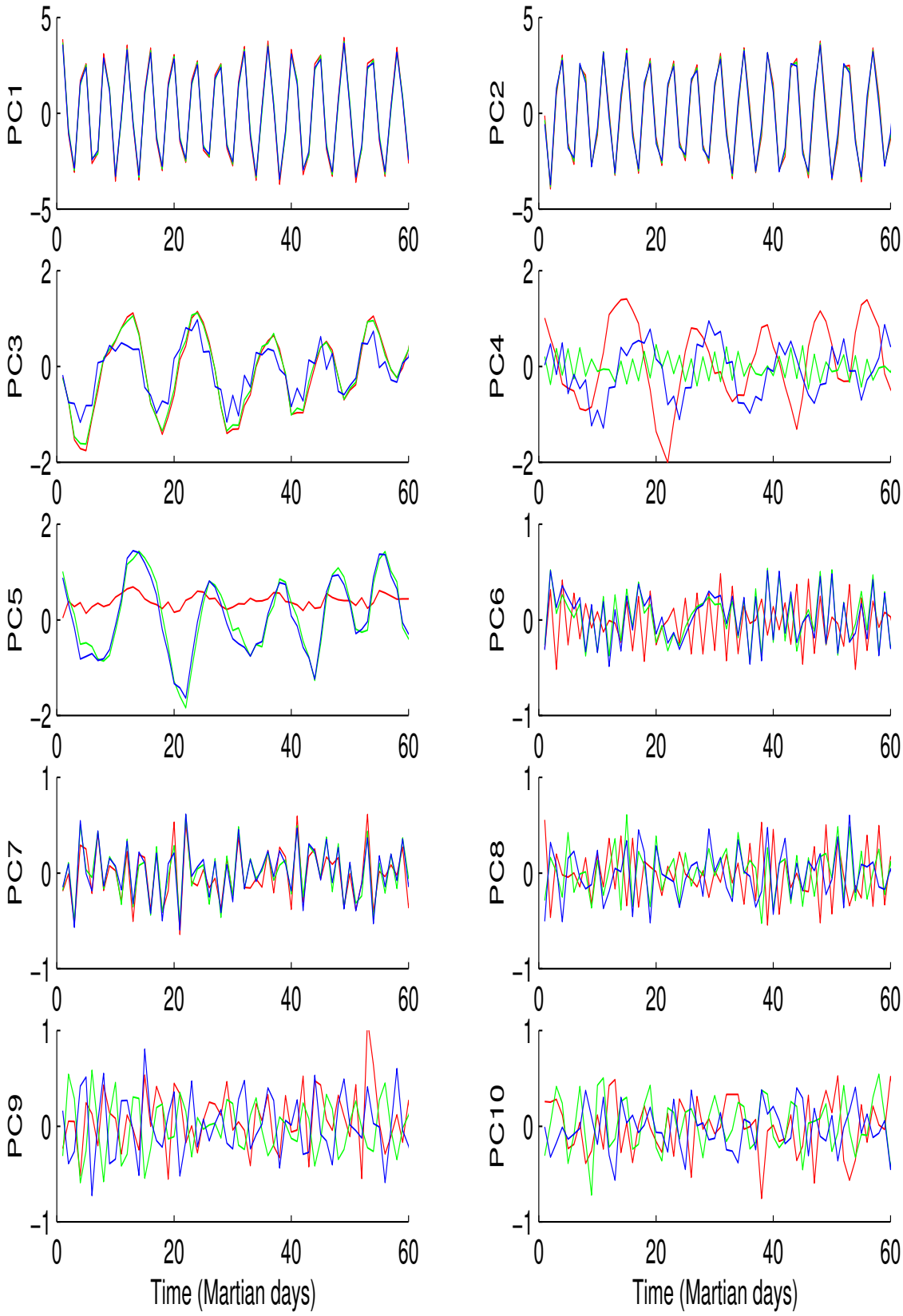
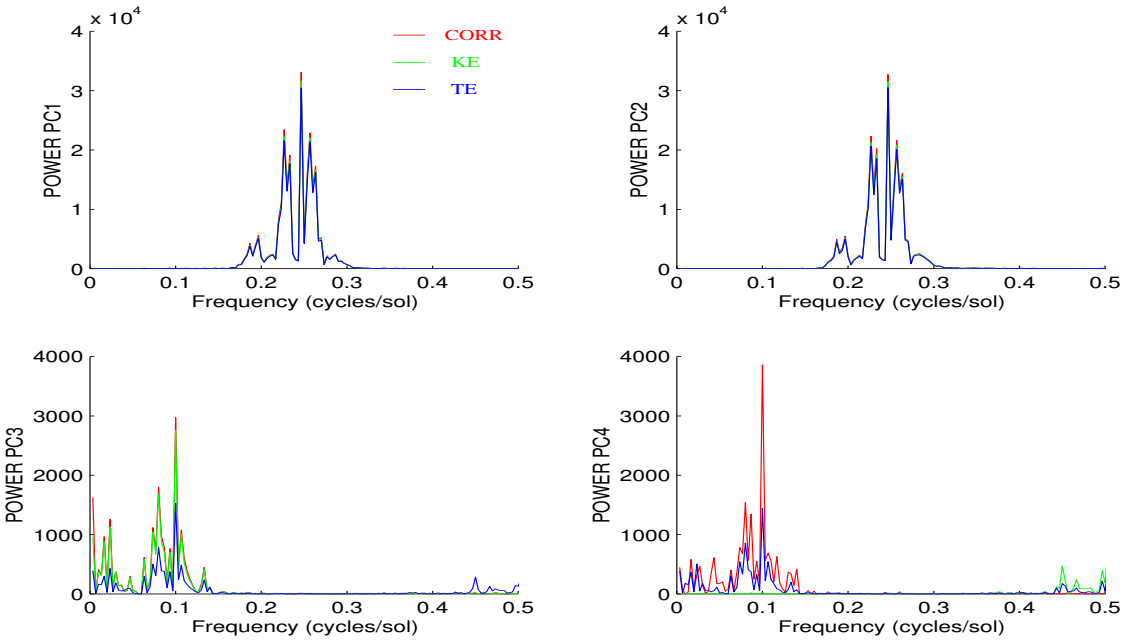
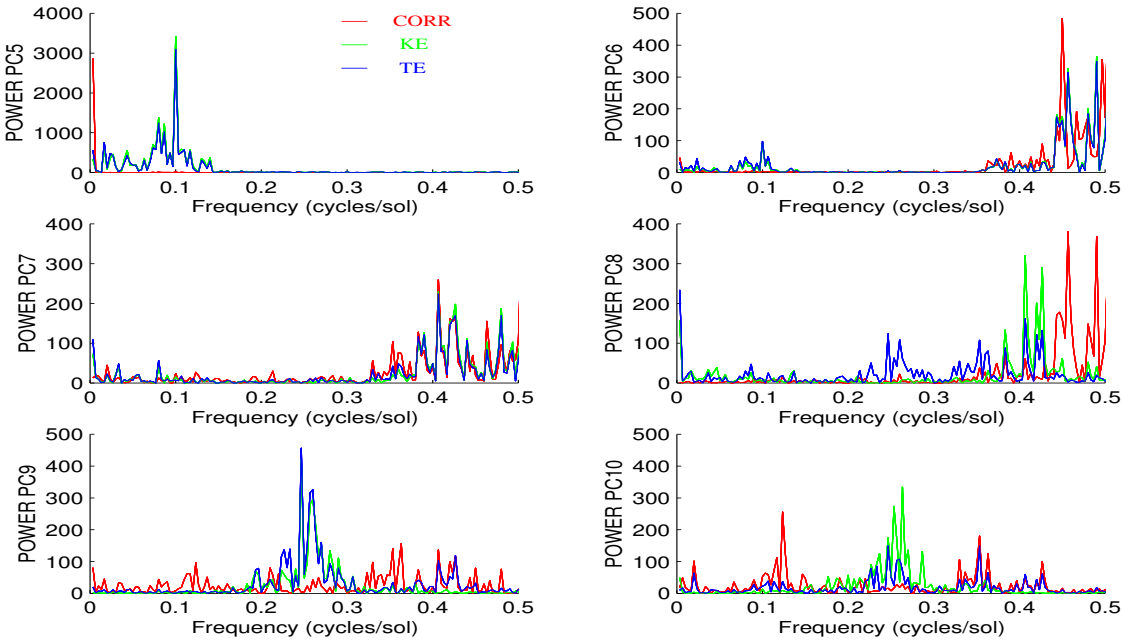


Figure 3.19: Original SGCM departure principal components (PCs) using the correlation (red), the KE (green) and the TE norms (blue) shown during an interval of 60 Martian days. PCs 1 and 2 show very little norm dependence as they are almost identical. The correlation and TE norms produce similar structures in PCs 3 and 4 whereas the higher order modes show little similarity.



(a) Power spectra profiles of the original SGCM departure PCs 1 to 4 using the correlation (red), the KE (green) and the TE norms (blue). There exists a clustering of frequencies about bands (0.23,0.27) and (0.08,1) cycles/sol for PCs 3 and 4. There is little difference in the power spectra profiles for PCs 1 and 2, confirming that POD modes 1 and 2 are almost norm independent.



(b) Power spectra profiles of the original SGCM departure PCs 6 to 10 using the correlation (red), the KE (green) and the TE norms (blue). The differences in the norms are visible as the dominant frequencies (particularly in the higher order modes) are considerably varied. Instead of one dominant frequency band, here there exists a floor of frequency values, suggesting that there is considerable interaction between many frequencies.

Figure 3.20: Power spectra profiles of the original SGCM departure PCs where (a) shows the frequencies of PCs 1 to 4 and (b) shows the frequencies of modes 5 to 10 using the correlation (red), KE (green) and TE (blue) norms.

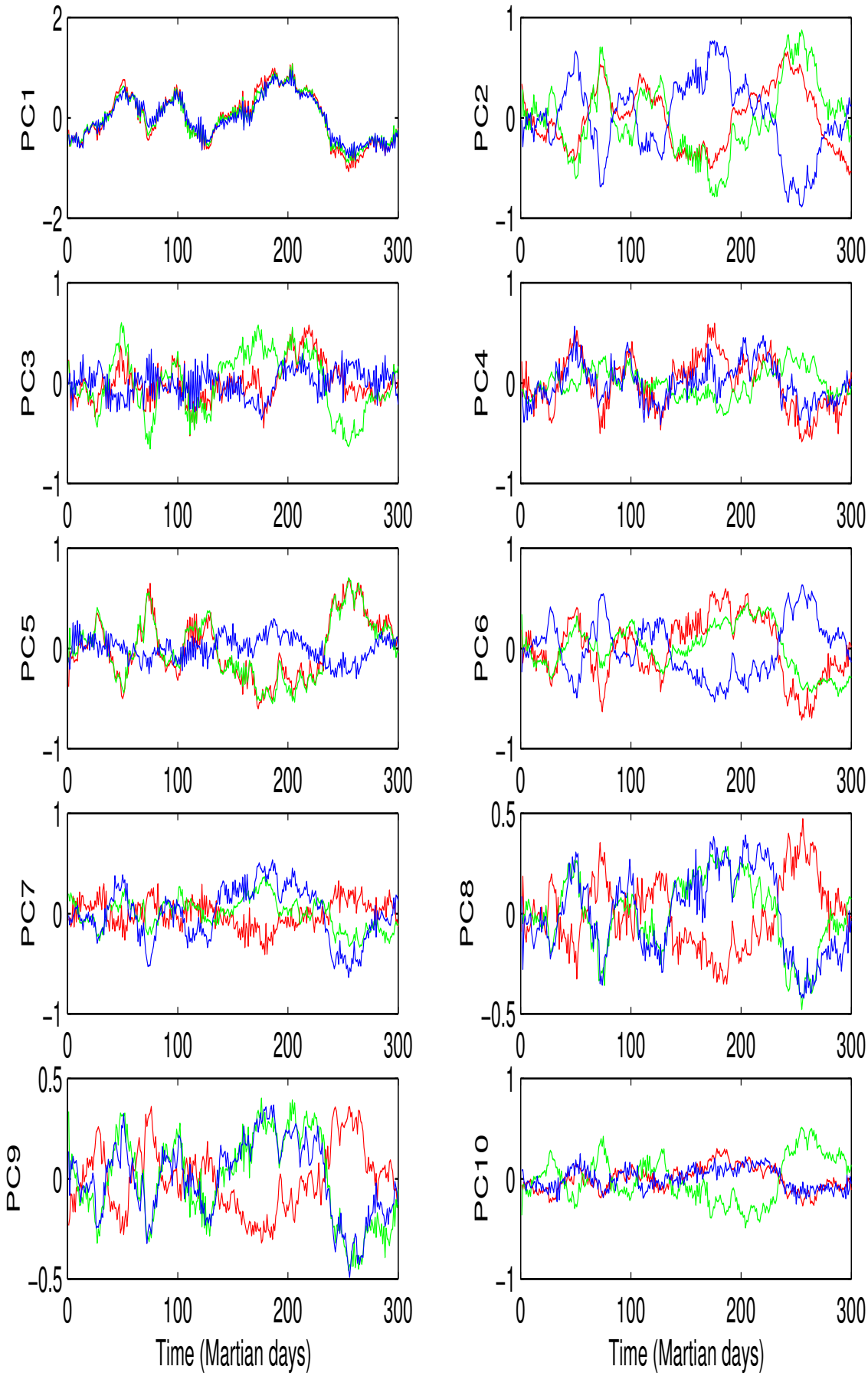


Figure 3.21: Original SGCM principal components (PCs) of the mean-flow correction field using the correlation (red), the KE (green) and the TE norms (blue) during an interval of 300 Martian days.

3.3.13 Energy in the SGCM eigenmodes

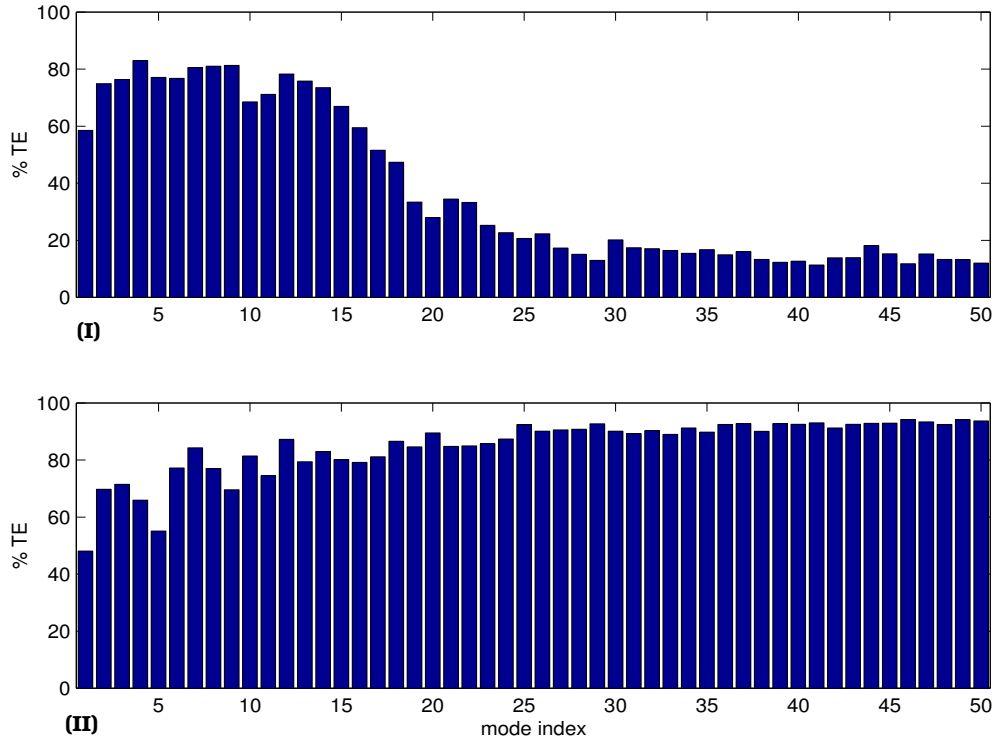
The streamfunction data were decomposed into a basic-state, mean-flow correction (MFC) and a wave flow. It has been calculated that the zonal flow (= basic state + MFC) accounts for 97.85 % of the total variance, 90.73 % of KE and 93.68 % of TE if the correlation, KE and TE norms respectively are used.

Figure 3.22(a) shows histograms of KE as a fraction of TE in the zonal POD modes using the (I) KE and (II) TE eigenvectors. It can be seen that the modes based on the TE norm extract the APE of the original system far more efficiently than the KE norm as this is illustrated by the amount of KE captured in the first zonal eigenmode. As the leading zonal mode accounts for over 90 % of the total energy in the system (see Figure 3.22(a)), then the amount of KE and APE captured by this mode will almost be equal to the total amount of KE and APE contained in the truncated system. If the KE eigenvectors are examined it can be seen that 59 % of the TE in the first zonal eigenmode is in the form of KE. This suggests that the KE norm has inefficiently captured the SGCM APE as the original model has almost an equal amount of KE and APE (see Figure 2.3). This is consistent with the fact that the KE norm optimises KE and not APE. Upon examination of the TE eigenvectors it is observed that 48 % of the TE in the leading zonal mode is in the form of KE and so the original balance between KE and APE has been retained.

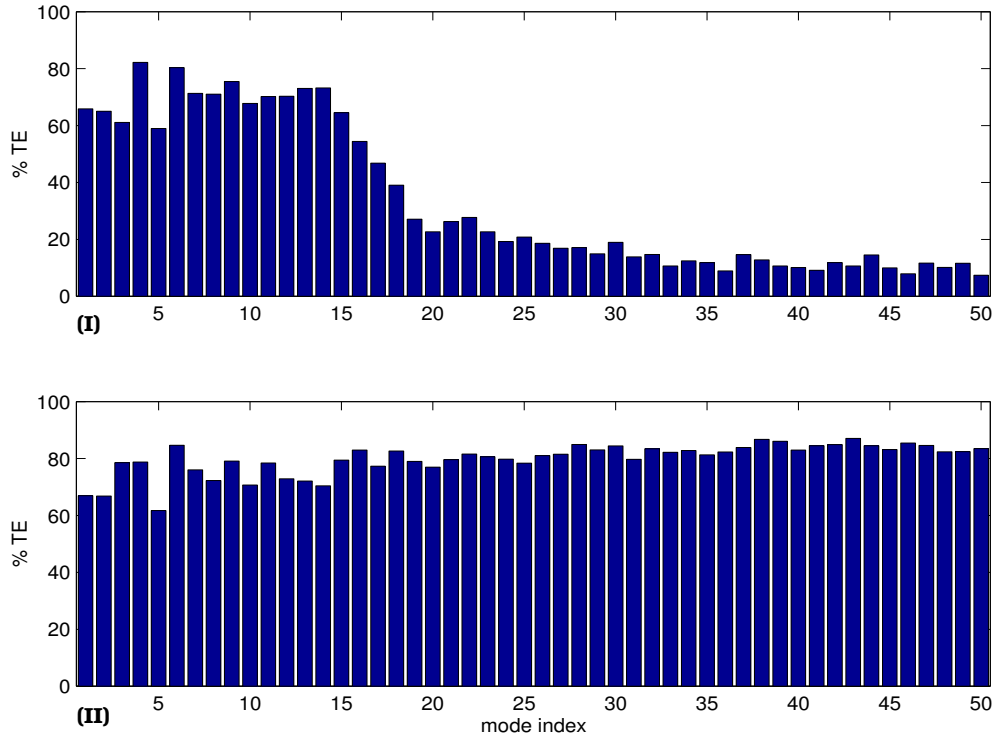
It is also evident that the KE norm pushes a large proportion of the system's KE into the leading order modes as there is a very fast tail-off in the amount of KE contained in the eigenmodes, particularly after the first 14 KE eigenvectors. This contrasts with the TE eigenmodes which contain a very significant proportion of KE in the higher-order modes (about 80 %). This seems to suggest that the TE eigenvectors have placed most of the APE in the leading order modes and so have effectively optimised the capture of APE.

Figure 3.22(b) shows KE as a fraction of TE in the departure POD modes using the (I) KE and (II) TE eigenvectors. The leading KE modes are dominated by KE (accounting for about 70 % of TE), but after 14 KE eigenvectors there is a steep tail-off in the KE fraction which implies that the low-order TE modes capture most of the APE. However, the higher-order TE eigenvectors contain mainly KE. Again, this indicates that the KE norm places the majority of its KE supply into the leading order modes whereas the TE norm pushes the APE into the lower-order TE eigenvectors.

Figure 3.22 agrees with what Selten (1993) found in the context of the Earth, for he concluded that even though the eigenvalues obtained from the different norms looked qualitatively the same, the corresponding sets of eigenvectors would behave differently in terms of how much KE and APE was captured by each of the eigenmodes. He also showed that the KE eigenvectors naturally distributed the majority of the KE into the leading order modes, whereas the higher-order KE eigenvectors were dominated by APE. It was also shown that the TE eigenvectors placed most of the APE in the leading order TE eigenmodes, leaving a greater fraction of KE in the higher-order modes.



(a) KE as a fraction of the TE contained in each of the zonal modes using the (I) KE and (II) TE eigenvectors.



(b) KE as a fraction of the TE contained in each of the wave modes using the (I) KE and (II) TE eigenvectors.

Figure 3.22: KE as a fraction of TE contained in (a) the zonal modes and (b) the wave modes for the (I) KE and (II) TE eigenvectors.

Summary

The last 30 KE eigenvectors contain mainly APE. In a truncated model based on the KE eigenvectors baroclinic patterns are ignored, so interactions involving the temperature field will be relatively poorly represented. This means that the energy conversion process, something which is essential in the model dynamics, will not be well described.

It is now our aim to carry out a similar analysis for data obtained from the full MGCM, which includes a comprehensive range of physical parameterisations (which are relevant to the Martian atmosphere) and topography.

3.4 The MGCM

The MGCM is a spectral primitive equation model which used a horizontal T21 truncation and uses 10 equally spaced σ levels in the vertical (shown in Table 3.2) and is described in detail by Collins et. al (1996). The difference is that this version of the GCM includes parameterisation schemes which represent the effects of infrared radiative transfer in the 15 μm CO_2 band, radiative effects of dust in both the long and short waves, dry convection, boundary layer processes, surface heating, subsurface conduction of heat, and the condensation and sublimation of CO_2 on the surface[4] (as described by Hourdin et al. (1993)). Furthermore, surface topographic height from the Digital Terrain Model's (DTM) data set (see Wu (1978)) is also included.

Figure 3.23(a) shows the time-averaged zonal winds with respect to latitude and height (km) during a seasonal run from northern autumn through to spring and summer (L_s = 157.3° to 53.7°). The zonal winds are dominant in the NH, ranging from 10-70 ms^{-1}. The winds have been temporally averaged over the northern autumn, winter and spring seasons, resulting in dominant winds occurring in the NH. Temperature is illustrated in Figure 3.23(b). The temperature profiles are almost symmetric about the equator, where the warmest isotherm of 220 K is found at ground level, declining almost linearly to 160 K at 30 km from the ground. Typical model streamfunction data from the MGCM is illustrated in Figure 3.24, where a stationary or slowly moving wave can be seen at about $\pm 50^o$ latitude, together with a rapidly moving diurnal tide which is seen as a travelling wave.

3.4.1 A winter run

It was found that there existed large seasonal variations in the amplitudes of the MGCM PCs from autumn through winter to spring, which were a consequence of the variations in the thermal tide and the diurnal cycle. The 'seasonal mixing' of the forcing terms caused complications to arise in the spatial and temporal modal configurations. It was decided to only use the northern winter season of the MGCM (L_s = 239.6° to 300.42°) so that the seasonal conditions would be comparable to those of the perpetual southern winter run of the SGCM (see Figure 3.25).

[4] CO_2 (the major constituent of the atmosphere) freezes out to form an immense polar cap, alternately at each pole, and forms a great cover of snow which then evaporates with the coming of spring in each hemisphere.

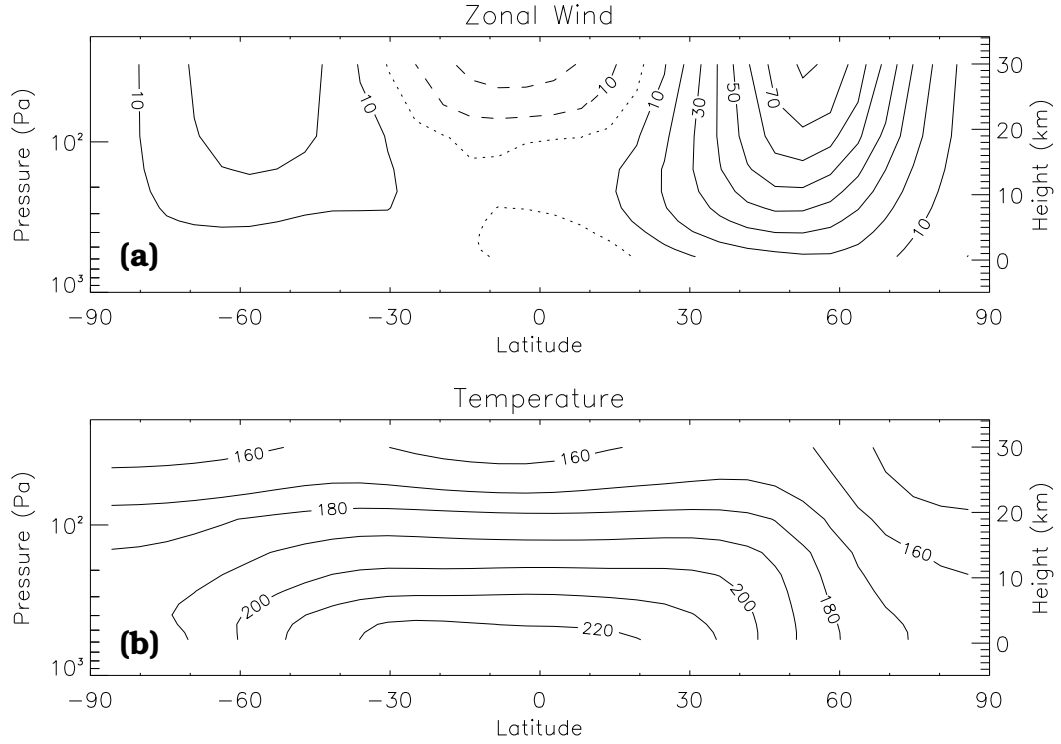


Figure 3.23: Plots of the (a) MGCM time-averaged zonal wind and (b) temperature (on σ=0.45) over 2000 time levels from northern autumn through to spring and summer ($L_s = 157.3°$ to $53.7°$).

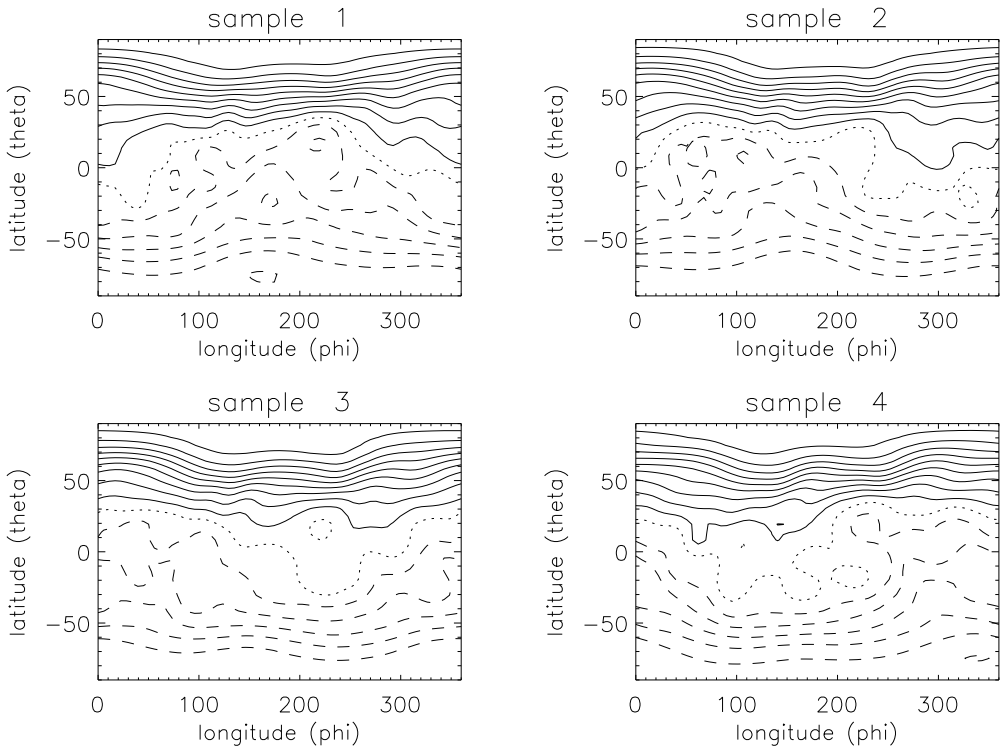


Figure 3.24: Snapshots of the MGCM modelled streamfunction data showing the first 4 of the 5 samples taken on the first day (intervals of $\frac{1}{5}$ of a Martian day) during the winter period on σ=0.45.

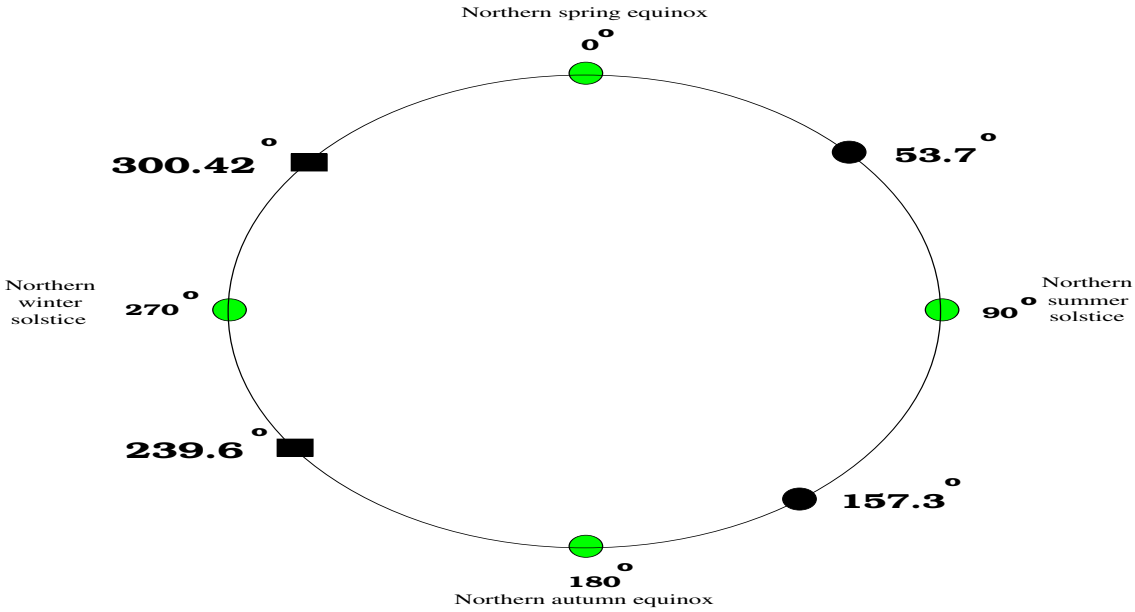


Figure 3.25: The MGCM data ($L_s = 157.3°$ to $L_s = 53.7°$) contained significant 'seasonal mixing'. It was decided that only the northern winter run would be used ($L_s = 239.6°$ to $300.42°$), denoted above by the section which bridges the two black squares.

Clearly we see that there are limitations to the conclusions which can be made from the MGCM study, but it is realised that dominant seasonal trends can potentially obscure important meteorological patterns.

3.4.2 MGCM eigenspectrum

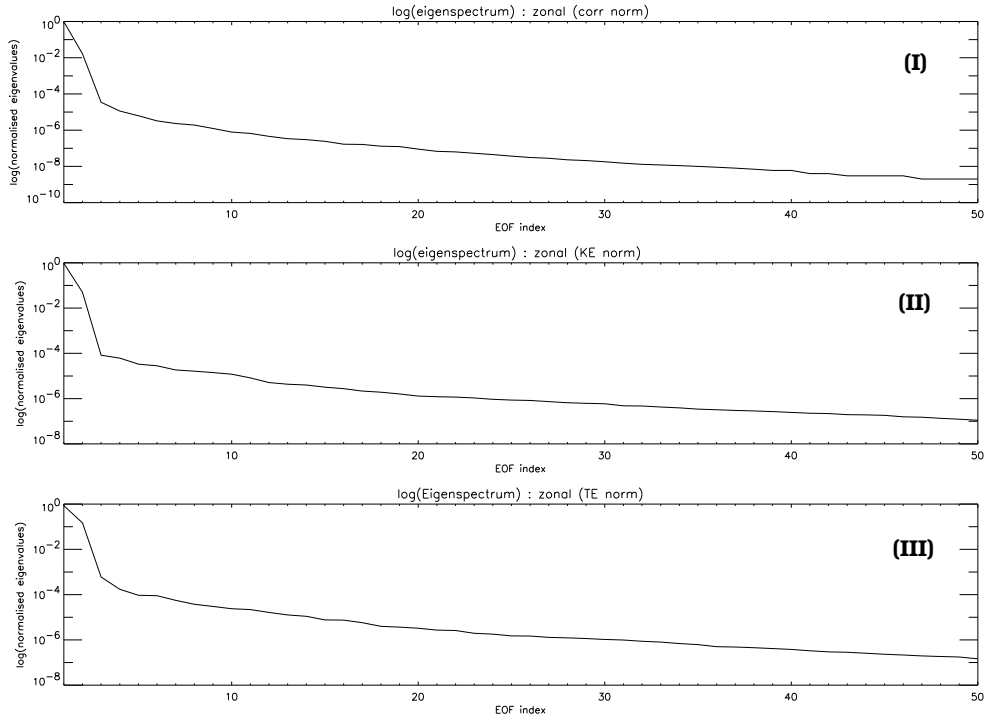
Figures 3.26 and 3.27 respectively illustrate the eigenspectra[5] of the zonal and departure fields using the correlation, KE and TE norms, where each logarithmic profile of the zonal and departure fields displays a rapid decrease in the magnitude of the eigenvalues.

Figure 3.26(b) shows that the first two eigenvalues of the zonal field dominates the flow and captures 99.993 % of the variance (in I), 99.968 % of KE (in II) and 99.878 % of TE (in III) if the correlation, KE and TE norms are respectively used. The fall-off rate of the correlation zonal eigenspectrum is faster than that observed in the energy eigenspectra, confirmed by the behaviour of the logarithmic eigenspectra profiles which are shown in Figure 3.26(a).

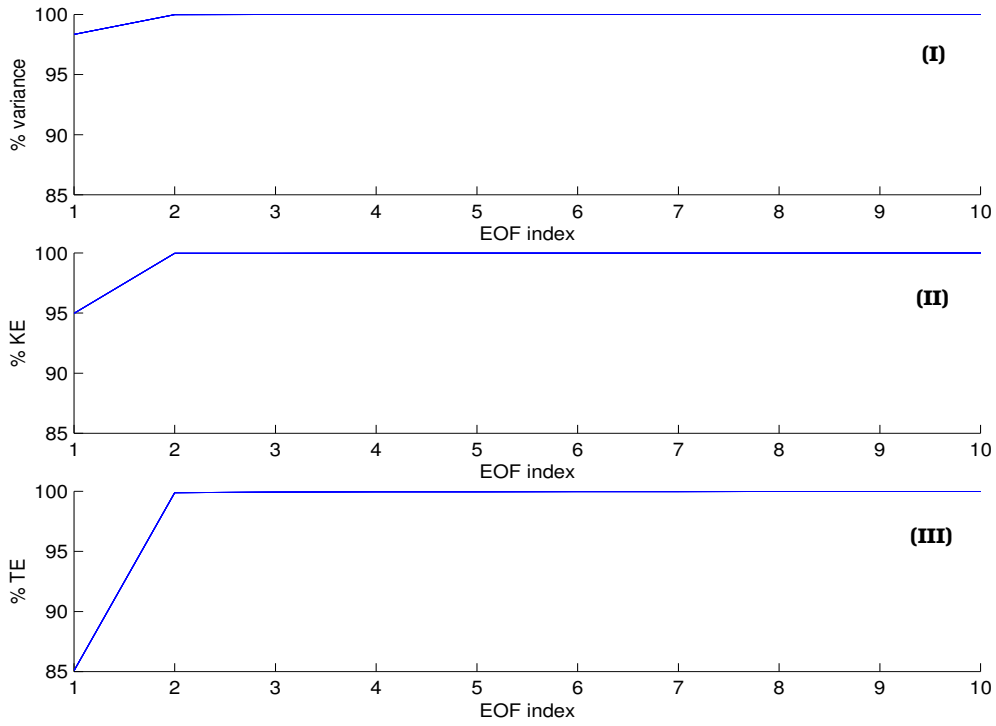
Figure 3.27(b) shows that the leading departure eigenvalue captures 46.84 % of the variance (in I), 52.29 % of KE (in II) and 43.09 % of TE (in III) of the wave energy if the correlation, KE and TE norms are respectively used. The bimodal behaviour which was found in the SGCM eigenspectra (Figure 3.9) is not observed as readily in the MGCM.

Table 3.5 shows that the subspaces spanned by 10 correlation, KE and TE eigenvectors respectively capture 98.64 %, 93.15 % and 96.13 % of the variance, KE and TE of the wave flow. Only 3 correlation or 3 TE eigenmodes are required to capture 90 % of the variance or

[5]The eigenvalues are normalised so that their sums equal 1.

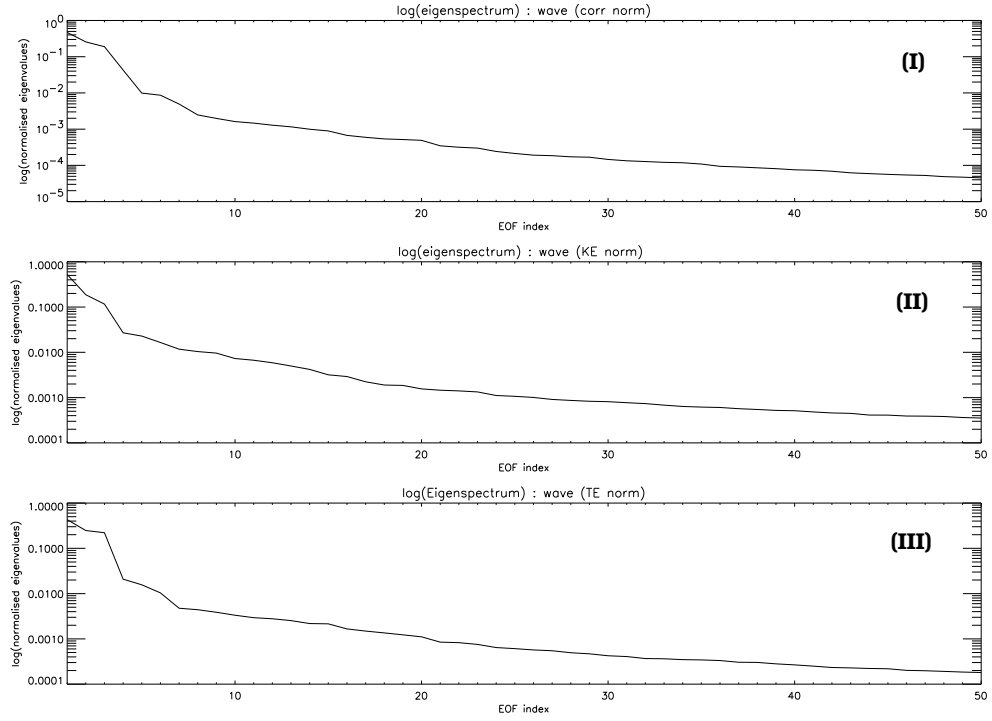


(a) The logarithmic eigenspectra zonal profiles using the (I) correlation, (II) KE and (III) TE norms.

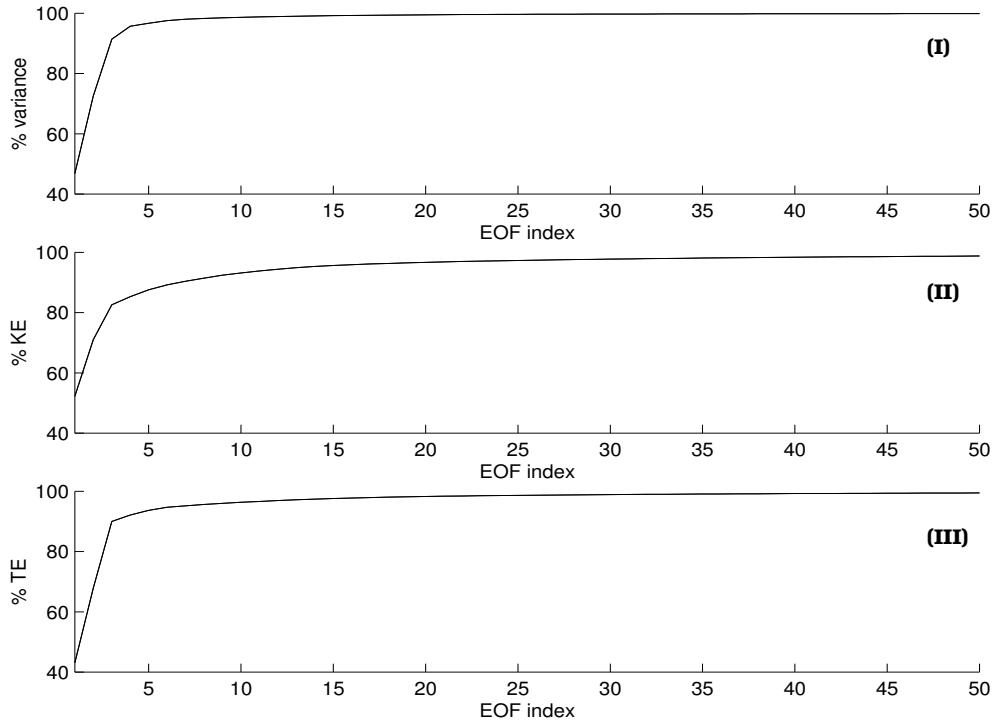


(b) The cumulative (I) variance, (II) KE and (III) TE capture of the zonal field using the (I) correlation, (II) KE and (III) TE norms respectively.

Figure 3.26: The (a) logarithmic and (b) cumulative eigenspectra of the MGCM zonal flow using the (I) correlation, (II) KE and (III) TE norms.



(a) The logarithmic eigenspectra departure profiles using the (I) correlation, (II) KE and (III) TE norms.



(b) The cumulative (I) variance, (II) KE and (III) TE capture of the departure field using the (I) correlation, (II) KE and (III) TE norms respectively.

Figure 3.27: The (a) logarithmic and (b) cumulative eigenspectra of the MGCM departure flow using the (I) correlation, (II) KE and (III) TE norms.

Norm	90 %	95 %	99 %	4 modes	10 modes	50 modes
Correlation	3 modes	4 modes	7 modes	95.68 %	98.64 %	99.89 %
KE	7 modes	14 modes	34 modes	85.30 %	93.15 %	98.79 %
TE	3 modes	7 modes	18 modes	92.11 %	96.36 %	99.46 %

Table 3.5: POD modal truncation against variance/energy capture in the departure flow.

90 % of the TE respectively, whereas 7 KE eigenmodes are needed to capture 90 % of the KE. If an additional 9 % of the variance, KE or TE of the system is sought then an extra 4, 27 or 15 correlation, KE or TE eigenvectors are required. This illustrates that the correlation eigenspectrum has the fastest fall-off rate and this can be seen in the logarithmic eigenspectra plots shown in Figure 3.27(a).

It can be seen in Figure 3.27(a) that a sudden decrease by a factor of 10 occurs beyond wave pattern 3. Therefore, only 3 wave modes would be sufficient to describe the model dynamics to the first order (see Kantz and Schreiber (1997)). However, if nonenergetic components are critical in the dynamics of the MGCM then truncating at 3 modes would mean that a low-dimensional MGCM model would not be capable of describing that part of the attractor (of the original MGCM) in which such modes become important. This part of the attractor might be highly significant in the overall dynamics of the MGCM.

3.4.3 MGCM POD modes

Figures 3.28 - 3.33 illustrate the barotropic and baroclinic components of the MGCM departure POD modes. We will now compare the various patterns produced by the correlation, KE and TE norms, examining the barotropic and baroclinic components of each of the modes. Again it will be seen that the choice of norm affects the ordering of the modes so for convenience, only the odd departure eigenmodes will be discussed here.

Modes 1 and 2 form a conjugate pair for each of the norms and thus denotes a travelling wave in the longitudinal direction. From barotropic mode 1 (Figures 3.28, 3.30 and 3.32) it can be seen that there are striking differences produced by the correlation and the energy norms (the positive contour lines are shaded in red, the negative contours in blue and the zero contour lines in green). The barotropic component from the correlation norm characterises a wavenumber 1, which is centred at the equator and depicts the diurnal tide (confirmed in the PC behaviour which will be discussed in § 3.4.4). The wavenumber 1 structure exists in each of the baroclinic components of correlation mode 1, and the structures are almost identical to those observed in the TE norm mode. The energy norms do not extract such a distinctive wavenumber 1 pattern in the barotropic component, and only the TE norm extracts a wavenumber 1 structure (in each of the baroclinic components of mode 1). The KE norm produces complex spatial patterns which are concentrated about the equator (this is also evident in the remaining KE modes). From the KE barotropic component it can be seen that a longitudinal wavenumber 2 structure exists at latitudes of 50^o and 20^o in the SH, as well as a complicated wavenumber 4 structure at a latitude of 20^o in the NH. It is also noted that the KE norm retrieves intricate longitudinal wavenumber 3 structures in the first and second baroclinic components, respectively centred

at latitudes of 15^o in the NH and 25^o in the SH. This suggests that the KE norm places greater emphasis upon smaller-scale structures. The TE norm produces modes of a similar nature to those generated by the correlation norm.

The diurnal cycle is evident in the KE and TE mode 3 (Figures 3.30 and 3.32 respectively). The TE norm produces a definite wavenumber 1 structure, centred at the equator, and the diurnal tide is clearly apparent in all the baroclinic components where clean wavenumber 1 structures are observed (confirmed by the PC behaviour in § 3.4.4). However, the KE norm only extracts a modulated wavenumber 1 pattern, as well as many small-scale structures, being particularly evident in baroclinic components 2 and 3, which again suggests that the KE norm tends to extract small-scale spatial structures. Both the KE and TE mode 3 differ strongly from the spatial patterns produced by the correlation norm. Here, in the barotropic component a modulated wavenumber 2 pattern exists, which is centred at a latitude of 50^o in the SH, as well as a wavenumber 3 structure at a latitude of 15^o in the NH. The first baroclinic component of the correlation norm contains a wavenumber 1 pattern centred at a latitude of 50^o in the SH, in addition to a wave number 2 structure at a latitude of 15^o in the NH. The second and third baroclinic components reveal a longitudinal wavenumber 3 structure centred at a latitude of 15^o in the SH.

The energy POD modes 4 and 5 form conjugate pairs in each case, whereas correlation modes 5 and 6 form a complex pair. When comparing mode 5 between the various norms (Figures 3.29, 3.31 and 3.33) it can be observed that the correlation norm modes display more extensive coherent structures than those produced by the energy norms. A definite wavenumber 1 is clearly visible in correlation mode 5, where the barotropic, the first baroclinic and the third baroclinic components respectively contain wavenumber 1 structures at latitudes 60^o, 30^o and 20^o in the SH. The second baroclinic component contains two parallel lines of longitudinal wave number 1 structures, centred at the equator and at a latitude of 35^o in the SH. However, the energy norms produce rather complicated spatial patterns. The KE and TE norms generate a modulated wavenumber 1 structure in the barotropic component, which is centred at a latitude of 55^o in the SH, together with smaller-scale structures which are rather intricate in form, concentrated in the NH. By making comparison between the baroclinic components of the KE and TE modes it is clear that the KE norm is biased towards smaller-scale patterns, whereas the TE norm produces larger spatial patterns, far less complex in design.

The barotropic component of correlation mode 7 (Figure 3.29) contains two parallel horizontal wavenumber 2 structures, centred at latitudes $\pm 25^o$. Similarly, the first and second baroclinic components of the correlation mode contain two parallel wave number 3 structures, respectively centred at latitudes of 25^o in the NH, 50^o in the SH and at latitudes of 40^o in the NH and 25^o in the SH. The third baroclinic component has a severely modulated wave number 2 structure, centred at a latitude of 45^o in the NH. The barotropic component of the KE norm possesses a longitudinal wavenumber 3 structure centred at a latitude of 60^o in the SH, whereas the first baroclinic component has a wavenumber 2 structure at a latitude of 50^o in the SH. The second and third baroclinic components of the KE mode contain modulated wavenumber 2 structures, centred at a latitude of 45^o in the NH, as well as numerous small-scale structures which exist in

the SH. The barotropic and the three baroclinic components of TE mode 7 holds wavenumber 2 structures, where the barotropic component reveals a wavenumber 2 pattern at a latitude of 70^o in the SH, and the entire NH consists of a complicated, large-scale pattern. Each of the baroclinic components possess two parallel horizontal wavenumber 2 structures, respectively centred at latitudes of $\pm 35^o$, $\pm 45^o$, 35^o in the NH and 15^o in the SH. So, the correlation norm extracts clear coherent structures, the TE norm produces similar spatial patterns, but the KE norm generates many small-scale complicated structures.

If mode 9 (Figures 3.29, 3.31 and 3.33) is examined it can be seen that the correlation norm produces a dominant wavenumber 2 pattern whereas the energy norms produce much smaller-scale wave structures. In each of the baroclinic components of correlation mode 9 a strong coupled wavenumber 2 pattern can be observed, centred at latitudes $\pm 50^o$. The KE norm generates many distinct, small-scale spatial structures, where a weak wavenumber 2 can be identified at a latitude of 50^o in the SH. The TE norm produces a stronger wavenumber 2 profile than that generated by the KE norm, centred at a latitude of 50^o in the SH. However, the KE norm does not extract the clear coherent structures which are visible in the correlation modes, but produces many disjoint and modulated spatial patterns.

Summary

Even though the eigenspectra profiles of the various norms are similar (Figure 3.27), the corresponding eigenvectors are very different. This was not observed in the SGCM data, where the two leading spatial modes were seemingly insensitive to the choice of norm. The correlation norm clearly extracts a spatial wavenumber 1 pattern whereas the energy norms place a greater emphasis towards smaller-scale structures, particularly the KE norm. Unlike the spatial patterns observed from the SGCM, the structures here are not sinusoidal, as their surface amplitude is significantly modulated by the orography. However, the spatial POD structures seem to suggest that only the leading three eigenmodes are required to capture the diurnal tide. This will be determined in § 3.4.4.

It has been shown that the use of the different norms affects the ordering of the POD modes. It is clear that modes 1 and 2 (for each of the norms) form a complex conjugate pair and thus depict a travelling diurnal tide (confirmed by the PC behaviour in § 3.4.4), whereas modes 5 and 6 for the correlation norm and modes 4 and 5 for the energy norms together depict a travelling wavenumber 1 structure, and quite possibly represent the leading order baroclinic mode. Hence two distinct travelling waves can be observed from the leading order POD modes. Clearly, the use of the various norms has affected the ordering of the spatial eigenmodes, and so direct comparisons between modes at the same truncation level would be pointless. Again, an understanding of how many modes form complex conjugate pairs for a given truncation level is crucial so that it is known how many travelling waves are retained in a given truncation level.

It is difficult to determine whether the higher order spatial POD modes evolve as travelling or standing waves, so to provide us with additional information regarding the behaviour of the spatial patterns we will study their temporal behaviour by analysing the principal components of the flow.

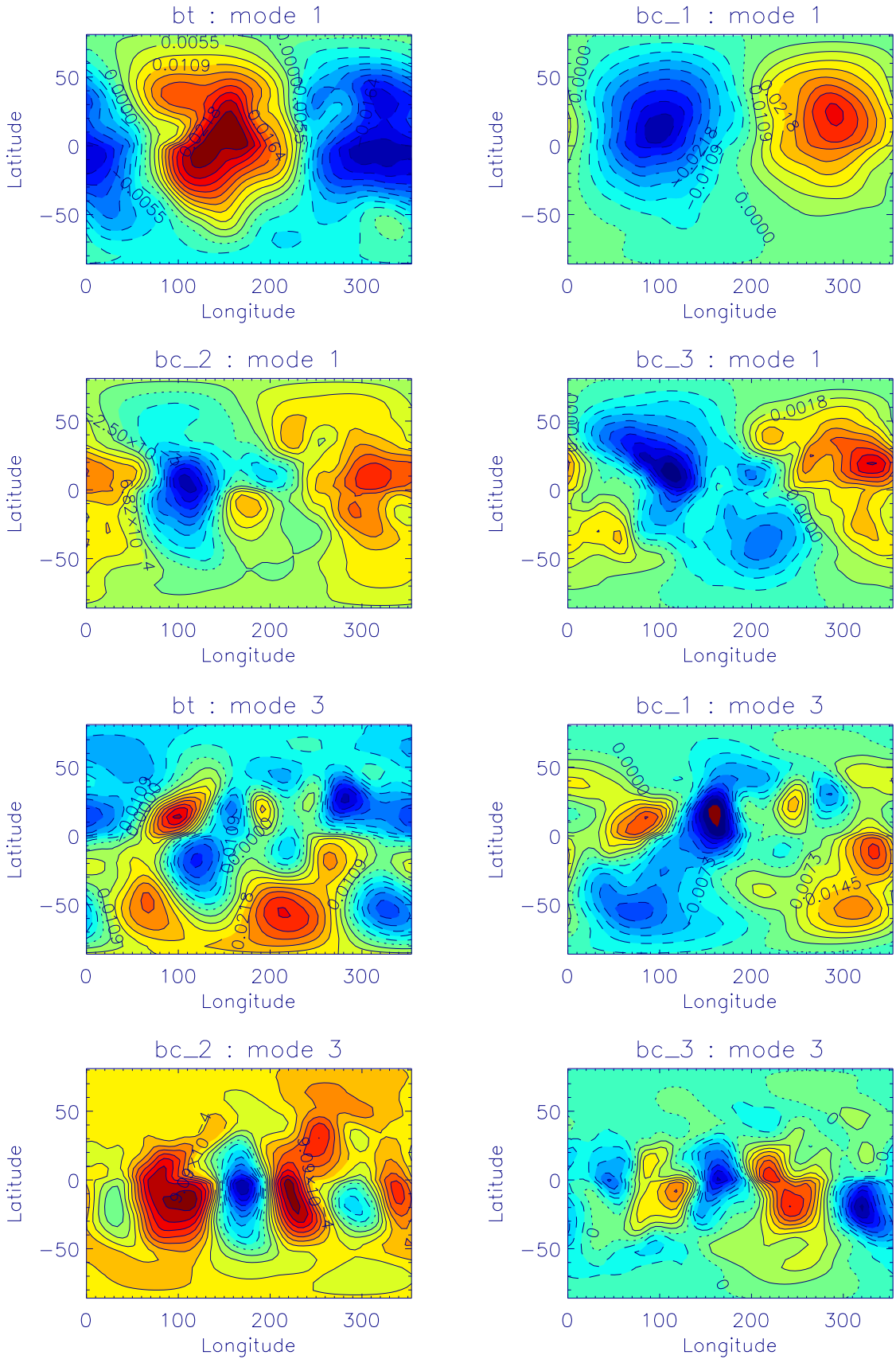


Figure 3.28: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the correlation norm.

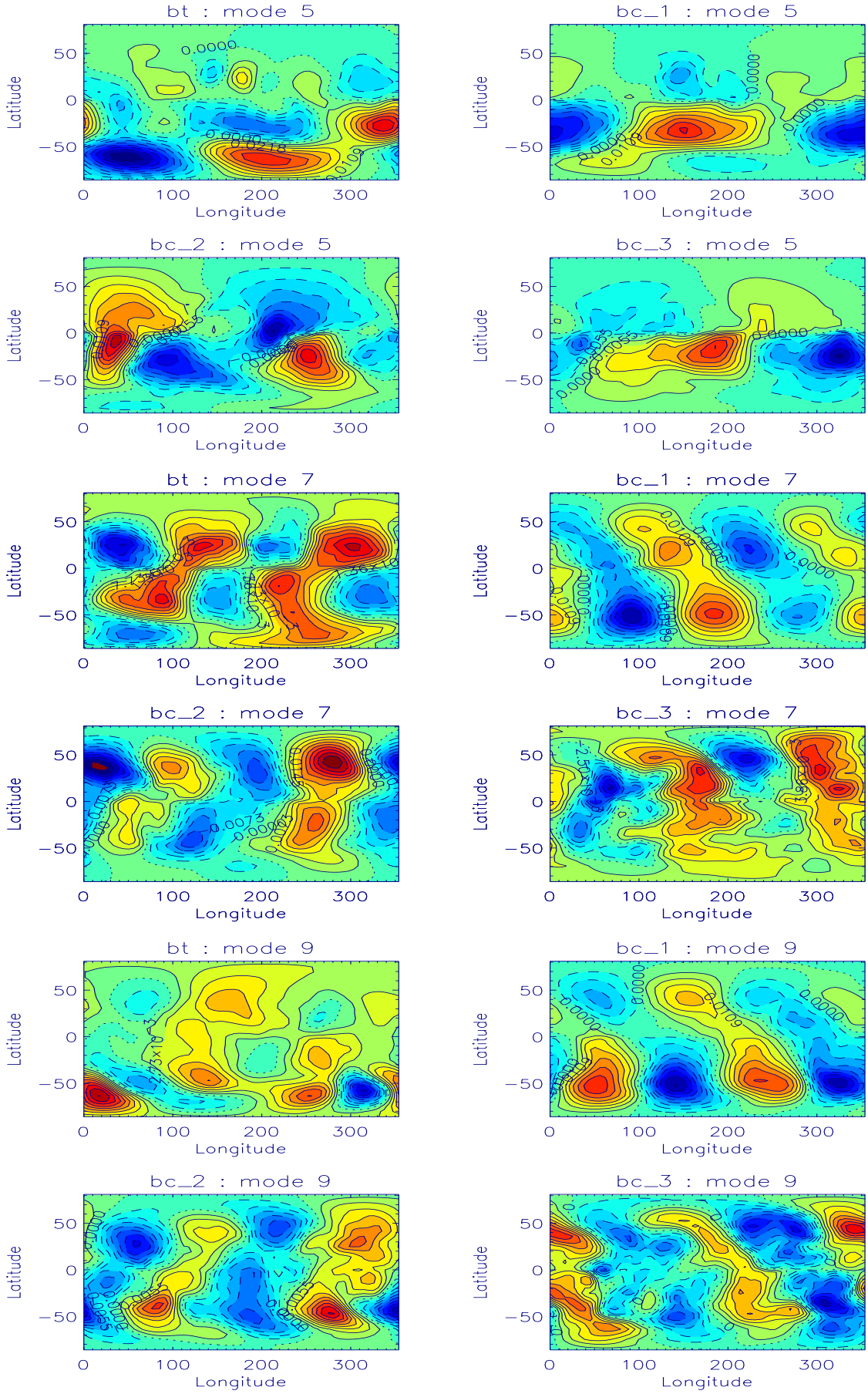


Figure 3.29: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the correlation norm.

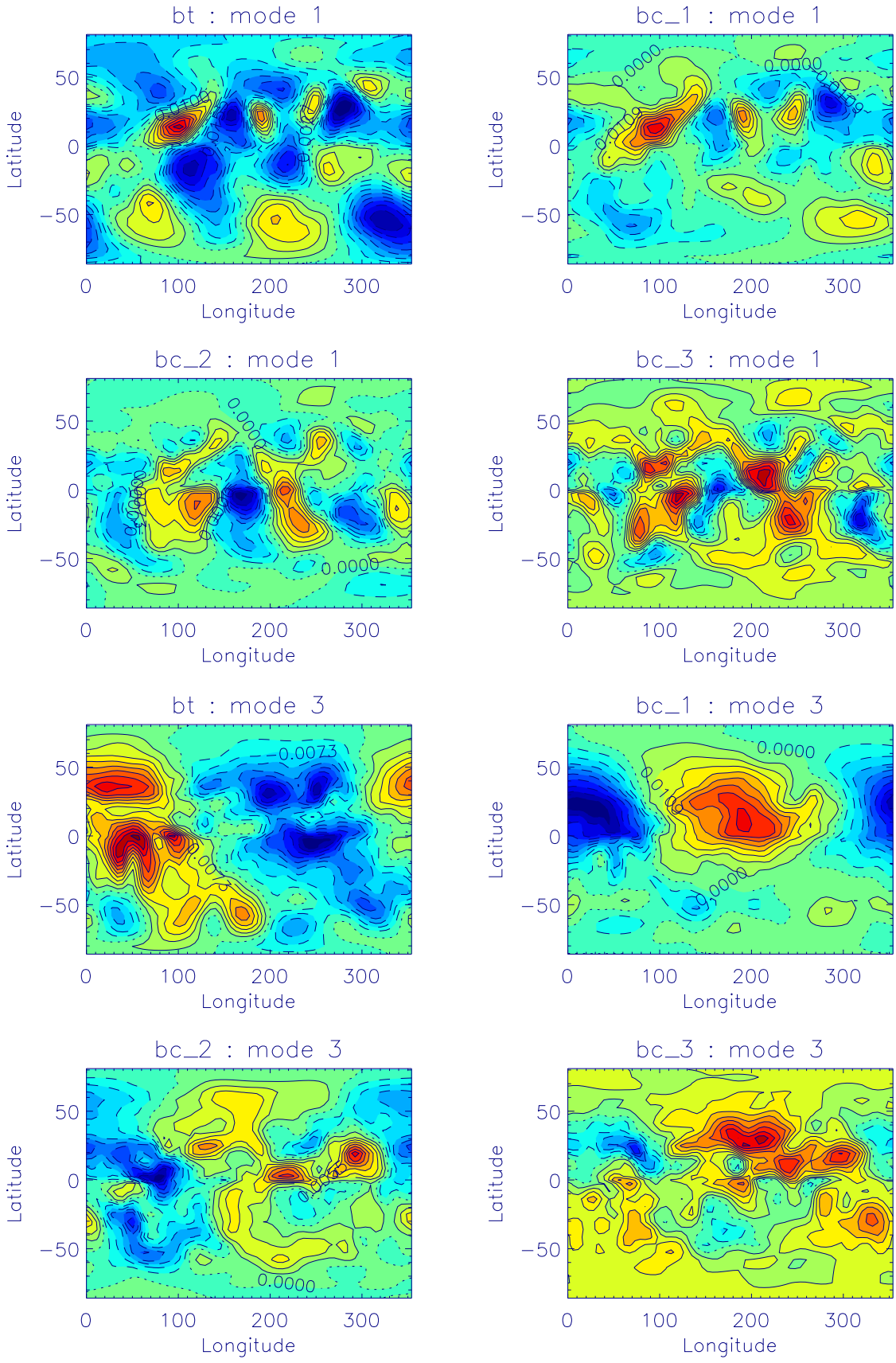


Figure 3.30: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the KE norm.

