On the Closure Problem in the Reduction of Complex Atmospheric Models by PIPs and EOFs: A Comparison for the Case of a Two-Layer Model with Zonally Symmetric Forcing

U. ACHATZ AND G. SCHMITZ

Institut für Atmosphärenphysik an der Universität Rostock, Kühlungsborn, Germany

ABSTRACT

The reduction of complex systems to their essential degrees of freedom (e.g., patterns spanning that part of state space that the system's trajectory passes in time) and development of reduced models based on these might be one tool for improving basic understanding of simulation results obtained from current, increasingly more complex, meteorological circulation models. Some successful work on the reduction of complex systems by principal interaction patterns (PIPs) or empirical orthogonal functions (EOFs) has already been done. However, the parameterization of the influence of unresolved modes onto the selected patterns, called closure in this context, is still an outstanding problem. Nonlinear closure schemes, although greatly improving prediction on shorter timescales, have so far been observed to lead to absolute instability of the reduced model, that is, explosive unbounded energy growth after some integration time. In this work a method is outlined for circumventing this problem. Energy conservation constraints are formulated that can be used in the extraction of a useful empirical closure from synthetic model data by minimizing the error between tendencies in the full and the reduced model with the closure parameters as variables. In the present context the computational size of the associated minimization calculations could be reduced by utilizing the zonal symmetry in the full model's forcing and boundary conditions. So it could be shown that each EOF or PIP must be an element of the subspace of a single zonal wavenumber. Coupling conditions for the closure coefficients are derived that further decrease the dimensionality of the problem. The method is tested by reducing multiple baroclinic wave life cycles in a quasigeostrophic two-layer model on the basis of both EOFs and PIPs. It is shown that the aforementioned stability problems connected with the nonlinearity of a closure are indeed avoided by the method. Furthermore, the closure greatly improves the simulation capabilities of the reduced model both on short and long timescales. In contrast to previous results on linear closure schemes, the authors find that in nonlinear closure schemes also the linear terms have to be handled carefully in order to ensure realistic behavior of the reduced model on longer timescales. In a comparison between reductions based on EOFs and PIPs substantial superiority of the latter in effectively extracting the essential degrees of freedom is demonstrated.

1. Introduction

Parallel to a dramatic increase in available computing power, meteorology has experienced a continuous rise in the complexity of standard models used for the study of atmospheric behavior. Thus, more and more realistic simulations of ever finer details have become possible. With their growth in complexity, such models, however, also lead to an increasing need for tools to help us gain basic understanding of simulation results. Among others, one class of such tools might be algorithms providing a reduction of complex systems to their essential degrees of freedom. They could considerably increase the transparency of the examined system and give us helpful clues to a deeper understanding of underlying

processes and of the interaction between dominant components contributing to atmospheric variability on various timescales.

Several methods have been suggested for the identification of meaningful basic patterns of atmospheric dynamics. Point-correlation maps of atmospheric data have proven very useful for the identification of major hemispheric-scale flow patterns such as the Pacific/North American and North Atlantic Oscillation patterns (Wallace and Gutzler 1981). Frederiksen (1982, 1983) and Frederiksen and Bell (1987) examined the most unstable normal modes of a quasigeostrophic two-layer model with orography and nonzonal forcing. They were able to associate these as well to elementary processes like cyclogenesis and onset of blocking as to the aforementioned teleconnection patterns. Simmons et al. (1983) could show that teleconnection patterns can already be identified among the normal modes of a similarly forced barotropic model. A useful method for extracting dynamically relevant empirical normal modes

Corresponding author address: Dr. Ulrich Achatz, Institut für Atmosphärenphysik an der Universität Rostock, Schloßstraße 4-6, 18225 Kühlungsborn, Germany.
E-mail: achatz@iap-kborn.d400.de

directly from observed data has been suggested by Brunet (1994). A set of patterns to be seen in contrast to normal modes, and possibly better suited for an insightful solution, especially of meteorological initial value problems, are optimal vectors (e.g., Farrell and Ioannou 1996 and references therein). Another interesting approach is EOF analysis. Rhinne and Karhilla (1975) and Schubert (1985, 1986) have investigated dynamical interactions between EOFs determined from atmospheric observations. They could show that a few dominant EOFs are able to describe the predominant part of the observed tendencies at a quality comparable to gridpoint or spectral models with much larger numbers of degrees of freedom. Another possibility is direct utilization of observed dynamics in the form of measured tendencies. This is the aim of a principal interaction pattern (PIP) analysis (Hasselmann 1988). Assuming some general form for the nonlinear equations determining the interaction between the different structures, both patterns and interaction coefficients are determined simultaneously by minimizing a function that measures the error between tendencies as given by the reduced model and the full system. Useful results have already been achieved in identifications of patterns with a predominantly linear behavior, so-called principal oscillation patterns (POP). Among numerous examples an interesting work is that of Schnur et al. (1993), where POPs have been identified in atmospheric data resembling baroclinic waves in their respective growth and decay phases. First successful nonlinear PIP analyses of atmospheric models have been performed by Achatz et al. (1995) and Kwasniok (1996). A further development of the technique using adjoint methods has recently been published by Kwasniok (1997).

Reduced models, that is, starting from a number of the identified patterns that should be much smaller than the number of degrees of freedom of the full system, evolution equations for the expansion coefficients of the system's state vector with respect to these, have been formulated and examined for the case of EOFs (Selten 1993, 1995a,b) and PIPs (see references above). In these cases reduced, projected, and truncated models have been obtained by approximating the system's state vector by a superposition of these patterns, inserting it into the known dynamical equations of the full system, and projecting these via a Galerkin projection onto the patterns under consideration. For the examined—relatively simple—systems it has been demonstrated that reduced models are able to simulate the full system faithfully on the basis of a much reduced number of degrees of freedom. Such models, however, cannot achieve the most effective possible reduction since they do not include any parameterization of the influence on the considered patterns by modes not explicitly taken into account in the reduction. In order to achieve faithful simulations such models therefore usually still must include comparably many patterns. This might be avoided by using a low-order model that differs from the purely projected

and truncated version by some so-called closure scheme taking the coupling between unresolved modes and leading patterns into account—resembling closely the attempt to parameterize in spectral circulation models the effect of neglected small-scale spherical harmonics by the usually added horizontal diffusion. In other words, the art of finding the best possible reduced model consists in finding the best possible closure. Certainly it might be most attractive to derive such a scheme in an analytical manner (e.g., as described by Lindenberg and West 1984). However, in most applications this seems to be extremely difficult. Therefore an empirical–statistical approach might have better prospects as a practical way to determine for a given set of patterns a useful closure parameterization. Some first successful steps in this direction have already been taken. Kwasniok (1996) included in his PIP algorithm a free fit of the linear tensor of the reduced model, thereby obtaining notable improvement in prediction and simulation capability over the truncated model. Similar success for EOF models has been reported by Selten (1995a). Selten (1995b) has even gone a step further by determining all interaction coefficients of his quadratic EOF model such that they are best able to reproduce the local tendencies of the full system. The reduced system thus identified is indeed much more powerful as for its short-term predictive capabilities. A drawback so far, however, is that long-term integrations of the model are not possible since after about 20 days its energy starts to grow without bounds. It has been suggested by the author that this problem might be removed by two different approaches. The first would be to determine the reduced model from data of the full system that are slightly perturbed away from its true climate. The system tendencies used for the determination of the low-order model are then hoped to contain enough pull toward this climate so that this might also be reflected in the tendencies simulated by the low-order model, thereby preventing it from moving away too far from the projected system's climate. The second approach consists in finding appropriate constraints for the closure corrections so that in the adiabatic case full energy conservation is ensured. In this case the model energy would be forbidden to grow without bounds. It can then be hoped that this property is not destroyed by moving to the more realistic nonadiabatic case.

