Reduced Atmospheric Models Using Dynamically Motivated Basis Functions

FRANK KWASNIOK

School of Engineering, Computer Science and Mathematics, University of Exeter, Exeter, United Kingdom

(Manuscript received and in final form 21 December 2006)

ABSTRACT

Nonlinear deterministic reduced models of large-scale atmospheric dynamics are constructed. The dynamical framework is a quasigeostrophic three-level spectral model with realistic mean state and variability as well as Pacific–North America (PNA) and North Atlantic Oscillation (NAO) patterns. The study addresses the problem of finding appropriate basis functions for efficiently capturing the dynamics and a comparison between different choices of basis functions; it focuses on highly truncated models, keeping only 10–15 modes. The reduced model is obtained by a projection of the equations of motion onto a truncated basis spanned by empirically determined modes. The total energy metric is used in the projection; the nonlinear terms of the low-order model then conserve total energy. Apart from retuning the coefficient of horizontal diffusion, no empirical terms are fitted in the dynamical equations of the low-order model in order to properly preserve the physics of the system. Using the methodology of principal interaction patterns (PIPs), a basis is derived that carefully compromises minimizing tendency error with maximizing explained variance in the resolved modes. A new PIP algorithm is introduced that is more compact and robust than earlier PIP algorithms; a top-down approach is adopted, removing modes from the system one by one.

The mean state and standard deviation of the streamfunction as well as transient momentum fluxes are well reproduced by a PIP model with only 10 modes. Probability density functions are accurately modeled and autocorrelation functions are captured fairly well using 15 modes. Reduced models based on PIPs are substantially superior to reduced models based on empirical orthogonal functions (EOFs). The leading PIPs have a higher projection onto the PNA and NAO teleconnection patterns than the corresponding EOFs. Both with EOFs and PIPs, the interactions between the resolved modes are predominantly linear and the improvement of PIP models on EOF models stems entirely from better modeling these linear interactions although the full nonlinear tendencies are optimized. There is considerable influence of smaller-scale modes on the large-scale modes due to nonlinear coupling that is not well captured by either EOFs or PIPs. This nonlinear backscattering possibly plays a role in generating the low-frequency variability of the model. The results call for a nonlinear and/or stochastic closure scheme in which PIPs may be suitable basis functions.

1. Introduction

The construction of reduced atmospheric models, that is, models that explicitly deal only with a limited number of essential degrees of freedom while keeping as much realism as possible, has attracted some attention in recent years. The motivation for such studies is twofold. First, one may regard such work as scientifically interesting in its own right from a theoretical point of view, as it addresses the fundamental question of how many degrees of freedom are necessary to capture the dynamics to some accuracy. This is related to the dimension and complexity of the underlying attractor. Second, one may hope that low-order models, because of their compactness, can provide conceptual insight into atmospheric processes and mechanisms. In current weather prediction or climate models, such insight is not possible because of the enormous complexity and richness in simulation detail. Reduced models then could be useful tools for climate studies, prediction purposes, or data assimilation. Admittedly, it is not yet clear whether this latter goal can be achieved.

We focus here on low-order models that are nonlinear and deterministic. The strategy for deriving such low-order models is to first identify the essential degrees of freedom of large-scale atmospheric dynamics;

Corresponding author address: Dr. Frank Kwasniok, School of Engineering, Computer Science and Mathematics, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, United Kingdom.
E-mail: f.kwasniok@exeter.ac.uk

DOI: 10.1175/JAS4022.1

the reduced model is then derived by projecting a dynamical model of atmospheric flow onto these essential structures; and if appropriate, a closure scheme is applied to account for the unresolved modes and processes.

A crucial issue in the context of reduced models is how to determine adequate basis functions for efficiently spanning the dynamics. Classical approaches based on spherical harmonics (e.g., Lorenz 1963; Charney and de Vore 1979) are not well suited here since only a very limited degree of realism can be expected from such models. Empirically determined basis functions achieve a much more efficient compression of phase space. The most obvious choice when searching for optimal empirical modes is empirical orthogonal functions (EOFs). They allow for an optimal representation of atmospheric fields with a given number of basis functions. EOFs have been successfully used as basis functions in atmospheric low-order models (Rinne and Karhila 1975; Selten 1995, 1997; Achatz and Branstator 1999; D'Andrea and Vautard 2001; Achatz and Opsteegh 2003).

However, EOFs are not optimized for the purpose of constructing reduced models since their derivation does not refer to the dynamics of the underlying system. A framework to arrive at more efficient dynamical descriptions is offered by the method of principal interaction patterns (PIPs; Hasselmann 1988; Kwasniok 1996, 1997a, 2004; Crommelin and Majda 2004). PIPs are based on a dynamically motivated variational principle taking the temporal evolution of the underlying system into account. They have been shown to lead to low-order models with fewer degrees of freedom than with EOFs. For a turbulent barotropic model, a PIP model using 40 patterns performs as well as an EOF model with 100 modes (Kwasniok 2004).

A separate class of reduced models is formed by purely linear models with stochastic forcing (Newman et al. 1997; Branstator and Haupt 1998; Whitaker and Sardeshmukh 1998; Zhang and Held 1999; Winkler et al. 2001). The linear operator may either be empirically determined or arise from a linearization of the equations of motion about some basic state. Such models are not explicitly dealt with in the present paper but the question of finding proper basis functions is also relevant in that context. It has recently been shown in the framework of linearized stochastically forced dynamics that truncated models based on EOFs are far from optimal for systems with strongly nonnormal linear operators, as is the case in shear flow problems (Farrell and Ioannou 2001).

Recently, nonlinear stochastic atmospheric models of very low order have been derived both in a barotropic and baroclinic framework (Franzke et al. 2005; Franzke and Majda 2006). They are based on a general stochastic mode reduction strategy for complex nonlinear systems (Majda et al. 1999, 2003). Finding optimal basis functions is also an issue there, as EOF-based models turn out to exhibit climate drift (Franzke and Majda 2006).

In the present paper, the work on PIPs is extended to a more complex atmospheric model. The potential of dynamically motivated basis functions is investigated in the context of a baroclinic model with realistic variability. A modified PIP approach is derived in order to make the PIP calculations feasible and more robust.

The paper is structured as follows. In section 2, the atmospheric model used as the dynamical framework for the present investigation is introduced. Then the procedure of arriving at reduced models is described. Section 4 addresses the question of deriving appropriate basis functions for a reduced atmospheric model. The results obtained for the baroclinic quasigeostrophic model considered here are given and discussed in section 5. The paper closes with a summarizing and concluding section.

2. The atmospheric model

a. The quasigeostrophic model equations

In the present study, a baroclinic quasigeostrophic (QG) model on the sphere is used as the dynamical framework. The adiabatic part of the model is given by the QG potential vorticity equation in pressure coordinates. Standard vertical discretization onto three equally spaced pressure levels here located at 250, 500, and 750 hPa yields the dynamical equations

$$\frac{\partial q_1}{\partial t} + J(\Psi_1, q_1) = D_1(\Psi_1, \Psi_2) + S_1, \tag{1}$$

$$\frac{\partial q_2}{\partial t} + J(\Psi_2, q_2) = D_2(\Psi_1, \Psi_2, \Psi_3) + S_2, \tag{2}$$

$$\frac{\partial q_3}{\partial t} + J(\Psi_3, q_3) = D_3(\Psi_2, \Psi_3) + S_3, \tag{3}$$

with

$$D_1(\Psi_1, \Psi_2) = k_N R_{1,2}^{-2}(\Psi_1 - \Psi_2) - k_H \nabla^8 \hat{q}_1, \tag{4}$$

$$\begin{aligned} D_2(\Psi_1, \Psi_2, \Psi_3) = {} & -k_N R_{1,2}^{-2}(\Psi_1 - \Psi_2) \\ & + k_N R_{2,3}^{-2}(\Psi_2 - \Psi_3) - k_H \nabla^8 \hat{q}_2, \end{aligned} \tag{5}$$

$$D_3(\Psi_2, \Psi_3) = -k_N R_{2,3}^{-2}(\Psi_2 - \Psi_3) - k_E \nabla^2 \Psi_3 - k_H \nabla^8 \hat{q}_3, \quad (6)$$

where Ψ_i denotes the streamfunction and q_i the potential vorticity at pressure level $i = 1, 2, 3$. All quantities are nondimensional using the earth's radius a as the unit of length and the inverse of the angular velocity of the earth's Ω as the unit of time. Here J and ∇^2 stand for the nondimensional Jacobian and Laplacian operators on the sphere, respectively. Potential vorticity is related to the streamfunction by

$$q_1 = \nabla^2 \Psi_1 - R_{1,2}^{-2}(\Psi_1 - \Psi_2) + 2\mu, \quad (7)$$

$$q_2 = \nabla^2 \Psi_2 + R_{1,2}^{-2}(\Psi_1 - \Psi_2) - R_{2,3}^{-2}(\Psi_2 - \Psi_3) + 2\mu, \quad (8)$$

and

$$q_3 = \nabla^2 \Psi_3 + R_{2,3}^{-2}(\Psi_2 - \Psi_3) + 2\mu + h, \quad (9)$$

where λ denotes the geographic longitude and μ is the sine of geographic latitude. The Rossby deformation radii $R_{1,2}$ and $R_{2,3}$ have dimensional values of 575 and 375 km, respectively. The h represents a nondimensional topography that is related to the actual dimensional topography of the earth h^* by $h = 2\mu_0 h^*/H$, where μ_0 is the sine of some average geographic latitude taken to be 45°N and H is a scale height set to 8 km. The three pressure levels can be considered midpoints of three layers of equal pressure thickness $\Delta p = 250$ hPa. Equations (1)–(3) incorporate the boundary conditions $(\partial\Psi/\partial p)|_{p=p_t} = (\partial\Psi/\partial p)|_{p=p_b} = 0$ at the upper boundary of the top layer ($p_t = 125$ hPa) and the lower boundary of the lowest layer ($p_b = 875$ hPa). The zonal and meridional velocities of the flow at level i are given by $u_i = -\sqrt{1-\mu^2}(\partial\Psi/\partial\mu)$ and $v_i = (1/\sqrt{1-\mu^2})(\partial\Psi/\partial\lambda)$. The model is considered on the Northern Hemisphere. The boundary condition of no meridional flow, that is, vanishing streamfunction, is applied at the equator on the three model levels: $\Psi_i(\lambda, 0) = 0$; $i = 1, 2, 3$.

The dissipative terms on the right-hand side of Eqs. (1)–(3) are Newtonian temperature relaxation, Ekman damping on the lowest level, and a strongly scale-selective horizontal diffusion of vorticity and temperature. The $\hat{q}_i$ is the time-dependent part of potential vorticity at level i, that is, $\hat{q}_1 = q_1 - 2\mu$, $\hat{q}_2 = q_2 - 2\mu$, and $\hat{q}_3 = q_3 - 2\mu - h$. The coefficient of temperature relaxation represents a radiative time scale of 25 days; the linear drag damps the wind at 750 hPa on a spindown time scale of 1.5 days. The terms S_1, S_2, and S_3 are sources of potential vorticity that are independent of time but spatially varying.

Introducing a streamfunction vector $\boldsymbol{\Psi} = (\Psi_1, \Psi_2, \Psi_3)^{\mathrm{T}}$ encompassing the three vertical levels, and analogously for all other fields, the relationship between streamfunction and the time-dependent part of potential vorticity can be expressed as

$$\hat{\mathbf{q}} = -\mathbf{T}\boldsymbol{\Psi}, \quad (10)$$

where $\mathbf{T}$ is the operator

$$\mathbf{T} = \begin{pmatrix} -\nabla^2 + R_{1,2}^{-2} & -R_{1,2}^{-2} & 0 \\ -R_{1,2}^{-2} & -\nabla^2 + R_{1,2}^{-2} + R_{2,3}^{-2} & -R_{2,3}^{-2} \\ 0 & -R_{2,3}^{-2} & -\nabla^2 + R_{2,3}^{-2} \end{pmatrix} \quad (11)$$

acting on all three vertical levels. The operator $\mathbf{T}$ is invertible on the hemisphere with the present boundary condition. The dynamical equations of the QG model for the streamfunction can be written in compact form as

$$\frac{\partial \boldsymbol{\Psi}}{\partial t} = \mathbf{G}(\boldsymbol{\Psi}) = \mathbf{T}^{-1}[\mathbf{J}(\boldsymbol{\Psi}, \mathbf{q}) - \mathbf{D}(\boldsymbol{\Psi}) - \mathbf{S}]. \quad (12)$$

The vector Jacobian operator $\mathbf{J}$ is to be understood as the Jacobian operator J applied on each of the vertical levels. Here $\mathbf{D} = (D_1, D_2, D_3)^{\mathrm{T}}$ is the vector of dissipation terms and $\mathbf{S} = (S_1, S_2, S_3)^{\mathrm{T}}$ is the vector of diabatic source terms.

The streamfunction is expanded into a series of spherical harmonics: $\Psi_i = \Sigma_{\alpha=1}^{N_0} \Psi_{i,\alpha} Y_\alpha$ and $i = 1, 2, 3$. The functions Y_α are the real spherical harmonics normalized in a way that $(1/4\pi) \int_\Omega Y_\alpha Y_\beta \, d\Omega = \delta_{\alpha\beta}$ holds where $\int_\Omega d\Omega = \int_0^{2\pi} \int_{-1}^{+1} d\mu \, d\lambda$ denotes integration over the surface of the sphere. The expansion is triangularly truncated at total wavenumber 30. Only odd modes are used because of the hemispheric domain and the boundary conditions at the equator. The number of degrees of freedom in the model is $N_0 = 465$ for each level and $N = 3N_0 = 1395$ in total.