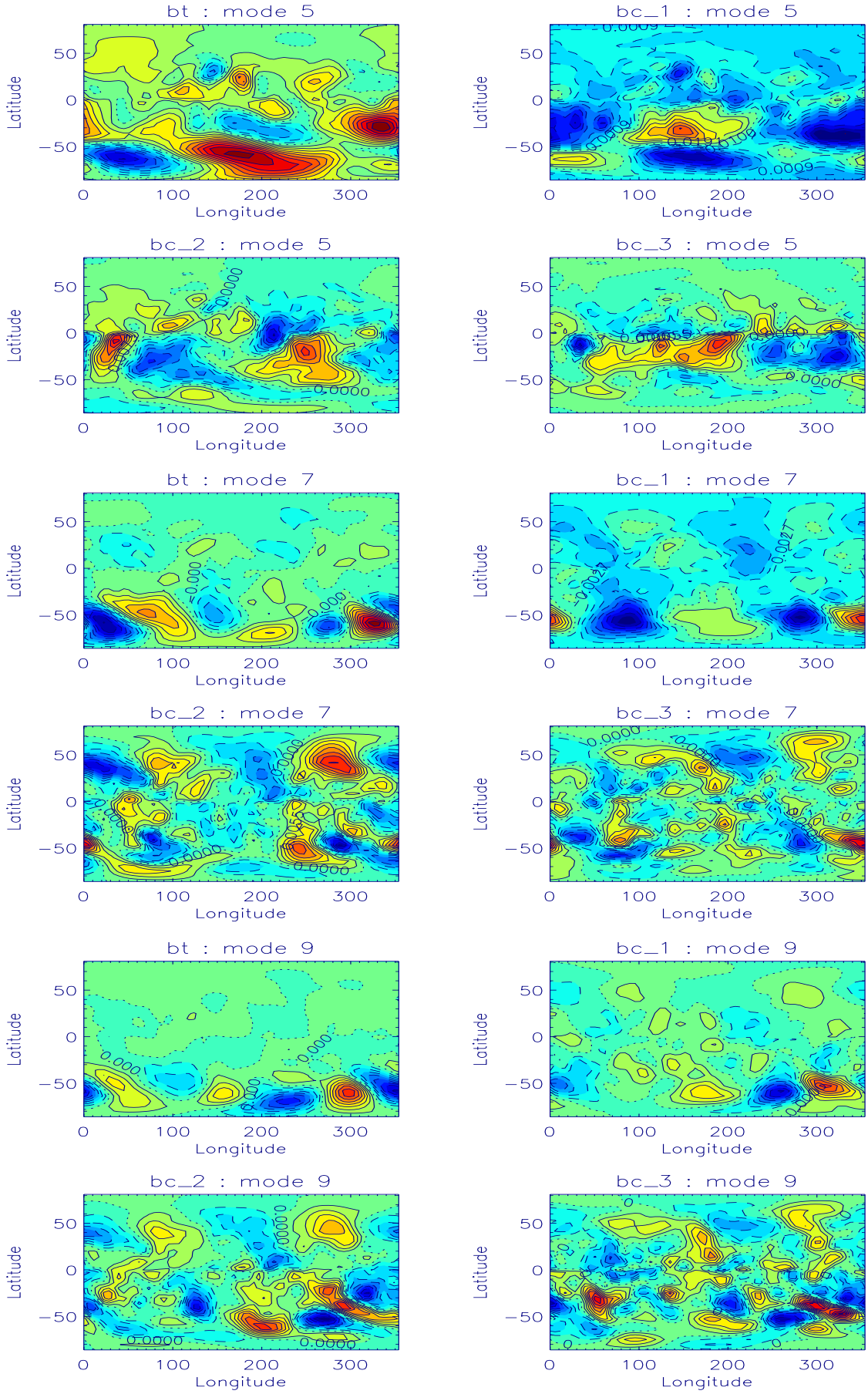


Figure 3.31: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the KE norm.

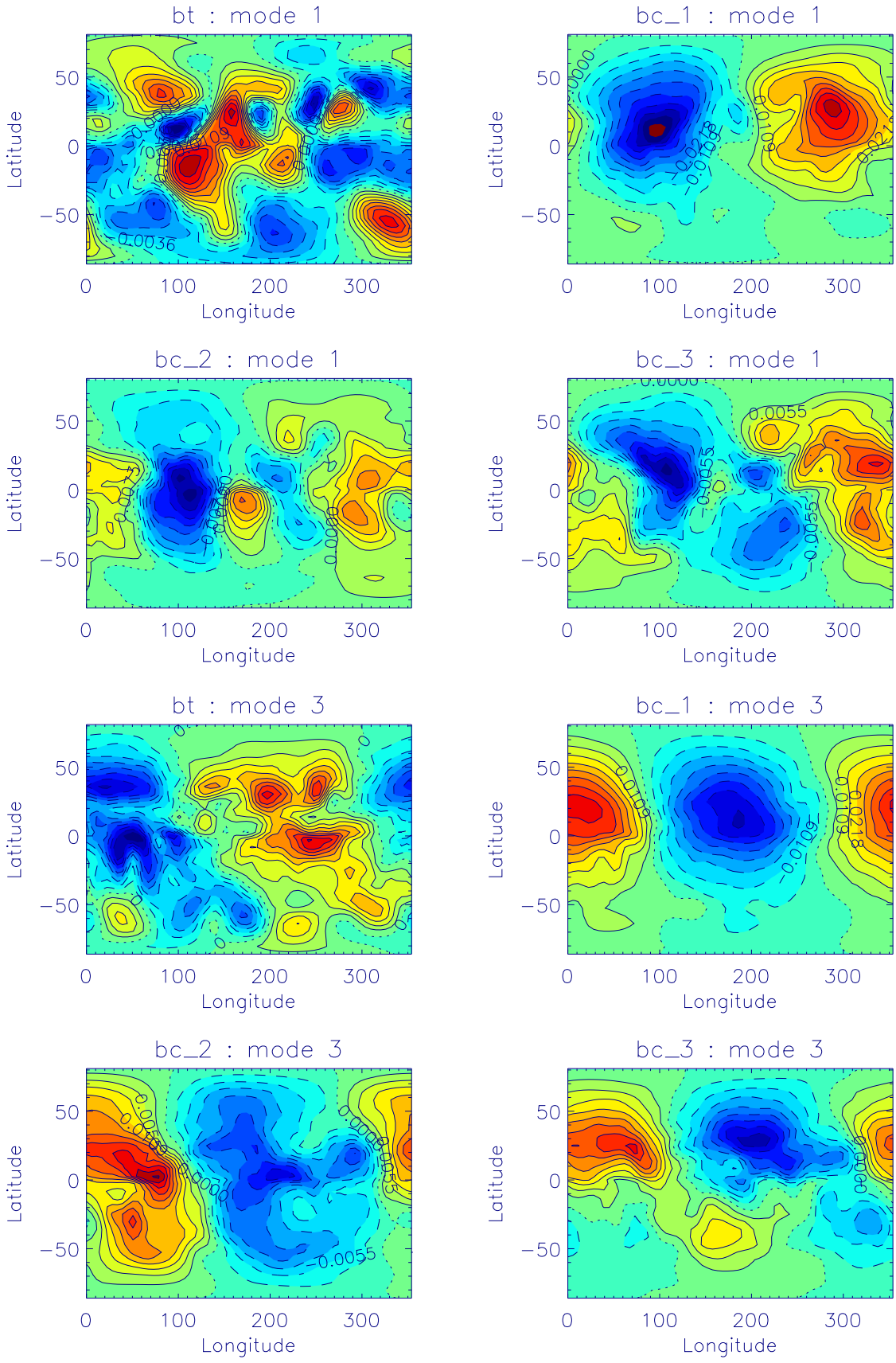


Figure 3.32: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 1 and 3 using the TE norm.

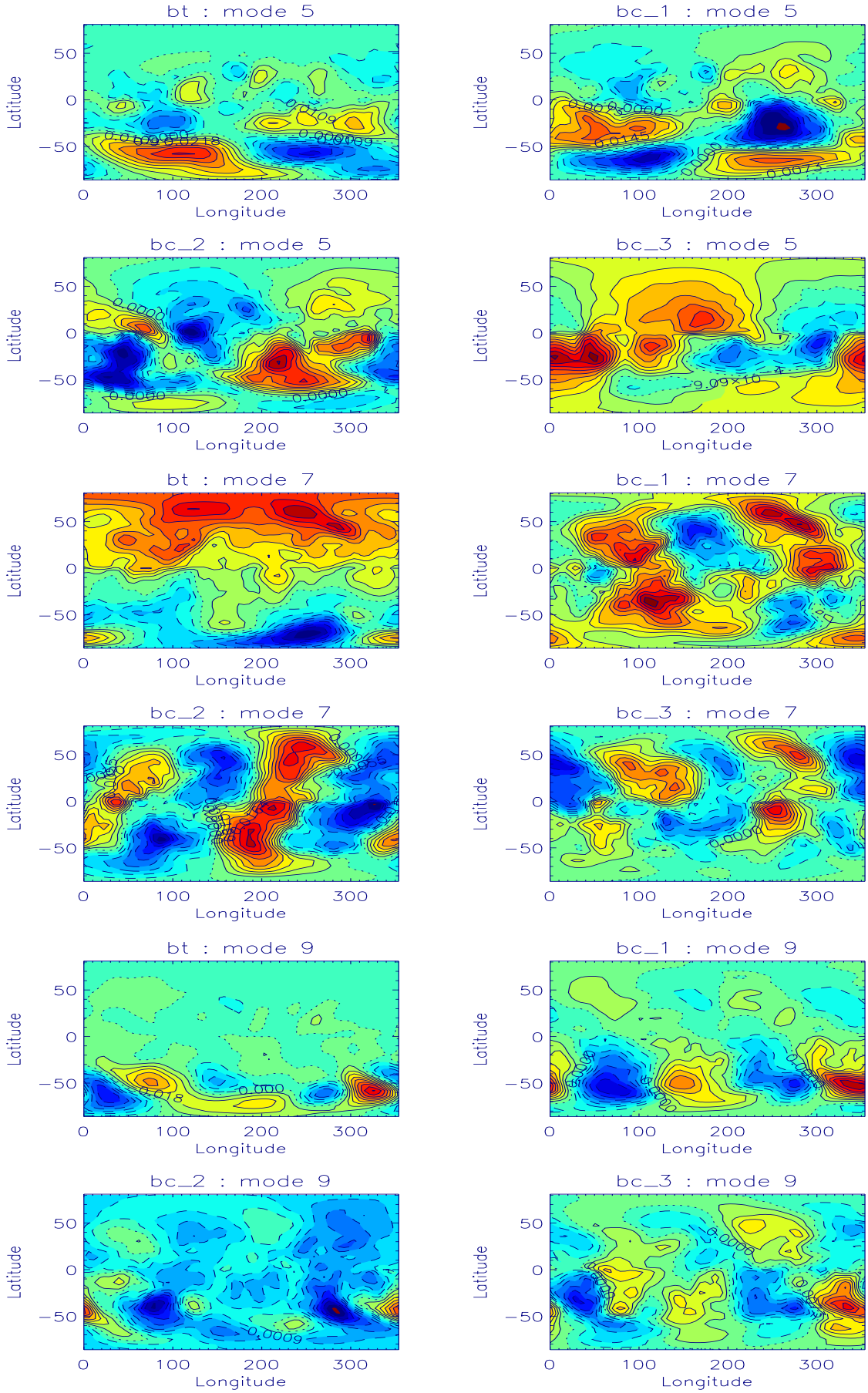


Figure 3.33: The barotropic (bt) and leading three baroclinic components (bc_1, bc_2 and bc_3) of modes 5, 7 and 9 using the TE norm.

3.4.4 MGCM Principal Components and Frequency Spectra

Figure 3.35 shows the temporal behaviour of the departure spatial eigenmodes for the different norms over 20 Martian days during a northern winter run ($L_s = 239.6°$ to $300.42°$). There is strong periodic behaviour in PCs 1 and 2, where the 1 day waves oscillate between the nondimensionalised values [-9,9]. The structures are wave-like, with the two leading eigenvectors forming conjugate pairs that have almost the same shape but are displaced one quarter of a period out of phase. This indicates that the pair explain a recurrent oscillation pattern in the time series (see Vautard et al. (1992)). The leading two oscillations are virtually unaffected by the choice of norm, but the higher-order modes appear to be more variable. However, the leading three temporal eigenmodes are very similar and contain evident periodic structures.

The power spectra profiles of PCs 1 and 2 are shown in Figure 3.36(a) where it can be seen that a large frequency peak occurs at 1 cycle/sol for each of the norms. This confirms that POD modes 1 and 2 represent a travelling wave, namely the diurnal tide which is a migrating wave and moves with the sun. The power spectra also pick out an additional frequency of about 0.18 cycles/sol, corresponding to a slowly oscillating 5.6 day oscillation which can easily be observed in Figure 3.34. Therefore, the two leading PCs oscillate with frequencies associated with the diurnal tide and a low frequency baroclinic wave. This suggests that there may exist some degree of nonlinear interaction between the high and low frequency oscillations. It can be concluded then that each norm efficiently extracts the diurnal tide in the leading two eigenmodes, which according to Figure 3.27 accounts for 73.72 % of the variance, 71.53 % of KE and 68.14 % of TE of the wave flow if the correlation, KE and TE norms are respectively used. It is observed that the shape and frequency of the leading temporal oscillations are weakly affected by the choice of norm.

PC 3 behaves almost identically to PCs 1 and 2, the only significant difference being that the correlation PC now oscillates with a reduced amplitude, being almost half the size of the correlation PCs 1 and 2. The regular periodic 1 day wave only exists in PCs 1 to 3 and thus suggests that the leading three POD modes are required to capture the diurnal tide. According to the eigenspectra profiles illustrated in Figure 3.27, the leading three spatial POD modes would account for 91.41 % of the variance, 82.61 % of the KE and 90.62 % of the TE of the wave field if respectively the correlation, KE and TE norm is used.

The KE and TE norms produce similar oscillatory behaviour and frequency spectra in PC 4 (seen in Figure 3.36(a)), which have periods ranging between 6-10 days (corresponding to frequencies of 0.1-0.18 cycles/sol). The correlation mode contains a small, distorted 1 day oscillation structure, which is evident by observing the small spectra peak at 1 cycle/sol in Figure 3.36(a), as well as a low frequency oscillation of about 0.06 cycles/sol which corresponds to about a 17 day wave. The form of the oscillations in the KE and TE PC 5 are almost identical to those observed in the KE and TE PC 4, which confirms that KE and TE modes 4 and 5 form complex conjugate pairs and thus describes a travelling wave. However, the 1 cycle/sol frequency (which corresponds to the diurnal tide) is extracted far more cleanly in PC 5 by each of the norms.

The correlation POD modes 5 and 6 form a complex conjugate pair, and this is evident in the frequency spectra (in Figure 3.36(b)), which shows a dominant peak at about 0.18 cycles/sol for both modes. The correlation and KE norms produce quasi-periodic oscillations in PC 6, where Figure 3.36(b) shows that the largest peak occurs at about 0.18 cycles/sol with an additional small peak at 1 cycle/sol. The TE PC contains oscillations which have frequencies of 1 and 2 cycles/sol, the largest peak occurring at the latter frequency. The TE POD mode 6 contains a dominant spatial wavenumber 2 structure, which indicates that the TE norm is extracting both the diurnal and the semi-diurnal tide, illustrating that the POD method selects the most physically significant modes in the MGCM data set. As the TE mode 6 has no conjugate pair this then suggests that this component of the semi-diurnal tide is a standing wave. Such findings are significant since it can be seen that the PODs (and PCs) can be separated into modes which represent the leading baroclinic mode, the diurnal and semi-diurnal tide, and it can also be determined whether the spatial patterns evolve as travelling waves, or whether they represent standing waves.

It can be seen from PCs 7 to 10 in Figure 3.36(b) that the norms produce oscillations with almost identical frequencies. In each of the power spectra figures of PCs 7 to 10 a cluster of peaks is generally found around about 0.05 cycles/sol, a frequency range of 0.2-0.8 cycles/sol, as well as at 1 (diurnal tide) and 2 cycles/sol (semi-diurnal cycle). In PCs 7 to 9 the dominant spectra peaks occur at a frequency of 2 cycles/sol, and it is interesting to note that POD modes 7 to 9 are dominated by spatial wavenumber 2 structures, revealing that the semi-diurnal cycle is efficiently extracted by the POD method.

PCs 7 and 8 in Figure 3.35 look somewhat complicated and erratic, however, the power spectra profiles observed in Figure 3.36(b) reveal that the dominant frequencies lie in the 0.2-0.8 cycles/sol range, as well as at 1 and 2 cycles/sol. Interestingly, each norm picks out these frequencies, which indicates that even though the PCs have rather complicated oscillatory forms the power spectra analysis reveals that they can be decomposed into nearly identical frequencies. The spatial patterns contained in POD modes 7 and 8 are again dominated by longitudinal wavenumber 2 structures, confirming the presence of the semi-diurnal tide. Even though POD modes 7 and 8 (for each of the norms) are spatially similar and have comparable PC behaviour, they do not form conjugate pairs and so each of the spatial patterns depict standing waves. Therefore, the semi-diurnal tide does not propagate in time.

PC 9 is very similar to PC 8, except that the power spectra peaks at 2 cycles/sol are considerably smaller in PC 9 if the KE or the TE norm is used. Significantly, the quality of the spatial wavenumber 2 structures in the KE and TE POD mode 9 are poorer than that which is observed in POD mode 8. The correlation POD mode 9 has a regular wave number 2 structure, and its PC contains a large power peak at a frequency of 2 cycles/sol. The TE norm picks out a dominant low frequency oscillation, confirmed by the large power peak at approximately 0.06 cycles/sol, which corresponds to a 17 day wave.

PC 10 is dominated by a low frequency oscillation of approximately 0.06 cycles/sol, with additional power peaks at about 0.3, 0.45, 0.8, 1 and 2 cycles/sol. Therefore, even the most

complicated PC oscillatory forms contain almost the same frequencies as those found in the leading eigenmodes. This suggests that the progressively more complicated PC oscillatory behaviour in the higher-order modes is more to do with the various excitation of these frequencies, and presumably their interaction has resulted in the higher-order temporal eigenmodes taking on seemingly erratic oscillatory forms.

Summary

The eigenspectra profiles of the MGCM wave flow (shown in Figure 3.27) illustrates that there exists three dominant eigenmodes, which accounts for most of the wave energy in the system (nearly 90 %). These modes have been identified as representing the diurnal cycle which implies that the diurnal tide is effectively captured by the leading three modes, irrespective of which norm is used.

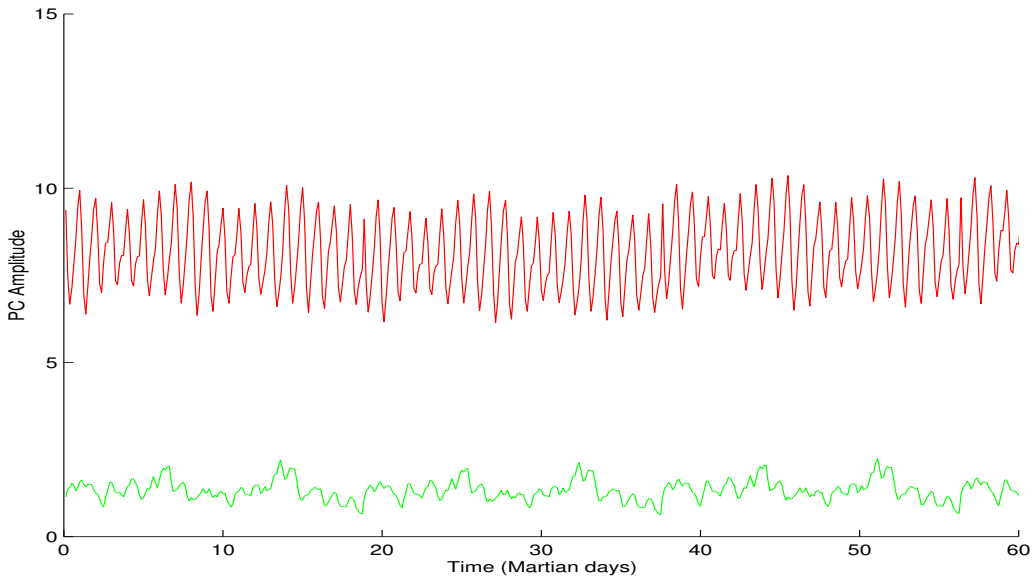


Figure 3.34: Root mean square principle component amplitudes for various eigenvectors. The figure shows plots of $\sqrt{(p_1)^2 + (p_2)^2 + (p_3)^2}$ (red line) and $\sqrt{(p_4)^2 + (p_5)^2}$ (green line) obtained from the TE norm during 60 days of the MGCM data.

The regular nature of the MGCM baroclinic waves is evident in the spatial fields. Two definite oscillatory patterns have been found to exist in the leading PCs, namely the diurnal tide (a high frequency oscillation) and a 6-10 day wave (a low frequency pattern), where the time evolution of the diurnal tide and the leading baroclinic mode (obtained from the TE norm) are shown in Figure 3.34. It can clearly be seen that the the leading wave (plotted in red) is a 6 day wave with a superimposed 1 day wave, whereas the other wave (in green) is a 1 day wave with a superimposed 10 day wave. The POD method extracts both forced and free waves[6], where the presence of the forced diurnal tide is induced by the diurnal fluctuations in the solar heating. The behaviour of the various PCs seems to suggest that there is some degree of interaction between these wave types, a mixing between the forced diurnal tide and the leading 6-10 day baroclinic wave. This contrasts with the work of Collins et al. (1996)

[6]Forced waves must be continually maintained by an oscillation mechanism of a given wave speed and wavenumber, whereas free waves are unmaintained.

who, using the MGCM data set, found a low frequency (7.2 day) and a high frequency (either a 3.4 or a 2.4 day) wave in the leading PCs, and argued that the observed Martian baroclinic waves in frequency space are bimodal i.e. there are two dominant oscillating states. It must be noted that in our analysis oscillations with frequencies of about 0.29 and 0.41 cycles/sol (corresponding to 3.4 and 2.4 day waves) are observed in the higher-order PCs, being particularly evident in PCs 7 to 10. This then suggests that the use of the vertical functions (the barotropic and leading three baroclinic modes) and their inclusion in the state vector have created discrepancies between our findings and those of Collins et al. (1996) in the spatial POD patterns and the oscillatory nature of the time varying PCs, resulting in only two propagating wave patterns to be found in the leading order modes, namely the diurnal tide and the leading 6-10 day baroclinic wave, as well as a 3.4 and 2.4 day baroclinic wave in the higher order modes.

It can be concluded that the nonlinear interactions between the baroclinic waves and the thermal tide are important in determining the spatial structure of the PODs and the temporal behaviour of the PCs. Recalling that the leading order PCs of the SGCM departure flow revealed two travelling waves, respectively propagating at almost pure frequencies of about 0.23 and 0.08 cycles/sol, then this contrasts with what has been found from the MGCM analysis, where the leading modes contain quasi-periodic structures, a diurnal tide as well as a 6-10 day baroclinic wave. This highlights the fact that the MGCM is a far more sophisticated model, which includes the forced diurnal tide and topography, suggesting that a successful low-order Galerkin model seems far more likely for the SGCM case. The spatial and temporal complications which arise from the thermal tide/baroclinic wave interaction in the MGCM hints that the reproduction of the implicit complexities of the data via a set of MGCM ODEs is beyond the scope of this thesis.

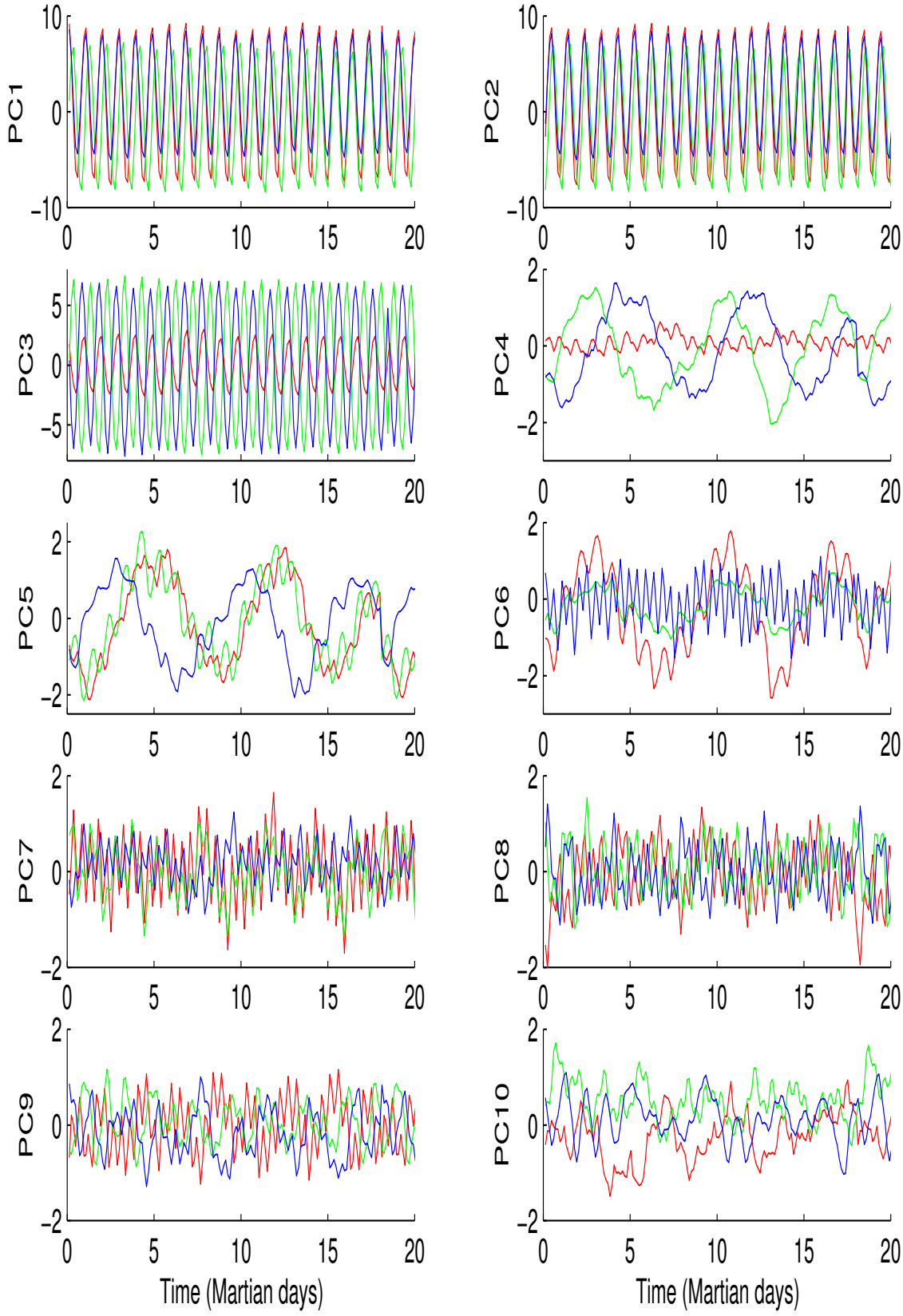
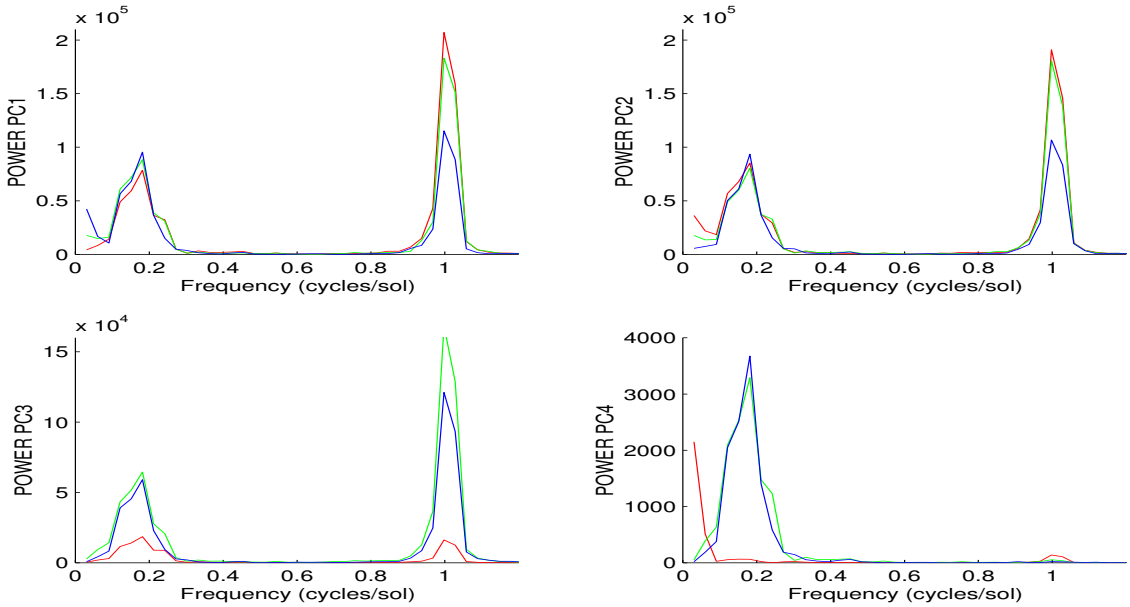
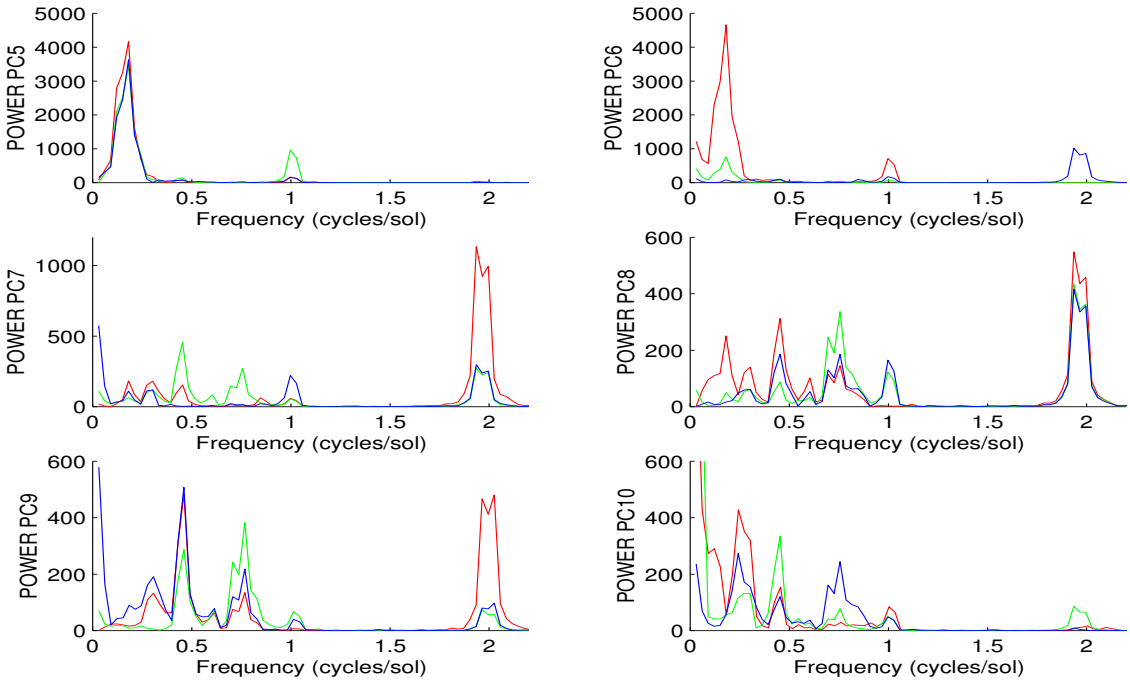


Figure 3.35: Original departure MGCM principal components (PCs) using the correlation (red), the KE (green) and the TE norms (blue) shown during an interval of 20 Martian days. PCs 1 and 2 represent the diurnal cycle, i.e. a 1 day wave, PCs 4 and 5 depict a low-frequency wave of period of about 10 days. The higher order modes show little similarity.



(a) Power spectra profiles of the original MGCM departure PCs 1 and 2 using the correlation (red), the KE (green) and the TE norms (blue). The diurnal tide is strongly evident in PCs 1, 2 and 3.



(b) Power spectra profiles of the original MGCM departure PCs 6 to 10 using the correlation (red), the KE (green) and the TE norms (blue). Instead of one dominant frequency band, here there exists a floor of frequency values, suggesting that there is considerable interaction between many frequencies.

Figure 3.36: Power spectra profiles of the original MGCM departure PCs where (a) shows the frequencies of PCs 1 to 4 and (b) shows the frequencies of modes 5 to 10 using the correlation (red), KE (green) and TE (blue) norms.

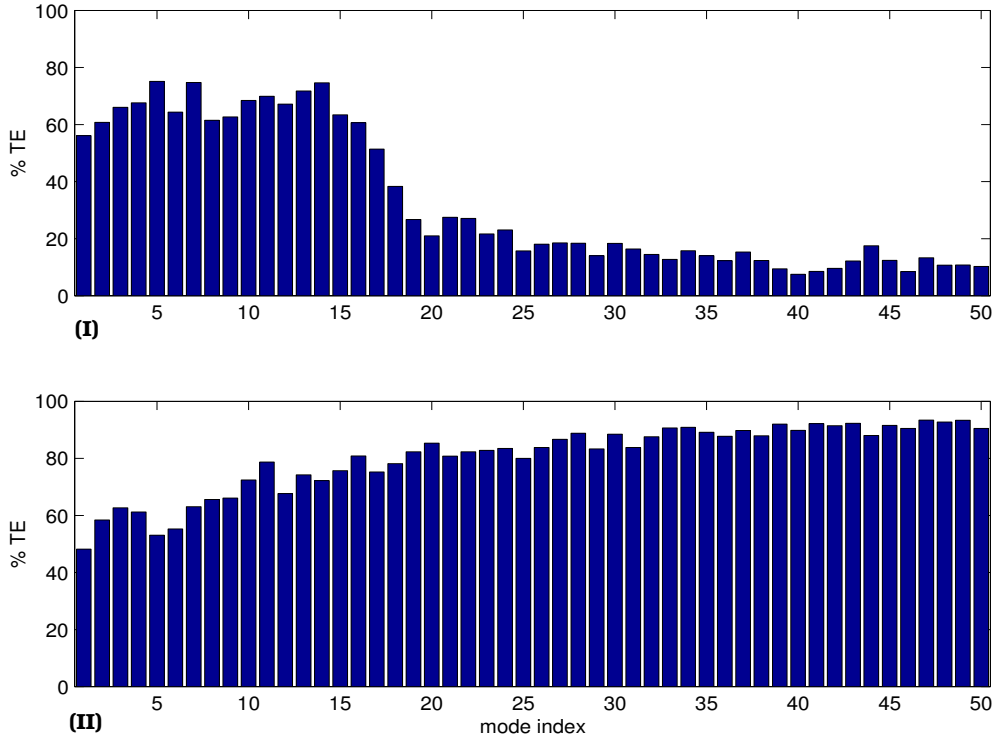
3.4.5 Energy in the MGCM eigenmodes

The streamfunction data was decomposed into a basic state, mean-flow correction (MFC) and a wave flow. It has been calculated that the zonal flow (= basic state + MFC) accounts for 99.68 % of the total variance, 99.07 % of KE and 96.89 % of TE if the correlation, KE and TE norms are respectively used.

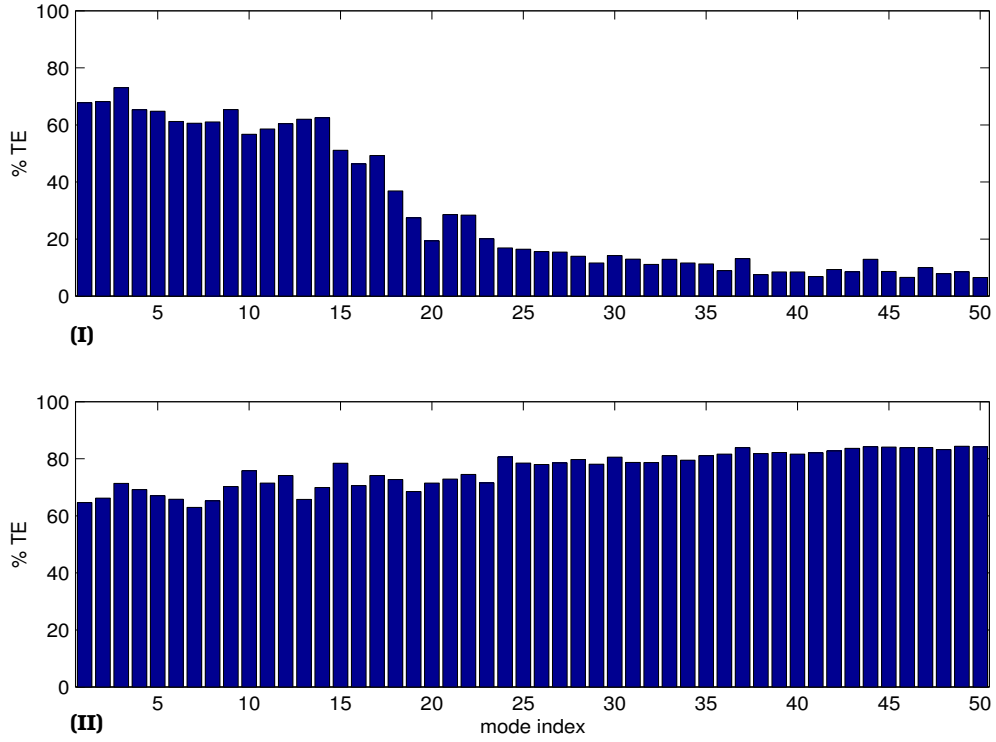
Figure 3.37(a) shows KE as a fraction of TE in the zonal POD modes using the (I) KE and (II) TE eigenvectors. It can be seen that the TE norm captures the APE of the original system far more efficiently than the KE norm, as is evident in the amount of KE contained in the leading two eigenmodes. As the two leading zonal modes account for 99.993 % of the variance, 99.968 % of KE and 99.878 % of TE (Figure 3.26) if the correlation, KE or TE norms are respectively used, then the amount of KE and APE represented in these two modes will almost be the amount of TE captured in the POD system. Figure 3.37(a) shows that the energy in the first two KE eigenvectors (in I) is predominantly KE, accounting for 56.9 and 62.4 % of TE in these two modes. This suggests that the KE norm has inefficiently extracted the MGCM APE supply as the original model contains almost equal amounts of KE and APE (see Figure 2.6). However, the energy content in the leading two TE eigenvectors (in II) shows that the retained KE and APE is in almost equal amounts as 48.2 and 58.4 % of TE is in the form of KE. This confirms that the KE norm has effectively captured the KE, but placed insufficient emphasis on capturing the APE supply.

Figure 3.37(a) also illustrates that the KE norm (in I) places a large proportion of the extracted KE into the leading order modes as there is a very fast tail-off in the amount of KE contained in the eigenmodes, particularly after the first 11 KE eigenvectors. This contrasts with the TE eigenmodes (in II) which contain a very significant proportion of KE in the higher-order modes (about 90 %). This suggests that the TE eigenvectors have placed most of the APE in the leading order modes and so have effectively optimised the capture of APE supply.

Figure 3.37(b) shows KE as a fraction of TE in the departure POD modes using the (I) KE and (II) TE eigenvectors. From the eigenspectra profiles of the wave flow in Figure 3.27 it was shown that the leading three eigenmodes captured 91.41 % of the variance, 82.61 % of KE and 90.02 % of TE of the wave flow if the three norms are respectively used, which implies that the amount of KE and APE represented in these three leading eigenmodes will almost account for the total amount of TE captured in the POD departure modes. The amount of KE retained in the leading three departure modes is very similar between the KE and TE eigenvectors (shown in I and II) which demonstrates that the TE norm also effectively extracts the KE from the original system. The leading KE wave modes are dominated by KE (accounting for about 70 % of TE), but after 13 KE eigenvectors there is a steep tail-off in the KE percentage, falling to under 20 % in the higher-order modes. This contrasts with the higher-order TE vectors which are dominated by KE. This again suggests that the KE norm pushes the majority of its KE supply into the leading order modes, whereas the TE norm pushes the APE into the lower-order modes, besides effectively capturing KE.



(a) KE as a fraction of the TE contained in each of the zonal modes using the (I) KE and (II) TE eigenvectors.



(b) KE as a fraction of the TE contained in each of the wave modes using the (I) KE and (II) TE eigenvectors.

Figure 3.37: KE as a fraction of TE contained in (a) the zonal modes and (b) the wave modes for the (I) KE and (II) TE eigenvectors.

3.5 Conclusions

By formulating the POD modes in spectral space a high resolution could be maintained and not at the expense of CPU time. By defining the state vector **V** to contain the barotropic and first three baroclinic functions then 97% of the TE of the original system can be captured. Having already truncated the system vertically, it was then important to implement the most effective method in capturing the remaining energy types of the system. Energy norms were used so as to maximise the amount of KE and TE retained. The KE and TE energy norms were the only norms capable of monitoring the energy exchanges and hence capturing the most dynamically important modes.

It has been shown that the PODs are not uniquely defined but are sensitive, for instance, to the choice of an inner product. We have discussed the differences between the eigenvectors which describe deviations from the mean state (as standard EOFs do), and those which describe the total state vector. It was shown that the KE norm places a large proportion of the extracted KE into the leading order eigenmodes but insufficiently captures APE. It was also shown that after about 13 KE eigenvectors there was a steep tail-off in the amount of KE contained in each of the eigenmodes, where the higher-order modes were dominated by APE. This can be explained by the fact that truncated models based on KE eigenvectors place very little emphasis on baroclinic patterns, so interactions involving the temperature field are poorly represented resulting in the energy conversion processes not being well described. In a truncated TE model, the energy cycle is better described since the TE norm places extra weight on the temperature field. This was suggested from the work of Selten (1993) who argued that a truncated TE model reproduces the energy cycle substantially better than the KE model (truncated to the same number of components) since the TE norm places extra weight on the temperature gradient.

Modal similarities were a feature of the SGCM lower-order modes, where the correlation and energy norms created spatial eigenmodes which were almost identical in amplitude and form. Structural differences were visible if the higher-order modes were compared. It was shown that the energy norms tended to extract symmetrical patterns and often produced smaller-scale spatial structures. An inspection of the corresponding PCs showed that the dominant oscillations of PCs 1 and 2 were roughly sinusoidal 4 day periodic structures (high frequency waves) whereas correlation PCs 3 and 4, and KE and TE PCs 3 and 5 were approximately 10 day waves (low frequency waves). The results presented showed that the frequency distribution of the highly regular baroclinic wave activity in the idealised Martian atmosphere was nearly bimodal.

The eigenspectra profiles of the MGCM wave flow illustrated that there were three dominant wave modes, which accounted for most of the wave energy of the system (about 90%). The loss of the bimodal behaviour which was observed in the SGCM may possibly be due to a loss of symmetry (argued by Aubry et al. (1991)), since the spatial wavenumber patterns were severely modulated by orography and the diurnal tide.

The regular nature of the MGCM baroclinic waves was evident in the POD spatial fields. The

waves were not simple sinusoidal modes, and their surface amplitude were significantly modulated by the orography. However, the various norms generated different spatial structures in the leading wave modes unlike that which was observed in the SGCM data (where the leading two wave eigenmodes were almost insensitive to the choice of norm). This seems to suggest that the greater baroclinicity and modal activity in the MGCM results in greater differences in the spatial patterns produced by the various norms.

Two definite wave patterns were found to exist in the leading MGCM PCs, namely the diurnal tide (a high frequency wave) and a 6-10 day wave (a low frequency pattern). The leading wave modes have been identified as representing the diurnal cycle, indicating that it was effectively extracted, irrespective of which norm was being used. The behaviour of the various PCs seemed to suggest that there was some degree of interaction between these two wave types.

We now want to test the idea of using PODs as a dynamical basis in a simple atmospheric model (the SGCM). Since a POD basis will effectively capture the variance of a system, then at any time the state of the system is accurately described with a relatively small number of PODs. The assumption is that the patterns onto which little variance is projected have a negligible influence on the evolution of the dominant patterns. The minimal requirements for such a model are that it has a vertical structure to allow for energy conversion by the baroclinic instability mechanism and that it is small enough to allow for extensive calculations.

As the SGCM data is thought to exhibit low-dimensional behaviour it is our aim to produce a Galerkin model which is capable of reproducing realistic dynamical flows (using the spherical QG PV equations). It is our aim to study the possibility of using PODs as a basis to obtain a reasonable low-order SGCM i.e. to determine whether a model based on a relatively small number of POD modes can reproduce the global structure of the atmospheric attractor.

The mathematical formulation of the Galerkin projection (which uses the spherical QG equations and the idealised SGCM data) will be discussed in Chapter 4.

Chapter 4

POD-Galerkin Projection

4.1 Spectral Galerkin methods

The dynamics of the weather and climate system is simulated by partial differential equations (PDEs) describing the underlying physical processes. These equations can be transformed to a set of n time dependent ordinary differential equations (ODEs), if the space variability is expanded into a set of n orthogonal functions. An example of this is the spectral truncation of the hydrodynamic equations for the atmospheric flow by Silberman (1954).

The aim is to obtain a system of ODEs by utilising the Galerkin technique. Having identified a suitable set of POD modes, the dependent variables are expanded as sums of products of temporally dependent amplitude coefficients and the spatially dependent eigenfunctions, and the series is truncated. Then it is necessary to develop a system of n first-order differential equations by taking inner products with each basis function in turn.

The immediate advantage of using the POD method in the Galerkin projection is that it yields an optimal set of basis functions (in the sense that it compresses the variance of the signal into the leading order modes more effectively than any other decomposition), and the resulting truncated system of ODEs captures the maximum amount of energy among all possible truncations of the same order.

Suppose the governing equations of the system under consideration are of the form:

$$\frac{\partial A}{\partial t} = \mathcal{L}A + \mathcal{N}A, \tag{4.1}$$

where $\mathcal{L}$ is a linear, $\mathcal{N}$ is some non-linear operator and A is the spatial and temporal function which must be determined. Retaining n POD modes the flow vector A is:

$$A(x,t)_n = \sum_{i=1}^{n} a_i(t)\phi_i(x), \tag{4.2}$$

where the $\{a_i(t)\}_{i=1}^n$ are the temporal amplitudes and $\{\phi_i(x)\}_{i=1}^n$ the spatial components of

the i-th POD mode. Defining the inner product of two functions $\phi(x)$,$\psi(x) \in L^2(x)$ as

$$(\phi, \psi) = \int_\Omega \phi(x)^* \psi(x) dx, \tag{4.3}$$

where $(\phi_i, \phi_j) = \delta_{ij}$. Substituting (4.2) into (4.1) and projecting the equation onto each POD mode in turn, we obtain

$$\left(\phi_j, \frac{\partial}{\partial t} A_n - \mathcal{L}(A_n - \mathcal{N}(A_n)) \right) = 0, \tag{4.4}$$

for $j = 1, .., n$. The POD model of (4.4) can be alternatively written as

$$\frac{d}{dt} a_i \quad = \quad \alpha_i + \beta_{ij} a_j + \gamma_{ijk} a_j a_k, \tag{4.5}$$

for $i = 1, ..., n$, where α_i are constants, β_{ij} are the coefficients of the linear terms and the interaction coefficients γ_{ijk} are coefficients of the quadratic terms. Here identical indices in a term imply a summation and the interaction coefficients are anti-symmetric, that is $\gamma_{ijk} = -\gamma_{ikj}$. In the case of spherical harmonics, the majority of the interaction coefficients γ_{ijk} are zero but for PODs this is not the case. The coefficients in the evolution equations of the POD amplitudes in (4.5) are determined numerically.

Since the data is represented on a regular discretised grid, the n-dimensional system in (4.5) can be written as a matrix equation:

$$\frac{d\mathbf{a}}{dt} = L\mathbf{a} + N\mathbf{a}, \tag{4.6}$$

where $\mathbf{a}$ $(= (a_1, .., a_n))$ is the approximate state vector and L and N are the linear and non-linear matrix operators respectively.

Thus the phase space containing the time evolution of the underlying process is spanned by the n different variables a_j, $j = 1, ..., n$ of the dynamical system. Portraits of the time evolution of the system are formed by trajectories in this n-dimensional phase space. They exhibit distinct patterns, possibly revealing attractors which trap trajectories after transients (originating from the initial states) have decreased. All initial values, whose trajectories are captured by the attractor, define the attractor basin. Fraedrich (1986) states that there are one or more steady states (i.e. zero dimensional points) which belong to related attractor basins; periodic orbits or limit cycles are one-dimensional closed lines and quasi-periodicity occurs on two-dimensional tori in phase space. Irregular or chaotic motions can appear on strange attractors in three and higher dimensional phase spaces (for further details see § 5.7.3).

4.2 Derivation of the Galerkin spherical QG model

Recalling that the adiabatic, spherical QG equation in isobaric coordinates can be written as

$$\left[\frac{\partial}{\partial t} + J(\psi, \cdot) \right] \left[\nabla^2 \psi + f_o^2 \frac{\partial}{\partial p} \left(\frac{1}{S} \frac{\partial \psi}{\partial p} \right) + f \right] = -r \nabla^2 \psi, \tag{4.7}$$

where ψ is the streamfunction; r is the Ekman dissipation parameter; $f = 2\Omega \sin\theta$ is the full Coriolis parameter; Ω is the rotational rate; f_o is a synoptic scale of motion for the Coriolis parameter; $S = \frac{N^2}{g^2\rho^2}$ is the stratification parameter where N is the Brunt-Väisälä frequency; J is the Jacobian operator for the advection terms and is given by

$$J(f,g) = -\frac{1}{\cos\theta}\left(\frac{\partial f}{\partial\theta}\frac{\partial g}{\partial\phi} + \frac{\partial f}{\partial\phi}\frac{\partial g}{\partial\theta}\right), \tag{4.8}$$

for two arbitrary functions f and g and the Laplacian operator in spherical coordinates is given by

$$\nabla^2 = \frac{1}{a^2\cos^2\theta}\left[\frac{\partial^2}{\partial\phi^2}(\quad) + \cos\theta\frac{\partial}{\partial\theta}(\cos\theta\frac{\partial}{\partial\theta}(\quad))\right]. \tag{4.9}$$

The streamfunction for the horizontal flow will be decomposed into a wave, a zonal mean-flow correction (MFC) and a basic state term. Taking $\psi(\theta,\phi,t)$ to be the total streamfunction, then

$$\psi(\theta,\phi,t) = \tilde{\psi}(\theta,\phi,t) + \bar{\psi}(\theta,t) + \hat{\psi}(\theta), \tag{4.10}$$

where the tilde denotes the wave mode, the overbar the MFC mode and the hat the time independent basic state (see § 3.3.1, § 4.2.1 and § 5.3.2 for further details). The departure flow ($\tilde{\psi}$) and the mean-flow correction ($\bar{\psi}$) are defined as:

$$\tilde{\psi} = \psi - \int_{\phi=0}^{2\pi} \psi\, d\phi, \quad \bar{\psi} = \int_{\phi=0}^{2\pi} \psi\, d\phi - \hat{\psi}, \tag{4.11}$$

and so the mean-flow correction is defined as the time dependent difference between the instantaneous zonal flow and the basic state, i.e. the component of the mean flow which varies in time. It is assumed that the background zonal flow (or basic state) is steady. The basic state may be unstable to waves, whose nonlinear interactions may feed back to alter the zonal flow.

The wave and mean-flow correction perturbations to the basic state are expanded in orthonormal bases which are derived by the POD method (detailed in § 1.4). Thus we expand

$$\psi(\theta,\phi,t) \simeq \hat{\psi}(\theta) + \sum_{i=1}^{n} a_i(t)\bar{\phi}_i(\theta) + \sum_{j=1}^{m} b_j(t)\tilde{\phi}_j(\theta,\phi), \tag{4.12}$$

where the overbar denote the mean-flow correction terms (n modes in total), the tildes the wave terms (m modes) and the hat the time-independent basic state.

The state vector $\mathbf{a}(t)$ for the system in (4.6) is defined as

$$\mathbf{a} = [a_1,\ldots,a_n,b_1,\ldots,b_m]^\top, \tag{4.13}$$

and the problem is to compute the ODEs in the form $\dot{\mathbf{a}} = L\mathbf{a} + N\mathbf{a}$.

4.2.1 Defining the basic state

The generation of the low-dimensional set of Martian ODEs is based upon a basic state being specified (which is consistent with the original data) and then analysing the subsequent evolution of wave perturbations. If the waves remain small, linear theory predicts the spatial form of the waves, whether they will grow or decay, besides their frequency. Growing perturbations may eventually equilibrate at finite amplitude through a process of nonlinear interactions, and this may feed back to change the background zonal flow.

There are two possible ways to define the above decomposition of its mean flow from the SGCM data. The simplest scheme is to take the time average of the zonally averaged flow, and define this to be the basic state. Alternatively, it is possible to run a 2-D axisymmetric version of the numerical model (by suppressing the waves). This was done with the SGCM for 400 sols to ensure that the state had equilibrated and then the steady state was used as the basic state. From our calculations, it was evident that there was very little difference in the form of the barotropic and baroclinic components of these two differently produced states (see § 3.3.1), and so for consistency, the zonally symmetric version was used as the basic state (see § 5.3.2).

4.2.2 PODs or spherical harmonics ?

According to Selten (1995) there are clear advantages for using spherical harmonics (instead of PODs) since both kinetic energy (KE) and enstrophy (ENS) are conserved for arbitrary truncations when forcing is used. This is because spherical harmonics are eigenfunctions of the Laplace operator and so when the Galerkin method is used the projection operator commutes with the Laplace operator.

However, PODs are not eigenfunctions of the Laplace operator, and the projection does not commute with the Laplace operator. Selten argues that if there is an interaction between two localised structures, then the truncation of the expansion in spherical harmonics misrepresents the interactions. By using spherical harmonics the existence of recurrent structures in the atmosphere are not exploited whereas the motivation for using PODs is that they describe the most energetic recurrent structures in the circulation.

Following Penland and Ghil (1993), it could be argued that since EOFs themselves are determined by the full nonlinear dynamics, a POD model automatically includes some of the nonlinear effects. In contrast, the dynamics has no influence in determining the shape of the spherical harmonics, which are simply eigenfunctions of Laplace's equation on a sphere. Both spherical harmonics and PODs separate large scales and small scales. However, PODs efficiently describe the dominant circulation structures and separate the slow and fast structures. In a truncated spectral model with spherical harmonics as basis functions, short waves are neglected that contribute to the description of large scale, localised circulation structures that evolve slowly in time. A low-order POD model resolves these slow flow structures and has incorporated that part of the short waves necessary for a good description of these structures. The small-scale PODs that are neglected evolve on short time scales and their effect on the

slow, large scale EOFs is mainly dissipative (see Selten (1995)).

4.2.3 Boundary conditions

Since the atmosphere is unbounded, it is difficult to implement a boundary condition (bc) which is physically realistic. We want the amplitude of the energy of the motion to decay to zero for large z (where z is the vertical height scale), taking the upper-parts of the atmosphere to correspond to $z \to \infty$. For the energy to be finite valued as $z \to \infty$, we require $\rho_s \psi^2$ to remain finite as $z \to \infty$, where ρ_s is the global average of the fluid density ρ at each level of z.

Alternatively, if the amplitude does remain finite then the energy flux of the motion must be in the vertical direction. If there are no sources of wave energy at infinity, then the radiation condition must be used so that all waves have an outward-directed energy flux.

The bcs have been addressed by Rempfer (1996), and can be categorised as either homogeneous and linear or nonhomogeneous. If the bcs are linear and homogeneous, then since the POD will each satisfy the bcs individually, any superposition of the eigenfunctions also satisfies the bcs. This is not the case for nonhomogeneous bcs where the use of additional constraints is necessary.

4.2.4 Calculating the vertical derivative term

It will now be shown how the vertical structure equation (discussed in Chapter 2) provides us with a useful identity, allowing the vertical activity of the original system to be modelled in the Galerkin process. It will be shown that it is unnecessary to explicitly calculate the vertical derivative term $f_o^2 \frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial \psi}{\partial p}\right)$ in (4.7).

By assuming the following separable form for the streamfunction data :

$$\psi(\theta, \phi, p, t) = \sum_{i=0}^{N-1} H_i(p)\tilde{\psi}_i(\theta, \phi, t), \tag{4.14}$$

where H_i is the i-th vertical baroclinic mode (i=0 is the barotropic mode), $\tilde{\psi}_i$ is the amplitude of the i-th baroclinic mode (as a function of θ, ϕ and t) and N is the number of vertically resolved modes. Recalling from (2.20) that the vertical structure equation in isobaric coordinates can be written as:

$$f_o^2 \frac{d}{dp}\left(\frac{1}{S}\frac{dH_l}{dp}\right) = -\lambda_l H_l, \tag{4.15}$$

and as (4.15) is satisfied for each of the vertical modes, we then have :

$$\sum_{l=0}^{N-1} f_o^2 \frac{d}{dp}\left(\frac{1}{S}\frac{dH_l}{dp}\right) = -\sum_{l=0}^{N-1} \lambda_l H_l. \tag{4.16}$$

If both sides of (4.16) are multiplied by $\tilde{\psi}_l$ we then obtain :

$$\sum_{l=0}^{N-1} f_o^2 \frac{d}{dp}\left(\frac{1}{S}\frac{dH_l}{dp}\right)\tilde{\psi}_l = -\sum_{l=0}^{N-1} \lambda_l H_l \tilde{\psi}_l, \tag{4.17}$$

and recalling that $\psi = \sum_{l=0}^{N-1} H_l\tilde{\psi}_l$ then (4.17) becomes :

$$f_o^2 \frac{d}{dp}\left(\frac{1}{S}\frac{d\psi}{dp}\right) = -\sum_{l=0}^{N-1} \lambda_l H_l \tilde{\psi}_l. \tag{4.18}$$

We note that the LHS of (4.18) is the term required in the original spherical QG equation given in (4.7). Recalling that the expansion for the i-th POD mode is of the form:

$$\Phi_i(\theta,\phi,p) = \sum_{k=0}^{N-1} \Phi_i^k(\theta,\phi,p) = \sum_{k=0}^{N-1} \alpha_{ik} H_k(p)\tilde{\psi}_k(\theta,\phi), \tag{4.19}$$

where N is the number of vertical modes, n is the horizontal truncation level for the POD mode expansion, Φ_i^k denotes the k-th vertical component of the i-th POD mode (taking account of the barotropic and baroclinic components), α_{ik} is the nondimensionalised coefficient of the k-th vertical component and the i-th mode and $\tilde{\psi}_k(\theta,\phi)$ is the amplitude of the time averaged k-th baroclinic mode. As the total streamfunction field ψ can be expanded in terms of POD modes, by using (4.19) we have the expression:

$$\psi(\theta,\phi,p,t) = \sum_{i=1}^{n} a_i(t)\ \Phi_i(\theta,\phi,p) = \sum_{i=1}^{n} a_i(t) \sum_{k=0}^{N-1} \alpha_{ik} H_k(p)\ \tilde{\psi}_k(\theta,\phi), \tag{4.20}$$

where a_i is the POD time-coefficient of the i-th eigenmode. Now, by using (4.15), (4.19) and (4.20) we obtain:

$$f_o^2 \frac{d}{dp}\left(\frac{1}{S}\frac{d\psi}{dp}\right) = -\sum_{i=1}^{n} a_i \sum_{k=0}^{N-1} \alpha_{ik}\lambda_k H_k \tilde{\psi}_k = -\sum_{i=1}^{n}\sum_{k=0}^{N-1} a_i \lambda_k \Phi_i^k. \tag{4.21}$$

We note that as the horizontal and vertical truncation limits n and N increase, the representations given in equations (4.19) and (4.21) will become more accurate.

This illustrates that the vertical derivative term in (4.7) can be represented by a POD mode expansion which is given in (4.21). Here each of the k-vertical components of the truncated set of spatial eigenfunctions are multiplied by their associated eigenvalue (or Froude number) which has been obtained from the vertical structure equation given in (4.15).

We note that the vertical derivative of the first vertical component produces a zero vector since $\lambda_1 = 0$ (associated with the barotropic component). This is because the barotropic mode is uniformly constant throughout the vertical levels and hence its vertical derivative must be zero.

4.2.5 Rescaling the spherical QG equations

The SGCM produces nondimensionalised data and so the dynamical fields and the independent variables of the spherical QG equations given in (4.7) must also be consistently nondimensionalised in order that the solutions of the Martian low-dimensional model may be directly compared to the original SGCM time series.

The spherical QG equations in (4.7) were nondimensionalised using

$$t \to \left(\frac{1}{\Omega}\right) t' \quad \text{and} \quad \nabla \to \left(\frac{1}{a}\right) \nabla', \tag{4.22}$$

where the nondimensional variables are denoted by primes and $a = 3.394 \times 10^6$m is the radius of the planet and $\Omega = 7.08822 \times 10^{-5} s^{-1}$ is the rotational rate. This allows the horizontal spatial scale of variation to be given by the length scale a, whose velocities are characterised by the velocity scale Ω. Thus the equations have simply been scaled so that the relative order of each of the terms is clearly measured by the nondimensional parameter multiplying it and no approximations have been made.

4.2.6 The Galerkin projection of the Martian equations

We will now discuss how the spherical QG equations were transformed into a set of ODEs by the Galerkin projection. If ψ is substituted into the spherical QG equations in (4.7), and if the vertical condition given in (4.15) is used then we obtain:

$$\frac{\partial}{\partial t}\left(\nabla^2\psi - \sum_{l=0}^{N-1} \lambda_l H_l \tilde{\psi_l}\right) = -r\nabla^2\psi - J\left(\psi, \nabla^2\psi - \sum_{l=0}^{N-1} \lambda_l H_l \tilde{\psi_l} + f\right). \tag{4.23}$$

The series expansion for ψ (which is given in (4.12)) is now substituted into (4.23), and then the meanflow correction basis functions $\bar{\Phi}_k^p$ (where $k = 1, ..., n$) are each projected in turn onto the equations, resulting in a system of n dimensional ODEs of the following form:

$$
\begin{aligned}
&\sum_{p=0}^{N-1}\left(\sum_{i=1}^{n}\dot{a_i}(\bar{\Phi}_k^p,\nabla^2\bar{\Phi}_i^p)+\sum_{i=1}^{m}\dot{b_i}(\bar{\Phi}_k^p,\tilde{\Phi}_i^p)-\sum_{i=1}^{n}\lambda_p\dot{a_i}(\bar{\Phi}_k^p,\bar{\Phi}_i^p)-\sum_{i=1}^{m}\lambda_p\dot{b_i}(\bar{\Phi}_k^p,\tilde{\Phi}_i^p)\right)\\
&\quad=\sum_{p=0}^{N-1}\Bigg(-\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\nabla^2\bar{\Phi}_j^p))-\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_j(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\nabla^2\tilde{\Phi}_j^p))\\
&\qquad-\sum_{i=1}^{n}a_i(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\nabla^2\hat{\Phi}^p))+\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\lambda_p(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\bar{\Phi}_j^p))\\
&\qquad+\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_pa_ib_j(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\tilde{\Phi}_j^p))+\sum_{i=1}^{n}\lambda_pa_i(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,\hat{\Phi}^p))\\
&\qquad-\sum_{i=1}^{n}a_i(\bar{\Phi}_k^p,J(\bar{\Phi}_i^p,f))-\sum_{i=1}^{m}b_i(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,f))-(\bar{\Phi}_k^p,J(\hat{\Phi}^p,f))\\
&\qquad-\sum_{i=1}^{m}\sum_{j=1}^{n}b_ia_j(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\nabla^2\bar{\Phi}_j^p))-\sum_{i=1}^{m}\sum_{j=1}^{m}b_ib_j(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\nabla^2\tilde{\Phi}_j^p))\\
&\qquad-\sum_{i=1}^{m}b_i(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\nabla^2\hat{\Phi}^p))+\sum_{i=1}^{m}\sum_{j=1}^{n}\lambda_pb_ia_j(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\bar{\Phi}_j^p))\\
&\qquad+\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_pb_ib_j(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\tilde{\Phi}_j^p))+\sum_{i=1}^{m}\lambda_pb_i(\bar{\Phi}_k^p,J(\tilde{\Phi}_i^p,\hat{\Phi}^p))\\
&\qquad-\sum_{j=1}^{n}\lambda_pa_j(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\nabla^2\bar{\Phi}_j^p))-\sum_{j=1}^{m}\lambda_pb_j(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\nabla^2\tilde{\Phi}_j^p))\\
&\qquad-\lambda_p(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\nabla^2\hat{\Phi}^p))+\sum_{j=1}^{n}\lambda_pa_j(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\bar{\Phi}_j^p))\\
&\qquad+\sum_{j=1}^{m}\lambda_pb_j(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\tilde{\Phi}_j^p))+\lambda_p(\bar{\Phi}_k^p,J(\hat{\Phi}^p,\hat{\Phi}^p))\\
&\qquad-r\sum_{i=1}^{n}a_i(\bar{\Phi}_k^p,\nabla^2\bar{\Phi}_i^p)-r\sum_{i=1}^{m}b_i(\bar{\Phi}_k^p,\nabla^2\tilde{\Phi}_i^p)-r(\bar{\Phi}_k^p,\nabla^2\hat{\Phi}^p)\Bigg),
\end{aligned}
\tag{4.24}
$$

and similarly if the spatial wave eigenmodes $\tilde{\Phi}_k^p$ (for $k=1,...,m$) are projected on equation (4.23) a further m ODEs are produced, thus in total creating a set of $(n+m)$ nonlinear coupled ODEs. Because the basis functions are orthonormal, this procedure yields a coupled system of ODEs describing the evolution of the expansion coefficients. Since in practice the expansion of the basis functions is finite, it leads to a reduced form of the spherical QG equations.

