HAL
open science

A simplified model of the Martian atmosphere - Part 2: a POD-Galerkin analysis

S. G. Whitehouse, S. R. Lewis, I. M. Moroz, P. L. Read

To cite this version:

S. G. Whitehouse, S. R. Lewis, I. M. Moroz, P. L. Read. A simplified model of the Martian atmosphere - Part 2: a POD-Galerkin analysis. Nonlinear Processes in Geophysics, 2005, 12 (5), pp.625-642. hal-00302634

HAL Id: hal-00302634

https://hal.science/hal-00302634

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonlinear Processes in Geophysics, 12, 625–642, 2005
SRef-ID: 1607-7946/npg/2005-12-625
European Geosciences Union
© 2005 Author(s). This work is licensed
under a Creative Commons License.

Nonlinear Processes in Geophysics

A simplified model of the Martian atmosphere – Part 2: a POD-Galerkin analysis

S. G. Whitehouse[1], **S. R. Lewis**[2], **I. M. Moroz**[1], **and P. L. Read**[2]

[1]Mathematical Institute, University of Oxford, UK
[2]Atmospheric, Ocanic & Planetary Physics, Clarendon Laboratory, University of Oxford, UK

Received: 9 December 2004 – Revised: 22 March 2005 – Accepted: 11 April 2005 – Published: 17 June 2005

Abstract. In Part I of this study Whitehouse et al. (2005) performed a diagnostic analysis of a simplied model of the Martian atmosphere, in which topography was absent and in which heating was modelled as Newtonian relaxation towards a zonally symmetric equilibrium temperature field. There we derived a reduced-order approximation to the vertical and the horizonal structure of the baroclinically unstable Martian atmosphere, retaining only the barotropic mode and the leading order baroclinic modes. Our objectives in Part II of the study are to incorporate these approximations into a Proper Orthogonal Decomposition-Galerkin expansion of the spherical quasi-geostrophic model in order to derive hierarchies of nonlinear ordinary differential equations for the time-varying coefficients of the spatial structures. Two different vertical truncations are considered, as well as three different norms and 3 different Galerkin truncations. We investigate each in turn, using tools from bifurcation theory, to determine which of the systems most closely resembles the data for which the original diagnostics were performed.

1 Introduction

The dynamics of the weather and climate systems are 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 spatial variation is expanded into a set of n orthogonal basis functions.

The current paper forms the companion paper to a diagnostic analysis (Whitehouse et al., 2005, hereafter denoted as I) of a simplified model of the Martian general circulation (the so-called SGCM), in which topography is excluded. In that paper a low-dimensional description of a baroclinically unstable atmosphere, under conditions appropriate to Mars, was derived via a reduced-order approximation of the vertical and horizontal structure of the system, retaining only the dominant baroclinic and barotropic modes. Our objective here is to derive a suitable set of basis functions as Proper Orthogonal Decomposition (POD) modes from an analysis of the streamfunction data. The POD approach puts most of the variance of the data into the leading order modes. We obtain systems of ODEs using the Galerkin technique in which the dependent variables are expanded as time-varying amplitude coefficients of the spatial POD eigenfunctions. The resulting truncated equations capture the maximum amount of energy among all possible truncations of the same order.

More specifically, suppose the governing equations of the system under consideration take the form

$$\frac{\partial A}{\partial t} = \mathcal{L}A + \mathcal{N}A, \tag{1}$$

where $\mathcal{L}$ is a linear operator, $\mathcal{N}$ is a non-linear operator and A is the spatial and temporal function to be determined. In terms of a POD expansion, A takes the form

$$A(x,t) = \sum_{i=1}^{n} a_i(t)\Phi_i(x), \tag{2}$$

where the $\{a_i(t)\}_{i=1}^n$ are the time-varying amplitudes of the POD modes $\{\Phi_i(x)\}_{i=1}^n$. Defining the inner product of two functions $\phi(x), \psi(x) \in L^2(x)$ as

$$(\phi, \psi) = \int_\Omega \phi(x)^* \psi(x) dx, \tag{3}$$

where $(\phi_i, \phi_j)=\delta_{ij}$, and substituting Eq. (2) into Eq. (1) gives

$$\frac{d}{dt}a_i = \alpha_i + \beta_{ij}a_j + \gamma_{ijk}a_j a_k, \tag{4}$$

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 repeated indices imply a summation and $\gamma_{ijk}=-\gamma_{ikj}$.

Correspondence to: I. M. Moroz
(moroz@maths.ox.ac.uk)

This paper is structured as follows. In Sect. 2 we discuss a general numerical scheme for constructing low-order systems of ODEs based on the spherical quasi-geostrophic (QG) equations. Analytical work was unable to deal with the nonlinearities which arose in the governing equations and so numerical techniques were employed. After the PODs were calculated in spectral space, they were fitted to a chosen set of basis functions, in order that they lay on the same grid as that of the original data.

Section 3 summarises our results with the two-variable (2-v) SGCM POD-Galerkin model, in which just two modes (the barotropic and first baroclinic) are taken to represent the vertical structure of the flow. In I we reported that the first barotropic mode accounted for 89% of the total kinetic energy (KE) with the first baroclinic mode reponsible for just 6.1%. The barotropic mode supplies no available potential energy (APE), while first baroclinic mode accounts for 48.1%.

In contrast, in Sect. 4 we describe our analysis of the four-variable (4-v) SGCM POD-Galerkin model, which comprises the barotropic and the first three baroclinic modes and accounts for 99% of the KE and 97% of the APE of the vertical structure. In both cases linear stability properties of the basic-state revealed that the most energetic wave (the wavenumber 3 structure) was the most unstable and that the total energy (TE) norm (described in I) is the optimum norm for computation of the POD-Galerkin equations. Three regimes were identified in a bifurcation analysis of the ODEs: two primary wave regimes and a mixed mode regime. It was within the mixed-mode regime that the behaviour of the truncated system modes were more closely identifiable with the original SGCM PCs. In addition, we also examined different levels of Galerkin truncation: 6-D, 10-D and 20-D.

2 Derivation of the Galerkin spherical QG model

The adiabatic, spherical quasi-geostrophic equation in isobaric coordinates is

$$\left[\frac{\partial}{\partial t}+J(\psi,\cdot)\right]\left[\nabla^2\psi+f_o^2\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\psi}{\partial p}\right)+f\right]=-r\nabla^2\psi, \quad (5)$$

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

$$J(f,g)=\frac{1}{\cos\theta}\left(\frac{\partial f}{\partial\theta}\frac{\partial g}{\partial\phi}-\frac{\partial f}{\partial\phi}\frac{\partial g}{\partial\theta}\right), \quad (6)$$

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

$$\nabla^2=\frac{1}{a^2\cos^2\theta}\left[\frac{\partial^2}{\partial\phi^2}(\quad)+\cos\theta\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial}{\partial\theta}(\quad)\right)\right]. \quad (7)$$

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, we write

$$\psi(\theta,\phi,t)=\widetilde{\psi}(\theta,\phi,t)+\overline{\psi}(\theta,t)+\widehat{\psi}(\theta), \quad (8)$$

where the tilde denotes the wave mode, the overbar the MFC mode and the hat the time-independent basic state, determined as follows:

$$\widetilde{\psi}=\psi-\int_{\phi=0}^{2\pi}\psi\,d\phi, \quad \overline{\psi}=\int_{\phi=0}^{2\pi}\psi\,d\phi-\widehat{\psi}. \quad (9)$$

The MFC term is the time-dependent difference between the instantaneous zonal flow and the (time-independent) basic state. Although steady, the basic state may be unstable to waves, whose nonlinear interactions feed back to alter the 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. The alternative is to run a 2-D axisymmetric version of the numerical model, suppressing the waves. This was done with the SGCM for 400 sols to ensure that the state had equilibrated and then the resulting state was chosen as the basic state. Our calculations, reported in I, demonstrated very little difference in the barotropic and baroclinic components of these two differently produced states, and so we follow the procedure adopted in I and use the second version as the basic state.

The wave and mean-flow correction perturbations to the basic state are then expanded in orthonormal bases which are derived by the POD method to obtain

$$\psi(\theta,\phi,t)\simeq\widehat{\psi}(\theta)+\sum_{i=1}^{n}a_i(t)\overline{\Phi}_i(\theta)+\sum_{j=1}^{m}b_j(t)\widetilde{\Phi}_j(\theta,\phi), \quad (10)$$

where the overbar denotes the MFC terms, the tildes the wave terms, and the hat the time-independent basic state.

2.1 PODs or spherical harmonics

Selten (1995) notes clear advantages for using spherical harmonics instead of PODs, since both kinetic energy (KE) and enstrophy are conserved for arbitrary truncations when forcing is used. This is because spherical harmonics are eigenfunctions of the Laplace operator and when the Galerkin method is used, the projection operator commutes with the Laplace operator. PODs, on the other hand, are not eigenfunctions of the Laplacian and neither does the projection operator commute. With spherical harmonics, the existence of recurrent structures in the atmosphere are not exploited, whereas the PODs describe the most energetic recurrent structures in the circulation (Selten, 1995).

Since EOFs (or empirical orthogonal eigenfunctions) 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 spherical harmonics. PODs efficiently describe the dominant structures, separating the wave scales. In a truncated spectral model with spherical harmonics, short waves are neglected; a low-order POD model resolves these. The small-scale PODs that are neglected evolve on short time scales and their effect on the large scales is mainly dissipative (see Selten, 1995).

It is difficult to implement a boundary condition which is physically realistic for an unbounded atmosphere. We require the amplitude of the energy of the motion to decay to zero for large z, where z is the vertical height scale. For the energy to be finite as $z\rightarrow\infty$, we require $\rho_s\psi^2$ to remain finite as $z\rightarrow\infty$, where ρ_s is the global average of the fluid density ρ at each level of z. If, however, the amplitude remains finite, then a radiation condition must be used so that all waves have an outward-directed energy flux. Such issues have been addressed by Rempfer (1996). For linear and homogeneous boundary conditions, any linear combination of POD eigenfunctions will satisfy them, if they do so individually. This is not the case for nonhomogeneous conditions.

2.2 On the vertical structure equation

The vertical structure equation, discussed in I, provides a useful identity, making it unnecessary to calculate the term $f_o^2\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\psi}{\partial p}\right)$ in Eq. (5).

We write the streamfunction as

$$\psi(\theta,\phi,p,t)=\sum_{i=0}^{N-1}H_i(p)\widetilde{\psi}_i(\theta,\phi,t), \tag{11}$$

where H_i is the i-th vertical baroclinic mode (i=0 is the barotropic mode), $\widetilde{\psi}_i$ 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. From I, the vertical structure equation in isobaric coordinates takes the form

$$f_o^2\frac{d}{dp}\left(\frac{1}{S}\frac{dH_l}{dp}\right)=-\lambda_l H_l, \tag{12}$$

for each vertical mode, where $a^2\lambda_0$ gives the Froude number of the barotropic mode, $a^2\lambda_l$ for $l\neq0$ that of the l-th baroclinic mode and a=3.394×10^6 m is the radius of the planet. Multiplying both sides of Eq. (12) by $\widetilde{\psi}_l$, summing over N modes and using Eq. (11), we obtain

$$f_o^2\frac{d}{dp}\left(\frac{1}{S}\frac{d\psi}{dp}\right)=-\sum_{l=0}^{N-1}\lambda_l H_l\widetilde{\psi}_l, \tag{13}$$

which is precisely the term arising in the spherical QG Eq. (5). The i-th POD mode, for i=1, ..., n, has the expansion

$$\begin{aligned}\Phi_i(\theta,\phi,p)&=\sum_{k=0}^{N-1}\Phi_i^k(\theta,\phi,p)\\&=\sum_{k=0}^{N-1}\alpha_{ik}H_k(p)\widetilde{\psi}_k(\theta,\phi),\end{aligned} \tag{14}$$

where N is the number of vertical modes; Φ_i^k is the k-th vertical component of the i-th POD mode (including both barotropic and baroclinic wave components); α_{ik} is the nondimensional coefficient of the k-th vertical component and the i-th POD mode; $\widetilde{\psi}_k(\theta,\phi)$ is the amplitude of the time-averaged k-th baroclinic mode. Expanding the total streamfunction field ψ in terms of POD modes, we obtain

$$\begin{aligned}\psi(\theta,\phi,p,t)&=\sum_{i=1}^{n}a_i(t)\Phi_i(\theta,\phi,p)\\&=\sum_{i=1}^{n}a_i(t)\sum_{k=0}^{N-1}\alpha_{ik}H_k(p)\widetilde{\psi}_k(\theta,\phi),\end{aligned} \tag{15}$$

where a_i is the time-dependent coefficient of the i-th POD eigenfunction, so that

$$\begin{aligned}f_o^2\frac{d}{dp}\left(\frac{1}{S}\frac{d\psi}{dp}\right)&=-\sum_{i=1}^{n}a_i\sum_{k=0}^{N-1}\alpha_{ik}\lambda_k H_k\widetilde{\psi}_k\\&=-\sum_{i=1}^{n}\sum_{k=0}^{N-1}a_i\lambda_k\Phi_i^k.\end{aligned} \tag{16}$$

The vertical derivative term in Eq. (5) can therefore be expressed in terms of a POD mode expansion.