The variability exhibited by the QG model depends crucially on the diabatic source terms S_1, S_2, and S_3. To get a model behavior close to that of the real atmo-

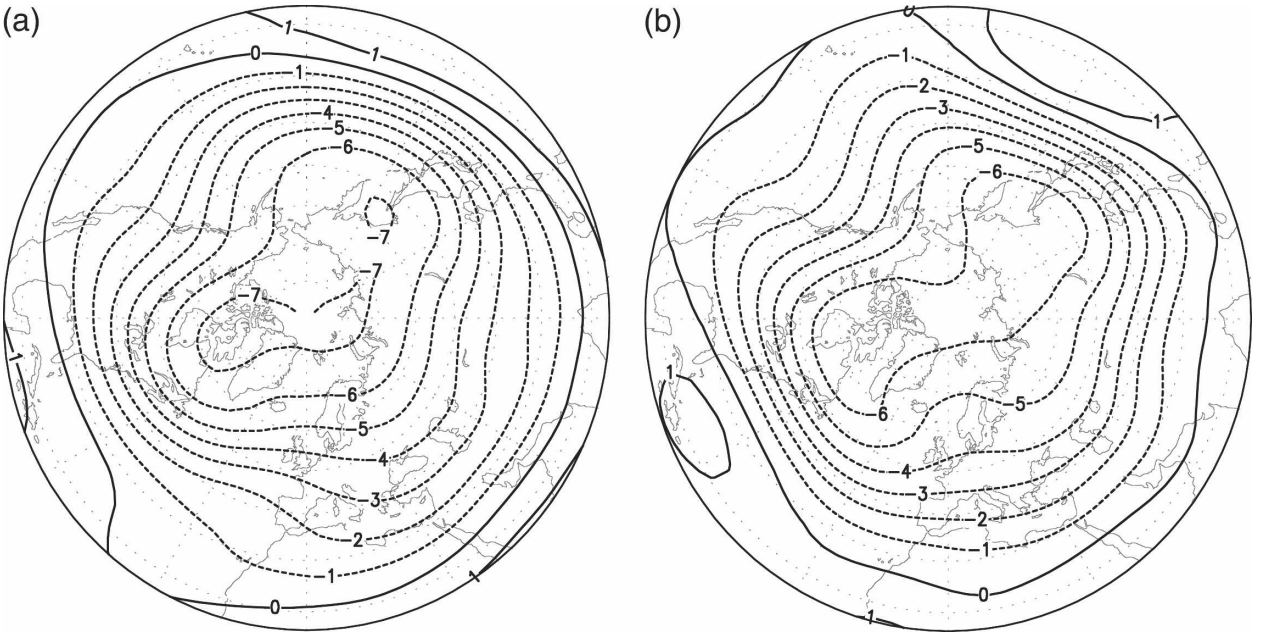

FIG. 1. Mean streamfunction at 500 hPa from (a) ECMWF reanalysis data and (b) the QG model. Units are 10^7 m^2 s^{-1}.

sphere, the forcing terms are determined by requiring that when computing potential vorticity tendencies from Eqs. (1)–(3) for a large number of observed atmospheric fields, the average of these tendencies must be zero (Roads 1987). The motivation behind this procedure is to allow the ensemble of fields used in such a computation to be representative of a statistically stable long-term behavior of the model. The source terms are then given as

$$\mathbf{S} = \langle \mathbf{J}(\boldsymbol{\Psi}, \mathbf{q}) \rangle_{\text{obs}} - \langle \mathbf{D}(\boldsymbol{\Psi}) \rangle_{\text{obs}}, \tag{13}$$

where $\langle \cdot \rangle_{\text{obs}}$ denotes the average over an ensemble of observed fields. The data used to estimate $\mathbf{S}$ were reanalysis data of vorticity in the Northern Hemisphere from the European Centre for Medium-Range Weather Forecasts (ECMWF) for each day in January from 1979 to 1994 at the three pressure levels. Actually, the values for 750 hPa were not available and thus obtained by linear interpolation between 700 and 775 hPa. The source terms S_1, S_2, and S_3 depend on the coefficients of the dissipative terms k_N, k_E, and k_H [see Eq. (13)]; k_N and k_E were fixed a priori to the values given above. An adequate value for the coefficient of horizontal diffusion was found empirically by demanding that the slope of the kinetic energy spectrum at large wavenumbers in the T30 model matches that observed in the reanalysis data. The model was integrated for 1000 days for several values of k_H and the resulting forcing terms. It was found that the slope matches best if the horizontal diffusion damps harmonics of total wavenumber 30 at a time scale of 1 day.

b. Model statistics versus reanalysis data

The model was integrated in time using the parameters and the forcing obtained as described above. The nonlinear terms were calculated using the spectral transform method; a variable-order, variable-step Adams routine from the Numerical Algorithms Group (NAG) library was used as the numerical scheme with error tolerance set to 10^{-7}. After discarding the first 5000 days in order to eliminate transient behavior, daily data for a period of 50 000 days were archived. All model statistics quoted in the present paper were calculated from this dataset.

Figure 1 shows the mean streamfunction at 500 hPa in the January reanalysis data from 1979 to 1994 and in the QG model. The agreement is very close. Also the mean states at 250 and 750 hPa are reproduced very well (not shown). In Fig. 2, the standard deviation of the streamfunction at 500 hPa is given. The most dominant features are the two areas of high variability in the Atlantic and the Pacific regions. These are well captured regarding their positions and amplitudes. The pattern of variability at 250 hPa is also dominated by these two centers of activity and is also very well modeled (not shown); however, the model has too much variance on the 250-hPa level. Table 1 summarizes the pattern correlation between the QG model and ECMWF reanalysis data for various fields. The mean streamfunction and mean zonal velocity show high pattern correlations on all levels. The standard deviation of streamfunction and also potential vorticity that has a

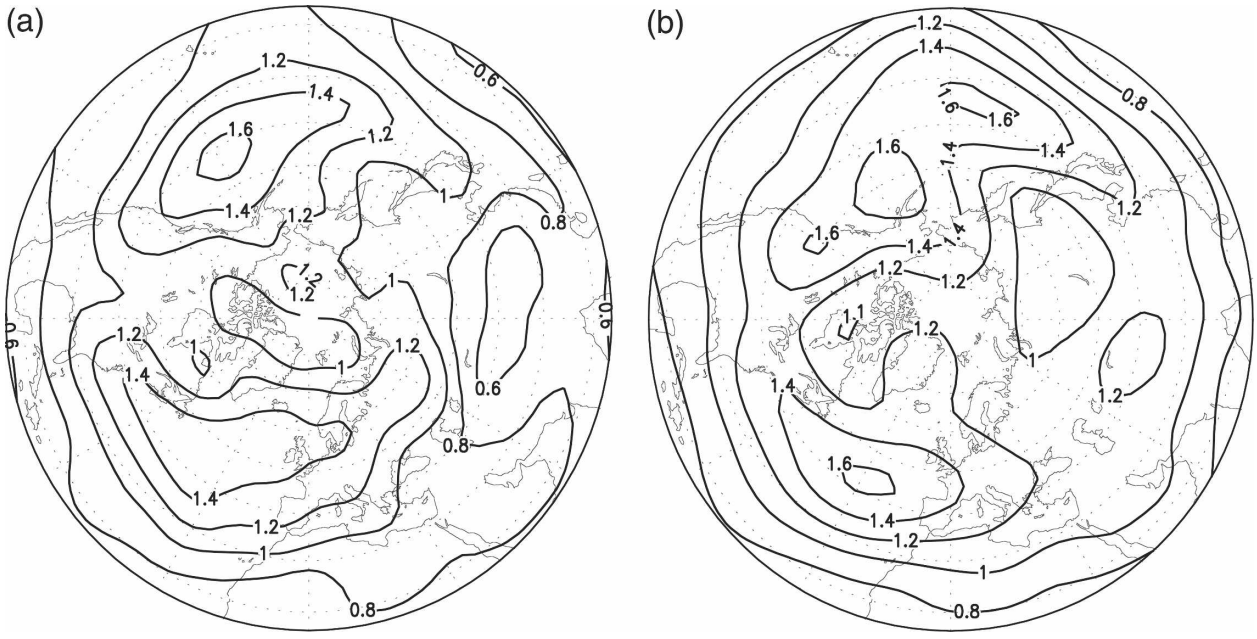

FIG. 2. Same as in Fig. 1 but for standard deviation of streamfunction.

considerably smaller spatial scale are extremely well reproduced. The transient momentum fluxes turn out to be harder to model, but at least on the two upper levels they are reasonably well captured. Figure 3 shows the correlation maps of the points (58°N, 173°W) and (58°N, 26°W), calculated from low-pass-filtered (Blackmon 1976) streamfunction at 500 hPa. The model exhibits realistic Pacific–North America and North Atlantic Oscillation patterns. The North Atlantic Oscillation pattern is actually very similar to the eastern Atlantic pattern in the original analysis by Wallace and Gutzler (1981). We conclude that the model reproduces the main features of the extratropical atmospheric variability well enough to serve as a framework to test the idea of PIP-based reduced models.

The QG model used here is very similar to that introduced by Marshall and Molteni (1993). The present model has a higher horizontal resolution. The Rossby deformation radii in the two models are different but so are the level spacings, and the corresponding profiles of static stability are very close. Moreover, it was found that the mean state, the variability, and the teleconnection patterns are quite insensitive with respect to the precise values of the Rossby deformation radii given that the diabatic forcing is calculated according to Eq. (13) for the chosen values. In the QG model by Marshall and Molteni (1993), the coefficient of the linear drag on the lowest level depends on the land–sea mask. This complication is avoided here, yet the present model is as close to reanalysis data as that by Marshall and Molteni (1993), in details even better. A comparison of the model by Marshall and Molteni (1993) with ECMWF reanalysis data is given by Franzke and Majda (2006).

TABLE 1. Pattern correlation of various fields in the QG model with the corresponding fields in ECMWF reanalysis data.

Level	$\langle\Psi\rangle$	$\langle u\rangle$	$\sqrt{\langle\Psi'^2\rangle}$	$\sqrt{\langle\hat{q}'^2\rangle}$	$\langle u'v'\rangle$
250 hPa	0.99	0.93	0.98	0.97	0.73
500 hPa	0.99	0.92	0.99	0.98	0.64
750 hPa	0.96	0.92	0.97	0.94	0.37

c. Quadratic adiabatic invariants

A class of scalar products for two sufficiently smooth functions $\mathbf{f} = (f_1, f_2, f_3)^{\mathrm{T}}$ and $\mathbf{g} = (g_1, g_2, g_3)^{\mathrm{T}}$ is introduced as

$$[\mathbf{f}, \mathbf{g}]_0 = \frac{1}{4\pi} \sum_{i=1}^{3} \int_{\Omega} f_i g_i \, d\Omega \tag{14}$$

and

$$[\mathbf{f}, \mathbf{g}]_\gamma = [\mathbf{f}, \mathbf{T}^\gamma \mathbf{g}]_0 \quad \gamma = 1, 2, 3, \ldots . \tag{15}$$

The operator $\mathbf{T}$ is self-adjoint with respect to the scalar product $[\cdot, \cdot]_0$: $[\mathbf{f}, \mathbf{Tg}]_0 = [\mathbf{Tf}, \mathbf{g}]_0$. Each of the scalar products has a norm associated with it:

$$\|\mathbf{f}\|_\gamma = [\mathbf{f}, \mathbf{f}]_\gamma^{1/2} \quad \gamma = 0, 1, 2, \ldots . \tag{16}$$

The QG equations possess an infinite number of adiabatic integral invariants. Among them are total en-

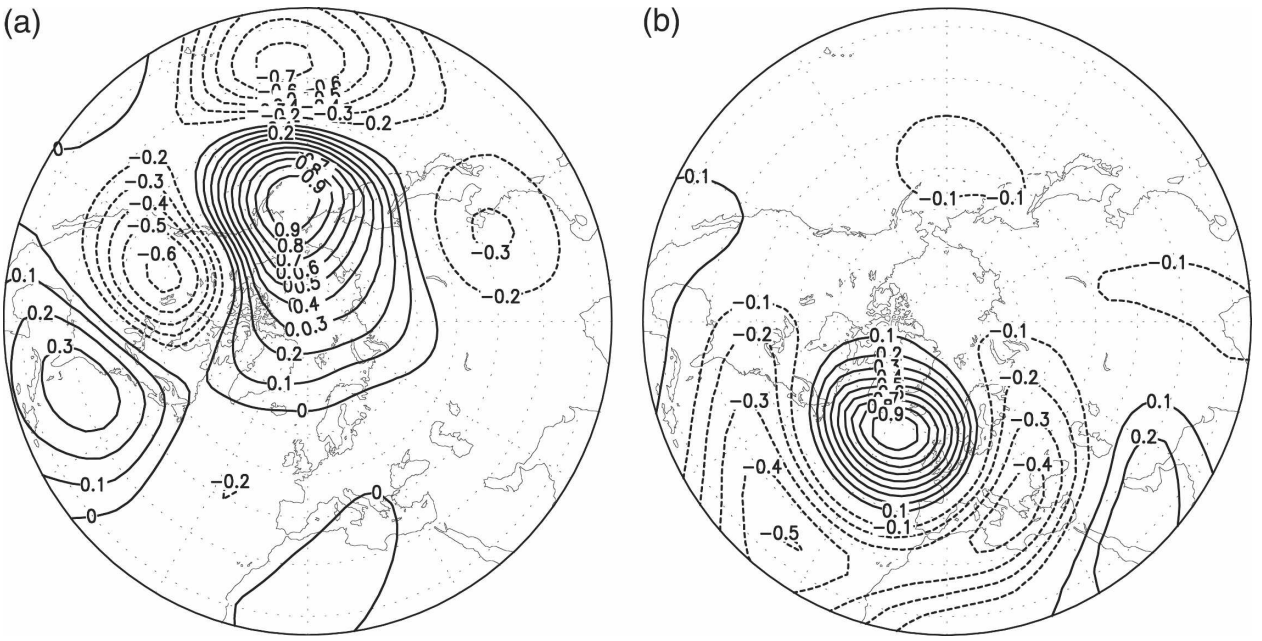

FIG. 3. PNA and NAO patterns in the QG model: Correlation maps of low-pass-filtered streamfunction at 500 hPa for the points (a) (58°N, 173°W) and (b) (58°N, 26°W).

ergy as the sum of kinetic and available potential energy,

$$E = -\frac{1}{2}[\boldsymbol{\Psi}, \hat{\mathbf{q}}]_0 = \frac{1}{2}[\boldsymbol{\Psi}, \boldsymbol{\Psi}]_1, \tag{17}$$

and potential enstrophy,

$$V = \frac{1}{2}[\hat{\mathbf{q}}, \hat{\mathbf{q}}]_0 = \frac{1}{2}[\boldsymbol{\Psi}, \boldsymbol{\Psi}]_2. \tag{18}$$

The unit of energy (more precisely energy per unit area) is $\Omega^2\alpha^2\Delta p/g$, where g is acceleration due to gravity; the unit of potential enstrophy (more precisely potential enstrophy per unit area) is $\Omega^2\Delta p/g$.

In the spectral basis, the vector of streamfunction $\boldsymbol{\Psi}$ is a vector of length N encompassing all spherical harmonic components at all three levels. The order of the components is given by $\boldsymbol{\Psi} = (\Psi_{1,1}, \Psi_{2,1}, \Psi_{3,1}, \cdots, \Psi_{1,N_0}, \Psi_{2,N_0}, \Psi_{3,N_0})^{\mathrm{T}}$ and accordingly for all other fields. The matrix representation of the operator $\mathbf{T}$ in the spectral basis is a symmetric, positive definite $N \times N$ matrix $\mathbf{M}$, the total energy metric of the spectral model. It has a tridiagonal structure; there is a 3×3 block for each spherical harmonic component on the diagonal of $\mathbf{M}$ given by

$$\mathbf{M}_\alpha = \begin{pmatrix} n_\alpha(n_\alpha+1) + R_{1,2}^{-2} & -R_{1,2}^{-2} & 0 \\ -R_{1,2}^{-2} & n_\alpha(n_\alpha+1) + R_{1,2}^{-2} + R_{2,3}^{-2} & -R_{2,3}^{-2} \\ 0 & -R_{2,3}^{-2} & n_\alpha(n_\alpha+1) + R_{2,3}^{-2} \end{pmatrix}, \tag{19}$$

where n_α is the total wavenumber of the spherical harmonic Y_α. All other elements of $\mathbf{M}$ are zero. This greatly facilitates the actual computations. The scalar products for two functions $\mathbf{f}$ and $\mathbf{g}$ in the spectral basis are $[\mathbf{f}, \mathbf{g}]_\gamma = \mathbf{f}^{\mathrm{T}}\mathbf{M}^\gamma\mathbf{g}$ for $\gamma = 0, 1, 2, \ldots$.