Here we follow the second approach in a test of the possibilities of energy-conserving closures in the reduction of multiple baroclinic wave life cycles in a quasigeostrophic two-layer model on the sphere. This specific scenario is of special meteorological interest by itself. For references on the phenomenon see, for example, Achatz et al. (1995). Its quasi-periodic character, however, also makes it easier to select useful criteria for an evaluation of the quality of the reduced model. Furthermore, the number of data to be analyzed under avoidance of sampling problems can still be rather modest. The aim of the analysis is twofold: First, to examine

several methods to obtain the best closed reduced model for our full system and, secondly, to compare the potential of EOFs and PIPs as basic patterns for low-order models with energy-conserving closure. Consideration is being given only to these two kinds of basis patterns because so far they have been shown to be well suited for model reductions. For others, for example, normal modes or POPs, no way has been described yet how to select among the complete basis provided by the pattern algorithm the small number of those patterns that are most important for and best apt to provide a good reduced model. The corresponding difficulty lies in the nonnormality of the basis, leading typically to considerable projection errors when only few of the patterns are selected. Additionally, because we are looking for *nonlinear* reduced models, the nice property of normal modes or POPs to diagonalize (in complex space) the operator describing the linear part (in the case of normal modes with respect to some reference state) of the tendencies of the system is not much helpful any more. The paper is thus structured as follows. In section 2 we will describe the two-layer model and the time series examined. Section 3 contains some useful results on EOFs and EOF models for systems with zonally symmetric boundary conditions and physics, describes both our approach to ensure that a closure for an EOF model does not destroy the reduced models stability, and reductions of the multiple baroclinic wave life cycles by the method. Section 4 repeats the same for PIPs and compares the results of PIP analysis with the ones for EOF models. Section 5 summarizes the essentials and discusses possible extensions and generalizations of this work.

2. Data

a. The two-layer model

The data used in this work have been obtained by numerical integration of the equations of a very simple two-layer model. Nevertheless, it has sufficient similarity to the real atmosphere so as to be able to reproduce the main features of baroclinic wave life cycles. Starting with the quasigeostrophic version of the spherical two-layer model as formulated by Lorenz (1960) surface friction and Newtonian cooling are included. Furthermore, we have incorporated forcing terms that make a chosen zonal wind configuration a stationary solution of the system. In order to numerically stabilize the integrations weak horizontal diffusion is also needed. Static stability is assumed to be a constant. The most drastic simplification is that the Coriolis parameter is replaced by a latitude-independent value in the thermal wind equation as well as in the term describing vertical advection [B-model; e.g., Baines and Frederiksen (1978)]. Thus we finally end up with two equations giving the necessary time derivatives:

$$\frac{\partial}{\partial t}\nabla^2\psi = -J(\psi, \nabla^2\psi) - J(\tau, \nabla^2\tau) - 2\frac{\partial\psi}{\partial\lambda}$$

$$+ k_s\nabla^2(\tau - \psi) + F_\psi$$

$$- \kappa\nabla^2(\nabla^2(\nabla^2\psi)) \qquad (1)$$

$$\frac{\partial}{\partial t}(\nabla^2 - r^2)\tau = -J[\psi, (\nabla^2 - r^2)\tau] - J(\tau, \nabla^2\psi)$$

$$- 2\frac{\partial\tau}{\partial\lambda} - k_s\nabla^2(\tau - \psi) - h_N r^2(\tau_f - \tau)$$

$$- \kappa(\nabla^2 - r^2)[\nabla^2(\nabla^2\tau)]. \qquad (2)$$

Barotropic and baroclinic streamfunctions are denoted by $\psi = (\phi_3 + \phi_1)/2$ and $\tau = (\phi_3 - \phi_1)/2$, where ϕ_3 and ϕ_1 are the upper and lower layer streamfunctions; J is the Jacobian operator. The equations are nondimensionalized using a_0, $1/\Omega$, and $(a_0^2\Omega^2)/(c_p b)$ as length, time, and potential temperature scale, respectively. Here a_0 is the radius of earth, Ω its angular velocity, c_p the heat capacity of air at constant pressure, and $b = 0.124$; k_s (in this work 0.25 d^{-1}) and h_N (0.1 d^{-1}) are Ekman surface friction and Newtonian cooling, respectively; F_ψ denotes zonal-mean vorticity forcing and τ_f a forced zonal shear corresponding to the temperature distribution forced by Newtonian cooling. Forcing was chosen such that the equilibrium configuration of the model is zonally symmetric with $u_1 = 0$ and $u_3 = (20$ m s^{-1}) sin$^2 2\varphi$. Here κ is the horizontal diffusion applied to the equations; its value is $\kappa = 1 \times 10^{-6}a^4/$d, which is about as small as possible to prevent the adiabatic limit of the baroclinic wave life cycles examined here from eventually piling up energy at small scales. The parameter $r = 2$ sin$\varphi_0/(\sqrt{\overline{\sigma}})$ contains the latitude φ_0 chosen for the evaluation of the Coriolis parameter (45°N) and the static stability $\overline{\sigma}$. In the cases presented here we have used a static stability $\overline{\sigma} = 0.01$ corresponding to a temperature difference between the layers of 34.5 K. For details of the derivation of (1) and (2) see Achatz et al. (1995). Because the quasigeostrophic approximation is invalid near the equator and variation of the Coriolis force with latitude is incompletely described by the model, the scenario examined was chosen to be equatorially symmetric. At least in middle latitudes the error made by the model should not be too dramatic.

For the determination of the synthetic dataset all variables are represented by a truncated expansion in terms of spherical harmonics. Horizontal resolution comprises a triangular spectral truncation at zonal and total wavenumber 42. An explicit time scheme using a variable-order, variable-step Adams method has been applied as the integration scheme for the spectral equations.

b. The time series

In order to obtain the desired multiple baroclinic wave life cycles the model has been initialized with a zonal-mean wind being equal to the equilibrium configuration
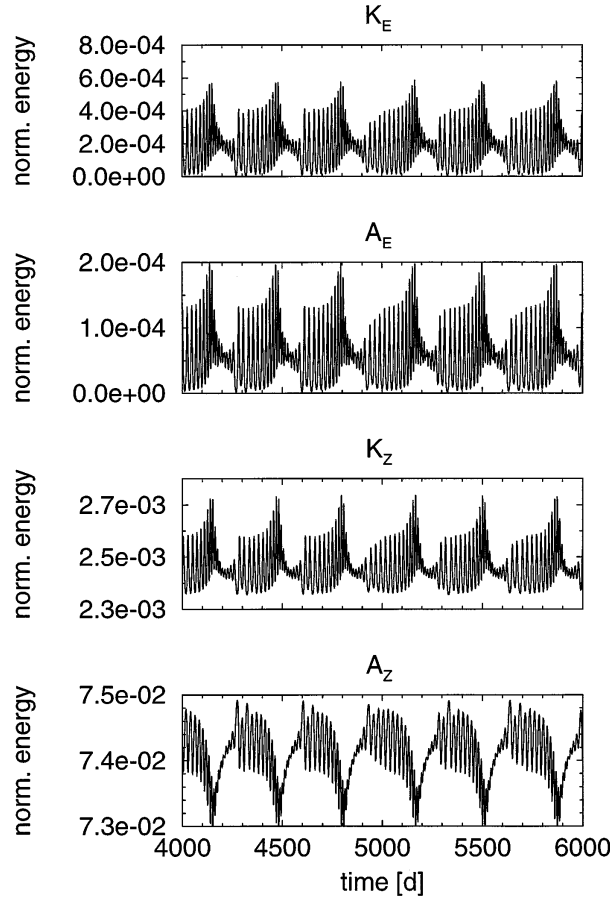
FIG. 1. The four Lorenzian energy types for days 4000–6000 out of the time series to be analyzed. Normalization constant is $E_0 = (p_0\Omega^2 a_0^4)/g$ ($=9.042 \times 10^{22}$ J), where p_0 is the surface pressure, Ω the earth's angular frequency, a_0 its radius, and g acceleration due to gravity. Here K and A denote kinetic and available potential energy of the eddies (index E) or the zonal-mean wind (index Z).

and a small contribution of the most unstable normal mode at zonal wavenumber $m = 6$ as determined from an instability analysis of the equilibrium profile. This initial condition restricts the state vector to be confined to the subspaces belonging to zonal wavenumbers 0, 6, 12, . . . , 42. Given the equatorial symmetry (so that spectral coefficients of both streamfunctions are only nonzero if the total wavenumber $n = m + 2k + 1$, $k = 0, 1, 2, \ldots$), the number of degrees of freedom in the problem is thus 42 (real) in zonally symmetric subspace and 126 (complex) in wave subspace. The model has then been integrated for 14 000 d. The resulting four Lorenzian energy types for days 4000 to 6000 are shown in Fig. 1. The multiple life cycles are clearly visible. Two relevant timescales are discernible: one characterizing the life cycles and one for a modulation of these with a period of about 400 d. The ability of the reduced model to reproduce this behavior will be one criterion in an evaluation of its quality in the further analysis.