The SGCM produces nondimensionalised data and so the dynamical fields and the independent variables of the spherical QG equations must also be made nondimensional to facilitate comparison with the numerical model. This is accomplished by using $t\rightarrow\left(\frac{1}{\Omega}\right)t'$ and $\nabla\rightarrow\left(\frac{1}{a}\right)\nabla'$ where dimensionless variables are denoted by primes and Ω=7.08822×10^{-5} s^{-1} is the rotational rate of the planet.

2.3 The Galerkin projection of the spherical QG equations

Substituting Eq. (13) into Eq. (5) and re-arranging, we obtain

$$\begin{aligned}&\frac{\partial}{\partial t}\left(\nabla^2\psi-\sum_{l=0}^{N-1}\lambda_l H_l\widetilde{\psi}_l\right)=\\&-r\nabla^2\psi-J\left(\psi,\nabla^2\psi-\sum_{l=0}^{N-1}\lambda_l H_l\widetilde{\psi}_l+f\right).\end{aligned} \tag{17}$$

Substituting the series expansion for ψ from Eq. (10) and projecting the MFC basis functions $\overline{\Phi}_k^p$ (where k=1, ..., n) and the spatial wave eigenmodes $\widetilde{\Phi}_k^p$ (for k=1, ..., m) onto the resulting equations, results in a system of n+m-dimensional ODEs, which is conveniently written in matrix form as

$$\mathcal{C}\dot{\mathbf{a}}=\mathbf{f}+\Lambda\mathbf{a}+\mathcal{A}\mathbf{a}+\mathcal{B}\mathbf{a}+\mathcal{D}\mathbf{a}+\mathcal{M}\mathbf{a}+\mathcal{W}\mathbf{a}, \tag{18}$$

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 depend upon the components of the state vector $\mathbf{a}=(a_1,...,a_n,b_1,...,b_m)^T$ (see Appendix for detailed defintions of these matrices).

Schubert (1985) argued that a thermal driving term enters implicitly into the analysis of the atmospheric model, due to

the assumption of a fixed base state, equal to the observed seasonal mean. The constant vector **f** in Eq. (18) describes the dissipation of the basic state. To maintain this state requires that this term be balanced by adding a forcing term equal to $-\mathbf{f}$, which is implemented by neglecting the constant vector. This scheme ensures that the trivial solution **a**=**0** will be a solution of the model equations.

3 The general numerical scheme

The generation of a coupled system of ODEs for the SGCM modelling requires many numerical calculations, which can be summarised as follows:

- 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.

It was necessary to generate the models in the most efficient way since the eigenspectral profiles of the SGCM data, derived in I, revealed that models of between 6 to 20-D were required to capture enough of the system's energy and so create models with over 400 terms (for a truncation level of n, 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, such as Mathematica to calculate the linear stability profiles for each of the wave mode pairs; C integration packages to integrate the nonlinear ODEs and AUTO (Doedel, 1981) to perform bifurcation analyses and to identify any stable solutions.

Each of these applications require the ODEs to be specified in a different format and so an automated procedure for the conversion of equations between the various formats was developed. For example, a parsing program was written to read the Mathematica output (i.e. the generated ODEs) and write a file of equation coefficients. In order to validate our methodology, we tested our codes on a two-layer quasi-geostrophic model of baroclinic instability as formulated by Mak (1985) and found excellent agreement.

3.1 Fitting the POD modes to spherical harmonics

The first task was to fit the streamfunction POD modes to a chosen set of basis functions using a spectral expansion technique. Each POD mode was fitted to spherical harmonics and Legendre polynomials in order that the spatial eigenmodes lay on the same grid as the original data.

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

$$\Phi_i(\theta, \phi, p) = \sum_{m=0}^{20} \sum_{n=m}^{N_m} V_{mni} P_{m,n}(\mu) e^{im\phi}, \tag{19}$$

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, ϕ is the longitudinal coordinate, μ $(=\sin\theta)$ is the latitudinal coordinate, and N_m=21 if m is even or N_m=20 if m is odd. The PODs were fitted using the numerical routine RGTOSP, which was written by one of us (SRL), based upon the method of Machenhauer and Daley (1972) to determine the spherical harmonic coefficients of a field given initially on an evenly spaced latitude-longitude grid.

Naughton et al. (1996) identified two types of error that can result from representing a smooth function on a sphere in terms of finite spherical harmonics: grid representation error and aliasing error. The grid representation error is due to the spherical harmonic series having a finite truncation of spectral coefficients, whereas the aliasing error is associated with the grid representation error and describes the effect of the coarseness of the grid used (due, perhaps, to an incorrect resolution of wavenumbers or their misrepresentation). The grid representation error vanishes provided the grid resolution is sufficient for the spectral coefficients retained. There is also spectral truncation error, caused by finite truncation of the exact spectral series for the function.

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.

3.2 Calculating the derivative fields

An accurate method to evaluate the derivatives of discretised functions is to approximate the functions by series which can be differentiated term by term. Often for this reason spectral space is used. One method which allows the transformation of a field from spectral space to grid space and back again via an inverse transformation was developed independently by Eliasen et al. (1970) and by Orszag (1970). Alternatively, values of the Legendre functions $P_{m,n}$ and their derivatives $(1-\mu^2)\frac{dP_{m,n}}{d\mu}$ can be calculated, with odd and even functions interspersed, in order of increasing n with increasing m, for the staggered triangular truncation.

Having computed the derivative fields in either spectral or grid space, the interaction coefficients, given by the inner product of the POD modes with the various derivative fields are calculated in grid space by applying the trapezoidal rule in both the latitudinal and longitudinal directions. The speed

of the process is greatly enhanced by the use of FFTs and Gaussian Quadrature.

3.3 Analysing the Martian ODE sets

The main restriction with our modelling method is that the Froude number is held fixed for any ODE set. This is because the time derivatives on the LHS in Eq. (18) explicitly depend upon the Froude number, and since the generation of the ODEs requires the numerical inversion of the matrix $\mathcal{C}$, any change in Froude number means that a different set of equations would result. This is why the vertical structure equation analysis of I was so vital, as the eigenvalues automatically calibrate the Froude number in the equations for subsequent use in $\mathcal{C}$, leaving dissipation as the only free parameter.

The ODEs were analysed by a number of numerical packages. Mathematica was used to construct all of the terms in the matrix Eq. (18); a fourth order Runge-Kutta integrator was obtained from Numerical Recipes (see Press et al., 1992) and AUTO was used to identify any periodic orbits and their stability properties. In addition numerical integration packages such as XPPAUT (Ermentrout, 1998) and Xigra (Jäckel, 1995) were also used, each one of which required the files to be written in C.

4 2-vertical mode SGCM Galerkin models

In this section we present the POD-Galerkin analysis of the SGCM model when the barotropic and the first baroclinic mode only are chosen to describe the vertical structure of the flow, and which we term the 2-v model. As mentioned in the introduction, this combination accounts for 95.1% of the total KE but just 48.1% of the total APE, and so represents a severe truncation of the actual vertical structure. We include it since it provides some interesting results, as well as providing a useful comparison with the POD-Galerkin model in which an additional 2 baroclinic modes are included (the 4-vertical mode system) to be discussed in Sect. 6.

We fix the Froude number at F=8.14, the eigenvalue associated with the first baroclinic mode. James (1994) states that for Mars, $R_d \sim 0.33a$, where R_d is the Rossby radius of deformation. This suggests that the scale of the baroclinic waves is large relative to the planetary radius. A global Froude number of 8.14 corresponds to a Rossby radius of deformation of 1.896×10^3 km, and gives $R_d \sim 0.35a$. F=8.14 is therefore a reasonable estimate of the global Froude number.

Collins and James (1995) showed that most of the variance (or energy) of the SGCM was explained by the two leading EOFs, which captured 75% of the variance, with no other EOF explaining more than 3%. They also found that the principal components of these two leading EOFs formed a limit cycle, suggesting that this could be thought of as representative of the system's attractor. 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 therefore evidence that a low-dimensional model of the SGCM (and indeed for the MGCM) is viable.

4.1 The choice of norm

The issue of what is the appropriate norm to use in POD modelling was motivated by the work of Selten (1993) for a two-level QG model of the Earth's atmosphere, formulated in spherical harmonics, truncated at T5 and using EOFs to describe the behaviour of the circulation. Selten found that the global structure could only be described with the TE norm. Both the standard correlation and the KE norms failed to reproduce the global attractor of the original model, since they were incapable of properly simulating the energy conversion processes. Baroclinic waves convert APE into KE and so a TE norm is a natural choice.

If the KE norm is used, the baroclinic patterns are usually poorly represented, which results in interactions involving the temperature field to be poorly represented as well. Baroclinic instability depends upon the existence of a vertical shear of the basic current, which is associated with a horizontal temperature gradient. Horizontal and vertical temperature gradients imply the existence of APE in the basic flow, and forms the energy source for wave growth. If this temperature gradient is poorly extracted, then the resulting model will produce poor results. Indeed when Selten (1993) compared models produced using 23 KE and TE eigenvector patterns, he found that the APE was considerably reduced than for the original system. On the other hand, for the TE norm, the energetics were far better.

Although we computed the EOFs using all three norms in I, and although these same norms were used in the POD-Galerkin modelling, the TE norm was found to be the ideal norm for modelling the baroclinic waves, although comparisons will be made with results obtained using the standard correlation and KE norms.

4.2 Some modelling issues

There are a number of important issues regarding an accurate modelling of experimental data. These include:

- What is the optimum vertical resolution in the POD formulation so that the dynamics of the low-dimensional models faithfully reproduce the original system?
- How does the horizontal truncation level affect the accuracy of the low-dimensional models?
- Which is the best norm?
- How large a fraction of the total energy should ideally be retained in the Galerkin model?