Total energy in the spectral model is

$$E = \frac{1}{2}\sum_{\alpha=1}^{N_0} n_\alpha(n_\alpha+1)(\Psi_{1,\alpha}^2 + \Psi_{2,\alpha}^2 + \Psi_{3,\alpha}^2) + \frac{1}{2}R_{1,2}^{-2}\sum_{\alpha=1}^{N_0}(\Psi_{1,\alpha} - \Psi_{2,\alpha})^2 + \frac{1}{2}R_{2,3}^{-2}\sum_{\alpha=1}^{N_0}(\Psi_{2,\alpha} - \Psi_{3,\alpha})^2 = \frac{1}{2}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{M}\boldsymbol{\Psi} \tag{20}$$

and potential enstrophy is given as

$$V = \frac{1}{2}\sum_{\alpha=1}^{N_0}(\hat{q}_{1,\alpha}^2 + \hat{q}_{2,\alpha}^2 + \hat{q}_{3,\alpha}^2) = \frac{1}{2}\hat{\mathbf{q}}^{\mathrm{T}}\hat{\mathbf{q}} = \frac{1}{2}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{M}^2\boldsymbol{\Psi}. \tag{21}$$

The nonlinear terms of the spectral model conserve both total energy and potential enstrophy as the continuous model does for the continuous forms of total energy and potential enstrophy. When formulating the model as an anomaly model about the time mean state, analogous statements to those given above hold for the anomaly total energy and the anomaly potential enstrophy.

3. Reduced models

Now reduced models are constructed within a perfect model approach; the QG three-level T30 model is regarded as the reference model and the goal is to derive a low-order model that captures the principal properties of the long-term behavior of the T30 model with as few degrees of freedom as possible.

a. Projection onto the essential degrees of freedom

An L-dimensional subspace of the N-dimensional phase space of the T30 model is considered that is spanned by only a limited number of spatial modes $\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$. The modes form an orthonormal system with respect to the scalar product $[\cdot, \cdot]_\gamma$: $[\mathbf{p}_i, \mathbf{p}_j]_\gamma = \delta_{ij}$. The patterns $\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$ are meant to represent the essential degrees of freedom of large-scale atmospheric dynamics and remain to be determined. The low-order model is formulated as an anomaly model. The streamfunction is separated into the mean state and the anomalies:

$$\boldsymbol{\Psi} = \langle\boldsymbol{\Psi}\rangle + \boldsymbol{\Psi}'. \tag{22}$$

Here and in the following, $\langle\cdot\rangle$ denotes the ensemble average over the attractor of the T30 model; in practice, it is replaced by the average over an ensemble of data points taken from the long-term integration of the model assuming ergodicity of the system. The dynamical equation for the streamfunction anomalies is

$$\frac{\partial\boldsymbol{\Psi}'}{\partial t} = \mathbf{T}^{-1}\mathbf{J}(\boldsymbol{\Psi}', \hat{\mathbf{q}}') + \mathbf{B}\boldsymbol{\Psi}' + \mathbf{C} \tag{23}$$

with the modified linear operator

$$\mathbf{B}\boldsymbol{\Psi}' = \mathbf{T}^{-1}[\mathbf{J}(\boldsymbol{\Psi}', \langle\mathbf{q}\rangle) + \mathbf{J}(\langle\boldsymbol{\Psi}\rangle, \hat{\mathbf{q}}') - \mathbf{D}(\boldsymbol{\Psi}')] \tag{24}$$

and the effective forcing of the anomalies given by

$$\mathbf{C} = \mathbf{T}^{-1}[\mathbf{J}(\langle\boldsymbol{\Psi}\rangle, \langle\mathbf{q}\rangle) - \mathbf{D}(\langle\boldsymbol{\Psi}\rangle) - \mathbf{S}]. \tag{25}$$

The streamfunction anomalies are expanded into a series of empirical patterns:

$$\boldsymbol{\Psi}' = \sum_{i=1}^{L} z_i\mathbf{p}_i. \tag{26}$$

The dynamical equations of the reduced model are generated by Galerkin projection of the QG model equations onto the modes $\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$ with respect to the scalar product $[\cdot, \cdot]_\gamma$:

$$\dot{z}_i = [\mathbf{p}_i, \mathbf{G}(\langle\boldsymbol{\Psi}\rangle + \sum_{i=1}^{L} z_i\mathbf{p}_i)]_\gamma. \tag{27}$$

The reduced model is a quadratically nonlinear system of L first-order differential equations

$$\dot{z}_i = \sum_{j,k=1}^{L} A_{ijk}z_jz_k + \sum_{j=1}^{L} B_{ij}z_j + C_i, \tag{28}$$

where the tensors of interaction coefficients are given by

$$A_{ijk} = \frac{1}{2}[\mathbf{p}_i, \mathbf{T}^{-1}\mathbf{J}(\mathbf{p}_j, -\mathbf{T}\mathbf{p}_k) + \mathbf{T}^{-1}\mathbf{J}(\mathbf{p}_k, -\mathbf{T}\mathbf{p}_j)]_\gamma, \tag{29}$$

$$B_{ij} = [\mathbf{p}_i, \mathbf{B}\mathbf{p}_j]_\gamma, \tag{30}$$

and

$$C_i = [\mathbf{p}_i, \mathbf{C}]_\gamma. \tag{31}$$

There are no additional empirical terms in the dynamical equations of the low-order model at this stage beyond those already introduced in setting up the QG model. For a given set of patterns, the low-order model is completely determined by the QG equations; this ensures that the physics of the QG model is properly preserved in the reduced model. We will discuss later the question of improving the low-order model by empirically fitting terms in the dynamical equations, often referred to as an empirical closure scheme.

b. Conservation properties of reduced models

In view of the conserved quantities of both the continuous QG equations and the spectral model, one may ask if these conservation properties are also adopted by reduced models. One may expect reduced models that inherit the conservation properties of the underlying equations of motion to perform better than models for which this is not the case. Particularly, conservation of at least one of the two quadratic integral quantities by the nonlinear terms guarantees that no nonlinear insta-

bilities occur when integrating the low-order model in time. Conservation properties have indeed been shown to be beneficial for the performance of atmospheric low-order models (e.g., Kwasniok 1996, 2004).

In the following, the conservation properties of the reduced model of Eqs. (27) and (28), respectively, are studied for the scalar products $[\cdot,\cdot]_0$, $[\cdot,\cdot]_1$, and $[\cdot,\cdot]_2$ that have a physical meaning as the norm streamfunction, the total energy, and the potential enstrophy scalar product, respectively. The deliberations of Selten (1995) for a barotropic model are extended to the baroclinic model used here; moreover, several different scalar products are considered whereas in the work of Selten (1995), attention is restricted to the norm streamfunction metric. A projection operator $\mathbf{P}_\gamma$ is introduced that projects a function $\mathbf{f} = (f_1, f_2, f_3)^{\mathrm{T}}$ onto the subspace spanned by the modes $\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$ with respect to the scalar product $[\cdot,\cdot]_\gamma$:

$$\mathbf{P}_\gamma \mathbf{f} = \sum_{i=1}^{L} [\mathbf{p}_i, \mathbf{f}]_\gamma \mathbf{p}_i. \tag{32}$$

The modes $\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$ are assumed to form an orthonormal system with respect to $[\cdot,\cdot]_\gamma$, that is, $[\mathbf{p}_i, \mathbf{p}_j]_\gamma = \delta_{ij}$. If necessary they have to be orthonormalized with respect to the scalar product considered using a Gram–Schmidt procedure. The projection operator $\mathbf{P}_\gamma$ is idempotent:

$$\mathbf{P}_\gamma^2 = \mathbf{P}_\gamma, \tag{33}$$

and self-adjoint with respect to the respective scalar product

$$[\mathbf{f}, \mathbf{P}_\gamma \mathbf{g}]_\gamma = [\mathbf{P}_\gamma \mathbf{f}, \mathbf{g}]_\gamma, \tag{34}$$

but in general not with respect to other scalar products. Moreover, we will subsequently make use of the facts that the operator $\mathbf{T}$ is self-adjoint with respect to all scalar products out of the class introduced here:

$$[\mathbf{f}, \mathbf{T}\mathbf{g}]_\gamma = [\mathbf{T}\mathbf{f}, \mathbf{g}]_\gamma, \tag{35}$$

and the Jacobian is orthogonal to its arguments with respect to the norm streamfunction scalar product

$$[\mathbf{f}, \mathbf{J}(\mathbf{f}, \mathbf{g})]_0 = [\mathbf{g}, \mathbf{J}(\mathbf{f}, \mathbf{g})]_0 = 0. \tag{36}$$

It turns out to be crucial whether the projection operator $\mathbf{P}_\gamma$ commutes with the operator $\mathbf{T}$, that is, $\mathbf{P}_\gamma\mathbf{T} = \mathbf{T}\mathbf{P}_\gamma$ or not. This condition has a physical interpretation. If $\mathbf{P}_\gamma$ commutes with $\mathbf{T}$, the anomaly potential vorticity $\hat{\mathbf{q}}' = -\mathbf{T}\boldsymbol{\Psi}'$ is represented by the same patterns as the corresponding streamfunction anomaly field, that is, $\mathbf{T}\mathbf{p}_i \in \mathrm{span}\{\mathbf{p}_1, \ldots, \mathbf{p}_L\}$ for any $i = 1, \ldots, L$. This is the case for an expansion in spherical harmonics since spherical harmonics are eigenfunctions of ∇^2 but generally not for empirically determined modes. If $\mathbf{P}_\gamma$ commutes with $\mathbf{T}$ it also commutes with all powers of $\mathbf{T}$ and it follows that the projection operators $\mathbf{P}_\gamma$ are identical for all γ. This means that all scalar products out of the class considered here generate the same reduced model. The anomaly total energy of a state in the reduced model is $E_r = \frac{1}{2}[\mathbf{P}_\gamma\boldsymbol{\Psi}', \mathbf{P}_\gamma\boldsymbol{\Psi}']_1$ and its tendency due to the nonlinear terms is $[\mathbf{P}_\gamma\boldsymbol{\Psi}', \mathbf{P}_\gamma\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_\gamma\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_\gamma\boldsymbol{\Psi}')]_1$. The anomaly potential enstrophy is $V_r = \frac{1}{2}[\mathbf{P}_\gamma\boldsymbol{\Psi}', \mathbf{P}_\gamma\boldsymbol{\Psi}']_2$ and its nonlinear tendency is $[\mathbf{P}_\gamma\boldsymbol{\Psi}', \mathbf{P}_\gamma\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_\gamma\boldsymbol{\Psi}', -\mathbf{T}\,\mathbf{P}_\gamma\boldsymbol{\Psi}')]_2$.

When using the norm streamfunction metric $[\cdot,\cdot]_0$ (anomaly) total energy is not conserved since

$$[\mathbf{P}_0\boldsymbol{\Psi}', \mathbf{P}_0\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_0\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_0\boldsymbol{\Psi}')]_1 = [\mathbf{P}_0\boldsymbol{\Psi}', \mathbf{T}\mathbf{P}_0\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_0\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_0\boldsymbol{\Psi}')]_0 \neq 0, \tag{37}$$

unless $\mathbf{P}_0$ commutes with $\mathbf{T}$; (anomaly) potential enstrophy is also not conserved since

$$[\mathbf{P}_0\boldsymbol{\Psi}', \mathbf{P}_0\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_0\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_0\boldsymbol{\Psi}')]_2 = [\mathbf{P}_0\boldsymbol{\Psi}', \mathbf{T}^2\mathbf{P}_0\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_0\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_0\boldsymbol{\Psi}')]_0 \neq 0, \tag{38}$$

unless $\mathbf{P}_0$ commutes with $\mathbf{T}$.

With the total energy metric $[\cdot,\cdot]_1$ total energy is conserved as

$$[\mathbf{P}_1\boldsymbol{\Psi}', \mathbf{P}_1\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_1\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}')]_1 = [\mathbf{P}_1\boldsymbol{\Psi}', \mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_1\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}')]_1 = [\mathbf{P}_1\boldsymbol{\Psi}', \mathbf{J}(\mathbf{P}_1\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}')]_0 = 0, \tag{39}$$

but not potential enstrophy unless $\mathbf{P}_1$ commutes with $\mathbf{T}$. When employing the total energy metric energy is given by $E_r = \frac{1}{2}\sum_{i=1}^{L} z_i^2$ and energy conservation is equivalent to the condition

$$A_{ijk} + A_{jik} + A_{kij} = 0 \tag{40}$$

for the nonlinear interaction coefficients, which are given by

$$A_{ijk} = \frac{1}{2}[\mathbf{p}_i, \mathbf{J}(\mathbf{p}_j, -\mathbf{T}\mathbf{p}_k) + \mathbf{J}(\mathbf{p}_k, -\mathbf{T}\mathbf{p}_j)]_0. \tag{41}$$

Equation (40) can be shown from Eq. (41) using integration by parts. The failure of the nonlinear terms of the reduced model to conserve anomaly potential enstrophy is not necessarily a disadvantage. The subspace of the low-order model can be expected to contain primarily large-scale structures; from the classical picture of an enstrophy cascade from large scales to small scales, it follows that under the nonlinear dynamics of the full spectral model potential enstrophy in the subspace of the reduced model is also not conserved, but there is an outflow to the small scales not resolved in the reduced model. So nonlinear terms not conserving potential enstrophy augmented by additional viscosity to capture the outflow of potential enstrophy may be appropriate. One may object that total energy in the subspace of the resolved patterns is also not conserved under the nonlinear dynamics of the spectral model. However, the situation is different from that with potential enstrophy. Due to the inverse cascade of kinetic energy occurring in two-dimensional turbulence, energy is somehow confined to the large scales. Conservation of energy by the reduced dynamics therefore seems appropriate.

Usage of the potential enstrophy metric $[\cdot, \cdot]_2$ leads to the conservation of potential enstrophy in the reduced model since

$$[\mathbf{P}_2\boldsymbol{\Psi}', \mathbf{P}_2\mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_2\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_2\boldsymbol{\Psi}')]_2 = [\mathbf{P}_2\boldsymbol{\Psi}', \mathbf{T}^{-1}\mathbf{J}(\mathbf{P}_2\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_2\boldsymbol{\Psi}')]_2 = [\mathbf{P}_1\boldsymbol{\Psi}', \mathbf{T}\mathbf{J}(\mathbf{P}_1\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}')]_0$$
$$= [\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}', \mathbf{J}(\mathbf{P}_1\boldsymbol{\Psi}', -\mathbf{T}\mathbf{P}_1\boldsymbol{\Psi}')]_0 = 0, \quad (42)$$

but not energy unless $\mathbf{P}_2$ commutes with $\mathbf{T}$. Using the scalar product $[\cdot, \cdot]_2$ is equivalent to expanding $\hat{\mathbf{q}}'$ rather than $\boldsymbol{\Psi}'$ into a series of empirical patterns and then using the scalar product $[\cdot, \cdot]_0$ in the projection.