3. Reduced models on the basis of EOFs

Before going into any details of the reduction of our life cycles on the basis of leading EOFs, we will first derive some useful properties of both patterns and reduced model from the zonal symmetry of the system's forcing and boundary conditions. After this we will discuss a general strategy to ensure a realistic time dependence of energy in the reduced model. Here the appropriate constraints on the closure will be specified. Finally in this section, we will examine two different approaches to determine a good closure from the data and discuss their respective merits and drawbacks.

a. Consequences of zonal symmetry in forcing and boundary conditions

The two-layer model has no orography. Its forcing is purely zonally symmetric. Therefore no longitude should be dynamically distinguishable from any other. As a consequence, averaging any product of the deviations of two variables from their time mean at different horizontal locations over long time should result in a covariance function that only depends on the difference between the two longitudes but not on the longitudes themselves. Taking, for example, these variables to be the barotropic and baroclinic streamfunctions, this can be written

$$\frac{1}{n_t}\sum_t [\psi(\lambda, \varphi, t) - \hat{\psi}(\varphi)][\tau(\lambda', \varphi', t) - \hat{\tau}(\varphi')]$$

$$= C(\lambda - \lambda', \varphi, \varphi'), \qquad (3)$$

where n_t denotes the number of times taken for the average. Due to the zonal symmetry in boundary conditions and forcing, the time means $\hat{\psi}$ and $\hat{\tau}$ must also be independent of longitude. Using the expansions,

$$\psi(\lambda, \varphi, t) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \psi_n^m(t) Y_n^m(\lambda, \varphi)$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \psi_n^m(t) e^{im\lambda} P_n^m(\varphi) \qquad (4)$$

$$\tau(\lambda, \varphi, t) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \tau_n^m(t) Y_n^m(\lambda, \varphi)$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \tau_n^m(t) e^{im\lambda} P_n^m(\varphi), \qquad (5)$$

and analogous ones for the time means, where Y_n^m and P_n^m are the well-known spherical harmonics and associated Legendre functions, together with the reality of ψ, $\hat{\psi}$, τ, and $\hat{\tau}$, we derive

$$C(\lambda - \lambda', \varphi, \varphi')$$

$$= \sum_{m,m'} e^{i(m\lambda - m'\lambda')} \sum_{n,n'} P_{n'}^{m'}(\varphi') P_n^m(\varphi)$$

$$\times \frac{1}{n_t}\sum_t \left[\overline{\psi_{n'}^{m'}(t)} - \overline{\hat{\psi}_{n'}^{m'}}\right][\tau_n^m(t) - \hat{\tau}_n^m]. \qquad (6)$$

The overbar denotes complex conjugation. Equation (6) can only be true if

$$\frac{1}{n_t} \sum_t \left[\overline{\psi_{n'}^{m'}}(t) - \overline{\hat{\psi}_{n'}^{m'}} \right][\tau_n^m(t) - \hat{\tau}_n^m] \sim \delta_{mm'}. \quad (7)$$

Similarly one can show that

$$\frac{1}{n_t} \sum_t \left[\overline{\psi_{n'}^{m'}}(t) - \overline{\hat{\psi}_{n'}^{m'}} \right][\psi_n^m(t) - \hat{\psi}_n^m] \sim \delta_{mm'}, \quad (8)$$

$$\frac{1}{n_t} \sum_t \left[\overline{\tau_{n'}^{m'}}(t) - \overline{\hat{\tau}_{n'}^{m'}} \right][\tau_n^m(t) - \hat{\tau}_n^m] \sim \delta_{mm'}, \quad (9)$$

and

$$\frac{1}{n_t} \sum_t \left[\overline{\tau_{n'}^{m'}}(t) - \overline{\hat{\tau}_{n'}^{m'}} \right][\psi_n^m(t) - \hat{\psi}_n^m] \sim \delta_{m,m'}. \quad (10)$$

The scatter matrix of our system does not couple different zonal wavenumbers.

This has consequences for the EOFs we can extract from our data. Let us line up for convenience all independent spectral coefficients (i.e., with nonnegative zonal wavenumber) of both streamfunctions in a complex state vector of our system. With $\underset{\sim}{M}$ being the metric of our state space and $\delta\Phi = \Phi - \hat{\Phi}$ the deviation of the state vector Φ from its time mean, EOFs of this turbulent part of the data are solutions of the eigenvalue problem

$$\underset{\sim}{S}\underset{\sim}{M}\mathbf{e}_\nu = \lambda_\nu^2 \mathbf{e}_\nu, \quad (11)$$

$$\underset{\sim}{S} = \frac{1}{n_t} \sum_t \overline{\delta\Phi}\, \delta\Phi \quad (12)$$

is the scatter matrix. The eigenvalues are identical with the variance associated with each pattern. From the above we see that *if forcing and boundary conditions are zonally symmetric, and the metric does not couple different zonal wavenumbers, then each EOF is completely contained in the subspace of one zonal wavenumber.* The interested reader can easily convince himself that this result holds for arbitrary atmospheric models with zonal symmetry in all processes entering its dynamic equations.

A dynamical interpretation of this property can be given as follows. Given the zonal symmetry, it should be expected that every EOF is translationally invariant with respect to longitude in such a way that any longitudinal shift of the pattern can be completely achieved by a change only in the degrees of freedom directly associated with this pattern. Indeed, since for each EOF $\mathbf{e}_\nu$ there is a single zonal wavenumber m_ν, starting from a state vector that is proportional to this EOF, a shift in longitude by a distance $\Delta\lambda$ is achieved by multiplying its expansion coefficient by a phase factor, $\exp(im_\nu\Delta\lambda)$. No change in the expansion coefficients of any other pattern is necessary.

We will now use our result for the derivation of some general properties of the dynamic equations of low-or-

der models for our system. However, first let us introduce some ordering among the patterns that will make the dynamical structure of our low-order models more transparent. From now on we will distinguish between zonally symmetric patterns $\mathbf{e}_\nu^z$ and pure wave patterns ($m_\nu > 0$) $\mathbf{e}_\nu^w$ with associated expansion coefficients z_ν and w_ν, respectively. The reality conditions for each streamfunction yield

$$\overline{\psi_n^m} = \psi_n^{-m} \quad (13)$$

$$\overline{\tau_n^m} = \tau_n^{-m}, \quad (14)$$

so that both the zonally symmetric patterns and their expansion coefficients are completely real. Avoiding any nonlinearities that are of an order larger than 2, the most general low-order model (including the closure) for the system on the basis of leading EOFs, observing the above reality condition, can now be written in two types of tendency equations:

$$\dot{z}_\nu = \sum_\mu A_{\nu\mu}(z_\mu - Z_\mu) + \sum_{\mu\rho} G_{\nu\mu\rho}(z_\mu - Z_\mu)(z_\rho - Z_\rho)$$
$$+ \sum_{\mu\rho} B_{\nu\mu\rho} w_\mu \overline{w_\rho} + \sum_\mu (H_{\nu\mu} w_\mu + \overline{H_{\nu\mu}}\, \overline{w_\mu})$$
$$+ \sum_{\mu\rho} (I_{\nu\mu\rho} z_\mu w_\rho + \overline{I_{\nu\mu\rho}} z_\mu \overline{w_\rho})$$
$$+ \sum_{\mu\rho} (J_{\nu\mu\rho} w_\mu w_\rho + \overline{J_{\nu\mu\rho}}\, \overline{w_\mu}\, \overline{w_\rho}) \quad (15)$$

and

$$\dot{w}_\nu = \sum_\mu C_{\nu\mu} w_\mu + \sum_{\mu,\rho} D_{\nu\mu\rho} z_\mu w_\rho + \sum_{\mu,\rho} E_{\nu\mu\rho} w_\mu w_\rho$$
$$+ \sum_{\mu,\rho} F_{\nu\mu\rho} w_\mu \overline{w_\rho} + \sum_\mu K_{\nu\mu} \overline{w_\mu}$$
$$+ \sum_{\mu,\rho} L_{\nu\mu\rho} z_\mu (z_\rho - Z_\rho) + \sum_{\mu,\rho} M_{\nu\mu\rho} z_\mu \overline{w_\rho}$$
$$+ \sum_{\mu,\rho} N_{\nu\mu\rho} \overline{w_\mu}\, \overline{w_\rho}. \quad (16)$$