Comparisons were made between models produced from the correlation, KE and TE norms with three different levels of truncation: 20-D, 10-D and 6-D. Here D denotes dimension, namely the maximum number of POD modes retained

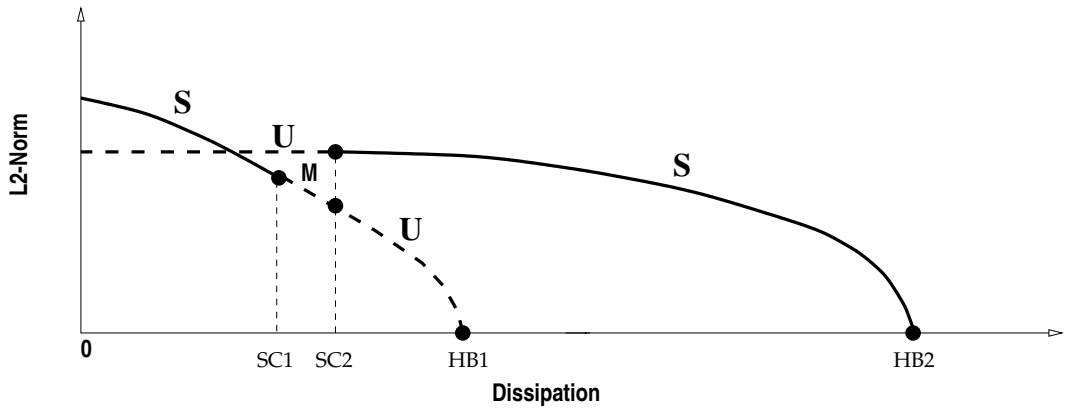

Fig. 1. Stability exchanges along the solution curves produced by AUTO. The figure shows the stable "subsidiary" periodic orbits arising from supercritical Hopf bifurcation HB_2 and the unstable orbits arising from subcritical HB_1. The region of the stable "mixed-mode" is identified by the letter "M" which exists between the subsidiary wave regime (to its right) and the principal wave regime (to its left).

in the Galerkin expansions, thereby providing nine different models for comparison. The dimensions of the models can be understood as follows. For the 6-D system, 2 MFC and 4 wave terms were retained; for the 10-D system, 4 MFC and 6 wave terms were retained; while for the 20-D system, we retained 10 of each.

4.3 The bifurcation analysis

The solutions to the various truncated systems were found using AUTO, beginning with the trivial solution and using dissipation r as the principal bifurcation parameter, starting at large values of r and then decreasing r, noting any bifurcation points encountered. In all cases two Hopf bifurcations (HB) were found and were labelled as HB_1 and HB_2 (the latter occurring at the higher dissipation value). These two were found to occur at different values of r for each of the nine different models. For example the values of r for the 20-D TE norm are $(HB_1, HB_2)=(0.3738, 0.4519)$, while for the 10-D TE system, we found $(HB_1, HB_2)=(0.3732, 0.4507)$. The corresponding values for the KE and the correlation norms were higher (see Whitehouse, 1999, for further details).

The AUTO results (not reproduced here, but see Whitehouse, 1999, for details) showed the 20-D correlation model to yield a larger value for the amplitude of the bifurcating periodic solution than either the 10-D or the 6-D correlation models. The amplitude was larger with both the KE and TE norms. In contrast to Schubert (1985), however, we found very little sensitivity to model truncation in the amplitude and period of the bifurcating limit cycles. 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, or to differences in our spatial eigenfunction formulation, our choice of norm or our vertical resolution scheme. This may indicate that our Galerkin process is very efficient in extracting the important 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.

4.4 Stability properties of the periodic orbits

We use the term "subsidiary periodic orbits or waves" to refer to limit cycles arising from the (supercritical) Hopf bifurcation HB_2. Such waves travel at a rate of about 4–5 cycles every 60 sols). We use the term "principal periodic orbits or waves" to refer to the unstable limit cycles which are produced from HB_1, with a frequency of about 12–14 cycles every 60 sols. A "mixed-mode" type comprises a superimposed "principal" and "subsidiary" wave.

5 The discovery

Figure 1 depicts the primary HB branches and the stability exchanges which occur with AUTO runs where "S" denotes a stable periodic orbit, "U" an unstable periodic orbit and "M" a stable "mixed-mode". Stability was determined by monitoring the Floquet multipliers.

Using dissipation as the control parameter and beginning at high dissipation (r=1) and decreasing, we found that the periodic orbit associated with HB_2 was always stable, while that associated with HB_1, and occurring at a smaller dissipation value, was always unstable. The amplitude of the stable limit cycle increased with decreasing dissipation, the periodic solution losing stability to a stable mixed mode at the point of stability exchange (SC), labelled SC_2 in Fig. 1. The unstable orbit, arising from HB_1, also grows in amplitude as dissipation is decreased, the unstable branch eventually stabilising at the termination of the mixed mode branch at SC_1. For example with the TE norm and the 10-D truncated system we have, in terms of the dissipation parameter values,

$(HB_2, HB_1, SC_2, SC_1)=$

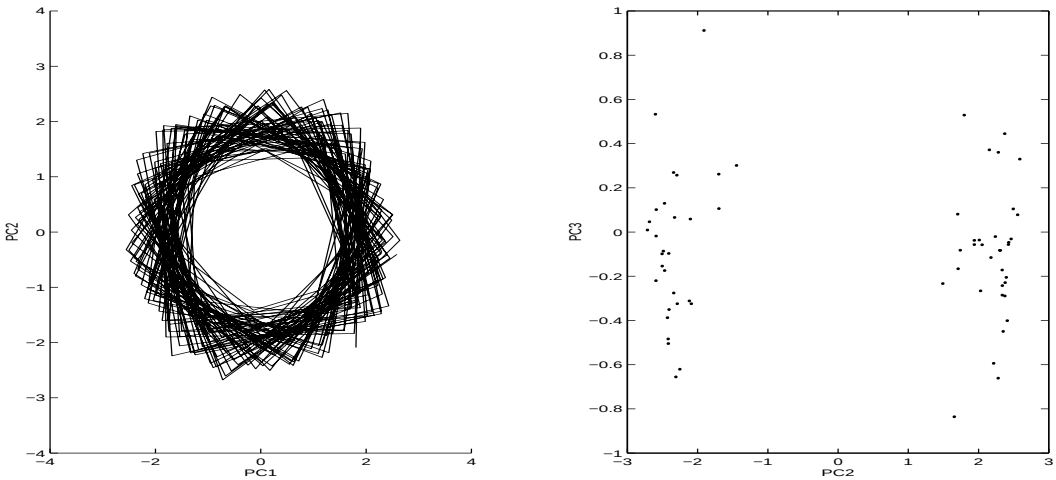

Fig. 2. A comparison of the phase portraits and Poincaré sections (at PC1=0) from the original SGCM data.

$$(0.4507, 0.3732, 0.3213, 0.2973). \tag{20}$$

Furthermore, for all cases investigated we found the following behaviour:

- $r \in$: $[HB_2,\ 1]$ no waves are found;
- $r \in$: $[HB_1,\ HB_2]$ the subsidiary wave is stable;
- $r \in$: $[SC_2,\ HB_1]$ the subsidiary wave is stable and the principal wave is unstable;
- $r \in [SC_1,\ SC_2]$: the subsidiary wave is unstable, the principal wave is unstable and the "mixed" wave is stable;
- $r \in [0,\ SC_1]$: the subsidiary wave is unstable and the principal wave is stable,

where [ ] denotes the interval under consideration.

5.1 Single wave regimes

For convenience, the following notation shall be used in our discussions: "Principal" and "subsidiary" will refer to those oscillations, arising respectively from HB_1 and HB_2, and travelling at rates of 13–14 and 4–5 cycles every 60 sols. The state vector of the system is defined as:

$$\mathbf{a} = [a_1, ..., a_n, b_1, ..., b_m], \tag{21}$$

where the a_i's are the n mean-flow correction terms (MFC modes) and the m b_i's are the wave departure terms (wave modes).

We begin with a discussion of the subsidiary wave regime $[SC_2,\ HB_1]$ at r=0.33, with initial conditions in all of the models taken to be the (unstable) principal wave. In all cases, the system evolved to the stable subsidiary wave after any transients had decayed away. For the 20-D systems, the correlation norm model took about 50 sols (model solar days) to reach its equilibrated amplitude and a further 80 sols to finally equilibrate; the KE norm model took just under 50 sols to reach its correct amplitude and a further 60 sols to stabilise; while the TE norm mode grew to its equilibrated amplitude within 20 sols and took a further 50 sols to completely equilibrate. Equivalent optimal performance of the TE norm was also observed in the 6-D and 10-D systems.

5.2 Phase space and attractor reconstructions

There are several reasons why the evolution of the Martian atmospheric circulation might take place on a relatively low-dimensional attractor.

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 is called the attractor. The phase space trajectories of the system always lie on the attractor. As Mo and Ghil (1987) pointed out, EOFs tend to point from the time mean to the most populated regions of the system's phase space.

Since 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 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 report on the extent to which it is possible to reproduce the global structure of the original attractor using less than 20 POD modes.

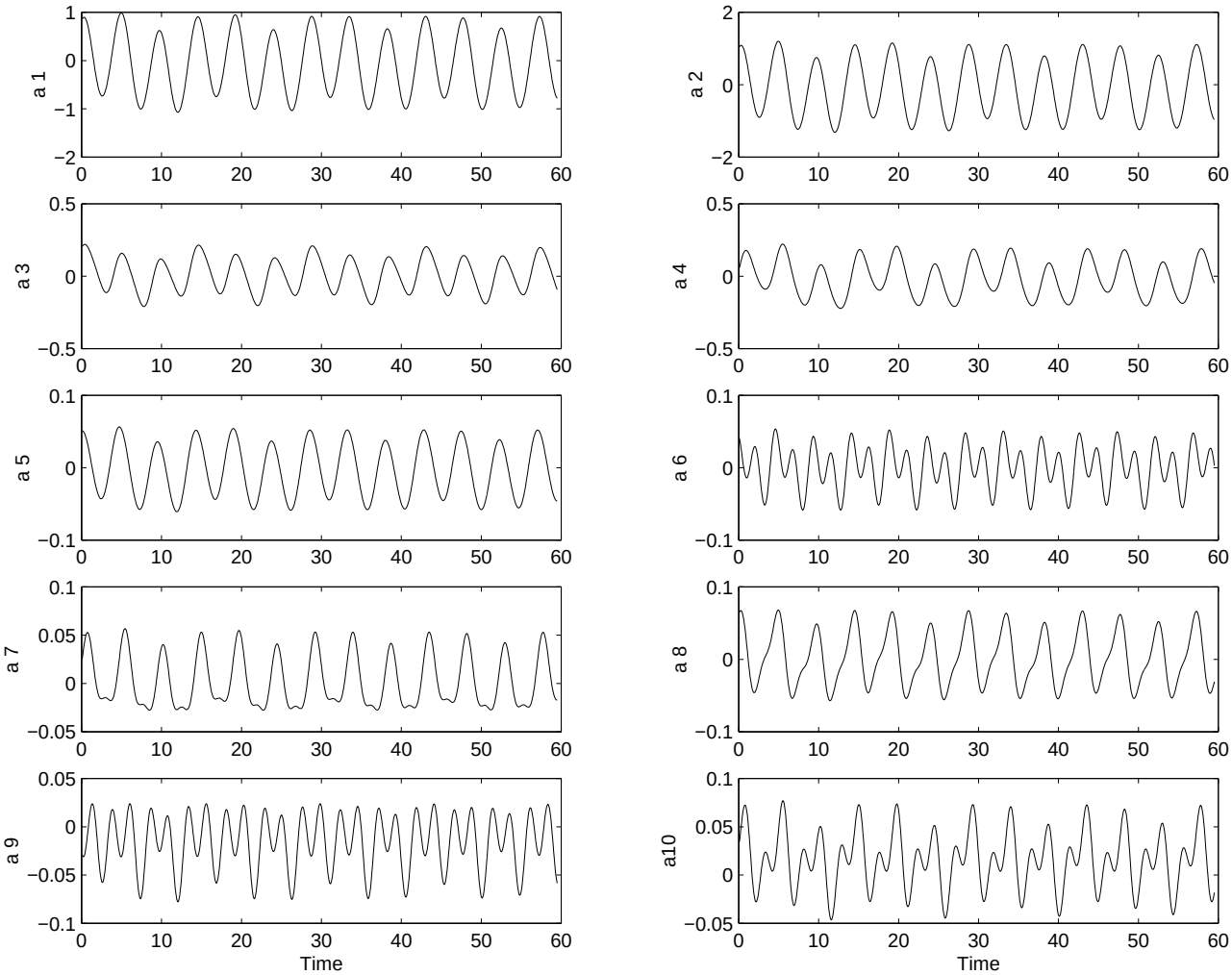

Fig. 3. Amplitudes of the ten MFC modes (a_1, ..., a_{10}) in the mixed-wave regime, r=0.395, from the 20-D correlation norm model.