It can be summarized that for low-order models based on spherical harmonics, all three scalar products $[\cdot, \cdot]_0$, $[\cdot, \cdot]_1$, and $[\cdot, \cdot]_2$ generate the same dynamical model that conserves both total energy and potential enstrophy; particularly, the full T30 model does. For empirical modes, the three models are generally all different and have different conservation properties. It is not possible to conserve both energy and potential enstrophy and a choice has to be made. Following the discussion above and keeping in mind that there are many geophysical fluid systems without an enstrophy-like invariant at all, even in the continuous equations, we feel that it is more important to conserve energy than potential enstrophy; therefore the total energy scalar product $[\cdot, \cdot]_1$ is used throughout this study.

For completeness, it is noted that there are alternative formulations of reduced models with different conservation properties (Selten 1995). They involve a projection of both streamfunction and potential vorticity inside and/or outside the Jacobian in the definition of the reduced model, thus enforcing that streamfunction and potential vorticity are represented by the same patterns. It is then possible to conserve both energy and potential enstrophy. However, these formulations leave the framework of linear Galerkin projection and have undesirable properties (e.g., the nonlinear coupling between a triad of modes depends on which other modes are used in the low-order model, which is unphysical). Therefore these alternative formulations are not considered in the present work.

4. Basis functions

Reduced models can be derived as outlined in the previous section for any orthonormal basis of patterns. Now the question of how to determine basis functions that capture the essential dynamics of the system with as few modes as possible arises.

a. Empirical orthogonal functions

When looking for empirically determined modes, the most straightforward choice is EOFs. EOFs offer an optimal representation of the anomaly streamfunction field in the mean least squares sense (here measured in the total energy metric) with a given number of modes L. They are calculated as eigenvectors of the eigenvalue problem

$$\boldsymbol{\Gamma}\mathbf{M}\mathbf{e}_i = \lambda_i\mathbf{e}_i, \quad (43)$$

where $\boldsymbol{\Gamma}$ is the covariance matrix of the spectral components of streamfunction in the T30 model. The eigenvalues λ_i give the variance accounted for by each EOF. The EOFs form an orthonormal system with respect to the total energy metric: $[\mathbf{e}_i, \mathbf{e}_j]_1 = \mathbf{e}_i^{\mathrm{T}}\mathbf{M}\mathbf{e}_j = \delta_{ij}$. It is computationally more convenient to solve the symmetric eigenvalue problem

$$\mathbf{M}^{1/2}\boldsymbol{\Gamma}\mathbf{M}^{1/2}\tilde{\mathbf{e}}_i = \lambda_i\tilde{\mathbf{e}}_i \quad (44)$$

and then to transform according to $\mathbf{e}_i = \mathbf{M}^{-1/2}\tilde{\mathbf{e}}_i$.

EOFs have two appealing properties: first, they have a clearly understandable theoretical basis, and second, they can be obtained quite easily and robustly even for high-dimensional systems since they involve only a symmetric eigenvalue problem. This makes them a prime candidate for basis functions in a reduced model.

EOFs have been used as basis functions in atmospheric low-order models in recent years (Selten 1995, 1997; Achatz and Branstator 1999; D'Andrea and Vautard 2001; Achatz and Opsteegh 2003). When proposing a more ambitious scheme for deriving basis functions, one clearly has to address the questions of whether that new scheme is still feasible and whether the improvement of that new scheme on EOFs is large enough to justify a more complicated computation.

b. Principal interaction patterns

A more advanced choice of basis functions is principal interaction patterns. The method of PIPs takes the dynamics of the underlying system into account in order to define the basis functions. A cost function is introduced that measures either the mean squared tendency error of the reduced model (Hasselmann 1988; Kwasniok 1996) or the mean-squared error between trajectories of the reduced and the full model (Kwasniok 1997a,b, 2001, 2004; Crommelin and Majda 2004). The PIPs are determined by minimizing this cost function with respect to the patterns subject to $[\mathbf{p}_i, \mathbf{p}_j]_1 = \delta_{ij}$. This poses a high-dimensional nonlinear minimization task. The approach based on trajectories is generally more powerful than a cost function referring only to the tendencies. However, the algorithm based on trajectories is computationally quite expensive unless the system under consideration is relatively simple. Therefore we stick to local tendencies in the present work. Even minimization of a tendency-based cost function is quite tedious for the model under investigation here since all patterns are treated at the same time and the calculation has to be repeated for a different truncation level L.

Therefore a novel PIP algorithm is proposed that is more compact and robust than earlier PIP algorithms. A top-down approach is adopted, removing modes from the system one by one. The large minimization problem is divided into several smaller ones. Having a subspace of K patterns left, the mode to be removed, the Kth PIP, is determined by minimizing the tendency error of the reduced model build on the remaining $K - 1$ modes. The PIPs then have a definite order as EOFs have and are independent of the truncation level L of the reduced model in which they are employed. Moreover, the minimization problem may be expected to be less ill conditioned when dealing only with one pattern at a time. Suppose the system is already projected onto a subspace of dimension K spanned by the modes $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$ that are orthonormal: $[\mathbf{r}_i, \mathbf{r}_j]_1 = \delta_{ij}$. The algorithm to determine the mode to be removed next, the Kth PIP $\mathbf{p}_K$, is as follows.

- Step 1: The basis $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$ is rotated to the basis $\{\mathbf{s}_1, \ldots, \mathbf{s}_K\}$, which is uncorrelated and ordered by decreasing variance. Here $\mathbf{s}_i$ is given by $\mathbf{s}_i = \sum_{j=1}^{K} U_{ji}\mathbf{r}_j$, where $\mathbf{U}$ is the orthogonal matrix with the eigenvectors of the covariance matrix of the modes $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$ as its columns, ordered by decreasing corresponding variance. The modes $\{\mathbf{s}_1, \ldots, \mathbf{s}_K\}$ span the same subspace as the modes $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$ do and are orthonormal: $[\mathbf{s}_i, \mathbf{s}_j]_1 = \delta_{ij}$.
- Step 2: Let $\mathbf{v} = \sum_{i=1}^{K} \alpha_i \mathbf{s}_i$ be an arbitrary linear combination of the modes $\{\mathbf{s}_1, \ldots, \mathbf{s}_K\}$. The cost function

$$\chi^2(\mathbf{v}) = \chi^2(\alpha_1, \cdots, \alpha_K) = \left\langle \sum_{i=1}^{K-1} (\dot{z}_i - \dot{z}_i^*)^2 \right\rangle \tag{45}$$

  is minimized with respect to the coefficients $\alpha = (\alpha_1, \cdots, \alpha_K)$, subject to the constraint $||\mathbf{v}||_1^2 = \sum_{i=1}^{K} \alpha_i^2 = 1$. Here, $\dot{z}_i$ is the tendency of the mode $\mathbf{x}_i$ in a reduced model according to Eq. (27) of the order $L = K - 1$ based on the modes $\{\mathbf{x}_1, \ldots, \mathbf{x}_{K-1}\}$ and $\dot{z}_i^* = [\mathbf{x}_i, (\partial\Psi'/\partial t)]_1$ is the (full nonlinear) tendency of the mode $\mathbf{x}_i$ in the QG model. The modes $\{\mathbf{x}_1, \ldots, \mathbf{x}_{K-1}\}$ span the $(K - 1)$-dimensional subspace orthogonal to $\mathbf{v}$. They are obtained by applying Gram–Schmidt orthogonalization to the set of modes $\{\mathbf{v}, \mathbf{s}_{K-1}, \ldots \mathbf{s}_1\}$, mapping it to the set $\{\mathbf{v}, \mathbf{x}_{K-1}, \ldots \mathbf{x}_1\}$. The modes $\{\mathbf{x}_1, \ldots, \mathbf{x}_{K-1}\}$ satisfy $[\mathbf{x}_i, \mathbf{x}_j]_1 = \delta_{ij}$ and $[\mathbf{v}, \mathbf{x}_i]_1 = 0$. The result of the minimization is denoted by $\mathbf{v}^{\min} = \sum_{i=1}^{K} \alpha_i^{\min} \mathbf{s}_i$.
- Step 3: Set $\mathbf{p}'_K = \mathbf{v}^{\min} + \eta\mathbf{s}_K$ and then $\mathbf{p}_K = \mathbf{p}'_K/||\mathbf{p}'_K||_1$ as the Kth PIP.
- Step 4: Gram–Schmidt orthogonalization is applied to the set of modes $\{\mathbf{p}_K, \mathbf{s}_{K-1}, \ldots, \mathbf{s}_1\}$, mapping it to the set $\{\mathbf{p}_K, \mathbf{y}_{K-1}, \ldots, \mathbf{y}_1\}$. The modes $\{\mathbf{y}_1, \ldots, \mathbf{y}_{K-1}\}$ span the $(K - 1)$-dimensional subspace orthogonal to $\mathbf{p}_K$; they satisfy $[\mathbf{y}_i, \mathbf{y}_j]_1 = \delta_{ij}$ and $[\mathbf{p}_K, \mathbf{y}_i]_1 = 0$.

The K value is reduced by 1 and steps 1–4 are repeated, taking the mode $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$ as the modes $\{\mathbf{y}_1, \ldots, \mathbf{y}_{K-1}\}$ from the previous stage until $K = 2$. The final stage $K = 2$ yields $\mathbf{p}_2$ and $\mathbf{p}_1$ as the remaining modes. The patterns generated by this algorithm are orthonormal ($[\mathbf{p}_i, \mathbf{p}_j]_1 = \delta_{ij}$); they are generally not uncorrelated.

The PIP algorithm becomes unpractical when applied to the full QG model; a prior reduction of the system using EOFs is advisable. Introducing a parameter $K_{\max}$, the PIPs are set equal to the EOFs for $K > K_{\max}$. The algorithm is started with $K = K_{\max}$ setting the modes $\{\mathbf{r}_1, \cdots, \mathbf{r}_{K_{\max}}\}$ as $\{\mathbf{e}_1, \cdots, \mathbf{e}_{K_{\max}}\}$, the first $K_{\max}$ EOFs.

The minimization of the cost function in step 2 poses a nonlinear minimization problem with K variables that

has to be solved numerically. The constraint $||\mathbf{v}||_1^2 = 1$ is treated easiest by including a normalization of $\mathbf{v}$ into the definition of the cost function, turning the minimization problem into an unconstrained one. This was then solved using a quasi-Newton method, actually the routine E04DGF from the NAG library. The required gradient of the cost function with respect to α was provided approximately using the computationally cheap method of first-order forward finite differences. It could be provided exactly, making the algorithm even faster. Details on forming the gradient of cost functions like the one in Eq. (45) are discussed by Kwasniok (1996, 1997a). The minimization was started with $\mathbf{v} = \mathbf{s}_K$, that is, $(\alpha_1, \ldots, \alpha_{K-1}, \alpha_K) = (0, \ldots, 0, 1)$ as an initial guess. Because of the rotation in step 1, $\mathbf{s}_K$ is the mode with the smallest explained variance in the space of the remaining modes. In the absence of other prior knowledge it seems reasonable to take the EOF-like approach of removing the least energetic mode as a starting point from which to optimize. The number of iterations in the minimization procedure was limited to five independent of K. This is clearly not enough to solve the minimization problem up to machine precision. However, the goal here is to roughly estimate the potential of the PIP method, particularly in comparison to EOF models. Moreover, the cost function is derived from a finite sample of data, inevitably causing some statistical uncertainty in its value. It is therefore not necessary to locate the minimum of the cost function very accurately.

Here η is a nonnegative parameter of the method. For $\eta = 0$, the PIPs are determined purely by minimizing the cost function. Contrastingly, for $\eta \to \infty$, the information of the cost function is neglected and for any K the least energetic of the remaining modes is removed from the system. The PIPs are then just the EOFs for all K. For finite values of η, the method compromises minimizing tendency error with maximizing explained variance. First of all, η is a regularization parameter as it is often introduced in inverse problems. The cost function of Eq. (45) involves higher-order moments of the system and can always be expected to be ill conditioned. This ill conditioning is resolved by mixing the cost function with the prior knowledge that the modes should not deviate too much from the EOFs (Kwasniok 1996), having in mind that EOFs perform well in a reduced model and can certainly serve as a starting point upon which to improve. But the meaning of the parameter η goes beyond this. Apart from ill conditioning and statistical uncertainties due to the empirical nature of the cost function, it is conceivable that the relationship between local tendency error and the behavior in a long-term integration is not monotonic. A mixture between minimizing tendency error and capturing variance in order to ensure that the large-scale structures are kept may be more desirable than strictly minimizing tendency error. This point will be expanded upon later when discussing the results.

5. Results and discussion

Reduced models have been derived from the QG model as outlined in sections 3 and 4. The parameter $K_{\max}$ was set to 25. For $K > 25$, the Kth PIP is identical to the Kth EOF. For $K \geq 25$, the subspace spanned by the first K PIPs is identical to the subspace spanned by the first K EOFs. The cost function χ^2 was evaluated using an ensemble of 5000 data points, separated by 5 days in order to reduce serial correlation. The calculation of the PIPs is then feasible on a current workstation within a couple of hours; on a high-performance computer it would be much faster. In the following, results are reported for $\eta = 1$; the dependence of the PIPs and corresponding low-order models on η will be discussed later.

a. Enhanced horizontal diffusion

When integrating the reduced models in time, it turns out that both PIP and EOF models have systematically too much variance. The mean squared amplitudes of the modes at the truncation limit, being many times larger than in the spectral model, are especially inflated. This problem is well known with low-order models (e.g., Selten 1995; Kwasniok 1996). A remedy is the introduction of additional dissipation into the linear operator of the system. Analogous to the horizontal diffusion necessary in the spectral model, an enhanced dissipation in the linear term of the reduced model acts as a crude closure scheme to account for the interactions of the resolved modes with the neglected modes. In the present study, the strategy is to stick as closely as possible to the governing equations of the system and to introduce only as few empirical components as possible into the dynamical equations of the reduced model. Additional dissipation is introduced by retuning only the coefficient of horizontal diffusion acting on the anomalies in the reduced model. From each of the three model levels, there are two terms in the low-order model involving horizontal diffusion: $-k_H\nabla^8\langle\hat{q}_i\rangle$ and $-k_H\nabla^8\hat{q}_i'$. The terms $-k_H\nabla^8\langle\hat{q}_i\rangle$ are left unchanged; in the anomaly terms $-k_H\nabla^8\hat{q}_i'$, the coefficient k_H is replaced by a value k_H^r in the reduced model. The changed coefficient is determined in the following manner. The low-order model is integrated in time for a couple of values k_H^r larger than k_H; transient behavior is

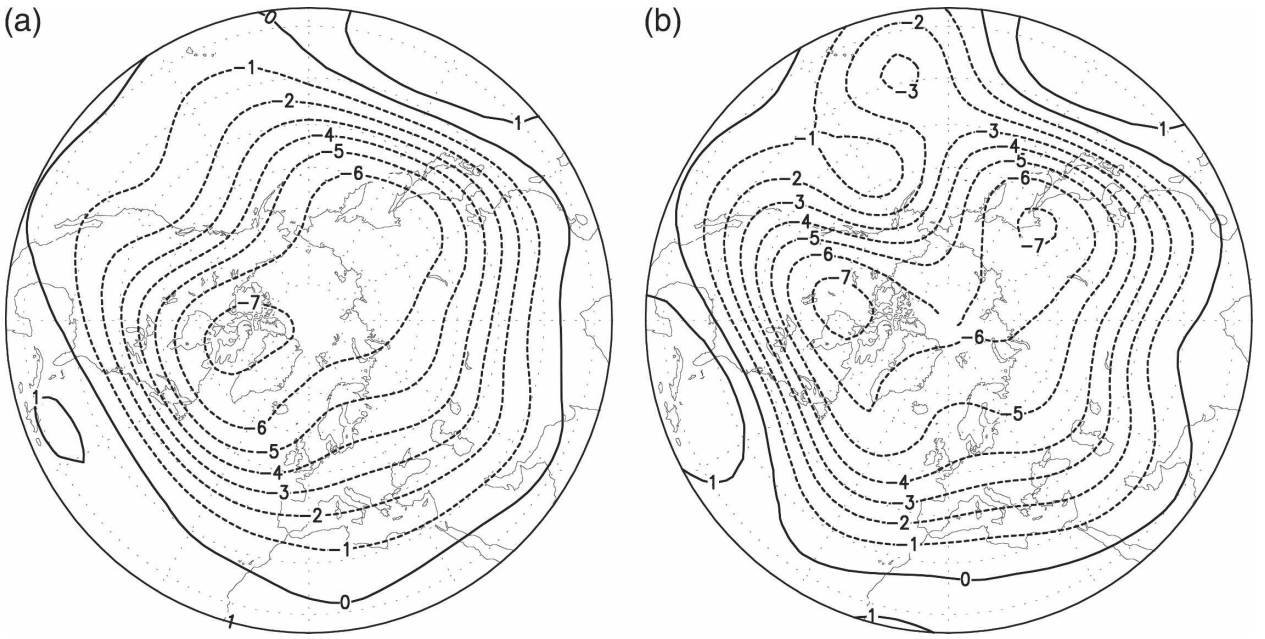

FIG. 4. Mean streamfunction at 500 hPa in the reduced models with (a) 10 PIPs and (b) 10 EOFs. Units are 10^7 m^2 s^{-1}.

discarded. The mean anomaly total energy is estimated; 1000 days of daily data are sufficient to obtain a moderately accurate estimate. It is empirically observed that the variance of the reduced model decreases monotonically when increasing the value of k_H^r; it eventually decreases to virtually zero when k_H^r takes large values. The appropriate value of k_H^r for the reduced model is determined as the value for which the mean anomaly total energy of the reduced model matches the mean anomaly total energy of the spectral model projected onto the retained PIPs or EOFs, respectively. The value of k_H^r depends on the truncation level L.