4.2.7 Notation for inner products

It is now necessary to include some simplifying notation so that the matrix elements which are generated by the above summations can be analysed and easily categorised. We define

$$\begin{aligned}
(\bar{\Phi}_k^p, \bar{\Phi}_i^p) &= \bar{\bar{\alpha}}_{ki}, & (\bar{\Phi}_k^p, \lambda_p \bar{\Phi}_i^p) &= \bar{\bar{\alpha}}_{ki}^{\lambda}, \\
(\bar{\Phi}_k^p, \nabla^2 \bar{\Phi}_i^p) &= \bar{\bar{\beta}}_{ki}, & (\bar{\Phi}_k^p, \nabla^2 \lambda_p \bar{\Phi}_i^p) &= \bar{\bar{\beta}}_{ki}^{\lambda}, \\
(\bar{\Phi}_k^p, J(\bar{\Phi}_i^p, \bar{\Phi}_j^p)) &= \bar{\bar{\bar{\delta}}}_{kij}, & (\bar{\Phi}_k^p, J(\lambda_p \bar{\Phi}_i^p, \bar{\Phi}_j^p)) &= \bar{\bar{\bar{\delta}}}_{kij}^{\lambda}, \\
(\bar{\Phi}_k^p, J(\tilde{\Phi}_i^p, f)) &= \tilde{\bar{\eta}}_{ki}, & (\bar{\Phi}_k^p, J(\tilde{\Phi}_i^p, \nabla^2 \bar{\Phi}_j^p)) &= \bar{\bar{\bar{\zeta}}}_{kij}.
\end{aligned}$$

This then allows the $(m+n)$-dimensional ODE system to be expressed in matrix form, where the inner products form the elements of the matrices.

4.2.8 The matrix problem

The $(n+m)$-dimensional nonlinearly coupled ODE set is more conveniently written in matrix form, where the n equations in (4.24) and a further m equations (which are obtained by using $\tilde{\Phi}_k^m$ $(k = 1, ..., m)$ as the projecting mode) can be written as:

$$\mathcal{C}\dot{\mathbf{a}} = \mathbf{f} + \Lambda\mathbf{a} + \mathcal{A}\mathbf{a} + \mathcal{B}\mathbf{a} + \mathcal{D}\mathbf{a} + \mathcal{M}\mathbf{a} + \mathcal{W}\mathbf{a}, \tag{4.25}$$

where $\mathbf{f}$ is a constant vector, $\{\mathcal{C}, \Lambda, \mathcal{A}, \mathcal{B}, \mathcal{D}\}$ are constant $(n+m)$-dimensional square matrices, and $\{\mathcal{M}, \mathcal{W}\}$ are $(n+m)$-dimensional square matrices which are dependent upon the components of the state vector $\mathbf{a} = (a_1, ..., a_n, b_1, ..., b_m)^T$.

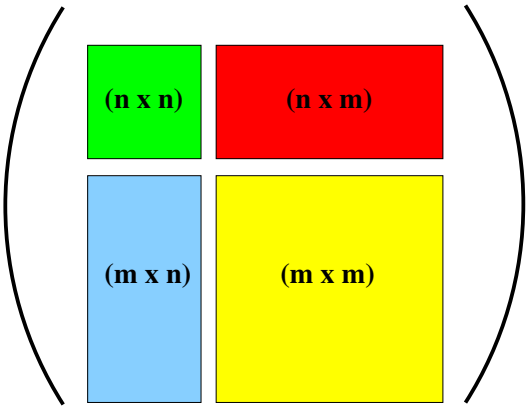


Figure 4.1: The block form of each of the matrices $\{\mathcal{C}, \Lambda, \mathcal{A}, \mathcal{B}, \mathcal{D}\}$ and $\{\mathcal{M}, \mathcal{W}\}$. Taking n to be the number of mean-flow correction modes and m the number of wave modes retained in the POD-Galerkin projection, then the resulting $(n+m)$ square matrices in (4.25) will have the diagonal blocks of matrix size $(n \times n)$ and $(m \times m)$ respectively, and the off diagonal blocks of matrix size $(n \times m)$ and $(m \times n)$.

Generally, matrix $\mathcal{C}$ couples the time derivatives in equation (4.25) through the vertical derivative term whereas the right hand side (RHS) of the system consists of a combination of dissipation terms, which are linear, and advection terms, which are both linear and nonlinear.

The dissipation of the wave and MFC terms are given by the linear matrix Λ and the linear advection terms which result from the basic state are included in matrices $\mathcal{A}$ and $\mathcal{B}$. The nonlinear advection terms arise from matrices $\mathcal{M}$ and $\mathcal{W}$ which respectively contain functions of the MFC and wave components of the state vector. In calculating the various interaction coefficients in matrix equation (4.25) it should be noted that since the meanflow correction terms and basic state are both independent of the zonal angle ϕ, then all these advection terms are necessarily zero.

Matrix $\mathcal{C}$ in (4.25) is an $(n+m)$-dimensional square matrix of the form:

$$\mathcal{C} = \begin{pmatrix} \bar{\bar{\beta}}_{ki} - \bar{\bar{\alpha}}^{\lambda}_{ki} & \tilde{\bar{\beta}}_{ki} - \tilde{\bar{\alpha}}^{\lambda}_{ki} \\ \bar{\tilde{\beta}}_{ki} - \bar{\tilde{\alpha}}^{\lambda}_{ki} & \tilde{\tilde{\beta}}_{ki} - \tilde{\tilde{\alpha}}^{\lambda}_{ki} \end{pmatrix}, \tag{4.26}$$

with the four submatrix blocks taking the form of that which is shown in Figure 4.1.

Vector $\mathbf{f}$ is a constant vector which arises from projecting all of the basis functions onto the basic state. It is defined as

$$\mathbf{f} = -r \begin{pmatrix} \hat{\bar{\beta}}_{k1} \\ \hat{\tilde{\beta}}_{k1} \end{pmatrix}, \tag{4.27}$$

and is an $(n+m)$ constant vector, generated by the Ekman dissipation terms of the basic state.

The matrix Λ in (4.25) arises from the linear dissipation meanflow and wave terms and is defined as

$$\Lambda = -r \begin{pmatrix} \bar{\bar{\beta}}_{ki} & \tilde{\bar{\beta}}_{ki} \\ \bar{\tilde{\beta}}_{ki} & \tilde{\tilde{\beta}}_{ki} \end{pmatrix}, \tag{4.28}$$

and matrix $\mathcal{A}$ in (4.25) arises from the advection of the wave terms by the basic state which gives rise to further linear terms and is defined as

$$\mathcal{A} = \begin{pmatrix} \mathbf{0} & -\hat{\bar{\zeta}}_{ki1} + \hat{\bar{\delta}}^{\lambda}_{ki1} \\ \mathbf{0} & -\hat{\tilde{\zeta}}_{ki1} + \hat{\tilde{\delta}}^{\lambda}_{ki1.} \end{pmatrix}. \tag{4.29}$$

Nonlinear terms are generated by the advection of the wave flow by the basic state in matrix $\mathcal{B}$ in (4.25) and is defined as

$$\mathcal{B} = \begin{pmatrix} B_{1,ki} & \mathbf{0} \\ \mathbf{0} & B_{2,ki} \end{pmatrix}, \tag{4.30}$$

which has a block diagonal structure, where the off diagonal blocks are zero and where the $(n \times n)$ and $(m \times m)$ sub-matrices $B_{1,ki}$ and $B_{2,ki}$ are defined as

$$\begin{aligned} B_{1,ki} &= -b_i \tilde{\bar{\zeta}}_{k1i} + b_i \tilde{\bar{\delta}}^{\lambda}_{k1i}, \\ B_{2,ki} &= -b_i \tilde{\tilde{\zeta}}_{k1i} + b_i \tilde{\tilde{\delta}}^{\lambda}_{k1i}. \end{aligned} \tag{4.31}$$

Advection of the wave flow by the MFC gives rise to further nonlinear terms, coupled by the matrix $\mathcal{M}$ which has the same block matrix structure as matrix $\mathcal{A}$. The only terms which appear are in the $i = (n+1), \ldots, (n+m)$ rows of the $(n+m)$ square matrix, where matrix

$\mathcal{M}$ is defined as

$$\mathcal{M} = \begin{pmatrix} \mathbf{0} & M_{1,ki} \\ \mathbf{0} & M_{2,ki} \end{pmatrix}, \tag{4.32}$$

and where the $n \times m$ and $m \times m$ sub-matrices $M_{1,ki}$ and $M_{2,ki}$ are defined as

$$\begin{aligned} M_{1,ki} &= -\sum_{j=1}^{n} a_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{n} a_j \tilde{\tilde{\delta}}^{\lambda}_{kij}, \\ M_{2,ki} &= -\sum_{j=1}^{n} a_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{n} a_j \tilde{\tilde{\delta}}^{\lambda}_{kij}. \end{aligned} \tag{4.33}$$

Additional nonlinear terms are created by the advection of the wave flow by the wave flow terms, which are coupled by the matrix $\mathcal{W}$ and has the following matrix block structure

$$\mathcal{W} = \begin{pmatrix} W_{1,ki} & W_{2,ki} \\ W_{3,ki} & W_{4,ki} \end{pmatrix}, \tag{4.34}$$

where the sub-matrices $W_{j,ki}, j = 1, ..4$ are defined as

$$\begin{aligned} W_{1,ki} &= -\sum_{j=1}^{m} b_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \tilde{\tilde{\delta}}^{\lambda}_{kij}, \\ W_{2,ki} &= -\sum_{j=1}^{m} b_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \tilde{\tilde{\delta}}^{\lambda}_{kij}, \\ W_{3,ki} &= -\sum_{j=1}^{m} b_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \tilde{\tilde{\delta}}^{\lambda}_{kij}, \\ W_{4,ki} &= -\sum_{j=1}^{m} b_j \tilde{\tilde{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \tilde{\tilde{\delta}}^{\lambda}_{kij}. \end{aligned} \tag{4.35}$$

Finally, the effects of the full Coriolis parameter f are included in matrix $\mathcal{D}$ which is defined as

$$\mathcal{D} = \begin{pmatrix} \mathbf{0} & -\tilde{\tilde{\eta}}_{ki} \\ \mathbf{0} & -\tilde{\tilde{\eta}}_{ki} \end{pmatrix}, \tag{4.36}$$

where the only terms which appear are in rows $i = (n+1), .., (n+m)$ of the $(n+m)$ square matrix.

4.2.9 Forcing the low-dimensional model

Schubert (1985) discussed one of the problems associated with using an atmospheric model, in how to introduce a driving or source term. He argued that a thermal driving term enters implicitly into the analysis due to the assumption of a fixed base state which is equal to the observed seasonal mean.

The constant vector **f** in (4.27) describes the dissipation of the basic state flow. Thus to maintain the basic state requires that this term be balanced by adding a forcing term equal to -**f** which is implemented by neglecting the constant vector. This forcing scheme ensures that the trivial solution $\mathbf{a} = \mathbf{0}$ (which corresponds to the basic state flow) will be a solution of the

model.

4.3 Overview of the general computational scheme

The generation of a coupled Martian ODE set requires many numerical calculations, and the various stages of the development of the low-dimensional model can be summarised as:

- The calculation of a set of POD modes from the SGCM streamfunction;
- Calculation of the various derivative fields;
- Implementation of the Galerkin projection scheme;
- Computation of the numerous interaction coefficients;
- Construction of the linear and nonlinear coupled matrices;
- Derivation of the nonlinear coupled ODE set;
- Conversion of the coefficients of the ODE sets to the relevant syntax;
- Integration of the ODE sets;
- Implementation of several numerical tools to analyse the data.

An overview of the general computational scheme of generating a low-dimensional Martian model is shown in Figure 4.2.

It was necessary to generate the models in the most efficient way since the eigenspectra profiles of the SGCM data in Figure 3.9 revealed that models between 6 to 20-D were likely to capture a sufficient amount of the system's energy and thus create models with over 400 terms (if the truncation level is n then the model may produce n linear and n^2 nonlinear terms). The software was designed in a modular fashion so as to allow the implementation of alternative methods of numerical calculation at each stage. Several tools were used to analyse the coupled ODEs, such as Mathematica, so as to calculate the linear stability profiles for each of the wave mode pairs; C integration packages to integrate the nonlinear ODE sets and AUTO (Doedel (1981)), to perform bifurcation analyses and to identify any stable periodic solutions which existed and where they occurred in parameter space.

Each of these applications required the ODEs to be specified in different formats and so it was felt that an automatic system for the conversion of equations between these formats should be developed. A parsing program was written so as to read the Mathematica output (i.e. the generated ODEs) and write a file of equation coefficients as is illustrated in Figure 4.3.

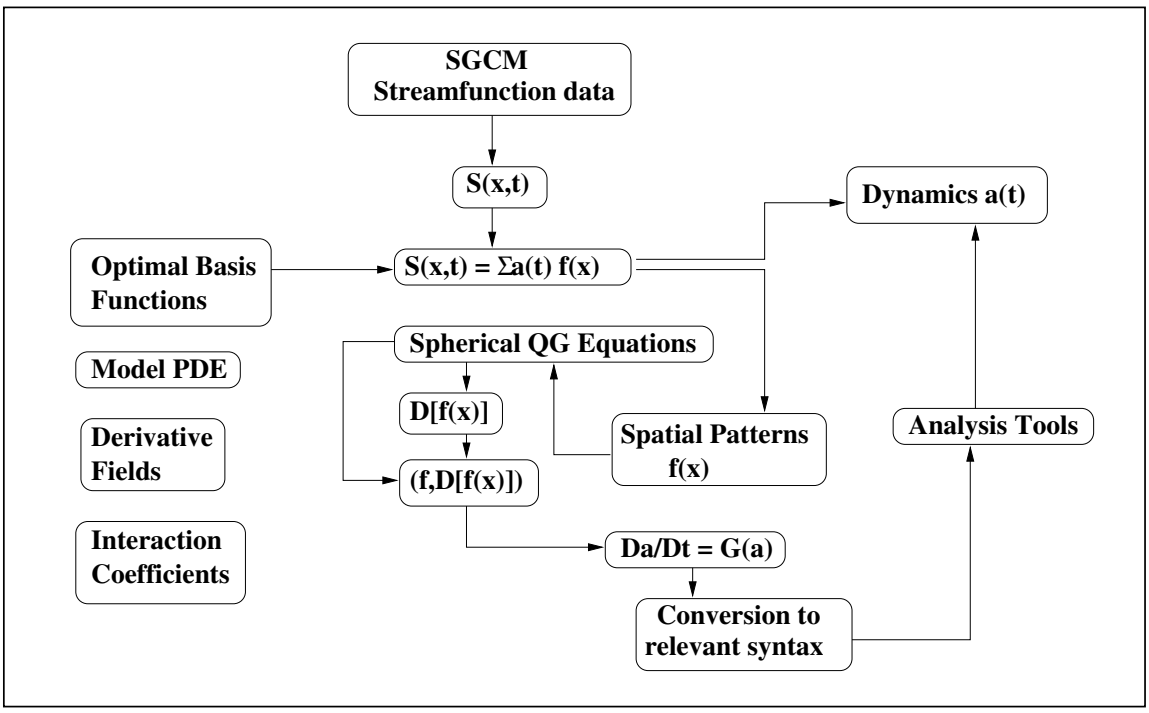


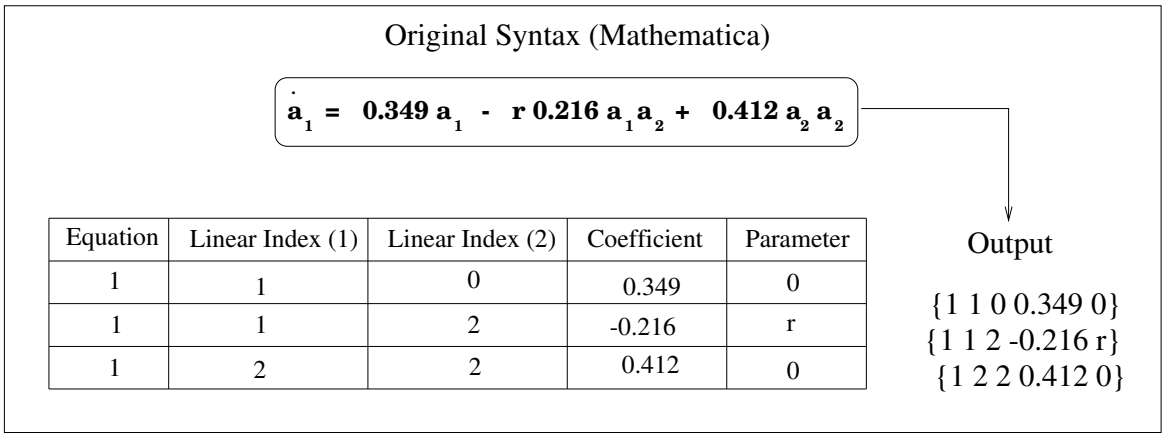
Figure 4.2: An overview of the general computational scheme of generating a low-dimensional POD Martian model. The Figure illustrates the many stages which are required to produce a dynamical system of the form $\dot{\mathbf{a}} = L\mathbf{a} + N\mathbf{a}$ (where L is the linear and N the nonlinear operators), which has hopefully retained some important information regarding the physics of the original SGCM system. This Chapter outlines the Galerkin projection of the spherical QG equations (§ 4.2.6); the computation of the derivatives (§ 4.3.3), of whether the calculation was made in spectral or Fourier space; the computation of the interaction coefficients (§ 4.3.4), i.e. the projection of each POD mode onto the appropriate derivatives. It was then necessary to develop a parsing program which read in the ODEs in Mathematica form and extracted the coefficients of each of the equations, labelling whether the term was linear or quadratic (§ 4.3.6). Once the ODE information was in parsed form, numerical packages were implemented to analyse the Martian ODE sets.

4.3.1 Mak's 1985 low-dimensional model

The two-layer QG model of baroclinic instability (with a channel geometry) which had been analytically formulated using a spectral model by Mak (1985) was checked by using a numerical spectral expansion. The motivation for such a procedure was to develop confidence in dealing with spectral nonlinear partial differential equation models and the implementation of the Galerkin method before it was used on the spherical QG equations using a spectral POD mode expansion.

Within the Mathematica code the spatial basis functions for the channel were defined; the orthonormal properties were verified and the barotropic and baroclinic flows expressed. The Galerkin projection was then performed and the ODEs tabulated and cross-checked against the results found by Mak (1985) (with reference also to Mak (1987)).

In the analytical spectral model the Ekman layers used (at the upper and lower levels) were



Equation	Linear Index (1)	Linear Index (2)	Coefficient	Parameter
1	1	0	0.349	0
1	1	2	-0.216	r
1	2	2	0.412	0

Figure 4.3: A parsing program was written to transform the numerically generated ODEs (created by Mathematica) to a file of equation coefficients so that the ODE sets could be read by other programming languages and numerical packages.

different and the flow was resolved into its barotropic (ψ) and baroclinic (τ) components, where

$$\psi = \frac{1}{2}(\psi_1 + \psi_2), \quad \tau = \frac{1}{2}(\psi_1 - \psi_2) + Uy. \tag{4.37}$$

Here the ψ_1, ψ_2 are the nondimensionalised streamfunctions in the upper and lower levels respectively and U is the forcing in the model. The appropriate boundary conditions are

$$\begin{aligned} \psi(0, y, t) &= \psi(\frac{2\pi}{\gamma}, y, t), \\ \frac{\partial \psi}{\partial x} = 0, \quad & \frac{\partial}{\partial t}\int_0^{\frac{2\pi}{\gamma}} \frac{\partial \psi}{\partial y} dx = 0, \end{aligned} \tag{4.38}$$

at $y = 0, \pi$. The same conditions are imposed on the baroclinic (τ) component.

Solutions for the non-dimensionalised governing potential vorticity equations were solved analytically by using a spectral expansion with basis functions defined as

$$S_{m,n} = \frac{\sqrt{\gamma}}{\pi} e^{im\gamma x} \sin(ny), \quad C_{0,n} = \frac{\sqrt{\gamma}}{\pi} \cos(ny), \tag{4.39}$$

where $m, n \in \mathbb{Z}/\{0\}$ and $\gamma = 1$ (as the study concentrated on synoptic-scale waves). The advantage of using the relations given in (4.39) is that those functions are the eigenfunctions of the Laplace operator satisfying (4.38) on the domain, where the eigenvalues satisfy the relation $\lambda_{m,n} = (m\gamma)^2 + n^2$.

The solutions for the barotropic and baroclinic components of the flow are written by Mak (1985) as

$$\begin{aligned} \psi &= \sum_{m=-M}^{M} \sum_{n=1}^{N} \psi_{m,n} S_{m,n} + \sum_{l=1}^{J} \psi_{0,l} C_{0,l}, \\ \tau &= \sum_{m=-M}^{M} \sum_{n=1}^{N} \tau_{m,n} S_{m,n} + \sum_{l=1}^{J} \tau_{0,l} C_{0,l}, \end{aligned} \tag{4.40}$$

where J is the zonal and N the wave component. It immediately follows that $\psi_{-m,n} = \psi^*_{m,n}$ and $\tau_{-m,n} = \tau^*_{m,n}$ (where * denotes complex-conjugation).

If (4.40) is substituted into the potential vorticity equation (expressed in the form of the barotropic and baroclinic components), a set of ODEs for the time evolution of the spectral coefficients $\psi_{m,n}, \psi_{0,l}, \tau_{m,n}$ and $\tau_{0,l}$ are found. Mak chose three waves $(i,j) = (\pm 1, 1), (\pm 1, 2), (\pm 2, 1)$ and one zonal component $(0, l) = (0, 1)$ for his truncated system, resulting in a set of 7 ODEs for the evolution of the spectral coefficients. By using the same expansion as in (4.40) and coding the expansions so that the truncations for the analytical and numerical models were identical, the 7-D ODE set found by Mak was reproduced, where each of the coefficients of the linear and nonlinear terms were compared and found to agree (for more details see the Appendix entitled 'Derivation of the Spectral Equations' in sections (A5) and (A6) in Mak (1985)).

4.3.2 Fitting the POD modes to spherical harmonics

The first task was to fit the streamfunction POD modes (for the SGCM case) to a chosen set of basis functions. This was achieved by using a spectral expansion technique, so that each POD mode is fitted to a series of spherical harmonics and Legendre polynomials in order that the spatial eigenmodes lay on the same grid as that of the original data (in T21 form).

Expanding the spatial eigenmodes $\Phi_i(\theta, \phi, p)$ as a jagged T21 spectral truncation, we have

$$\Phi_i(\theta, \phi, p) = \sum_{m=0}^{20} \sum_{n=m}^{N_m} V_{mni} P_{m,n}(\mu) e^{im\phi}, \tag{4.41}$$

where V_{mni} are the spectral coefficients for the i-th POD mode, $P_{m,n}(\mu)$ are the associated Legendre polynomials of the first kind (for further details see § 3.3.3), and the summation N_m is defined as:

$$N_m = \begin{cases} 21 & m \quad \text{even} \\ 20 & m \quad \text{odd.} \end{cases} \tag{4.42}$$

The PODs were fitted as in (4.41) by using the numerical routine RGTOSP which was written by Dr S. R. Lewis from AOPP (priv. comm. (1998)) who used the method of Machenhauer and Daley (1972) to determine the spherical harmonic coefficients of a field initially given on an evenly spaced latitude-longitude grid.

Accuracy of the representation of smooth fields on a sphere

A fundamental constraint on the accuracy of the low dimensional model is the accuracy with which the finite spectral representation can represent the features in the flow field. According to Naughton et al. (1996), the structures which can appear in the atmosphere such as Potential Vorticity features in the middle latitudes are often localised. It has been shown by Hortal (1991), Hortal et al. (1991b) and Simmons (1991) that there is a need to keep extra points in the last several latitudes, near the poles, to avoid the occurrence of a 'noise problem' (this

phenomenon has been explained by Courtier and Naughton (1994)).

Naughton et al. (1996) comment on two types of error that result from representing a smooth function on a sphere expressed in terms of finite spherical harmonics, namely *grid representation* and *aliasing* errors. The grid representation error is due to the spherical harmonic series having a finite truncation of spectral coefficients. The aliasing error is associated with the grid representation error and describes the effect of the coarseness of the grid used (when the wavenumbers of the function may not be able to be correctly resolved, and may be misrepresented or aliased into lower wavenumbers). The spectral truncation error is the error caused by finite truncation of the exact spectral series for the function, and the grid representation error being the error in representing the finite spherical harmonic series on a particular finite grid. In the latter case the grid representation error vanishes provided the grid resolution is sufficient for the spectral coefficients retained (Swarztrauber (1979)).

Naughton et al. (1996) concluded that the triangular truncation had superior accuracy for the representation of smooth fields on a sphere over a range of resolutions.

4.3.3 Calculating the derivative fields

An accurate method to evaluate the derivatives of discretised functions is to approximate the function by some series which can be differentiated term by term and for this reason the derivative fields are often calculated in spectral space.

A method which allows the transformation of a field from spectral to grid space and then an inverse transform to revert back to the spectral domain was developed independently by Eliasen et al. (1970) and Orszag (1970). (For a comprehensive summary and discussion of general spectral methods the reader is referred to Machenhauer (1991)). A first implementation of such a scheme fitted each POD mode to a series of spherical harmonics and Legendre polynomials so that the eigenfunctions lay on the same grid as that of the original data (in T21 form). Since spherical harmonics are eigenfunctions of the Laplacian then

$$\nabla^2 P_{m,n} e^{im\phi} = -n(n+1) P_{m,n} e^{im\phi} \tag{4.43}$$

holds.

Taking ϕ to be the longitudinal and μ ($= \sin\theta$) the latitudinal coordinate, and introducing the derivative operators $\frac{\partial}{\partial\phi}$ and $\frac{\partial}{\partial\mu}$ which are defined as:

$$\{\ \}^{(\phi)} = \frac{\partial}{\partial\phi}\{\ \}, \quad \{\ \}^{(\mu)} = -(1-\mu^2)\frac{\partial}{\partial\mu}\{\ \}, \tag{4.44}$$

then from § 4.2.7 the derivatives to be calculated are:

$$\begin{array}{rclrcl} \psi^{(\phi)}(\theta,\mu) & = & \sum_m \psi_m^{(\phi)}(\mu) e^{im\phi}, & \psi^{(\mu)}(\theta,\mu) & = & \sum_m \psi_m^{(\mu)}(\mu) e^{im\phi}, \\ \zeta^{(\phi)}(\theta,\mu) & = & \sum_m \zeta_m^{(\phi)}(\mu) e^{im\phi}, & \zeta^{(\mu)}(\theta,\mu) & = & \sum_m \zeta_m^{(\mu)}(\mu) e^{im\phi}, \end{array}$$

where

$$\begin{aligned} \psi_m^{(\phi)}(\mu) &= \mathrm{Im}\sum_{mn}\psi_{mn}P_{mn}(\mu), \quad \psi_m^{(\mu)}(\mu) = \sum_n \psi_{mn}H_{mn}(\mu), \\ \zeta_m^{(\phi)}(\mu) &= \mathrm{Im}\sum_n \zeta_{mn}P_{mn}(\mu), \quad \zeta_m^{(\mu)}(\mu) = \sum_n \zeta_{mn}H_{mn}(\mu), \end{aligned}$$

and where Im denotes the imaginary part and ζ_{mn} and H_{mn} are defined as

$$\zeta_{mn} = -\frac{n(n+1)}{a^2}\psi_{mn} \quad \text{and} \quad H_{mn} = -(1-\mu^2)\frac{dP_{mn}}{d\mu}. \tag{4.45}$$

By using the same recurrence relation as used by Machenhauer (1991), an expression for $\frac{dP_{m,n}}{d\theta}$ can be obtained:

$$\begin{aligned} -\cos\theta\frac{dP_{m,n}}{d\theta} &= -(1-\mu^2)\frac{dP_{m,n}}{d\mu} \\ &= nD_{m,n+1}P_{m,n+1} - (n+1)D_{m,n}P_{m,n-1}, \end{aligned} \tag{4.46}$$

where $D_{m,n} = \left(\frac{n^2-m^2}{4n^2-1}\right)^{\frac{1}{2}}$, and the Legendre coefficient for $P_{m,n}$ is defined as:

$$\begin{aligned} \sin\theta P_{m,n} &= \mu P_{m,n} \\ &= D_{m,n+1}P_{m,n+1} + D_{m,n}P_{m,n-1}. \end{aligned} \tag{4.47}$$

Therefore, the derivative

$$\frac{d\psi}{d\theta} = \sum_{m=0}^{M}\sum_{n=0}^{N}\psi_{mn}e^{im\phi}\frac{dP_{m,n}}{d\theta}. \tag{4.48}$$

can now be determined if the recurrence relations in (4.46) and (4.47) are used.

Alternatively, the subroutine LGNDRE, written by Dr. S. R. Lewis (priv. comm. (1998)), from the multi-level spectral model (for details refer to Blackburn (1985)) can be used. This routine calculates values of the Legendre functions $P_{m,n}$ and their derivatives $(1-\mu^2)\frac{dP_{m,n}}{d\mu}$, where both odd and even functions are returned, interspersed, in order of increasing n with increasing m, for the staggered triangular truncation.

So in summary, there are two possible ways to calculate the dynamical fields required for the calculation of the interaction coefficients in § 4.2.7.

4.3.4 Calculating the Interaction coefficients

For a chosen truncation level, a large number of derivative fields have to be computed as shown by the inner products in § 4.2.7. Once these have been evaluated, the projection of each POD mode onto the appropriate derivatives determine the entries of the matrix (i.e. the interaction coefficients) which define the model. The interaction coefficients are given by the inner product of the POD modes with the various derivative fields that arise in the governing equations. The inner product is implemented by applying the trapezoidal rule in both the latitudinal and longitudinal directions.

The interaction coefficients are calculated in two different ways, depending upon which derivative is to be calculated in the projection. Below is a summary of how each of the inner products were calculated and so how each of the interaction coefficients were determined.

METHOD I

- The derivatives $\frac{\partial}{\partial\phi}$, $\frac{\partial^2}{\partial\phi^2}$, ∇^2 are calculated in spectral space.
- The derivatives are transformed from spectral space onto a longitude-latitude grid.
- The inner products of the derivative fields are calculated in grid space.

METHOD II

- The streamfunction field is transformed from spectral to grid space.
- The derivatives $\frac{\partial}{\partial\theta}$, $\frac{\partial^2}{\partial\theta^2}$ are calculated on this grid using the subroutine LGNDRE.
- The inner products of the derivative fields are then calculated in grid space.

The speed of the process is greatly enhanced by the Fast Fourier Transforms (FFTs) and Gaussian Quadrature (for more details see Press et al. (1992) and Arfken (1985)). The advantages inherent in usage of a Fast-Fourier algorithm have been discussed by Orszag (1970).

4.3.5 Test example: verification of computational scheme

It was necessary to verify the accuracy of the computational scheme which has been outlined in Figure 4.2. The numerical routines which defined the derivative fields and the Galerkin projection (Fortran code) as well as the Mathematica code which constructed the matrices and which generated the ODE sets were all verified by using a simple test function for the linearised problem, whose derivatives and interaction coefficients could be solved analytically.

A method had to be decided upon which allowed the vertical form of the spatial modes to be analytically feasible. The vertical functions $H_1(p)$ and $H_2(p)$ were chosen to be

$$H_1(p) = \frac{1}{\sqrt{(p_n - p_1)}}, \quad H_2(p) = \left(\frac{2}{p_n - p_1}\right) \sin\left(\frac{2\pi(p - p_1)}{p_n - p_1}\right), \tag{4.49}$$

where H_1 is the barotropic and H_2 the first baroclinic mode. Here the normal modes satisfy the vertical structure equation (2.5) with the condition that the stratification parameter $S(\sigma)$ is constant, as well as satisfying the necessary vertical integral conditions

$$\int_{p_1}^{p_n} H_2(p)dp = 0 \quad \text{and} \quad \int_{p_1}^{p_n} H_i^2(p)dp = 1, \tag{4.50}$$

for $i = \{1, 2\}$. The latitudinal zonal flow was taken to be $\sin 2\theta$ (a zonal jet in the northern hemisphere), the basic state $-U\theta$ (where U is the magnitude of the velocity of the zonal flow)

and the longitudinal wave component to be $\sin n\phi$ so that

$$\begin{aligned}
\tilde{\Phi}_1 &= (H_1 \sin 2\theta \sin n\phi, H_2 \sin 2\theta \sin n\phi)\,, \\
\tilde{\Phi}_2 &= (H_1 \sin 2\theta \cos n\phi, H_2 \sin 2\theta \cos n\phi)\,, \\
\hat{\Phi} &= -\tfrac{U}{2}\,(H_1\theta, H_2\theta)\,, \\
f &= 2\sin\theta,
\end{aligned}$$

where $\tilde{\Phi}_1$ and $\tilde{\Phi}_2$ are the wave modes, $\hat{\Phi}$ is the basic state and f is the Coriolis parameter, also noting that the wave modes satisfy the condition $\tilde{\Phi}_i \cdot \tilde{\Phi}_j = \delta_{ij}$.

Matrix $\mathcal{C}$ in (4.26) couples the time derivative of the state vector and it is necessary to compute the determinant of this matrix since the Jacobian matrix ($\mathbf{J}$) of the linearised problem is

$$\mathbf{J} = \frac{1}{\det(\mathcal{C})}\,\mathcal{C}^{-1}\mathcal{L}, \tag{4.51}$$

where $\mathcal{L}$ is the matrix of the linear system. From (4.26) the determinant of matrix $\mathcal{C}$ is of the form:

$$\det = \left(\tilde{\tilde{\beta}}_{11} - \tilde{\tilde{\alpha}}^{\lambda}_{11}\right)\left(\tilde{\tilde{\beta}}_{22} - \tilde{\tilde{\alpha}}^{\lambda}_{22}\right) - \left(\tilde{\tilde{\beta}}_{12} - \tilde{\tilde{\alpha}}^{\lambda}_{12}\right)\left(\tilde{\tilde{\beta}}_{21} - \tilde{\tilde{\alpha}}^{\lambda}_{21}\right), \tag{4.52}$$

and so the Jacobian matrix of the linearised problem of (4.51) can be written as

$$\mathbf{J} = \begin{pmatrix} \tilde{\tilde{\beta}}_{22} - \tilde{\tilde{\alpha}}^{\lambda}_{22} & \tilde{\tilde{\alpha}}^{\lambda}_{12} - \tilde{\tilde{\beta}}_{12} \\ \tilde{\tilde{\alpha}}^{\lambda}_{21} - \tilde{\tilde{\beta}}_{21} & \tilde{\tilde{\beta}}_{11} - \tilde{\tilde{\alpha}}^{\lambda}_{11} \end{pmatrix} \cdot \left(\frac{1}{\det(\mathcal{C})}\right) \cdot \begin{pmatrix} L_1 & L_2 \\ L_3 & L_4 \end{pmatrix}, \tag{4.53}$$

for $i = \{1, ..4\}$ and where the linear entries L_i are defined as:

$$\begin{aligned}
L_1 &= -\tilde{\tilde{\eta}}_{11} - r\tilde{\tilde{\beta}}_{11} - \hat{\tilde{\tilde{\zeta}}}_{111} + \hat{\tilde{\tilde{\delta}}}^{\lambda}_{111} - \tilde{\tilde{\zeta}}_{111} + \tilde{\tilde{\delta}}^{\lambda}_{111}, \\
L_2 &= -\tilde{\tilde{\eta}}_{12} - r\tilde{\tilde{\beta}}_{12} - \hat{\tilde{\tilde{\zeta}}}_{121} + \hat{\tilde{\tilde{\delta}}}^{\lambda}_{121} - \tilde{\tilde{\zeta}}_{112} + \tilde{\tilde{\delta}}^{\lambda}_{112}, \\
L_3 &= -\tilde{\tilde{\eta}}_{21} - r\tilde{\tilde{\beta}}_{21} - \hat{\tilde{\tilde{\zeta}}}_{221} + \hat{\tilde{\tilde{\delta}}}^{\lambda}_{211} - \tilde{\tilde{\zeta}}_{211} + \tilde{\tilde{\delta}}^{\lambda}_{211}, \\
L_4 &= -\tilde{\tilde{\eta}}_{22} - r\tilde{\tilde{\beta}}_{22} - \hat{\tilde{\tilde{\zeta}}}_{221} + \hat{\tilde{\tilde{\delta}}}^{\lambda}_{221} - \tilde{\tilde{\zeta}}_{212} + \tilde{\tilde{\delta}}^{\lambda}_{212}.
\end{aligned} \tag{4.54}$$

The non zero entries of the linearised Jacobian matrix in (4.53) were calculated analytically and found to be:

$$\begin{aligned}
\tilde{\tilde{\alpha}}_{11} &= \pi^2 = \tilde{\tilde{\alpha}}_{22}, \\
\tilde{\tilde{\alpha}}^{\lambda}_{11} &= \frac{\pi^2}{2} = \tilde{\tilde{\alpha}}^{\lambda}_{22}, \\
\tilde{\tilde{\beta}}_{11} &= -\pi^2\left[4 - n^2\right] = -\tilde{\beta}_{22}, \\
\tilde{\tilde{\beta}}^{\lambda}_{11} &= -\frac{\pi^2}{2}\left[4 + n^2\right] = -\tilde{\beta}^{\lambda}_{22}, \\
\hat{\tilde{\tilde{\delta}}}_{121} &= \frac{nu\pi^2}{4} = -\hat{\tilde{\tilde{\delta}}}_{112} = -\hat{\delta}_{211} = \hat{\tilde{\tilde{\delta}}}_{211}, \\
\tilde{\tilde{\zeta}}_{112} &= \frac{nu\pi^2}{4}\left[4 + n^2\right] = -\tilde{\tilde{\zeta}}_{211},
\end{aligned} \tag{4.55}$$

where the remaining terms in (4.53) were identically zero because of the orthogonality properties of the trigonometric functions.

Letting $\upsilon = 4 - n^2$ and $\omega = \frac{nU\pi^2}{4}(4 + n^2)$ for simplicity, then as

$$\begin{pmatrix} \dot{b}_1 \\ \dot{b}_2 \end{pmatrix} = \mathbf{J} \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \tag{4.56}$$

the coupled ODEs for the linearised problem can be written as

$$\begin{pmatrix} \dot{b}_1 \\ \dot{b}_2 \end{pmatrix} = \frac{-1}{(\upsilon + \frac{1}{2})(\upsilon - \frac{1}{2})} \begin{pmatrix} (\upsilon - \frac{1}{2})(r\upsilon b_1 - \omega b_2) \\ -(\upsilon + \frac{1}{2})(\omega b_1 + r\upsilon b_2) \end{pmatrix}. \tag{4.57}$$

If the magnitude of the velocity of the basic state is taken to be $U = 1$ and the longitudinal wavenumber of the spatial mode is taken to be $n = 2$, i.e. a wavenumber 2 profile, it was found that the true coefficients of the ODE set (given in (4.57)) only differed from the numerically generated coefficients by between 0.5 - 0.9% i.e. the computational scheme had less than 1% error. This suggested that the general scheme for processing the data which included the stages of (1) computing the derivatives and the interaction coefficients, (2) the translation of the coefficients into several formats, (3) the manipulation of these coefficients and (4) the construction of numerous matrices as well as the generation of the ODE set, was accurate to at least 99%.

4.3.6 Analysing the Martian ODE set

The main restriction with our modelling method is that the Froude number is fixed for any ODE set. The reason for this is that the time derivatives on the LHS in (4.25) explicitly depend upon the Froude number, and since the generation of the ODE set requires the numerical inversion of the coupling matrix $\mathcal{C}$ then any change in Froude number means that an additional set of equations would have to be computed. It is for this reason why the vertical structure equation discussed in Chapter 2 was so vital to our analysis, as the eigenvalues of the calculation automatically calibrate the Froude number in the equations and these values can be used in matrix $\mathcal{C}$, otherwise a time consuming parameter space search would have been required to determine the appropriate Froude number for the system.

The ODE sets were analysed by a number of numerical packages. Mathematica was used to construct the matrix equation in (4.25); a Runge-Kutta fourth order integrator scheme was obtained from Numerical Recipes (see Press et. al (1992)) and the numerical bifurcation tool AUTO (see Doedel (1981)) was primarily used to identify any periodic orbits and their stability properties. Additionally, numerical integration packages such as XPPAUT (Ermentout (1998)) and Xigra (Jäckel (1995)) were also used, each of these requiring the files to be written in C.

4.4 Conclusions

In order to solve the equations that govern the evolution of the Martian atmospheric flow, a basis in phase space must be chosen. Often a gridpoint description or a spectral basis consist-

ing of spherical harmonics would be used, even though it is recognised that EOFs are more efficient in describing large-scale atmospheric dynamics (Selten (1997)). According to Selten (1993), a basis of spherical harmonics is even capable of representing circulation patterns that never occur, i.e. they are redundant and inefficient.

In this Chapter a method for generating a low-dimensional model using the spherical QG equations has been outlined. Particular detail has focused on the computation of the derivative fields and the interaction coefficients. It has been shown that since the spherical harmonics have the advantage of being the eigenfunctions of the Laplacian then the longitudinal and latitudinal derivatives also have simple forms. Each POD mode is fitted to a spherical-Legendre series in order to compute the derivatives of the various fields. The code which generates the ODE sets has been verified by the use of a simple test function. The linearised problem was shown to be at least 99 % in agreement with the analytically calculated values.

It has been shown that the vertical derivative term in the spherical QG equations (in (4.7)) can be represented by a POD mode expansion. Not only does the vertical structure equation in (4.15) produce eigenmodes which seek to minimise APE in the vertical direction (as discussed in Chapter 2), and thus captures the buoyancy driven processes (since these instabilities are governed by a conversion of APE into KE), but also has provided us with a useful identity, making it unnecessary to explicitly calculate the vertical derivative term in (4.7).

In Chapter 3 the investigation of the time series of the SGCM streamfunction data revealed that the evolution of the atmospheric flow could quite possibly be reduced to a low-dimensional system and hence the evolution of the Martian atmospheric circulation could lie on a relatively low-dimensional attractor. As POD patterns effectively describe the attractor of low-dimensional systems (see Preisendorfer (1988)) these spatial patterns shall be used. Reducing the system to its essential degrees of freedom could considerably increase the transparency of the SGCM and provide useful insight into the underlying processes, as well as interactions between dominant components which contribute to atmospheric variability on various timescales.

The intention now is to generate low-dimensional models of the Martian atmospheric circulation, which accurately reproduce the dynamics observed in the original SGCM. The aim is to investigate the qualitative behaviour of low-order ODE sets using dynamical systems theory. The steady solutions of the governing equations will be of special importance as the stable states are those which will be observed in a physical system. Bifurcation theory will be used to monitor the solutions of the nonlinear systems and determine how their qualitative characters change as the dissipation parameter is varied. The occurrence of periodic orbits will be identified and their stability properties will be closely analysed.

Chapter 5

2-v SGCM Galerkin models

Since the SGCM data depicts a steady wavenumber 3 pattern and given the dominant energy content which exists in the few leading order modes, then it is conceivable that the SGCM data could be consistent with a low-dimensional set of ODEs. The aim is then to try to generate the time-series of the SGCM's principal components (PCs) using such a minimal equation set and to determine whether the dynamics of the SGCM is actually low dimensional. If the system is in fact low dimensional then a thorough examination of the equation set may provide us with an understanding of some of the underlying dynamics of the SGCM data set. In determining the accuracy of the Galerkin model it is easy to compare the size of the wave and mean-flow correction (MFC) amplitudes and also to contrast the frequencies of each of the modes with the original system. By analysing the set of ODEs it might be possible to identify the important energy exchanges which occur between the modes, besides specifying which modes are influential in determining a particular flow regime.

In Chapter 3 it was shown that, given an estimate of the buoyancy frequency profile, the solution of the vertical structure equation (2.5) gives a series of eigenvectors which can be used to decompose the flow. The eigenvalues of the calculation automatically calibrate the Froude number in the equations, and leaves dissipation as the only free parameter in the modelled equations. It is realised that fixing the Froude number is a significant limitation of the model. It would be desirable if the Froude number could be left as a free parameter but in the present context this is impractical. The explicit form of the nonlinear ODEs depends on inverting the coupling matrix, which is itself actually dependent on the Froude number. This would mean that every time the Froude number was varied, the inverse of the coupling matrix would have to be recalculated.

Notwithstanding this limitation, in this Chapter the robustness and accuracy of the numerical implementation of the Galerkin projection will be tested. By selecting the barotropic and first baroclinic mode[1] (the '2-v' scheme) to model the vertical activity of the original system and fixing the Froude number[2] to be F =8.14 (the eigenvalue associated with the first baroclinic

[1] We will denote '2-v' and '4-v' to respectively denote 2 or 4 vertical modes retained in the modelling scheme.

[2] James (1994) states that $\frac{Rd}{a} \sim 0.33$ for Mars (where Rd is the Rossby radius of deformation), which suggests that the scale of the baroclinic waves is large relative to the planetary radius. If the global Froude number is taken to be 8.14 then according to Table 2.1 this corresponds to a Rossby radius of deformation of 1.896×10^3 km, and thus a ratio of $\frac{Rd}{a} \sim 0.35$. This indeed suggests that $F = 8.14$ is a reasonably accurate estimation of the system's Froude number.

mode) various Galerkin models will be discussed and contrasted. Only two vertical modes have initially been retained as it was thought that such a study would be analogous to a two-level model, and thus providing an interesting comparison with the models in Chapter 6 (which contain an additional two baroclinic modes in the vertical direction). By making these comparisons it can be determined what effect the improved vertical resolution scheme has had upon the behaviour of the generated Galerkin modes, and thus how vital it is to retain as much information as is possible about the vertical activity of the original model. The way in which the different norms affect the solutions of the generated low-dimensional models is discussed in detail, together with a discussion of the influence of different truncation levels.

5.1 Background to baroclinic wave studies

Many attempts have been made to model and study the dynamics of fully developed baroclinic waves, ranging from laboratory studies to extensive theoretical work which has focused on linear and weakly nonlinear mathematical models, often based on QG theory (see Pedlosky (1979)).

Baroclinic instability often arises in rotating stratified fluids with vertically sheared basic currents. Such instabilities play a role in the dynamics of atmospheres and oceans, and is the fundamental instability mechanism in large-scale midlatitude atmospheric and oceanic circulations, and their nonlinear evolution has been the subject of much study. Under appropriate conditions of lateral heating and rotation, the interior fluid density and pressure surfaces are orientated such that the potential energy in the density field can be converted into kinetic energy. Hide (1953, 1956) studied the role of thermal convection by using a differentially heated rotating annulus, and showed that regular baroclinic waves occurred under a wide range of experimental conditions (for reviews of recent work see Hide and Mason (1975)).

Work over the years has concentrated on Couette flow between concentric cylinders; Bénard convection in a horizontal layer heated from below and shear flow, particularly the boundary layer and the free shear layer. The annulus experiments are comparable to studies of Couette flow and Bénard convection. The flow in the annulus experiments exhibit a variety of distinct regimes, and depend upon the imposed external conditions. There are two main types of vacillation which have been seen in studies of rotating annulus convection, namely 'amplitude vacillation' (AV) and 'structural vacillation' (SV). In AV there are periodic variations in the amplitude of the waves where in SV[3] the spatial form of the waves change periodically or irregularly. Transitions may occur between states, and this is possible by slowly varying the external control parameters (such as the imposed temperature contrast $\triangle T$ and the background rotation rate Ω). This is of immediate interest to our studies since many authors have suggested (e.g. Barnes (1986) and Tillman et al. (1979)) that Martian baroclinic waves may be similar to the steady waves observed in the annulus.

[3]SV appears to indicate that the fully irregular flow is developing in the system. Work by Buzyna et al. (1984) has shown that in such cases the dominant wavenumber and the flow structure fluctuate in a form of 'geostrophic turbulence'.

In an attempt to understand the various regimes exhibited in the rotating annulus (e.g. see Hignett et al. (1985)), much work has concentrated on linear and weakly nonlinear mathematical models. Linear theory has shown that baroclinic waves will grow via an instability of the axisymmetric flow, having wave structures which are often similar to those which are observed in the annulus experiments. Two highly idealised models which have been analysed are the Eady model and the Phillips model (see Pedlosky (1979)). Eady (1949) investigated the stability of an inviscid fluid on an f-plane[4] for a uniform zonal flow with constant vertical shear, whereas Phillips (1954) assumed two layers of a homogeneous fluid. The Phillips model has been extended by the inclusion of dissipation and the β-effect, higher-order terms, or the use of different basic states. Such analysis has provided important information regarding the various conditions under which the axisymmetric regime will occur.

Weakly nonlinear theory, however, focuses upon supercritical conditions which are close to the onset of instability, where only one or two wave components may be linearly unstable. Here the equations are expanded in powers of supercriticality about the marginally stable solution for a small (but finite) amplitude wave yielding a set of equations which describe the amplitude of the wave and its interactions with the mean flow. Hart (1981) studied a weakly nonlinear system in the presence of Ekman dissipation and found that the equilibrated wave had a longer wavelength than the linearly most unstable mode. He found that the longer waves extracted energy more slowly and dominated at finite amplitude (as compared with the shorter, more unstable waves). These findings have been substantiated by similar results of Mak (1985) and Klein and Pedlosky (1986).

Much work has concentrated on weakly nonlinear and truncated spectral models of baroclinic instability, in an attempt to understand experimental observations, as well as to study the basic bifurcation structures of baroclinic flows. Similar studies have centred on amplitude vacillations, involving relatively few spatial structures (such as the work of Pedlosky (1971)) and the equilibration of unstable waves, either to steady or periodically vacillating states (the reader is referred to Hart (1979) for a thorough review).

Hart (1976) presented experimental observations from the two-layer annulus which showed that the wave disturbances were well explained by a two-layer model in cylindrical geometry, on an f-plane. Pedlosky and Frenzen (1980) used a full numerical two-layer spectral model, and concentrated on the moderate dissipation region of parameter space and observed that Lorenz type chaos could be produced. This was further substantiated by the work of Gibbon and McGuinness (1980), and Brindley and Moroz (1980), who indicated ways in which flows which had simple spatial structures may undergo transitions to aperiodic or chaotic behaviour. Pedlosky (1981) showed that the inclusion of the β-effect smoothed out the wave activity within Lorenzian chaotic structures, and for large values of β the chaotic solutions were destroyed via period halving bifurcations. Gibbon and McGuinness (1982), and Pedlosky and Frenzen (1980) showed that the amplitude equations (under certain conditions) reduced to the Lorenz

[4]On the Earth and Mars, the planetary contribution to the potential vorticity varies with latitude. The qualitative effect of spherical geometry may be included via the β-plane parametrisation which represents the linearised variation of the Coriolis parameter $f = f_0 + \beta y$, where $\beta = \frac{df}{dy}|_{f=f_0}$. Models which ignore the effects of planetary sphericity are referred to as f-plane models, and those which include it are called β-plane models.

equations (1963a), which describe two-dimensional convection in a horizontal layer of fluid heated from below.

Spectrally truncated, QG numerical models are capable of producing flow regimes which are highly comparable with those observed in laboratory experiments (Lorenz (1963b)), such as regular waves, periodic vacillations and irregular flows. Such highly truncated models represent the behaviour of the waves, whose amplitudes are not necessarily small, with simplified physical processes and a very limited number of interactions with other wave components. Modifications of the physics, such as including the β-effect (Pedlosky (1981)), or the inclusion of wave modes (Mak (1985), Klein and Pedlosky (1986), Yoshida and Hart (1986)) severely alter the behaviour and the bifurcations of the systems under study.

Mak (1985) used a highly truncated spectral model of the two-layer β-plane channel model to study the influences of wave-wave and wave-zonal flow interactions on baroclinic instability. It was shown that when no wave-wave interactions were included the wave which governed the flow had a longer zonal wavelength than that of the most unstable wave according to linear theory, and was not the most linearly unstable wave. If wave-wave interactions were included in the model, then in contrast, the general equilibration leads to one of four distinctly different equilibrated states (depending upon the baroclinic forcing and the dissipation parameters): (1) a steady single-wave state, (2) a steady-multiple wave state, (3) a triad-limit-cycle (found to be robust across the parameter space) or (4) a fluctuating state. However, it has been found that severely truncated models may be strongly dependent upon the level of truncation (e.g. Nese, Dutton and Wells (1987); Curry et al. (1984)), particularly if the regimes contain evidence of low-dimensional chaotic motions.

Considerable progress has been made in laboratory studies, particularly in the investigations of transitions to 'baroclinic chaos'[5] in rotating two-layer experiments, where Hart (1985, 1986) and Ohlsen and Hart (1989a, b) have tried to relate their experimental findings to flows exhibited by various low-order models, providing considerable experimental evidence of routes which involve period-doubling cascades and quasi-periodicity[6]. However, other similar studies have concentrated on the spatial and temporal properties of the flow in the well-known irregular regime, such as the work done by Buzyna, Pfeffer and King (1984).

Recently Read et al. (1992) investigated baroclinic flows under moderate forcing conditions and showed that there existed two distinct regimes of flow in the thermally driven rotating annulus. They identified low-dimensional chaos which exited near the transition from one steady wave to a wave of a different zonal wavenumber. It was suggested that nonlinear competition between the adjacent zonal wave numbers was responsible for the transition from a

[5]'Baroclinic chaos' or 'geostrophic turbulence' (Charney (1971), Hart (1985)) is found at high rotation rates where all the temporal and spatial coherence is lost. This regime may be thought of as a characteristic of the Earth's unpredictable weather patterns.

[6]Quasi-periodicity is quite a common route to chaos. A stable fixed point turns into a limit cycle via a Hopf bifurcation. The limit cycle goes through a second Hopf bifurcation into a 2-frequency torus. Under further changes of parameter the torus becomes unstable and may either develop into a 3-torus via another Hopf bifurcation or into a chaotic attractor with a dimension slightly larger than two, which can have some overall structure resembling the torus. In particular the two frequencies of the limit cycle can dominate the chaotic motion.

quasi-periodic amplitude vacillation to a chaotically modulated, amplitude vacillation. It was interesting that a quasi-periodic behaviour (m=3) vacillation (on a 2-torus) was found to be a low-dimensional (D $\sim$ 3) chaotically modulated vacillation at very low frequency (apparently organised by a 3-torus). The flows obtained very much depended upon the initial conditions and the way the experimental conditions were approached, and it was argued that the periodic vacillations and apparent transitions to low-dimensional chaotic behaviour indicated that the initial stages in the development of disordered baroclinic wave motions were a form of 'weak turbulence'.

Much study has concentrated on the development of periodic vacillations leading to chaos in highly truncated spectral models of baroclinic waves. A form of amplitude vacillation in a weakly nonlinear model has been studied by Pedlosky (1970b, 1971), which consisted of a single zonal wave mode interacting with a zonal flow with two-layer vertical discretisation. Pedlosky's findings created a great interest in attempting to identify the various transitions which occur before chaotic behaviour is reached. The work of Gibbon and McGuinness (1980) and Pedlosky and Frenzen (1980) provided significant contributions to the understanding of such processes, as they found period-doubling cascades, a chaotic region and a sequence of period-halving bifurcations as dissipation was varied.

Boville (1982) discussed some mixed-wave aperiodic vacillations in a fully nonlinear two-layer QG model with moderate spectral resolution. The parameter region under consideration was such that many waves were unstable simultaneously. It was found that for parameter values which were consistent with atmospheric flows, vacillating waves were observed, including amplitude, wavenumber and structural vacillations. Low-dimensional models (about 10-D) have been shown to exhibit both amplitude and structural vacillations (such as Weng et al. (1986)), similar to what is observed in the annulus experiments. However, there have been many unsuccessful attempts to use similar low-order models to describe laboratory results (such as Hart (1986)), where it has been shown that at the modest degree of supercriticality used in the experiments, the resolution of the used low-dimensional model is insufficient to capture the actual dynamics. Thus, recent attempts to compare directly with the two-layer experiments have used high-resolution numerical simulations of the governing QG equations.

Collins and James (1995) have used a fully nonlinear GCM with simplified parameterisations of physical processes, namely Newtonian cooling and Rayleigh friction to examine the regular nature of baroclinic waves in the context of Mars. They found that for certain parameter settings, corresponding to those representative of the Martian atmosphere, large amplitude regular waves occurred. These regular waves contrast with the irregular waves seen in similar terrestrial studies (James and James (1992)), and seen in the Earth's troposphere. Collins and James concluded that the short radiative time scale in the Martian atmosphere, and hence large heating and cooling rates, resulted in a complicated nonlinear balance between advection and diabatic heating, thus maintaining the large amplitude waves. It is the radiative time scale therefore which seems to be at least one of the crucial parameters that leads to Martian baroclinic waves being regular and predictable.

Motivation

The SGCM data which is used in this thesis describes a wave 3 regime and is similar to single wavenumber flows which are observed in the differentially heated annulus experiments (see Hide and Mason (1975)). It was important to choose a system which can be reduced to a low-order system, but was still capable of retaining the original system's qualitative behaviour. It was shown by Collins and James (1995) that most of the SGCM's variance (or energy) could be explained by a couple of modes (the two leading EOFs were found to explain 75% of the variance of the system, with no other EOF explaining more than 3%). They also found that the corresponding principal components of the leading two EOFs formed a limit cycle in PC space, and suggested that this could be thought of as representative of the system's attractor. These findings all imply that a low-dimensional ODE model of the SGCM is viable. Furthermore, the work done by Collins et al. (1996) demonstrated that for the simulated Viking Lander 2 pressure field the dominant four modes captured either 50 or 60% of the variance of the system, depending upon whether the diurnal cycle is included or not. There is then evidence that a low-dimensional model for the MGCM is also attainable.

Read et al. (1992) provided us with sufficient confidence that a low-dimensional model of the Martian model could be developed. It was stated for the annulus waves that *"The essence of the dynamics, including typical sequences of flow transitions, should therefore be capable, at least in principle, of being captured in a 'low-order' model with a small number of degrees of freedom."* They felt that there were certain indicators in their findings which pointed to the feasibility of low-dimensional modelling of the annulus. This has motivated us in attempting to describe the wave 3 regime obtained from the SGCM by using a low-dimensional model. It is hoped that it will be possible to identify *'sequences of flow transitions'* which occur in the original data set. Optimistically, the models will be capable of reproducing some of the amplitude and structural vacillations which occur in the original data besides selecting the dynamically most important modes.

5.2 The advantages of using the total energy norm

The choice of the norm used in POD modelling is a major issue. The standard correlation norm and two different energy norms (KE and TE) will be used in calculating the POD modes from the SGCM data. These spatial eigenfunctions will then be substituted into the spherical QG equations, which will be used to create the set of ODEs via a Galerkin procedure.

The issue of the choice of a norm was motivated by the work of Selten (1993) which involved a two-level QG model of the Earth's atmosphere. The model was formulated in spherical harmonics and truncated at T5, where EOFs were used to describe the evolution of the circulation. It was shown that the global structure can only be described by using the TE norm. Selten argued that the use of the standard correlation and the KE norms in low order models both failed to reproduce the global attractor of the original model because they were incapable of simulating properly the energy conversion processes. Selten also stressed the importance of using sufficient vertical resolution to allow the modelling of energy conversions via the baroclinic instability mechanism.

This is an important result to note since the SGCM (and MGCM) data sets contain many modal interactions and so potentially many opportunities of energy exchange. Baroclinic waves depend upon APE which can be released in the form of a disturbance and thus allowing the possibility of its conversion to KE, so using a TE norm (TE=KE+APE) seems an obvious choice. This illustrates the importance of using POD modes which take account of these energy types if a highly truncated low-dimensional model is to be generated.

If the KE norm is used, the baroclinic patterns are usually poorly represented. This results in interactions which involve the temperature field to be poorly represented. The implications of a poor extraction of the baroclinic structures are fairly widespread, particularly if the essential energy conversion processes of the model are insufficiently depicted, as these strongly influence the dynamics of the model. This was illustrated in the cases considered by Selten (1993) where comparisons were made between models produced using 23 KE and TE eigenvector patterns, which were called KE[23] and TE[23] respectively. The amount of KE in the KE[23] model was comparable to that of the complete model, but the PE was considerably lower than that of the original system. Unlike the complete model, large fluctuations in the amount of TE over time were exhibited in the KE[23] model. This meant that Selten's KE model inaccurately represented the TE in the model, even though the KE was well accounted for. Clearly, the problem with the KE norm was in the poor representation of the APE supply.

Selten (1993) showed that in his truncated TE[23] model the energy cycle was much better described. The reason for this was argued to be due to the TE norm giving sufficient weight to the temperature field. It was highlighted that there existed a reasonable qualitative agreement between the TE[23] and the complete model.

In this thesis the TE norm will also be identified as the ideal norm for modelling baroclinic instability. The reason for this is that it places a heavy emphasis on the temperature field. As we are aware, baroclinic instability depends upon the existence of the vertical shear of the basic current which is associated with a horizontal temperature gradient. The presence of the horizontal temperature and vertical gradients imply the existence of APE in the basic flow, which is the energy source for baroclinic instability. If this temperature gradient is poorly extracted by use of the norm, then it is not surprising that a Galerkin model (generated by the spatial eigenmodes) produces poor results as it will have insufficient knowledge of the energy conversion processes which have taken place in the original model. It is highly probable that in such models both the KE and APE terms are important to the overall dynamics of the system.

Some of Selten's models displayed intermittent behaviour. Even though these were infrequent, there were occasions where the flow underwent a transition to another regime, with very different predictability properties. He found that the main transition was an example of the flow changing from a chaotic to a more regular regime. By producing phase space reconstructions it was shown that the truncated TE model reproduced both parts of the attractor of the complete model. The only obvious deficiency was that the transitions seen in the complete model were not as dominant in the TE[23] model. It was concluded that the only way to describe

the global structure of the attractor with a truncated model was by using the TE norm.

It is for the reasons discussed that a TE norm will be used in this thesis and comparisons shall be made with results obtained from models using the standard correlation and KE norms.

5.3 Linear Stability Method

Suppose the POD-Galerkin method generates a system of ODEs of the form:

$$\dot{\boldsymbol{a}} = \boldsymbol{\mathcal{F}}(\boldsymbol{a}), \tag{5.1}$$

then the trivial fixed point solution $\boldsymbol{a}_0 = \mathbf{0}$, corresponds to the basic state. If the equations are linearised about $\boldsymbol{a}_0$ then the stability of the equilibrium state can be determined. Defining the perturbation about the fixed point to be $\boldsymbol{x}$, and taking a Taylor expansion about $\boldsymbol{a}_0$, then

$$\dot{\boldsymbol{a}} = \boldsymbol{\mathcal{F}}(\boldsymbol{a}_0) + \boldsymbol{x}\mathcal{D}\boldsymbol{\mathcal{F}}(\boldsymbol{a}_0) + \mathcal{O}(\|\boldsymbol{x}\|^2). \tag{5.2}$$

Therefore, in a neighbourhood of $\boldsymbol{a}_0$, the dynamics are governed by the linear system

$$\dot{\boldsymbol{x}} = \mathcal{D}\boldsymbol{\mathcal{F}}(\boldsymbol{a}_0)\boldsymbol{x}, \tag{5.3}$$

which has the solution

$$\boldsymbol{x}(t) = e^{\lambda t}\boldsymbol{x}_0, \tag{5.4}$$

where λ are the eigenvalues of the Jacobian matrix, and

$$\mathcal{D}\boldsymbol{\mathcal{F}}(\boldsymbol{a}_0) = \begin{bmatrix} \frac{\partial F_1}{\partial a_1} & \cdots & \frac{\partial F_1}{\partial a_N} \\ \cdots & & \cdots \\ \frac{\partial F_N}{\partial a_1} & \cdots & \frac{\partial F_N}{\partial a_N} \end{bmatrix}. \tag{5.5}$$

This solution will be asymptotically stable if the real parts of the eigenvalues are negative. For the trivial equilibrium solution, $\boldsymbol{a}_0 = \mathbf{0}$, the only contributions to the Jacobian matrix come from the linear problem:

$$\mathcal{C}\dot{\boldsymbol{a}} = \mathcal{L}\boldsymbol{a}, \tag{5.6}$$

where $\mathcal{C}$ is the coupling matrix of the linear problem which depends explicitly upon the Froude number (F) and $\mathcal{L}$ is the matrix for the linear problem (refer to § 4.2.8).

It is now necessary to formulate the Jacobian matrix $\mathcal{J}$ of the linear system, where

$$\dot{\boldsymbol{a}} = \mathcal{C}^{-1}\mathcal{L}\boldsymbol{a} = \mathcal{J}\boldsymbol{a}. \tag{5.7}$$

Matrix $\mathcal{J}$ was computed explicitly using Mathematica. The expression is usually complex, since the inversion of the coupling matrix produces a product of quadratic expressions in the Froude number.

The aim is to compute the locus of critical Froude number points where the real part of

any eigenvalue of the Jacobian passes through zero. A simple numerical root finding search was implemented (in Mathematica) which used a bisection algorithm and located the critical Froude number for each of the dissipation values used by returning the maximum real part of the eigenvalues of the Jacobian matrix. The only parameter that varied in the Jacobian matrix was dissipation (Γ) and by repeating this process for different dissipation values the linear stability curves were constructed.