Figure 2 shows the phase portrait and the Poincaré section (at PC1=0) of the SGCM PCs from the original data, using the TE norm; the equivalent figures for the correlation and KE norms are very similar (Whitehouse, 1999). Since the two leading POD modes explain about 80% of the variance, the trajectories spend most of their time very close to the plane spanned by these two POD modes and trace out the form of a fuzzy circle in (PC1,PC2)-space, as exemplified by the Poincaré section, and substantiating the findings of Collins et al. (1996).

5.3 Mixed-mode regime

We next consider integrations of the 20-D systems within the mixed mode regime of $[SC_1,\ SC_2]$ at dissipation values of r=0.395, 0.35 and 0.34 for the correlation, KE and TE norm models respectively, over 60 Martian sols after discarding an initial 200 sols to remove transient behaviour. We note that the leading wave pairs of the Galerkin models are comparable in amplitude to the original SGCM PCs at these parameter values.

Figures 3 and 4 show the MFC and wave modes of the 20-D correlation norm model. From I, we see that the frequency and amplitude of wave modes b_1 and b_2 are almost identical to those of PCs 1 and 2, travelling at a rate of about 13 cycles per 60 sols. Wave modes b_3 and b_4, similar to b_1 and b_2 in amplitude and shape represent mixed modes, which contradicts the original findings in I that PCs 1, 2 and 3 and 4 respectively behave like principal and subsidiary waves.

Wave modes b_6 to b_{10} propagate at exactly double the frequency of principal modes b_1 and b_2, and travel at a rate of about 26 cycles/60 sols. They are consistent with both the amplitudes and the high frequency behaviour of the higher order PCs. For convenience we shall refer to such waves as "principal-doubled", denoting waves propagating with exactly double the frequency of the principal waves.

The MFC modes differ from the corresponding PCs as the Galerkin modes have definite frequencies unlike the original data patterns, although they do have very similar amplitudes. The leading order modes travel at a rate of 13 cycles/60 sols, whereas the higher order modes propagate at double this rate.

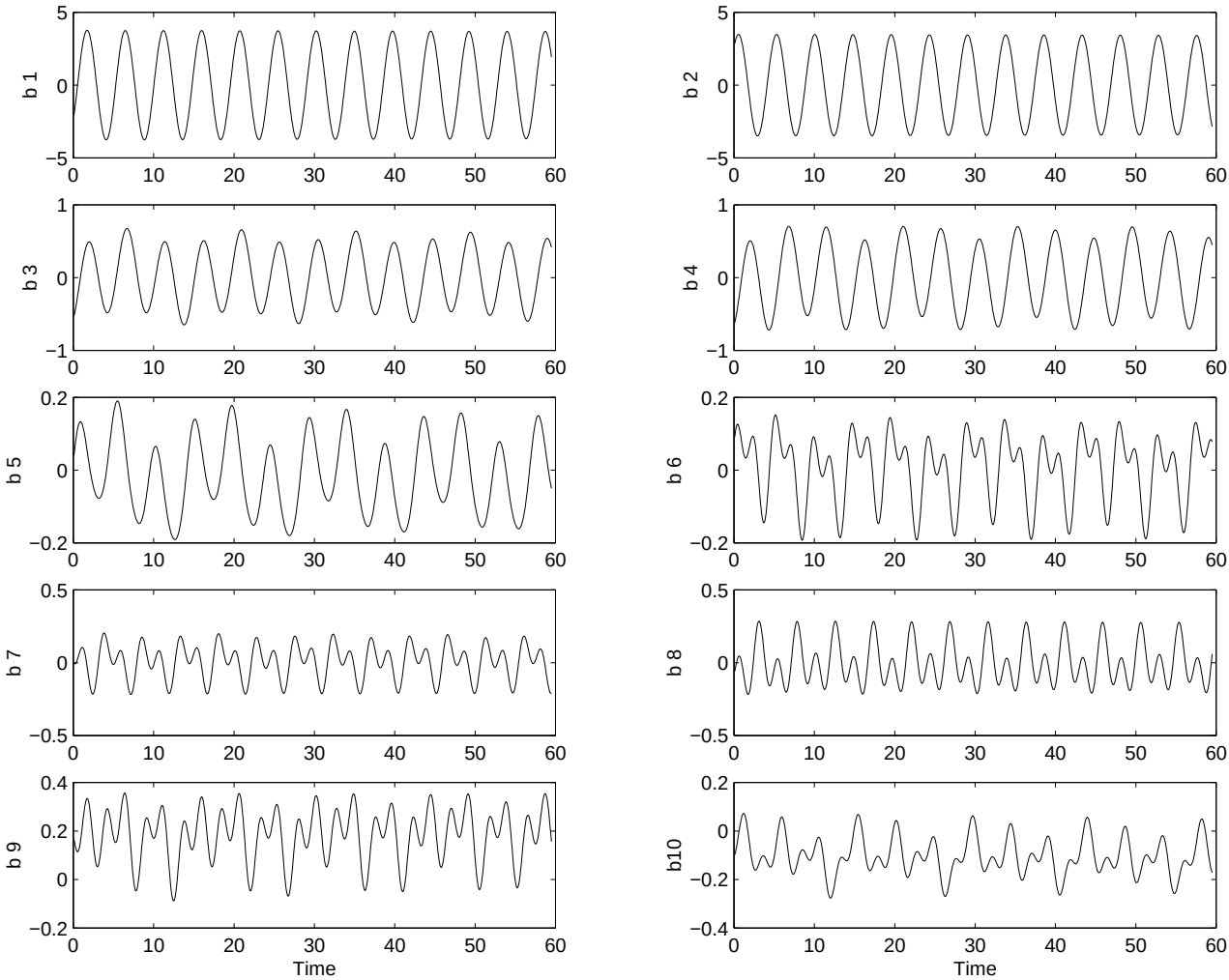

Fig. 4. Amplitudes of the ten wave modes (b_1, ..., b_{10}) in the mixed-wave regime, r=0.395, from the 20-D correlation norm model.

The wave solutions of the 20-D KE model (not shown here), give the two leading wave modes b_1 and b_2 travelling at a rate of over 13 cycles every 60 sols. Modes b_3 and b_4 are mixed modes, having a dominant principal wave with a superimposed subsidiary wave with distortions in b_4 being far more pronounced. The phrase "mixed principle and principal-doubled wave" here refers to a principal wave with a superimposed principal-doubled wave. This demonstrates that the KE norm picks out the subsidiary wave, visible in the original system, and is thereby superior to the correlation model.

Wave mode b_5 is a mixed principal-subsidiary wave, while b_6 is a superimposed principal and principal-doubled oscillation, also observed in the original system (see I and Whitehouse, 1999, for further details), where the original PC 5 contains subsidiary waves but PC 6 behaves in a far more complicated way with a much higher frequency.

The KE MFC modes are very similar to the corresponding wave modes with mixed principal-subsidiary waves, mixed principal and principal-doubled or mixed subsidiary and principal-doubled waves. The behaviour is again different to that of the original MFC PCs for the same reasons outlined above for the correlation norm.

The 20-D TE model substantially improves on the 20-D KE model, see Figs. 5 and 6. 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 b_6 can be seen. The two leading wave modes b_1 and b_2 travel at a rate of over 14 cycles every 60 sols, almost identical to the frequency at which the SGCM PCs 1 and 2 propagate. There is now an obvious presence of the subsidiary waves, travelling at a rate of 5 cycles every 60 sols, in b_3, b_4, b_5 and b_6. The inclusion of APE in the optimisation with the TE model has created striking improvements over the previous two systems, particularly for the higher order wave modes. The two principal modes b_1 and b_2 seem insensitive to changes in these norms. Mode b_6 provides evidence of some degree of interaction between the

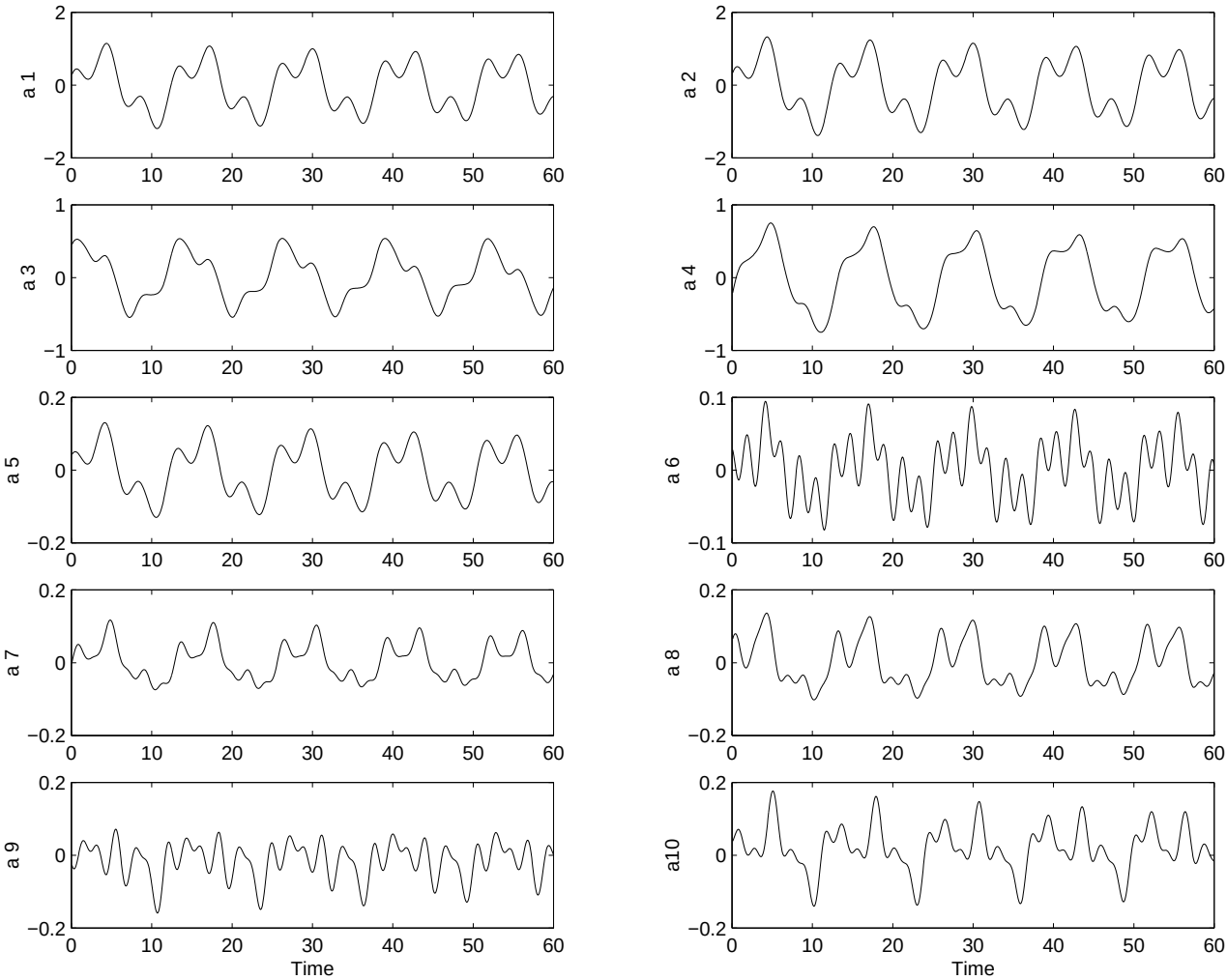

Fig. 5. Amplitudes of the ten MFC modes (a_1, ..., a_{10}) in the mixed-wave regime, r=0.395, from the 20-D TE norm model.

subsidiary wave and a principal-doubled wave. This was observed in the original PCs, which showed a dramatic change in oscillatory structure from PC 5 to 6, the latter being dominated by an erratic, high frequency wave. This high frequency behaviour of b_6 may be a result of the interaction of the subsidiary and principal waves. The amplitudes of the individual TE wave modes correspond accurately to those of the original PCs. Wave distortions in the higher order modes also increase 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 behaviour of the MFC modes in the 20-D TE norm model are similar to those found in the 20-D KE model, the main difference being the strong presence of the subsidiary wave, particularly evident in modes a_1 to a_4. The TE model has 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 shows that there exists a fine balance between the principal and the subsidiary wave considered the dominant mode of the system, with this dominance depending upon the amount of KE and APE retained. Since the original SGCM contains a principal wave in PCs 1 and 2, and a subsidiary wave in PCs 3 and 4, we conclude that the most accurate reproduction is by the 2-v TE model. This suggests that the subsidiary wave in b_3 and b_4 relies upon the optimisation of both KE and APE.