Here only a single parameter that has a clearly understandable meaning is fitted empirically in the low-order model; everything else is determined by the underlying equations of motion. This is in contrast to recent work on nonlinear reduced models where the complete linear operator of the reduced model is determined empirically (Achatz and Branstator 1999; Achatz and Opsteegh 2003; Kwasniok 2004). We do not argue against such extensive empirical closure schemes in general. They may be necessary and appropriate, particularly when leaving the perfect model approach, for then unresolved processes as well as unresolved scales have to be accounted for. Fitting the whole linear operator has been shown to improve the simulation of selected quantities; however, it renders a physical interpretation of the low-order model doubtful. It is certainly of value to first determine basis functions capturing the genuine physics of the system optimally; even when applying an extensive empirical closure scheme later, such basis functions can be expected to need smaller closure corrections and to perform better than other choices. Especially when using the reduced model for climate change studies, it should be beneficial to keep any terms fitted to the current climate to a minimum.

b. Long-term behavior of the reduced models

Long-term integrations of PIP and EOF models were made at various truncation levels using enhanced horizontal diffusion determined as described above; 50 000 days of daily data were archived after discarding 5000 days of integration in order to get rid of transient motion. All statistics of the reduced models quoted here are based on this ensemble of data. In the following, results are discussed for the truncation levels $L = 10$ and $L = 15$.

The climatological mean state of streamfunction at 500 hPa in reduced models with 10 PIPs and 10 EOFs is given in Fig. 4. The correspondence with the mean state of the QG model (Fig. 1b) is very close for both models; the EOF model has some distortions.

In Figs. 5 and 6, the standard deviation of streamfunction at 500 hPa in PIP and EOF models with 10 modes and the T30 model projected onto the respective patterns is displayed. The PIP model captures the pattern of variability and also the amplitudes very well. With the EOF model, there are considerable errors both in position and amplitude of the maxima.

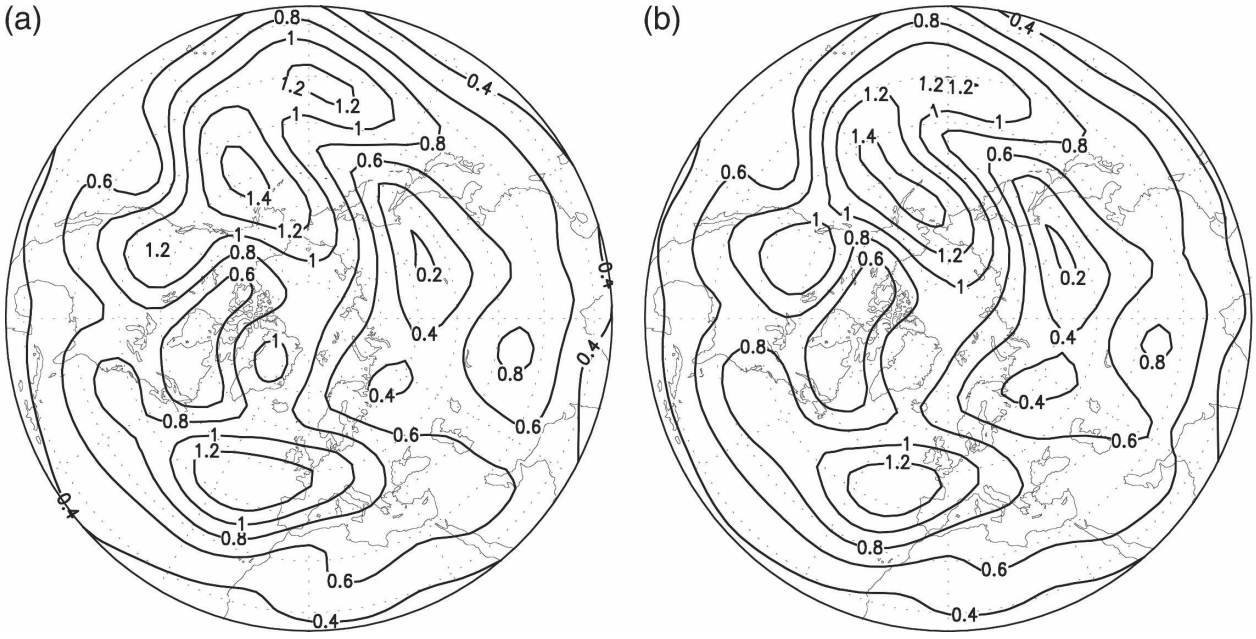

FIG. 5. Standard deviation of streamfunction at 500 hPa in the (a) QG model projected onto 10 PIPs and (b) reduced model using 10 PIPs. Units are 10^7 m^2 s^{-1}.

Figures 7 and 8 show the transient momentum fluxes $\langle u_1' v_1' \rangle$ at 250 hPa in low-order models based on 10 PIPs and 10 EOFs, respectively, and in the T30 model projected onto the respective patterns. The PIP model is quite accurate; the EOF model performs clearly worse.

In Table 2, the performance of PIP and EOF models with 10 and 15 modes is summarized for all three vertical levels. The pattern correlation of the mean streamfunction, the streamfunction standard deviation, and the transient momentum fluxes in the various reduced models with the corresponding fields in the QG model projected onto the respective modes used in the reduced model are given. It should be noted that the pattern correlation between two standard deviation

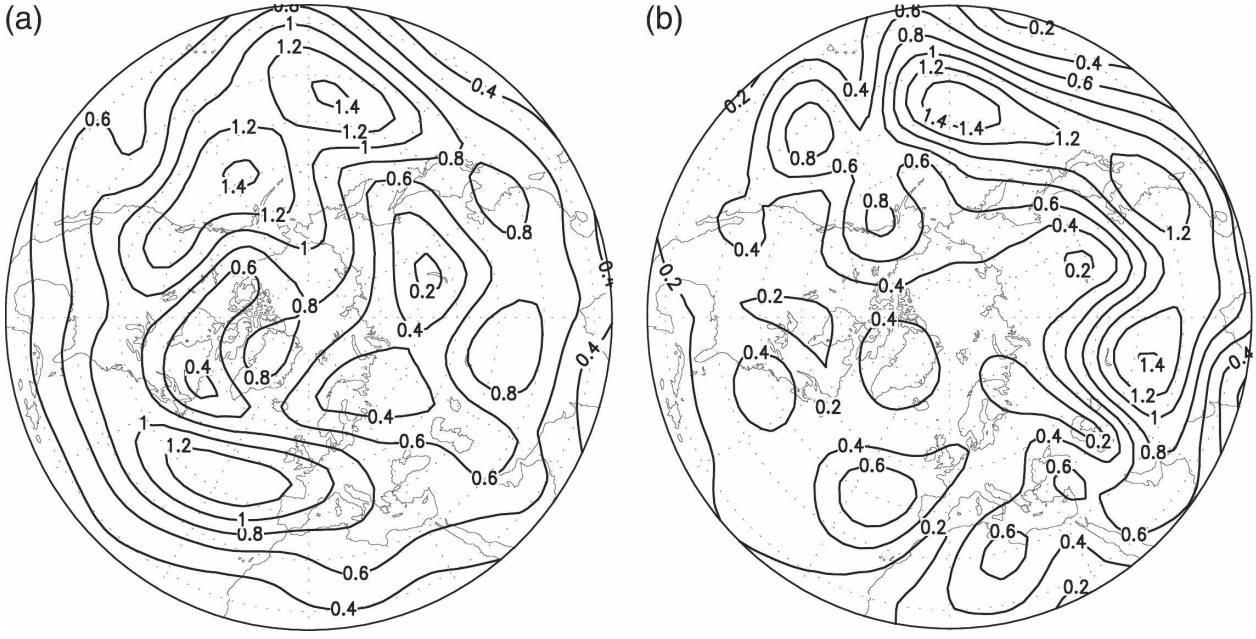

FIG. 6. Same as in Fig. 5 but for the (a) QG model projected onto 10 EOFs and (b) reduced model using 10 EOFs.

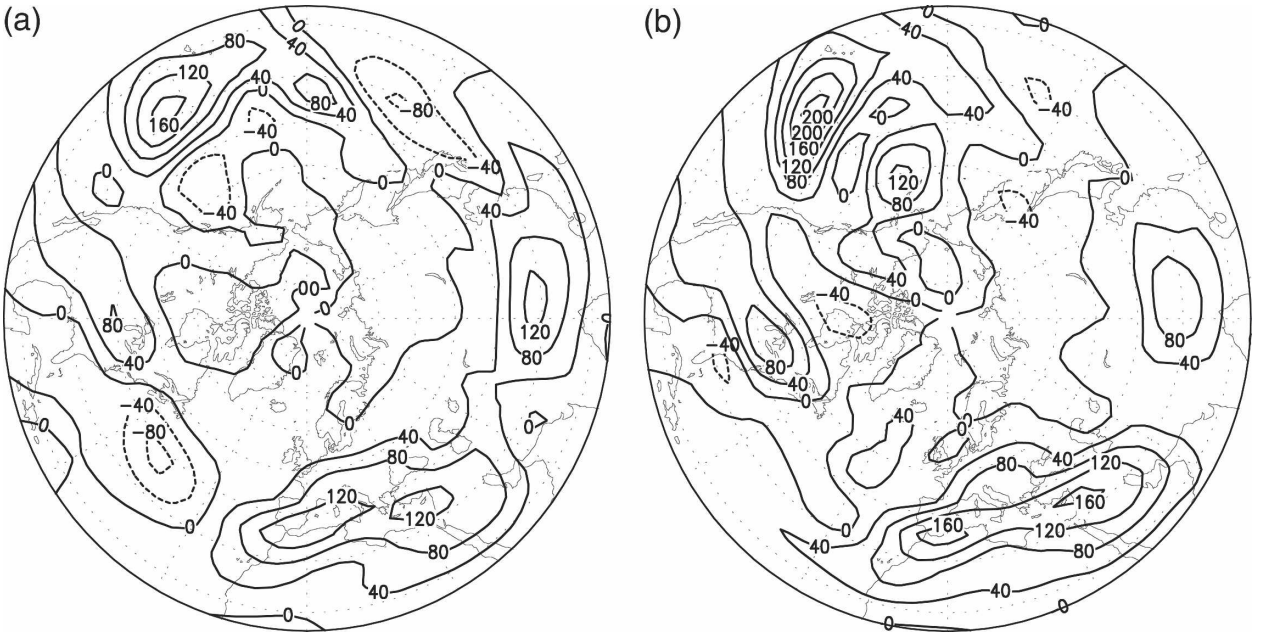

FIG. 7. Transient momentum fluxes at 250 hPa in the (a) QG model projected onto 10 PIPs and (b) reduced model using 10 PIPs. Units are $m^2 s^{-2}$.

patterns is strongly biased toward high positive values since both fields are by construction positive everywhere. This leads to a high correlation even for a relatively poor agreement; only a value of 0.95 or higher indicates a really good accordance. A similar bias occurs when comparing two climatological mean states using the pattern correlation because the mean streamfunction field has negative sign almost everywhere on all three levels. Both with 10 and 15 modes, the PIP model outperforms the corresponding EOF model in all quantities. Moreover, the PIP model with 10 modes already outperforms the EOF model with 15 modes in all quantities. The improvement is most prominent in the transient momentum fluxes, which are harder to

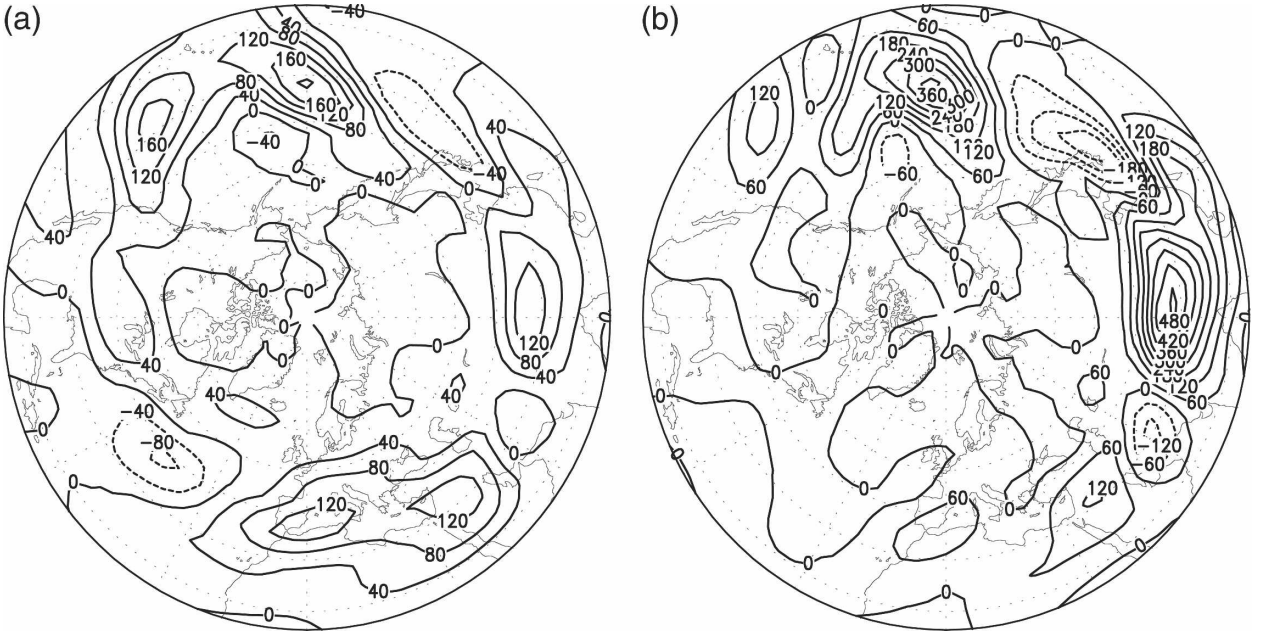

FIG. 8. Same as in Fig. 7 but for the (a) QG model projected onto 10 EOFs and (b) reduced model using 10 EOFs.