Here Z_ν denotes the expansion coefficients of the zonally symmetric equilibrium configuration of the low-order model, which is assumed to exist in analogy to the full system. Since the full two-layer model does not contain wave forcing, an analogous term is not included in (16) either. A great simplification can be derived from the observation that the dynamics must be invariant under every longitudinal translation. Shifting the whole system by a longitudinal distance $\Delta\lambda$ corresponds to multiplying each w_ν and its tendency in (15) and (16) by a phase factor $\exp(im_\nu\Delta\lambda)$. The coefficients of zonally symmetric patterns remain unchanged as well as their tendencies. Invariance of the dynamics implies that lhs and rhs of both equations remain identical under the translation. Since $m_\nu \geq 0$ this can, however, only be ensured for every possible translation if only tensors A–G are nonzero. Additionally they must fulfill the coupling conditions

$$B_{\nu\mu\rho} = 0 \quad \text{if } m_\mu \neq m_\rho, \tag{17}$$

$$C_{\nu\mu} = 0 \quad \text{if } m_\nu \neq m_\mu, \tag{18}$$

$$D_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\rho, \tag{19}$$

$$E_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\mu + m_\rho, \tag{20}$$

and

$$F_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\mu - m_\rho. \tag{21}$$

An extension of this to low-order models with higher nonlinearity and to systems other than the two-layer model is straightforward. That only tensors A–G are nonzero and that they must satisfy conditions (17)–(21) will be used for a substantial reduction of the number of free variables in the optimization problem, which is at the heart of the identification of the best empirical closure.

b. Low-order models conserving turbulent energy in the nonlinear terms

Two results of the work by Selten (1993, 1995a,b), Achatz et al. (1995), and Kwasniok (1996) seem to be important for the closure problem:

- Purely projected and truncated models (i.e., models without parameterization of the feedback from unresolved modes) seem to be always stable in the sense that no explosive unbounded growth of energy has been observed in such models.
- Allowing the tensor describing the linear dynamics of a reduced model to be different from the projected one in order to account for the aforementioned feedback does not destroy stability.

It therefore seems to be especially important that the nonlinear terms of a low-order model are constructed such that they do not violate the conservation of some type of energy. Actually, with a total energy metric, this is just the case for any projected and truncated EOF model of our two-layer system. Total energy, that is, the sum of kinetic and available potential energy, is defined by

$$E_{\text{tot}} = \sum_n \left[\frac{n(n+1)}{2}(\psi_n^0)^2 + \frac{n(n+1)+r^2}{2}(\tau_n^0)^2 \right]$$
$$+ \sum_{m>0,n} \{n(n+1)|\psi_n^m|^2$$
$$+ [n(n+1)+r^2]|\tau_n^m|^2\}. \tag{22}$$

Therefore the appropriate metric is $M_{ij} = \tilde{M}_i\delta_{ij}$ with

$$\tilde{M}_i = \begin{cases} \dfrac{n(n+1)}{2} & \text{if } \Phi_i = \psi_n^0 \\[2mm] \dfrac{n(n+1)+r^2}{2} & \text{if } \Phi_i = \tau_n^0 \\[2mm] n(n+1) & \text{if } \Phi_i = \psi_n^m, \ m > 0 \\[1mm] n(n+1)+r^2 & \text{if } \Phi_i = \tau_n^m, \ m > 0. \end{cases} \tag{23}$$

If the two-layer model tendency equations in state space are written as

$$\dot{\mathbf{\Phi}} = \mathbf{T}(\mathbf{\Phi}), \tag{24}$$

the equations of the projected (nonclosed) EOF model are given by

$$\dot{z}_\nu = \sum_i e_{i\nu}^z \tilde{M}_i T_i(\mathbf{\Phi}_{\text{tr}}) \tag{25}$$

and

$$\dot{w}_\nu = \sum_i \overline{e_{i\nu}^w} \tilde{M}_i T_i(\mathbf{\Phi}_{\text{tr}}), \tag{26}$$

where the truncated state vector is

$$\mathbf{\Phi}_{\text{tr}} = \hat{\mathbf{\Phi}} + \sum_\nu z_\nu \mathbf{e}_\nu^z + \sum_\nu w_\nu \mathbf{e}_\nu^w. \tag{27}$$

Equations (25)–(27) yield, together with (1) and (2),

$$\dot{z}_\nu = \sum_\mu A_{\nu\mu}^{\text{pr}}(z_\mu - Z_\mu) + \sum_{\mu,\rho} B_{\nu\mu\rho}^{\text{pr}} w_\mu \overline{w}_\rho \tag{28}$$

and

$$\dot{w}_\nu = \sum_\mu C_{\nu\mu}^{\text{pr}} w_\mu + \sum_{\mu,\rho} D_{\nu\mu\rho}^{\text{pr}} z_\mu w_\rho$$
$$+ \sum_{\mu,\rho} E_{\nu\mu\rho}^{\text{pr}} w_\mu w_\rho + \sum_{\mu,\rho} F_{\nu\mu\rho}^{\text{pr}} w_\mu \overline{w}_\rho. \tag{29}$$

Here Z denotes the projected zonally symmetric equilibrium configuration of the two-layer model. The interaction tensors A^{pr}, B^{pr}, C^{pr}, D^{pr}, E^{pr}, and F^{pr} together with some details of the derivation can be found in the appendix.

A quantity possibly conserved in the adiabatic limit of this model is turbulent energy, that is,

$$E_{\text{turb}} = \sum_{i,j} \overline{\delta\Phi}_i M_{ij} \delta\Phi_j$$
$$= \sum_\nu z_\nu^2 + \sum_\nu |w_\nu|^2. \tag{30}$$

The conditions for conservation, that is, $\dot{E}_{\text{turb}} = 0$, can be derived from (28) and (29) to be

$$A_{\mu\nu}^{\text{pr}} = -A_{\nu\mu}^{\text{pr}}, \tag{31}$$

$$\sum_\mu A_{\nu\mu}^{\text{pr}} Z_\mu = 0, \tag{32}$$

$$C_{\mu\nu}^{\text{pr}} = -\overline{C_{\nu\mu}^{\text{pr}}}, \tag{33}$$

$$2B_{\nu\mu\rho}^{\text{pr}} + D_{\rho\mu\nu}^{\text{pr}} + D_{\mu\rho\nu}^{\text{pr}} = 0, \tag{34}$$

and

$$E_{\nu\mu\rho}^{\text{pr}} + E_{\nu\rho\mu}^{\text{pr}} + \overline{F_{\mu\nu\rho}^{\text{pr}}} + \overline{F_{\rho\nu\mu}^{\text{pr}}} = 0. \tag{35}$$

In checking the validity of these conditions let us first look at the adiabatic case with zero time-averaged state. There (A17) shows that A^{pr} vanishes so that (31) and (32) are fulfilled. From (A19) we see that C^{pr} assumes the anti-Hermetian character as required by (33). Using the antisymmetry of the Jacobian operator and that for any three functions, f, g, and h, $fJ(g, h) + gJ(f, h) =$

$J(fg, h)$ and $fJ(g, h) + hJ(g, f) = J(g, fh)$, one can convince oneself from (A18) and (A20)–(A22) that (34) and (35) also hold. In the diabatic case and/or if the general time mean state does not vanish, the linear tensors no longer satisfy the required conditions. Nonetheless (34) and (35) hold even then so that the nonlinear terms do not affect the energy balance irrespective of whether we are in an adiabatic or diabatic situation or whether the time mean vanishes or not. This is the main reason why, as will be shown later, the projected model is stable in the sense that its turbulent energy always keeps within certain limits.