5.3.1 Earlier Results

Linear instability studies of Martian baroclinic waves have been the work of Gadian (1978) and Barnes (1984). They both showed that zonal wavenumbers 3-4 were generally the most linearly, most unstable modes with the growth rates of the order of days, similar to the growth rates found in studies of terrestrial baroclinic waves (Simmons and Hoskins (1977,1978), Hollingsworth (1975), Wyatt (1981)). It is interesting to note that similar stability studies for terrestrial zonal flows are typically double these values[7] (e.g. Simmons and Hoskins (1976,1977)). Highly truncated nonlinear models have been used (Barnes (1986)), and diagnostic studies of full GCMs (Pollack et al. (1981) and Barnes et al. (1993)) diagnosed the NASA Ames Martian GCM and found baroclinic waves with wave numbers and periods similar to those seen in the observations and predicted by linear theory.

Barnes (1984)

Barnes examined the baroclinic instability of zonal-mean flows using a spherical QG model, having both vertical and meridional shear, similar to those in the wintertime Martian atmosphere under both relatively non-dusty and highly dusty conditions. The used basic states were idealised, but were based closely upon the Mariner 9 and Viking observations. The most unstable mode at any given zonal wavenumber was calculated by using the initial value approach which was employed by Brown (1969).

Barnes used a 'tilted jet' zonal flow, intended to resemble the wintertime Martian zonal-mean circulation under non-dust storm conditions, which was observed by Mariner 9 IRIS (Conrath (1981)). It was shown that the largest growth rates occurred at zonal wavenumbers 3 and 4, depending upon the friction. Figure 5.1 shows a growth rate curve from one of the calculations, where the curve has a maximum growth rate at wavenumber 3, being of similar magnitude to the terrestrial growth rates. The growth rates and phase speeds of the most unstable modes for realistic Martian zonal flows on a sphere were found to be generally consistent with the observed properties of transient baroclinic disturbances in the Martian atmosphere (Barnes (1980,1981,1983)). It was found that the largest growth rates occurred at zonal wavenumber 3 and 4, with phase speeds $\sim$ 10-20 ms^{-1} in middle latitudes. Barnes showed that the inclusion of zonal-mean topography (similar to that in the Martian northern hemisphere) resulted in a substantial reduction in the growth rates. Barnes (1984) commented that the dominant wavenumber does not necessarily have to have the largest linear growth, as nonlinear interactions can alter the behaviour of the modes.

[7]Primarily, the reason why the most unstable waves on Mars are smaller than those calculated using Earth models is because Mars has a smaller planetary radius, the actual wavelengths of maximum instability being almost equal to those for the Earth.

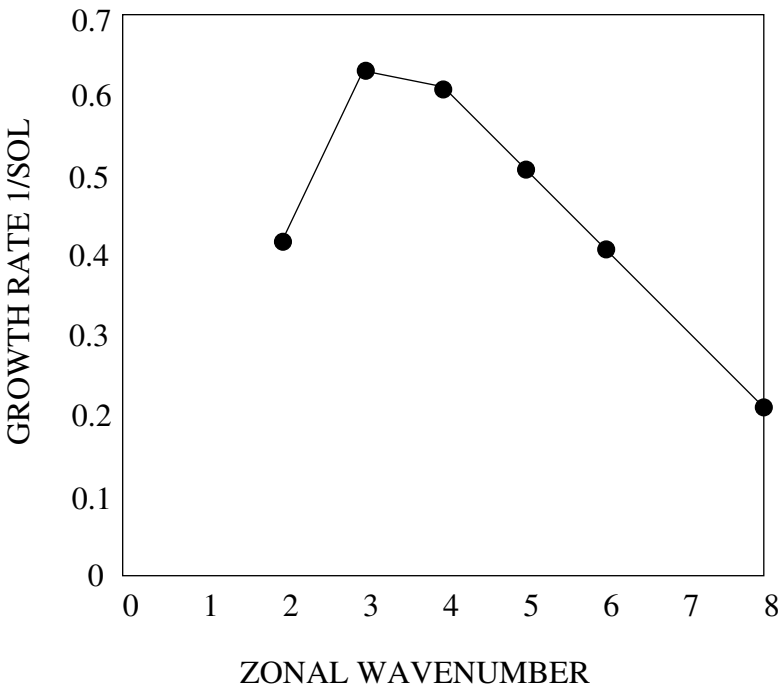


Figure 5.1: A growth rate curve for the most unstable normal modes of a baroclinic jet, corresponding to non-dusty conditions on Mars. Adapted from Barnes (1984).

5.3.2 2-v linear stability results

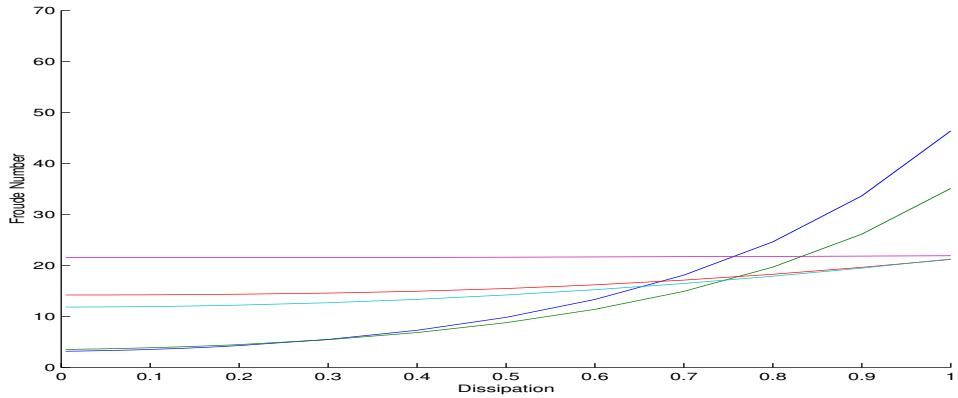
Table 5.1 shows the different POD modes[8] and their associated zonal wave numbers[9]. By taking consecutive pairs of modes and computing the locus of critical Froude number points it is found that for all mode pairs there are relative increases in the critical Froude number as dissipation increases (Figure 5.2). The mode which dominates the fully nonlinear flow i.e. the wavenumber 3 is the most unstable mode at the inviscid limit for all cases and is the most stable for $\mathcal{O}(1)$ dissipation. Interestingly, the ordering of the individual stability curves at the inviscid limit corresponds to the energy content of the modes, i.e. the energy dominant mode is the most unstable at the inviscid limit, and so on. At larger dissipation values, the critical Froude numbers for the first and second pairs of modes begin to increase nonlinearly. The stability curves for the third and fourth mode pairs increase almost linearly as the $\mathcal{O}(1)$ dissipation values are approached. The fifth pairing only change slightly over the dissipation range.

We note that there are very few differences in the leading four wave modes between the linear stability curves from the models determined from either the correlation, KE or TE decompositions illustrated in Figure 5.2.

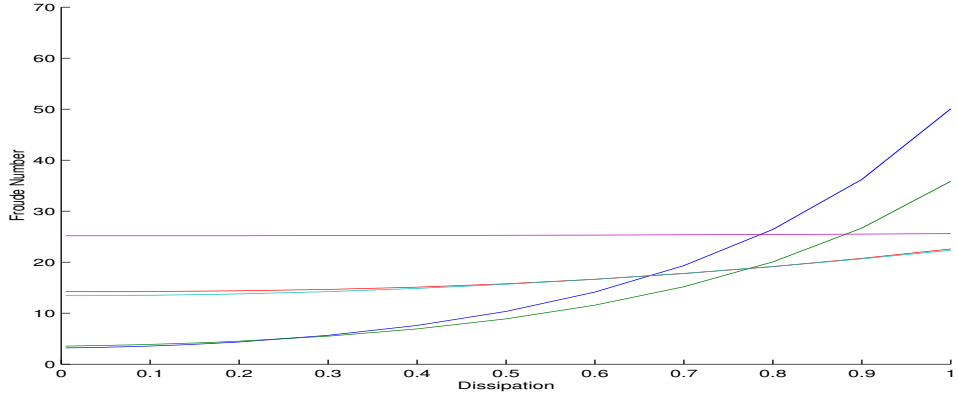
It can be seen that each of the individual linear stability curves of the correlation model appear in the same order in the energy models (except for the third mode pair (mode 5, mode 6) in the TE model) with the main difference being that the individual curves have been shifted

[8]The state vector of the 2-v PODs only included the barotropic and the first baroclinic modes. However, the 2-v spatial structures were found to be almost identical to the 4-v modes (Figures 3.12 - 3.17), i.e. the barotropic and the first baroclinic components of the 2-v and 4-v modes were almost indistinguishable from each other.

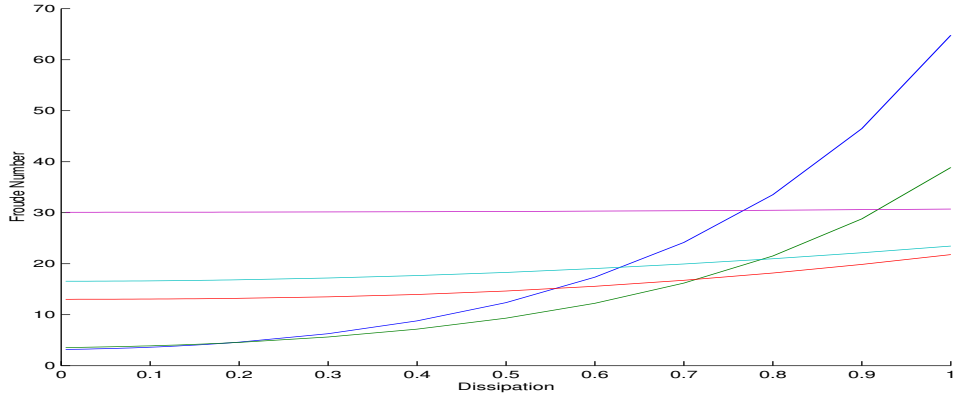
[9](w_1, w_2) means that the first mode has a zonal wavenumber w_1 and the second mode a wavenumber w_2.



(a) CORRELATION GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 5, mode 6) in light blue, (mode 7, mode 8) in red and (mode 9, mode 10) in purple.

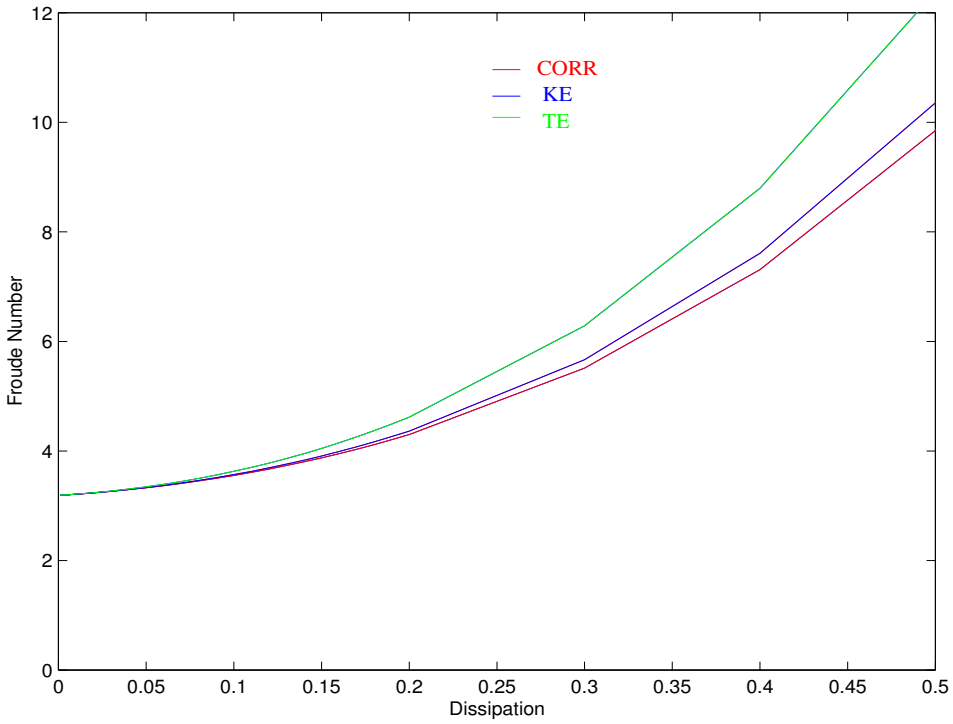


(b) KE GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 5, mode 6) in light blue, (mode 7, mode 8) in red and (mode 9, mode 10) in purple.

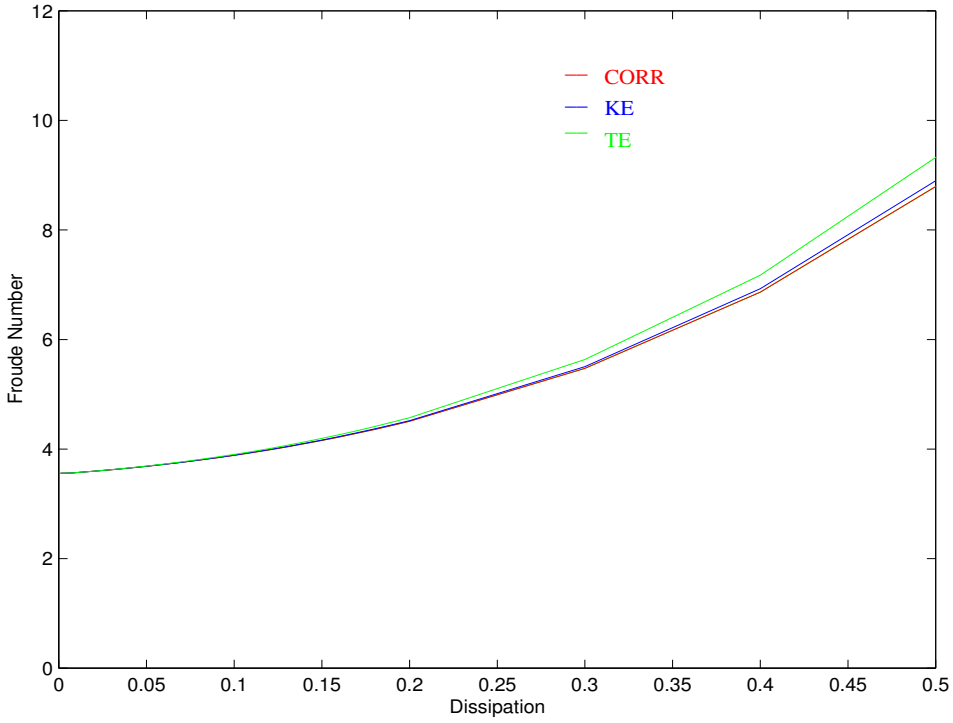


(c) TE GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 7, mode 8) in red, (mode 5, mode 6) in light blue and (mode 9, mode 10) in purple.

Figure 5.2: Linear stability curves for the (a) correlation, (b) KE and (c) TE norm models. The dominant energy mode pair (mode1,mode2) is the most unstable at the inviscid limit, whilst being the most stable at $\mathcal{O}(1)$ dissipation. Note that the curves have been displaced vertically for the energy norm models.



(a) Linear stability profiles for the first wavenumber pair i.e. (mode 1, mode 2) for the correlation, KE and TE norm models.



(b) Linear stability profiles for the second wavenumber pair i.e. (mode 3, mode 4) for the correlation, KE and TE norm models.

Figure 5.3: Linear stability curves for (a) the first wavenumber pair, (b) the second wavenumber pair for the correlation, KE and TE norm models. Note that as $\mathcal{O}(1)$ dissipation is approached the individual stability curves produced by the energy norms become somewhat steeper.

MODES	CORR	KE	TE
(mode 1, mode 2)	(3, 3)	(3, 3)	(3, 3)
(mode 3, mode 4)	(2, 2)	(2, 6)	(2, 6)
(mode 5, mode 6)	(zonal flow, 6)	(2, 6)	(2, 6)
(mode 7, mode 8)	(4, 6)	(4, 4)	(4, 4)
(mode 9, mode 10)	(4, 4)	(3, 3)	(4, 4)

Table 5.1: POD modal pairings and their corresponding spatial wavenumbers.

upwards to higher critical Froude numbers in the correlation model, particularly in the case of the TE norm model.

The energy norms have shifted the third wave pair upwards to higher critical Froude values. This can be seen in Figures 5.2(b) and 5.2(c) where the light blue line (associated with the third wave pair) creeps to greater critical Froude values. This shift is to such an extent that in the TE model the wave pair has become more stable than the fourth pair at the inviscid limit. This highlights that the third wave pair is particularly sensitive to changes in the spatial structures of the POD modes produced by the different norms.

Figure 5.3 shows the linear stability curves for (a) the first and (b) the second wavenumber pairs for the correlation, KE and TE norm models. Comparing the curves produced by the various norms it is observed that the energy stability curves have steeper gradients and have been vertically shifted away from the correlation curves. This suggests that the stability properties of the energy curves are more sensitive to changes in Γ.

Collins (1993) and Collins and James (1995)

According to Collins (1993), the time average of the zonally averaged SGCM flow was found to be baroclinically stable (had only one significant unstable mode, a zonal wavenumber 5) and so the zonally symmetric version of the SGCM was used as the basic state instead. By changing the definition of the basic state, baroclinic modes were found with growth rates of order 1/2 day^{-1}, the largest growth rate occurring at zonal wavenumber 5. Collins and James (1995) commented on this result and stated that the model behaviour could not be simply explained by linear theory, they argued that nonlinear wave and wave-mean flow interactions must be crucial in maintaining a wave 3 flow.

It should be noted that the linearly most unstable mode is not necessarily that which will dominate in the fully nonlinear flow. This was illustrated by Elliot (1995) who studied the linear stability problem for a flow dominated by wavenumber 3 using a high resolution, axisymmetric QG solver and found that the most unstable mode had a wavenumber 15 structure and not the expected wavenumber 3.

Obviously, the findings of Collins (1993) differ with our results as we have generated low-dimensional models based on the SGCM data and have identified the spatial wavenumber 3 pattern as the linearly most unstable mode of the system. This discrepancy with the results of Collins and James (1995) may be due to the severe vertical truncation which we used. As

high wavenumbers tend to have vertically confined structures, then these waves may not be well represented in our truncation, possibly resulting in the wavenumber 3 mode of the wave 3 regime being the most unstable of the system.

5.4 Robustness of the numerical nonlinear models

Since the flow is decomposed into a basic state, mean-flow correction and wave terms then the evolution of the components is influenced by nonlinear terms which represent mean flow-wave interactions or wave-wave interactions. The initially exponential growth of waves from the baroclinically unstable zonal flow is limited by such nonlinear effects, which leads to equilibration to steady or vacillating waves, or to more complex, or possibly chaotic flow. Even in low-dimensional models several possible mechanisms represented by different nonlinear terms can be active and may cause complex flow.

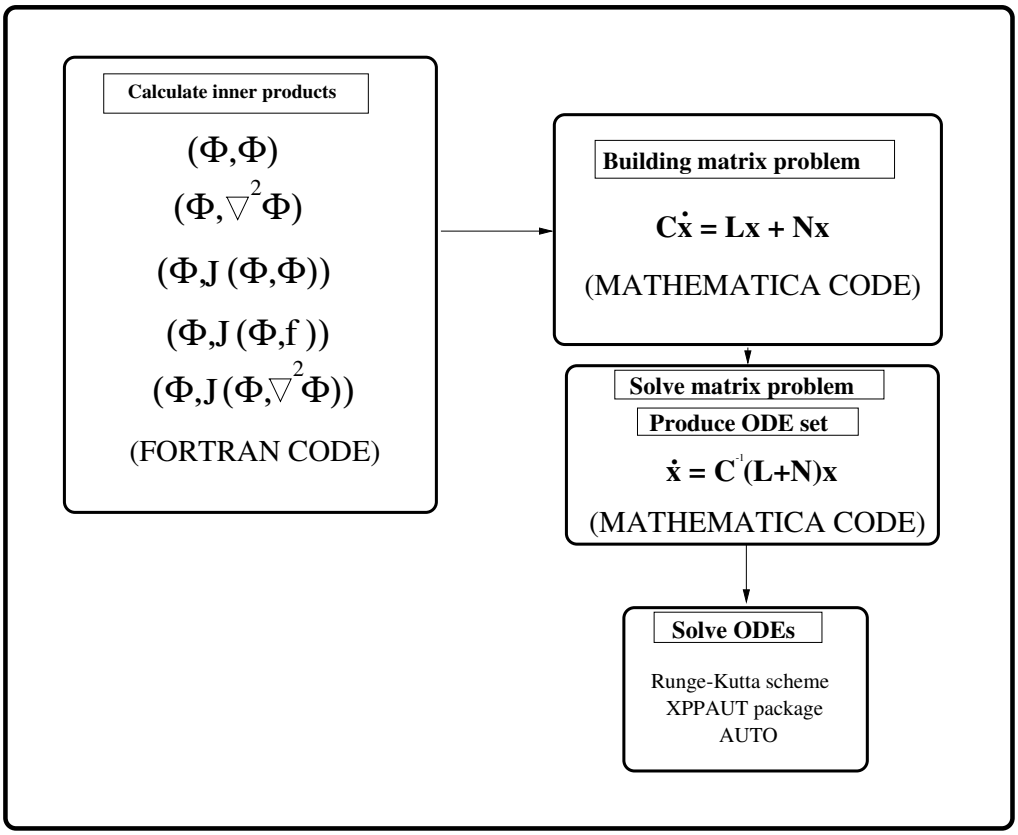


Figure 5.4: The process of producing the Galerkin ODE sets. The coefficients were calculated in Fortran and the matrices (both linear and nonlinear) were built up using Mathematica. The matrix problem was then solved using Mathematica.

The various stages in the development of the Galerkin nonlinear ODE set is schematically shown in Figure 5.4. The necessary Galerkin projections were coded in Fortran. The code read in the various data fields and calculated the different projections. The resulting projection data was then fed into several Mathematica programs, enabling the construction of the required matrices in the calculating of both the linear and nonlinear terms. The matrix problem was then solved, producing a set of ODEs.

For a direct comparison to be made between the solutions of the ODE set and the original SGCM PCs it was essential for the integral time variable t to have units in Martian days. Since the dimensionalised Martian time quantity satisfies $t \rightarrow \frac{1}{\Omega}t'$ (where t' is non-dimensionalised time) and as the rotational rate $\Omega = 7.08822 \times 10^{-5} s^{-1}$ then:

$$\begin{aligned} t &\rightarrow \left(\frac{10^5}{7.08822}\right) \text{secs} \times t' \\ &\rightarrow 0.16081 \text{ Martian sols} \times t'. \end{aligned} \tag{5.8}$$

The Galerkin evaluations in calculating the projections onto the derivative fields and particularly the nonlinear advective terms were computationally intensive. It was therefore necessary to increase the speed of computation. A filter was used in order that the terms which were smaller than a certain threshold value could be ignored.

It was also necessary to ascertain how sensitive the set of ODEs were to certain threshold values. The dimensionless threshold values used were $\{10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}\}$. In order to check the sensitivity of the models to these threshold values it was decided that a separate set of model equations would be generated for each of the given thresholds and comparisons would then be made between them.

Dissipation (Γ)	**Threshold** $= 10^{-8}$	$= 10^{-6}$	$= 10^{-4}$	$= 10^{-2}$
1.0	65.34	65.34	65.32	65.13
0.5	12.72	12.72	12.69	12.41

Table 5.2: Comparison of the critical Froude number for the first POD wave mode with four numerical threshold levels, at two dissipation values for the TE model.

Linear stability curves were computed and then compared for these four thresholds. Table 5.2 shows two critical Froude numbers for two different dissipation values, using the four given threshold values. It can be seen that values smaller than 10^{-6} can be ignored in the filtered models as they do not affect the position of the linear stability curve to 2 decimal places.

To address this issue of robustness four TE 10-D models were produced, each one having a different threshold value. The models were then integrated using the same parameters and initial conditions and results are shown in Figure 5.5 (except for threshold value 10^{-8} as it was identical to the solution curves of 10^{-6}). It can be seen that the two models which have the coefficient thresholds of $\{10^{-4}, 10^{-6}\}$ are almost identical. The model which uses a threshold value of 10^{-2} clearly behaves differently. This confirms that values smaller than the threshold value of 10^{-6} can be ignored and so the threshold value 10^{-6} is used.

5.5 Outline of modelling aims

Certain obvious questions had to be addressed during this analysis, including:

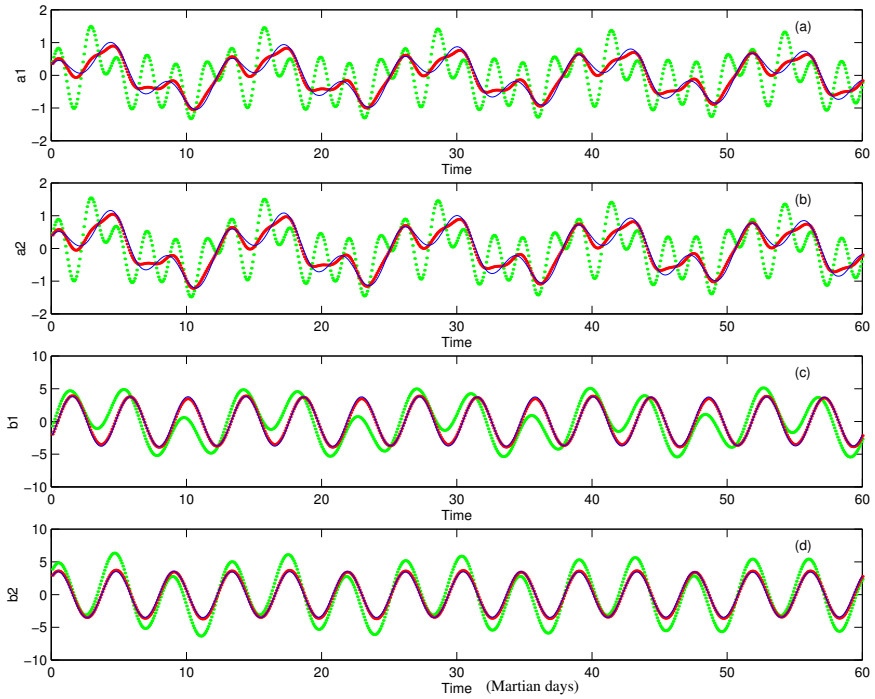


Figure 5.5: Comparing one solution component for each of the different coefficient thresholds $\{10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}\}$. Plots show the mean-flow correction (MFC) modes 1 and 2 (in plots (a) and (b)) and the wave amplitudes 1 and 2 (in plots (c) and (d)) respectively. The plots are red for threshold value 10^{-6}, blue for 10^{-4} and green for 10^{-2}. Threshold value 10^{-8} is omitted, being found to be identical to 10^{-6}.

(1). What is the required vertical resolution in the POD formulation so that the dynamics produced by the low-dimensional models are similar to that seen in the original system?
(2). How does the horizontal truncation affect the accuracy of the low-dimensional models?
(3). Which norm should be selected?
(4). Is it important how large a fraction of the total energy is retained in the Galerkin model? If so, what is a 'sufficient' amount?

Comparisons were made between models which were produced from the correlation, KE and TE norms. It was felt that by using three models with different truncations the question of how the truncation levels affect the quality of the generated models could be addressed.

Truncation	No. of MFC terms	No. of wave terms
(10,10)	10	10
(4,6)	4	6
(2,4)	2	4

Table 5.3: Notational form for the various Galerkin truncations which will be considered in this thesis.

The three different truncation levels used in this thesis (for comparisons) are shown in Table 5.3, where their abbreviated notational forms are explained. By contrasting the qualitative differences in the amplitude and frequency of the generated modes of the 20-D, 10-D and 6-D systems allows the investigation of which modes need to be retained and how many modes are

required in attempting to reproduce the original dynamics.

5.6 Assessment of bifurcation structures

A thorough analysis of the ODEs was required so as to provide us with an understanding of some of the underlying dynamics of the reduced SGCM. This was not possible by simply integrating the Galerkin models for different dissipation values and so the numerical continuation package AUTO (Doedel (1981)) was used to analyse the equations by determining their bifurcation[10] and stability properties. This enabled a comprehensive investigation to be made into the behaviour of the Galerkin models.

As three different truncation levels and three different norms were used, this provided us with nine different models for comparison. The solutions to the various ODEs were found by using AUTO, beginning from the trivial solution and using dissipation as the principal bifurcation parameter. AUTO was run by starting at high dissipation values (hence a stable state) and by decreasing dissipation the various bifurcation points were identified.

AUTO traced the stability of the solution by computing the eigenvalues of the Jacobian for every dissipation value in a given range of interest. In all cases two Hopf bifurcations[11] (related to periodic solutions) were found on the trivial solution. By calling the two Hopf bifurcation points HB_1 and HB_2 (the latter being at the higher dissipation value) it was found that the Hopf bifurcations occurred at different dissipation values for different norms.

Table 5.4 shows the locations of the Hopf bifurcation points on the dissipation axis using the correlation, KE and TE norms in the 6-D, 10-D and 20-D models. It is noted that if the KE or TE norms are used then the Hopf bifurcations are shifted along to smaller values of dissipation. The greatest displacement occurs if the TE norm is used.

All the Hopf bifurcations occur in the region $\Gamma \in [0.37, 0.46]$ and so the use of the different norms has produced only subtle differences in the positioning of the Hopf bifurcations.

The different truncation levels have resulted in very minor differences in the positioning of the Hopf points for a given norm. These differences can only be seen in the third decimal place of the dissipation value (seen in Table 5.4).

In all AUTO runs only two Hopf bifurcation points for the trivial solution were found if F was taken to be 8.14 and Γ was used as the free parameter. AUTO computed the imaginary part of the eigenvalues at the two Hopf bifurcations. As the differences between the 6-D, 10-D and 20-D models for each of the norms were minimal, only the 10-D model is shown for the

[10] A system of ODEs has a **bifurcation** where variation of an external parameter alters the solution's *topological character* - i.e. the overall arrangement of orbits and the sense of flow on them. Especially significant are existence, number and stability of critical points and closed orbits.

[11] A Hopf bifurcation is defined by a pair of eigenvalues crossing the imaginary axis. The Hopf bifurcation involves a non-hyperbolic stationary point with linearised eigenvalues $\pm iw$, and thus a two-dimensional centre manifold, and the bifurcating solutions are periodic rather than stationary. Note that the imaginary part of the eigenvalue pair provide information concerning the period of the limit cycle, where $\tau = \frac{2\pi}{\mathrm{Im}(\lambda)}$.

NORM	**6-D** HB_1	**6-D** HB_2	**10-D** HB_1	**10-D** HB_2	**20-D** HB_1	**20-D** HB_2
correlation	0.4332	0.4663	0.4331	0.4668	0.4337	0.4671
KE	0.4195	0.4617	0.4197	0.4621	0.4206	0.4626
TE	0.3729	0.4501	0.3732	0.4507	0.3738	0.4519

Table 5.4: The location of the Hopf bifurcation points on the dissipation axis at constant F(=8.14) for the 6-D, 10-D and 20-D models.

correlation, KE and TE norms in Table 5.5.

Table 5.5 shows the periods of the regular wave patterns produced by the different Galerkin models which arise from the two Hopf bifurcations at $F = 8.14$. A frequency value of 0.23 cycles/sol describes waves which oscillate at a rate of about 14 cycles every 60 sols and a value of 0.08 cycles/sol a rate of about 5 cycles in 60 sols. Corresponding waves with these frequencies can be identified in the original SGCM (see Figure 3.20).

It is also noted that the imaginary part of the eigenvalues at the Hopf bifurcations are smaller if the correlation norm is used. This results in the periods of the energy norm modes being smaller than those produced from the standard correlation norm. This simply means that the energy norm models produce periodic orbits with greater frequencies.

NORM	**Im(λ_1)**	**Period τ_1** (sols)	**Im(λ_2)**	**Period τ_2** (sols)
correlation	0.2369	26.52	0.0850	73.88
KE	0.2482	25.32	0.0879	71.52
TE	0.2609	24.08	0.0922	68.17

Table 5.5: Two periodic orbits produced from the Hopf bifurcation points HB_1 and HB_2 in the various 10-D models.

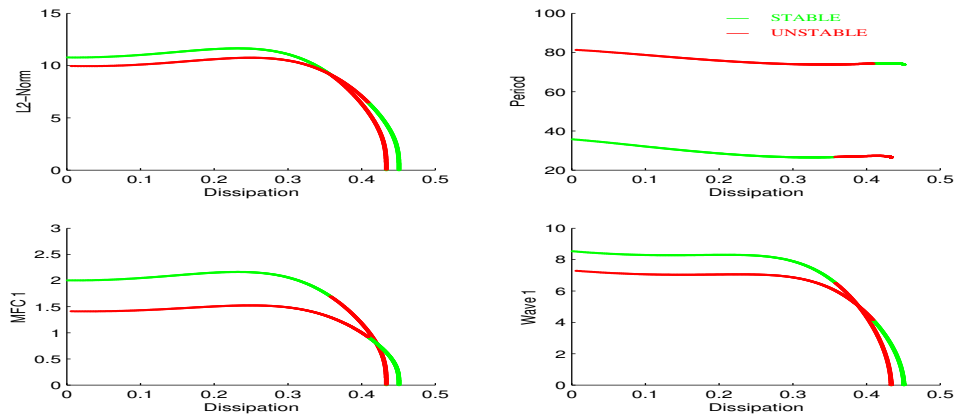
It was then necessary to branch switch at the Hopf bifurcation so that the resulting periodic orbits could be located. To identify the stability properties of the periodic orbits AUTO was set to monitor the Floquet multipliers (see Appendix B).

5.6.1 Comparisons between the models

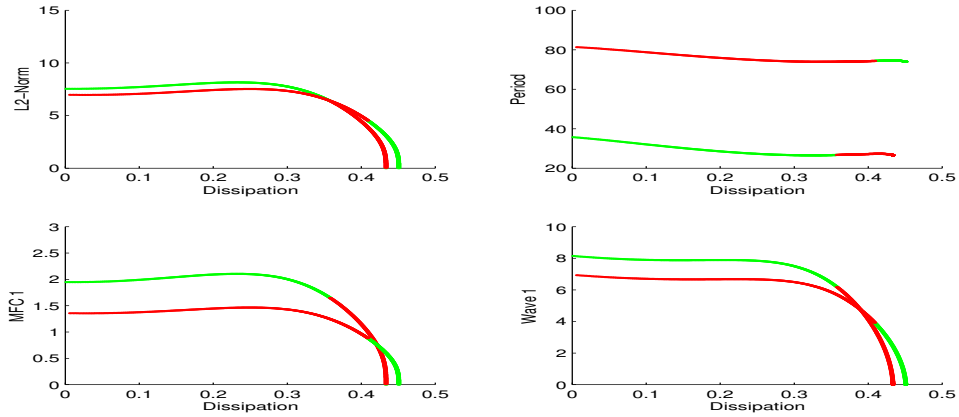
To understand the differences between the correlation, KE and TE norms it was important to produce AUTO runs for each of the generated models and to compare and contrast them in order that the following questions can be addressed :

- How does the truncation level affect the solution's behaviour?
- What differences in the solutions are due to the use of a different norm ?

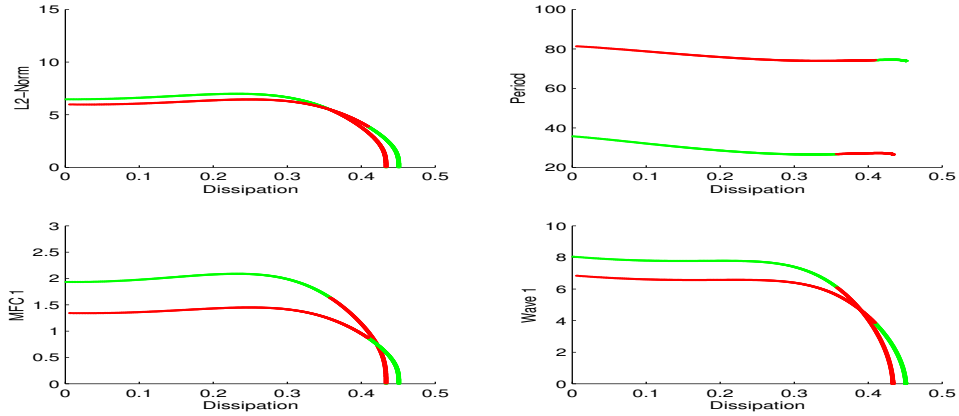
By comparing the correlation norm models (illustrated in Figure 5.6) it is clear that there are only small differences between the various models of truncation. The AUTO results show that



(a) 20-D **CORRELATION** MODEL. Figures of L2-norm, Period, amplitude of MFC mode 1 and amplitude of wave mode 1 against Dissipation.



(b) 10-D **CORRELATION** MODEL. Figures of L2-norm, Period, amplitude of MFC mode 1 and amplitude of wave mode 1 against Dissipation.



(c) 6-D **CORRELATION** MODEL. Figures of L2-norm, Period, amplitude of MFC mode 1 and amplitude of wave mode 1 against Dissipation.

Figure 5.6: AUTO runs for the (a) 20-D, (b) 10-D and (c) 6-D correlation models. Figures show branches of periodic orbits which are shown as green (for stable) and red (for unstable) arising from the subcritical (HB_1=0.4331) and supercritical (HB_2=0.4668) Hopf bifurcations.

the 20-D model has larger L2-norm[12] values than either the 10-D and the 6-D.

The 20-D model does produce stable periodic orbits of a slightly larger amplitude than those generated from the 10-D and 6-D models. This is illustrated by comparing the stable and unstable amplitude branches of the MFC and wave solutions. Even though the discrepancies are small, nevertheless, the leading modes have larger amplitudes in the larger dimensional models.

The KE and TE norms produced slightly different branch structures. Not only were the L2-norm values and the individual amplitude values larger for the KE and TE models than those found by using the correlation norm, but the periods of the energy norm modes (which arose from each of the Hopf bifurcations) were slightly less, i.e. the frequency of the energy modes is greater (for details see Table 5.5). These changes in propagation rates are due to the norms picking out different spatial structures and the way in which they project onto the background flow and the Potential Vorticity gradient.

5.6.2 Model truncation

The necessary model truncation is a significant matter to consider. Even though the SGCM data seems to indicate that there exists some kind of low-dimensional behaviour occurring it may be found that an insufficient amount of retained spatial eigenmodes may seriously damage the quality of our findings.

Schubert (1985) addresses this issue of truncation by discussing how EOFs can be used as a dynamical basis for understanding the nature of the dominant modes of variability of the northern hemisphere winter circulation on Earth. The model was based on equivalent barotropic dynamics, where the usual spectral modes (in this case spherical harmonics) are replaced with EOFs. He found that the major disadvantage of using EOFs was the apparent *'strong sensitivity to the model truncation'*. This lead to a conflict between the statistical results which suggested a possible strong model truncation (12 EOFs which explained 76% of the variance of the system) and dynamical convergence which required a larger number of degrees of freedom (more than 40 EOFs needed in order that 97% of the variance could be explained).

It must be noted in our case, however, the AUTO plots shown in Figures 5.6 suggest that there is very little *'sensitivity to the model truncation'* in amplitude and period. If this happens to be the case then our results will seriously contrast with those generated by Schubert's EOF models of the Earth's atmosphere (1985). This difference may be due to the SGCM Galerkin models being generated in a physical regime which has fewer modes competing for dominance and thus being intrinsically lower dimensional, and also due to the fact that the Earth's wintertime flow is more complicated than the SGCM dynamics.

Such discrepancies may also be due to the differences in our spatial eigenfunction formulation, our use of norm or our vertical resolution scheme. If our models are insensitive to truncation then this may indicate that our Galerkin process is very efficient in extracting the important

[12]AUTO uses the period and the L2-norm (which is defined as the square-root of the sum of the squares of each of the amplitudes) as solution measures.

information and/or the SGCM data is low-dimensional, resulting in a 6-D model doing rather well at reproducing the first four SGCM wave modes.

5.6.3 Stability properties of the periodic orbits

We will denote '**subsidiary waves**' as those waves which arise from the Hopf bifurcation HB_2 (which travel at a rate of about 4-5 cycles every 60 sols) and '**principal waves**' as those which are produced from HB_1 (with a frequency of about 12-14 cycles every 60 sols). A '**mixed-mode**' type will be found to be a *superimposed* 'principal' and 'subsidiary' wave.

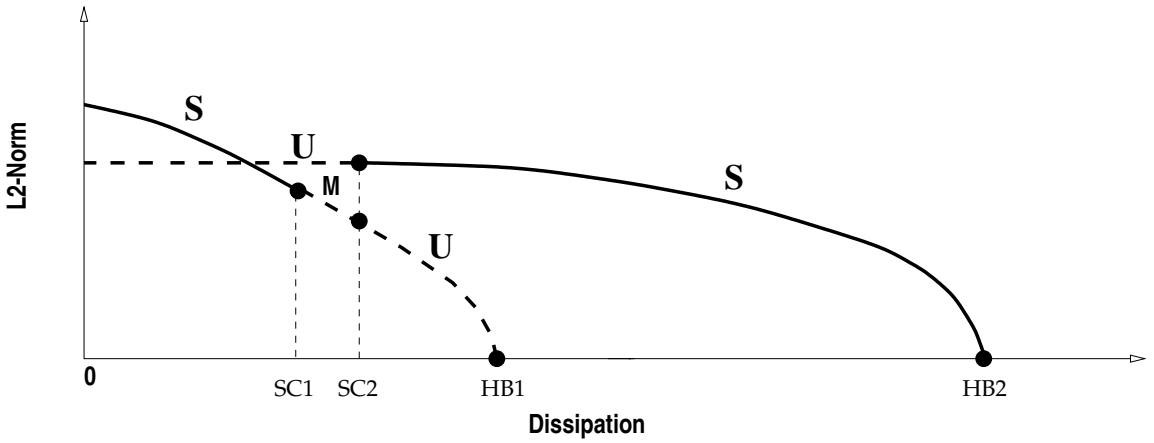


Figure 5.7: Stability exchanges along the solution curves produced by AUTO. Figure shows the stable 'subsidiary' periodic orbits arising from supercritical Hopf bifurcation HB_2 and the unstable orbits arising from subcritical HB_1. The region of the stable 'mixed-mode' is identified by the letter 'M' which exists between the subsidiary wave regime (to its right) and the principal wave regime (to its left), and is shown in Figure 5.8.

Figure 5.7 illustrates the stability exchanges which occur along the solution curves of the AUTO runs where '**S**' denotes a stable periodic orbit, '**U**' unstable and '**M**' the appearance of a stable 'mixed-mode' type. By branch switching at the Hopf bifurcation resulting periodic orbits were located. AUTO determined whether the orbits were stable or unstable by monitoring the Floquet multipliers. The locations of the various Hopf bifurcations (the '**HB**' points) and the various stability exchanges (at points '**SC**') for the numerous 10-D models are shown in Table 5.6.

TE	**KE**	**correlation**
HB_1=0.3732	HB_1=0.4197	HB_1=0.4331
HB_2=0.4507	HB_2=0.4621	HB_2=0.4668
SC_1=0.2973	SC_1=0.3439	SC_1=0.3571
SC_2=0.3213	SC_2=0.3892	SC_2=0.4125

Table 5.6: The location of the various Hopf bifurcations (HB_1 and HB_2) as well as points of stability exchange (SC_1 and SC_2) for the 10-D correlation, KE and TE Galerkin models.

By using dissipation as the control parameter, beginning at high dissipation ($\Gamma = 1.0$) and decreasing, it was found that the periodic orbits which arose from the Hopf bifurcation HB_2 were always stable i.e. they were *supercritical*. In all cases considered there exists another Hopf

bifurcation at a smaller dissipation value (denoted HB_1) which was always *subcritical* in nature.

From HB_2 the amplitudes grow with decreasing dissipation, whilst the period remains almost constant. Interestingly, the steady periodic waves lose their stability at the point at which the stable mixed mode appears.

The unstable orbits which arise from HB_1 also grow in amplitude as dissipation is decreased and the unstable branch eventually stabilises at the point at which the orbits arise from HB_2 becomes unstable.

In all cases considered the following behaviour occurs:

- $\Gamma \in$: [HB_2, 1] NO WAVES are found;
- $\Gamma \in$: [HB_1, HB_2] the subsidiary wave is **stable**;
- $\Gamma \in$: [SC_2, HB_1] the subsidiary wave is **stable** and the principal wave is **unstable**;
- $\Gamma \in [SC_1, SC_2]$: the subsidiary wave is **unstable**, the principal wave is **unstable** and the 'mixed' wave is **stable**;
- $\Gamma \in [0, SC_1]$: the subsidiary wave is **unstable** and the principal wave is **stable**,

where [ ] denotes the interval under consideration.

Figure 5.8 shows the various regimes which exist for constant $F = 8.14$ and varying values of Γ. At high Γ (=1.0) there are no waves and the trivial solution is always stable to any perturbations. The dissipation parameter decreases until the Hopf bifurcation HB_2 is reached. At this point stable subsidiary waves are produced and as Γ is decreased further these stable waves get larger in amplitude. Through this subsidiary wave region the other Hopf bifurcation point (HB_1) is passed through. Within this region, and to the left of HB_1 the unstable principal waves become stable subsidiary waves. A transition occurs to a mixed-mode regime once point SC_2 has been reached. At this point (and to its left) mixed-waves occur until point SC_1 is met. It is at SC_1 where the mixed-wave mode becomes a stable principal mode. Principal waves are evident throughout the principal-wave regime region right through to zero dissipation.

Lovegrove (1998)

Lovegrove (1998) combined theoretical and experimental investigations to study baroclinic instability in a rotating two-layer fluid. A set of spectral amplitude equations were used to describe the two-layer fluid in a rectangular channel and were examined for bifurcations as certain model parameters were varied. The bifurcation parameters used were the dissipation Γ; the rotational Froude number F; and β, a measure of the planetary vorticity gradient. It was shown that the presence of β altered the symmetry of the spectral amplitude equations, fundamentally altering the observed bifurcation structures. It was found that as the value of β was increased, the dynamics of the spectral model became simpler. In particular, the torus bifurcation which had been subcritical at small β (in agreement with the multiple scales results of Fowler and McGuinness (1984)), became supercritical at higher values of β and for sufficiently large β (larger than 1.84) no secondary Hopf bifurcations were found.

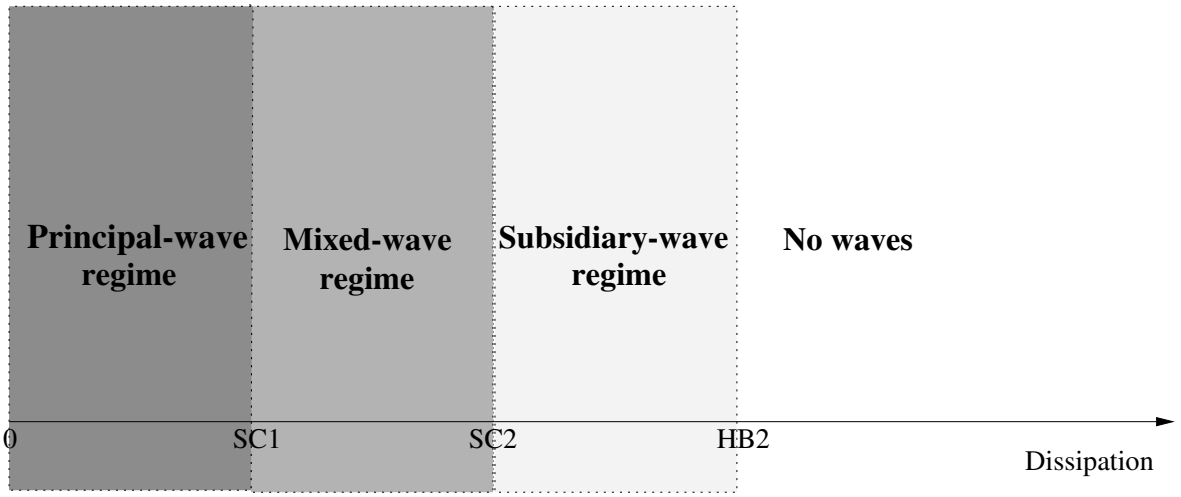


Figure 5.8: The various regimes which occur if Γ is varied and $F(=8.14)$ is kept fixed. Due to the existence of two Hopf bifurcations (HB_1 and HB_2) the waves change from a stable subsidiary wave (for high Γ) to a stable mixed-mode (superimposed principal and subsidiary waves) to finally a stable principal wave (for low Γ) if Γ if decreased. The positions of points SC_1, SC_2 and HB_2 are shown in Figure 5.7.

Since the two-layer studies of Lovegrove are comparable with the two-mode models which have been developed in this section, it is certainly worth comparing our observations and determining whether there are any similarities. Figure 5.7 reveals the solution curves which arise from the two Hopf bifurcations, as well as their stability properties. As shown, the stable subsidiary periodic orbits arise from the supercritical Hopf bifurcation HB_2 and the unstable orbits arise from subcritical HB_1. These periodic branches form three distinct regimes, namely a subsidiary, mixed and principal wave regime (see Figure 5.8). However, it should be stressed that in none of our AUTO calculations was a secondary Hopf (a torus bifurcation) observed, and according to Lovegrove (1998) this may possibly be the result of β being sufficiently large. This may certainly be a possible explanation as to why such simple regimes have been observed in our ODE sets. The fact that the full Coriolis parameter has been retained in our governing set of equations (in order that the full effects of the planet's sphericity is retained) may be suggesting that the Coriolis parameter has effectively eliminated any possible double Hopf bifurcation which may have occurred.

5.7 Single wave regimes

The following results now presented are from integrations of the 6-D, 10-D and 20-D models, using the correlation, KE and TE norms, for given dissipation values and for a fixed Froude number value of 8.14. The regions in which particular modes dominate the system will be highlighted, and evidence of the 'mixed' mode will be illustrated.

Notation

For convenience, the following notation shall be used in our discussions:

• '**Principal**' and '**subsidiary**' will refer to those oscillations which respectively arose from

Hopf bifurcations HB_1 and HB_2, namely those which travel at rates of 13-14 and 4-5 cycles every 60 sols respectively (see Figure 5.7).

• The state vector of the system was defined as:

$$\mathbf{a} = [a_1, ..., a_n, b_1, ..., b_m]\,. \tag{5.9}$$

Here the a_i's are the mean-flow correction terms (n in total) and the b_i's are the departure terms (m in number). Thus, the terms '**MFC mode**' and '**wave mode**' will respectively refer to the temporal descriptions of the mean-flow correction and the departure terms. For further details refer to § 5.5.

5.7.1 Numerical integrations

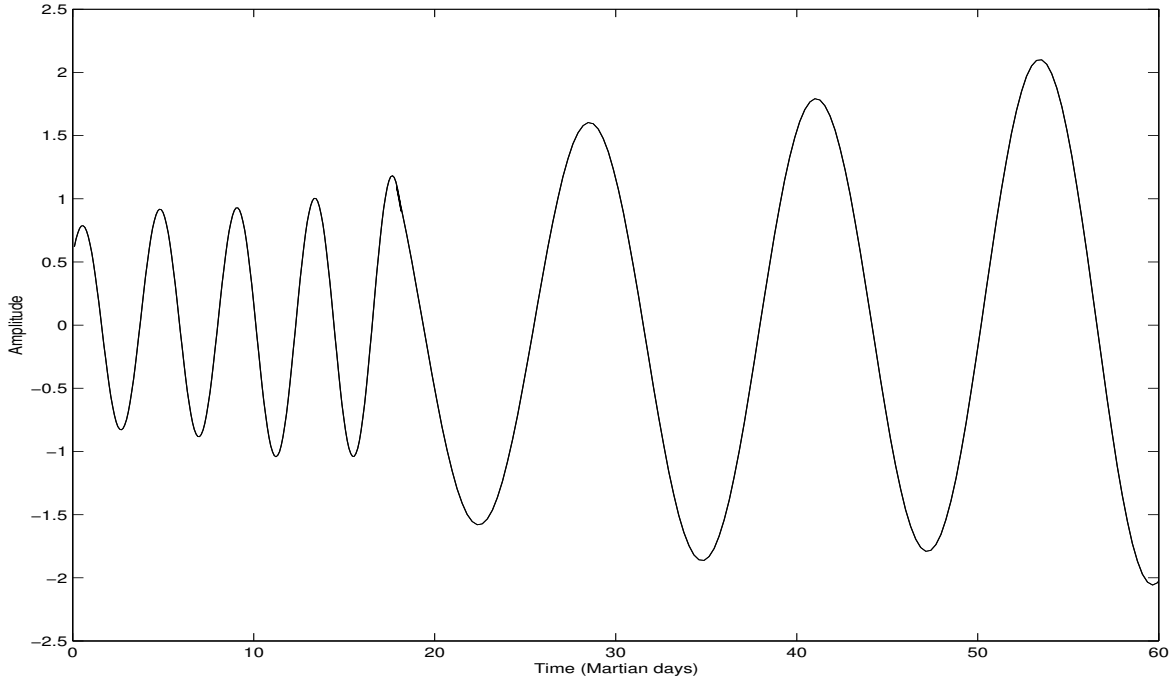
Figure 5.9(a) shows a transient behaviour of a typically spontaneous time dependent integration of the 20-D TE model in the subsidiary wave regime region [SC_2, HB_1] at $\Gamma = 0.33$, where initially each of the unstable modes are travelling with the frequency of the principal wave. Each mode is growing in time and finally stabilises, travelling at the frequency of the subsidiary wave which is shown in Figure 5.9(b). This demonstrates that within region [SC_2, HB_1] (see Figure 5.7) the system is initially unstable, but the unstable principal wave equilibrates to a stable subsidiary wave when the system has finally stabilised. So in the physical system, after a sufficient time period (about 50 Martian days), only a stable subsidiary wave is observed. This process is consistent with what was seen in the AUTO analyses shown in Figure 5.6 where the unstable principal branch is below the stable subsidiary branch and so the mode grows until it finally reaches the upper branch. This illustrates how the two wavenumbers associated with the most unstable wave modes interact with each other.

Figure 5.10 highlights the differences in transient times between the different norm models, where the Root-Mean Square (RMS) values of the first wave mode in the correlation, KE and TE norm models are shown. The top diagram (a) shows that the correlation norm mode takes about 50 days to reach its equilibrated amplitude and a further 80 days to finally equilibrate. This compares with the middle figure (b) which shows that the KE norm mode takes a little less than 50 days to reach its correct amplitude and a further 60 days to stabilise. From the lowest panel (c) it is shown that the TE norm mode grows to its equilibrated amplitude very quickly. Within 20 days it has reached its final state amplitude and after a further 50 days the TE mode has equilibrated. This means that it takes approximately 80 days for the TE mode to grow from an arbitrarily small size, to reach its stable amplitude and equilibrate, compared with about 130 and 110 days for the correlation and KE norm modes respectively. In all of our integrations this phenomenon has been observed.

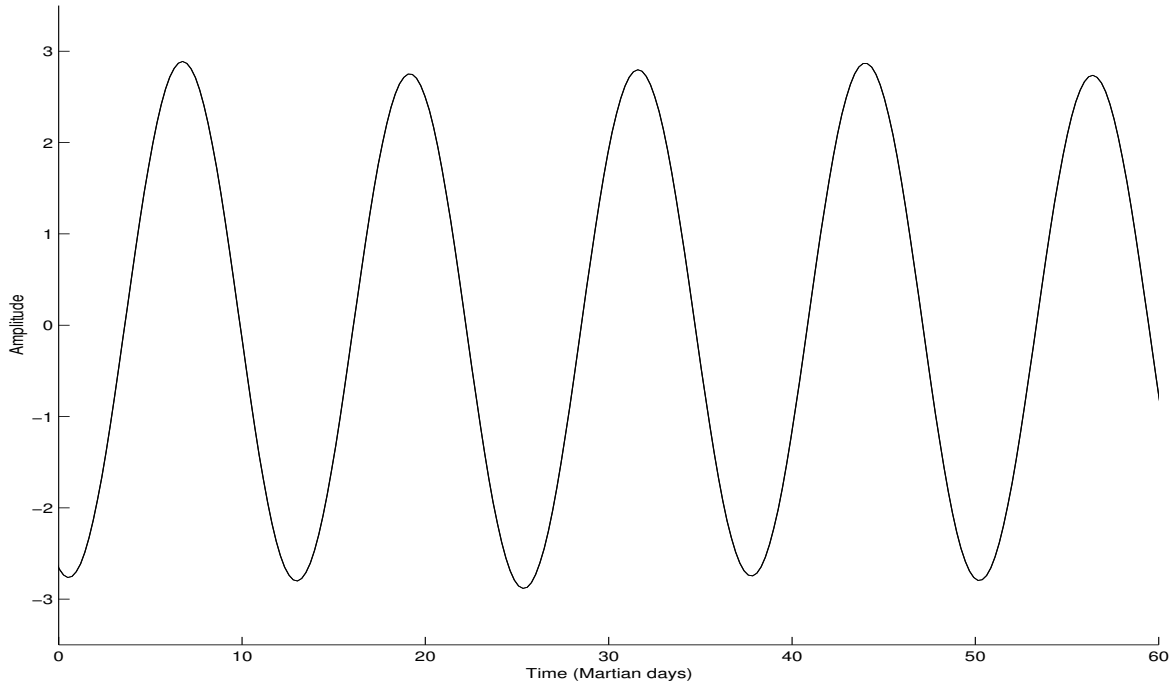
5.7.2 Power Spectra Results

The power spectra for the principal wave regime (analogous to the wave 3 regime as these stable waves arise from HB_1) is very similar to that found by Collins[13] (1993) from the original SGCM data set. Collins showed that for the wave 3 regime there existed one dominant

[13]It must be noted that Collins used the power spectrum developed by Chatfield (1980) which included a smoothing process for more accurate calculations.



(a) 20-D **TE** MODEL. Integrating from small initial conditions in the subsidiary wave regime (at Γ=0.33). Figure shows wave mode b1 originating as an unstable principal wave (on the unstable branch which arises from HB_1) changing to an unstable subsidiary wave, growing over the first 60 Martian days.



(b) 20-D **TE** MODEL. The unstable subsidiary waves have finally equilibrated after 100 Martian days. In other words, mode b1 has reached the upper stable branch (which emerges from HB_2) and has stabilised (see Figure 5.6 for example).

Figure 5.9: A small unstable principal wave growing and finally equilibrating (after 100 Martian days) to a stable subsidiary wave in the subsidiary wave regime (at Γ=0.33) in the region [SC_2,HB_1]. These figures highlight the process of equilibration and confirms the results from AUTO.

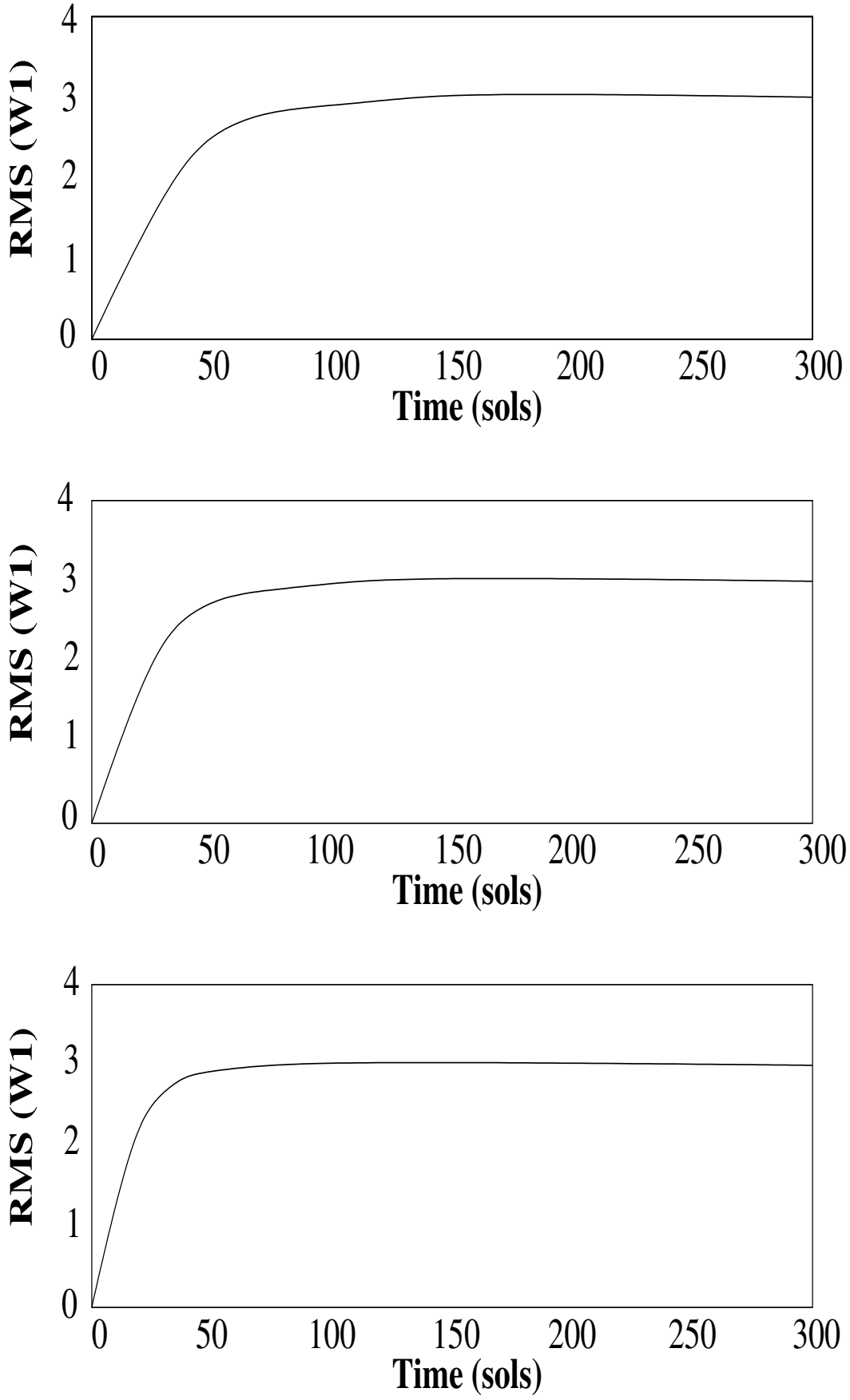
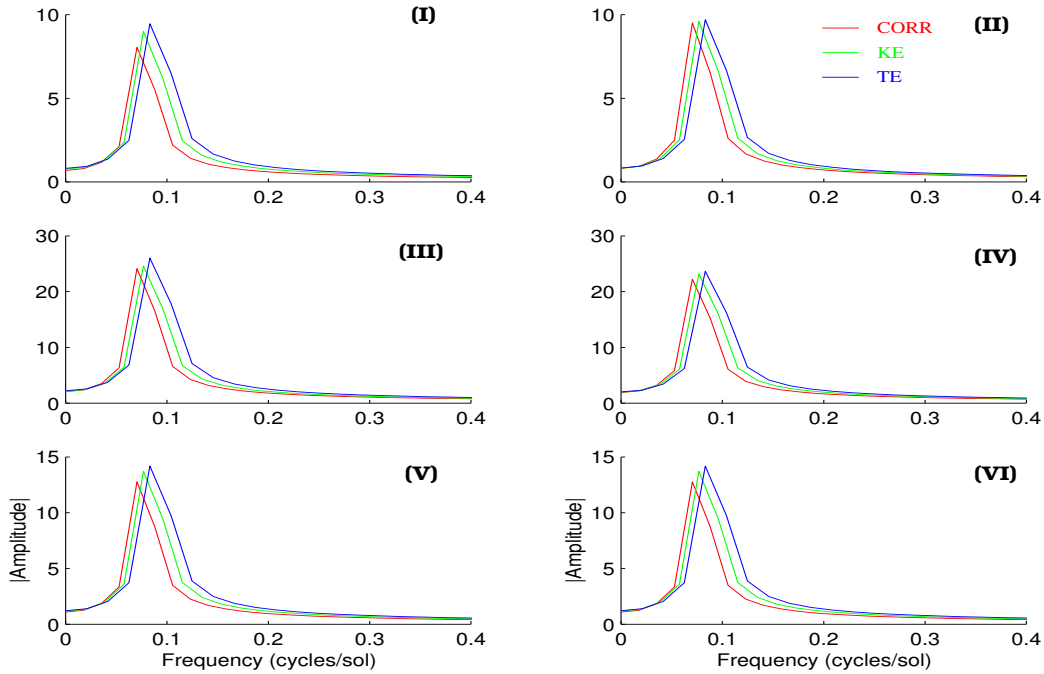
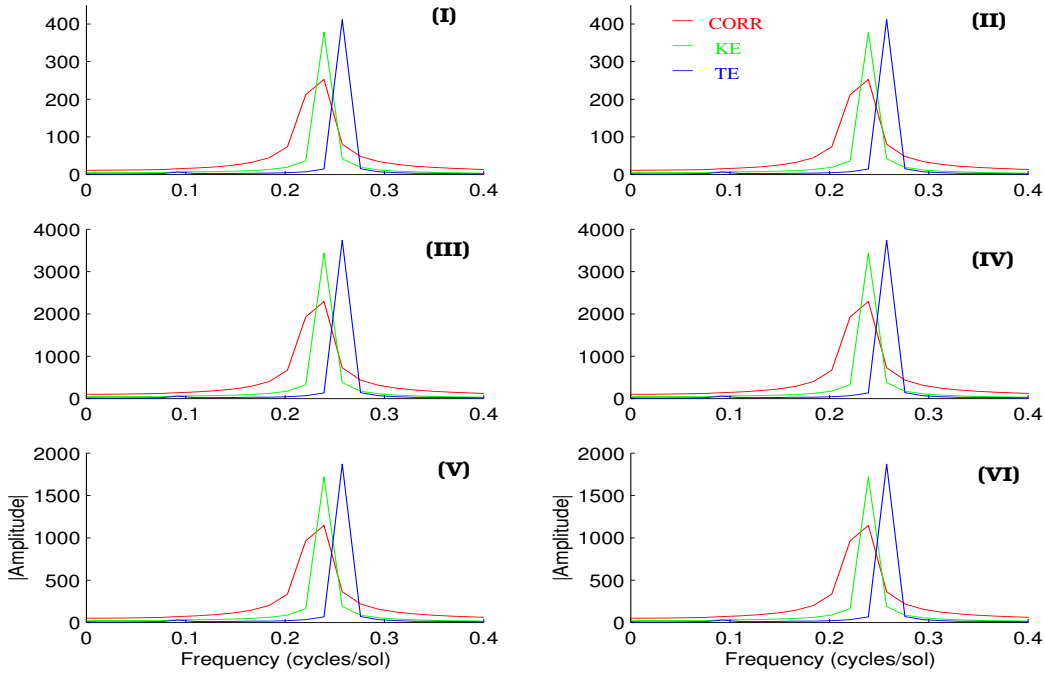


Figure 5.10: Comparisons of the equilibration times for the 20-D correlation, KE and TE norm models. Figures show the Root-Mean Square (RMS) of Wave 1 (b1) for (a) the correlation (b) the KE and (c) the TE norm. It is noted that the equilibration time for the 20-D TE norm model is considerably faster than the others.