TABLE 2. Pattern correlation of various fields in a reduced model based on 10 (numbers on left) or 15 (numbers on right) PIPs or EOFs with the corresponding fields in the QG model projected onto the respective modes used in the reduced model.

	$\langle\Psi\rangle$		$\sqrt{\langle\Psi'^2\rangle}$		$\langle u'v'\rangle$	
Level	10/15 PIPs	10/15 EOFs	10/15 PIPs	10/15 EOFs	10/15 PIPs	10/15 EOFs
250 hPa	1.00/1.00	0.99/0.99	0.98/0.99	0.89/0.97	0.88/0.92	0.67/0.85
500 hPa	1.00/1.00	0.98/0.99	0.98/0.99	0.89/0.97	0.76/0.81	0.56/0.71
750 hPa	1.00/1.00	0.95/0.99	0.98/0.99	0.88/0.96	0.77/0.81	0.56/0.70

capture in a low-order model than the mean state and the standard deviation of streamfunction.

The probability density functions of the first four PIPs in the full QG model and a reduced model using 15 PIPs are shown in Fig. 9, and accordingly for EOFs in Fig. 10. The probability densities of the leading EOFs as well as PIPs are very close to Gaussian. The PIP model with 15 modes models them quite accurately; the EOF model with 15 modes is clearly worse, having drifts in the mean, errors in variance, and skews not present in the QG model.

In Fig. 11, the autocorrelation functions of the first 4 PIPs in the QG model and in a reduced model using 15 PIPs are displayed. The corresponding information for EOFs is given in Fig. 12. The slowly decaying autocorrelation of the first and the fourth PIP is reproduced but the correlation time scale is too long in the low-order model. The relatively fast decay of the autocorrelations of the second and third PIP are captured; however, the reduced model exhibits some oscillations that are not present in the QG model. The model based on 15 EOFs performs clearly worse than the model based on 15 PIPs. The autocorrelation of the first EOF is well captured up to a lag of 15 days but then turns to negative values, which is not the case in the QG model. In the second, third, and fourth EOF, the reduced model

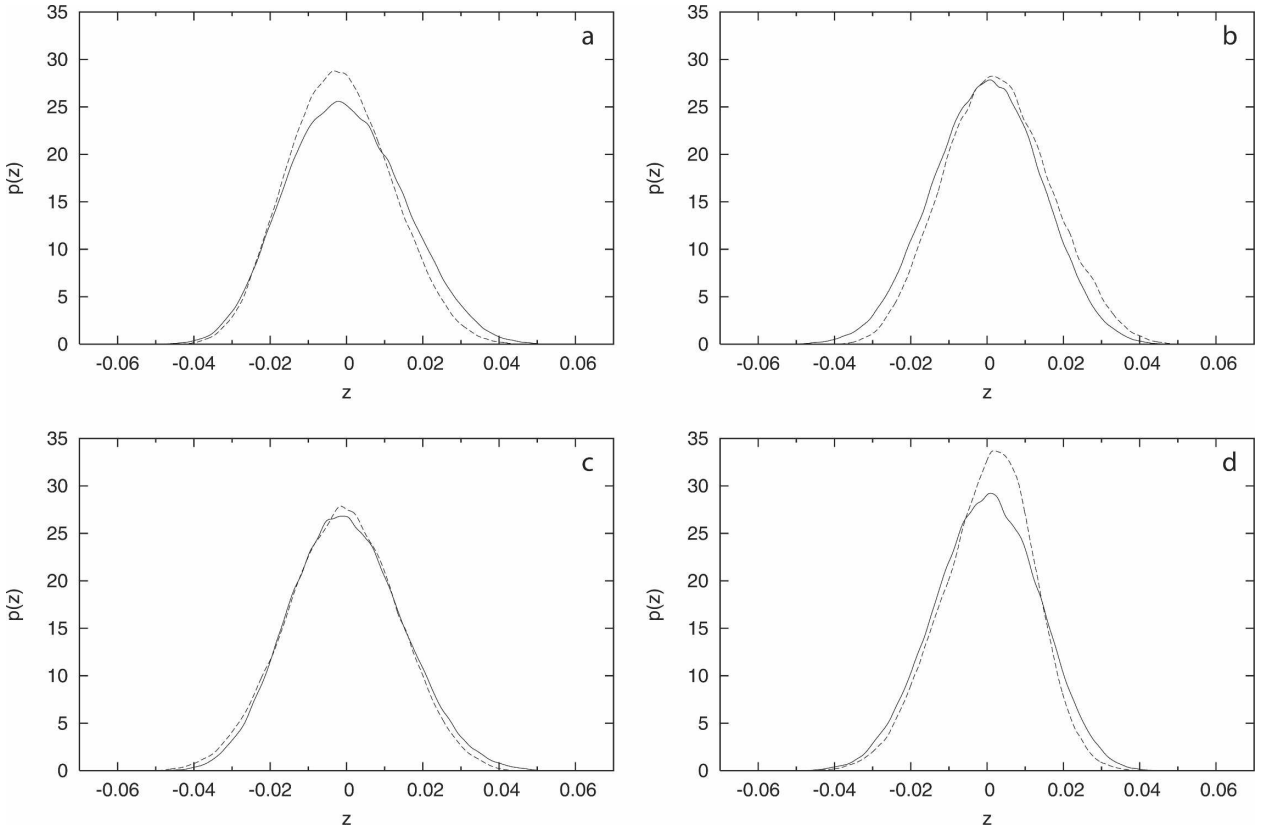

FIG. 9. Probability density function of the (a) first, (b) second, (c) third, and (d) fourth PIP in the QG model (solid) and the reduced model using 15 PIPs (dashed).

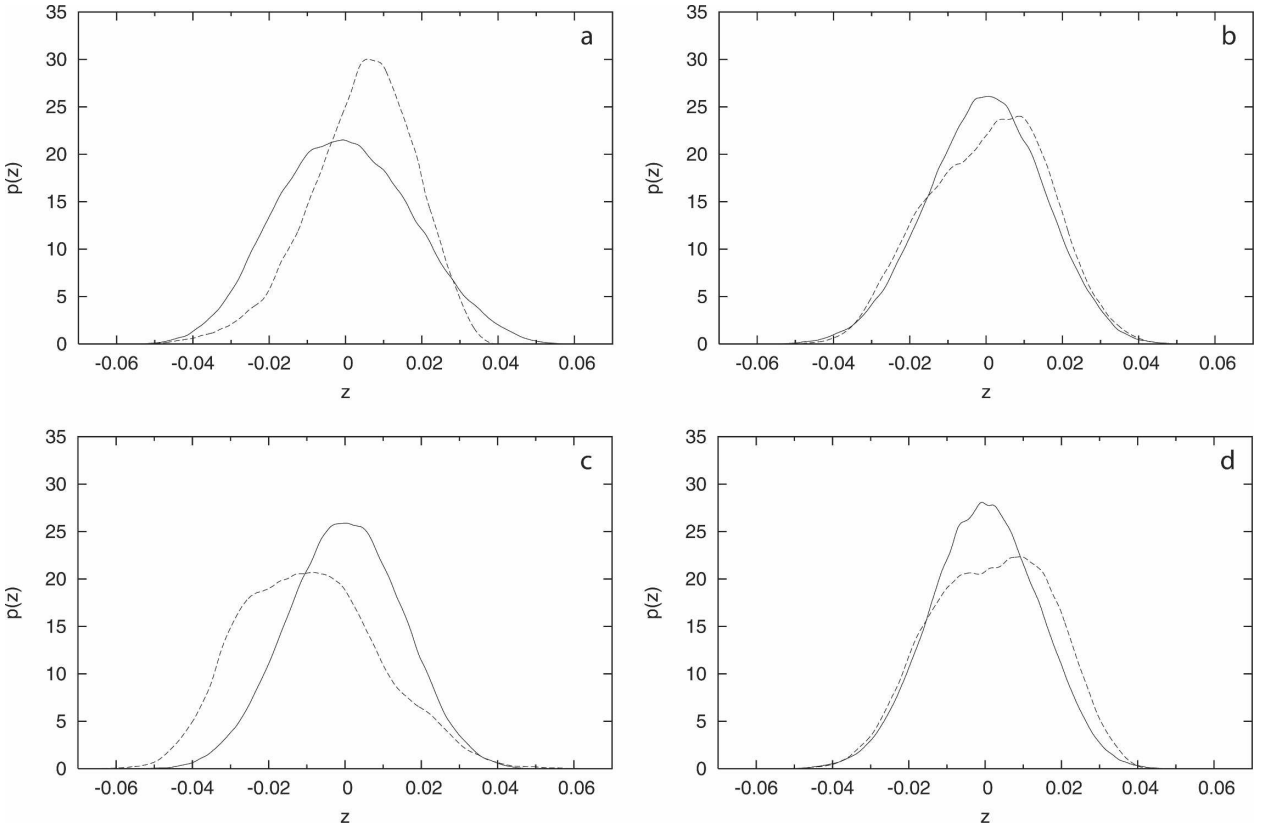

FIG. 10. Probability density function of the (a) first, (b) second, (c) third, and (d) fourth EOF in the QG model (solid) and the reduced model using 15 EOFs (dashed).

exhibits pronounced spurious oscillatory behavior. The low-frequency variability appears to be very hard to model correctly in a highly truncated deterministic low-order model; it seems to be generated by the interaction of quite a large number of modes on different spatial scales.

The PIP algorithm was run for $\eta = 0$, $\eta = 0.5$, $\eta = 1$, $\eta = 2$, $\eta = 5$, and $\eta = 10$. For $\eta \to \infty$, the PIPs coincide with the EOFs. Actually, with $\eta = 5$ the PIPs are already very close to and with $\eta = 10$ virtually indistinguishable from the EOFs and so is the performance of the corresponding reduced models. The patterns with $\eta = 1$ discussed above perform best in a long-term integration. In the following, these patterns are meant when referring to PIPs. With $\eta = 0.5$ and $\eta = 2$, the models are slightly better than EOF models but worse than PIP models. When setting $\eta = 0$, the patterns are determined purely by the cost function of Eq. (45); they will be referred to as purely tendency-based modes. Interestingly, the low-order models built on them perform worse than with any other choice of η discussed here. In particular, they are clearly worse than PIP models and considerably worse than EOF models (not shown). This holds for all statistical quantities shown above and both truncation levels $L = 10$ and $L = 15$.

c. *Tendency analysis*

The question arises of which processes or flow situations are better captured by PIPs than by EOFs leading to their superior performance. To shed some light on why PIP models outperform EOF models, a tendency analysis was done. Three empirical bases are considered: EOFs, PIPs ($\eta = 1$), and purely tendency-based modes ($\eta = 0$). Figure 13 illustrates the correlation between the tendency vectors in the reduced model and the full QG model for the three bases as a function of the number of retained modes. The correlation is given for the full tendency and separately for the linear and the nonlinear part. Figure 14 discusses the magnitude of tendencies in the reduced model and the full QG model for the different modes; again, a separate look at the linear and nonlinear parts of the tendencies is informative. The statistics presented in Figs. 13 and 14 are based on 25 000 days of daily data different from the data sample used to calculate the PIPs and also the EOFs. The statistical uncertainty in

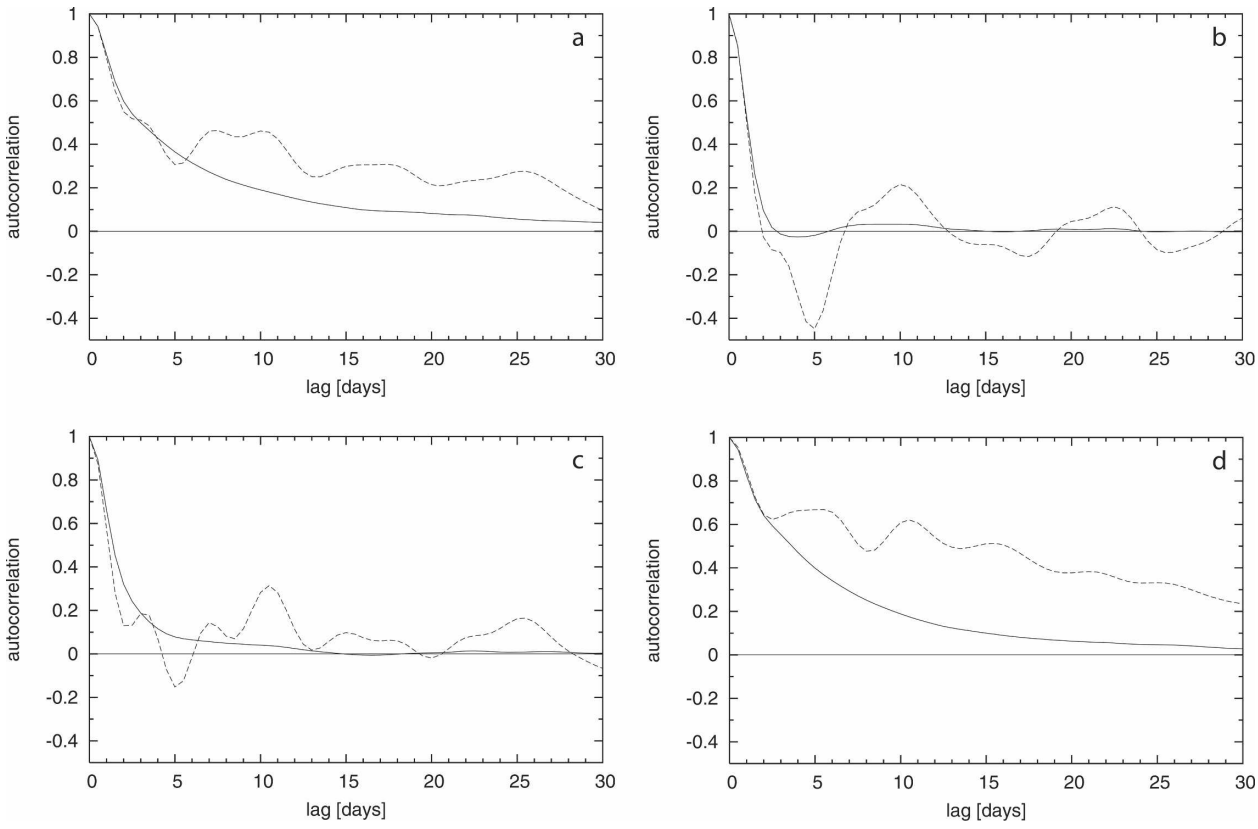

FIG. 11. Autocorrelation function of the (a) first, (b) second, (c) third, and (d) fourth PIP in the QG model (solid) and the reduced model using 15 PIPs (dashed).

all quantities is negligible and therefore not indicated in detail. PIPs have a considerably higher tendency correlation than EOFs; purely tendency-based modes again reach a clearly higher tendency correlation. For the linear part of the tendencies, the results are very similar to those for the full tendencies but the level of correlation is higher for all three bases. In the nonlinear tendencies there is virtually no difference between the three bases, and the level of correlation is much poorer than for the linear part and the full tendencies. More than 30 modes are necessary to reach a modest correlation of 0.6. PIPs are characterized by somewhat larger tendencies than EOFs in the QG model and by clearly larger tendencies in the reduced models. The purely tendency-based modes have much larger tendencies than PIPs and EOFs in both the QG model and the reduced models. The picture for the linear tendencies is very much the same as for the full tendencies. In the magnitude of the nonlinear tendencies there is hardly any difference between PIPs and EOFs both in the T30 and the low-order models. The largest part of the tendencies of the leading modes in the QG model comes from the linear operator, but there are considerable nonlinear contributions. The ratio of the mean squared moduli of the linear and nonlinear tendencies lies between about 4.5 with $L = 3$ and 2.8 with $L = 30$ for both EOFs and PIPs. In the reduced models, the dominance of the linear terms in the tendencies is much more prominent, particularly for highly truncated models. The ratio between the magnitudes of linear and nonlinear terms in the PIP models is as large as about 53, 20, and 11 for $L = 5$, $L = 10$, and $L = 15$. The corresponding figures for EOFs are 44, 13, and 9.