It can be expected that a closed model that follows the energetic behavior of the projected (nonclosed) version faithfully will also exhibit this stability. Let the closed model be given by the most general quadratic model admitted by zonally symmetric boundary conditions and forcing, that is, by the tendency equations (15) and (16) with only tensors A–G being nonzero and fulfilling the coupling conditions (17)–(21). If we now demand that the energy balance is the same as in the projected model, that is, that $\dot{E}_{\text{turb}} = \dot{E}_{\text{turb}}|_{\text{pr}}$ where $\dot{E}_{\text{turb}}|_{\text{pr}}$ is the turbulent energy's time derivative as given in the projected model, we find as conditions for the closure corrections, denoted by $T_{...}^{\text{cl}} = T_{...} - T_{...}^{\text{pr}}$ for any of the tensors A–F,

$$A_{\mu\nu}^{\text{cl}} = -A_{\nu\mu}^{\text{cl}}, \qquad (36)$$

$$\sum_{\mu} A_{\nu\mu}^{\text{cl}} Z_{\mu} = 0, \qquad (37)$$

$$C_{\mu\nu}^{\text{cl}} = -\overline{C_{\nu\mu}^{\text{cl}}}, \qquad (38)$$

$$\sum_{\mu,\rho} G_{\nu\mu\rho} Z_{\mu} Z_{\rho} = 0, \qquad (39)$$

$$\sum_{\rho} (G_{\nu\mu\rho} + G_{\mu\nu\rho} + G_{\nu\rho\mu} + G_{\mu\rho\nu}) Z_{\rho} = 0, \qquad (40)$$

$$\sum_{\text{perm } \nu,\mu,\rho} G_{\nu\mu\rho} = 0, \qquad (41)$$

$$2B_{\nu\mu\rho}^{\text{cl}} + D_{\rho\mu\nu}^{\text{cl}} + D_{\mu\rho\nu}^{\text{cl}} = 0, \qquad (42)$$

and

$$E_{\nu\mu\rho}^{\text{cl}} + E_{\nu\rho\mu}^{\text{cl}} + \overline{F_{\mu\nu\rho}^{\text{cl}}} + \overline{F_{\rho\nu\mu}^{\text{cl}}} = 0. \qquad (43)$$

Among the above conditions, (39)–(43) are most important since they prevent all nonlinear terms—consisting of closure contributions and those obtained from the projection—from leading to catastrophic growth in turbulent energy. All conditions together ensure that explosive energy growth cannot occur if this is not observed in the nonclosed, purely projected, and truncated model either.

c. Simulation of multiple baroclinic wave life cycles by closed EOF models

Following our statistical–empirical ansatz we have examined how well-reduced EOF models perform using an optimal closure parameterization extracted directly from the data produced by our two-layer model (and as described in section 2b). This has been done in reductions consisting of finding the minimum of a squared error sum between the tendencies of the full two-layer model, projected onto the explicitly considered EOFs, and the closed model's tendencies, that is,

$$\varepsilon_t(A^{\text{cl}}, G, C^{\text{cl}}, D^{\text{cl}}, F^{\text{cl}})$$

$$= \frac{1}{\sigma_t} \sum_{t} \left[\sum_{\nu} (\dot{z}_{\nu} - \dot{z}_{\nu}^{\text{pr}})^2 + \sum_{\nu} |\dot{w}_{\nu} - \dot{w}_{\nu}^{\text{pr}}|^2 \right], \quad (44)$$

where

$$\sigma_t = \sum_{t} \left[\sum_{\nu} \dot{z}_{\nu}^{\text{pr}2} + \sum_{\nu} |\dot{w}_{\nu}^{\text{pr}}|^2 \right], \qquad (45)$$

$$\dot{z}_{\nu}^{\text{pr}} = \sum_{i} e_{i\nu}^{z} \tilde{M}_i T_i(\mathbf{\Phi}_{\text{tr}}), \qquad (46)$$

and

$$\dot{w}_{\nu}^{\text{pr}} = \sum_{i} \overline{e_{i\nu}^{w}} \tilde{M}_i T_i(\mathbf{\Phi}_{\text{tr}}). \qquad (47)$$

In order to avoid the discussed stability problems with respect to unbounded energy growth constraints (36)–(43) have been applied in two different minimizations by varying degrees of completeness. In both reduction approaches the nonlinear terms have been forced to satisfy (39)–(43). For reasons of simplicity we have set

$$G_{\nu\mu\rho} = 0, \qquad (48)$$

which keeps the structure of advective terms as it is in the full-scale model.

The difference in the two approaches (described below), which lies in the stringency of the application of constraints to linear closure terms, has been motivated by the following consideration. It has been shown by Kwasniok (1996) and Selten (1995a,b) that model reductions by PIPs or EOFs might be improved in quality by appropriately adjusting the linear tensor of their models. None of the two authors has applied any constraints like (36)–(38). Nevertheless, energy did not grow without bounds in any of the cases considered by them. Therefore it might indeed be a good way just to connect the nonlinear terms to constraints and let the linear terms be free variables of the optimization problem. On the other hand, at least for a comparison, it might be good to look in addition at optimized models in which also the linear closure terms do not change the turbulent energy balance. This is also attractive because in the latter approach a closer link of the reduced to the known true dynamics is ensured by keeping the influence of all diabatic effects as it can be determined by pure projection. We have thus performed at given numbers of leading EOFs optimizations of two types. In the first optimization, henceforth called LC, the linear terms are forbidden to violate the turbulent energy balance. This has been done by setting

$$A_{\nu\mu}^{\text{cl}} = 0 \qquad (49)$$

and calculating all $C_{\nu\mu}^{\mathrm{cl}}$ with $\mu < \nu$ via (38) from $C_{\mu\nu}^{\mathrm{cl}}$. The more general conditions (36) and (37) have been found to lead to only negligible differences in the resulting closure. Because they potentially destroy the symmetry of the projected linear tensor for the zonally symmetric EOFs, they have therefore not been used in their general form any further. The free variables of the optimization are then all $C_{\nu\mu}^{\mathrm{cl}}$ with $\mu \geq \nu$, $D_{\nu\mu\rho}^{\mathrm{cl}}$ and $F_{\nu\mu\rho}^{\mathrm{cl}}$. In the second optimization, called LU, no constraints have been applied to linear closure terms. Optimization variables are there all $A_{\nu\mu}^{\mathrm{cl}}$, $C_{\nu\mu}^{\mathrm{cl}}$, $D_{\nu\mu\rho}^{\mathrm{cl}}$, and $F_{\nu\mu\rho}^{\mathrm{cl}}$.

Since all constraints are linear, in the practical optimization one has to solve a linear regression problem that can be done by standard methods. Because it was at hand and had to be used in the PIP context (discussed below) anyway, we have used a nonlinear sequential quadratic programming algorithm as available in the NAG library (Ford and Pool 1984). The minimization has always been initialized with vanishing closure corrections. Coupling conditions (17)–(21) hold for the projected tensors and thus must do so also for the closure corrections. This helps to reduce the number of free variables substantially. In order to give the reader some impression of the magnitude of the problem, we offer the following. In the case of an LC optimization of an EOF model based on 15 EOFs, that is, seven zonally symmetric patterns and eight in wave space (as usual, the variance criterion has been chosen to decide which patterns are to be picked as the leading ones), the number of free real variables in the minimization (imaginary and real parts are counted separately) is 744, instead of 1984 without the coupling conditions.

Only a subset of the whole time series has been used for the actual reduction. This is the time span between days 2001 and 4000. For every day one data point in state space has been considered. A test for the statistical sufficiency of the analyzed dataset has been performed by recalculating ε_t after the minimization, using the previously determined EOFs and model coefficients (from days 2001 to 4000), but for days 4001 to 6000. Virtually no difference could be seen between the two thus calculated ε_t. This means that, at least with respect to local tendencies, a reduced model determined from the first subset of data is as well able to simulate the second one. We therefore consider the analyzed subset statistically sufficient enough so that relevant sampling errors are avoided.