(a) POWER SPECTRA FOR THE SUBSIDIARY WAVE REGIME. A clustering of peaks are found about 0.09 cycles/sol. The energy norm spectra generally peak at higher frequencies than for the correlation norm and are greater in amplitude.



(b) POWER SPECTRA FOR THE PRINCIPAL WAVE REGIME. A concentration of peaks exist about the frequency band (0.23,0.25) cycles/sol. The energy norm spectra generally peak at higher frequencies than for the correlation norm and are greater in amplitude.

Figure 5.11: Power spectra for the (a) subsidiary and (b) principal wave regimes using the 10-D correlation (red), KE (green) and TE (blue) norm models, respectively for MFC modes a1 and a2 (I,II), wave modes b1 and b2 (III, IV) and wave modes b3 and b4 (V,VI). The modes travel at one dominant frequency.

frequency around 0.21-0.22 cycles per sol for the surface pressure at 47^o in the southern hemisphere.

Figures 5.11(a) and 5.11(b) show the power spectra profiles for the modes in the subsidiary and principal wave regimes respectively. Here we see examples of 1-frequency flows, where all the modes travel at the same frequency, either a subsidiary or a principal wave respectively.

Figure 5.11(a) shows the subsidiary wave regime for the first two MFCs and the first four wave modes. The frequencies cluster around 0.08-0.09 cycles/sol. Again the effect of using the energy norm can be seen in the frequency spectra, where the TE modes consistently travel with higher frequencies than the correlation norm modes.

Figure 5.11(b) shows the power spectra of the first two MFCs and the first four wave modes in the principal wave regime. It is noted that the frequencies of the principal wave range between 0.23-0.25 cycles/sol depending upon which norm has been used. It is clear that the use of the TE norm has increased the frequency of the travelling modes.

As with Figure 5.11(b) the power spectra profiles in Figure 5.11(a) are almost identical between the different modes. The differences between the correlation and energy norms are not so clear. Again there is a consistency with the eigenvalues obtained from the Jacobian at Hopf bifurcation point HB_2 which were 0.0850, 0.0879 and 0.0922 respectively for the 10-D correlation, KE and TE norm models shown in Table 5.5.

Summary

Evidence has been provided that the models derived using the energy norm equilibrate much faster than models produced using a correlation norm. Even though the power spectra of the energy norm models are almost identical, the speed of equilibration separates the TE from the KE norm models.

The power spectra from the Galerkin models have indicated that the correct frequency of the spatial wavenumber 3 structure can be picked out. It has also been shown that the energy norms produce modes with greater frequencies than those obtained from using the standard correlation norm.

5.7.3 The system's attractor

We will consider the dynamics of an explicit system of m first order ODEs,

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t \in \mathbb{R}. \tag{5.10}$$

If $\mathbf{f}$ does not explicitly depend on t then the flow is autonomous. A sequence of points $\mathbf{x}(t)$ solving (5.10) is a trajectory of the dynamical system where $\mathbf{x}(0)$ is the initial condition. Typical trajectories will either run away to infinity as time proceeds or stay in a bounded area, which is the case we are interested in.

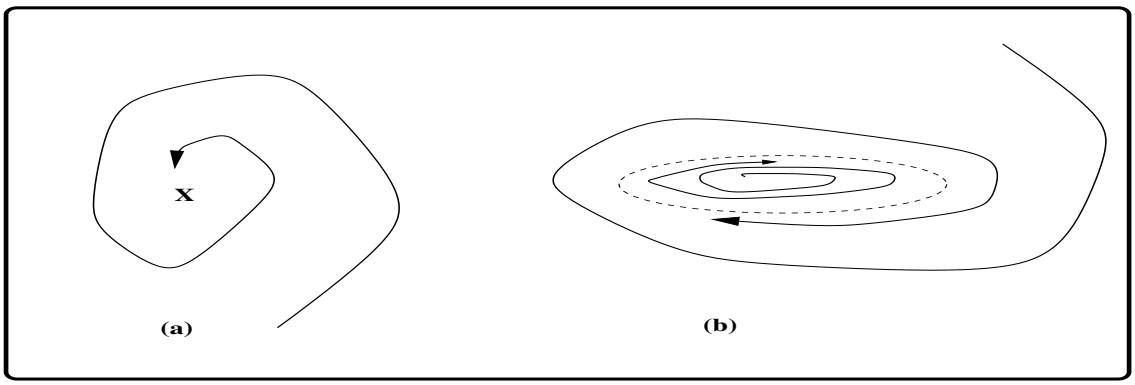


Figure 5.12: Attractors in the plane. Here (a) describes a fixed point and (b) a limit cycle. Depending on the initial condition, a trajectory approaches the limit cycle in (b) either from within or without.

If the dynamical system is dissipative (which means that on average a phase space volume containing initial conditions is contracted under the dynamics) then, on average we have $\nabla \cdot \mathbf{f} < 0$ or $\det|\mathbf{J}_f| < 1$ where $\mathbf{J}_f$ is the Jacobian matrix of the derivatives of $\mathbf{f}$ defined as:

$$(\mathbf{J}_f)_{ij} = \frac{\partial}{\partial x^{(j)}} \mathbf{f}_i(\mathbf{x}). \tag{5.11}$$

For such systems, a set of initial conditions of positive measure will, after some transient time, be attracted to some subset of phase space that contains all of the asymptotes (as $t \to \infty$). This set itself is invariant under the dynamical evolution and is called the *global attractor*[14] (Constantin et al. (1988)). Loosely, the attractor is a set to which all neighbouring trajectories converge. Simple examples of (non-chaotic) attractors are fixed points (after the transient time the system settles to a stationary state) and limit cycles (the system approaches a periodic motion) shown in Figure 5.12. More interesting attractors can occur for flows in at least 3-D where the mechanism of stretching and folding can produce chaos (a detailed description of the various types of observed attractors is given by Kantz and Schreiber (1997)).

5.7.4 Phase space and attractor reconstructions

There are several reasons to believe that the evolution of the atmospheric circulation on Mars takes place on a relatively low-dimensional, chaotic attractor. Conveniently, EOFs are patterns that efficiently describe the attractor (Preisendorfer (1988)). Thus our Martian POD modes should provide an efficient way in formulating a model of the attractor of the original SGCM.

Lorenz (1984) postulated that the state of the atmosphere can be depicted as a point in an infinite dimensional phase space, whose coordinates are given by the values of the state variables at every point in physical space. As the circulation evolves, this point traces out a trajectory in phase space. The geometrical object, constituted by the set of points which the trajectory repeatedly approaches arbitrarily closely in due time, thus is called the attractor. The phase space trajectories of the system always lie on the attractor. As Mo and Ghil (1987) pointed

[14]The term attractor is difficult to define in a rigorous way. There is still disagreement what the exact definition should be. The reader is referred to Guckenheimer and Holmes (1983, p. 256), Eckmann and Ruelle (1985) and Milnor (1985) for discussions.

out, EOFs tend to point from the time mean to the most populated regions of the system's phase space.

EOF analysis has also been used for dimension estimates by Broomhead et al. (1987). They claim that the number of significant eigenvalues of the covariance matrix mirrors the dimension of the subspace which contains the attractor. This implies the definition of the noise floor. All eigenvalues smaller than this are considered to reflect mere noise directions. Numerical simulations show that for many models there are more eigenvalues larger than the noise floor than are expected from the known dimension of the attractor. One explanation is that the manifolds which contain the attractors are not usually linear subspaces and typically a strange attractor has a complicated structure and many folds (Abraham and Shaw (1983)). The dimension of the attractor may be low, but because of the localised curvature, an EOF basis may not be able to take much advantage of the low dimensionality since it is just a linear transformation of the original basis in phase space. For an application of using EOFs as an embedding procedure see Albano et al. (1988).

As many nonlinear dissipative systems embed attractors whose dimensions are considerably less than the dimension of their entire phase space (Kantz and Schreiber (1997)), it might then be possible to identify low-dimensional attractors present in both the original SGCM and the Galerkin models. As EOFs point to the regions of large variance (and covariance), an understanding of how these span the phase space is essential if the attractor is to be identified (Mo and Ghil (1987)). By making comparisons between the original and reconstructed attractors a good indication of whether these low-dimensional models are accurately picking out the important dynamical modal behaviours can be made.

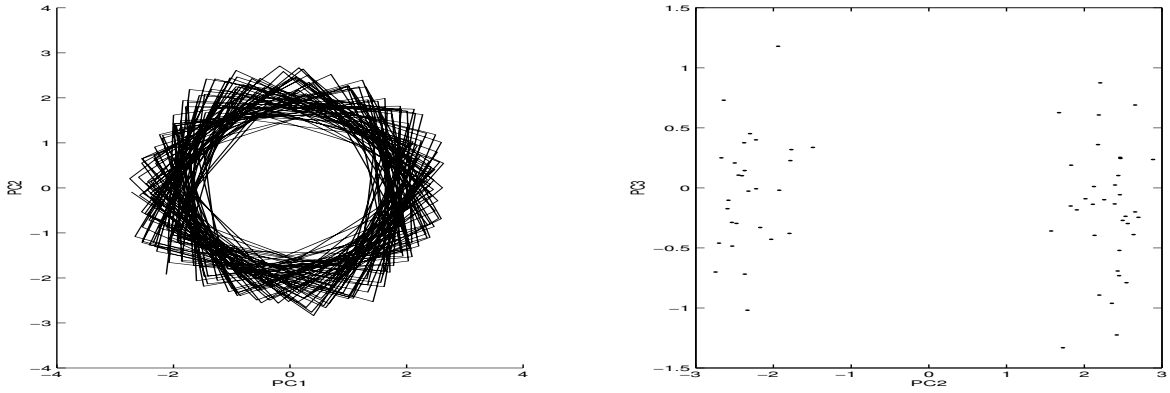
In this section we will investigate to what extent it is possible to reproduce the global structure of the original attractor using less than 20 POD modes.

5.7.5 Phase portraits

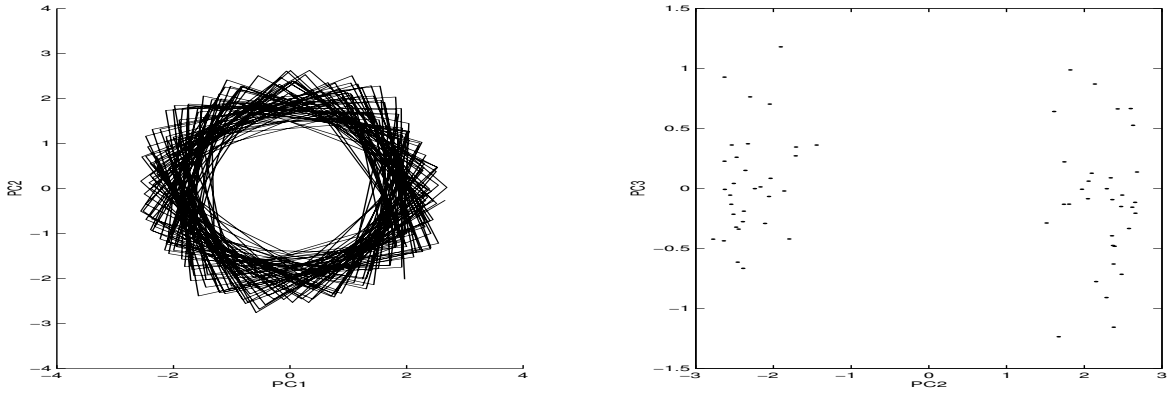
Figures 5.13(a), 5.13(b) and 5.13(c) show the phase portraits and Poincaré sections (at PC1=0) of the SGCM PCs from the original data, using the correlation, KE and TE norms respectively.

As the two leading POD modes explain about 80 % of the variance (see § 3.3.9), the system's trajectories will spend most of their time very close to the plane spanned by these two POD modes. The trajectories trace out a circle in PC space, very close to a limit cycle (see Guckenheimer and Holmes (1983)).

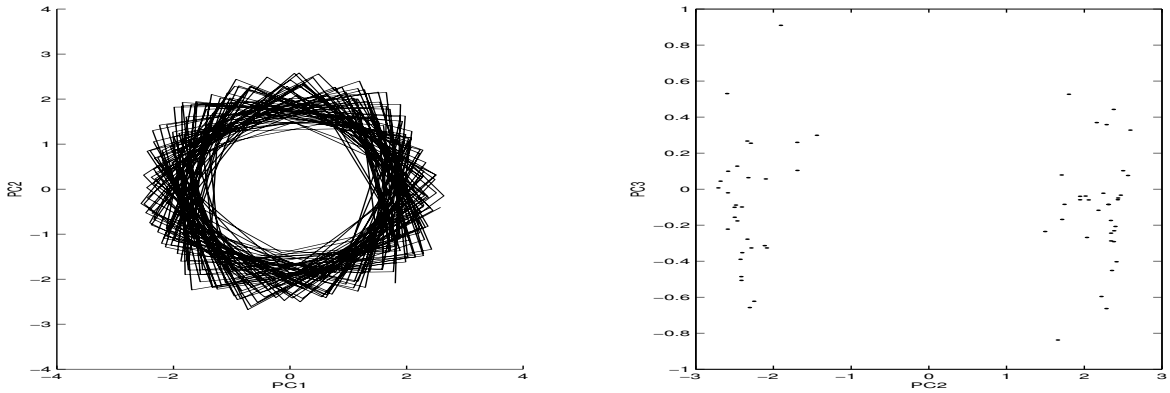
The Poincaré sections do not reveal any further structure, and instead of observing a possibly noisy toroidal surface, the trajectories appear to cover the entire interior of the surface. This confirms that (PC1,PC2)-space describes a periodic orbit, albeit 'fuzzy' in nature (due to the modulations in the leading PCs and the fact that the time interval between successive PC values is 1 sol), and thus substantiates the findings of Collins and James (1995). The phase portraits of the leading two wave modes are almost identical between the norms confirming the behaviour of the PCs seen in Figure 3.18.



(a) PHASE PORTRAIT AND POINCARÉ SECTION (AT PC1=0) OF THE SGCM DATA USING THE CORRELATION NORM.



(b) PHASE PORTRAIT AND POINCARÉ SECTION (AT PC1=0) OF THE SGCM DATA USING THE KE NORM.



(c) PHASE PORTRAIT AND POINCARÉ SECTION (AT PC1=0) OF THE SGCM DATA USING THE TE NORM.

Figure 5.13: A comparison of the phase portraits and Poincaré sections (at PC1=0) from the original SGCM data. The phase portraits depict structures which are close to periodic orbits (analogous to a limit cycle), and are shown to be norm independent. As the two leading modes explain so much of the energy then the system's trajectories will spend most of their time very close to the periodic structure and can be thought of as being the attractor of the system. The Poincaré sections do not reveal any further structure, where the trajectories appear to cover the entire interior of the surface.

5.8 Galerkin model comparisons

5.8.1 20-D models

Each of the following integrations are within the mixed mode range of [SC_1, SC_2], illustrated in Figure 5.7, at dissipation values $\Gamma = 0.395$, 0.35 and 0.34 for the correlation, KE and TE norm models respectively, over 60 Martian days (or 60 sols) after a transient of 200 days. It should be noted that the leading wave pairs of the various Galerkin models are comparable in amplitude to the original SGCM PCs at these selected values of Γ.

In Figures 5.14 and 5.15 the MFC and wave modes of the 20-D correlation norm model are illustrated respectively. By using Figure 3.19 as a reference for the original form of the SGCM PCs it can be seen that the frequency and amplitude of wave modes b1 and b2 are almost identical to those of PCs 1 and 2.

Figure 5.15 shows the dominant modes b1 and b2 travelling at a rate of about 13 cycles per 60 sols. Wave modes b3 and b4 appear to be similar to wave modes b1 and b2 in terms of their amplitude and waveform, even though wave modes b3 and b4 represent mixed modes, clearly being superimposed principal and subsidiary waves (where the term 'subsidiary' here denotes those waves which travel at a rate of about 4 cycles every 60 sols). This is significant since the original PCs (illustrated in Figure 3.19) show that PCs 1, 2 and PCs 3 and 4 respectively behave like principal and subsidiary waves.

It is of interest to note that wave modes b6 to b10 in Figure 5.15 are more complicated in nature than the leading order modes. These modes propagate at exactly double the frequency of principal modes b1 and b2, travelling at a rate of about 26 cycles/60 sols, and are consistent with the high frequency behaviour of the higher order PCs. For convenience, these oscillatory types from now shall be referred to as '**principal-doubled**' waves, and denote waves which propagate with exactly double the frequency of the principal waves. We also note that the amplitudes of each of the Galerkin modes correspond rather accurately to those of the original PCs.

The MFC modes in Figure 5.14 do not behave like their corresponding PCs as the Galerkin modes have definite frequencies unlike the original data patterns, although they do have very similar amplitudes. Generally, the leading order modes travel at a rate of 13 cycles/60 sols, whereas the higher order modes propagate at double this rate.

The wave solutions of the 20-D KE model illustrated in Figure 5.17 show the two leading wave modes b1 and b2 travelling at a rate of over 13 cycles every 60 sols. Here it can be seen that modes b3 and b4 are mixed modes, having a dominant principal wave with a superimposed subsidiary wave. The distortions in b4 are far more pronounced than those seen in the corresponding correlation mode, where a mixed principal and principal-doubled wave[15] is observed. This demonstrates that the norm which extracts the KE in an optimal way picks out the

[15]The phrase 'mixed principle and principal-doubled wave' for example, refers to a principal wave with a superimposed principal-doubled wave.

subsidiary wave which is visible in the original system, and thus is superior to the correlation model.

It is noted that wave mode b5 is a mixed principal-subsidiary wave, but b6 is more erratic, being a superimposed principal and principal-doubled oscillation. This is of interest because within the original system a similar type of behaviour is observed (see Figure 3.19). The original PC 5 contains subsidiary waves but PC 6 behaves in a far more complicated way, having a much higher frequency. Similar behaviour is also seen in the KE model.

The KE norm model not only reproduces wave modes with comparable frequencies to those observed in the original model, but the amplitudes of the corresponding modes are almost identical. The modes become progressively more distorted and complicated as the higher order modes are approached. Again, this is what is seen in the original system, where PCs 7 to 10 are high frequency, erratic modes with no definite frequencies.

The KE MFC modes (shown in Figure 5.16) behave very similarly to their corresponding wave modes (Figure 5.17). Generally, the leading order modes are mixed principal-subsidiary waves, and the higher order modes contain either mixed principal and principal-doubled or mixed subsidiary and principal-doubled waves. Clearly, the behaviour is different to that of the original MFC PCs (Figure 3.21), as the Galerkin modes have definite wave frequencies (except perhaps for the higher order modes). The amplitudes of the corresponding Galerkin MFC modes and the original PCs are comparable though.

It can then be concluded that the 2-v KE model behaves very well. The generated wave modes are similar to those of the original system, in terms of both their amplitudes and frequencies. There is a noticeable improvement from the correlation 20-D model, particularly in reproducing the subsidiary waves in modes b3 and b4. It can then be concluded that the additional KE which has been maintained in the KE model is essential for the reproduction of these subsidiary waves. This is significant, particularly since we know that baroclinic waves depend upon energy conversion of APE to KE, and maybe a similar process is occurring here, one which is heavily reliant upon a supply of KE from the mean flow.

The differences between the 20-D correlation and KE models sheds light upon what may be occurring in the original system. The KE mixed wave modes b3 and b4 illustrate that there may be a considerable amount of competition between the subsidiary and principal waves. In the correlation model, the principal wave clearly dominates and this may be accounted for by an insufficient amount of retained KE. In the original systems, there exists a strong subsidiary wave in PCs 4 and 5. The appearance of the subsidiary wave may simply depend upon whether the model contains sufficient energy to generate it. The principal wave appears regardless of which norm is used, present in wave modes b1 and b2 for both the correlation and KE models.

By studying the results of the 20-D TE model, the previous conclusions are somewhat substantiated. From Figure 5.19 it is noted that the leading two wave modes b1 and b2 (i.e. the principal waves) are now travelling at a rate of over 14 cycles every 60 sols, which is almost

identical to the frequency at which the SGCM PCs 1 and 2 propagate at. There is now an obvious presence of the subsidiary waves (i.e. waves which travel at a rate of 5 cycles every 60 sols) in wave modes b3, b4, b5 and b6. The additional APE which has been optimised in the TE model has created striking differences to that which was observed in the KE model. It can then be concluded that the optimisation of both KE and APE are essential in reproducing the subsidiary waves of the original PCs, whereas the principal wave of modes b1 and b2 do not seem to be sensitive to changes in these energy types.

The 20-D TE model substantially improves on the 20-D KE model. Not only are the amplitudes and frequencies of the TE wave modes highly comparable with those of the original PCs, but also the decomposition of the subsidiary wave in b6 can be seen. This provides evidence that there exists some degree of interaction between the subsidiary wave and a higher frequency wave (a wave which has been identified as a principal-doubled wave). This is what is observed in the original PCs, where there is a dramatic change in oscillatory structure from PC 5 to 6, the latter being dominated by an erratic, high frequency wave. From the results of the correlation, KE and TE models, there is substantial evidence to argue that the high frequency behaviour of wave mode b6 may simply be due to the presence of a 'sufficient' amount of KE and APE in the reduced system, and may be a result of the interaction of the subsidiary and principal waves. Certainly the amplitudes of the individual TE wave modes correspond accurately to those of the original PCs. Wave distortions in the higher order modes also increases if more modes are retained, indicating that the complicated oscillatory structures which are observed in the original PCs requires a considerable degree of interaction between the leading order and the higher order Galerkin modes if similar behaviour is to be reproduced.

The MFC modes of the 20-D TE model (shown in Figure 5.18) are similar to those found in the 20-D KE model, with the main difference being the strong presence of the subsidiary wave, particularly evident in MFC modes a1 to a4. The TE model has certainly drawn out the subsidiary wave so that it features in almost all of the modes, whereas the principal wave dominates in the correlation model. This illustrates that there exists a fine balance between the principal or the subsidiary wave being the dominant mode of the system, and this dominance depends upon the amount of retained KE and APE. As the original SGCM contains a principal wave in PCs 1 and 2, and a subsidiary wave in PCs 3 and 4, it is concluded that the most accurate reproduction is by the 2-v TE model. This implies that the presence of the subsidiary wave in modes b3 and b4 relies upon the optimisation of both KE and APE.

Summary

Three important oscillatory structures have been observed in the mixed-wave regime, namely a principal ($\sim$13-14 cycles/60 sols) wave, a subsidiary wave ($\sim$4-5 cycles/60 sols) and a principal-doubled wave ($\sim$26-28 cycles/60 sols). It is clear from Figures 5.14 - 5.19 that there is considerable interaction between these wave types, and that this behaviour may shed light on what occurs in the original SGCM. Since obviously the principal-doubled wave is a harmonic of the leading principal wave, it can be seen then that each of the Galerkin mean-flow correction and wave modes are fundamentally composed of different combinations of principal and subsidiary waves. Maybe, the original SGCM behaves similarly, but because of its greater degrees of

freedom, such simple interactions are obscured by the high degree of mixing between these two wave types.

The use of the energy norms in deriving POD modes has resulted in increasing the frequency of the individual modes. This is particularly evident if the TE modes are examined. Interestingly, in the original system, the leading two PCs travel at about 14 cycles every 60 Martian days which is identical to the frequency of the two leading wave modes in the 2-v TE Galerkin model.

As the original system has a principal wave in PCs 1 and 2, and a subsidiary wave in PCs 3 and 4, it is concluded that the most accurate reproduction is by the TE model. This implies that the presence of the subsidiary wave in modes b3 and b4 relies upon both KE and APE being well represented. What is interesting from this conclusion is that it agrees with the work of Selten (1993), who used a model which operated in a regime comparable to the Earth's climate, and argued that only the TE norm could reproduce the original dynamics of this system.

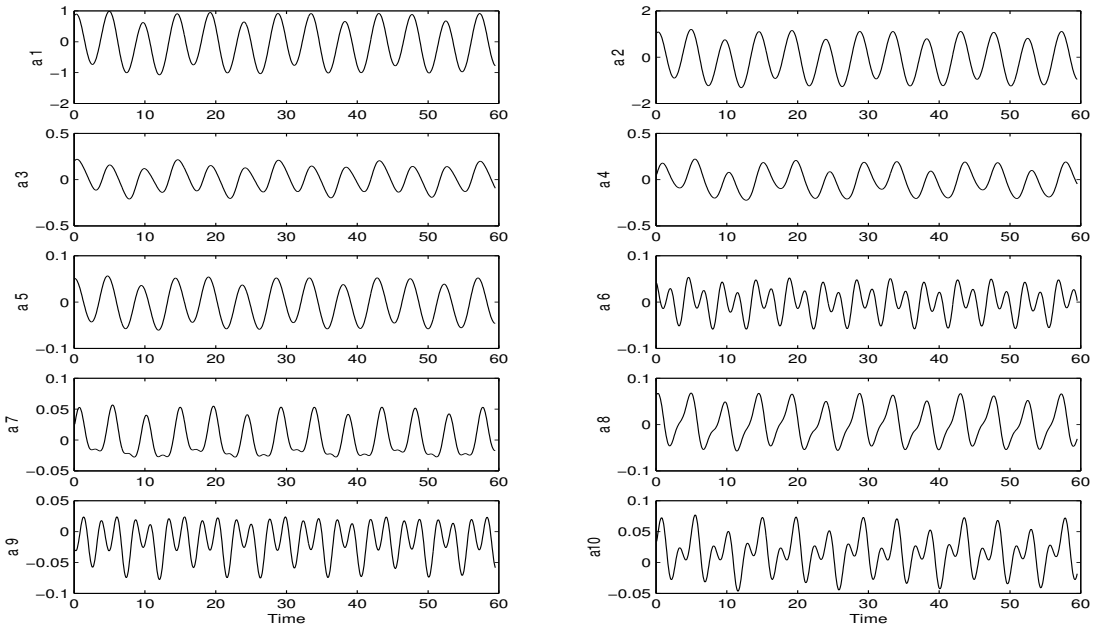


Figure 5.14: 20-D **CORRELATION** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.395) over 60 Martian days.

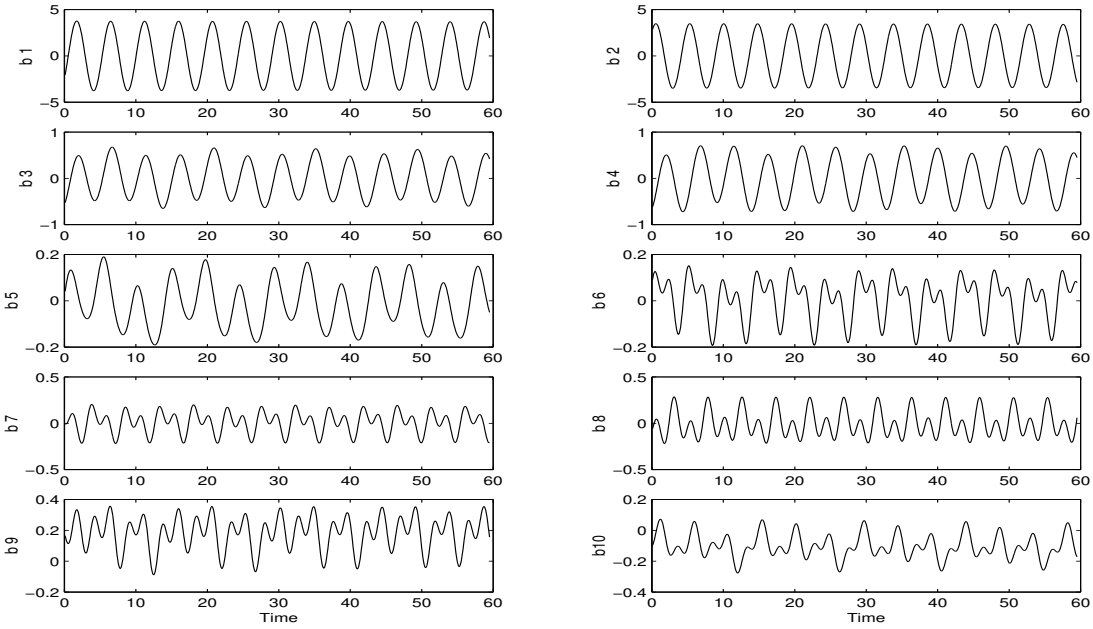


Figure 5.15: 20-D **CORRELATION** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.395) over 60 Martian days.

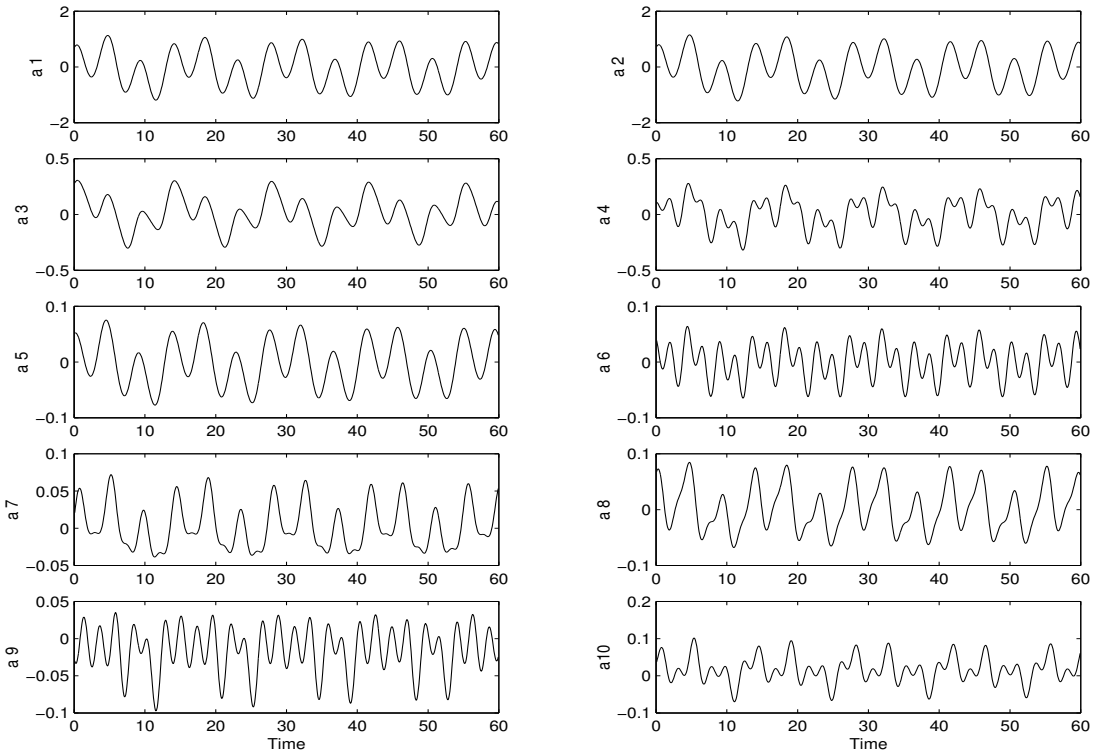


Figure 5.16: 20-D **KE** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.35) over 60 Martian days.

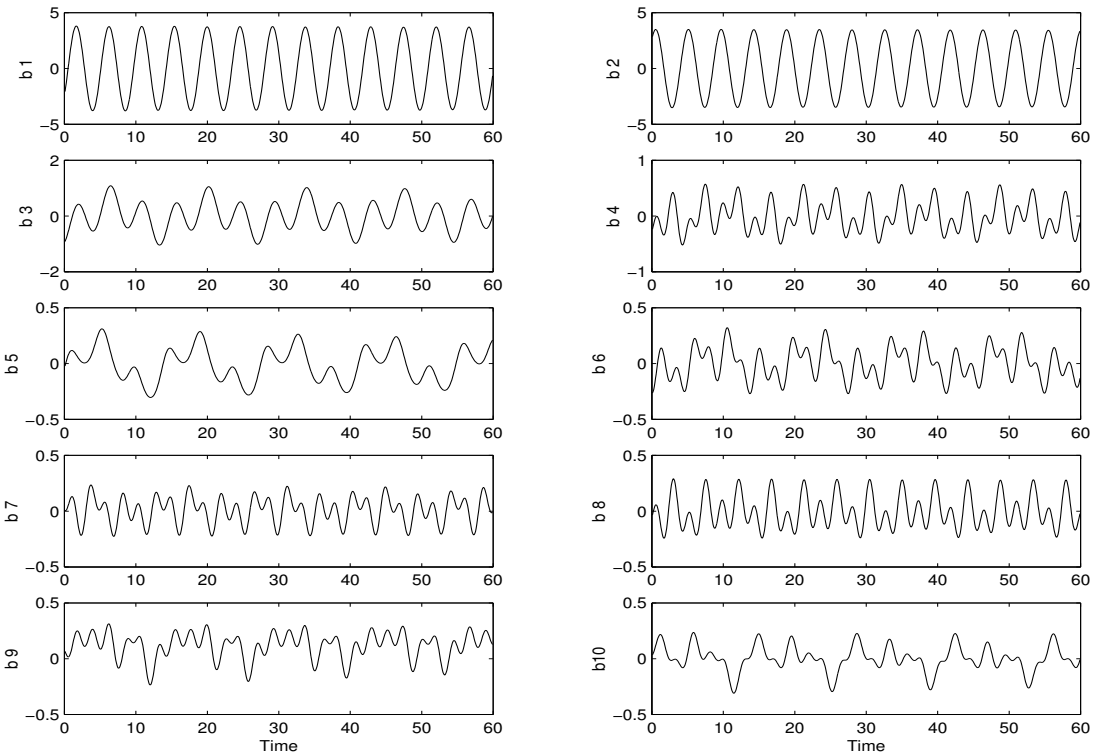


Figure 5.17: 20-D **KE** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.35) over 60 Martian days.

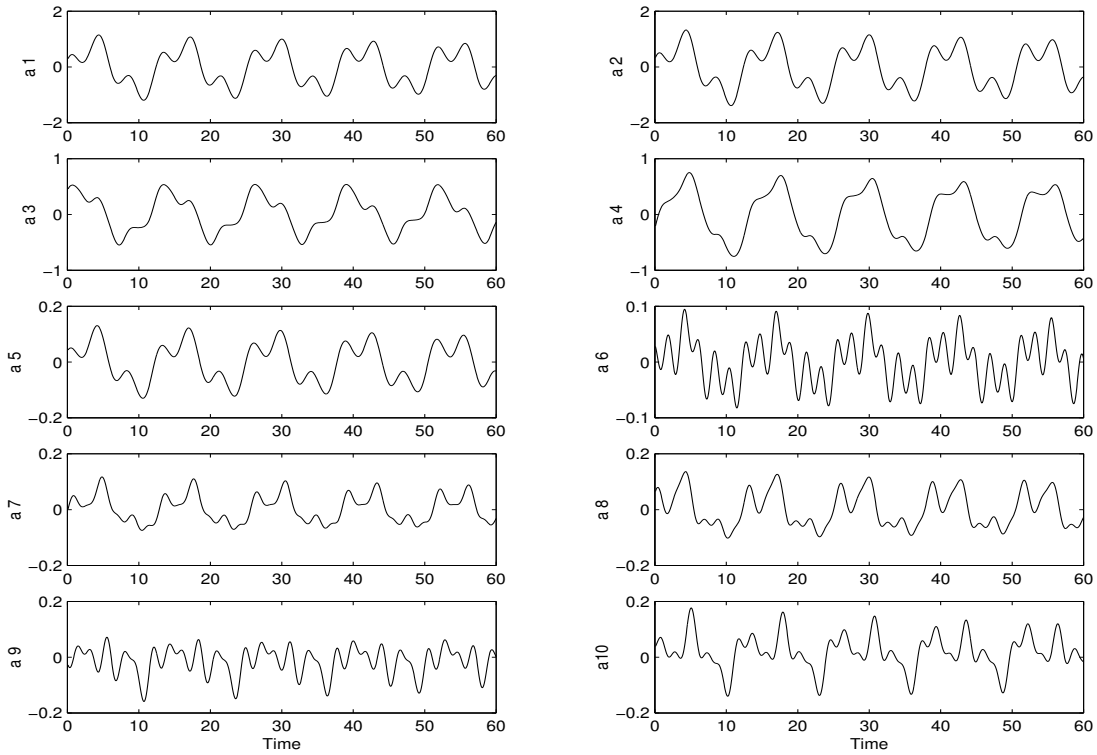


Figure 5.18: 20-D **TE** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.34) over 60 Martian days.

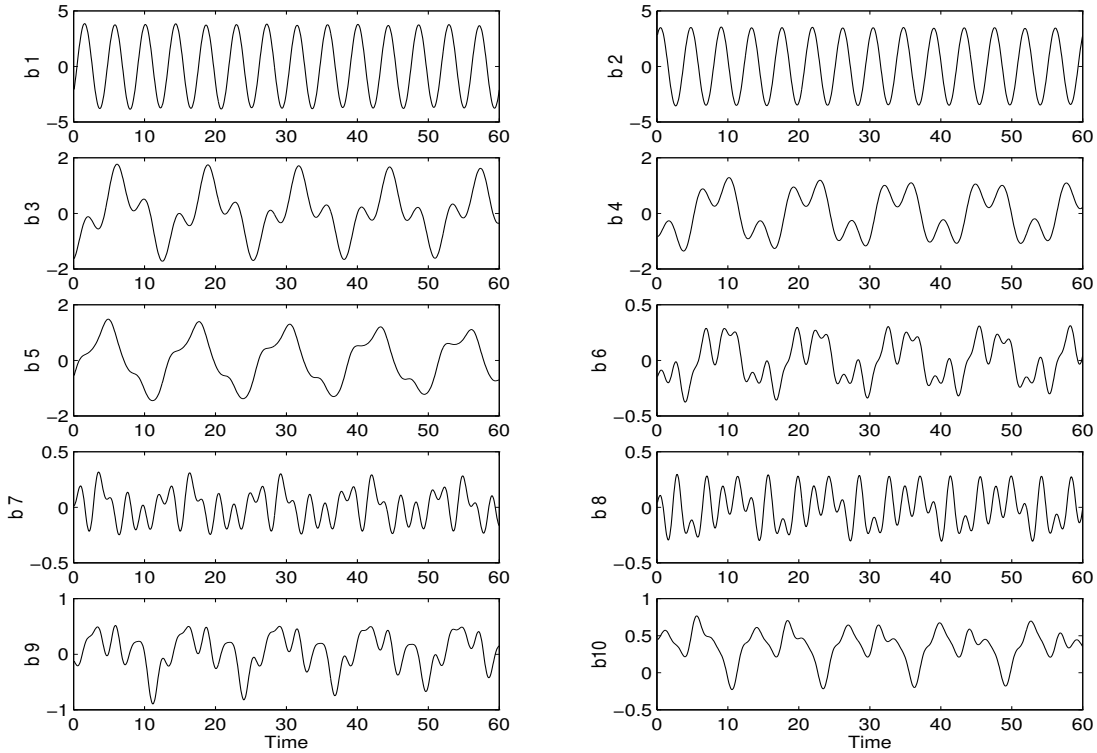


Figure 5.19: 20-D **TE** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.34) over 60 Martian days.

5.8.2 10-D Galerkin models

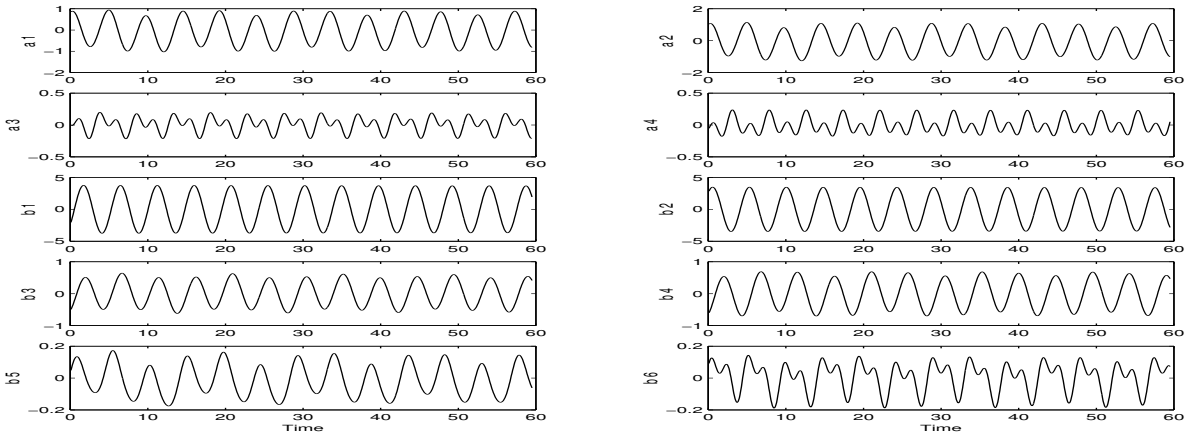
We have investigated how the choice of norm affects the behaviour of the solutions of the models, but we must now address the issue of how the truncation level affects the accuracy of the model behaviour. What we want to consider is the influence various truncation levels have on the accuracy of the solutions obtained from the Galerkin models in moving from a 20-D to a 10-D truncation, and eventually to a 6-D.

Each of Figures 5.20(a), 5.20(b) and 5.20(c) show plots of MFC modes a1 to a4 then wave modes b1 to b6 respectively. Figure 5.20(a) illustrates the 10-D correlation model, in which the corresponding wave modes are almost identical in amplitude, frequency and shape to the 20-D model. However, there is more evidence of a subsidiary wave ($\sim$4 cycles/60 sols) in the 20-D model, particularly in wave modes b3 and b4.

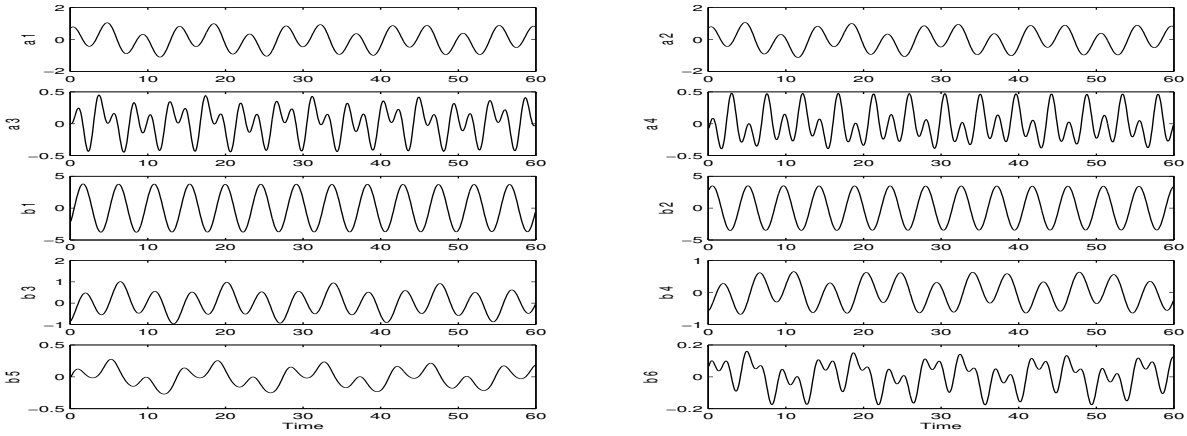
MFC modes a3 and a4 of the 20-D and 10-D models are also different. It appears that MFC modes a3 and a4 in the 10-D correlation model are comparable in behaviour with modes a6 and a10 in the 20-D model. By only retaining four MFC modes the model has changed their ordering. So the truncation size has slightly affected the form of the wave modes as well as the ordering of the mean-flow correction terms.

Figure 5.20(b) illustrates the integration runs for the 10-D KE model. What is immediately obvious when comparing the results of the 20-D KE model in Figure 5.17 with those of Figure 5.20(b) is that the principal-doubled wave in wave mode b4 is actually a mixed principal-subsidiary wave in the 10-D model. Again, a reordering of the MFC modes has occurred due to a change in the truncation level. It seems that MFC modes a3 and a4 in the 10-D model appear as modes a6 and a9 respectively in the 20-D model (shown in Figure 5.16). Again, the reduction in the truncation has affected the quality of the subsidiary wave behaviour as well as the order of the MFC modes.

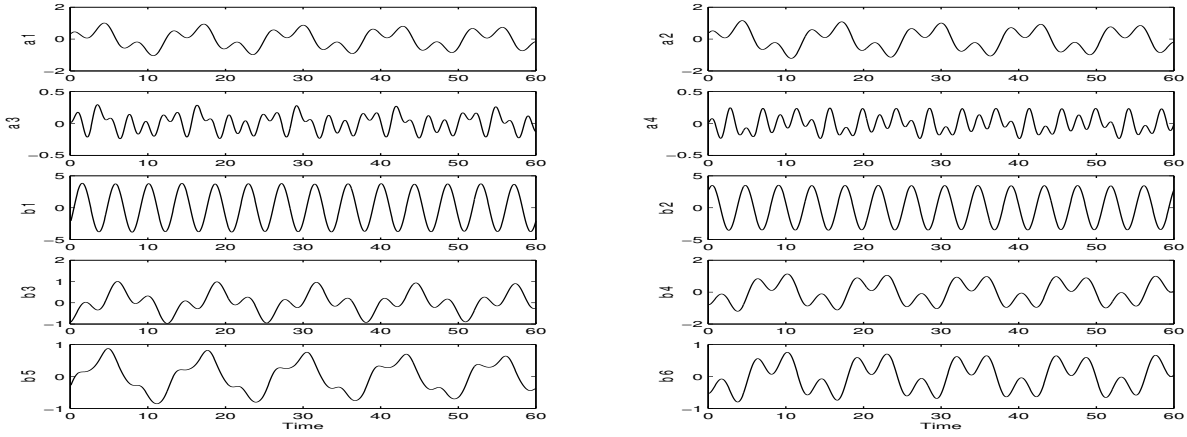
Figure 5.20(c) shows results from the 10-D TE model. The quality of wave mode b6 is severely affected by the truncation. In the 10-D model b6 is no longer distorted, being originally a mixed subsidiary and principal-doubled wave in the 20-D version, but now is simply a mixed principal-subsidiary wave, thus becoming more unlike the original system. The MFC terms have again been reordered, where MFC modes a3 and a4 in the 10-D system are similar to modes a6 and a9 in the 20-D system.



(a) 10-D **CORRELATION** MODEL : Equilibrated amplitudes of the leading four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.395) over 60 Martian days.



(b) 10-D **KE** MODEL : Equilibrated amplitudes of the first four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.35) over 60 Martian days.



(c) 10-D **TE** MODEL : Equilibrated amplitudes of the the first four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.34) over 60 Martian days.

Figure 5.20: Amplitudes of the leading MFC and wave modes in the mixed-wave regime from the 10-D (a) correlation, (b) KE and (c) TE norm models.

5.8.3 6-D models

On comparing the 6-D correlation model in Figure 5.21(a) with the corresponding 20-D model in Figure 5.15, there appears only to be slight differences between the respective modes. The subtle difference is seen in the mixed modes (wave modes b3 and b4), where the presence of the subsidiary wave (∼4 cycles/60 sols) is more evident in the 20-D model.

The 6-D KE model in Figure 5.21(b) has no evidence of modal distortions such as those which appear in b4 in the 20-D version, as it has changed from a mixed principal and principal-doubled wave to simply a mixed principal-subsidiary wave (see Figure 5.17). The truncation reduction has produced smooth wave patterns without the accompanying higher frequency wave structures which are observed in the 20-D model.

The 6-D TE model in Figure 5.21(c) is far more efficient in the reproduction of the original patterns as the principal waves travel at 14 cycles/60 sols and the mixed principal-subsidiary waves (contained in modes b3 and b4) propagate at about 5 cycles/60 sols. However, no high frequency behaviour, which is so evident in the original SGCM, is observed in any of the 6-D models. This seems reasonable since in the original system the higher order PCs are far more erratic than the leading dominant oscillatory structures.

Summary

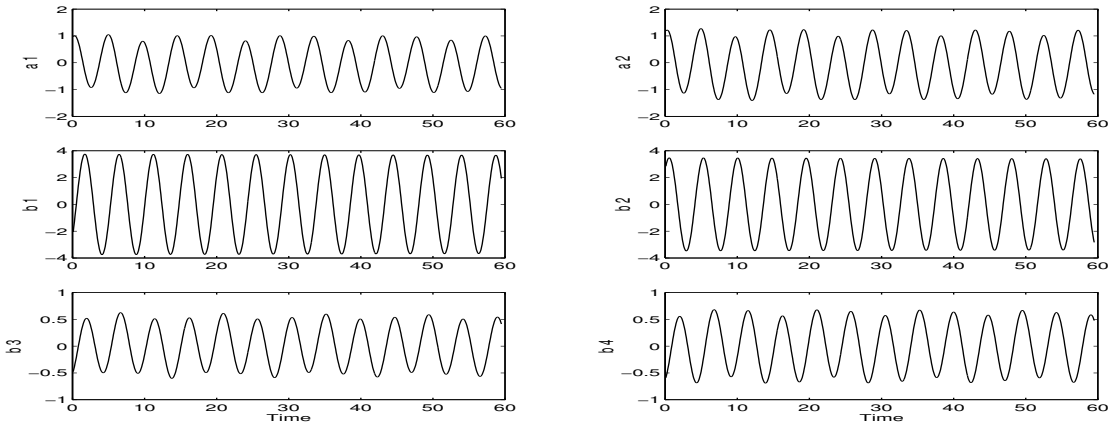
What has been found is that the principal-doubled waves, which were so apparent in the 20-D systems, cease to exist in the 6-D versions. Such high frequency behaviour was also observed in the original SGCM PCs. Thus a severe reduction in the horizontal truncation has resulted in a system being totally governed by two wave types, namely the principal and subsidiary waves. By reducing the horizontal truncation, necessary energy sources which feed the principal-doubled disturbances are lost in the 6-D model, i.e. the severe reduction causes the ODE set to be incapable of generating the high-frequency harmonic.

5.9 The mixed wave regime

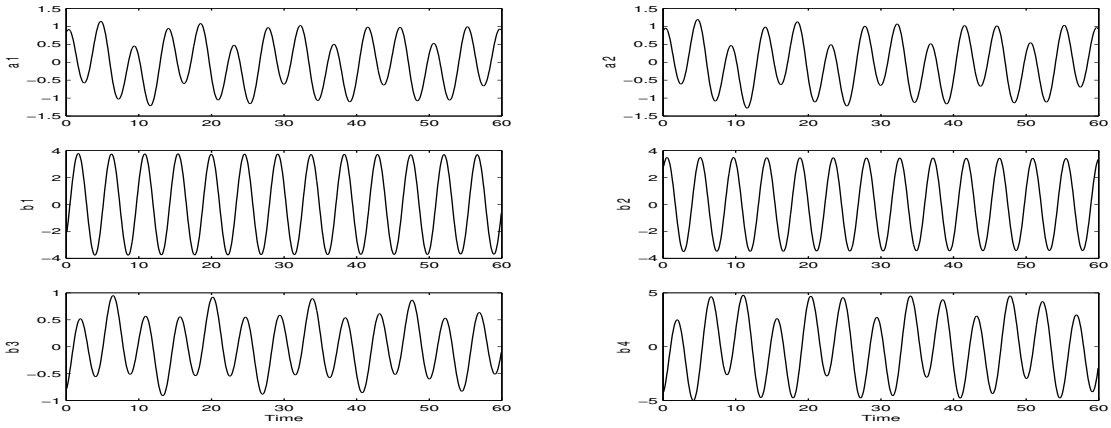
5.9.1 Phase Portraits

Figures 5.22(a), 5.22(b) and 5.22(c) show phase portraits and Poincaré sections (at b1=0) obtained from the 20-D correlation (at Γ =0.395), KE (Γ =0.35) and TE (Γ =0.34) Galerkin models respectively (the 6-D and 10-D models show little difference), over 300 Martian days and after a 200 day transient, using the same Γ values as are used in § 5.8.1.

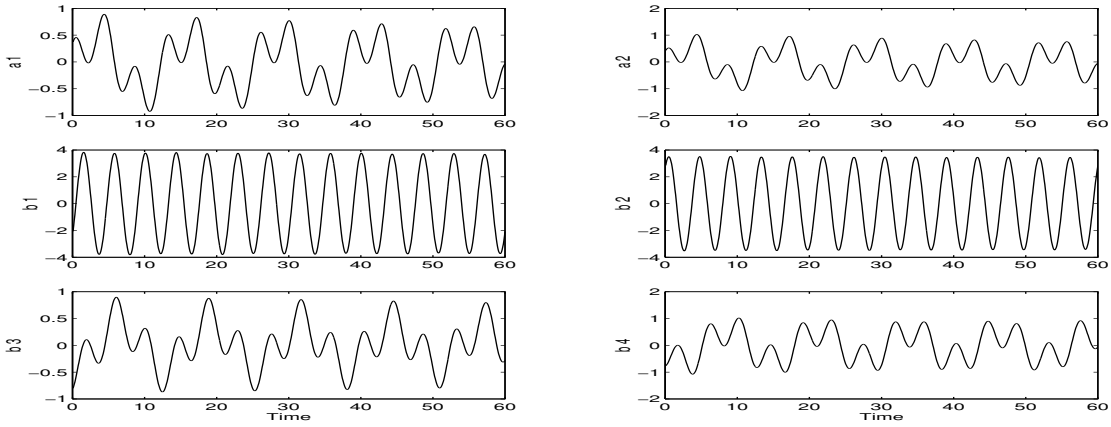
The Galerkin (b1,b2) phase portrait show structures which are close to periodic orbits, and are almost identical to the (PC1,PC2)-space of the original data seen in Figure 5.22. This is evidence that the Galerkin models have been successful in extracting the principal wave of the original SGCM data (analogous to the time evolutionary behaviour of the zonal wavenumber 3 mode). It is also noted that the energy norms have produced a thicker banded periodic structure which is consistent with greater amplitude modulation in the leading two wave modes



(a) 6-D **CORRELATION** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.395) over 60 Martian days.



(b) 6-D **KE** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.35) over 60 Martian days.



(c) 6-D **TE** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.34) over 60 Martian days.

Figure 5.21: Amplitudes of the leading MFC and wave modes in the mixed-wave regime from the 6-D (a) correlation, (b) KE and (c) TE norm models.

b1 and b2. This is particularly true for the TE norm case. This demonstrates that the energy norms try to capture the original wave modulations which occur in the two leading PCs whereas the correlation norm is somewhat less successful in extracting this behaviour. The TE norm does very well in picking this out.

The Poincaré sections in Figures 5.22(a) - 5.22(b) do not reveal a standard limit cycle, nor a period-doubled cycle, but instead a trajectory which has been stretched into a period-doubled 'ribbon'. The Poincaré section in Figure 5.22(c) depicts a period-tripled 'ribbon', illustrating the subtle differences in the modal behaviour of the TE norm modes. Unlike the Poincaré sections of the original system in Figure 5.13, the trajectories do not cover the entire interior of the surface (describing a limit cycle which has been stretched into a flat ribbon), but the Galerkin Poincaré sections (Figure 5.22) reveal additional structure, providing evidence that there is a phase relationship between Galerkin wave modes b2 and b3. Indeed, the SGCM is low dimensional if this is revealing that the original PCs are in fact related in a similar way.

Summary

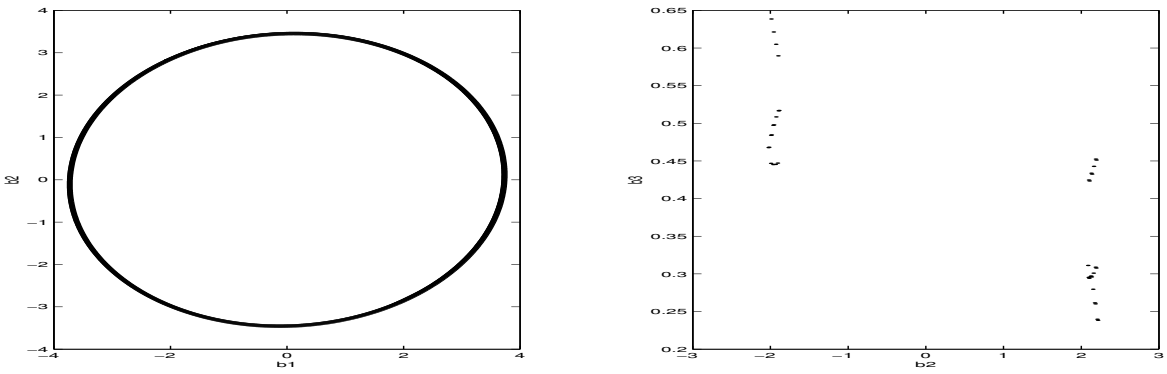
Our results are encouraging as the attractor reconstruction using the two leading POD modes (b1 and b2) is almost identical to that found by Collins and James (1995), and what was shown in Figures 5.13(a) - 5.13(c) from the original SGCM data.

It must be appreciated that the Galerkin models do not produce structures which are identical to those observed in the original model. This difference is due to greater variability in amplitude modulation in the two leading PCs in the original model. But even though the Galerkin models are not reproducing the highly complicated distortions seen in the SGCM flow, it is believed that they are still modelling the most significant underlying dynamics of the original system. This is why the Galerkin phase portraits are relatively simple structures as they are picking out the most important dynamical patterns.

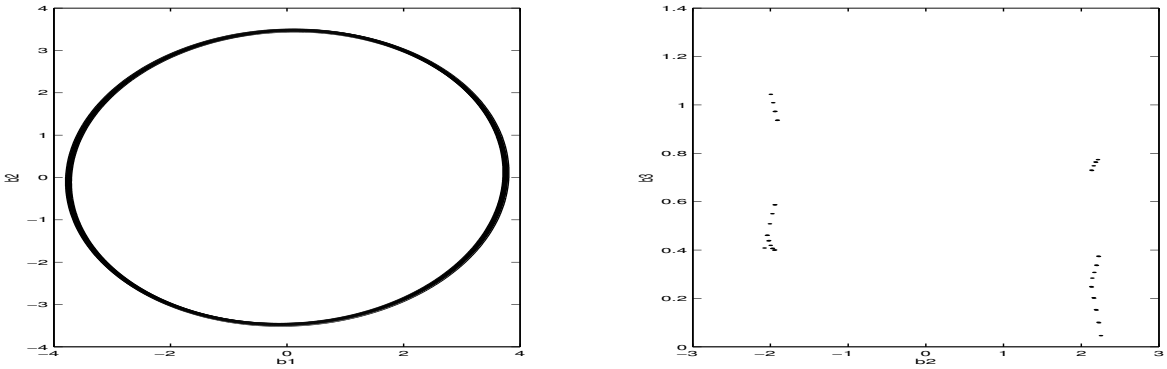
However, the Galerkin Poincaré sections (Figure 5.22) are highly comparable with those of the original SGCM (Figure 5.13), but instead of a stretched, flat ribbon structure which is observed in the original system, ribbons which are period-doubled and period-tripled are found. This suggests that there might exist a possible phase relationship between the two leading SGCM PC conjugate pairs. If this happens to be the case, then there would be a considerable amount of evidence to support the hypothesis that the original SGCM is in fact low-dimensional.

5.9.2 Power spectra analysis

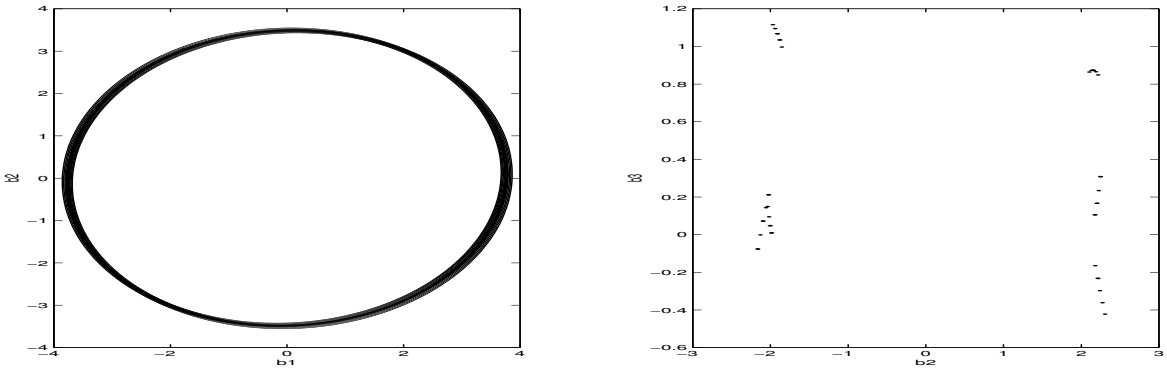
Figure 5.23 shows the power spectra profiles for the mixed wave regime. The top two diagrams ((a),(b)) show MFC modes a1 and a2 where the dominant frequencies are clustered about the 0.08-0.1 and 0.23-0.25 cycles/sol ranges. The middle pictures ((c),(d)) show the power spectra of leading wave modes b1 and b2. Here it can be seen that there is only one obvious frequency which exists and this centres around the 0.23-0.25 cycles/sol range. Finally, the lower plots ((e),(f)) show the power spectra of wave modes b3 and b4 which are very similar to those of the MFC plots.



(a) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **CORRELATION** MODEL AT $\Gamma = 0.395$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for over 80 % of the wave energy in the system.



(b) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **KE** MODEL AT $\Gamma = 0.35$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for 80 % of the wave energy in the system.



(c) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **TE** MODEL AT $\Gamma = 0.34$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for nearly 80 % of the wave energy in the system.

Figure 5.22: Phase space reconstructions and Poincaré sections (at b1=0) using the 20-D (a) correlation, (b) KE and (c) TE norm models. A structure which is close to a periodic orbit is obtained in (b1,b2)-space and becomes thicker if the energy norms are used (note the thicker banded orbit if the TE norm is used).