The phase portrait and Poincaré section (at b_1=0) for the TE 20-D Galerkin model at (r=0.34) shows a banded periodic orbit, almost identical to the (PC1,PC2)-space of the original data, shown in Fig. 2, while the corresponding Poincaré section shows a “ribbon” like structure. This provides encouraging evidence that the Galerkin truncation has successfully extracted the principal wave of the original SGCM data. Indeed the two leading POD modes (b_1 and b_2) are almost identical to that found by Collins and James (1995) in Fig. 2. Moreover they are capable of capturing the most significant underlying dynamics of the original system.

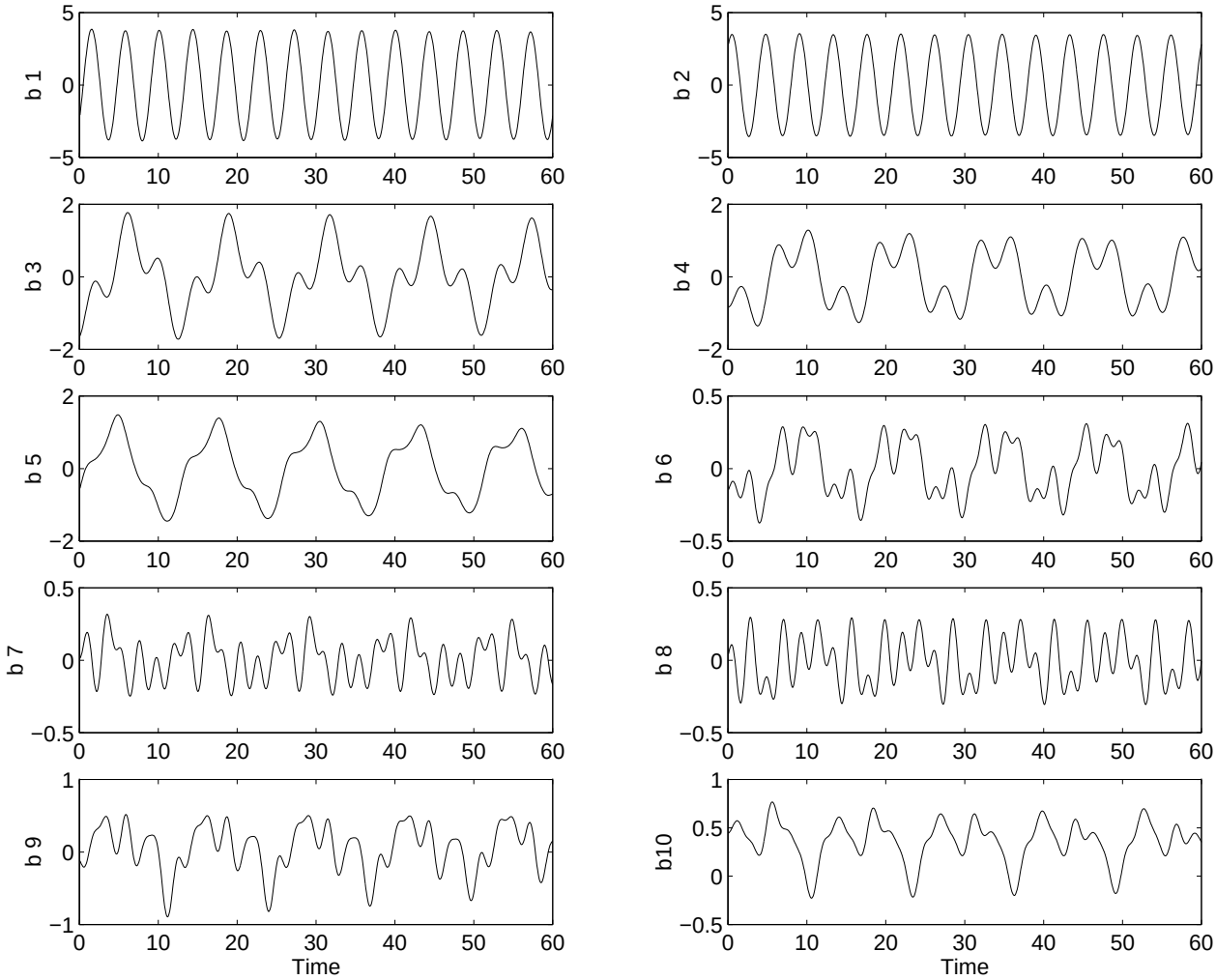

Fig. 6. Amplitudes of the ten wave modes (b_1, ..., b_{10}) in the mixed-wave regime, r=0.395, from the 20-D TE norm model.

5.4 10-D and 6-D Galerkin models

Having investigated how the choice of norm affects the behaviour of the 20-D POD-Galerkin model solutions, we now address the issue of how the level of truncation affects model accuracy. In this paper we shall focus on the TE 10-D and 6-D model for comparison purposes. The reader is referred to Whitehouse (1999) for details of the other truncated systems.

Figure 7 shows wave and MFC modes from the 10-D TE model. In the 10-D model, the wave mode b_6 is no longer distorted, unlike in the 20-D system where it was a mixed subsidiary and principal-doubled wave. Here it is simply a mixed principal-subsidiary wave. The MFC terms have been reordered with modes a3 and a4 in the 10-D system now similar to modes a6 and a9 in the 20-D system. The 6-D TE model in Fig. 8 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 b_3 and b_4) propagate at about 5 cycles/60 sols. However, no high frequency behaviour, which is so evident in the original SGCM, is observed.

The effects of truncation have been to remove the principal-doubled waves, which appear in the 20-D systems, but which cease to exist in the lower dimensional versions.

5.5 Power spectra analysis

For the 10-D POD-Galerkin models, the dominant frequencies of the MFC modes a_1 and a_2 are clustered about the ranges 0.08–0.1 and 0.23–0.25 cycles/sol, while the frequency of the leading wave modes b_1 and b_2 is centred around the 0.23–0.25 cycles/sol range.

The frequencies of the mixed-wave regime compare very well with the frequencies extracted from the original system. The original SGCM PCs cluster about the 0.23–0.25 cycles/sol frequency band, which is also the case here. The most revealing difference between the original system and the truncated models is seen in wave modes b_3 and b_4. 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 ad-

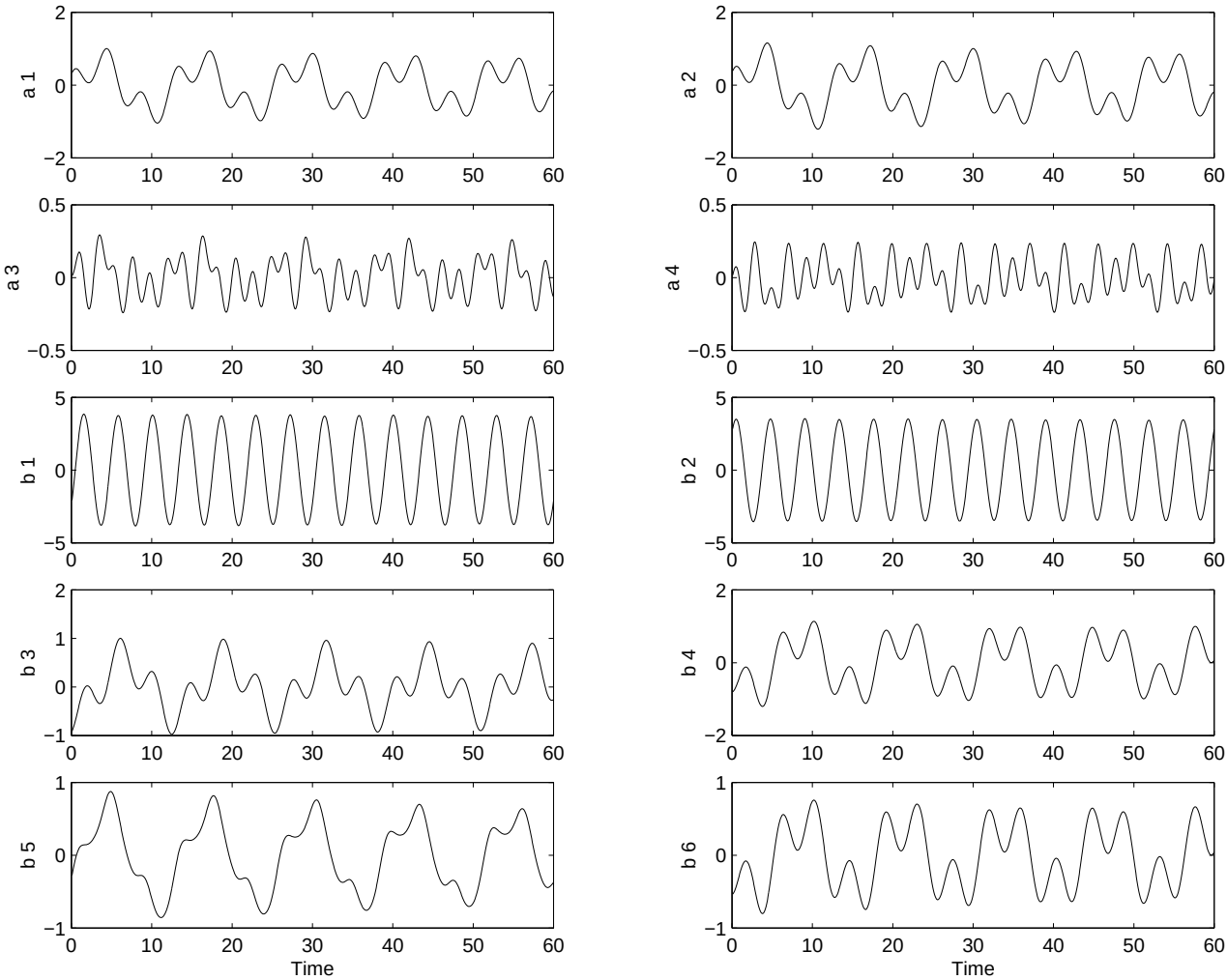

Fig. 7. Amplitudes of the leading MFC (a_1, ..., a_4) and wave modes (b_1, ..., b_6) in the mixed-wave regime from the 10-D TE norm model.

ditional 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, b_1 and b_2.

The energy norms displace the frequencies towards higher values, compared with the correlation norm, bringing them more into line with the values found in the original system.

For wave modes b_3 and b_4, the correlation norm model produces a very small frequency peak at about 0.08 cycles/sol as compared to a much larger peak at 0.23 cycles/sol. This indicates that the principal wave dominates the flow. Similar behaviour is found with the KE norm, the corresponding peaks having larger amplitudes. If the TE norm is used, however, the dominant frequency is at about 0.09 cycles/sol, implying that the TE norm produces the subsidiary wave as the dominant wave. The peak corresponding to the principal wave is also smaller than that produced by the KE norm.