The relative variance spectrum for the first 20 EOFs determined from this subset is shown in Fig. 2. The decay in variance with growing EOF index is rather steep. The dominating EOF, a pattern with zonal wavenumber $m = 6$ is very similar to the normal mode that had been used for the initialization of the time series. In spite of their small variance contribution, however, it will be seen in the following that quite a few of the following patterns are indeed needed for a realistic simulation of the time series with all its major character-
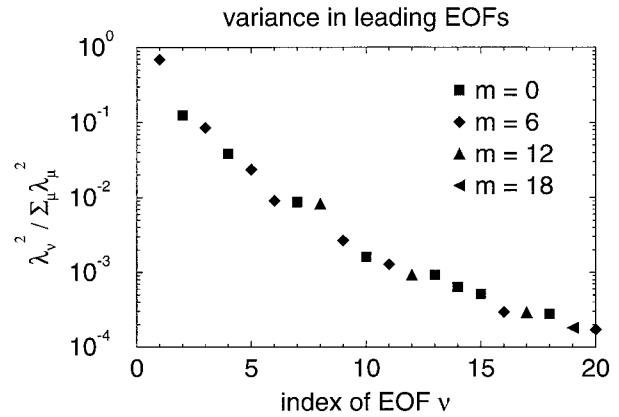

FIG. 2. Relative variance associated with the 20 leading EOFs of days 2001–4000 in the analyzed time series, together with the zonal wavenumber associated with each pattern.

istics. The spatial structure of the three leading zonally symmetric and wave EOFs, respectively, is shown in Figs. 3–5.

For the predictive behavior of a reduced model on shorter timescales (in our case $\Delta t \leq 60$ d), the relative tendency error ε_t itself is a measure since it quantifies the local similarity between projected tendencies and tendencies in the reduced model. Figure 6 contains this number for all reductions on the basis of up to 13 leading EOFs. The relative tendency variance error of the corresponding purely projected and truncated (nonclosed) model is also given. Especially at larger numbers of considered EOFs the improvement in local accuracy by the closure in comparison to the nonclosed model is quite substantial. About an order of magnitude can be gained. Generally, models that do not conserve turbulent energy in the linear closure terms have smaller local errors. This is natural since the number of minimization variables is larger than in the energy conserving case. More light can be shed on the short time predictive qualities of the resulting model by performing standard prediction tests. As an example we show the case of 11 EOFs, that is, four zonally symmetric patterns and seven in wave space. Two thousand integrations have been performed over 100 days, starting from day 4001 up to 6000. From this set of predictions a mean anomaly correlation,

$$c_a(\Delta t) = \left\langle \frac{\mathrm{Re} \sum_{\nu} \overline{a_\nu}\big|_{\mathrm{pred}}(\Delta t) a_\nu\big|_{\mathrm{pr}}(\Delta t)}{\sqrt{\sum_{\nu} |a_\nu|_{\mathrm{pred}}(\Delta t)|^2} \sqrt{\sum_{\nu} |a_\nu|_{\mathrm{pr}}(\Delta t)|^2}} \right\rangle, \quad (50)$$

and a root-mean-squared error of the prediction,

$$\varepsilon_{\mathrm{pred}} = \sqrt{\frac{\left\langle \sum_{\nu} |a_\nu|_{\mathrm{pred}}(\Delta t) - a_\nu|_{\mathrm{pr}}(\Delta t)|^2 \right\rangle}{\left\langle \sum_{\nu} |a_\nu|_{\mathrm{pr}}(\Delta t)|^2 \right\rangle}}, \quad (51)$$
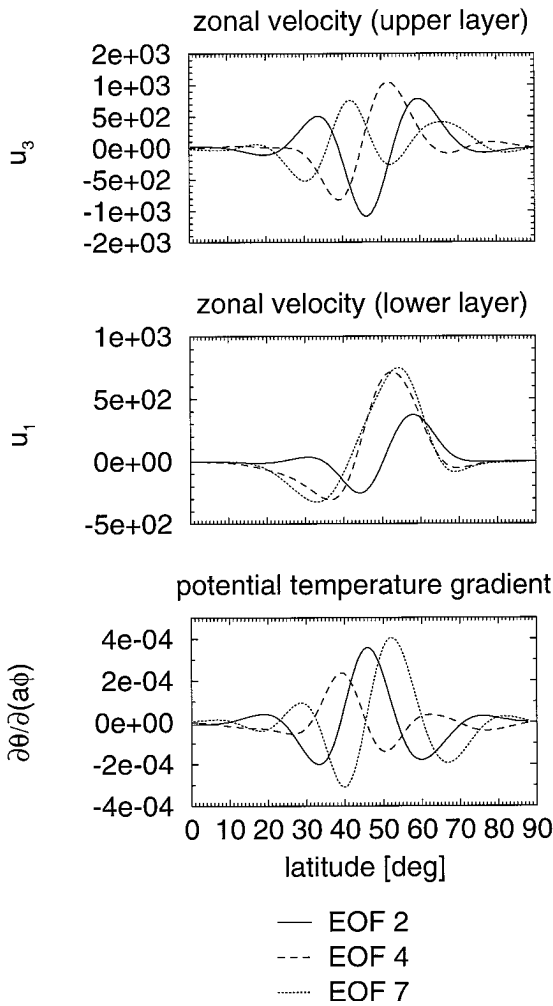
FIG. 3. Spatial structure of the three leading zonally symmetric EOFs of days 2001–4000 in the analyzed time series. Units are arbitrary.
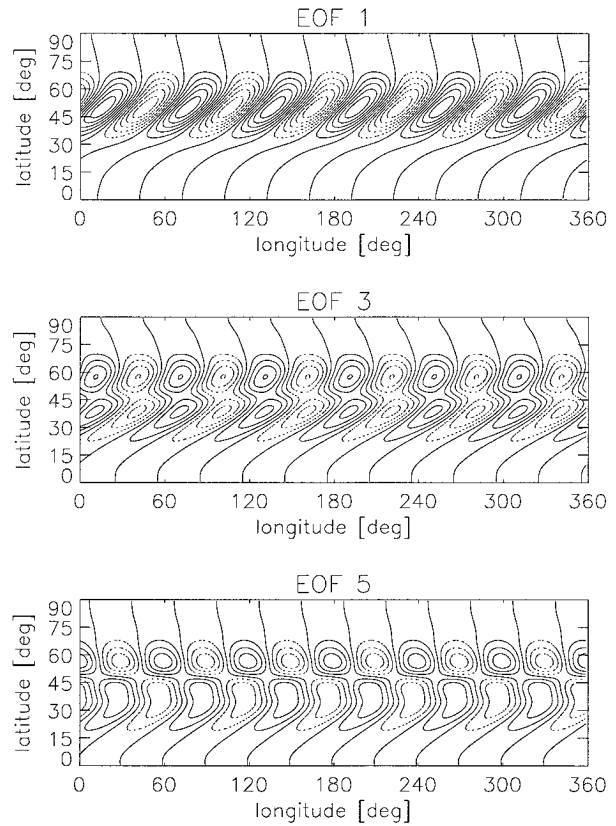

FIG. 4. Lower-layer streamfunction of the three leading wave EOFs of days 2001–4000 in the analyzed time series. Dashed lines indicate negative values, solid lines positive ones. Contour intervals are constant but in arbitrary units.

have been calculated. Here $a_\nu|_{\text{pred}}(\Delta t)$ is the expansion coefficient of the νth EOF (both for zonally symmetric or wave patterns) as predicted by the reduced model for time Δt after beginning the integration, and $a_\nu|_{\text{pr}}(\Delta t)$ is the corresponding coefficient as determined by projection on the respective data point of the full dataset. Angle brackets denote averaging over all predictions. Figure 7 shows the results. Persistence corresponds to $a_\nu|_{\text{pred}}(\Delta t) = a_\nu|_{\text{pr}}(0)$. As expected from the relative local tendency error, all closed models perform much better than the projected (nonclosed) model. The model from the LU optimization has best predictive capability. As also diagnosed by Selten (1995b), due to the best fit of local tendencies, closed models are indeed able to give rather good short-term predictions. An interesting aspect is that the full two-layer model, integrated from truncated initial conditions obtained by projection onto the leading EOFs, performs worse for shorter prediction periods. This results from missing the effect of unre-

solved modes in the beginning of the integration by using truncated initial conditions.

For an estimate of the capability of the EOF models to simulate the longtime behavior of the full system, we have projected for each model both day 2001 and 4001 of the data onto the respective EOFs and integrated from there over 2000 days. Various time and zonal means of the simulation have then been compared to the corresponding quantities of the full dataset. Additionally, the time dependence of all four energy types has been looked at in order to check whether the reduced model is able to simulate the prominent modulation of energy in the original data. This additional check of the actual time development of the system, not possible so easily for more irregular datasets like, for example, that produced by the much-studied barotropic model with zonally dependent forcing and orography, must be seen as an especially stringent test. As will be seen below, correct simulation of the modulation necessitates significantly more patterns than just faithful reproduction of first and second moments of the full system.