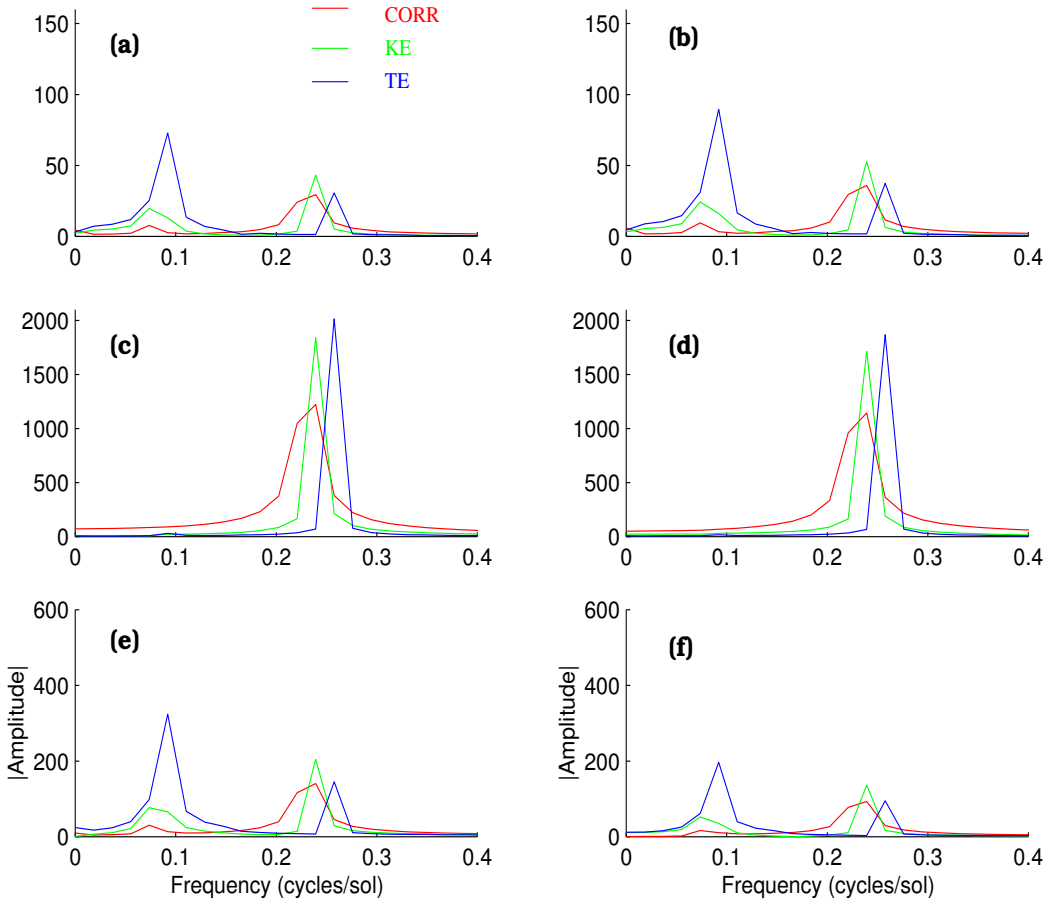


Figure 5.23: POWER SPECTRA PROFILES IN THE MIXED-WAVE REGIME. The figure shows results from the 10-D correlation (red), KE (green) and TE (blue) norm models of the first two MFCs a1 and a2 in (a) and (b); the wave modes b1 and b2 in (c) and (d); and wave modes b3 and b4 in (e) and (f). Except for the b1 and b2 profiles, there are clearly two dominant frequencies associated with each of the modes. This illustrates that the mixed-waves are in fact superimposed principal and subsidiary waves as the frequency peaks gather about the (0.08,1) and (0.23,0.26) cycles/sol frequency bands.

The frequencies of the mixed-wave regime compare very well with the frequencies extracted from the original system, shown in Figure 3.20(a). The original SGCM PCs cluster about the 0.23-0.25 cycles/sol frequency band which is what is seen in Figure 5.23. The most revealing difference between the frequencies of the low-dimensional modes and that observed in the original system is seen in wave modes b3 and b4. In the original system these wave modes have one dominant frequency value, distributed around 0.1 cycles/sol. This same frequency (associated with the subsidiary wave) is also extracted in the low-dimensional model, as well as an additional frequency (the principal wave) of about 0.23 cycles/sol. This extra frequency component is not apparent in the original model. This is apparently an artifact of the mixed-wave regime as the model has produced two important frequency values for all the modes except in the two leading wave modes, b1 and b2.

The effect of using the energy norms can be seen in Figure 5.23 as they have slightly displaced the frequencies towards higher frequency values as compared with those produced by the correlation norm. It can be concluded here that the various norms have produced different spatial

POD structures, which has resulted in increasing the frequency rate of the Galerkin modes. It should also be noted that this upwards shifting of the frequencies has brought the frequency values more into line with the values found in the original system.

It can be seen that for wave modes b3 and b4 of the correlation norm model a very small frequency peak occurs at about 0.8 cycles/sol as compared with a significantly larger peak at 0.23 cycles/sol. This indicates that the principal wave pattern dominates the flow (seen in Figure 5.15). The power spectra obtained from the KE norm produce a dominant peak at 0.23 cycles/sol as well as a much larger peak at 0.08 cycles/sol than was seen in the correlation norm model. This illustrates that the KE norm produces a dominant principal wave as well as an accompanying subsidiary wave.

If the TE norm is used, the largest frequency peak is at about 0.09 cycles/sol in wave modes b3 and b4. This implies that the TE norm produces the subsidiary wave as the dominant wave type. It is also noticed that the peak which corresponds to the principal wave is actually smaller than that produced by the KE norm.

Summary

Even though the original SGCM PCs produced by the different norms show very little difference in behaviour and structure, the behaviour of the Galerkin modes are clearly affected by the choice of norm. The differences found in using the energy norms in the POD modes are seen in the different power spectra profiles.

The frequencies of the low-dimensional modes in the mixed-wave regime correspond very well with those of the original system. This is particularly true for the case of wave modes b1 and b2. Even though the power spectra reveal an additional important frequency in wave modes b3 and b4, it still picks out the importance of the subsidiary wave.

The use of the energy norm has suggested that the two-frequency mode (seen in b3 and b4) which exists in the mixed-wave regime is capable of changing to a single-wave frequency mode, i.e. the mixed principal-subsidiary oscillatory structures can be changed to a single subsidiary wave if the energy norms are used. If this was shown to be the case then the energy norms would each be capable of extracting the first two wave pairs respectively as principal and subsidiary waves, and thus be reproducing the behaviour of the leading four SGCM PCs.

Chapter 6 will consider models which have retained a greater amount of vertical energy. It will be shown that the accuracy of the individual models improves if four vertical modes are used instead of two (and hence retaining more than 99% of KE and 97% APE in the vertical direction).

5.10 Conclusions

• It has been shown that a 6-D TE Galerkin model is capable of reproducing the amplitude and frequency of the leading two SGCM PCs (when operating in a mixed-wave regime), i.e. in the reproduction of oscillatory structures which travel at a rate of about 14 cycles/60 sols. Of all the low-order models considered in this section the 20-D TE model was found to be superior, as it was capable of producing both 'principal' and 'subsidiary' waves (waves which respectively travel at a rate of about 14 and 5 cycles/ 60 sols). The accuracy of the TE norm models in reproducing the leading wave structures demonstrated the importance of the barotropic and baroclinic energy conversions which occur within the original system since the TE norm is the only norm capable of simulating energy conversion processes (as it gives sufficient weight to the baroclinic fields). Our conclusions agree with the work of Selten (1993) who argued that the use of the TE norm was essential in reproducing the global structure of the attractor and hence in reproducing the original dynamics of the system.

• Each model produced principal waves in the leading two wave modes (i.e. an oscillation of about 14 cycles/60 sols), and was thus capable of reproducing the temporal behaviour of the spatial zonal wavenumber 3 mode. However, the subsidiary wave structures ($\sim$5 cycles/60 sols waves) were only observed in the KE and TE norm models. The presence of the subsidiary wave (in the mixed-wave regime) greatly depended upon which norm was used, and was only observed to be a dominant oscillatory structure if the TE norm was selected. The result of selecting the standard correlation (Euclidean) norm in the modelling process was that the principal wave governed the mixed-wave regime. Therefore, the appearance of both a principal and subsidiary wave in the mixed-wave regime demands the use of an energy norm.

• The amplitudes and frequencies of the Galerkin modes were shown to be robust to the truncation level (although the quality of the extracted subsidiary wave was affected). The robustness may possibly be due to few modes competing for dominance in the physical regime of the SGCM, i.e. the original system is intrinsically low-dimensional. However, it was shown that the form of the oscillations were affected by the horizontal truncation level, particularly in terms of the modulation and the erratic disturbances of the individual oscillations. The small distortions which are present in the original PCs were similar to oscillatory structures observed in the 20-D systems. These wave types were referred to as 'principal-doubled' waves, and were only evident for a sufficiently high level of truncation.

• Quasi-periodic flow observed in the low-order models has been shown to be due to the competition of two oscillatory structures, fighting for dominance, namely the 'principal' and 'subsidiary' waves. Regimes where either of these waves dominated the flow were observed in the principal and subsidiary wave regimes, where all the modes travelled at one frequency. The mixed-wave regime contained both waves types. However, by using a low level of truncation and a non-energy norm it was found that the principal wave dominated the quasi-periodic flow. It has been shown that the occurrence of the various regimes were highly sensitive to dissipation changes and can be explained by the close proximity of the two Hopf bifurcations.

Chapter 6

Extended 4-v Galerkin models

Chapter 5 analysed models which were constructed from the SGCM data. These generated low-dimensional ODE sets used only two vertical modes (the barotropic and first baroclinic) for their vertical formulation (i.e. '2-v' models) and consequently could capture only at most 48.1% of the APE of the system (shown in § 2.3.5). The reason why only two vertical modes were initially considered is because such a system is highly comparable with a classical two-level model. By including an additional two vertical modes (the new '4-v case') and making comparisons with the results of Chapter 5, an understanding of how the improved vertical resolution scheme might affect the accuracy of the models can be made, besides determining whether a two-level scheme has the required vertical resolution to accurately reproduce the dynamics observed in the original system.

In this Chapter we will use four vertical modes to model the vertical activity and hence capture 99% of KE and 97% of APE in the vertical direction (see § 2.3.5). It is now necessary to examine whether the accuracy of the modal behaviour which exists in the mixed-mode regime can be improved upon. Comparisons will be made with the findings of Chapter 5 and the question of whether the additional vertical modes improves the accuracy of our modelling scheme will be addressed.

The model truncations considered in this Chapter will again be the (10,10), (4,6) and (2,4) truncations for the correlation, KE and TE norm models which were explained in § 5.7. For each of the models, a combination of numerical integration experiments and bifurcation continuation calculations will be used to search for travelling wave solutions.

6.1 Linear Stability method

As was done in § 5.3 the coupling matrix ($\mathcal{C}$) and the matrix for the linear problem ($\mathcal{L}$) were calculated in order to formulate the Jacobian matrix ($\mathcal{J}$) of the linear system. This was computed explicitly using Mathematica, in which dissipation (Γ) was used as a free parameter. The locus of critical Froude number points were then calculated (when the real part of any eigenvalue of the Jacobian passes through zero). This allowed the construction of the various linear stability curves (for more details refer to § 5.3.2).

6.1.1 4-v linear stability results

The different POD modes and their associated zonal wavenumbers for the 4-v case are the same as for the 2-v, shown in Table 5.1. The results are similar to those which were found in § 5.3.2. However, differences are found in the growth rates of the leading stability curves.

For all mode pairs there are relative increases in the critical Froude number as Γ increases (Figure 6.1). Again, the mode which dominates the flow, i.e. the wavenumber 3, is the most unstable at the inviscid limit but is the most stable for $\mathcal{O}(1)$ dissipation. It is noted that there are a few differences between the various stability curves. The ordering of the linear stability profiles is again according to the amount of energy contained in the mode pairs (except for the third pair (mode 5, mode 6) in the energy models). It can be concluded that the use of the energy norms have the effect of stabilising the modes (seen in the vertical shifts towards higher critical Froude values).

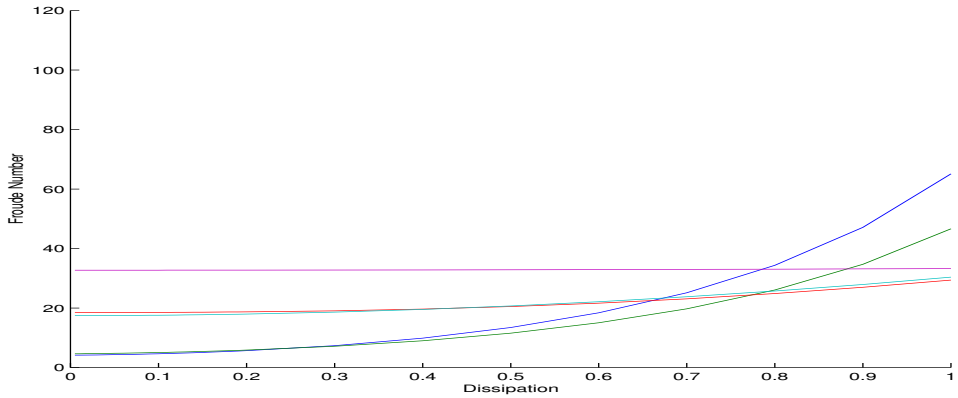
Comparing Figure 6.1 with Figure 5.2 it is evident that the additional energy retained in the 4-v models has also had the effect of moving the stability curves upwards. The change in the vertical formulation has had the greatest effect on the leading two wave pairs in the KE and TE models where the critical Froude numbers rapidly increase as Γ increases, rising in excess of $F = 100$ for $\mathcal{O}(1)$ dissipation. This suggests that the stability of the energetic two POD mode pairs is highly sensitive to the level of vertical truncation. The steep gradients of the energy curves for (a) the first and (b) the second mode pairs are illustrated in Figure 6.2. Clearly, the TE norm modes are the most stable at the inviscid limit and also increase at the most rapid rate as Γ increases.

MODELS	$\mathbf{F}_c$	Γ	MODELS	$\mathbf{F}_c$	Γ
4-v CORR	6.501	0.2485	**2-v CORR**	5.311	0.2824
4-v KE	7.994	0.2341	**2-v KE**	5.011	0.2491
4-v TE	7.989	0.2322	**2-v TE**	4.487	0.1902

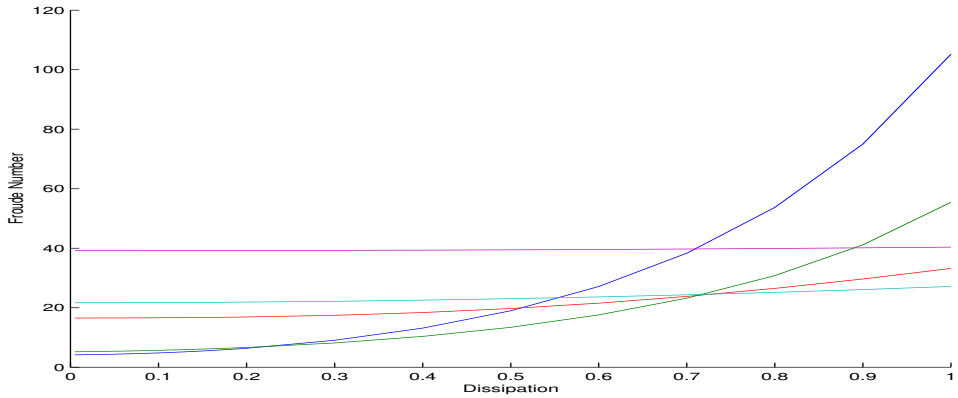
Table 6.1: Double Hopf locations in the different Galerkin models.

The first double Hopf bifurcation point DHB_1 (defined at the intersection of the linear stability curves of the first two wavenumber pairs) lies on the lines F =7.994 and 7.989 in the 4-v KE and TE models respectively (see Table 6.1). Recalling that the Hopf bifurcations HB_1 and HB_2 sit on the line $F = 8.14$, it is then apparent that the double Hopf may have a significant role in the dynamical behaviour of the low-dimensional ODE sets (see Figure 6.3). Similarly, from Table 6.1 it can be seen that the Froude numbers at which the double Hopfs occur in the 4-v models are considerably closer to F =8.14 than their 2-v counterparts. Therefore the presence of the double Hopf may result in significant nonlinear effects, and thus may have a substantial role in the dynamical behaviour of the 4-v Galerkin modes.

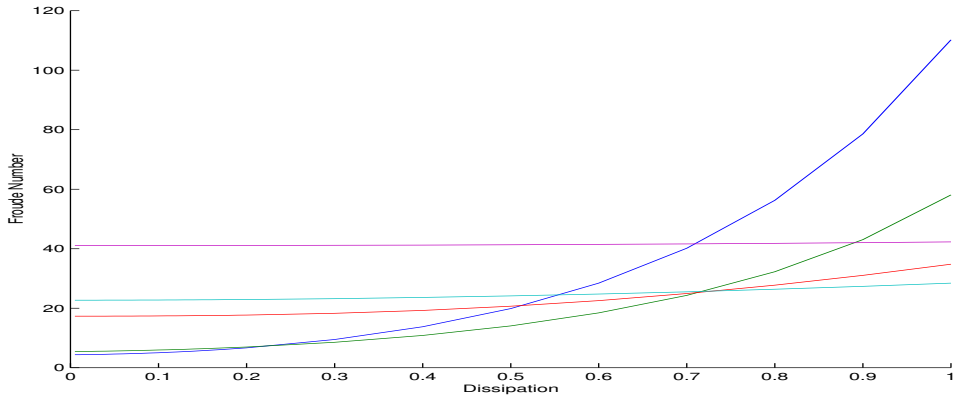
Since the double Hopf point DHB_1 lies in close proximity to the Hopf bifurcations HB_1 and HB_2 of the 4-v energy models, it may be found that DHB_1 plays a particularly important role in determining the dynamical behaviour of the energy low-order ODE sets. The nonlinear



(a) CORRELATION 4-v GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 5, mode 6) in light blue, (mode 7, mode 8) in red and (mode 9, mode 10) in purple.

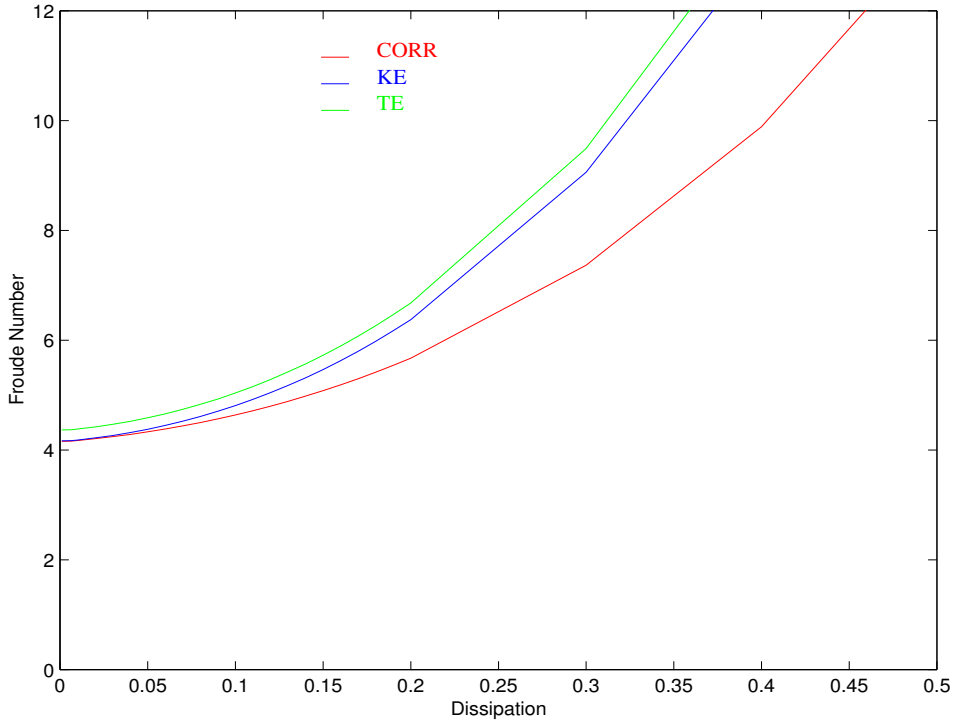


(b) KE 4-v GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 7, mode 8) in red, (mode 5, mode 6) in light blue and (mode 9, mode 10) in purple.

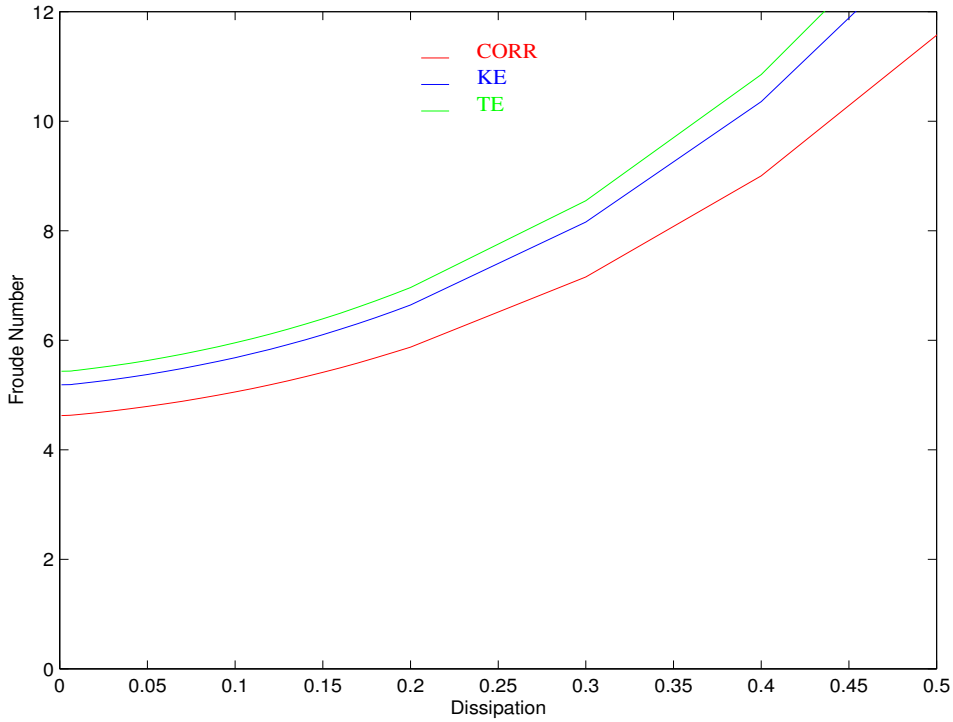


(c) TE 4-v GALERKIN MODEL : Individual linear stability curves for mode pairs (mode 1, mode 2) in navy, (mode 3, mode 4) in green, (mode 7, mode 8) in red, (mode 5, mode 6) in light blue and (mode 9, mode 10) in purple.

Figure 6.1: Linear stability curves for the (a) correlation, (b) KE and (c) TE norm models. The dominant energy mode pair (mode 1, mode 2) is the most unstable at the inviscid limit, whilst being the most stable at $\mathcal{O}(1)$ dissipation. Note that the curves have been displaced vertically for the energy norm models.



(a) Linear stability profiles for the first wavenumber pair i.e. (mode 1, mode 2) for the correlation, KE and TE norm models.



(b) Linear stability profiles for the second wavenumber pair i.e. (mode 3, mode 4) for the correlation, KE and TE norm models.

Figure 6.2: Linear stability curves for (a) the first wavenumber pair, (b) the second wavenumber pair for the correlation, KE and TE norm models. Note that as $\mathcal{O}(1)$ dissipation is approached the individual stability curves produced by the energy norms become somewhat steeper. In effect, the energy norms have stabilised the system.

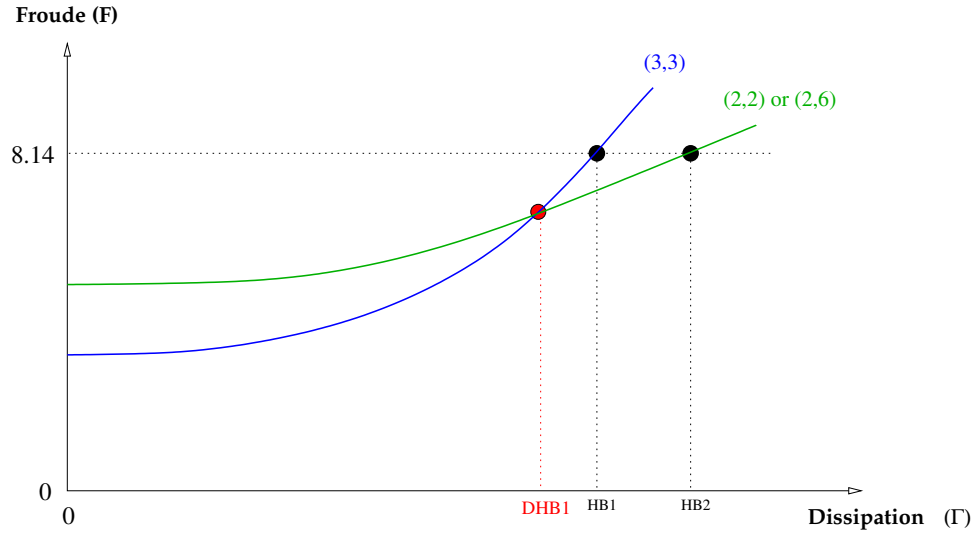


Figure 6.3: Linear stability curves for the first and second wavenumber pairs. The Hopf bifurcation points HB_1 and HB_2 appear at the intersections of the line F =8.14 with the stability curves for the first and second wavenumber pairs respectively. The double Hopf bifurcation DHB_1 occurs at the intersection of these two stability curves.

effects of the double Hopf may be found to be instrumental in reproducing dynamical behaviour similar to that which is observed in the original SGCM. Concerning its possible influential role in the energy models, it will be seen in this section that the energy ODE sets produce modes which are far more complicated and erratic (i.e. nonlinear) than those which are found from using the standard correlation norm. This might be indicating that the double Hopf bifurcation point is in fact dynamically influential. A thorough discussion of the role of the double Hopf will be made in § 6.5.

6.2 Assessment of bifurcation structures

Again the use of AUTO was essential in identifying the existence and positions of the Hopf bifurcations, and the resulting periodic orbits. AUTO was run initially from a high dissipation value (Γ=1.0) in a stable state and by decreasing the principal bifurcation parameter the bifurcation points were identified.

In all AUTO runs only two Hopf bifurcations for the trivial solution were found (taking F =8.14). By comparing the locations of the Hopf bifurcations with those of Chapter 5 (shown in § 5.6) it can be seen that there has been a movement in their respective positions. This suggests that the positioning of the Hopf points (influenced by the arrangement of the linear stability curves) are affected by the quality of the model's vertical resolution.

AUTO computed the imaginary part of the eigenvalues at the two Hopf bifurcations (HB_1 and HB_2). As the differences between the 6-D, 10-D and 20-D models for each of the norms were minimal, only the 10-D model is shown for the correlation, KE and TE norms in Table 6.3. Table 6.3 shows the periodicity of the wave patterns produced by the different Galerkin models, arising from the two Hopf bifurcations at F =8.14. Interestingly, if the frequency values of Table 6.3 are compared with the 2-v cases in Table 5.5 it can be seen that the corresponding

NORM	6-D HB_1	6-D HB_2	10-D HB_1	10-D HB_2	20-D HB_1	20-D HB_2
correlation	0.3211	0.3498	0.3213	0.3503	0.3218	0.3507
KE	0.2653	0.2983	0.2657	0.2986	0.2659	0.2989
TE	0.2563	0.2820	0.2564	0.2822	0.2568	0.2827

Table 6.2: The location of the Hopf bifurcation points on the dissipation axis at constant F(=8.14) for the 4-v 6-D, 10-D and 20-D models.

NORM	$\mathbf{Im(\lambda_1)}$	**Period** τ_1 (sols)	$\mathbf{Im(\lambda_2)}$	**Period** τ_2 (sols)
correlation	0.2487	25.26	0.0893	70.36
KE	0.2511	25.02	0.0901	69.70
TE	0.2628	23.91	0.0936	67.16

Table 6.3: Two periodic orbits produced from the Hopf bifurcation points HB_1 and HB_2 in the various 4-v 10-D models.

entries in the latter are smaller. This implies that the additional vertical energy retained (by using four vertical modes) has consistently resulted in an increased frequency for each of the travelling modes. Again, this demonstrates that modal behaviour is at least somewhat sensitive to energy input or modal truncation.

6.2.1 Comparisons between the 2-v and 4-v AUTO runs

The AUTO runs of the 4-v models produced almost identical structures to those found in Figure 5.2, and so are not shown here. However, it was found that the:

- L2-norm values are greater in the 4-v model;
- Frequencies are greater in the 4-v version.

The inclusion of two additional baroclinic modes (effectively increasing the retained TE in the model) has caused the frequencies and amplitudes of the leading principal and subsidiary waves to increase. This is evident if comparisons are made with the 2-v models.

6.2.2 Stability properties of the periodic orbits

The Hopf bifurcations at HB_2 were found to be initially stable i.e. the Hopf bifurcations were *supercritical*. Tracing their continuation with dissipation as a control parameter, the results were similar to those of the 2-v models where the amplitudes of the periodic orbits which arose from HB_2 always grew with decreasing dissipation, while the period remained almost constant.

The unstable orbits which arose from the *subcritical* Hopf bifurcation HB_1 also grew in amplitude with decreasing dissipation and the unstable branch finally stabilises at the point of stability exchange SC_1. The solution curves of the AUTO runs were almost identical to those found for the 2-v models, shown in Figure 5.7.

The locations of the various Hopf bifurcations and stability exchanges for the numerous 10-D 4-v models are shown in Table 6.4. This indicates that the regions of stability exchange are similar to those which were observed in the 2-v models. The mixed-wave regime still exists in

TE	KE	correlation
HB_1=0.2564	HB_1=0.2657	HB_1=0.3213
HB_2=0.2822	HB_2=0.2986	HB_2=0.3503
SC_1=0.1714	SC_1=0.1812	SC_1=0.2402
SC_2=0.2212	SC_2=0.2119	SC_2=0.2813

Table 6.4: The location of the various Hopf bifurcations (HB_1 and HB_2) as well as points of stability exchange (SC_1 and SC_2) for the 4-v 10-D correlation, KE and TE Galerkin models.

between the subsidiary (to its right) and principal wave regimes (to its left) and the dissipation ranges for the various regimes of the 2-v and 4-v models are summarised in Table 6.5.

MODEL	**SUBSIDIARY** **(SC_2, HB_2)**	**MIXED** **(SC_1, SC_2)**	**PRINCIPAL** **(0, SC_1)**
2-v corr	(0.4125, 0.4668)	(0.3571, 0.4125)	(0.0, 0.3571)
2-v KE	(0.3892, 0.4621)	(0.3439, 0.3892)	(0.0, 0.3439)
2-v TE	(0.3213, 0.4507)	(0.2973, 0.3213)	(0.0, 0.2973)
4-v corr	(0.2813, 0.3503)	(0.2402, 0.2813)	(0.0, 0.2402)
4-v KE	(0.2119, 0.2986)	(0.1812, 0.2119)	(0.0, 0.1812)
4-v TE	(0.2212, 0.2822)	(0.1714, 0.2212)	(0.0, 0.1714)

Table 6.5: The location of stability exchanges for the 10-D correlation, KE and TE models.

By comparing Tables 6.4 and 5.6 and observing the entries in Table 6.5 it can be seen that the positions of the Hopf bifurcations are sensitive to both the norm and the level of vertical truncation. The energy norms shift both Hopf points HB_1 and HB_2 towards smaller dissipation values. The same is found if four vertical modes are retained instead of two. This illustrates that both the energy norms (particularly the TE norm) and the improved vertical scheme have stabilised the 10-D Galerkin systems.

6.3 Galerkin model comparisons

The following results now presented are from integrations of the 6-D, 10-D and 20-D models, using the correlation, KE and TE norms for given dissipation values and for a fixed Froude number (=8.14). The terms 'principal' and 'subsidiary' as well as 'MFC' and 'wave mode' are explained in § 5.7.

6.3.1 20-D models

Each of the following integrations are within the mixed mode range of [SC_1, SC_2], at dissipation values $\Gamma = 0.28$, 0.21 and 0.22 for the correlation, KE and TE norm models respectively, over 60 Martian days after a transient of 200 days. It should be noted that the leading wave pairs of the various Galerkin models are comparable in amplitude to the original SGCM principal components at these selected values of Γ.

Comparing the wave modes in the 4-v and 2-v versions of the 20-D correlation model (Figure 6.5 with Figure 5.15), it is noticeable that the inclusion of the two additional vertical modes (baroclinic modes 3 and 4) has resulted in a cleaner extraction of the subsidiary wave

(i.e. waves which travel at about 4 cycles every 60 sols), which is particularly obvious when comparisons are made with the corresponding modes b3 and b4. This is consistent with our previous conclusions when it was stated that the mixed modes depended upon the presence of a sufficient amount of KE and APE. The APE in the 4-v 20-D model is somewhat more than that which was retained in the 2-v version (97.1 % compared with 48.1 %) and a far more evident subsidiary wave in wave modes b3 and b4 is a result.

The actual form of the wave and MFC modes are almost identical when comparisons are made between the two models. The frequency of the principal wave ($\sim$13 cycles every 60 sols) remains unchanged.

By comparing the 4-v 20-D KE model with its 2-v counterpart (Figure 6.7 with Figure 5.17), the existence of the subsidiary wave ($\sim$4 cycles every 60 sols) is more evident in the latter. This is evident when comparing modes b4, b5 and b6 between the models. It is observed that the frequencies of the modes are only slightly affected by improving the vertical resolution.

When comparing the 4-v 20-D TE model with its corresponding 2-v version (Figure 6.9 with Figure 5.19), it can be seen that the improved vertical resolution of the new POD modes has produced a dominant subsidiary oscillation ($\sim$5 cycles every 60 sols) in modes b3 and b4, as well as a set of time coefficients which are highly comparable with the original PCs. It is noted that mode b5 is much smoother in form than its 2-v counterpart, suggesting that any improvement in the vertical resolution scheme results in an enhancement of the subsidiary wave so that it almost becomes a pure frequency, i.e. changing from a mixed principal-subsidiary wave to a pure subsidiary wave. Smooth subsidiary wave structures are also seen in the MFC modes, being particularly evident in MFC modes a1 to a5.

The frequencies and amplitudes of modes b1 and b2 generated by the 4-v 20-D TE model (14 cycles every 60 sols) are almost identical to those of the original SGCM PCs (Figure 3.19). Additionally, modes b3 and b4 travel at a frequency of about 5 cycles every 60 sols, which is what is observed in the original system. It is noticeable that wave mode b5 also has a strong subsidiary wave form, being almost a pure frequency, but b6 contains fine perturbations, i.e. a mixed subsidiary and principal-doubled wave. Remarkably, this is also observed in the original system. Modes b7 and b8 behave very similarly to the SGCM PCs, both in terms of their high frequency structures and their amplitudes.

Interestingly, both TE modes b9 and b10 are highly distorted subsidiary wave types (i.e. mixed subsidiary-principal as well as a mixed subsidiary and **subsidiary-doubled**[1] wave), which is not observed in the original PCs. This disagreement might be revealing the nature of the high frequency modes in the original system, that they are in fact subsidiary waves.

To consolidate our previous work on the effects of the horizontal truncation we will now examine the 4-v 10-D and 6-D models in turn.

[1] 'Subsidiary-doubled' refers to those waves which propagate at double the frequency of the subsidiary waves.

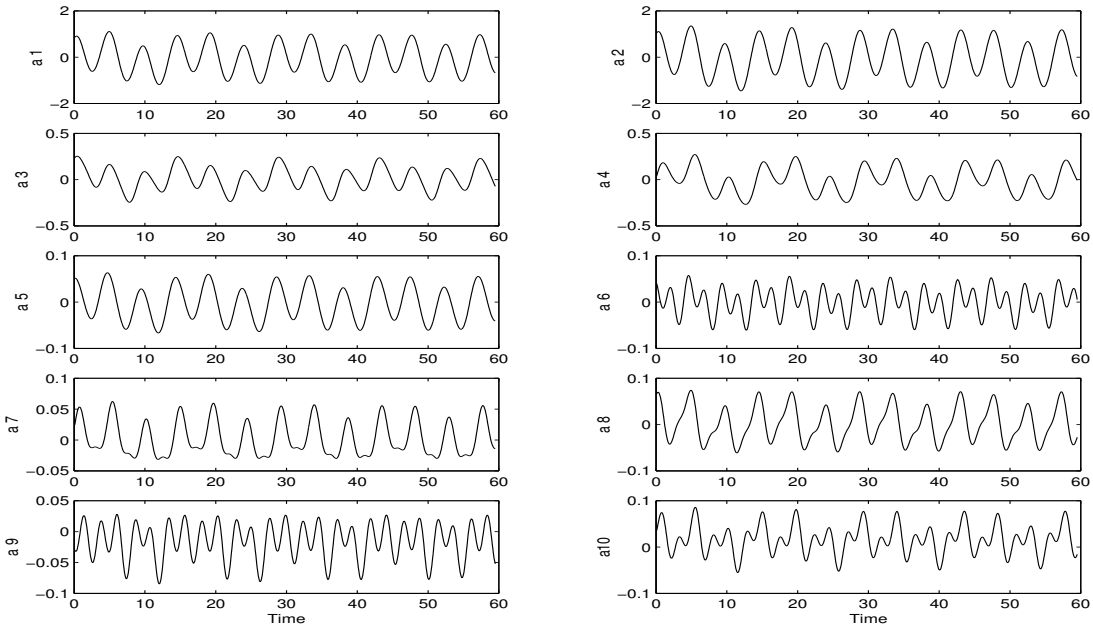


Figure 6.4: 20-D **CORRELATION** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.28) over 60 Martian days.

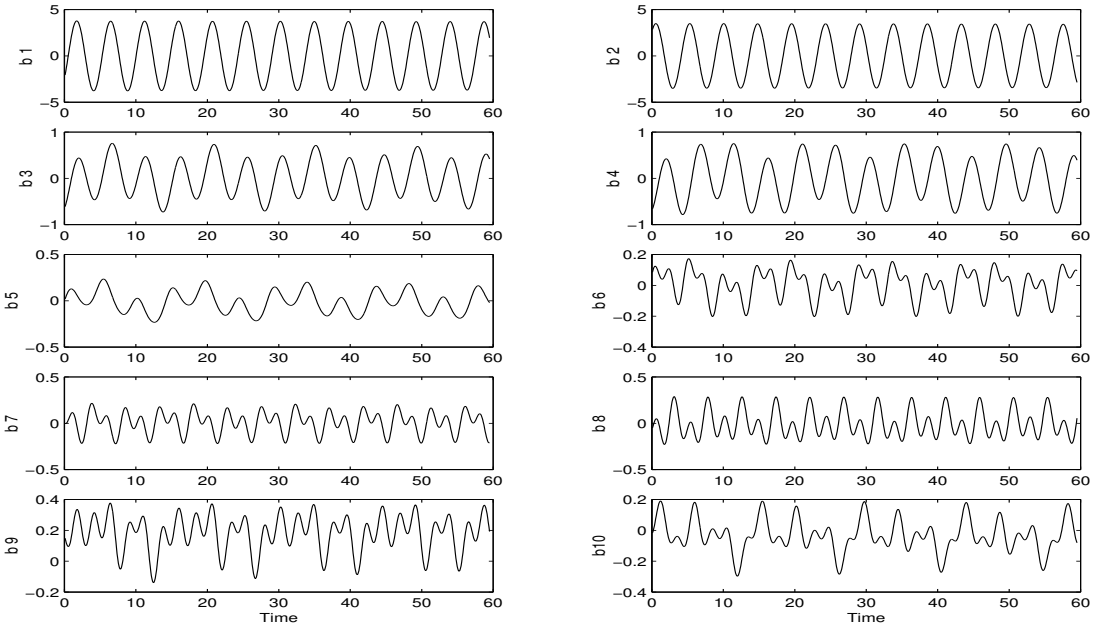


Figure 6.5: 20-D **CORRELATION** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.28) over 60 Martian days

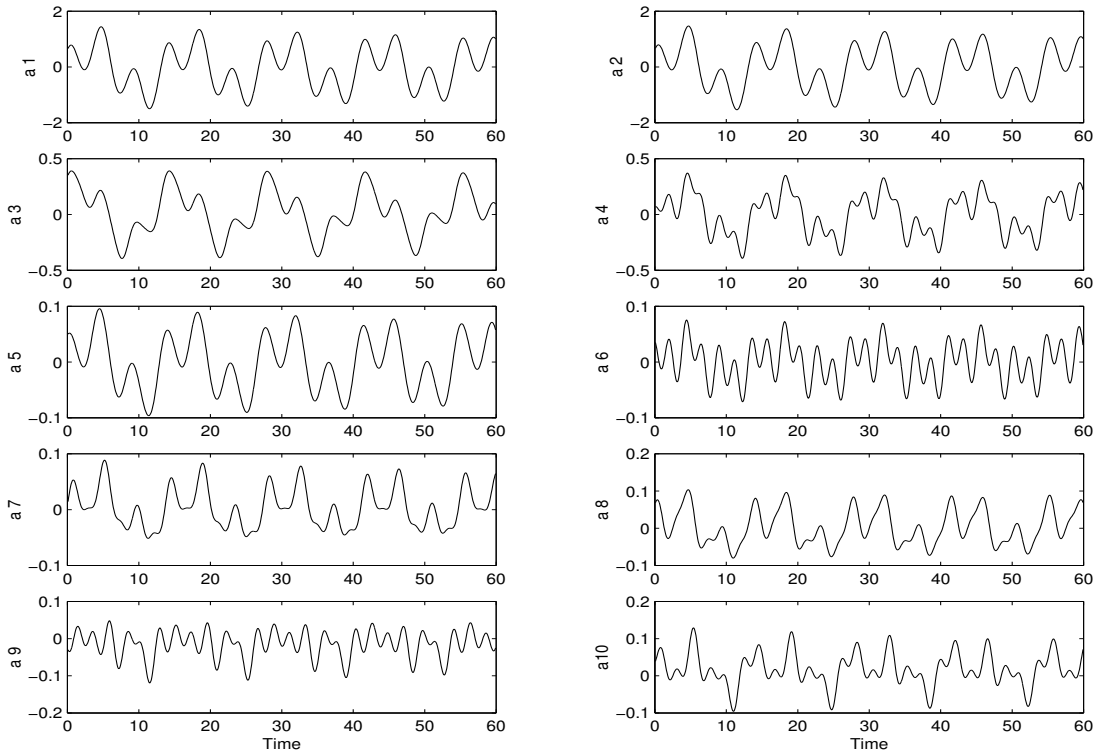


Figure 6.6: 20-D **KE** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.21) over 60 Martian days.

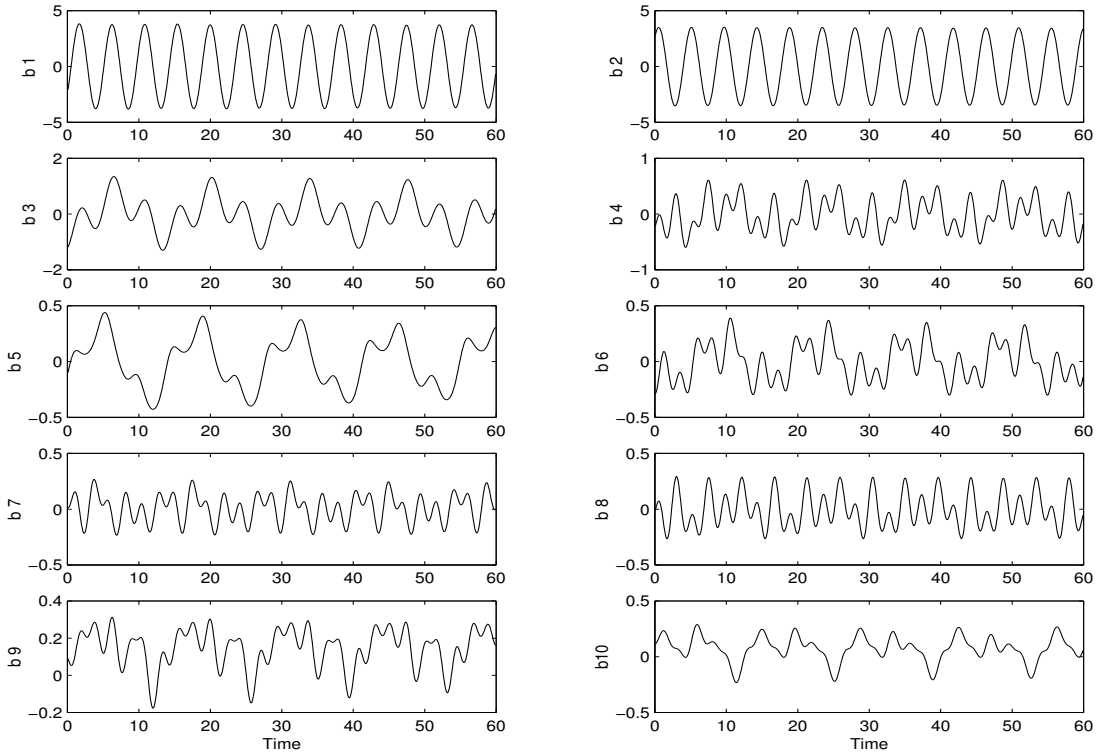


Figure 6.7: 20-D **KE** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.21) over 60 Martian days.

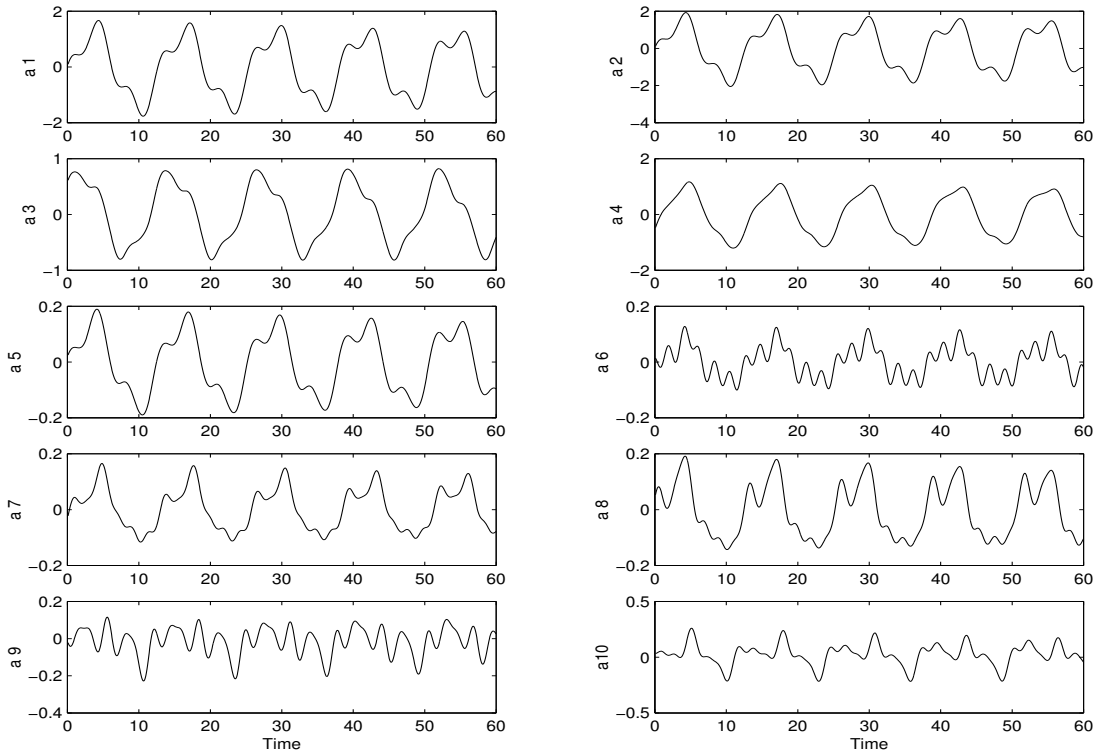


Figure 6.8: 20-D **TE** MODEL : Equilibrated amplitudes of the ten MFC modes in the mixed-wave regime (Γ=0.22) over 60 Martian days.

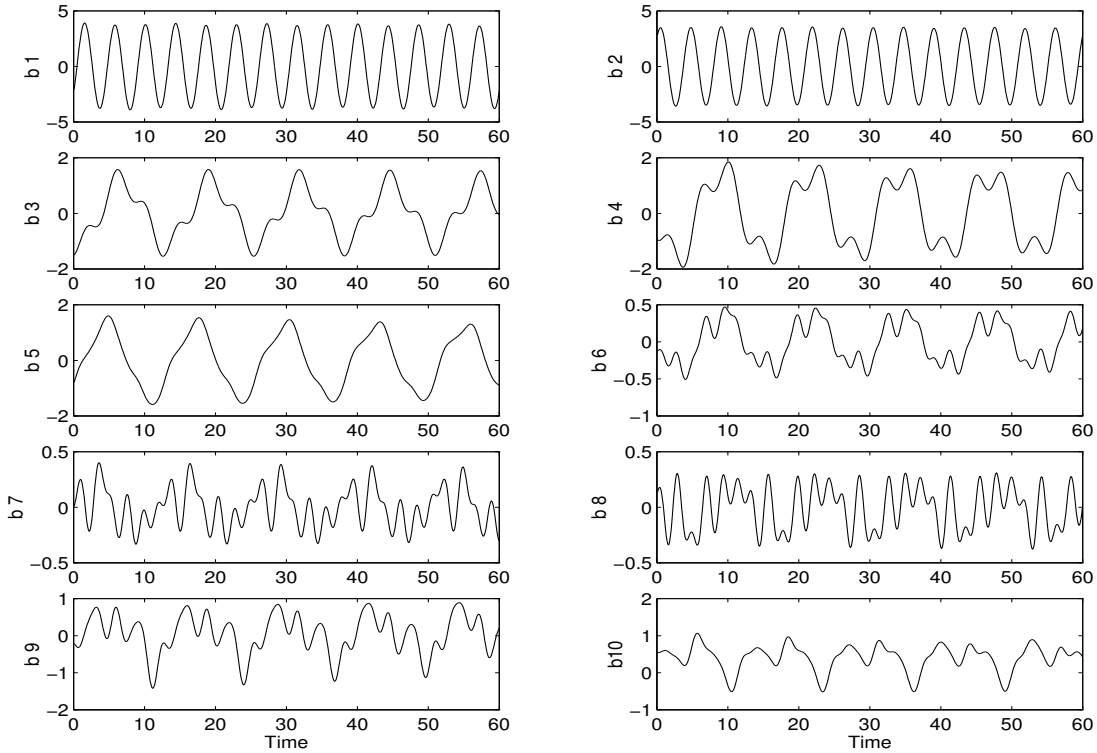


Figure 6.9: 20-D **TE** MODEL : Equilibrated amplitudes of the ten wave modes in the mixed-wave regime (Γ=0.22) over 60 Martian days.

6.3.2 10-D models

When comparing the 4-v 10-D correlation model (Figure 6.10(a)) with the 4-v 20-D version (Figure 6.5), the main difference between the two models is the frequency of the leading modes b1 and b2. In the 20-D version, b1 and b2 travel at a rate of about 13 cycles every 60 sols whereas the 10-D model produces modes which move at a rate of just over 12 cycles every 60 sols. It can be concluded here that nonlinear interactions with higher order modes changes the frequencies of the energy dominant modes.

A re-ordering of the MFC terms has occurred in the 10-D model. Retaining only four correction modes has resulted in the model re-ordering the modes so that a3 and a4 in the 10-D model appear as a6 and a9 in the 20-D version. So changes in the horizontal truncation affects the dynamical behaviour of the mean-flow correction modes. This has been seen before when we considered the different horizontal truncations of the 2-v models.

The 4-v 10-D KE model (shown in Figure 6.10(b)) produces a smooth mixed mode b4 (a mixed principal-subsidiary wave), without any of the distortions of its 20-D counterpart, i.e. there are no interactions with the principal-doubled wave (see Figure 6.7). So here the severe truncation reduction has resulted in the loss of the erratic structures which are also observed in the original model.

Again, a truncation reduction has produced an MFC reordering and thus a change in their dynamical behaviour. Here, correction modes a3 and a4 are very similar to modes a6 and a9 in the 20-D KE model.

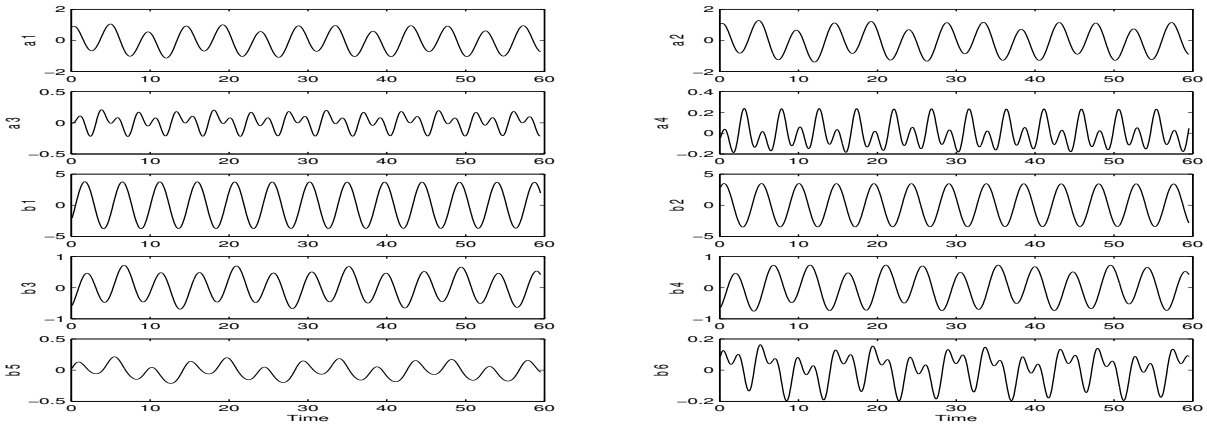
The 4-v 20-D TE model (Figure 6.9) produces a subsidiary oscillation ($\sim$5 cycles/60 sols) in mode b6 which also has high frequency perturbations (a superimposed principal-doubled wave). It is noted that the 10-D version (Figure 6.10(c)) does not duplicate this, suggesting that there is a need for the lower order modes to interact with the higher order modes to reproduce this behaviour. Again, we would argue that there is an insufficient amount of retained information in the 10-D model for such complicated structures to be generated, requiring interactions between the leading (modes b1 to b6) and higher order modes (b7 to b10).

The MFC modes have been reordered in the 10-D version (Figure 6.10(c)). Again, modes a3 and a4 appear to be very similar to modes a6 and a9 in the 20-D TE model (Figure 6.8).

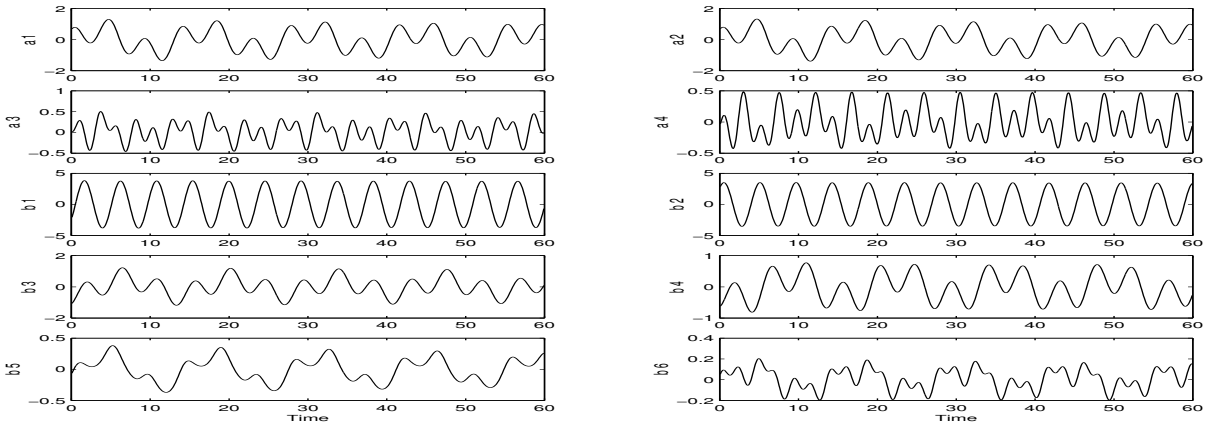
Summary

Even in the 4-v models, a severe horizontal truncation results in a crucial loss of information. This is supported by an absence of any erratic high frequency oscillatory behaviour in the 10-D models, identified as principal-doubled waves. The amount of vertical energy has affected the frequencies rates, where changes of $\sim$5%, 2% and 1% were respectively found for the 4-v correlation, KE and TE models from their 2-v counterparts.

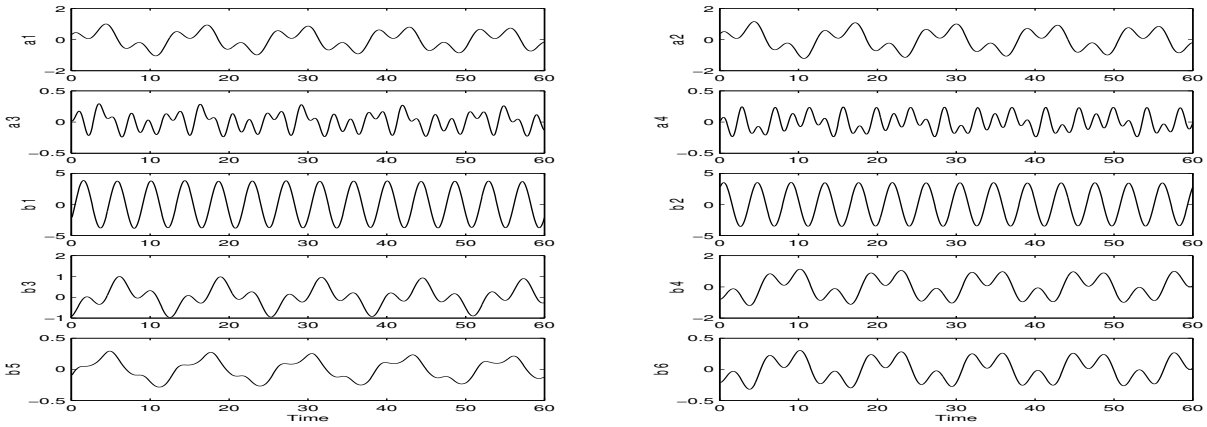
Clearly, for the most accurate modal reproduction a 4-v 20-D TE model is required.



(a) 10-D **CORRELATION** MODEL : Equilibrated amplitudes of the leading four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.28) over 60 Martian days.



(b) 10-D **KE** MODEL : Equilibrated amplitudes of the first four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.21) over 60 Martian days.



(c) 10-D **TE** MODEL : Equilibrated amplitudes of the first four MFCs (a1 - a4) and six wave modes (b1 - b6) in the mixed-wave regime (Γ=0.22) over 60 Martian days.

Figure 6.10: Amplitudes of the leading MFC and wave modes in the mixed-wave regime from the 10-D (a) correlation, (b) KE and (c) TE norm models.

6.3.3 6-D models

The leading modes b1 and b2 in the 4-v 6-D and 20-D correlation models (Figures 6.11(a) and 6.5) travel at a rate of just over 12 cycles every 60 sols. The form of the mixed modes b3 and b4 have been weakly affected by the severe truncation. However, unlike in the 20-D version, the 6-D model produces dominant principal behaviour in the mixed-wave regime.

When comparing the 4-v KE 6-D model (see Figure 6.11(b)) with its 20-D counterpart (Figure 6.7) it is noticeable that the distortions which exist in mode b4 (i.e. a superimposed principal-doubled wave) are absent in the 6-D version, simply now being a mixed principal-subsidiary wave. Significantly, the 4-v 6-D TE model (Figure 6.11(c)) generates modes which are very similar to the 20-D version (Figure 6.9), extracting a clear subsidiary dominant mixed wave in modes b3 and b4.

Summary

Both the vertical truncation level and the choice of norm affects the (1) frequency, (2) amplitude, (3) modulation and the (4) erratic disturbances of the waves. However, it has been shown that the amplitudes and frequencies of the modes are robust to changes in the horizontal truncation level.

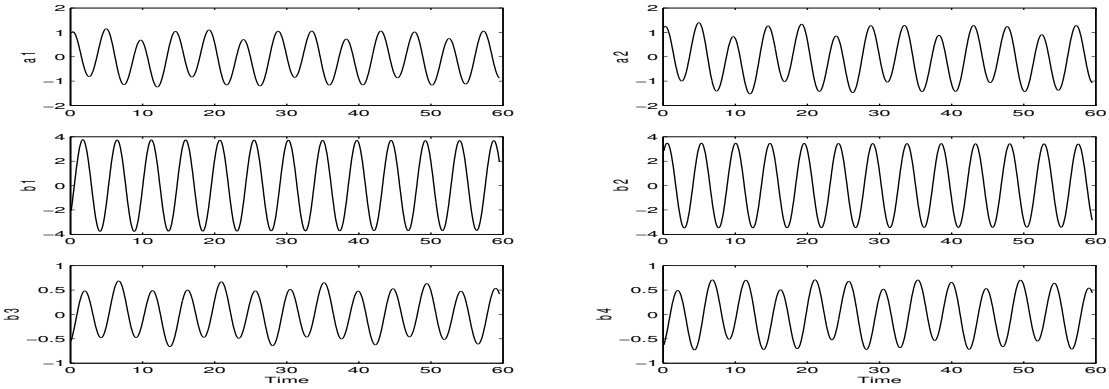
6.4 The mixed wave regime

6.4.1 Power spectra results

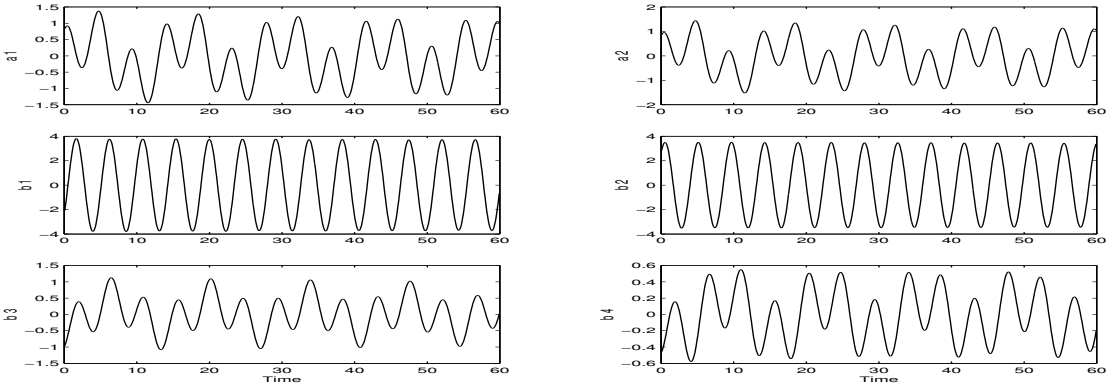
Figure 6.12 shows the power spectra profiles for the mixed-wave regime using the 10-D correlation, KE and TE norm models. The top two diagrams ((a) and (b)) show MFC modes a1 and a2 where the TE spectra peaks (shown in blue) depict a dominant frequency at about 0.1 cycles/sol (a subsidiary wave) and a much weaker peak at about 0.26 cycles/sol (a principal wave). The KE norm peaks (in green) depict a superimposed mixed wave, with a dual frequency rate of about 0.08 and 0.25 cycles/sol. The correlation norm (red) shows a weak peak at 0.08 cycles/sol and a larger peak at 0.25 cycles/sol. Similar behaviour is seen in the bottom two diagrams ((e) and (f)) for wave modes b3 and b4.

Within the mixed-wave regime the TE norm has separated a pure principal oscillation (in modes b1 and b2) from a pure subsidiary oscillation (in modes b3 and b4). The KE norm extracts a superimposed principal-subsidiary oscillation in modes b3 and b4, whereas the correlation norm places greater emphasis on the principal wave. Again, an almost pure principal wave is seen in wave modes b1 and b2, irrespective of which norm is used.

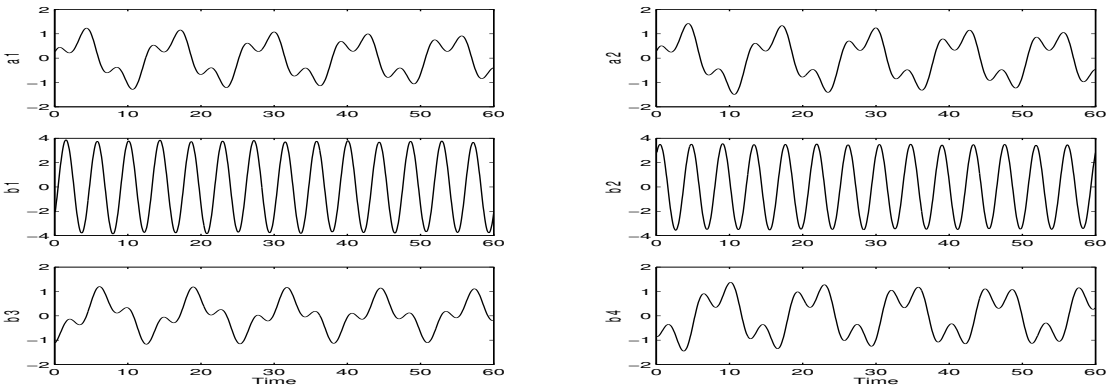
Comparing the power spectra results with those of the 2-v models in the mixed-wave regime (Figure 3.20(a)) reveals that the additional information retained in the vertical direction has resulted in a more dominant subsidiary oscillation in wave modes b3 and b4, particularly in the TE norm models where the principal wave in MFC modes a1, a2 and wave modes b3 and b4 has almost disappeared.