5.6 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 models is clearly affected by the choice of norm. The frequencies of the low-dimensional modes in the mixed-wave regime correspond very well with those of the original system, especially for b_1 and b_2. With the energy norm, the mixed principal-subsidiary oscillatory structures can be changed to a single subsidiary wave, thereby showing the energy norms capacity to extract the first two wave pairs as principal and subsidiary waves respectively and so reproducing the behaviour of the leading four SGCM PCs.

6 4-v POD-Galerkin models

In the previous section we analysed POD-Galerkin models which retained just two vertical modes – the barotropic and first baroclinic – and consequently captured only 48.1% of

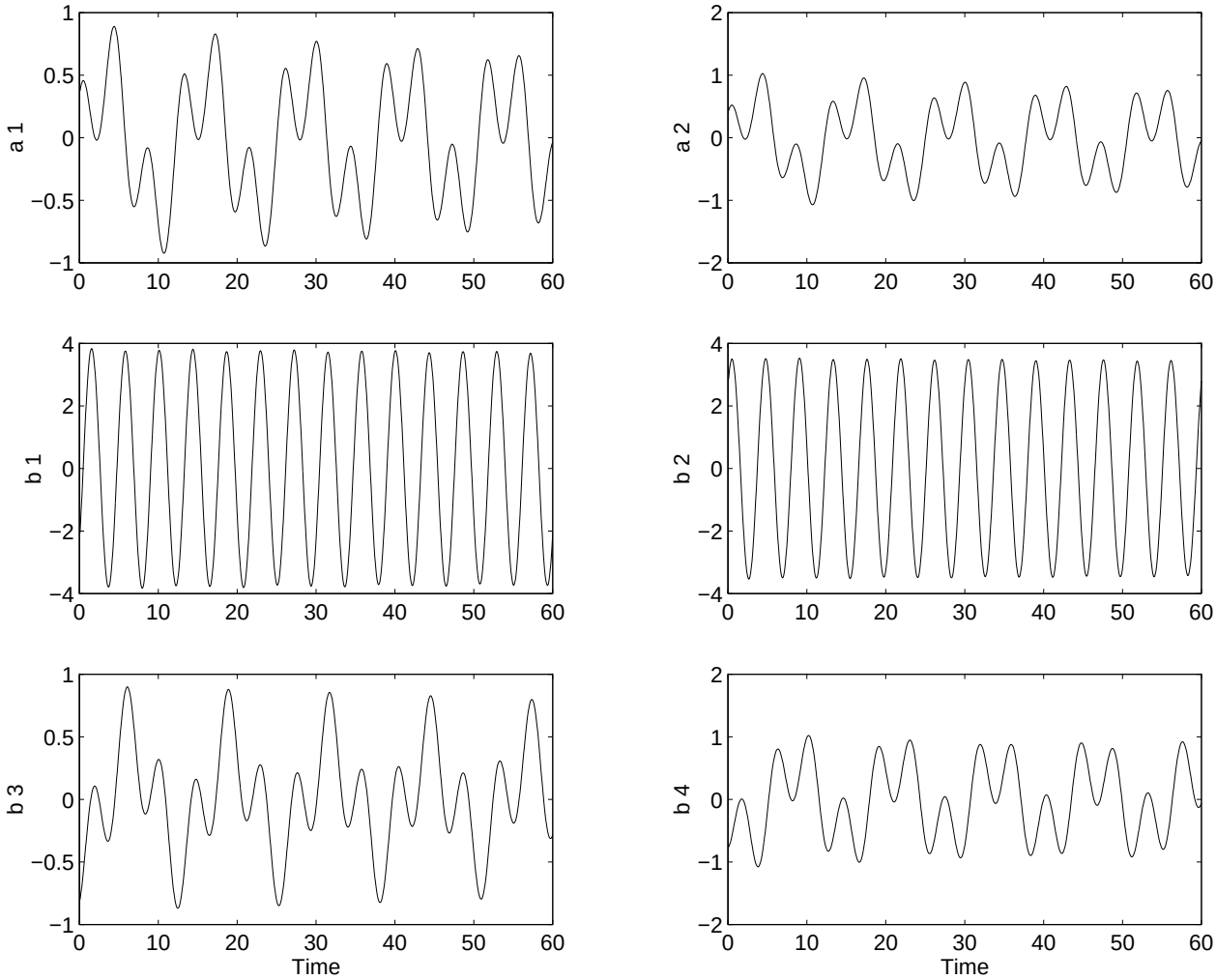

Fig. 8. Amplitudes of the leading MFC (a_1, ..., a_2) and wave modes (b_1, ..., b_4) in the mixed-wave regime from the 6-D TE norm model.

the APE and 95.1% of the KE of the system. In this section we include an additional two vertical baroclinic modes and consider the "4-v" POD-Galerkin models, which now capture 99% of KE and 97% of APE in the vertical. Comparisons will be made with the findings of Sect. 5, including the issue of whether the additional vertical modes improve the accuracy of our modelling scheme. We shall again consider the (2, 4), (4, 6) and the (10, 10) truncations for the correlation, KE and TE norm models, following the same procedure as for the 2-v systems.

6.1 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 case. The mode which dominates the flow is again the wavenumber 3 and is the most unstable mode at the inviscid limit but is the most stable for $\mathcal{O}(1)$ dissipation (see Fig. 9 for the linear stability curves for the TE norm; the corresponding curves with the KE and the correlation norms are similar). The ordering of the linear stability curves is again according to the amount of energy contained in the mode pairs (except for modes 5 and 6).

The additional energy retained in the 4-v models has the effect of moving the stability curves to larger values of Froude number, rising in excess of F=100 for $\mathcal{O}(1)$ dissipation. The first double Hopf bifurcation point DHB_1 (defined as the intersection of the linear stability curves of the first two wavenumber pairs) occurs at F=7.994 and F=7.989 in the 4-v KE and TE models, respectively. Since HB_1 and HB_2 sit on the line F=8.14, the double Hopf bifurcation may play a significant role in the dynamical behaviour of the low-dimensional ODE sets. Moreover these Froude numbers for the double Hopf bifurcations occur considerably closer to F=8.14 than do their 2-v counterparts of F=5.011 and F=4.487 for the KE and TE norms respectively. The presence of the double Hopf may therefore play a more substantial role in the behaviour of the 4-v model than in the 2-v model, and also in reproducing dynamical behaviour which is observed in the original SGCM.

6.2 AUTO integrations

In all AUTO runs we again chose F=8.14. Only two Hopf bifurcations for the trivial solution were found, their locations being shifted from those of the 2-v models. For example with the TE norm, the 20-D system had (HB_1, HB_2)=(0.2568, 0.2827), the 10-D system had (HB_1, HB_2)=(0.2564, 0.2822), while for the 6-D system (HB_1, HB_2)=(0.2563, 0.2820).

The inclusion of two additional baroclinic modes has also caused the frequencies of the leading principal and subsidiary waves to increase. If comparisons are made between the 10-D TE norm models, the frequencies for the 2-v system are (Ω_1, Ω_2)=(0.2609, 0.0922) as opposed to (Ω_1, Ω_2)=(0.2628, 0.0936) for the 4-v system.

The Hopf bifurcations at HB_2 were found to be initially stable and so supercritical. Tracing their continuation with dissipation as the control parameter, the results were similar to those for the 2-v models where the amplitudes of the periodic orbits bifurcating 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 increased in amplitude with decreasing dissipation and the unstable branch finally stabilised at the exchange of stabilities with the mixed mode solutions at SC_1. The solution curves of the AUTO runs were almost identical to those for the 2-v models.

The locations of the various Hopf bifurcations and stability exchanges for the 10-D 4-v TE model are, in terms of the dissipation parameter,

$$(HB_2, HB_1, SC_2, SC_1) =$$

$$(0.2822, 0.2564, 0.2212, 0.1714). \tag{22}$$

The mixed-wave regime exists between the subsidiary (to its right) and principal wave regimes (to its left), although all the bifurcation points occur at smaller dissipation values.

6.3 20-D POD-Galerkin models

For the purpose of comparisons with the 2-v POD-Galerkin models, we focus upon the 20-D truncated systems, referring the interested reader to Whitehouse (1999) for a discussion of the 10-D and the 6-D systems. Each of the integrations were conducted within the mixed mode range of $[SC_1, SC_2]$, at dissipation values of r=0.28, r=0.21 and r=0.22 for the correlation, KE and TE norm models respectively, again over 60 Martian sols after discarding an initial 200 sols. The leading wave pairs of the various Galerkin models are comparable in amplitude to the original SGCM principal components at these selected values.

In the 20-D correlation model, the inclusion of two additional baroclinic modes leads to a cleaner extraction of the subsidiary wave, traveling at about 4 cycles every 60 sols. This is consistent with our previous conclusions that the mixed modes depended upon the presence of a sufficient amount of KE and APE. The actual forms of the wave and MFC modes are almost identical between the 2-v and the 4-v models (and so are not shown here), the frequency of the principal wave at ∼13 cycles every 60 sols remains unchanged.

With the TE norm the improved vertical resolution of the new POD modes produces a dominant subsidiary oscillation of ∼5 cycles every 60 sols in modes b_3 and b_4, with time variations comparable with the original PCs. Mode b_5 is smoother in shape than its 2-v counterpart, suggesting that the improved vertical resolution results in change from it representing a mixed principal-subsidiary wave to a pure subsidiary wave. Smooth subsidiary wave structures are also evident in the MFC modes, especially in MFC modes a_1 to a_5.

The frequencies and amplitudes of modes b_1 and b_2 in the 4-v TE model are almost identical at 14 cycles every 60 sols to those of the original SGCM PCs. In addition modes b_3 and b_4 travel at a frequency of about 5 cycles every 60 sols, as observed in the original system. Wave mode b_5 also has a strong subsidiary wave form, while b_6 has a mixed subsidiary and principal-doubled wave structure, also observed in the original system. Modes b_7 and b_8 behave very similarly to the SGCM PCs, both in terms of their frequencies and their amplitudes.

Both TE modes b_9 and b_{10} are highly distorted subsidiary wave types, i.e. mixed subsidiary-principal as well as a mixed subsidiary and subsidiary-doubled wave, where "subsidiary-doubled" refers to those waves which propagate at double the frequency of the subsidiary waves. Such behaviour was not observed in the original PCs.

6.4 The attractor

Collins and James (1995) showed that by changing the parameters (the radiative and the friction timescales) there exist two flow regimes in the original SGCM, namely a wave 3 flow at (τ_D, τ_E)=(3,2) or a weak wave at (τ_D, τ_E)=(2,2). In our AUTO integrations we identified two Hopf bifurcation point, together with a mixed wave regime, enabling the flow to change from a subsidiary wave to a quasi-periodic flow and finally to a principal wave regime as r is decreased. The wave 3 flow is analogous to the principal wave in our model, while the weak wave flow can be identified with the subsidiary wave.

Collins and James (1995) state 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 low-order POD-Galerkin models. For the KE and TE norm models, the wavenumber 3 attractor takes on the form of a thicker limit cycle, referred to as "fuzzy" by Collins and James (1995), agreeing with the suggestion that the energy models describe the original flow far more accu-

rately because of the emphasis they give to the energy processes within the original system.

The quasi-periodic transients in the work of Mundt and Hart (1994) grew exponentially fast, necessitating an introduction of dissipation of the form $\frac{da_i}{dt} \propto -\rho M^{\alpha} a_i$ to each term, where ρ is an arbitrary scale factor, M is the total wavenumber and α is a scaling constant. In the models discussed here, dissipation was parameterised in terms of Ekman friction and no quasi-periodic solution was ever observed to grow without bound exponentially.