As one major result it has been observed that all integrations are stable in the sense that the energy never grew without bounds. This is not due to some special
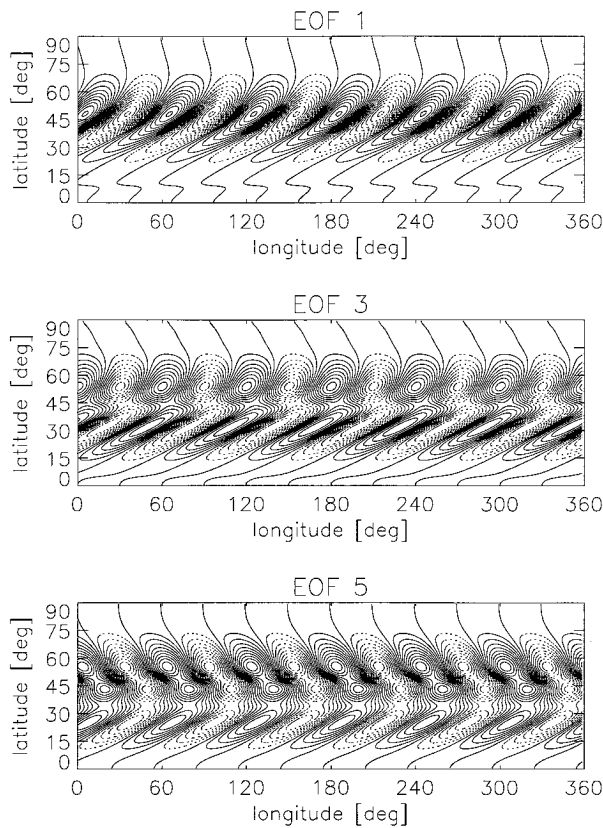
EOF 1

EOF 3

EOF 5

FIG. 5. Same as Fig. 4 but for the upper layer. Contour intervals are as in Fig. 4.

mean anomaly correlation
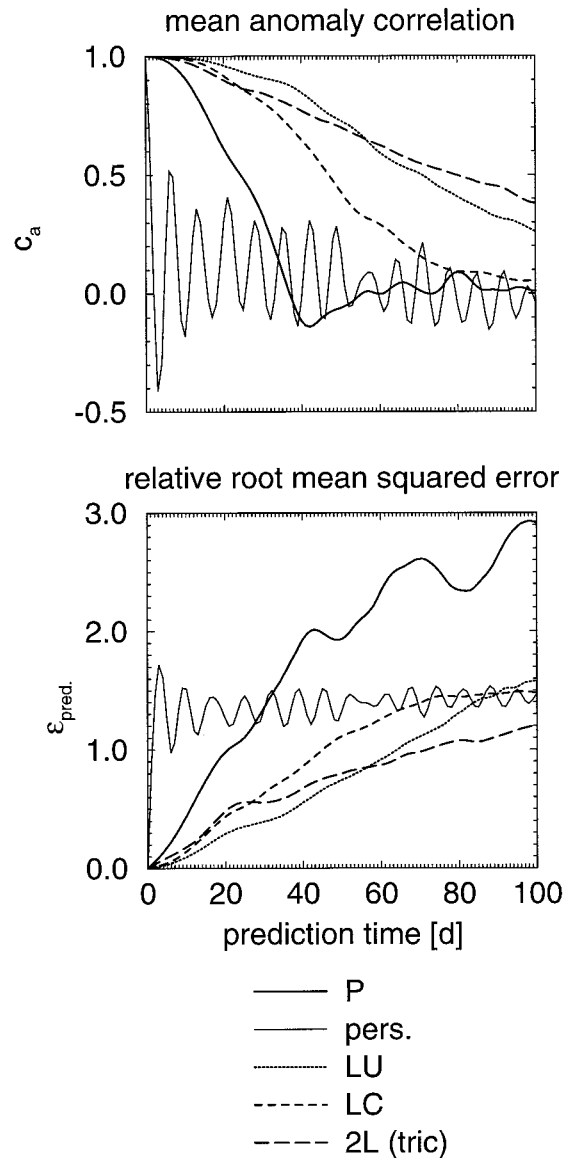
relative root mean squared error

FIG. 7. Mean anomaly correlation and relative root-mean-squared error of the various EOF models (LU denotes a model in which linear closure terms do not observe energy conservation constraints, whereas in the LC model they do) including the purely projected (i.e., nonclosed) version (P) and a persistence assumption (pers.) for 11 EOFs. Here 2L (tric) denotes the corresponding result obtained by integrating the full two-layer model from truncated initial conditions obtained by projection onto the 11 leading EOFs.

properties of our data, as could be argued from the difference of our dataset from the one analyzed by Selten (1995b), but to turbulent energy conservation in the nonlinear closure terms. Without energy conserving constraints on the interaction tensors we find closed models with the same runaway effects as discussed earlier.

relative tendency variance error

FIG. 6. Relative tendency variance error of the two considered closed models for up to 13 leading EOFs. Here LU denotes a model in which linear closure terms do not observe energy conservation constraints, whereas in the LC model they do. The errors of the corresponding purely projected (i.e., nonclosed) model (P) are also given.

Additionally, however, we find that the complete freedom in fitting the linear tensors as is allowed in models with projected nonlinear interaction tensors no longer exists. Model LU exhibits enormous problems in the way that, for most integrations, wave amplitudes especially are generally much too large. Therefore, it seems that turbulent energy conservation by linear closure terms not only ensures a more realistic link to the known diabatic dynamics of the full system but is also

essential for ensuring well-behaved simulations with wave amplitudes within realistic limits.

In an evaluation of the specific qualities of the LC closure with respect to longtime simulations we first observe that in nearly all of simulations performed a substantial improvement was found by using the closed instead of the projected (nonclosed) model. There is also a clear tendency for the simulations to be improved by increasing the number of patterns. Generally it must also be said that this cannot be set equal to a statement that increasing this number by one also leads to improvement. There are opposite cases. As an example, 16 EOFs perform worse than 15 and 17 (especially concerning faithful reproduction of the longtime modulation). Such occurrences are not as surprising as one might think immediately. The relative tendency variance error does decrease as the number of patterns is increased from 15 to 16 and then 17, and as a consequence the short-term predictive capability also increases. Nonetheless, although good longtime simulations might be hoped to be achieved by models with good short-time predictive capabilities (and mostly this is the case), this is by no means always to be expected. Rather, as a consequence, the strategy to be taken here is to find a lowest number of patterns simulating one or several aspects of the original data well without drawing from this the conclusion that, with the patterns performing so well there, these will also guarantee good performance of every LC model containing them. Keeping this in mind, we can summarize our findings about the quality of LC simulations in the following way. Among the cases considered (the number of EOFs was always larger than 10), the lowest number of patterns to yield good time and zonal mean zonal wind and meridional potential temperature gradient distributions in both simulations is 11 (4 zonally symmetric patterns, 7 in wave space). The same holds for the zonal-mean Eliassen–Palm flux (EPF) divergences. In order to find some clear longtime energy modulation in the simulations, we need at least 13 patterns (5 zonally symmetric, 8 in wave space). For a quantitatively good simulation of the period of the modulation, 17 patterns are needed (7 zonally symmetric, 10 in wave space). For this case Fig. 8 shows the resulting zonal-mean zonal wind in both layers and zonal-mean meridional potential temperature gradient, averaged over day 4001–6000. In Fig. 9, for the same integration, time and zonal mean of the EPF divergence in both layers are given. Comparison with the same quantities as determined from the original data reveals good agreement. As for the longtime modulation, the reader can conclude from the time dependence of the four energies in the simulations by the projected model (Fig. 10) and by the LC model (Fig. 11) that the characteristic oscillation in the latter case matches the one in Fig. 1 rather well—especially much better than in the former case.
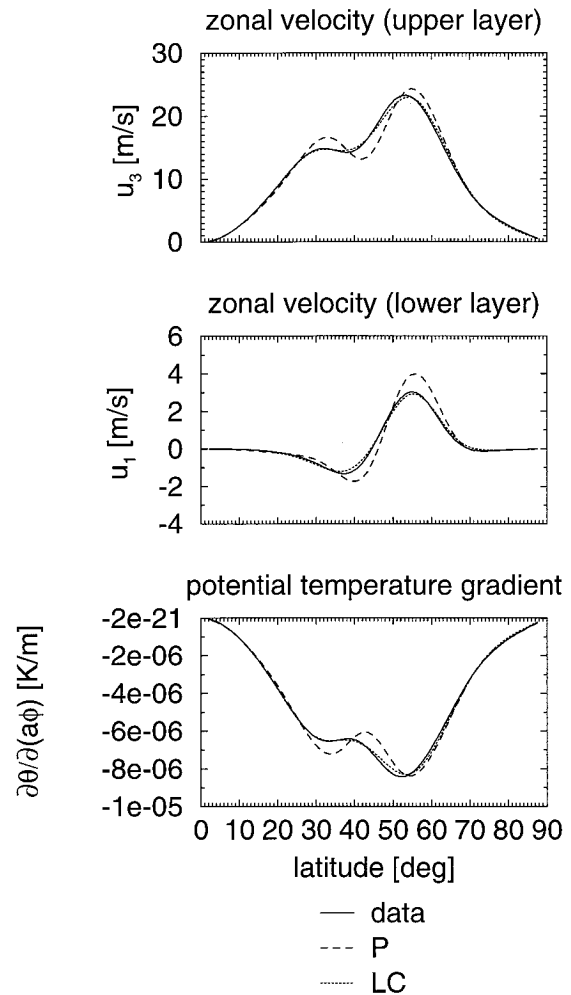