(a) 6-D **CORRELATION** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.28) over 60 Martian days.



(b) 6-D **KE** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.21) over 60 Martian days.



(c) 6-D **TE** MODEL : Equilibrated amplitudes of the first two MFCs (a1,a2) and four wave modes (b1 - b4) in the mixed-wave regime (Γ=0.22) over 60 Martian days.

Figure 6.11: Amplitudes of the leading MFC and wave modes in the mixed-wave regime from the 6-D (a) correlation, (b) KE and (c) TE norm models.

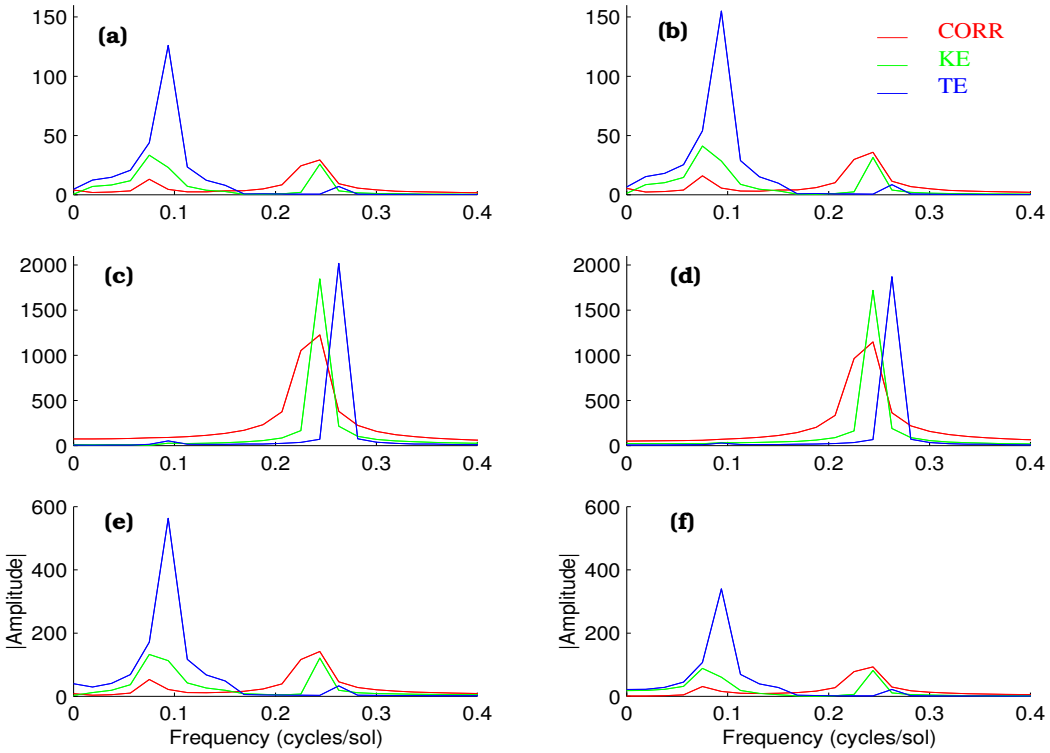


Figure 6.12: POWER SPECTRA OF THE MIXED-WAVE REGIME. The figure shows results from the 10-D 4-v correlation (red), KE (green) and TE (blue) norm models of the first two MFC modes a1 and a2 in (a) and (b); the wave modes b1 and b2 in (c) and (d); and wave modes b3 and b4 in (e) and (f). Again, there is only one dominant frequency found in wave modes b1 and b2. MFC modes a1, a2 and wave modes b3 and b4 show a strong presence of a subsidiary wave, particularly the TE norm model which has a dominant frequency of about 0.1 cycles/sol and a negligible principal wave frequency at about 0.25 cycles/sol. The improvement in the vertical resolution scheme has produced mixed waves with increasingly dominant subsidiary wave structures.

It can be concluded that the 4-v energy models produce modes which have frequencies almost identical to those observed in the original SGCM data. The power spectra profiles for the original SGCM data (shown in Figure 3.20(a)) illustrate that only one frequency is dominant in the leading four modes (a frequency of about 0.23 cycles/sol for wave modes b1 and b2, and a frequency of about 0.09 cycles/sol for modes b3 and b4). This was not reproduced in the 2-v models where modes b3 and b4 had two dominant frequencies (identified as a superimposed principal-subsidiary wave). This mixed-wave becomes a one-frequency subsidiary wave if additional vertical energy is retained in the TE model.

Comparing the frequencies in Figure 6.12 with those shown in Figure 5.23 it is noticed that the use of the additional two baroclinic modes has resulted in all the frequency values shifting to the right (towards higher frequency values) of their 2-v counterparts, having moved closer to the true frequency values. Therefore, the use of an energy norm or an improved vertical scheme both have the effect of increasing the frequency rates of the modes.

Summary

The 4-v 10-D TE Galerkin model produces leading wave modes which behave like pure principal or subsidiary waves and thus reproduces the original system with remarkable accuracy.

6.4.2 Phase Portraits

Figure 6.13 show phase portraits and Poincaré sections (at b1=0) obtained from the correlation, KE and TE Galerkin models respectively, over 300 Martian days and after a 200 day transient, using the same Γ values as are used in § 6.3.1.

The phase portraits shown in Figure 6.13 are very similar to those produced in Figure 5.22 for the 2-v case. An important distinction is seen in the new phase portraits of energy wave modes b1 and b2 (Figures 6.13(b) and 6.13(c)). It is evident that the use of the energy norms as well as an improved vertically resolved scheme have resulted in a thickening of the periodic structure. This accounts for a greater amplitude modulation in the leading two wave modes and thus a greater affinity with the original SGCM model. This seems to suggest that there will be a greater amount of amplitude modulation if more of the original energy is extracted and retained in the Galerkin model.

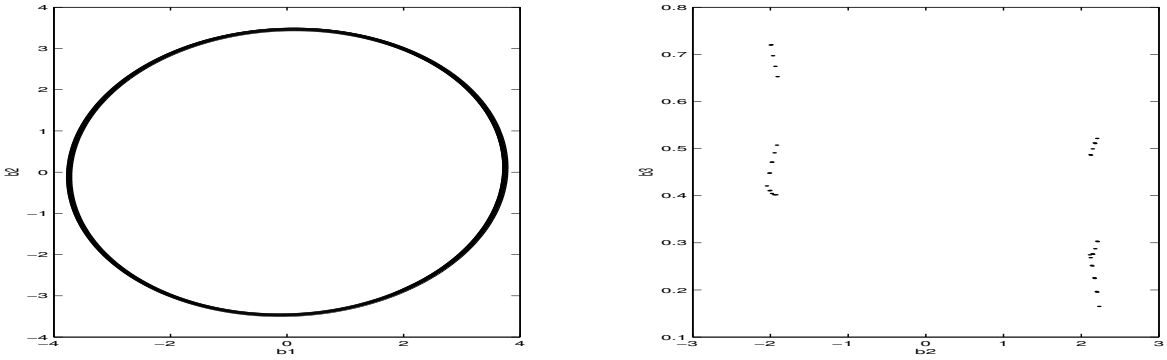
The Poincaré section in Figure 6.13(a) does not describe a standard limit cycle, but a period-doubled 'ribbon'. The Poincaré sections in Figures 6.13(b) and 6.13(c) depict a period-tripled 'ribbon'. Unlike the Poincaré sections of the original system in Figure 5.13, the trajectories do not cover the entire interior of the surface, but the Galerkin Poincaré sections (Figure 6.13) reveal additional structure. It can be seen that the 4-v KE and TE models behave similarly to the 2-v TE model (Figure 5.22(c)), and indicates that the period-tripled 'ribbons' are the most accurate reproduction of the original system.

6.4.3 Transient behaviour

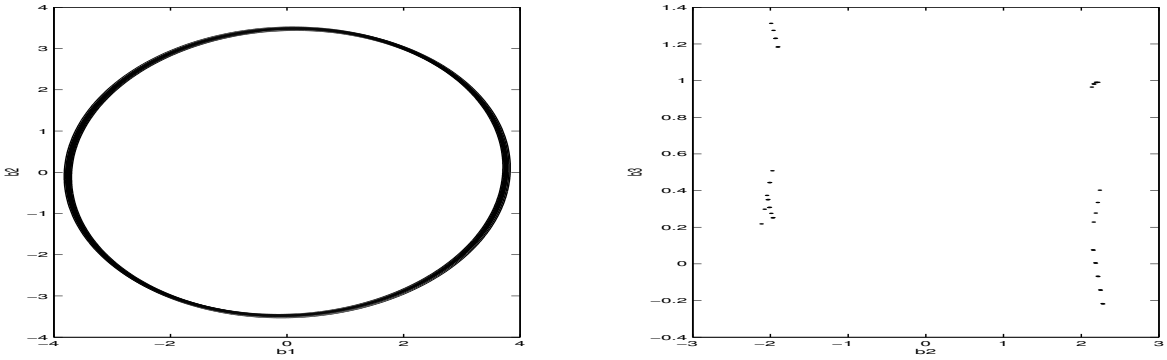
It is must be noted that when integrating the set of ODEs it is essential to observe the *transient* behaviour. By definition, transient behaviour will disappear after some time. Therefore, at first, it might seem that neither are relevant nor interesting for time series analysis. However, both impressions are wrong.

By choosing an initial condition that does not lie on the attractor, we might have to wait some time until the trajectory has settled down there. During this time the motion may have completely different properties than on the attractor itself. The transient time can be extremely short for highly dissipative systems. Numerical simulations show that systems with a large number of degrees of freedom can have extremely long transients (see Kantz and Schreiber (1997)).

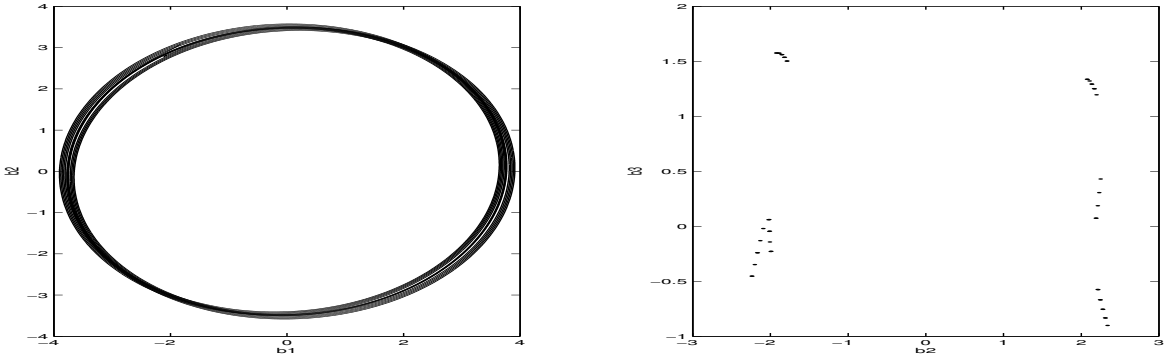
Figure 6.14 shows the transient trajectory moving towards the thick banded limit cycle from an arbitrary small initial condition (IC). The attractor here thus behaves very similarly to a limit cycle. Collins and James (1995) used the leading two PCs and projected them onto a plane by plotting one PC against the other (as has been done in Figure 6.14). They used the initial spin up of the model to represent the wave 3 transient and they showed that this transient almost immediately moved to a limit cycle which they said could be thought of as the attractor of the system. We have almost reproduced this result using the 4-v 20-D TE model (see Figure 6.14), where it can be seen that from small ICs the transient trajectory increases



(a) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **CORRELATION** MODEL AT $\Gamma = 0.28$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for over 80 % of the wave energy in the system.



(b) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **KE** MODEL AT $\Gamma = 0.21$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for 80 % of the wave energy in the system.



(c) PHASE PORTRAITS AND POINCARÉ SECTIONS (AT b1=0) OF THE LEADING GALERKIN WAVE MODES FROM THE 20-D **TE** MODEL AT $\Gamma = 0.22$. The trajectory is nearly a limit cycle, but the transverse (ribbon) structure manifested by the (b1,b2) section on the right is suggestive of motion on a torus and this accounts for nearly 80 % of the wave energy in the system.

Figure 6.13: Phase space reconstructions and Poincaré sections (at b1=0) using the 4-v 20-D (a) correlation, (b) KE and (c) TE norm models. A structure which is close to a periodic orbit is obtained in (b1,b2)-space and becomes thicker if the energy norms are used (note the thicker banded orbit if the TE norm is used).

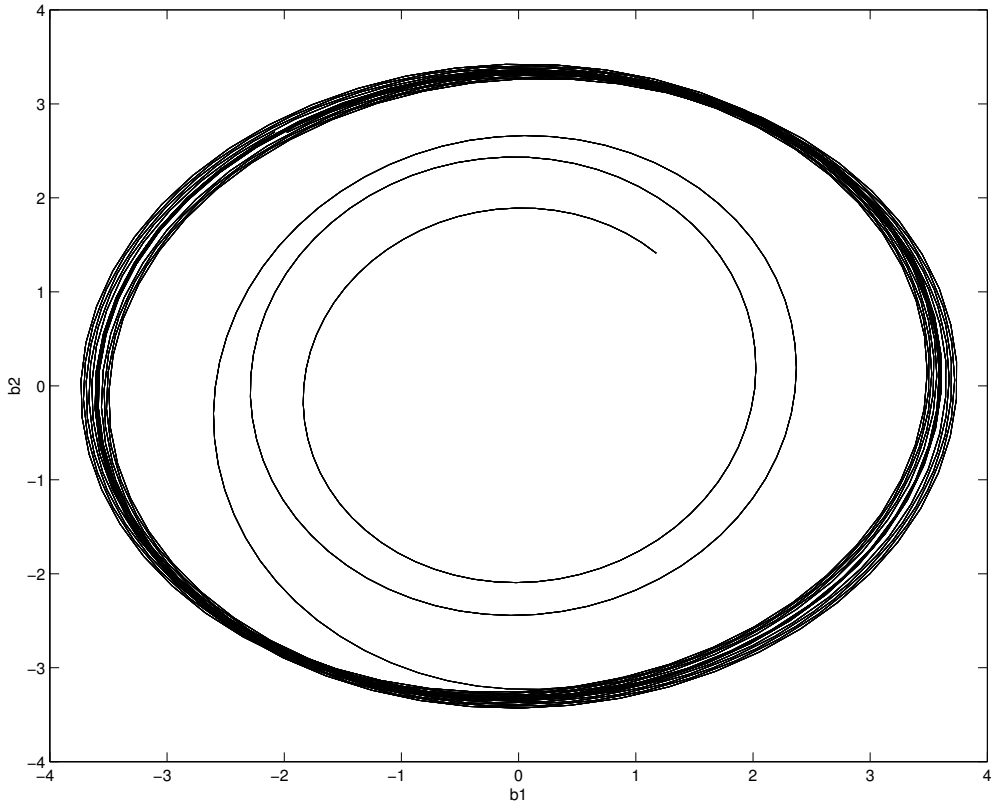


Figure 6.14: Transient attractor construction using the 4-v (10,10) TE model. The figure shows the attractor trajectories, i.e. wave mode b1 plotted against b2 for the mixed-wave regime from small ICs.

in amplitude and feeds outwards until the periodic structure is reached.

Similar small ICs were used to check the robustness of the model. In all cases comparable results were found. The phase portrait of wave modes b1 and b2 can certainly be identified as some kind of limit cycle and thought of as the low-dimensional attractor of the system.

In Figure 6.15 we tested whether the same periodic structure (identified in Figure 6.14) was obtained if the ICs were chosen to sit outside the attractor. The results showed that the large transients were sucked towards the wave 3 attractor. Infact, the waves equilibrated to the principal wave structure much faster than the growing waves shown in Figure 6.14.

The model is evidently robust also to large ICs and this is shown in Figure 6.16 where all the ICs outside the attractor move onto the wave 3 limit cycle within 5 cycles. Figures 6.17 and 6.18 illustrate transient behaviour and just how quickly the models equilibrate. As was found in Chapter 5, the energy norm models equilibrate much faster.

Figure 6.17 shows the transient behaviour from a small IC. It is observed that after 3 cycles the wave has equilibrated. In Figure 6.18 the transient behaviour for a large IC is shown. It can be seen that the wave equilibrates after only 2 cycles.

Many ICs were used which lay outside the periodic attractor. In each case, following rather short transients, the trajectories returned to the attractor. The dominance of the periodic

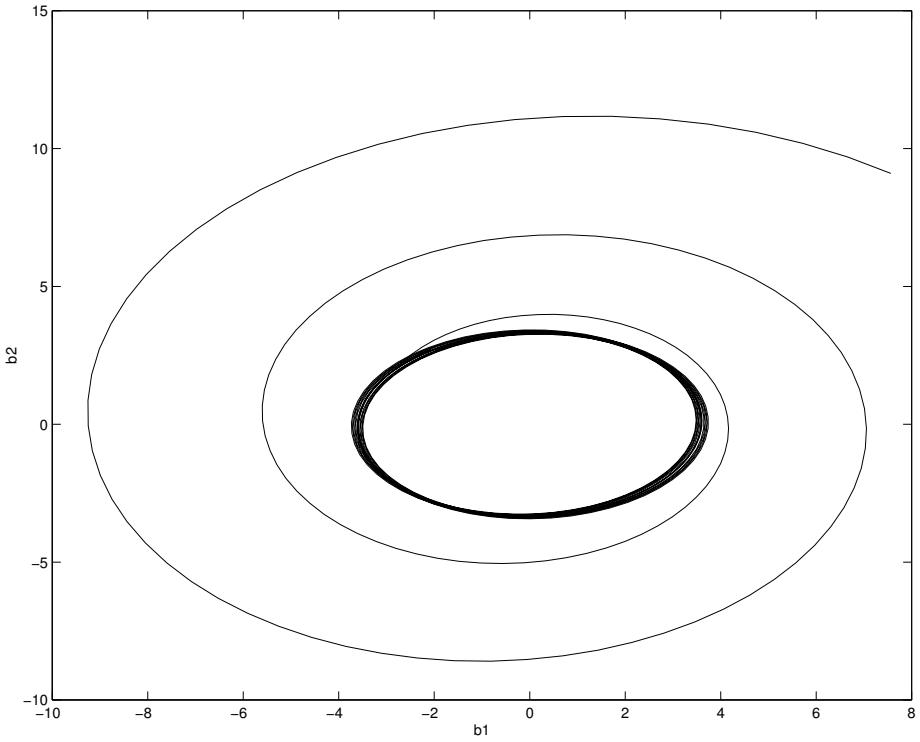


Figure 6.15: Transient attractor construction using the 4-v (10,10) TE model. The figure shows the attractor trajectories, i.e. wave mode b1 plotted against b2 for the mixed-wave regime from large ICs.

nature of the leading two wave modes (b1 plotted against b2) is clearly visible. Irrespective of the chosen ICs the system stabilises to this attractor.

6.4.4 The attractor

Collins and James (1995) showed that by changing the parameters (the radiative and the friction timescales) there existed two flow regimes in the original SGCM, namely a wave 3 flow at (τ_D, τ_E)=(3,2) or a weak wave at point (2,2) (see the SGCM regime diagram in Figure 3.1). It was shown that the weak wave attractor had trajectories which wandered around near the origin, contrasting with the limit cycle of the wave 3 regime. It was argued that because various parameters had been changed then this process was similar to a bifurcation as the nature of the attractor had changed. They did not suggest that this transition was a Hopf bifurcation but stated that there were some obvious similarities (a fixed point changing to a limit cycle).

Significantly, we have identified two Hopf bifurcation points which cause the flow to change from a subsidiary wave to a quasi-periodic flow and finally to a principal wave regime for decreasing Γ. The wave 3 flow is analogous to the principal wave and may be the weak wave flow can be identified with the subsidiary wave (i.e. an absent wavenumber 3 flow). This then suggests that the process which Collins and James observed has been reproduced and identified in our low-dimensional ODE sets, i.e. the process of a weak wave attractor, where the trajectories wander around near the origin (or a fixed point), changing to a wave 3 attractor (or a limit cycle) via a Hopf bifurcation.

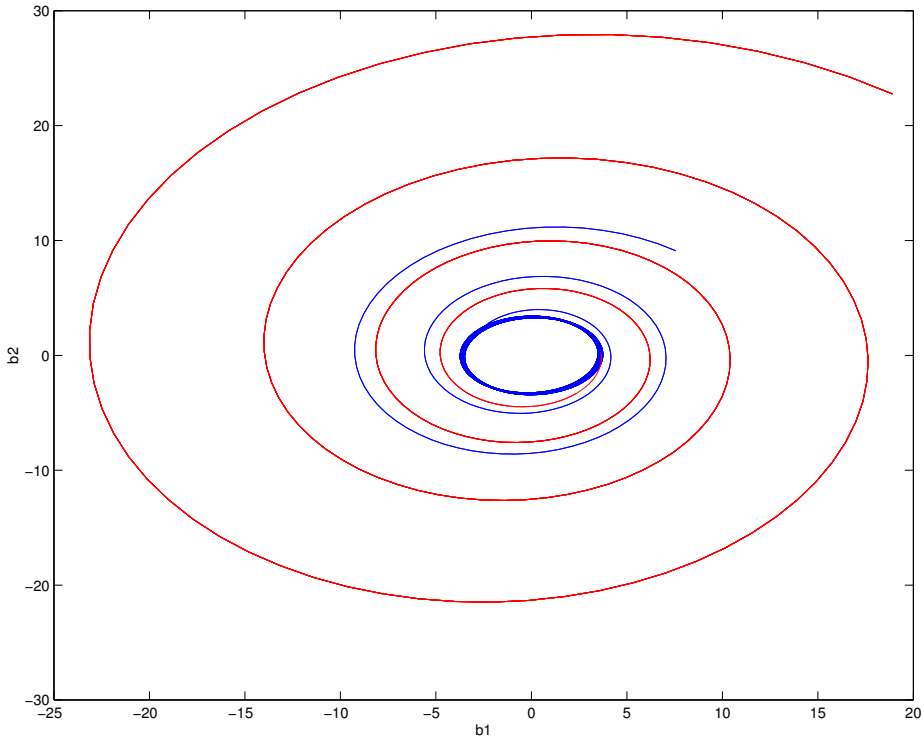


Figure 6.16: Robustness of the attractor using the 4-v (10,10) TE model. The figure shows the attractor trajectories, i.e. wave mode b1 plotted against b2 for the mixed-wave regime from large ICs.

Attractor reconstruction summary

Collins and James (1995) stated that *"the EOF behaviour ... does highlight the most important aspects of the (Martian) circulation and provides a link between a geophysically plausible system and the low-order models of dynamical systems."* They go on to say that *"the wave 3 flows presented in this paper are thought to show a low-dimensional behaviour."*

We have shown that the amplitude, frequency and amplitude modulation of this flow can be accurately reproduced by a 6-D spherical QG Galerkin TE model.

It has been shown that by using the energy norms and higher truncation levels the oscillations of the produced Galerkin modes are more comparable with those observed in the original system. It is also evident that for the KE and TE norm models the wave 3 attractor takes on the form of a thicker limit cycle (referred to as "fuzzy" by Collins and James (1995)), which appears to agree with the suggestion that the energy models describe the original wave 3 regime far more accurately because of the emphasis they give to the energy processes within the original system, i.e. they successfully pick out the original modulations in the flow.

6.4.5 Success in reproducing quasi-periodic behaviour

The systems which exhibited quasi-periodic transients in Mundt and Hart's (1994) work always "blew up" i.e. solutions grew exponentially fast. To solve this problem Mundt and Hart had to include a simple form of dissipation (following after Smith et al. (1991)), where dissipation was added to each mode by using the relationship $\frac{da_i}{dt} \propto -\rho M^{\alpha} a_i$, where ρ is an arbitrary scale factor, M is the total wavenumber (e.g. M =1 corresponds to the largest energy mode), and

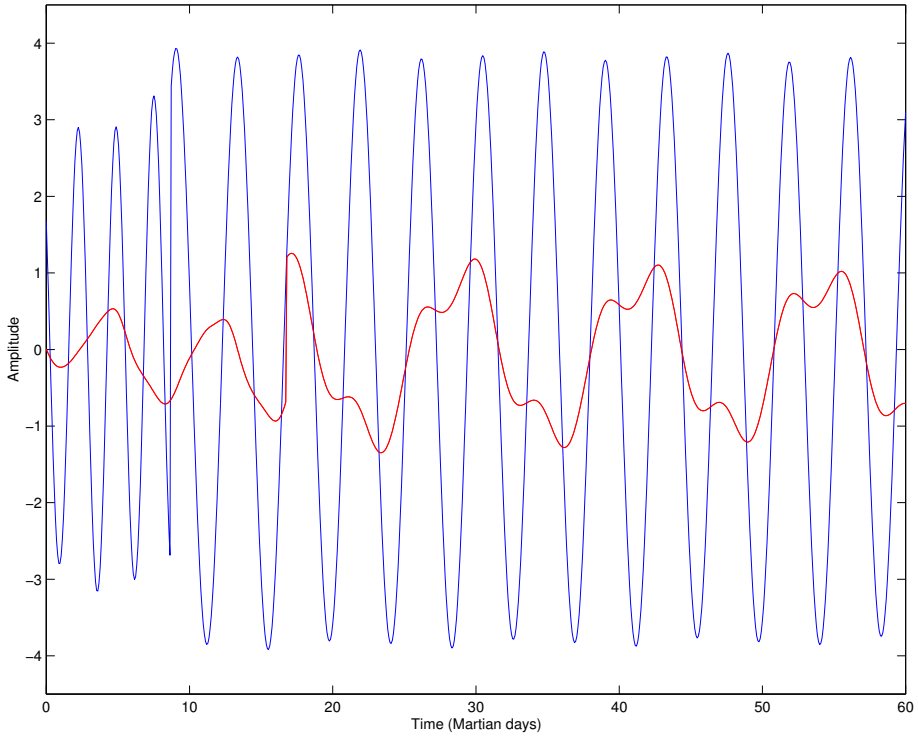


Figure 6.17: Transient behaviour in the 4-v (10,10) TE model at $\Gamma = 0.22$. Wave modes b1 (blue) and b3 (red) equilibrating to principal (∼14 cycles/60 sols) and subsidiary waves (∼5 cycles/60 sols) respectively from small ICs.

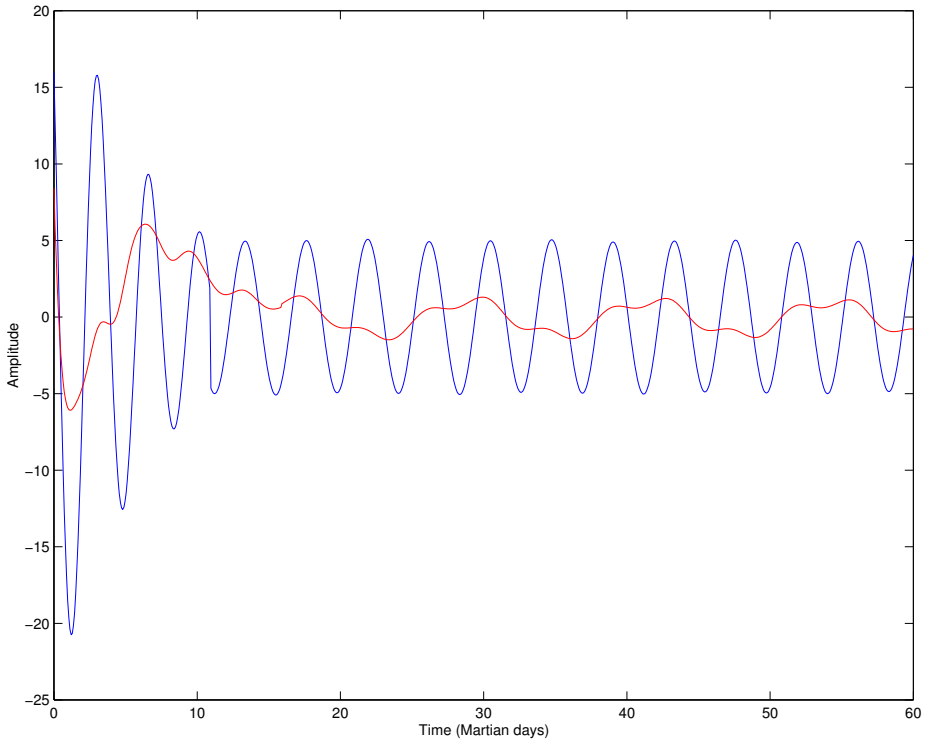


Figure 6.18: Transient behaviour in the 4-v (10,10) TE model at $\Gamma = 0.22$. Wave modes b1 (blue) and b3 (red) equilibrating to principal (∼14 cycles/60 sols) and subsidiary waves (∼5 cycles/60 sols) respectively from large ICs.

α is a scaling constant. Such a scheme had to be introduced if quasi-periodic structures were to remain stable.

In the models discussed in this thesis dissipation has been parameterised in terms of Ekman friction. In none of the integrations made was there an observed case in which a quasi-periodic regime 'blew up' with any choice of initial conditions. It has meant that it has not been necessary to add ad-hoc dissipation in order to achieve accurate results. Mundt and Hart concluded that *"a relatively low-dimensional dynamical system which qualitatively captures the mixing between barotropic and baroclinic instabilities results in quasi-periodic behaviour."* As our used POD state vector included both the barotropic and baroclinic components of the flow, the construction of the correlation matrix resulted in the mixing of these modal components. A successful 'mixing' process may provide one possible explanation of why such accurate quasi-periodic oscillatory structures were observed in our low-order ODE sets.

Schubert (1985) used a barotropic model to understand the nature of the dominant modes of variability of the Earth's northern hemisphere winter circulations and concluded that "*one must look into the vertical structure of the atmosphere to learn more*" about the physical processes (and energy exchanges) which occur in the system being modelled. He continued to argue that "*perhaps combined barotropically/baroclinically unstable modes need to be studied further.*" Is this suggesting that without a 'sufficient' mixing between the barotropic and baroclinic modes the quasi-periodic regimes are not extracted? If this is the case, then it is crucial to model the vertical structures as accurately as possible. It has been shown in Chapter 2 that the used vertical modes have been obtained from an equation which seeks to minimise APE, and since atmospheric circulations are widely interpreted as arising from the release of APE, then this minimisation principle naturally describes these baroclinic and barotropic energy processes. This may suggest that data sets which have dominant mixing processes (as well as energy exchanges) cannot be accurately described by a two-mode model as a satisfactory vertical resolution scheme is required for the existence of a quasi-periodic flow.

Arguing in this way begs the following question: 'Why are our generated modes so comparable to those of the original system when only two vertical modes have been used ?' Our findings suggest that the process of barotropic and baroclinic mixing in the SGCM does not occur to a significant degree. Does this simply mean that we have been rather fortunate in choosing a data set that does not require an important role for APE? No. It has been shown that APE almost represents 50 % of the total energy. Clearly, sufficient vertical information (of the behaviour of the steady wave 3 flow) is conveniently captured by using only two vertical modes (the barotropic and first baroclinic mode), but this might not be the case with other data sets.

6.5 The dynamical role of the double Hopf bifurcation

Collins and James (1995) stated regarding the SGCM that *"the model behaviour cannot be simply explained using linear instability theory. Nonlinear wave and wave-mean flow interactions must be important in the genesis and maintenance of the wave 3."*

Due to the close proximity of the double Hopf bifurcation points in relation to the Hopf points at Froude number 8.14 (shown in § 6.1), it was thought that the double Hopf may affect the two-wave interactions and so an understanding in the theory of co-dimension 2 points[2] may help in explaining the various modal stability exchanges as well as the quasi-periodic motions observed in the data.

The aim is to try to explain the mixed-mode regime, why it occurs and why in its observed form. Thus the quasi-periodic[3] flow will be an important regime in the forthcoming analysis.

6.5.1 A superposition of two waves

In this section we will investigate a modified hypothesis of Pfeffer and Fowlis (1968) and that considered in Moroz and Holmes (1984); namely that wave dispersion is a *nonlinear* superposition of two small, but finite, amplitude wave patterns of different zonal wavenumbers (but the same azimuthal wavenumber), having distinct frequencies and vertical structures.

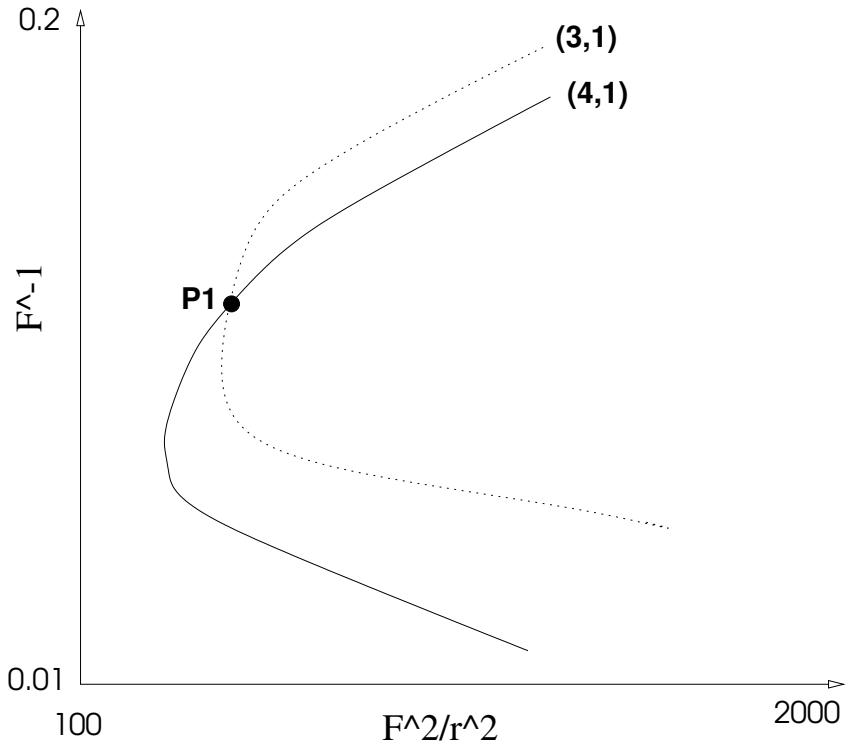


Figure 6.19: Neutral stability curves in $(F^2/r^2, F^{-1})$ parameter space from Moroz and Holmes (1984), where F is the internal rotational Froude number and r is the viscosity parameter (a measure of Ekman suction effects). The intersection of the waves with wavenumbers (3,1) and (4,1) define the double Hopf bifurcation point at P_1.

The model considered by Moroz and Holmes was the QG two-layer model on a β-plane as discussed by Pedlosky (1970). It was shown that while mixed modes do occur (in which both waves are present and interact to produce a quasi-periodic motion), they could be either

[2]The terminology used here essentially means that two parameters have to be tuned to achieve this type of bifurcation.

[3]The representation of such flows as superpositions of two waves have been proposed as a possible model for amplitude vacillation (see Lindzen et al. (1982) and Moroz and Brindley (1982)) and Moroz and Holmes (1984) for wave dispersion.

stable or unstable. It was concluded that in cases where the mixed mode is stable the hypothesis of Pfeffer and Fowlis is acceptable, whereas in the case for unstable mixed modes such an assumption cannot be used as unstable modes would not generally be physically observable.

From the marginal stability curves shown in Figure 6.19 it can be seen that it is possible to find points in parameter space at which two different modes simultaneously lose their stability according to linear theory (i.e. at a double Hopf bifurcation point). Moroz and Holmes argued that the wave interactions between wavenumber pairs of dispersive waves and the importance of the point of intersection of the stability curves at P_1 (shown in Figure 6.19) were considerably influential in the dynamics observed in the neighbourhood of the double Hopf point. They suggested that the pair of modes have distinct frequencies and vertical structures and therefore are possible candidates for the superposition scenario suggested by Pfeffer and Fowlis (1968). We shall therefore study the *nonlinear* development of a wave state which is a linear sum of the two waves destabilised at these points and hope that this will help in explaining some of the regimes observed in the Galerkin models.

6.5.2 Unfolding the double Hopf bifurcation

Moroz and Holmes (1984) used the following truncated set of equations to describe the interaction between a pair of waves that are simultaneously marginally stable at a double Hopf bifurcation point P_i in $(F^2/r^2, F^{-1})$ parameter space:

$$\begin{aligned}
\dot{R}_1 &= \mu_1 R_1 - R_1(N_1^r R_1^2 + M_1^r R_2^2) + O(|R_1|^5); \\
\dot{R}_2 &= \mu_2 R_2 - R_2(N_2^r R_1^2 + M_2^r R_2^2) + O(|R_2|^5); \qquad (6.1) \\
\dot{\Psi}_1 &= \nu_1 - (N_1^i R_1^2 + M_1^i R_2^2) + O(|R_1|^4); \\
\dot{\Psi}_2 &= \nu_2 - (N_2^i R_1^2 + M_2^i R_2^2) + O(|R_2|^4); \qquad (6.2)
\end{aligned}$$

where the μ_j, ν_j are linear functions of the bifurcation parameters and $\mu_j = \nu_j = 0$ at the bifurcation point P_i. As the amplitude and phase variables of (6.1) and (6.2) decouple (up to $O(|R_i|^4)$) then this means that the solutions of this system can be studied using just the 2-D amplitude equations (6.1).

Takens (1974) has proved all of the necessary stability theorems for this problem. Moroz and Holmes (1984) summarised the major implications of these results (refer to Guckenheimer and Holmes, 1983, Chap. 7, for more details) and we will follow their work, concentrating on equation (6.1) and neglecting the $O(|R_i|^5)$ terms.

(1). If equation (6.1) has an asymptotically stable (respectively unstable) equilibrium point $(\bar{\mathbf{R}}_\mathbf{1}, \mathbf{0})$ or $(\mathbf{0}, \bar{\mathbf{R}}_\mathbf{2})$ on either axis, then equations (6.1) - (6.2) have an asymptotically stable (respectively unstable) periodic orbit of frequency $w_1 + O(\sqrt{\mu_1})$ or $w_2 + O(\sqrt{\mu_2})$.

(2). If (6.1) has an asymptotically stable (respectively unstable) equilibrium point $(\bar{\mathbf{R}}_\mathbf{1}, \bar{\mathbf{R}}_\mathbf{2})$ in the interior of the positive quadrant, then (6.1) - (6.2) have an asymptotically stable (re-

spectively unstable) invariant torus, i.e. (6.1) - (6.2) have quasi-periodic solutions.

If additionally the centre manifold theorem (Carr, 1981; Guckenheimer and Holmes, 1983) is used then the above results suggest that, sufficiently close to the degenerate bifurcation point P_i the full spherical QG equations in (4.7) will produce either periodic (for equilibrium points $(\bar{R}_1, 0)$ or $(0, \bar{R}_2)$) or quasi-periodic (for the interior point $(\bar{R}_1, \bar{R}_2)$) behaviour. Moroz and Holmes argued that if combinations of the parameters μ_j, N_j and M_j can be found which give equilibria $(\bar{R}_1, \bar{R}_2)$, $\bar{R}_j \neq 0$, then it can be concluded that stable quasi-periodic motions (which can be observed) should occur for corresponding parameter values in the full problem.

We first re-scale (6.1) to reduce the number of coefficients. Letting $r_1 = R_1\sqrt{|N_1^r|}$, $r_2 = R_2\sqrt{|M_2^r|}$ to obtain

$$\begin{aligned} \dot{r_1} &= \mu_1 r_1 - r_1\left(r_1^2 + b r_2^2\right) = f_1(r_1, r_2; b, \mu_1), \\ \dot{r_2} &= \mu_2 r_2 - r_2\left(c r_1^2 + d r_2^2\right) = f_2(r_1, r_2; c, \mu_2), \end{aligned} \tag{6.3}$$

where $b = \frac{M_1^r}{M_2^r}$, $c = \frac{N_2^r}{N_1^r}$ and $d = \pm 1$.

Examining the equilibrium solutions of (6.3) (i.e. where $f_1 = f_2 = 0$), then in the positive quadrant we obtain

$$\begin{aligned} (r_1, r_2) &= 0, \\ (r_1, r_2) &= (\sqrt{\mu_1}, 0) \text{ for } \mu_1 > 0, \\ (r_1, r_2) &= (0, \sqrt{\mu_2/d}) \text{ for } \mu_2/d > 0, \\ (r_1, r_2) &= \left(\sqrt{\frac{d\mu_1 - b\mu_2}{A}}, \sqrt{\frac{\mu_2 - c\mu_1}{A}}\right) \text{ for } \frac{d\mu_1 - b\mu_2}{A}, \frac{\mu_2 - c\mu_1}{A} > 0, \end{aligned} \tag{6.4}$$

where $A = d - bc$.

It is now helpful to recall when solving a linear problem of the form $\dot{x} = Ax$, where $A \in \mathbb{R}^2$, then

$$\text{Trace}[A] = \sum_{i=1}^{n} \lambda_i, \quad \text{Determinant}[A] = \prod_{i=1}^{n} \lambda_i, \tag{6.5}$$

where the $\lambda_i' s$ are the eigenvalues of a (2×2) matrix A. We recall that if $\text{Det}[A] < 0$ there exists a saddle and if $\text{Det}[A] > \frac{1}{4}\text{Trace}[A]^2$ a focus (attracting if $\text{Trace}[A] < 0$ and repelling if $\text{Trace}[A] > 0$).

It is necessary therefore to calculate the trace and determinant of the linearised system if the stability properties of the solutions are to be monitored. Calculating the linearised matrix at some point $(\bar{r}_1, \bar{r}_2)$ gives:

$$\left(\begin{array}{cc} \frac{d\dot{r}_1}{dr_1} & \frac{d\dot{r}_1}{dr_2} \\ \frac{d\dot{r}_2}{dr_1} & \frac{d\dot{r}_2}{dr_2} \end{array}\right)_{(\bar{r}_1, \bar{r}_2)} = \left(\begin{array}{cc} \mu_1 - 3\bar{r}_1^2 - b\bar{r}_2^2 & -2b\bar{r}_1\bar{r}_2 \\ 2c\bar{r}_1\bar{r}_2 & \mu_2 - c\bar{r}_1^2 - 3d\bar{r}_2^2 \end{array}\right). \tag{6.6}$$

The behaviour remains relatively simple as long as the Hopf bifurcations do not occur from the interior fixed point $(\bar{r}_1, \bar{r}_2) = \left(\sqrt{\frac{d\mu_1 - b\mu_2}{A}}, \sqrt{\frac{\mu_2 - c\mu_1}{A}}\right)$, where it is found that:

$$\begin{aligned} \text{Trace} &= \frac{2}{A}\left[\mu_1 d(c-1) + \mu_2(b-d)\right], \\ \text{Determinant} &= \frac{4}{A}\left[(b\mu_2 - d\mu_1)(c\mu_1 - \mu_2)\right]. \end{aligned} \tag{6.7}$$

It is now possible to 'unfold' the degenerate case by varying the linear functions of the bifurcation parameters μ_1 and μ_2 in a full neighbourhood of (0,0). The analysis of (6.3) follows the workings of Guckenheimer and Holmes (1983). Simple linear analysis reveals the following equilibria (not distinguishing between modes and foci):

(1). $(r_1, r_2) = (0, 0)$ is always an equilibrium. It is a stable sink if $\mu_1, \mu_2 < 0$; a saddle if $\mu_1 > 0 > \mu_2$ or $\mu_2 > 0 > \mu_1$ and an unstable source if $\mu_1, \mu_2 > 0$.

(2). $(r_1, r_2) = (\sqrt{\mu_1}, 0)$ is an equilibrium if $\mu_1 > 0$. If in addition $c\mu_1 > \mu_2$ it is a sink; otherwise it is a saddle.

(3). $(r_1, r_2) = (0, \sqrt{\mu_2/d})$ is an equilibrium if $\mu_2/d > 0$. If in addition $b\mu_2 > d\mu_1$ it is a sink; otherwise it is a saddle.

(4). $(r_1, r_2) = \left([(d\mu_1 - b\mu_2)/(d - bc)]^{\frac{1}{2}}, [(\mu_2 - c\mu_1)/(d - bc)]^{\frac{1}{2}}\right)$ is an equilibrium if both radicals are positive. If $D < 0$ it is a saddle; if $D > 0, T < 0$ it is a sink; if $D > 0, T > 0$ it is a source.

The invariant lines (which satisfy $(r_1, r_2) = 0$) are $\mu_2 = \frac{d}{b}\mu_1$ and $\mu_2 = c\mu_1$. Hence the bifurcation to single wave modes $(\sqrt{\mu_1}, 0)$ and $(0, \sqrt{\mu_2/d})$ occur on lines $\mu_1 = 0$ and $\mu_2 = 0$, whereas bifurcations to the mixed mode occurs on lines $\mu_2 = \frac{d}{b}\mu_1$ and $\mu_2 = c\mu_1$ (but only if the roots are real in case **(4)**).

From the conditions for the existence of a fixed point given in (6.4), the Hopf bifurcations can only appear on the line:

$$\mu_2 = \frac{\mu_1 d(1-c)}{b-d}, \tag{6.8}$$

when

$$A = d - bc > 0, \tag{6.9}$$

i.e. the necessary condition is $\left(\frac{d}{b}\right) > c$. Furthermore, these bifurcations can only arise if the slope $\frac{d(1-c)}{b-d}$ of the line given in (6.8) lies between the slopes of the pitchfork lines:

$$\mu_2 = c\mu_1 \quad \text{and} \quad \mu_2 = \frac{d\mu_1}{b}, \tag{6.10}$$

and in the appropriate sector of the (μ_1, μ_2)-plane.

6.5.3 The SGCM bifurcation set

Attempting to identify a bifurcation set which allowed for Hopf bifurcations is fairly straightforward. Since the AUTO plots shown in Figure 5.6 showed that there existed both a subcritical and supercritical Hopf bifurcation at points HB_1 and HB_2 respectively, this implied that $d = -1$ in (6.3). However, it is important to determine whether there exists a phase portrait which is consistent with our numerical findings and also satisfies the above criteria. This situation was found and it corresponded (up to a reversal of time) to case IA of Section 7.4 in Guckenheimer and Holmes (1983) and is shown in Figure 6.20.

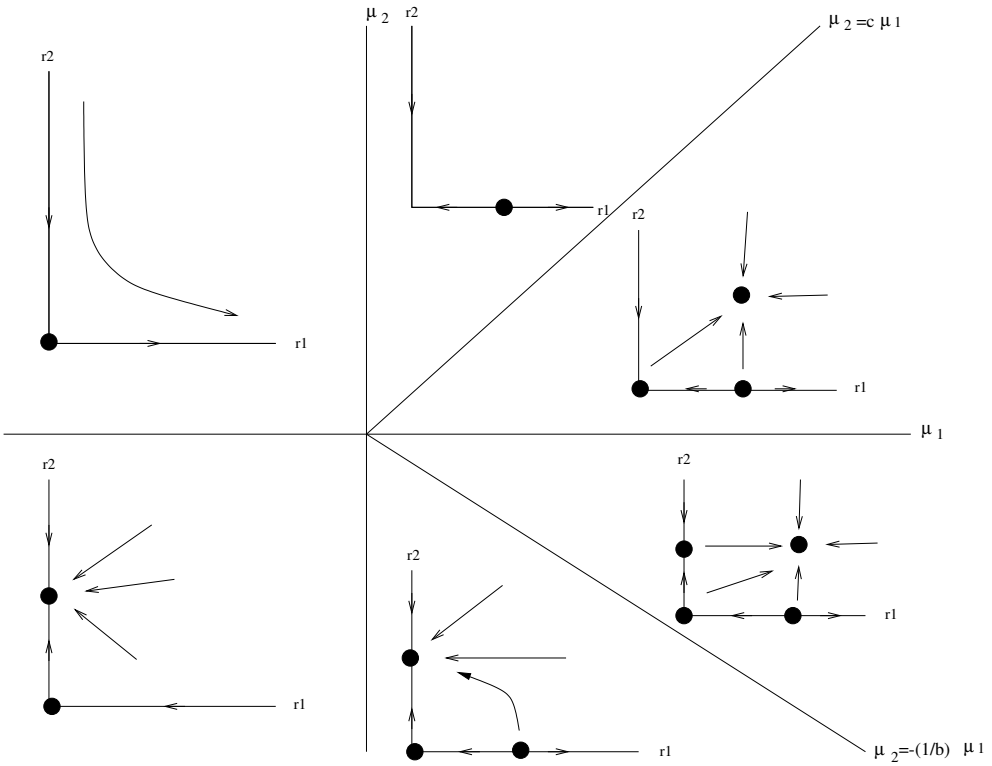


Figure 6.20: Bifurcation set for the SGCM Galerkin model in (μ_1, μ_2)-parameter space. The μ_1 axis corresponds to the linear stability boundary of the subsidiary wave and μ_2 for the principal wave at the double Hopf bifurcation point. A change in value of μ_1, μ_2 corresponds to changes in the Froude number (F) and dissipation (Γ).

It is therefore concluded that near the double Hopf point of interest (DHB_1, which is defined at the intersection of the linear stability curves of the first and second POD pairs) the zonal wavenumbers of wave modes b_1 (associated with the principal wave) and b_3 (associated with the subsidiary wave) interact to generate either stable pure or stable mixed modes[4]. This is illustrated in Figure 6.20 where the wedge in parameter space ($c\mu_1 > \mu_2 > \frac{d\mu_1}{b}$) contains stable quasi-periodic motions and this is what is observed within the interval [SC_1,SC_2] in the mixed-mode regime, illustrated by the AUTO curves in Figure 5.6 and the stability profiles of the periodic orbits in Figure 5.7.

[4]Here, we are considering the behaviour of composite modes, where all the modes are enslaved together. A pure principal or a pure subsidiary mode means that all the modes are travelling as a 1-frequency flow (as was observed in the principal and subsidiary wave regimes). Quasi-periodic behaviour means that the modes are mixed principal-subsidiary waves, as was observed in the mixed-wave regime.

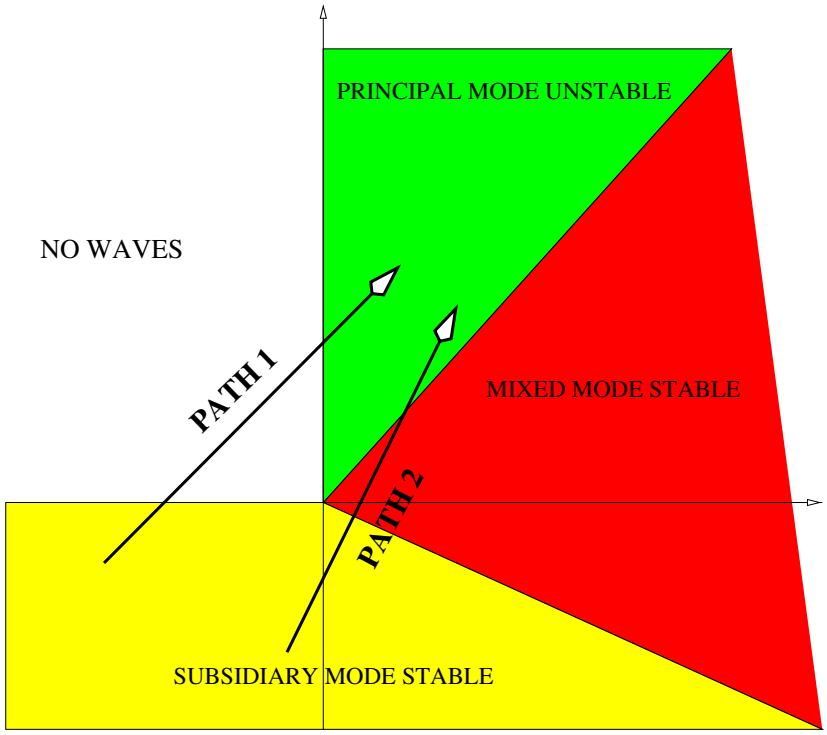


Figure 6.21: Paths in (μ_1, μ_2)-parameter space for increasing μ_1 and μ_2. Figure shows two possible paths (Path 1 and Path 2) which commence in a stable subsidiary regime and expire in an unstable principal regime.

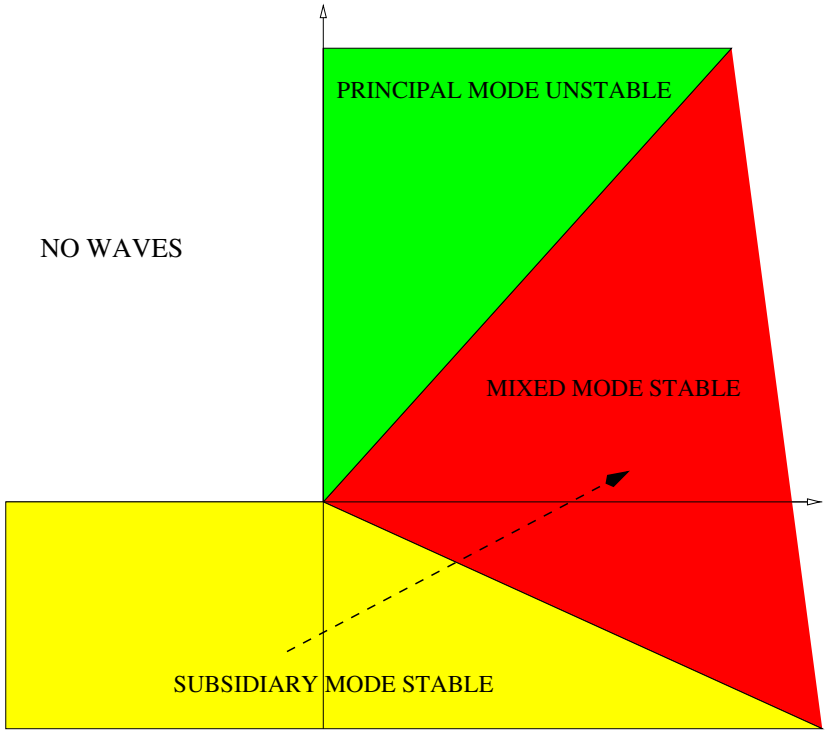


Figure 6.22: Path in (μ_1, μ_2)-parameter space for decreasing Γ and keeping F fixed at a value of 8.14. This corresponds to the bifurcation set described in Figure 6.20

It is noted that there are two optional paths which can be taken in Figure 6.20 and these are illustrated in Figure 6.21 by paths 1 and 2 (where the lines $\mu_2 = c\mu_1$ and $\mu_2 = \frac{d\mu_1}{b}$ separate the different regimes).

PATH 1:
(1). Only the subsidiary mode exists and is stable;
(2). No waves;
(3). Subsidiary mode is unstable.

PATH 2:
(1). Only the subsidiary mode exists and is stable;
(2). Subsidiary mode is stable, principal mode is unstable;
(3). Subsidiary mode is unstable, principal mode is unstable and the mixed-mode is stable;
(4). Principal mode is unstable, the mixed-mode is stable;
(5). Principal mode is unstable.

Recalling our discussion on the stability exchanges found along the various solution curves of the AUTO runs in § 6.2.2, it can easily be seen that Path 2 agrees with our model's behaviour. But it is noticed that stage **(5)** on path 2 is never observed in any of the AUTO or integration runs. There is a simple explanation to this. The analysis of Moroz and Holmes used the linear functions μ_1 and μ_2 which were varied in a neighbourhood of the degenerate bifurcation and depended upon the free parameters F and Γ. This contrasts with our analysis as only Γ is allowed to vary and F is kept fixed (at a value of 8.14). However a path which agrees with what is physically observed can still be identified, and is illustrated in Figure 6.22 (i.e. a route which depicts a stable subsidiary wave changing to a mixed wave).

Collins and James (1995) pointed out that an understanding of the SGCM wave 3 regime required more than linear stability analysis. The occurrence of the various regimes have been partly explained by the nonlinear effects of the double Hopf bifurcation. Such an investigation has been crucial in trying to understand why the low-dimensional SGCM models behave as they do. This then leads to the obvious questions of whether any Hopf and double Hopf bifurcation points exist in the original SGCM, and whether it is their existence that causes the different regimes to occur such as the wave 3 and the weak wave regimes.

Summary

The influential role of the Hopf bifurcation point DHB_1 has been examined and the two-wave interactions and the presence of the double Hopf bifurcation have possibly helped to explain the quasi-periodic motions observed in the Galerkin models. Essentially, the mixed-modes can be viewed as superimposed principal ($\sim$14 cycles/60 sols) and subsidiary ($\sim$5 cycles/60 sols) waves. The dominance of one of these waves has been shown to depend upon the norm and the quality of the vertical resolution.

The two-wave interaction and the presence of the double Hopf bifurcation possibly explains (to a certain extent) the quasi-periodic motions observed in the Galerkin models.

6.6 Conclusions

• The SGCM can be reduced to a low-order system, whilst maintaining the original system's qualitative behaviour. A 6-D model has been shown to be capable of reproducing the frequencies and amplitudes of the large-scale motions of the original system, but unable in producing the complicated distortions which are apparent with the original PCs. The models have been found to be incapable of reproducing any of the transitional behaviour which occurs in the leading SGCM PCs, i.e. a change from 13 to 15 cycles in 60 sols. This skill was expected to be beyond the capability of our low-dimensional models. However, the results of this thesis are very encouraging. Of all the low-order models considered the 4-v 20-D TE model was found to be superior, as it was capable of reproducing the amplitude and frequency of the leading four SGCM PCs (when operating in the mixed-wave regime). The effectiveness of the TE norm over the correlation and KE norms agrees with previous work of Schubert (1985) and Selten (1993). Thus, the spectral-POD-Galerkin technique is an efficient means of extracting the dynamics of the original flow, and consequently the SGCM wave 3 flow has been shown to exhibit 'low-dimensional' behaviour.

• It is believed from the investigations in this thesis that it is essential to accurately model **both** the vertical and horizontal activity of the original system, otherwise it is highly probable that the low-dimensional model will be incapable of providing important information regarding energy conversions, stability exchanges and regime transitions. It has been shown that the amplitudes and frequencies of the modes were robust to the horizontal truncation level. However, the improved vertical scheme clearly affected the behaviour of the modes. The used state vector included both the barotropic and baroclinic components of the flow, and thus resulted in a mixing of the barotropic and baroclinic components during the formulation of the POD modes. It is believed that this has been crucial in extracting important information regarding the vertical activity of the original flow. It was found that the 4-v modes travelled at higher frequencies than their 2-v counterparts, and contained far cleaner subsidiary waves ($\sim$5 cycles/60 sols). The inclusion of an additional two vertical modes resulted in the suppression of the principal ($\sim$14 cycles/60 sols) wave in wave modes b3 and b4, and caused the leading four wave modes to travel at (almost) identical frequencies to those observed in the SGCM.

• The wave 3 is analogous to the principal wave and possibly the weak wave can be identified with the subsidiary wave (i.e. an absent wavenumber 3 flow). This suggests that maybe the process which Collins and James (1995) observed has been reproduced in the ODE sets, i.e. the process of a "messy attractor" (a fixed point) changing to a wave 3 attractor (or limit cycle) via a Hopf bifurcation. It has also been shown that weakly nonlinear theory can explain the existence of the various flow regimes, particularly the mixed-mode wave (i.e. the superimposed principal-subsidiary wave) and thus confirmed that nonlinear phenomena is important in the genesis and maintenance of the wave 3. The double Hopf bifurcation (DHB_1) was found to neighbour the Hopf bifurcation points (HB_1 and HB_2) in the 4-v energy models and suggested that the nonlinear effects (arising from the double Hopf) would be more prominent. The results of this Chapter seem to confirm this since the energy norm modes have been found to be far more erratic and complex (i.e. nonlinear) than those created by the correlation norm.

• The low-dimensional models considered in this thesis are each capable of producing mixed-modes, frequency-locking regimes and limit-cycles. Various flow regimes and transitions have been captured, namely from a subsidiary wave to a quasi-periodic flow and finally to a principal wave regime. This confirms what Read et al. (1992) believed as they thought that sequences of flow transitions could be captured in a low-order model. We have been able to reproduce similar transitions using only a small number of degrees of freedom. This suggests that the dynamics observed in the annulus experiments (such as Früh and Read (1997)) can also be reproduced by these ODE sets. The excellent reproduction of the original system has been shown to be due to (1) an accurate vertical formulation scheme, (2) the use of the correct norm, (3) a sufficiently high level of truncation and (4) the fact that the original system is a steady wave 3 flow.

Chapter 7

Summary and Conclusions

In this thesis, we have constructed low-dimensional models of the SGCM streamfunction which were formulated using the Proper Orthogonal Decomposition (POD) method. The motivation for using PODs is that they effectively describe the most energetic, recurrent structures in the circulation. PODs were initially used to determine a basis for the modal decomposition of climate simulations of Mars obtained using two General Circulation Models, namely the SGCM and MGCM. The SGCM represents an idealised case, in which the primitive equations are solved on a sphere with simplified physical parameters (see Collins and James (1995) for details). The MGCM is a more realistic model, in which a comprehensive range of the relevant Martian physical parameters and topography are represented (see Hourdin et al. (1993) and Collins et al. (1996)). Reduced low-dimensional POD models that replicated the full SGCM streamfunction simulations were formulated by projecting the spherical QG equations onto the PODs of the large-scale calculations by using the Galerkin method. The successful construction of such models used relatively few PODs and were capable of reproducing the amplitude, frequency and behaviour of the leading oscillatory structures of the SGCM within a mixed-mode regime.

Summary of thesis

The governing equations were decided upon after much thought, even though both the primitive equations (Hoskins and Simmons (1975), Bourke (1972)) and the semigeostrophic equations (which were introduced by Eliassen and discussed by Hoskins and Bretherton (1972) and Hoskins (1975)) were also considered. The difficulty was that the semigeostrophic equations could not be formulated on a sphere as the Coriolis parameter was taken to be constant and so did not vary with latitude[1]. Mawson (1996) discussed a semigeostrophic approximation to the 2-D shallow water equations in which the Coriolis parameter f was allowed to vary; spherical geometry was also included. The QG equations with spherical variables were finally chosen (being based on work of Thuburn (1988)). Assuming that the QG assumption would hold in the modelling of the SGCM streamfunction flow was likely to be a possible weakness (if any) in the process of reducing the primitive equation data to a small set of coupled, nonlinear ODEs. In effect it was assumed that any local structures which existed near the equator were

[1]This is an area of extensive research in low-order and balanced equation modelling, where attempts are being made to formulate the semigeostrophic equations on a sphere but the form of the equation set has yet to be finalised.

irrelevant in explaining the dynamical behaviour of the original atmospheric model.

It was shown in Chapter 2 that the standard vertical structure equation (applicable to any planetary atmosphere) arises e.g. from the linearised, adiabatic quasi-geostrophic vorticity equation (where streamfunction is assumed to be vertically separable) as an immediate consequence of an extremum principle which minimises the functional of APE in the vertical direction. Such an extremum principle would seem to describe the natural tendency of an atmosphere to seek states of minimum APE through the release of KE in the form e.g. of baroclinic transient eddies. This principle was illustrated with reference to climate simulations using the SGCM and MGCM. The results implied that the Martian atmosphere places the majority of the APE in the leading order minimum APE modes. The results were presented and discussed in terms of the dominant physical processes responsible for the generation and release of APE in the Martian atmosphere.

The vertical activity of the model could have been alternatively modelled by the use of 3-D PODs which, by definition, would maximise the variance, KE or TE capture of the system, depending upon which norm was used. But if the atmosphere is naturally in a state of APE minimisation in order that KE can be released in the form of baroclinic waves, the optimisation of both APE and KE would be unsatisfactory, since this would not take into account this phenomenon, whereas the use of vertical modes from Flierl's equation (2.20) would model this inherent process relatively naturally. The maximisation of variance or energy capture is therefore not always favourable, particularly if there is a strong presence of baroclinic instabilities in the atmosphere.

Chapter 3 showed how the POD method has been applied to streamfunction fields from both the SGCM and MGCM data sets. The state vector $\mathbf{V}$ was defined to contain either the barotropic and the first baroclinic mode (called '2-v') or the barotropic and the first three baroclinic functions (called '4-v') in order that 41 or 97 % respectively of the APE of the original system could be respectively captured (for details see § 3.3.4). Having already truncated the system vertically, it was important to implement the most effective method in capturing the remaining energy types of the system and so the energy norms were used. It was shown that the PODs are not uniquely defined, for instance they depend on the choice of the inner product. Energy norms were used so as to maximise the amounts of KE and TE retained. An advantage of using a total energy metric over the more often applied Euclidean metric is that it weights contributions from different fields in a dynamically meaningful way; for example, the POD eigenvalues are an indication of the total energy variance represented by each POD.