The information provided by Figs. 13 and 14 can be summarized into three main points. First, in all three bases, the interactions among the leading modes are predominantly linear and the improvement of PIPs on EOFs stems entirely from better capturing these linear interactions although the full nonlinear tendencies enter into the cost function. The linear dynamics of the system is quite concentrated onto a relatively small number of large-scale patterns, resulting in high correlations for the linear part of the tendencies even in severely truncated low-order models. The nonlinear tendencies are very small and only weakly correlated with the nonlinear tendencies in the QG model. This is consistent with the picture of large-scale atmospheric flow as stochastically driven linear wave dynamics

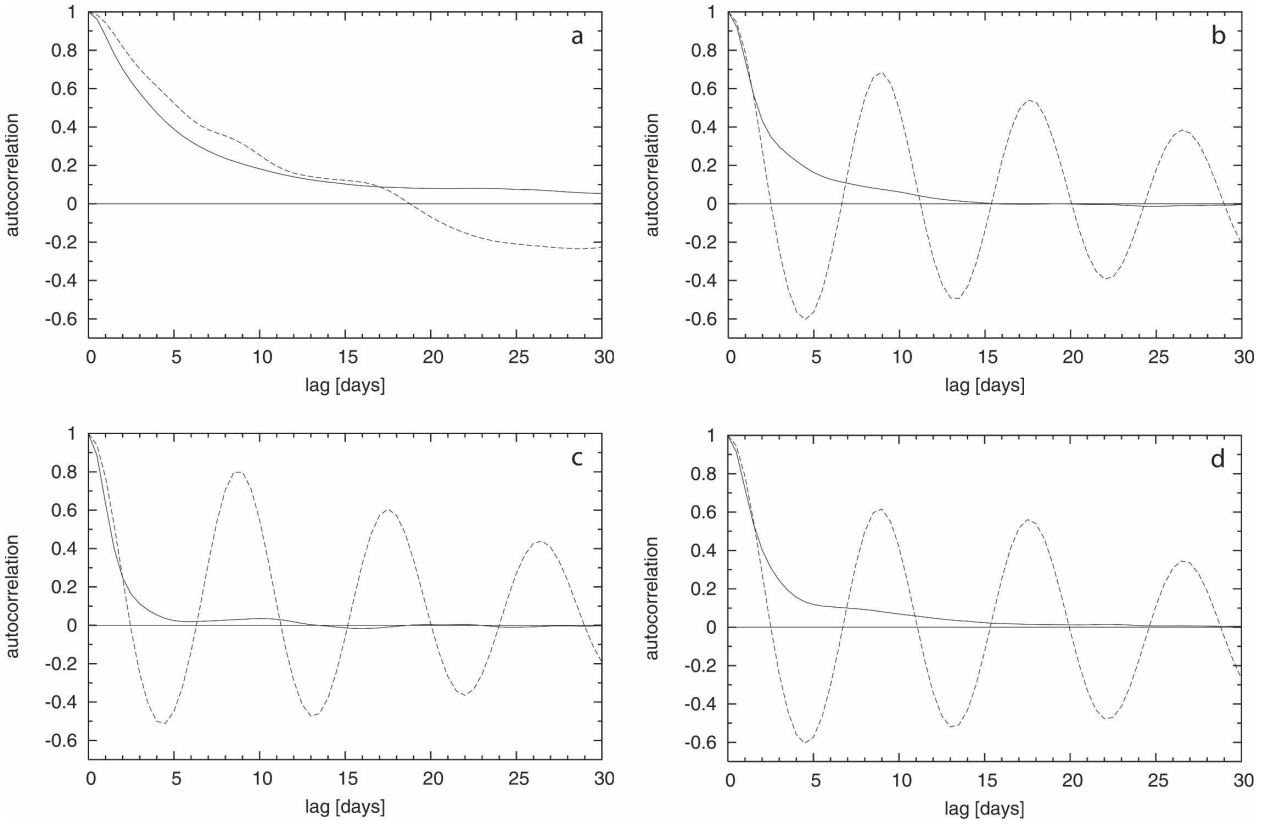

FIG. 12. Autocorrelation function of the (a) first, (b) second, (c) third, and (d) fourth EOF in the QG model (solid) and the reduced model using 15 EOFs (dashed).

(Newman et al. 1997; Branstator and Haupt 1998; Whitaker and Sardeshmukh 1998; Zhang and Held 1999; Winkler et al. 2001). Capturing the linear dynamics appears to be sufficient to model the climatological mean state, variance pattern, and transient momentum fluxes.

Second, there are significant nonlinear contributions in the tendencies of the large-scale modes in the T30 model; this is consistent with a recent study on mean phase space tendencies by Branstator and Berner (2005). This nonlinear part does not stem from interactions between the resolved modes but from nonlinear coupling of the resolved modes with unresolved modes. This nonlinear backscattering cannot be attributed to a few particular patterns but involves a large number of modes at intermediate and small spatial scales and can therefore not be captured in a highly truncated basis either by PIPs or EOFs. Given that even with PIPs, considerable errors in the autocorrelation functions remain, it may be concluded that the nonlinear backscattering from smaller scales to the large scales plays a role in generating the low-frequency variability of the QG model.

Third, the purely tendency-based modes have a considerably higher tendency correlation and also a smaller mean squared tendency error (not shown) than PIPs, yet in a long-term integration they perform worse than PIPs and even EOFs. On the other hand, for example, at $L = 15$ PIPs are only slightly better than EOFs in the tendencies but are considerably better in a long-term integration. This shows that the cost function based on tendencies has to be taken with caution. A smaller tendency error leads to better long-term behavior, but only up to a certain point; strict minimization of the tendency error proves detrimental. PIPs have a careful balance between minimizing tendency error and maximizing explained variance in the resolved modes. This is in accordance with the results of Kwasniok (1996). It is an interesting question whether a cost function based on trajectories rather than tendencies (Kwasniok 1997a, 2001, 2004; Crommelin and Majda 2004) leads to an improvement on the present PIP models, particularly regarding the nonlinear terms and the low-frequency variability.

All PIP calculations presented here use the parameter setting $K_{\max} = 25$. The dependence of the reduced models on the choice of $K_{\max}$ has not been investigated in detail. There may be potential for further improve-

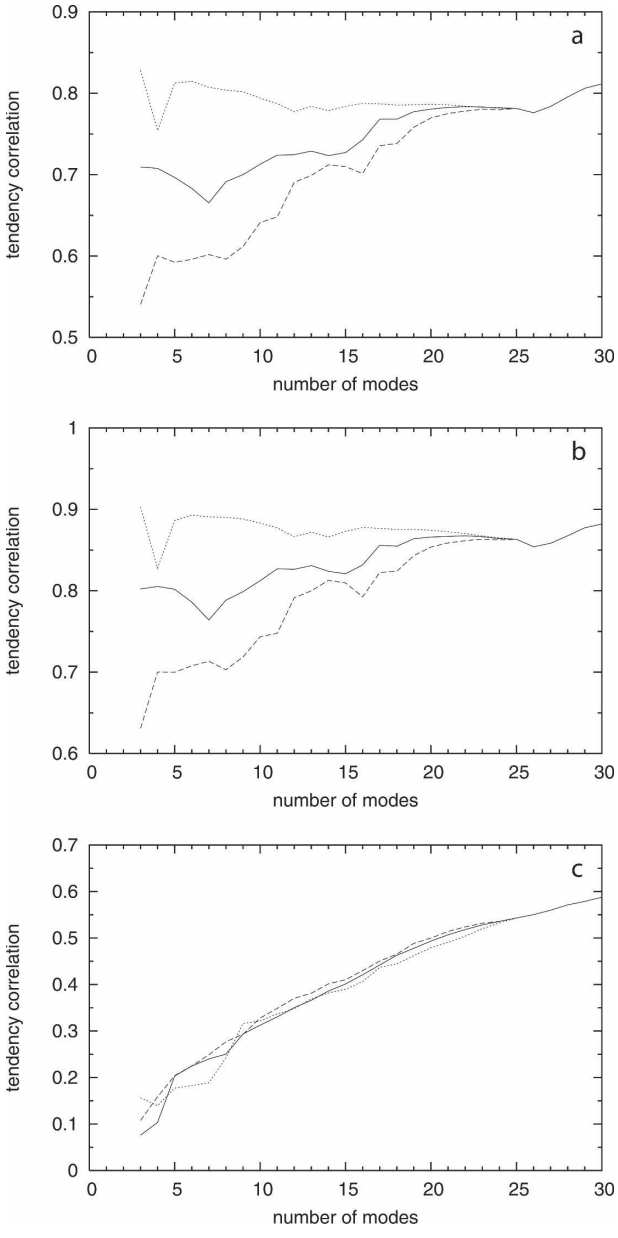

FIG. 13. (a) Correlation between tendency vectors in the QG model and the reduced model for PIPs (solid), EOFs (dashed), and purely tendency-based modes (dotted). (b) As in (a) but for the linear part of tendencies. (c) As in (a) but for the nonlinear part of tendencies.

ment on the present PIP models with a larger value of $K_{\max}$.

When dealing with nonlinear minimization problems the question of suboptimal local minima arises. Surprisingly, this seems to be more a problem for smaller values of K than for larger ones. The inhomogeneities in the tendency correlation at $K = 7$ with PIPs and at $K = 4$ with purely tendency-based modes (Fig. 13) indicate that the algorithm probably got stuck in a local minimum of the cost function there, whereas for larger K it does not seem to be a problem.

Given the dominance of the linear terms in the reduced models, the question arises if a cost function measuring only the mean squared error between the linear tendency in the reduced model and the linear tendency in the QG model already yields the same results. This actually turned out to be the case but with some caution. The parameter η then has to be shifted to higher values. Minimizing a cost function based on only linear tendencies with $\eta = 1$ leads to reduced models that are superior to EOF models but not as good as the PIP models with $\eta = 1$, based on full nonlinear tendencies (not shown); with $\eta = 1.5$ one gets results very close to those discussed above (not shown). Although the minimization is dominated by the linear terms the nonlinear tendencies in the cost function appear to have the effect of a regularization. When using only the linear tendencies from the start, the cost function is of second order in the data and can be expressed as a function of only the covariances of the QG model, making the algorithm much faster. However, the dependence of the cost function on the patterns is still of higher order and the problem remains a nonlinear minimization task. It should be noted that even when using only the linear tendencies in the cost function, the present approach is conceptually different from the method of principal oscillation patterns as reviewed by von Storch et al. (1995). Principal oscillation patterns are the normal modes of an empirical linear operator in the subspace of the leading EOFs whereas in the present work a subspace different from the EOF subspace is identified.

The linear operator occurring in shear flow problems such as the large-scale atmospheric circulation is generically strongly nonnormal (e.g., Farrell and Ioannou 1993). A method of constructing reduced representations of nonnormal linear operators is given by balanced truncation (Farrell and Ioannou 2001). Balanced truncation provides a nearly optimal solution to the problem of finding a truncated dynamical model of a system governed by a stable linear operator under unitary stochastic forcing. A substantial improvement of balanced truncation on EOF truncation has been demonstrated for such systems (Farrell and Ioannou 2001). The QG model formulated as an anomaly model about the climatological mean state [Eq. (23)] has a 19-dimensional unstable manifold; hence, balanced truncation is not applicable here. The fastest growth rate corresponds to an e-folding time of 7.0 days. The relative roles of transient perturbation growth versus perturbation growth due to unstable modes are not clear. It

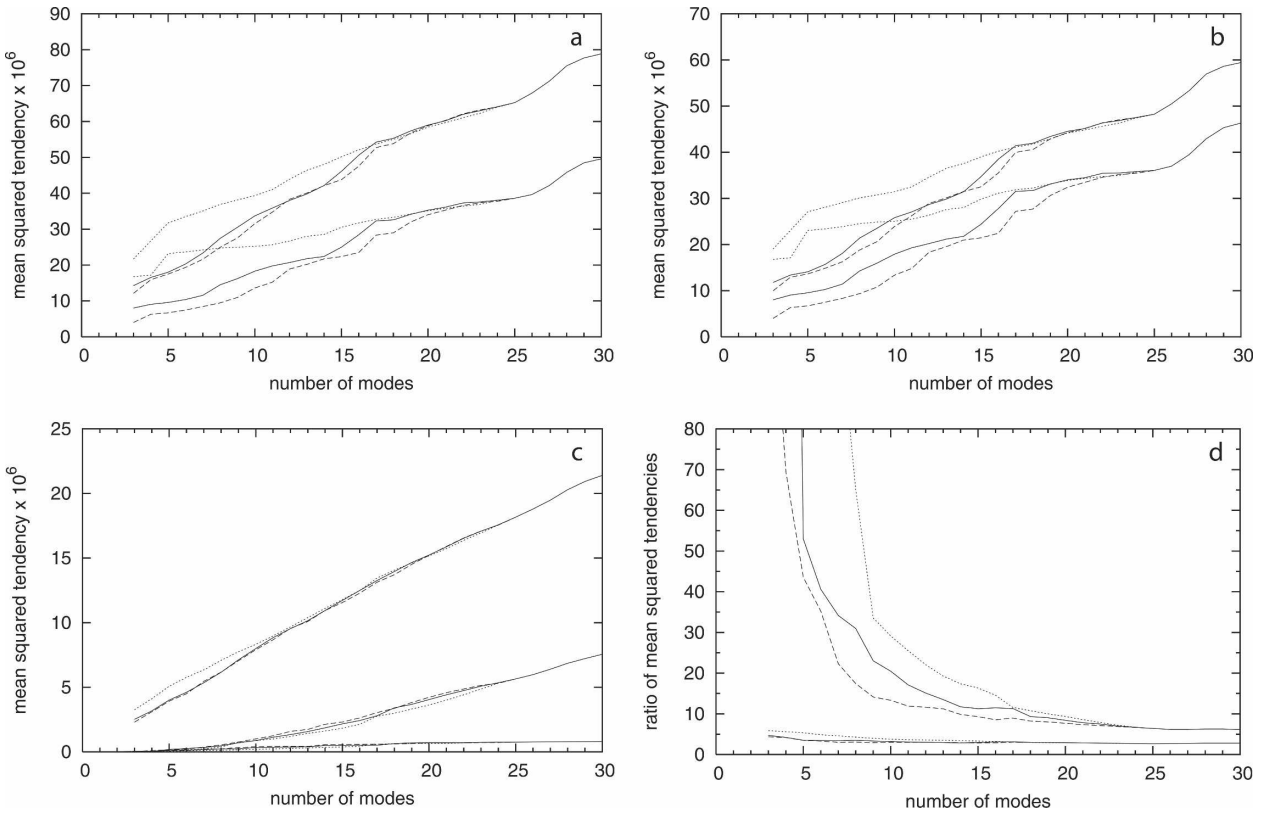

FIG. 14. (a) Mean squared modulus of tendency vector for PIPs (solid), EOFs (dashed), and purely tendency-based modes (dotted). Upper curves refer to the full QG model, lower curves to the reduced model based on the respective modes. (b) As in (a) but for the linear part of tendencies. (c) As in (a) but for the nonlinear part of the tendencies. Additionally, the lowermost curves give the squared modulus of the forcing terms in the different bases. (d) Ratio between mean squared linear tendency and mean squared nonlinear tendency for PIPs (solid), EOFs (dashed), and purely tendency-based modes (dotted). Lower curves refer to the full QG model, upper curves to the reduced model based on the respective modes.