6.5 The dynamical role of the double Hopf bifurcation

We now explore how the close proximity of the double Hopf bifurcation to the single Hopf points at Froude number F=8.14 could explain the transition sequence as well as the dynamics in the models and the SGCM. Linear stability theory shows that it is possible to find points at which two different modes simultaneously lose their stability in a double Hopf bifurcation.

Moroz and Holmes (1984) used a two-layer QG model on a β-plane to investigate whether wave dispersion could be viewed as a consequence of a double-Hopf bifurcation by calculating the normal form (see also Guckenheimer and Holmes, 1983). They argued that wave interactions between the two simple waves could lead either to an unstable mixed wave flow and so to hysteresis, or to a stable mixed wave flow and no hysteresis, depending upon the signs and relative magnitudes of the coefficients in the normal form. If r_j for j=1, 2 denotes the modulus of the amplitudes of the two Hopf bifurcations (writing these amplitudes in modulus-phase form), the following re-scaled set of equations describes the interaction between a pair of waves that are simultaneously marginally stable at a double Hopf bifurcation and in which the frequencies are determined solely by the modulus evolution:

$$\dot{r_1} = \mu_1 r_1 - r_1(r_1^2 + b r_2^2) = f_1(r_1, r_2; b, \mu_1), \tag{23}$$

$$\dot{r_2} = \mu_2 r_2 - r_2(c r_1^2 + d r_2^2) = f_2(r_1, r_2; c, \mu_2), \tag{24}$$

where μ_1, μ_2, b and c are given by the specifics of the problem and d=±1 (see Whitehouse, 1999; Guckenheimer and Holmes, 1983, for further details). In the following we shall identify r_1 with the subsidiary mode and r_2 with the principal mode. There are four equilibrium solutions:

$$(r_1, r_2) = 0, \tag{25}$$

$$(r_1, r_2) = (\sqrt{\mu_1}, 0) \quad \text{for} \quad \mu_1 > 0, \tag{26}$$

$$(r_1, r_2) = (0, \sqrt{\mu_2/d}) \quad \text{for} \quad \mu_2/d > 0, \tag{27}$$

$$(r_1, r_2) = \left(\sqrt{\frac{d\mu_1 - b\mu_2}{A}}, \sqrt{\frac{\mu_2 - c\mu_1}{A}}\right) \quad \text{for} \quad \frac{d\mu_1 - b\mu_2}{A}, \frac{\mu_2 - c\mu_1}{A} > 0, \tag{28}$$

where A=$d-bc$. A stability analysis gives the following:

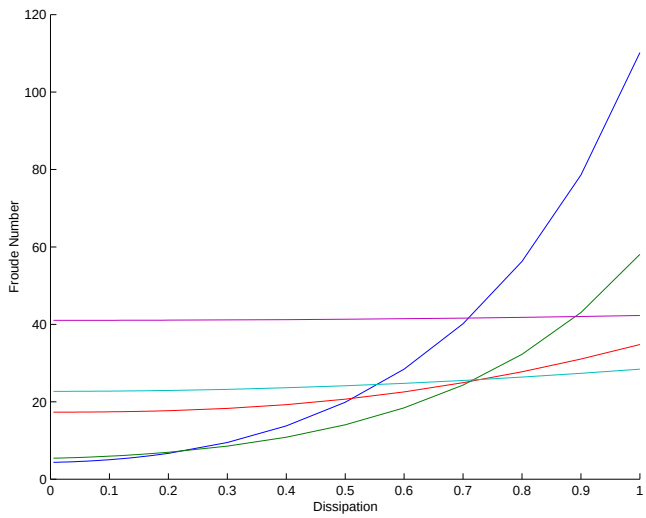

Fig. 9. Linear stability curves for the TE norm model. The dominant energy mode pair (mode 1, mode 2) is the most unstable at the inviscid limit, whilst being the most stable at $\mathcal{O}(1)$ dissipation. Modes 1 and 2 are shown in navy, modes 3 and 4 in green, modes 5 and 6 in red, modes 7 and 8 in light blue and modes 9 and 10 in purple.

- Equation (25) is always an equilibrium, corresponding to a state of no waves when the frequencies are taken into account. It is a stable sink if μ_1, μ_2<0; a saddle if μ_1>0>μ_2 or μ_2>0>μ_1 and an unstable source if μ_1, μ_2>0.

- Equation (26) is an equilibrium and corresponds to a periodic solution in the full system. If in addition $c\mu_1$>μ_2 it is a sink; otherwise it is a saddle.

- Equation (27) is an equilibrium and corresponds to a periodic solution in the full system. Its stability is determined by the two eigenvlaues $-2\mu_2$ and $\mu_1 - b\mu_1/d$. To exist, μ_1/d>0, so that if d=−1, we require μ_1<0 and the equilibrium is always unstable. However if d=+1, $-2\mu_2$<0 and we have a sink for $b\mu_2$>$d\mu_1$, otherwise it is a saddle.

- Equation (28) is an equilibrium and corresponds to quasi-periodic flow in the full system. If D<0 it is a saddle, if D>0, T<0 it is a sink and if D>0, T>0 it is a source, where $T=\frac{2}{A}[\mu_1 d(c-1)+\mu_2(b-d)]$ and $D=\frac{4}{A}[(b\mu_2-d\mu_1)(c\mu_1-\mu_2)]$.

The derivation of the equivalent normal form for the SGCM POD-Galerkin models was deemed a formidable task and was not undertaken. Instead we used the classification of Guckenheimer and Holmes (1983) to identify whether the behaviours we observed were in any way consistent with the cases outlined there. Our AUTO integrations showed the transition sequence as r was decreased to be: no waves, a stable subsidiary wave, a stable subsidiary wave coexisting with an unstable principal wave, a stable mixed mode state,

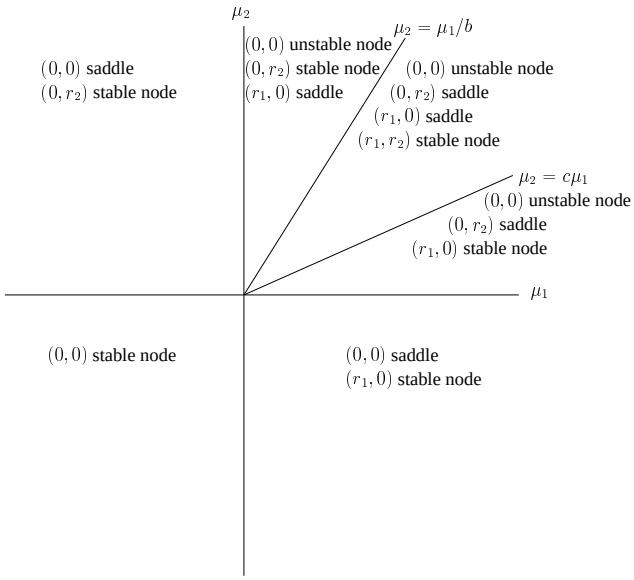

Fig. 10. Bifurcation set for the SGCM Galerkin model in (μ_1, μ_2)-parameter space. The μ_1 axis corresponds to the linear stability boundary of the subsidiary wave and μ_2 for the principal wave at the double Hopf bifurcation point. A change in the values of μ_1, μ_2 correspond to changes in the Froude number (F) and dissipation parameter (r).

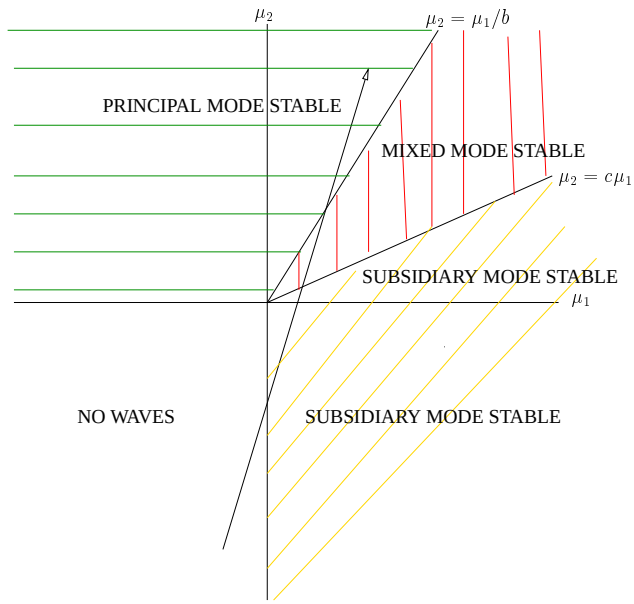

Fig. 11. A possible path in (μ_1, μ_2)-parameter space for increasing μ_1 and μ_2. The path commences in a state of no waves, passes through the stable subsidiary regime and stable moxed mode state and terminates in the stable principal regime.

coexisting with unstable principal and subsidiary waves and finally a stable principal wave.

We have explored the various possibilities for unfolding the double-Hopf bifurcation and have concluded that the likeliest scenario, which reproduces each of these transitions in order is the case $d{=}{+}1$ with $b, c{>}0$. This means that both principal and subsidiary waves are supercritical Hopf bifurcations, with the principal wave being an (unstable) saddle in the interval $[SC_2,\ HB_1]$. The specific classification of each of the equilibria is shown in Fig. 10 for the case of interest, while the stability regimes themselves are depicted in Fig. 11.

Near the double Hopf point of interest, 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. A pure principal or a pure subsidiary mode means that the modes correspond to 1-frequency flows (as in the principal and the subsidiary wave regimes), while quasi-periodic behaviour means that the modes are mixed principal-subsidiary waves, as was observed in the mixed-wave regime, and can be viewed as superimposed principal ∼14 cycles/60 sols and subsidiary ∼5 cycles/60 sols waves.

This is illustrated in Fig. 10 where the wedge $c\mu_1{>}\mu_2{>}d\mu_1/b$ in parameter space contains stable quasi-periodic solutions.

Figure 11 reproduces Fig. 10 in terms of stability and instability regions. One possible path through parameter space is illustrated:
(1) no waves;
(2) only the subsidiary mode exists and is stable;
(3) subsidiary mode is stable, principal mode is unstable;
(3) subsidiary and principal modes are unstable, while the mixed-mode is stable;
(4) principal mode is stable, the subsidiary mode is unstable.

Our discussion on the stability exchanges found along the various solution curves of the AUTO runs indicates that this path agrees with our model's behaviour.

Collins and James (1995) pointed out that an understanding of the SGCM wave 3 regime required more than linear stability theory. The presence of the double Hopf bifurcation in the POD-Galerkin models leads to the obvious questions of whether double Hopf bifurcation points exist in the original SGCM, and whether their existence influences the flow to the same extent as the models described here.

7 Conclusions

In this paper we have demonstrated that one of the wave 3 flows observed in the SGCM model for Mars is capable of being modelled by low-order systems of coupled nonlinear ordinary differential equations, derived via POD-Galerkin methods. Drawing upon the numerical investigations of a companion paper I, eighteen different POD-Galerkin models were investigated, depending upon the number of vertical modes included (2-v or 4-v), the level of horizontal truncation (6-D, 10-D or 20-D) and the type of norm used to compute the vertical structure (correlation, KE or TE).

Of all the low-order models considered the 4-v 20-D TE model was found to be the best, 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 (1995). We showed that the amplitudes and frequencies of the modes were robust to the horizontal truncation level. However, the improved vertical scheme (2-v versus 4-v) clearly affected the behaviour of the modes. The state vector included both the barotropic and baroclinic components of the flow, and so resulted in a mixing of barotropic and baroclinic wave components during the calculation of the POD modes. We believe that this was crucial in extracting important details of the vertical activity of the original flow.

We found that the 4-v modes travelled at higher frequencies than their 2-v counterparts, and contained far cleaner subsidiary waves (∼5 cycles/60 sols). The presence of two additional vertical modes resulted in the suppression of the principal (∼14 cycles/60 sols) wave in modes b_3 and b_4, and caused the leading four wave modes to travel at (almost) identical frequencies to those observed in the SGCM. The weakly nonlinear theory presented here can also explain the existence of the various flow regimes, in particular the mixed-mode wave (i.e. the superimposed principal-subsidiary wave) via the codimension-two double Hopf bifurcation in the 4-v TE model.