Discussions were made between the eigenvectors which describe deviations from the mean state (as standard EOFs do), and those which describe the total state vector. It was shown that the KE norm places a large proportion of the extracted KE into the leading order eigenmodes, but insufficiently captures the APE. This can be explained by the fact that truncated models based on KE eigenvectors place very little emphasis on baroclinic patterns, so interactions involving the temperature field are poorly represented resulting in the energy conversion processes not being well described. This was highlighted when it was shown that the last 30 KE eigenvectors

contained mainly APE, and so baroclinic patterns had been poorly extracted in the truncated models which were based on KE eigenvectors, suggesting that the energy conversion processes would not be well described in the low-order models.

The question was addressed as to what extent a low-dimensional POD model could reproduce the dynamics observed in the SGCM. The spectrum of POD eigenvalues was very steep, which indicated the existence of preferred, energetic structures in the SGCM. Modal similarities were a feature of the lower-order SGCM POD modes, where the correlation and energy norms created spatial eigenmodes which were almost identical in amplitude and form. However, structural differences were visible if the higher-order eigenmodes were compared. The leading spatial modes showed evidence of zonal wavenumbers 3 and 2, and an inspection of the corresponding PCs showed that the dominant oscillations of PCs 1 and 2 were roughly sinusoidal 4 day periodic structures (high frequency waves), whereas PCs 3 and 4 for the correlation norm and PCs 3 and 5 for the KE and TE norms were approximately 10 day waves (low frequency waves). The results presented showed that the frequency distribution of the highly regular baroclinic wave activity in the idealised Martian atmosphere was nearly bimodal, i.e. governed by two leading travelling waves. The fact that the analysis picked out two dominant frequency oscillations suggested that similar behaviour could be duplicated using a low-dimensional POD model.

The MGCM simulation was analysed, and the northern winter season ($L_s = 239.6^o$ to 300.42^o) was selected so that the seasonal conditions were comparable to the perpetual southern winter run ($L_s = 90^o$) of the SGCM. The POD eigenspectra revealed one leading eigenvalue, with a steep fall-off, and the bimodal behaviour which was was apparent in the SGCM eigenspectra was no longer observed. It was shown that even though the POD eigenspectra profiles of the various norms were similar, the corresponding eigenvectors were different. The MGCM POD modes were not simple sinusoidal modes, and their surface amplitudes were significantly modulated by the orography. However, the various norms generated different spatial structures in the leading wave modes unlike that which was observed in the SGCM data (where the leading three wave eigenmodes were almost insensitive to the choice of norm). The correlation norm clearly extracted a spatial wavenumber 1 pattern, whereas the energy norms placed a greater emphasis towards smaller-scale structures. It was apparent that zonal wavenumber 1 and 2 dominated the barotropic and baroclinic components of the spatial eigenmodes.

Two distinct oscillatory structures were found in the MGCM PCs, namely the diurnal tide (a high frequency 1 day wave) and a 6-10 day wave (a low frequency baroclinic wave). There was considerable evidence from the behaviour of the PCs that a degree of interaction existed between the high and low frequency waves, a mixing between the forced diurnal tide and the leading 6-10 day baroclinic wave. The behaviour of the spatial and temporal modes resulting from the thermal/baroclinic wave interaction in the MGCM suggested that the reproduction of the implicit complexities of the data in a very low-dimensional POD model was beyond the scope of this thesis.

Chapter 4 discussed a general numerical scheme for constructing low-order models from the

PODs based on the spherical QG equations, given a basis to represent perturbations to that flow. Analytical work was unable to deal with the full nonlinearities which arose in the governing equations and so further progress relied on the use of numerical techniques. PODs were calculated in spectral space (following the work of Schubert (1985)) and were then fitted to a chosen set of basis functions. This was achieved by using a spectral expansion technique, so that each POD mode was fitted to a series of spherical harmonics and Legendre polynomials in order that the spatial eigenmodes lay on the same grid as that of the original data (in T21 form). Particular attention focused on the computation of the derivative fields and the interaction coefficients. It was shown that since the spherical harmonics have the advantage of being the eigenfunctions of the Laplacian then the longitudinal and latitudinal derivatives all had simple forms.

Chapters 5 and 6 discussed the results of the 2-v and 4-v SGCM low-order POD-Galerkin models. The linear stability properties of the basic-state revealed that the most energetic wave (the wavenumber 3 structure) was the most unstable in both the 2-v and 4-v models, and that the energy norms stabilised the linear systems. The nonlinear ODEs (which had extracted the large-scale motions) were numerically integrated and in all models considered three regimes were found, namely a subsidiary, mixed and principal wave regime (for decreasing Γ), and it was within the mixed-mode regime that the modes were similar in shape and frequency to the original SGCM PCs. The quality of the solutions of the low-dimensional models can be attributed to the following features. Firstly, PODs serve as its horizontal basis functions and so ensure that the most energetic patterns of the general circulation model are represented. Secondly, those PODs are calculated using a total energy metric and according to Selten (1993) the TE norm is the only norm capable of describing the global attractor of the original model, and thus capable of simulating properly the various energy conversion processes which exist in the original SGCM data.

It was found that the energy norms had the effect of increasing the frequencies of the Galerkin temporal modes, particularly if the TE norm was used. These changes in the propagation rates were due to the norms picking out different spatial structures and how these projected onto the background flow and the Potential Vorticity gradient. As three different truncation levels and three different norms were used, this provided us with nine different models for comparison. It was observed that in the original SGCM the leading two PCs travelled at a frequency of about 14 cycles every 60 Martian days and that as little as a 6-D TE Galerkin model was required (operating in the mixed-wave regime) to reproduce this frequency (i.e. 14 cycles/60 sols) as well as the amplitudes of the two leading SGCM PCs.

Of all the low-order models considered in this thesis, the 4-v 20-D TE norm model was found to be superior, as it was capable of reproducing the amplitude and frequency of the leading four SGCM PCs (when operating in the mixed-wave regime) with remarkable accuracy, i.e. an error of 0.4% in the frequency of the principal waves (which travelled at a rate of $\sim$14 cycles/60 sols) and an error of 1.7% in the frequency of the subsidiary wave ($\sim$5 cycles/60 sols). Interestingly, it was found that the amplitudes and periods of the waves were insensitive to the horizontal truncation. This may be explained by the suggestion that few modes might

be competing for dominance in the physical regime of the SGCM, and thus the original system is intrinsically low-dimensional. However, the form of the oscillations were affected by vertical and horizontal truncations, particularly the modulation and the erratic disturbances of the individual oscillations.

These findings demonstrated that the original SGCM was in fact low-dimensional and that the spherical QG equations were successful in modelling the wave 3 regime which had been produced by a spectral primitive equation model (using simplified parameterisations for heating and friction). This implies that the retained geostrophic components have been instrumental in reproducing the behaviour of the original SGCM streamfunction flow, and that the neglected nongeostrophic modes have not had an adverse effect upon the behaviour of the Galerkin modes (due to possible distortions in the interactions between the various geostrophic modes). This suggests that the local spatial structures near the equator (which were observed in the POD modes) have not been influential in the modelling of the wave 3 regime even though QG theory breaks down at the equator.

It was thought that if the system was in fact low-dimensional then a thorough examination of the ODE sets may provide us with an understanding of some of the underlying dynamics of the SGCM data. The bifurcation studies of the SGCM ODE sets revealed how a stable steady solution (a point attractor in phase space) became a periodic one (a closed curve $\mathbb{S}^1$) at a Hopf bifurcation, and how the periodic solution became unstable and was succeeded by a stable quasi-periodic solution with two fundamental frequencies (on the surface of a torus $\mathbb{T}^2$) as the parameter Γ varied. The quasi-periodic flow with frequencies f_1 and f_2 say (corresponding to the principal and subsidiary waves) have been detected, but no additional Hopf bifurcation has been encountered which would have resulted in a frequency f_3. Analysing the solutions using AUTO showed that the stable trivial solution lost its stability to a *supercritical* Hopf bifurcation i.e. as Γ decreased, the trivial solution lost its stability to stable periodic orbits. These wave solutions were shown to be robust to the choice of initial conditions. It was also shown that there always existed another Hopf bifurcation for smaller Γ which was *subcritical* in nature.

It was possible to identify low-dimensional attractors present in both the original SGCM and the Galerkin models. Comparisons were made between the original and reconstructed attractors and it was found that these low-dimensional models were accurately picking out the most important dynamical behaviour of the original system and producing comparable attractors. The attractor in the mixed-wave regime behaved essentially like a limit cycle, and were almost identical to that found by Collins and James (1995) from the original SGCM data. This was confirmed by constructing Poincaré sections and by testing its robustness to various initial conditions. In each case the transient trajectories moved towards the limit cycle, finally reaching it and remaining there once the waves had equilibrated.

The importance of the co-dimension 2 point (the double Hopf bifurcation, defined at the intersection of the linear stability curves of the first two POD pairs) upon the behaviour of the waves has been discussed. Collins and James (1995) pointed out that an understanding of the wave 3 regime in the SGCM required more than linear stability analysis. Essentially, the

mixed-modes have been viewed as superimposed principal and subsidiary waves. The influential role of the Hopf bifurcation point has been examined and the two-wave interactions and the presence of the double Hopf bifurcation have explained the quasi-periodic motions observed in the Galerkin models to a certain extent. The occurrence of the various regimes (associated with decreasing Γ) has been explained by the nonlinear effects of the double Hopf bifurcation and the dominance of one of these waves has been shown to depend upon the norm and the vertical resolution. Such an investigation has been crucial in trying to understand why the low-dimensional SGCM models behave as they do. This then leads to the obvious questions of whether any Hopf and double Hopf bifurcation points exist in the original SGCM, and whether it is their existence which causes the different regimes to occur such as the wave 3 and the weak wave regimes.

The question is asked whether the low-dimensional models have revealed the true behaviour of the original SGCM? Two Hopf bifurcations have been identified (taking the Froude number to be F=8.14) which cause the flow to change from a subsidiary wave to a quasi-periodic flow and finally to a principal wave regime (for decreasing Γ). The wave 3 is analogous to the principal wave and possibly the weak wave flow can be identified with the subsidiary wave (i.e. an absent wavenumber 3 flow). This then suggests that maybe the process which Collins and James (1995) observed has been reproduced and identified in our low-dimensional ODE sets, i.e. the process of a fixed point changing to a wave 3 attractor (or a limit cycle) via a Hopf bifurcation and thus confirms the thoughts of Collins and James that the wave 3 flow is low-dimensional.

We further ask whether the low-dimensional models are capable of reproducing any of the regimes observed in the original SGCM? The regime diagram of the SGCM in Figure 3.1 (taken from Collins and James (1995)) illustrated that at the point $(\tau_D, \tau_E) = (3, 2)$ in regime space a wave 3 flow occurs, but if the friction timescale (τ_D) is varied (and keeping the radiative timescale constant at $\tau_E = 2$) then it is noted that either side of the wave 3 regime there exists a block of weak wave flows. The streamfunction data which was fed into the low-order ODE sets was obtained from the $(3, 2)$ setting, and it was observed (Figure 5.7) that for high Γ a subsidiary wave regime occurred (weak wave flow), and as Γ decreased a dominant wave 3 flow would appear. Does this indicate that the low-dimensional POD models are capable of describing flows in a neighbourhood of $(\tau_D, \tau_E) = (3, 2)$ in the original SGCM regime diagram? An ability to confidently answer these questions (i.e. substantiated by further numerical findings) would provide us with a much clearer understanding of how the SGCM flow behaves, and how accurately the SGCM models are performing in reproducing the original flow.

In all our investigations the Froude number of the SGCM has been taken to be $F = 8.14$. Clearly, this fixing of the Froude number is a constraint in our studies. However, this value of F was obtained from the vertical structure equation (2.20), and has been found to give solutions which are qualitatively similar in behaviour to that of the original flow. Nevertheless, low-order models have been generated for other Froude numbers and their solutions observed. It is worth pointing out that in no other model, for any other Froude number, was the behaviour of the Galerkin modes more comparable to the SGCM than those considered in this thesis. In fact,

for Froude numbers > 10, erratic oscillatory structures were observed, which bore no relation to the behaviour seen in the original system.

Comparisons with other work

Similar studies have been made over recent years, each with differing degrees of success. We will now briefly discuss their studies and make comparisons (if possible) with our findings.

Selten (1997) used a three-level hemispheric QG, baroclinic T21 model (described in Marshall and Molteni (1993)) to test how accurately an EOF basis could forecast the northern hemisphere climate in winter. Selten's conclusions on his baroclinic study were that the results were *"not very encouraging"*. He stated that the "*choice of norm certainly deserves more investigation*" and concluded that the poor results were possibly due to selecting the incorrect norm. Selten argued that if a realistic spectral model in three dimensions[2] could be formulated using theoretical basis functions, then it would be worthwhile using EOFs to represent the vertical structure as it has been known since the work of Holmström (1963), that the large-scale vertical structures of the atmosphere can be described by only a few EOFs.

Although the energy norm has been successful in assisting the reproduction of the dynamics of the original SGCM flow, it should be realised that it does not always guarantee better results. If nonenergetic components were critical in the dynamics of the system then these modes would be ignored in a severely truncated model, and as the energy conversion depends on subtle phase differences between the vertical levels of the streamfunction patterns, the crucial structures involved are not necessarily energetically significant. The outcome could then be that the truncated POD model would not correctly describe that part of the attractor of the complete model in which such modes become important. This part of the attractor however, might be highly significant in the overall dynamics of the system. It may even form a bridge between different regimes, and these components are then crucial for the simulation of occasional regime transitions.

Even though the energy norm primarily selects significant energy modes, it has been shown in this thesis that a 6-D TE norm model is capable of producing modes which are accurate in frequency and amplitude. This implies that the modes which contain an insignificant amount of energy do not have a major dynamical effect upon this flow. Comparisons in our studies between the several norms have suggested that the TE norm is the only norm capable of accurately reproducing the various vacillations, besides picking out the dynamically most important modes from the original system.

Recently, Stephen (1998) used a two-level QG Galerkin model to investigate baroclinic wave flows in the thermally-driven rotating annulus experiment, via low-dimensional sets of coupled, nonlinearly ODEs. Stephen suggested that a possible improvement would be to implement the vertical two-mode representation used by Flierl (1978) (as is used in Chapter 5). He claimed

[2]The first attempts to formulate a three dimensional spectral representation were not very successful (Machenhauer and Daley (1974)) .

that the vertical formulation which he adopted was poor and may be the reason why the produced Galerkin modes in his studies were inaccurate in terms of their amplitudes and frequencies.

It has been argued in this thesis that, if the behaviour of the original system is greatly influenced by its vertical activity, then a two-level model may be grossly inadequate in modelling the dynamics. This was demonstrated in Chapter 2 where the first baroclinic mode from the SGCM accounted for only 41.7 % of APE, whereas baroclinic modes 3 and 4 took up a further 31.4 and 17.3 % of APE respectively. Consequently, if a two-level model is used (equivalent to only retaining the barotropic and first baroclinic modes) then effectively nearly 50 % of the system's APE is being ignored. It is therefore not surprising that if such a scheme is implemented then the original dynamics cannot be reproduced by the model equations as modal activity is energy-dependent.

Mundt and Hart (1994) used a two-layer QG f-plane model of baroclinic instability in a straight channel with equal viscosities and layer depths, and a basic zonal current that had even parity about the channel centreline. Their EOF analysis indicated that in the periodic and simpler quasi-periodic regimes a relatively small number of EOFs ($\approx$ 10-30) was required to meet the variance threshold. They found that, for more complex quasi-periodic and chaotic regimes, a larger number of mode pairs were needed to meet the nominal variance requirement, but the first 20 mode pairs were often sufficient in capturing a large proportion of the variance (e.g. 90%). These finds suggested that may be it was possible to construct a low-order dynamical system (of dimension 10-20) that extracted the spatial and temporal behaviour of the entire flow.

Mundt and Hart constructed reduced dimension dynamical models (produced by a Galerkin procedure) and showed that, as the number of active unstable modes increased, the ability of the EOF low-dimensional system to model the full spatio-temporal dynamics efficiently diminished. It was concluded that EOF-based models were most useful in the transition regime from a simple (i.e. steady) flow to chaotic or weakly-turbulent behaviour. However, besides their use in low-order modelling, the EOFs were important in reproducing the travelling wave nature of periodic structural vacillation and the mode-mixing behaviour that occurred for the quasi-periodic solutions. It was concluded that EOF analysis was a useful diagnostic and prognostic tool in moderately complex systems.

The results of Mundt and Hart (1994) are highly comparable with our findings. In this thesis, the mode-mixing regime has been shown to be due to two modes competing for dominance. It has been shown in our work that a 6-D POD model can reproduce periodic and simple quasi-periodic regimes, which also exhibit periodic structural vacillations besides the amplitude-structural vacillation mode mixing behaviour in the mixed-mode regime. It can therefore be concluded that in the cases considered here, the use of POD analysis has been an excellent diagnostic tool in the modelling of a complex system.

The mixed-mode regimes which have been observed in our studies are also very similar to that

which was found by Früh and Read (1997) in their annulus experiments, which further confirmed the robust existence of experimental low-dimensional dynamics. Their studies focused on a flow dominated by a wavenumber 3, concentrating on nonlinear aspects of mixed-mode states at moderate values of the forcing parameters within the regular wave regime. They found four mixed-mode states in well-defined regions of parameter space, two of which were very similar to our findings - namely a nonlinear interference vacillation (defined as a superposition of two steady waves with different zonal wavenumbers and phase speeds) which resulted in quasi-periodic behaviour with a strong phase locking through higher harmonics and an attractor switching flow, where the dominant wavenumber switched at irregular intervals between two possible wavenumbers.

Früh and Read (1997) also observed that, when interference vacillation occurred, the flow lay on a clean $\mathbb{T}^2$ torus (linked to the two frequencies) and it was shown that this broke up to a more 'fuzzy' irregular torus if the stability parameter (or Burger[3] number) was reduced. They stated that when all frequencies were locked[4], the flow reduced to a limit cycle (which became unstable through an incomplete period-doubling cascade). Most of their observed regimes behaved as though governed by low-dimensional dynamics, involving only a limited number of domain-filling modes.

Früh and Read concluded that the multi-mode states arose from nonlinear wave interactions or from the merging of coexisting attractors due to the multiplicity of states in the baroclinic annulus. They said that *"the multi-mode flows can be described as a competition between two or more domain-filling travelling wave modes, resulting in low-frequency modulation of the wave amplitudes. This variability arises from nonlinear coupling of the waves through wave-wave or wave-zonal flow interactions or interactions of the fluid with the boundary conditions, sometimes leading to phase and frequency locking. These ingredients allow for a range of distinct behaviours, at least some of which are consistent with generic bifurcations of low-dimensional systems."*

The findings of Früh and Read are very similar to those which have been discussed in Chapter 6. The mixed-mode has been identified as a superimposed principal and subsidiary wave, each wave fighting for dominance. Wavenumber switching between the two dominant wavenumbers has also been shown to exist within certain dissipation intervals (seen in the principal, mixed and subsidiary wave regimes), as well as attractor switching at regime transitions. It is within the principal and subsidiary wave regimes that the phenomenon of frequency-locking is observed and essentially the flow also reduces to a limit cycle. Limit cycles have also been found to exist in the mixed-wave regime and it has been argued that because the two leading wave modes represent about 80 % of the wave energy then this limit cycle can be viewed as the attractor of the system.

[3]The Burger number (Θ) is alternatively called the thermal Rossby number which is related to the rotational Froude number, and is defined by $\Theta = \frac{g\alpha D\Delta T}{\Omega^2 L^2}$. The Burger number measures the relative strength of the buoyancy forces compared to the Coriolis acceleration. L is the gap width and D the depth of the annulus, ν is the kinematic viscosity, Ω is the rotation rate, ΔT the imposed temperature contrast, g the gravitational acceleration and α is the thermal expansion coefficient of the fluid.

[4]Frequency-locked states were also observed by James et al. (1981) in similar experiments, in which a rotating fluid annulus was subject to a radial temperature gradient.

The formulation of better models

It has been shown in this thesis that purely projected and truncated low-order POD models (without parametrisation of the feedback from unresolved modes) have considerable potential for reductions of standard models of atmospheric behaviour. A problem that had not been addressed is the related closure problem, that is, how the influence of patterns not explicitly resolved in the reduced model could be parametrised as a function of explicitly considered patterns. Thus, in the reduced models the number of degrees of freedom may be still much higher than is theoretically necessary. Selten (1995,1997) was successful in including empirical linear closure schemes, by adding an empirically determined damping for each EOF. The proposition was made that the mean effect of the neglected interactions was to damp the resolved modes and that it can be adequately described by a linear damping. The damping rates for each component were adjusted to compensate for the accumulation of energy due to the neglected interactions.

An important attempt at using a nonlinear closure was made by Selten (1995b,1997b). Selten (1995b) proposed some kind of energy conservation constraints on the closure parameters, which were themselves determined empirically from the full system. These suggestions were followed up by Achatz and Schmitz (1997), and were tested by reducing multiple baroclinic wave life cycles in a QG two-layer model. It was shown that closure greatly improved the simulation capabilities of the reduced model both on short and long timescales. Without exception, it was found that nonlinearly closed models were always superior to non-closed models, both with respect to short-time prediction and longtime climate simulations. In other words, empirical approaches have been found to be viable methods for representing processes which would otherwise not be present in simple models.

Future work

The vertical functions discussed in Chapter 2 are obtained from the vertical structure equation (2.5) with homogeneous boundary conditions $\frac{dH}{dp}|_{p_1} = \frac{dH}{dp}|_{p_n} = 0$. This then implies that the atmosphere is bounded at some upper boundary height. These functions were used as the original SGCM has an artificial 'sponge' top and so the vertical modes were consistent with this condition. However, in reality the atmosphere is infinite and disturbances should remain bounded as z (i.e. altitude) $\rightarrow \infty$. It was shown by Lindzen et al. (1968) that bounded atmospheres reproduce oscillations of the infinite atmosphere with an accuracy which depends upon height. More particularly, they found that the bounded atmosphere models produced spurious free oscillations and they concluded that bounded atmospheres do not properly respond to oscillations which propagate vertically. To address this problem we shall impose a radiation condition instead (at the top of the model) and comparisons will be made between the vertical modes of the bounded and infinite atmospheres. It is intended to calculate what percentage of TE is captured by using only the barotropic and three leading baroclinic modes from the newly revised vertical structure equation, as well as determining what effect their inclusion in the POD state vector has upon the solutions of the low-order ODE sets.

Within the original SGCM there is the conversion of baroclinic and barotropic energy. A

detailed analysis of the energetics within the low-order ODE sets would be interesting. By determining the accuracy of the reproduced energy cycle within the low-order ODE sets a greater understanding of what processes contribute to the evolution of the original SGCM flow field would be attained. This will be achieved by solving the energy equations (Holton (1979)) for the SGCM and the low-order models, allowing comparisons to be made between the two, and establishing whether the evolution of APE and KE in the two models are comparable. From such an investigation it will be found whether the energy conversions are well described in the reduced models, and if not then this will imply that the global behaviour of a truncated POD model is quite different from that of the complete model. By undergoing such a study it will be possible to clarify whether the TE model does in fact model the energy cycle better.

To address the effects of the diurnal tide and topography[5] it would be of interest to produce an MGCM low-dimensional POD model, using the spherical QG equations with additional terms to describe the forced tides as well as a term to represent topography. This would allow direct comparisons to be made between the two ODE sets, and by such a process it would be hoped that an understanding of the most important aspects of the atmosphere's circulation would be made, as well as an appreciation of how the physical processes effect the equation sets. The MGCM data would be restricted to NH mid-winter (for reasons already discussed) so that an almost direct comparison could be made between the SGCM and MGCM ODE sets. However, this raises the question of how the thermal tide with a diurnal disturbance could be represented, since it is not a QG mode. This is important since the diurnal tide provides a periodic perturbation to the mid-latitude circulation, and according to Collins et al. (1996) it causes the flow to flip erratically between two coexisting, otherwise insensitive baroclinic wave regimes.

The spherical QG equations do not include a forcing term, which would mean that a term would have to be included to drive the forced tides. This would obviously be a subject of much consideration, i.e. how should the forced nongeostrophic modes be described? A component of the thermal tide could possibly be represented as a forced Rossby wave which would create a spherical QG response as a forced baroclinic wave. It is recognised that the tides would have to be described by a forced term, migrating towards the west and the Galerkin method must project these tides onto the streamfunction. It would then be essential that the model took account of the interactions between the forced Rossby and barotropic instabilities.

One possible approach in the modal decomposition of the MGCM flow would be to use a similar mean-wave decomposition method as was used in this thesis, so that the POD modes were expressed as a zonally averaged component and a spatially varying departure term. The Galerkin method would be then implemented on both components as usual. A consequence is that the Jacobian and Laplacian operators would be somewhat simplified when operating on the zonal mean terms (due to the lack of longitudinal dependence). By including an axisymmetric forcing term to represent the seasonal diurnal tide, no direct forcing term would have to be applied to the non-axisymmetric part of the expression. This results in the nonlinear components only coupling with the axisymmetric components. With such a method the forcing

[5]Significant orography will certainly modify atmospheric circulation. Since Mars has a considerably greater orography range than that on Earth, then the effects of steady and forced eddies on the global heat transport would have to be carefully modelled (see James (1994)).

terms could be individually applied only to the relevant terms.

The complexity of the governing set of equations which would be used to model the MGCM streamfunction is greater than that used for the SGCM. Not only are there additional forcing terms to describe the thermal tide with a diurnal disturbance, but there is also the problem of modelling the internal heating Q say, of the atmosphere. Initially Q could be modelled so that the heating function is only dependent upon height/latitude so that there is no diurnal heating. Possibly Q could contain a relaxation time such as $Q \sim \frac{\overline{\theta}-\overline{\theta}^*}{\tau}$, where $\overline{\theta}$ is the potential temperature, $\overline{\theta}^*$ is the equilibrium potential temperature and τ is the relaxation time. The question of whether Q should be flow-dependent would have to be addressed in this study.

It has been shown in this thesis that the MGCM PCs contained two definite oscillatory structures, namely a diurnal tide and a 6-10 day baroclinic wave. It was observed that there was considerable interaction between the high and low frequency waves, a mixing between the forced diurnal tide and the leading low frequency baroclinic wave. Collins et al. (1996) commented that the diurnal tide in the MGCM data is spatially resonant with one of the baroclinic wave states, which results in nonlinear interactions occurring between the resonant and non-resonant modes, causing considerable complications in the analysis. It is also said that since an unfolding of the 1:1 spatially resonant mode interaction has not been solved in detail (corresponding to the thermal tide/baroclinic wave interaction) then this study would be difficult. Nevertheless, such an investigation would be worthwhile since Collins et al. (1996) thought that the nonlinear interactions between the baroclinic waves and the thermal tide was influential in determining the *"degree of spatial and temporal irregularity during the northern baroclinic wave season."*

Concluding remarks

Unlike the Earth models of Selten (1997), our baroclinic studies of a planetary atmosphere have been rather successful. However, this may be a result of differences in the dynamical complexities of the two planetary circulation flows, rather than in the way the low-order ODE sets were generated. This agrees with what is observed in reality, since baroclinic disturbances of the Earth's atmosphere are highly irregular (and thus the tropospheric flow is inherently unpredictable), whereas observations from Mars suggest that Martian baroclinic waves are far more regular and predictable, i.e. the atmospheric circulation on Mars is intrinsically lower dimensional than that which is found on Earth. Such differences in the complexities of the two planetary circulations may well explain why Martian low-dimensional models have been successful, and Earth models have not.

In conclusion, the considerations of this thesis have shown a viable approach for extracting data from the SGCM to produce a reduced model on the basis of PODs with an energy metric. The models which were generated were found to give qualitative agreement with the observed dynamics of the SGCM flow. The behaviour of the various regimes observed in the low-order ODE sets were comparable with observations from studies of large-scale waves and instabilities in planetary atmospheres, including a range of hydrodynamical experiments on baroclinic wave interactions in a stratified fluid in cylindrical containers. However, there is plenty of scope for

minor modifications and improvements in the modelling process, beginning with the inclusion of a nonlinear closure scheme which would parameterise the impact of unresolved components of the dynamics. Further studies would centre around the investigation of whether the low-order POD models approximate the energy cycle of the complete model. Closed models may be found to be more powerful than purely projected (non-closed) POD models and thus be more accurate in the reproduction of the SGCM dynamics.

Appendix A

Variational Principle for the POD

The following derivation of the variational principle underlying the POD was given by Berkooz (1991) in his Ph.D thesis.

Specifically, we have a signal $u(x,t)$ or an ensemble of realisations of a field $u(x)$. We restrict our attention to functions which are square integrable, $u(x,t_0) \in L^2(X) \ \ \forall t_0$. Defining the inner product of two functions $\phi(x), \psi(x) \in L^2(X)$,

$$(\phi, \psi) = \int_{-\infty}^{\infty} \phi(x)^* \psi(x) dx, \tag{A.1}$$

and the norm,

$$|\phi|^2 = (\phi, \phi)\,, \tag{A.2}$$

we seek to maximise, on average, the projection of the field onto these functions.

$$\max_{\psi(x) \in L^2(X)} \ \left\langle \frac{(u,\psi)}{(\psi,\psi)} \right\rangle = \left\langle \frac{(u,\phi)}{(\phi,\phi)} \right\rangle, \tag{A.3}$$

where $\langle \cdot \rangle$ denotes time average.

We compute this maximum by applying the calculus of variations. A necessary condition for ϕ to be a maximum is that,

$$\forall\ \psi \in L^2(X) \ \ \frac{d}{d\delta} \ \left. \frac{\left\langle |(u, \phi + \delta\psi)|^2 \right\rangle}{(\phi + \delta\psi, \phi + \delta\psi)} \right|_{\delta=0} = 0. \tag{A.4}$$

Changing the order of averaging and differentiating, expanding the numerator and taking the derivative, we obtain,

$$\begin{aligned}\frac{d}{d\delta}\frac{\left\langle|(u,\phi+\delta\psi)|^2\right\rangle}{(\phi+\delta\psi,\phi+\delta\psi)} &= \left\langle\frac{d}{d\delta}\frac{(u,\phi+\delta\psi)(\phi+\delta\psi,u)}{(\phi+\delta\psi,\phi+\delta\psi)}\right\rangle \\ &= \left\langle\frac{[(u,\phi+\delta\psi)(\psi,u)+(u,\psi)(\phi+\delta,\psi)]\,(\phi+\delta\psi,\phi+\delta\psi)}{(\phi+\delta\psi,\phi+\delta\psi)^2}\right. \\ &\quad\left.-\frac{(u,\phi+\delta\psi)(\phi+\delta\psi,u)\,[(\phi,\phi+\delta\psi)+(\phi+\delta\psi)]}{(\phi+\delta\psi,\phi+\delta\psi)^2}\right\rangle\end{aligned}$$

Evaluating at $\delta=0$ and equating to zero, we obtain

$$\left\langle\frac{[(u,\phi)(\psi,u)+(u,\psi)(\phi,u)]-(u,\phi)(\phi,u)\,[(\psi,\phi)+(\phi,\psi)]}{(\phi,\phi)^2}\right\rangle=0. \tag{A.5}$$

Equating the numerator to zero, writing the inner products as integrals, and noting the result,

$$(u,\phi)(\psi,u)+(u,\psi)(\phi,u)=2Re[(u,\phi)(\psi,u)], \tag{A.6}$$

we obtain:

$$\left\langle 2Re\int_X u^*\cdot\phi\int_X\psi^*\cdot u\int_X\phi\cdot\phi^*-2Re\int_X u\cdot\phi^*\int_X\psi\cdot u^*\int_X\psi\cdot\phi^*\right\rangle=0.$$

Rearranging the integrals by Fubini's theorem, and taking the averaging operators under the integrals to act only on the signal u, we obtain:

$$2Re\left[\int_X\left(\int_X\langle uu^*\rangle\cdot\phi\right)\cdot\phi^*\int_X\phi\cdot\phi^*-\int_X\left(\int_X\langle uu^*\rangle\cdot\phi\right)\cdot\phi^*\int_X\psi\cdot\phi^*\right]=0.$$

A sufficient condition for this to hold is

$$\int_X\langle uu^*\rangle\cdot\phi=\lambda\phi, \tag{A.7}$$

since this gives,

$$2Re\left[\int_X\lambda\phi\cdot\psi^*\int_X\phi\cdot\phi^*-\int_X\lambda\phi\cdot\phi^*\int_X\psi\cdot\phi*\right]=0. \tag{A.8}$$

Appendix B

Floquet Theory

Considering the autonomous ODE of the following form :

$$\dot{\mathbf{x}} = f(\mu, \mathbf{x}), \tag{B.1}$$

where $\mathbf{x} = (x_1, x_2, ..., x_N)^T \in \mathbb{R}^N$ and $\mu = \{\mu_1, \mu_2, ..., \mu_M\}$. Here the dot denotes differentiation with respect to time, $\mathbf{x}$ are the variables and μ are the parameters of the system.

By substituting a Taylor expansion in (B.1) of the form :

$$\mathbf{x} = \mathbf{X} + \mathbf{x}' \tag{B.2}$$

where $\mathbf{X}$ is the fixed point and $\mathbf{x}'$ is the small perturbation. This results into the following set of ODEs :

$$\dot{\mathbf{x}'} = \mathbf{Df}(\mathbf{X})\mathbf{x}' + O(\mathbf{x}'^2), \tag{B.3}$$

where $\mathbf{Df}(\mathbf{X}) = \frac{\partial f_i}{\partial x_j}|_{\mathbf{x}=\mathbf{X}}$ is the Jacobian matrix.

It may be possible that periodic solutions $\mathbf{X}(t)$=$\mathbf{X}(t+T)$ exist to the system of equations (B.1). A method for determining the stability of $\mathbf{X}(t)$ is known as Floquet theory (Drazin (1992)).

We note that the Jacobian matrix $\mathbf{Df}(\mathbf{X})(t)$ has coefficients which are functions of time, having period T. This then means that there must exist N linearly independent solutions to equation (B.3) being of the form :

$$\Phi_j(t) = p_j(t)e^{s_j(t)}, \tag{B.4}$$

where Φ is the $N \times N$ fundamental matrix, defined as :

$$\frac{\partial \mathbf{\Phi}}{\partial t} = \mathbf{Df}(\mathbf{X})(t)\mathbf{\Phi}. \tag{B.5}$$

The Floquet multipliers are the factors $e^{s_j(t)}$. We must note that one Floquet multiplier will

always be equal to unity because this takes account of the perturbations along the direction of the periodic orbit. For the system to be Poincaré stable the remaining Floquet multipliers must have a modulus less than 1, i.e. they must exist within the unit circle. When at least one of the Floquet multipliers leaves the unit circle a bifurcation occurs.

References

[1] R. H. Abraham and C. D. Shaw. *Dynamics : The Geometry of Behaviour. Part 2.* Aerial Press, Santa Cruz, 1983.

[2] 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.

[3] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp. Singular value decomposition and the grassberger-procaccia algorithm. *Phys. Rev. A.*, **38**:3017, 1988.

[4] G. Andrews, D., J. R. Holton, and C. B. Leovy. *Mid Atmospheric Dynamics.*

[5] G. Arfken. *Mathematical Methods for Physicists.* Academic Press, 1985.

[6] N. Aubry. On the hidden beauty of the proper orthogonal decomposition. *Theoret. Comput. Fluid Dynamics*, **2**:339–352, 1991.

[7] N. Aubry, R. Guyonnet, and R. Lima. Spatiotemporal analysis of complex signals: Theory and applications. *J. Statist. Phys.*, **64**:683–739, 1991.

[8] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, **192**:115–173, 1988.

[9] J.R. Barnes. Time spectral analysis of midlatitude disturbances in the martian atmosphere. **37**:2002–2015, 1980.

[10] J.R. Barnes. Midlatitude disturbances in the martian atmosphere: A second mars year. **38**:225–234, 1981.

[11] J.R. Barnes. *Baroclinic waves in the atmosphere of Mars: Observations, linear instability, and finite-amplitude evolution.* PhD thesis, Washington University, 1983.

[12] J.R. Barnes. Linear baroclinic instability in the martian atmosphere. **41**:1536–1550, 1984.

[13] J.R. Barnes. Finite-amplitude behaviour of a single baroclinic wave with multiple vertical modes: Effects of thermal damping. **43**:58–71, 1986.

[14] J.R. Barnes, J. B. Pollack, R. M. Haberle, R. W. Zurek, C. B. Leovy, H. Lee, and J. Schaeffer. Mars atmospheric dynamics as simulated by the nasa ames general circulation model: Ii transient baroclinic eddies. **98**:3125–3148, 1993.

[15] P. Berge, Y. Pomeau, and C. Vidal. *Order with Chaos.* Hermann, 1984.

[16] G. Berkooz. *Turbulence, Coherent Structures and Low-dimensional models.* PhD thesis, Cornell University, 1991.

[17] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Ann. Rev. Fluid Mech.*, **25**:539–575, 1993.

[18] G. Berkooz and E.S. TiTi. Galerkin projections and the proper orthogonal decomposition for equivariant equations. *Phys. Lett. A*, **174**:94–102, 1993.

[19] M. Blackburn. *Program description for the multi-level global spectral model.* University of Reading, Dept. of Meteorology, Atmospheric Modelling Group, 1985.

[20] W. Bourke. An efficient, one-level, primitive-equation spectral model. *Mon. Wea. Rev.*, **100**:683–689, 1972.

[21] B. A. Boville. Strongly nonlinear vacillation in baroclinic waves. *J. Atmos. Sci.*, **39**:1227–1240, 1982.

[22] A. Brandstäter, J. Swift, H. L. Swinney, A. Wolf, Farmer J. D., E. Jen, and P. J. Crutchfield. Low-dimensional chaos in a hydrodynamic system. *Phys. Rev. Letters*, **51**:1442–1445, 1983.

[23] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Physica D*, **20**:217–236, 1986.

[24] D.S. Broomhead, J. P. Huke, and G.P. King. Topological dimension and local coordinates from time series data. *Jour. Phys. A.*, **20**:L563, 1987.

[25] J. A. Brown. A numerical investigation of hydrodynamic instability and energy conversions in the quasi-geostrophic atmosphere: Part i. *J. Atmos. Sci.*, **26**:352–365, 1969.

[26] R. Buizza and T. N. Palmer. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**:1434–1456, 1995.

[27] R. Buizza, J. Tribbia, F. Molteni, and T. N. Palmer. Computation of optimal unstable structures of a numerical weather prediction model. *Tellus*, **45A**:388–407, 1993.

[28] G. Buzyna, R. L. Pfeffer, and R. Kung. Transition to geostrophic turbulence in a rotating differentially heated annulus of fluid. *J. Fluid Mech.*, **145**:377–403, 1984.

[29] B. J. Cantwell. Organised motion in a turbulent. *Ann. Rev. Fluid Mech.*, **13**:457–515, 1981.

[30] J. Carr. *Applications of Centre Manifold Theory.* Springer-Verlag, 1981.

[31] J. G. Charney. The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**:135–162, 1947.

[32] J. G. Charney. Geostrophic turbulence. *J. Atmos. Sci.*, **28**:1087–1095, 1971.

[33] J. G. Charney and F. G. DeVore. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**:1205–1216, 1979.

[34] C. Chatfield. *The Analysis of Time Series: An introduction.* Chapman and Hall, 1980.

[35] C. C. Chen, E . E. Wolf, and H. C. Chang. Low-dimensional spatiotemporal thermal dynamics on nonuniform catalytic surfaces. *J. Phys. Chem.*, **97**:1055–1064, 1993.

[36] M. Collins. *The Meteorology of Mars.* PhD thesis, Reading, 1993.

[37] M. Collins and I.N. James. Regular baroclinic transients waves in a simplified global circulation model of the martian atmosphere. *J. Geophys. Res.*, **100 (E7)**:14421–14432, 1995.

[38] M. Collins, S. R. Lewis, and Read. Regular and irregular baroclinic waves in a martian general circulation model: a role for diurnal forcing? *Adv. Space Res.*, **16 (6)**:3–7, 1995.

[39] M. Collins, S. R. Lewis, P. L. Read, and F. Hourdin. Baroclinic wave transitions in the martian atmosphere. *Icarus*, **120**:344–357, 1996.

[40] B. J. Conrath. Planetary-scale wave structure in the martian atmosphere. *Icarus*, **48**:246–255, 1981.

[41] P. Constantin, C. Foias, B. Nicolaenko, and R. Teman. *Integral manifolds and inertial manifolds for dissipative partial differential equations.* Springer., Berlin, 1988.

[42] P. Courtier and M. Naughton. A pole problem in the reduced gaussian grid. *Qr. J. Roy. Met. Soc.*, **120**:1389–1407, 1994.

[43] J. H. Curry, J. H. Herring, J. Loncaric, and S. A. Orszag. Order and disorder in two- and three-dimensional bénard convection. *J. Fluid Mech.*, **147**:1–38, 1984.

[44] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag. Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders. *Phys. Fluids A*, **3**:2337–2354, 1991.

[45] E. Doedel. AUTO, a program for the automatic bifurcation analysis of autonomous systems. *Cong.Numer.*, **30**:265–284, 1981.

[46] P. G. Drazin. *Nonlinear systems.* Cambridge University Press, 1992.

[47] E.T. Eady. Long waves and cyclone waves. *Tellus*, **1**:35–52, 1949.

[48] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57**:617–656, 1985.

[49] E. Eliasen, B. Machenhauer, and E. Rasmussen. On a numerical method for integrations of the hydrodynamical equations with a spectral representation of the horizontal fields. In *Report No. 2*, 1970.

[50] A. Eliassen. The quasi-static equations of motion. In *Geophys. Pub, 17, no. 3*, 1948.

[51] S. S. Elliot. *Numerical studies of baroclinic instability.* PhD thesis, Oxford, 1995.

[52] G.B. Ermentrout. XPPAUT X-Windows Phase-Plane plus AUTO. ftp://ftp.math.pitt.edu/pub/bardware/tut/start.html, 1998.

[53] C. B. Farmer and P. E. Doms. Global seasonal variation of water vapour on mars and implications for permafrost. *J. Geophys. Res.*, **84**:2881–2888, 1978.

[54] G. R. Flierl. Models of vertical structure and the calibration of two-layer models. *Dyn. Atmos. Oceans.*, **2**:341–381, 1978.

[55] A.C. Fowler and M. J. McGuinness. On the nature of the torus in the complex lorenz equations. *SIAM J. Appl. Math.*, pages 681–700, 1984.

[56] K. Fraedrich. Estimating the dimension of weather and climate attractors. *J. Atmos. Sci.*, **43**:419–432, 1986.

[57] W.-G. Früh and P. L. Read. Wave interactions and the transition to chaos of baroclinic waves in a thermally driven rotating annulus. *Phil. Trans. R. Soc. Lond. A*, **355**:101–153, 1997.

[58] A. M. Gadian. The dynamics of and the heat transport by baroclinic eddies and large scale stationary topographically forced long waves in the martian atmosphere. *Icarus*, **33**:454–465, 1978.

[59] D. Gibbon, J. and J. McGuinness, M. The real and complex lorenz equations in rotating fluids and lasers. *Physica D*, **5**:108–122, 1982.

[60] P. J. Gierash and R. M. Goody. The effect of dust on the temperature of the martian atmosphere. *J. Atmos. Sci.*, **29**:400–402, 1972.

[61] A. E. Gill. *Atmospheric-Oceanic Dynamics.* Academic Press, San Diego, 1982.

[62] G. H. Golub and C. F. Van Loan. *Matrix Computations.* North Oxford Academic, Oxford, 1983.

[63] R. M. Goody and M. J. S. Belton. Radiative relaxation times for mars. *Planet. Space Sci.*, **15**:247–256, 1967.

[64] M. D. Graham and I. G. Kevrekidis. Alternative approaches to the Karhunen-Loève decomposition for model reduction and data analysis. *Computers Chem. Engng.*, **20**:495–506, 1996.

[65] M. D. Graham, S. L. Lane, and D. Luss. Proper orthogonal decomposition analysis of spatitemporal patterns. *J. Phys. Chem.*, **97**:889–894, 1993a.

[66] M. D. Graham, P. H. Steen, and E. S. Titi. Computational efficiency and approximate inertial manifolds for a bènard convection system. *J. Nonlinear Sci.*, **3**:153–167, 1993b.

[67] J. Guckenheimer and G. Buzyna. Dimension measurements for geostrophic turbulence. *Phys. Rev. Lett*, **51**:1438–1441, 1983.

[68] R. M. Haberle, C. B. Leovy, and J. B. Pollack. Some effects of global dust storms on the atmospheric circulation of mars. *Icarus*, **50**:322–367, 1982.

[69] J. E. Hart. The modulation of an unstable baroclinic wave field. *J. Atmos. Sci.*, **33**:1874–1889, 1976.

[70] J. E. Hart. Finite amplitude baroclinic instability. *Ann.Rev.Fluid Mech.*, **11**:147–172, 1979.

[71] J. E. Hart. Wavenumber selection in baroclinic instability. *J. Atmos. Sci.*, **38**:400–408, 1981.

[72] J. E. Hart. A laboratory study of baroclinic chaos on the f-plane. *Tellus*, **37 A**:286–296, 1985.

[73] J. E. Hart. A model for the transition to baroclinic chaos. *Physica D*, **20**:350–362, 1986.

[74] S. L. Hess. Some aspects of the meteorology of mars. *J. Meteorol.*, **7**:1–13, 1950.

[75] R. Hide. Some experiments on thermal convection in a rotating liquid. *Qr. J. Roy. Met. Soc.*, **79**:161, 1953.

[76] R. Hide. An experimental study of thermal convection in a rotating fluid. *Phil. Trans. R. Soc. Lond.*, **A250**:441–478, 1958.

[77] R. Hide and P. J. Mason. Sloping convection in a rotating fluid. *Adv. Phys.*, **24**:47–100, 1975.

[78] P. Hignett, A. A. White, R. D. Carter, W. D. N. Jackson, and R. M. Small. A comparison of laboratory measurements and numerical simulations of baroclinic waveflows in a rotating cylindrical annulus. *Qr. J. Roy. Met. Soc.*, **111**:131–154, 1985.

[79] A. Hollingsworth. Baroclinic instability of a simple flow on the sphere. *Quart. J. R. Met. Soc.*, **101**:495–528, 1975.

[80] I. Holmström. On a method for parametric representation of the state of the atmosphere. *Tellus*, **15**:127–149, 1963.

[81] J. R. Holton. *An Introduction to Dynamic Meteorology.* Academic Press, 1979.

[82] M. Hortal. Formulation of the ecmwf model. **2**, pages 261–280, 1991.

[83] M. Hortal and A. J. Simmons. Use of reduced gaussian grids in spectral models. *Mon. Wea. Rev.*, **119**:1957–2074, 1991.

[84] B. J. Hoskins. The geostrophic momentum approximation of the sem-geostrophic equations. *J. Atmos. Sci.*, **32**:233–242, 1975.

[85] B. J. Hoskins and F. P. Bretherton. Atmospheric frontogenesis models: mathematical formulation and solution.. *J. Atmos. Sci.*, **47**:1714–1720, 1972.

[86] B. J. Hoskins and A. J. Simmons. A multilayer spectral model and the semi-implicit method. *Qr. J. Roy. Met. Soc.*, **109**:771–783, 1975.

[87] F. Hourdin, P. Le Van, F. Forget, and O. Talagrand. Meteorological variability and the annual surface pressure cycle on mars. *J. Atmos. Sci.*, **50**:3625–3640, 1993.

[88] M.Y. Hussaini and T.A. Zang. Spectral methods in fluid dynamics. *Ann. Rev. Fluid Mech.*, **19**:339–367, 1987.

[89] P. Jäckel. Xigra : an interactive ode integration package. 1995.

[90] I. N. James. *Introduction to Circulating Atmospheres.* Cambridge University Press, 1994.

[91] I. N. James and L. J. Gray. Concerning the effect of surface drag on the circulation of a baroclinic planetary atmosphere. *Quart. J. R. Met. Soc.*, **112**:1231–1250, 1986.

[92] I. N. James and P. M. James. Spatial structure of ultra-low-frequency variability of the flow in simple atmospheric circulation model. *Quart. J. R. Met. Soc.*, **118**:1211–1234, 1992.

[93] M. Joshi. *Orographic influences in the atmosphere of Mars.* PhD thesis, Oxford, 1994.

[94] H. Kantz and T. Schreiber. *Nonlinear time series analysis.* Cambridge University Press, Cambridge Nonlinear Science Series 7, 1997.

[95] H. H. Kieffer, B. M. Jakosky, C. W. Snyder, and M. S. Matthews. *Mars.* The University of Arizona Press, Tucson and London, 1992.

[96] G.P. King and I. Stewart. Phase space reconstruction for symmetric dynamical systems. *Physica D*, **58**:216–228, 1992.

[97] M. Kirby. Minimal dynamical systems from pdes using sobolev eigenfunctions. *Physica D*, **57**:466–475, 1992.

[98] P. Klein. Transitions to chaos in unstable baroclinic systems: a review. *Fluid Dyn. Res.*, **5**:235–254, 1990.

[99] P. Klein and J. Pedlosky. A numerical study of baroclinic instability at large supercriticality. *J. Atmos. Sci.*, **43**:1243–1262, 1986.

[100] A. J. Kliore, G. Fjeldbo, B. L. Siedel, M. J. Sykes, and P. M. Woiceshyn. S-band radio occultation measurements of the atmosphere and topography of mars with mariner 9: Extension mission coverage of polar and intermediate latitudes. *J. Geophys. Res.*, **78**:4331–4351, 1973.

[101] K. Krischer, R. Rico-Martinez, I. G. Kevrekidis, H. H. Rotermund, G. Ertl, and J. L. Hudson. Model identification of a spatiotemporally varying catalytic reaction. *A. I. Ch. E. J.*, **39**:89–98, 1993.

[102] C. B. Leovy. Mars: Theoretical aspects of meteorology. *Appl. Opt*, **8**:1279–1286, 1969.

[103] C. B. Leovy. Martian meteorology. *Annual Review of Astronomy and Astrophysics.*, **17**:387–413, 1979.

[104] C. B. Leovy and Y. Mintz. Numerical simulations of the weather and climate of mars. *J. Atmos. Sci.*, **26**:1169–1190, 1969.

[105] S.L. Lewis. Private communication. 1998.

[106] G. F. Lindal, H. B. Holtz, D. N. Sweetnam, Z. Shippony, J. P. Brenkle, G. V. Hartsell, and R. T. Spear. Viking radio occultation measurements of the atsmophere and topography of mars: Data acquired during 1 martian year of tracking. *J. Geophys. Res.*, **84**:8443–8456, 1979.

[107] R. S. Lindzen, E. S. Batten, and J. W. Kim. Oscillations in atmospheres with tops. *Mon. Weath. Rev.*, **96 (3)**:133–140, 1968.

[108] R. S. Lindzen, B. Farrell, and D. Jacqmin. Vacillation due to wave interference: Applications to the atmosphere and to annulus experiments. *J. Atmos. Sci.*, **39**:14–23, 1986.

[109] M. Loève. *Probability Theory.* Van-Nostrand, New York, 1955.

[110] E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. In *Technical Report I, MIT, Department of Meteorology Statistical Forecasting Project.*, 1956.

[111] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**:130–141, 1963a.

[112] E. N. Lorenz. Irregularity: A fundamental property of the atmosphere. *Tellus*, **36A**:98–110, 1984.

[113] A. F. Lovegrove. *Bifurcations and instabilities in rotating two-layer fluids.* PhD thesis, University of Oxford, 1998.

[114] J. L. Lumley. The structure of inhomogeneous turbulent flows. In A. M. Yaglom and V. I. Takarski, editors, *Atmospheric Turbulence and Radio Wave Propagation*, pages 166–178. Nauka, Moscow, 1967.

[115] J. L. Lumley. *Stochastic tools in turbulence.* Academic Press., New York, 1970.

[116] J. L. Lumley. *Transition and Turbulence.* Academic Press, New York, 1981.

[117] B. Machenhauer. Spectral methods. In *Proceedings Numerical Methods in Atmospheric Models*, pages 3–85, 1972.

[118] B. Machenhauer and R. Daley. A baroclinic primitive equation model with a spectral representation in three dimensions. In *Report No. 4*, 1972.

[119] M. Mak. Equilibration in nonlinear baroclinic instability. *J. Atmos. Sci.*, **42**:2764–2782, 1985.

[120] M. Mak. Dissipative structure of a non-linear baroclinic system. *J. Atmos. Sci.*, **44**:2613–2627, 1987.

[121] J. Marshall and F. Molteni. Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.*, **50**:1792–1818, 1993.

[122] M. H. Mawson. A shallow-water semi-geostrophic model on a sphere. *Qr. J. Roy. Met. Soc.*, **122**:267–290, 1996.

[123] S. G. Mikhlin. *Variational Methods in Mathematical Physics.* Pergamon Press, 1964.

[124] J. Milnor. On the concept of the attractor. *Commun. Math. Phys.*, **99**:177, 1985.

[125] I.M. Moroz and J. Brindley. An example of two-mode interaction in a three-layer model of baroclinic instability. *Phys. Lett. A*, **91A**:226–230, 1982.

[126] I.M. Moroz and P. Holmes. Double hopf bifurcation and quasi-periodic flow in a model for baroclinic instability. *J. Atmos. Sci.*, **41**:3147–3160, 1984.

[127] M. D. Mundt, N. H. Brummell, and J. E. Hart. Linear and nonlinear baroclinic instability with rigid sidewalls. *J. Fluid Mech.*, **291**:109–138, 1995.

[128] M. D. Mundt and J. E. Hart. Secondary instability, EOF reduction, and the transition to baroclinic chaos. *Physica D*, **78**:65–92, 1994.

[129] M. Naughton, P. Courtier, and W. Bourke. Representation errors in various grid and spectral truncations for a symmetric feature on the sphere. *Qr. J. Roy. Met. Soc.*, **122**:253–265, 1996.

[130] J. M Nese, J. A. Dutton, and R. Wells. Calculated attractor dimensions for low-order spectral models. *J. Atmos. Sci.*, **44**:1950–1972, 1987.

[131] D. R. Ohlsen and J. E. Hart. The transition to baroclinic chaos on the β-plane. *J. Fluid Mech.*, **203**:23–50, 1989a.

[132] D. R. Ohlsen and J. E. Hart. Nonlinear interference vacillation. *Geophys. Astrophys. Fluid Dyn.*, **45**:213–235, 1989b.

[133] S. A. Orszag. Transform method for calculation of vector coupled-sums : Applications to the spectral form of the vorticity equation. *J. Atmos. Sci.*, **27**:890–895, 1970.

[134] H. Park and L. Sirovich. The proper orthogonal decomposition in the analysis of turbulent flows. *Phys. Fluids A*, **2**:1659–1668, 1990.

[135] R. P. Pearce. On the concept of available potential energy. *Quart. J. R. Met. Soc.*, **104**:737–755, 1978.

[136] J. Pedlosky. Finite-amplitude baroclinic waves. *J. Atmos. Sci.*, **27**:15–30, 1970b.

[137] J. Pedlosky. Finite-amplitude baroclinic waves with small dissipation. *J. Atmos. Sci.*, **28**:587–597, 1971.

[138] J. Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag, Berlin, 1979.

[139] J. Pedlosky. The effect of β on chaotic behaviour of unstable baroclinic waves. *J. Atmos. Sci.*, **38**:717–731, 1981.

[140] C. Penland and M. Ghil. Forecasting northern hemisphere 700-mb geopotential height anomalies using empirical normal modes. *Mon. Wea. Rev.*, **121**:2355–2372, 1993.

[141] R. L. Pfeffer and W. W. Fowlis. Wave dispersion in a rotating differentially heated cylindrical annulus of fluid. *J. Atmos. Sci.*, **25**:361–371, 1968.

[142] N. A. Phillips. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level quasi-geostrophic model. *Tellus*, **6**:273–286, 1954.

[143] J. B. Pollack, D. S. Colburn, F. M. Flasar, R. Kuhn, L. E. Carlston, and D. C. Pidek. Properties and effects of dust particles suspended in the martian atmosphere. *J. Geophys. Res*, **84**:2929–2945, 1979.

[144] J. M. Pollack, C. B. Leovy, P. W. Greiman, and Y. Mintz. A martian general circulation model with large topography. *J. Atmos. Sci.*, **38**:3–29, 1981.

[145] R.W. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, Amsterdam, 1988.

[146] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetlerling. *Numerical Recipes*. Cambridge University Press, 1992.

[147] P. L. Read, M. J. Bell, D. W. Johnson, and R. M. Small. Quasi-periodic and chaotic flow regimes in a thermally driven,rotating fluid annulus. *J. Fluid Mech.*, **238**:599–632, 1992b.

[148] D. Rempfer. Investigations of boundary layer transition via galerkin projections on empirical eigenfunctions. *Phys. Fluids*, **8**:175–188, 1996.

[149] J. Rinne and V. Karhila. A spectral barotropic model in horizontal empirical orthogonal functions. *Qr. J. Roy. Met. Soc.*, **101**:365–382, 1975.

[150] S. D. Schubert. A statistical-dynamical study of empirically determined modes of atmospheric variability. *J. Atmos. Sci.*, **42**:3–17, 1985.

[151] S. D. Schubert. The structure, energetics and evolution of the dominant frequency-dependent three-dimensional atmospheric modes. *J. Atmos. Sci.*, **43**:1210–1237, 1986.

[152] W.D. Sellers. A statistical-dynamical approach to numerical weather prediction. *Sci. Rpt. No. 2, Statistical Forecasting Project, Dept. of Meteorology, MIT, Cambridge, Mass.*, 1957.

[153] F. M. Selten. Towards an optimal description of atmospheric flow. *J. Atmos. Sci.*, **50**:861–877, 1993.

[154] F. M. Selten. An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.*, **52**:915–936, 1995a.

[155] F. M. Selten. *An efficient description of large-scale atmospheric dynamics.* PhD thesis, Amsterdam, 1995b.

[156] F. M. Selten. Baroclinic emprirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.*, **54**:2099–2114, 1997.

[157] F. M. Selten. A statistical closure of a low-order barotropic model. *J. Atmos. Sci.*, **54**:1085–1093, 1997b.

[158] I. Silbermann. Planetary waves in the atmosphere. *J. Meteorol.*, **11**:27–34, 1954.

[159] A. J. Simmons. Development of a high resolution semi-lagrangian version of the ecmwf forecast model. **2**, pages 281–324, 1991.

[160] A. J. Simmons and B. J. Hoskins. Baroclinic instability on the sphere: Normal modes of the primitive and quasi-geostrophic equations. *J. Atmos. Sci.*, **33**:1454–1477, 1976.

[161] A. J. Simmons and B. J. Hoskins. Baroclinic instability on the sphere: Solutions with a more realistic tropopause. *J. Atmos. Sci.*, **34**:581–588, 1977.

[162] A. J. Simmons and B. J. Hoskins. The life cycle of some non-linear baroclinic waves. *J. Atmos. Sci.*, pages 414–432, 1978.

[163] L. Sirovich. Turbulence and the dynamics of coherent structures. part i: Coherent structures. *Quart. Appl. Math.*, **XLV**:561–571, 1987.

[164] L. Sirovich. Chaotic dynamics of coherent structures. *Physica D*, **37**:126–145, 1989.

[165] L. Sirovich, B. W. Knight, and J. D. Rodriguez. Optimal low-dimensional dynamical approximations. *Quart. Appl. Math.*, **48**:535–548, 1990.

[166] L. Sirovich and H. Park. Turbulent thermal convection in a finite domain: Part i. theory. *Phys. Fluids A*, **2**:1649–1658, 1990.

[167] G. R. Smith, E. D. Boerner, R. H. Cohen, and W. A. Newcomb. Empirical orthogonal functions and spectral models of the atmosphere. In *the 1991 American Geophysical Union Fall Meeting (9-13 December, 1991)*, 1991.

[168] A. V. Stephen. *POD Methods in Baroclinic Flows.* PhD thesis, Oxford, 1998.

[169] G. Strang. *An Introduction to Applied Mathematics.* Wellesley - Cambridge, 1986.

[170] P. N. Swarztrauber. On the spectral approximation of discrete scalar and vector functions on the sphere. *SIAM J.Numer.Analy*, **16**:934–949, 1979.

[171] F. Takens. Singularities of vector fields. **43**, pages 47–100, 1974.

[172] H. Tanaka. Global energetics analysis by expansion into three-dimensional normal mode functions during the fgge winter. *J. Met. Soc. of Japan*, **63 (2)**:180–200, 1985.

[173] H. Tanaka, E. C. Kung, and W.E. Baker. Energetics analysis of the observed and simulated general circulation using three-dimensional normal mode expressions. *Tellus*, **38A**:412–428, 1986.

[174] J. E. Tillman, R. M. Henry, and S. L. Hess. Frontal systems during the passage of the martian north polar hood over the viking lander 2 site prior to the first 1977 dust storm. *J. Geophys. Res.*, **84**:2947–2955, 1979.

[175] R. Vautard, P. Yiou, and M. Ghil. Singular-spectrum analysis: A toolkit for short, noise chaotic signals. *Physica D*, **58**:95–125, 1992.

[176] Y. Weng, H, A. Barcilon, and J. F. Magnan. Transitions between baroclinic flow regimes. *J. Atmos. Sci.*, **43**:1760–1777, 1986.

[177] S. Whitehouse, I. M. Moroz, and P. L. Read. An extremum principle for linearised vertical structure equations. In preparation, 1999.

[178] S. S. C. Wu. Mars synthetic topography mapping. *Icarus*, **33**:417–440, 1978.

[179] L. R. Wyatt. Linear and nonlinear baroclinic instability of the northern hemisphere winter zonal flow. *J. Atmos. Sci.*, **36**:2121–2129, 1981.

[180] R. W. Zurek. Free and forced modes in the martian atmosphere. *J. Geophys. Res.*, **93**:652–670, 1988.

Acknowledgements

I would like to thank my supervisors, Dr. I.M. Moroz and Dr. P.L. Read, for their advice, criticism and patience over the past three years. It has been greatly appreciated. Working with them has proven to be an enjoyable, eventful and interesting experience.

Thanks to all the OCIAM graduate students, for providing good company and a stimulating atmosphere in which to work. Special mention is due to Kiat and 'Big' Jim for having to share a room with me at some point in their Oxford sojourn, and to Kiat for his detailed analyses and informative remarks on the previous night's footy game, for warning me monthly on the potential dangers of the millennium bug and for teaching me the powers of the web; Pat, who has helped with all my 'computer concerns', and who was always successful at reaching out into the Martian atmosphere, finding me and then bringing me back down, safely to reality; and to those Radcliffe lunches, which over the months became a great daily tradition; Adam, for all his assistance and advice during this 'POD' experience and Keith for being a fountain of knowledge. A big thank you to Drs. Steve Lewis and Mat Collins for being so patient during those early days, particularly since I had no idea what a GCM was. Thanks also to Mat for proof-reading two of my rather long chapters. I would also like to thank the numerous colleagues who, over the years, have helped and assisted me in my studies. There are too many of these people to name. But I'd particularly like to mention Mr. Wilson, Dr. Johnson and Dr. Bolton from my undergraduate days, back in lovely Durham, who inspired and guided me during those formative years. I also wish to thank EPSRC for their financial support and Pembroke college, for providing funds to attend conferences in Vienna and Nice.

Special thanks are due to many friends, but particularly to Mark and Thad for the great two years I had at Buckingham 'Palace' (I'll never forget that kitchen), and for all those early morning debates on DNA, domestic politics, world affairs and what not - we certainly put the world right; and to Tony, who has been a great friend and the perfect housemate, as well as being the finest of cooks. Our Wednesday nights will live long in the memory. I now turn my attention to my dear friends Sarah and William, who have always given a listening ear to my troubles and woes, even though, at times, they were experiencing minor crises of their own. I'd specifically like to thank Luke, my not so little brother, who has cheerfully put up with me being away from home for too long now. Something good has come of all this since Luke has already decided, at the ripe old age of 11, that mathematics is not for him after closely observing my life, over the last 3 years, as a full-time mathematician. Now Veterinary science awaits him.

I would particularly like to thank my parents for their understanding and encouragement. But for them, I would not have been able to undertake this task. I thank them for allowing me the freedom to pursue my own interests, and for that wonderful experience back at 130, Bethabara.

And finally, I thank my other (and better) half Lindsey, for her love, support and continual encouragement over these long years, as well as reminding me that there's a world outside mathematics and that I'm not too old to venture out into it. We have done this together.