is conceivable that the superior performance of PIP models compared to EOF models is related to better capturing the nonnormal structure of the linear operator. It may be informative to investigate the linear operators of the EOF and PIP models using frequency response functions and pseudospectra and to compare them with each other and the linear operator of the full QG model.

d. Explained variance and spatial structure of PIPs

It is naturally interesting how much variance is explained by the various empirical modes. Figure 15 shows the explained variance of the individual modes and the cumulative variance for EOFs, PIPs ($\eta = 1$), and purely tendency-based modes ($\eta = 0$). The EOFs are by definition optimal in this measure and any other basis is worse. PIPs have only a moderate loss in explained variance compared to EOFs. The compression of phase space is still very good and the leading modes still carry a high explained variance. For example, the first 10 PIPs capture 29.7% of the mean anomaly total energy whereas the first 10 EOFs explain 31.9%. For 15 patterns, the figures are 38.9% and 40.8%, respectively. With the purely tendency-based modes, quite a lot of explained variance is lost; especially, the leading patterns, except for the first, are not high-variance modes any more.

To quantify the similarity or dissimilarity in the spatial structure of the different empirical bases used in the present paper, a measure of distance between two spaces is introduced. For two linear spaces $\mathcal{P}_L^{(1)} = \mathrm{span}\{\mathbf{p}_1^{(1)}, \cdots, \mathbf{p}_L^{(1)}\}$ and $\mathcal{P}_L^{(2)} = \mathrm{span}\{\mathbf{p}_1^{(2)}, \cdots, \mathbf{p}_L^{(2)}\}$, each of dimension L, with $[\mathbf{p}_i^{(1)}, \mathbf{p}_j^{(1)}]_1 = \delta_{ij}$ and $[\mathbf{p}_i^{(2)}, \mathbf{p}_j^{(2)}]_1 = \delta_{ij}$, the distance between them is defined as

$$\mathrm{dist}(\mathcal{P}_L^{(1)}, \mathcal{P}_L^{(2)}) = 1 - \frac{1}{L} \sum_{i,j=1}^{L} [\mathbf{p}_i^{(1)}, \mathbf{p}_j^{(2)}]_1^2. \qquad (46)$$

The measure dist is symmetric: $\mathrm{dist}(\mathcal{P}_L^{(1)}, \mathcal{P}_L^{(2)}) = \mathrm{dist}(\mathcal{P}_L^{(2)}, \mathcal{P}_L^{(1)})$. The value of dist always lies between zero and one. It is zero if and only if the spaces are

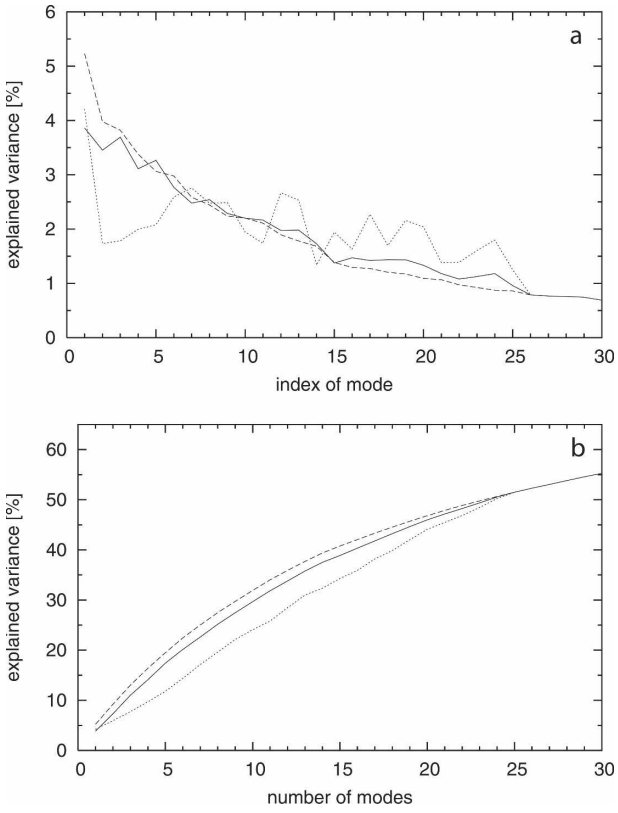

FIG. 15. (a) Explained variance and (b) cumulative explained variance of PIPs (solid), EOFs (dashed), and purely tendency-based modes (dotted).

identical; it is one if and only if the spaces are orthogonal with respect to the total energy scalar product. It is important to note that the measure dist compares spaces rather than individual modes. For any two bases spanning the same space (e.g., EOFs and rotated EOFs), the distance is zero although the individual modes may all be different. This is appropriate in the context of reduced models since a reduced model is determined by the subspace it is based on; a rotation of the basis functions leads to a dynamically equivalent model due to the tensor structure of the interaction coefficients (Kwasniok 1996). A distance between two linear spaces can be defined with respect to any scalar product in exactly the same way as above.

In Fig. 16, the distance between EOF and PIP space and the distance between EOF space and the space spanned by the purely tendency-based modes are given as a function of dimension. The PIP space is noticeably different from the EOF space, particularly for low dimensions, although the loss in explained variance is only moderate. The purely tendency-based modes are clearly more distinct from the EOFs than the PIPs are, except for the first pattern, which is relatively close to the first EOF.

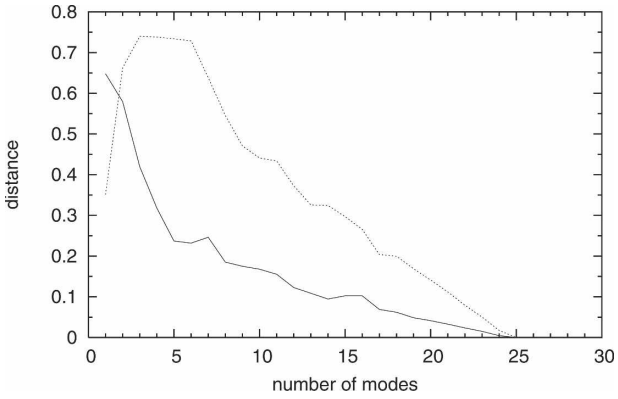

FIG. 16. Distance of PIP space (solid) and space of purely tendency-based modes (dotted) from EOF space as a function of dimension (see text for explanation).

Finally, some light is shed on the question of to what extent the well-known teleconnection patterns of low-frequency variability project onto the leading PIPs and EOFs. The components of the PIPs and EOFs at 500 hPa were orthonormalized with respect to the norm streamfunction metric on that level using a Gram–Schmidt procedure. Then the Pacific–North America (PNA) and North Atlantic Oscillation (NAO) teleconnection patterns at 500 hPa shown in Fig. 3 were projected onto each PIP or EOF, respectively, with respect to norm streamfunction, and subsequently reconstructed from their PIP and EOF expansions. The pattern correlation between the original PNA and NAO patterns and their reconstructions is given in Table 3 for various numbers of modes. Both teleconnection patterns have a higher projection onto the leading PIPs than onto the leading EOFs. For the PNA pattern, the advantage of PIPs is quite small (except with 5 modes) since the PNA pattern is already distinguished by its high explained variance. For the NAO pattern, the improvement of PIPs on EOFs is more pronounced. The NAO pattern does not carry a particularly high variance but seems to play an important role in the linear planetary wave dynamics. This is in accordance with a recent diagnostic dynamical study on teleconnection patterns (Franzke and Feldstein 2005).

6. Conclusions

Nonlinear deterministic reduced atmospheric models are constructed in the framework of a QG three-level model with realistic variability and reasonable PNA and NAO patterns. The reduced model is generated by Galerkin projection of the QG model equations onto empirically determined basis functions. The total energy metric is employed in the projection; the nonlinear

TABLE 3. Pattern correlation of PNA and NAO teleconnection patterns at 500 hPa in the QG model with their projections onto PIP and EOF subspaces of various dimensions.

	PNA		NAO	
No. of modes	PIPs	EOFs	PIPs	EOFs
5	0.95	0.87	0.68	0.50
8	0.96	0.94	0.79	0.66
10	0.96	0.94	0.85	0.73
12	0.97	0.96	0.93	0.83
15	0.97	0.97	0.95	0.94

terms of the low-order model conserve anomaly total energy. No empirical terms are introduced into the dynamical equations apart from retuning the coefficient of horizontal diffusion. Basis functions for efficiently spanning the dynamics are derived using the method of PIPs and compared with the more conventional EOFs. A novel PIP algorithm is introduced that removes modes from the system one by one and holds a careful balance between minimizing tendency error and maximizing explained variance in the resolved modes. The mean streamfunction, the variance of the streamfunction, as well as the transient momentum fluxes are well reproduced by a PIP model with 10 degrees of freedom. With 15 PIPs, probability density functions are accurately modeled; however, in the autocorrelation functions a considerable error remains. Models based on PIPs markedly outperform models based on EOFs. A tendency analysis reveals that, both with EOFs and PIPs, the interactions between the leading patterns are predominantly linear and the improvement of PIPs on EOFs comes entirely from better capturing these linear interactions. In the full QG model, there is a significant nonlinear contribution in the tendencies of the leading EOFs and PIPs. It stems from nonlinear coupling of the large-scale modes with a large number of unresolved smaller-scale modes and is not captured in a highly truncated model, either with EOFs or with PIPs. This nonlinear backscattering appears to play a role in forming the low-frequency variability in the QG model.

For future research, it would be most interesting to combine the PIP approach with a stochastic closure scheme in order to arrive at a nonlinear stochastic low-order model. This may be either a systematic strategy for stochastic mode reduction as proposed recently (Majda et al. 1999, 2003; Franzke et al. 2005; Franzke and Majda 2006) or an empirical scheme.

Acknowledgments. Part of this work has been done during employment at the Leibniz Institute for Atmospheric Physics in Kühlungsborn, Germany.

REFERENCES

Achatz, U., and G. Branstator, 1999: A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *J. Atmos. Sci.,* **56,** 3140–3160.

——, and J. D. Opsteegh, 2003: Primitive-equation-based low-order models with seasonal cycle. Part I: Model construction. *J. Atmos. Sci.,* **60,** 465–477.

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the northern hemisphere. *J. Atmos. Sci.,* **33,** 1607–1623.

Branstator, G., and S. E. Haupt, 1998: An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *J. Climate,* **11,** 2645–2667.

——, and J. Berner, 2005: Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: Phase space tendencies. *J. Atmos. Sci.,* **62,** 1792–1811.

Charney, J. G., and J. G. de Vore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.,* **36,** 1205–1216.

Crommelin, D. T., and A. J. Majda, 2004: Strategies for model reduction: Comparing different optimal bases. *J. Atmos. Sci.,* **61,** 2206–2217.

D'Andrea, F., and R. Vautard, 2001: Extratropical low-frequency variability as a low-dimensional problem. I: A simplified model. *Quart. J. Roy. Meteor. Soc.,* **127,** 1357–1374.

Farrell, B. F., and P. J. Ioannou, 1993: Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.,* **50,** 4044–4057.

——, and ——, 2001: Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.,* **58,** 2771–2789.

Franzke, C., and S. B. Feldstein, 2005: The continuum and dynamics of Northern Hemisphere teleconnection patterns. *J. Atmos. Sci.,* **62,** 3250–3267.

——, and A. J. Majda, 2006: Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.,* **63,** 457–479.

——, ——, and E. Vanden-Eijnden, 2005: Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.,* **62,** 1722–1745.

Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

Kwasniok, F., 1996: The reduction of complex dynamical systems using principal interaction patterns. *Physica D,* **92,** 28–60.

——, 1997a: Optimal Galerkin approximations of partial differential equations using principal interaction patterns. *Phys. Rev. E,* **55,** 5365–5375.

——, 1997b: Low-dimensional models of complex systems using principal interaction patterns. *Nonlinear Anal.,* **30,** 489–494.

——, 2001: Low-dimensional models of the Ginzburg-Landau equation. *SIAM J. Appl. Math.,* **61,** 2063–2079.

——, 2004: Empirical low-order models of barotropic flow. *J. Atmos. Sci.,* **61,** 235–245.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 1999: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA,* **96,** 14 687–14 691.

——, ——, and ——, 2003: Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.,* **60,** 1705–1722.

Marshall, J., and F. Molteni, 1993: Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.,* **50,** 1792–1818.

Newman, M., P. D. Sardeshmukh, and C. Penland, 1997: Stochastic forcing of the wintertime extratropical flow. *J. Atmos. Sci.,* **54,** 435–455.

Rinne, J., and V. Karhila, 1975: A spectral barotropic model in horizontal empirical orthogonal functions. *Quart. J. Roy. Meteor. Soc.,* **101,** 365–382.

Roads, J. O., 1987: Predictability in the extended range. *J. Atmos. Sci.,* **44,** 3495–3527.

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

——, 1997: Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.,* **54,** 2099–2114.

von Storch, H., G. Burger, R. Schnur, and J. S. von Storch, 1995: Principal oscillation patterns—A review. *J. Climate,* **8,** 377–400.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–804.

Whitaker, J. S., and P. D. Sardeshmukh, 1998: A linear theory of extratropical synoptic eddy statistics. *J. Atmos. Sci.,* **55,** 237–258.

Winkler, C. R., M. Newman, and P. D. Sardeshmukh, 2001: A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *J. Climate,* **14,** 4474–4494.

Zhang, Y., and I. M. Held, 1999: A linear stochastic model of a GCM's midlatitude storm tracks. *J. Atmos. Sci.,* **56,** 3416–3435.