In all of our investigations the Froude number was taken to be F=8.14. Clearly, this fixing of the Froude number is a constraint in our studies. However, this choice for F was obtained from the vertical structure equation in I and was found to give solutions which are qualitatively similar in behaviour to that of the original flow. Low-order models have also been generated for other values of F and their solutions observed. In no other case was the behaviour of the POD-Galerkin modes more comparable to the SGCM than those described here.

We close with a few remarks concerning some of the underlying assumptions made in our numerical integrations, as well as indicating further lines for future study. The vertical eigenfunctions derived in I were obtained from the vertical structure equation using homogeneous boundary conditions $\frac{dH}{dp}|_{p_1}=\frac{dH}{dp}|_{p_n}=0$, implying that the atmosphere is bounded. In reality the atmosphere is infinite and disturbances should remain bounded as the altitude $\to\infty$. Indeed Lindzen et al. (1968) showed that bounded atmospheric models can produce spurious free oscillations. To address this problem one could impose a radiation condition at the upper boundary and repeat the current investigations. This was, however, beyond the scope of our research remit.

The SGCM takes no account of the effects of the diurnal tide and topography. Since Mars has considerably greater orography than Earth, the effects of steady and forced eddies on the global heat transport would have to be carefully modelled (see James, 1994). Mars also experiences significant temperature variations between the daytime and the nightime temperatures, and this is responsible for the creation of a diurnal tide, providing a periodic perturbation to the mid-latitude circulation. While such effects were outside the scope of the current investigations, these more realistic features will form the basis for future research (Martinez-Alvarado et al., in preparation).

Appendix A Notation

If we define

$$(\overline{\Phi}^p_k, \overline{\Phi}^p_i) = \overline{\overline{\alpha}}_{ki}, \tag{A1}$$

$$(\overline{\Phi}^p_k, \lambda_p\overline{\Phi}^p_i) = \overline{\overline{\alpha}}^{\lambda}_{ki}, \tag{A2}$$

$$(\overline{\Phi}^p_k, \nabla^2\overline{\Phi}^p_i) = \overline{\overline{\beta}}_{ki}, \tag{A3}$$

$$(\overline{\Phi}^p_k, \nabla^2\lambda_p\overline{\Phi}^p_i) = \overline{\overline{\beta}}^{\lambda}_{ki}, \tag{A4}$$

$$(\overline{\Phi}^p_k, J(\overline{\Phi}^p_i, \overline{\Phi}^p_j)) = \overline{\overline{\delta}}_{kij}, \tag{A5}$$

$$(\overline{\Phi}^p_k, J(\lambda_p\overline{\Phi}^p_i, \overline{\Phi}^p_j)) = \overline{\overline{\delta}}^{\lambda}_{kij}, \tag{A6}$$

$$(\overline{\Phi}^p_k, J(\widetilde{\Phi}^p_i, f)) = \widetilde{\overline{\eta}}_{ki}, \tag{A7}$$

$$(\overline{\Phi}^p_k, J(\widetilde{\Phi}^p_i, \nabla^2\overline{\Phi}^p_j)) = \widetilde{\overline{\zeta}}_{kij}. \tag{A8}$$

then

$$\mathcal{C} = \begin{pmatrix} \overline{\overline{\beta}}_{ki} - \overline{\overline{\alpha}}^{\lambda}_{ki} & \widetilde{\overline{\beta}}_{ki} - \widetilde{\overline{\alpha}}^{\lambda}_{ki} \\ \widetilde{\overline{\beta}}_{ki} - \widetilde{\overline{\alpha}}^{\lambda}_{ki} & \widetilde{\widetilde{\beta}}_{ki} - \widetilde{\widetilde{\alpha}}^{\lambda}_{ki} \end{pmatrix}, \tag{A9}$$

is an $(n+m)$-dimensional square matrix, coupling the time derivatives in Eq. (18).

$$\mathbf{f} = -r \begin{pmatrix} \widehat{\overline{\beta}}_{k1} \\ \widehat{\widetilde{\beta}}_{k1} \end{pmatrix}, \tag{A10}$$

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

The matrix Λ in Eq. (18) arises from the linear dissipative mean flow and wave terms, and is defined as

$$\Lambda = -r \begin{pmatrix} \overline{\overline{\beta}}_{ki} & \widetilde{\overline{\beta}}_{ki} \\ \widetilde{\overline{\beta}}_{ki} & \widetilde{\widetilde{\beta}}_{ki} \end{pmatrix}, \tag{A11}$$

while $\mathcal{A}$ arises from the advection of the wave terms by the basic state, and is given by

$$\mathcal{A} = \begin{pmatrix} \mathbf{0} & -\widehat{\widetilde{\overline{\zeta}}}_{ki1} + \widehat{\widetilde{\overline{\delta}}}^{\lambda}_{ki1} \\ \mathbf{0} & -\widehat{\widetilde{\zeta}}_{ki1} + \widehat{\widetilde{\delta}}^{\lambda}_{ki1.} \end{pmatrix}. \tag{A12}$$

Nonlinear terms are generated by the advection of the wave flow by the basic state in the matrix $\mathcal{B}$, defined as

$$\mathcal{B} = \begin{pmatrix} B_{1,ki} & \mathbf{0} \\ \mathbf{0} & B_{2,ki} \end{pmatrix}, \tag{A13}$$

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

$$B_{1,ki} = -b_i\widetilde{\overline{\zeta}}_{k1i} + b_i\widetilde{\overline{\delta}}^{\lambda}_{k1i},$$
$$B_{2,ki} = -b_i\widetilde{\zeta}_{k1i} + b_i\widetilde{\delta}^{\lambda}_{k1i}. \tag{A14}$$

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

$$\mathcal{M} = \begin{pmatrix} \mathbf{0} & M_{1,ki} \\ \mathbf{0} & M_{2,ki} \end{pmatrix}, \tag{A15}$$

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

$$M_{1,ki} = -\sum_{j=1}^{n} a_j \overline{\overline{\zeta}}_{kij} + \sum_{j=1}^{n} a_j \overline{\overset{\approx}{\delta}}_{kij},$$
$$M_{2,ki} = -\sum_{j=1}^{n} a_j \overset{\approx}{\zeta}_{kij} + \sum_{j=1}^{n} a_j \overset{\approx}{\delta}_{kij}. \tag{A16}$$

Additional nonlinear terms are generated by the advection of the wave flow in the matrix $\mathcal{W}$:

$$\mathcal{W} = \begin{pmatrix} W_{1,ki} & W_{2,ki} \\ W_{3,ki} & W_{4,ki} \end{pmatrix}, \tag{A17}$$

where $W_{j,ki}$, j=1, ..4 are defined as

$$W_{1,ki} = -\sum_{j=1}^{m} b_j \overline{\overline{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \overline{\overset{\approx}{\delta}}_{kij},$$
$$W_{2,ki} = -\sum_{j=1}^{m} b_j \overline{\overset{\approx}{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \overline{\overset{\approx}{\delta}}_{kij},$$
$$W_{3,ki} = -\sum_{j=1}^{m} b_j \overset{\approx}{\overline{\zeta}}_{kij} + \sum_{j=1}^{m} b_j \overset{\approx}{\overline{\delta}}_{kij},$$
$$W_{4,ki} = -\sum_{j=1}^{m} b_j \overset{\approx}{\zeta}_{kij} + \sum_{j=1}^{m} b_j \overset{\approx}{\delta}_{kij}. \tag{A18}$$

Finally, the effects of the Coriolis parameter f are included in $\mathcal{D}$:

$$\mathcal{D} = \begin{pmatrix} \mathbf{0} & -\overline{\overline{\eta}}_{ki} \\ \mathbf{0} & -\overset{\approx}{\eta}_{ki} \end{pmatrix}, \tag{A19}$$

where the only non-zero terms appear in rows $i=(n+1)$, .., $(n+m)$.

Edited by: T. Chang
Reviewed by: one referee

References

Collins, M. and James, I.: Regular Baroclinic Transient Waves in a Simplified Global Circulation Model of the Martian Atmosphere, J. Geophys. Res., 100, E7, 14 421–14 432, 1995.

Collins, M., Lewis, S. R., Read, P. L., and Hourdin, F.: Baroclinic wave transitions in the Martian atmosphere, Icarus, 120, 344–357, 1996.

Doedel, E.: AUTO, a program for the automatic bifurcation analysis of autonomous systems, Cong. Numer., 30, 265–284, 1981.

Eliasen, E., Machenhauer, B., and Rasmussen, E.: On a numerical method for integrations of the hydrodynamical equations with a spectral representation of the horizontal fields, in Report No. 2, 1970.

Ermentrout, G.: XPPAUT X-Windows Phase-Plane plus AUTO, ftp://ftp.math.pitt.edu/pub/bardware/tut/start.html, 1998.

Guckenheimer, J. and Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields, Springer-Verlag, 1983.

Jäckel, P.: Xigra : an interactive ODE integration package, 1995.

James, I. N.: Introduction to Circulating Atmospheres, Cambridge University Press, 1994.

Kantz, H. and Schreiber, T.: Nonlinear time series analysis, Cambridge University Press, Cambridge Nonlinear Science Series 7, 1997.

Lindzen, R. S., Batten, E. S., and Kim, J. W.: Oscillations in atmospheres with tops, Mon. Weath. Rev., 96, 3, 133–140, 1968.

Lorenz, E. N.: Irregularity: A fundamental property of the atmosphere, Tellus, 36A, 98–110, 1984.

Machenhauer, B. and Daley, R.: A baroclinic primitive equation model with a spectral representation in three dimensions, in Report No. 4, 1972.

Mak, M.: Equilibration in nonlinear baroclinic instability, J. Atmos. Sci., 42, 2764–2782, 1985.

Mo, K. G. and Ghil, M.: Statistical and dynamical of persistent anomalies, J. Atmos. Sci., 44, 877–901, 1987.

Moroz, I. and Holmes, P.: Double Hopf bifurcation and quasi-periodic flow in a model for baroclinic instability, J. Atmos. Sci., 41, 3147–3160, 1984.

Mundt, M. D. and Hart, J. E.: Secondary instability, EOF reduction, and the transition to baroclinic chaos, Physica D, 78, 65–92, 1994.

Naughton, M., Courtier, P., and Bourke, W.: Representation errors in various grid and spectral truncations for a symmetric feature on the sphere, Quart. J. R. Met. Soc., 122, 253–265, 1996.

Orszag, S. A.: Transform method for calculation of vector coupled-sums : Applications to the spectral form of the vorticity equation., J. Atmos. Sci., 27, 890–895, 1970.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetlerling, W. T.: Numerical Recipes, Cambridge University Press, 1992.

Rempfer, D.: Investigations of boundary layer transition via galerkin projections on empirical eigenfunctions, Phys. Fluids, 8, 175–188, 1996.

Schubert, S. D.: A statistical-dynamical study of empirically determined modes of atmospheric variability, J. Atmos. Sci., 42, 3–17, 1985.

Selten, F. M.: Towards an optimal description of atmospheric flow, J. Atmos. Sci., 50, 861–877, 1993.

Selten, F. M.: An efficient description of the dynamics of barotropic flow, J. Atmos. Sci., 52, 915–936, 1995.

Whitehouse, S. G.: POD-Galerkin Modelling of the Martian Atmosphere, Ph.D. thesis, University of Oxford, 1999.

Whitehouse, S. G., Moroz, I. M., Lewis, S. R., and Read, P. L.: A simplified model of the Martian atmosphere – Part 1: a diagnostic analysis, Nonlin. Processes Geophys., 12, 603–623, 2005, **SRef-ID: 1607-7946/npg/2005-12-603**.