Fig. 8. Time and zonal means of the zonal wind in both layers and of the meridional potential temperature gradient, as obtained by an LC model (in which all closure terms observe energy conservation constraints) on the basis of 17 EOFs, integrated from day 4001 to day 6000, in comparison with the corresponding values in the two-layer model data and with the results of the projected (i.e., nonclosed) model (P).

4. Reduction by PIPs

In spite of their strengths closed EOF models are not necessarily the ultimate tool for reducing complex dynamical systems. As has been shown by Kwasniok (1996), models based on principal interaction patterns yield further improvement in comparison to empirical orthogonal functions in the reduction of a barotropic model by projected (nonclosed) and linearly closed models. This refers both to short-time predictions and to long-time climate simulations. The question therefore arises immediately whether in the case of nonlinearly closed models discussed here PIPs do not exhibit the same advantages. In an attempt to answer to this, PIP analyses of our multiple baroclinic wave life cycles have also been performed. We will approach the topic in this section in the following way. First, we will give a short
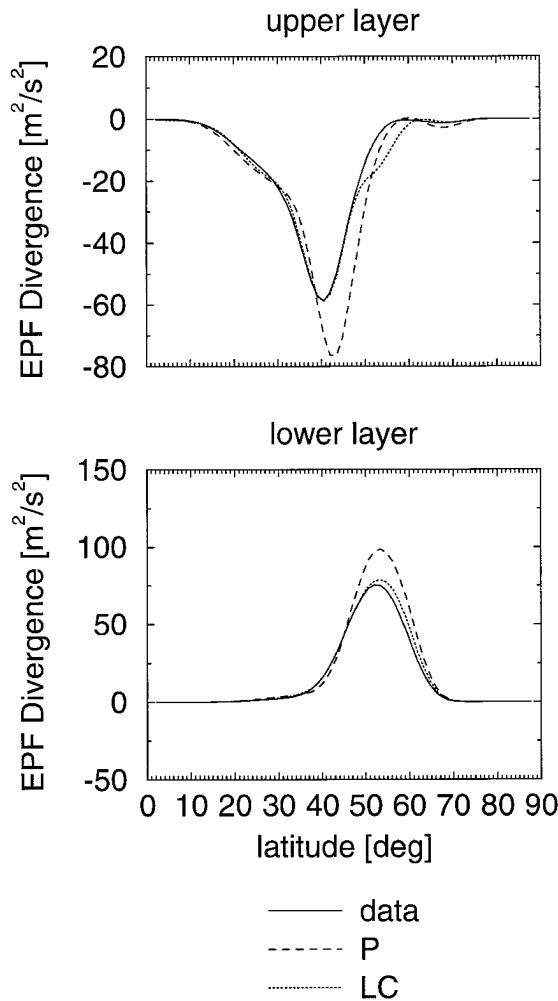
FIG. 9. Same as Fig. 6 but for the Eliassen–Palm flux (EPF) divergence in both layers.
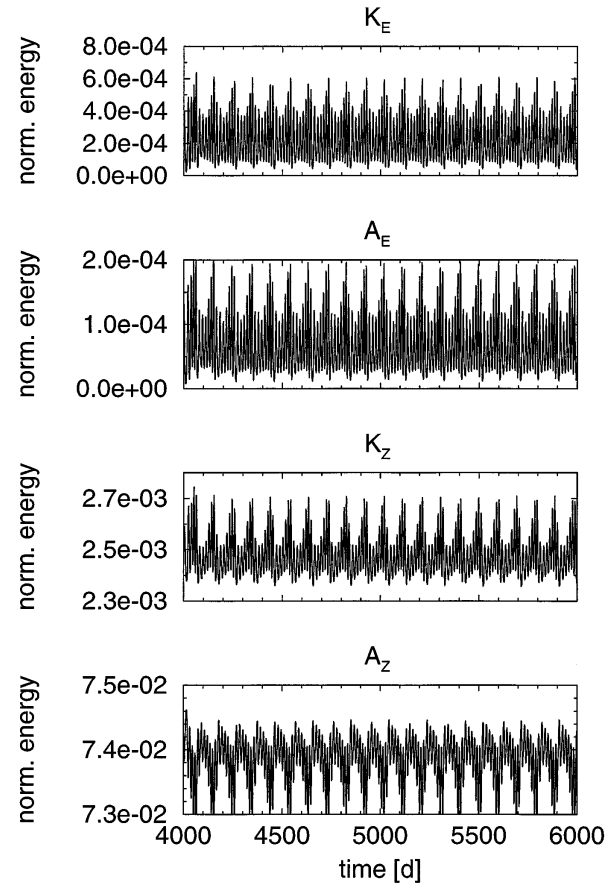

FIG. 10. Time dependence of all Lorenzian energies in a simulation by a projected (i.e., nonclosed) model based on 17 EOFs. Normalization constant is 9.042×10^{22} J.

overview about PIPs in general. Then we will specify the PIP model used. Finally, we will summarize the PIP optimizations we have done and discuss the results.

a. Principal interaction patterns in general

The term, principal interaction patterns, is somewhat misleading since what is really looked for in a PIP analysis is an optimal subspace of the full state space for representing the dynamics of the system on the basis of as few degrees of freedom as possible. Within this subspace gauge invariance of the model with respect to changes of the basis patterns exists (Kwasniok 1996); that is, as long as two sets of basis patterns span the same space the corresponding dynamical equations—obtainable from each other by application of standard transformation procedures—are equivalent in the sense that integrating them with the same initial condition results in completely identical system trajectories. PIP analysis is a generalization of the optimization proce-

dure described in the previous sections for EOFs in that now not only the interaction coefficients but also the reduced subspace used for the projection is an object of the optimization. There are several possibilities for defining a suitable objective function to be minimized. In accordance with the approach taken by Kwasniok (1996), we choose for it

$$\varepsilon_t^{\mathrm{PIP}} = \frac{1}{\sigma_t^{\mathrm{PIP}}} \sum_t \sum_i \left(\overline{\delta \dot{\Phi}_i^{\mathrm{PIP}}} - \overline{\delta \dot{\Phi}_i^{\mathrm{pr}}} \right) \tilde{M}_i (\delta \dot{\Phi}_i^{\mathrm{PIP}} - \delta \dot{\Phi}_i^{\mathrm{pr}}),$$

(52)

with

$$\sigma_t^{\mathrm{PIP}} = \sum_t \sum_i \overline{\delta \dot{\Phi}_i^{\mathrm{pr}}} \tilde{M}_i \delta \dot{\Phi}_i^{\mathrm{pr}}.$$

(53)

Here $\delta \dot{\Phi}^{\mathrm{pr}}$ is the tendency of the deviation of the state vector from the time mean, projected onto the PIP subspace and $\delta \dot{\Phi}^{\mathrm{PIP}}$ is the PIP model result for the latter. This approach ensures best agreement between the full and reduced model in the very subspace spanned by the PIPs. More details will be given below. In general, the advantage of the PIP method could be that it is not so
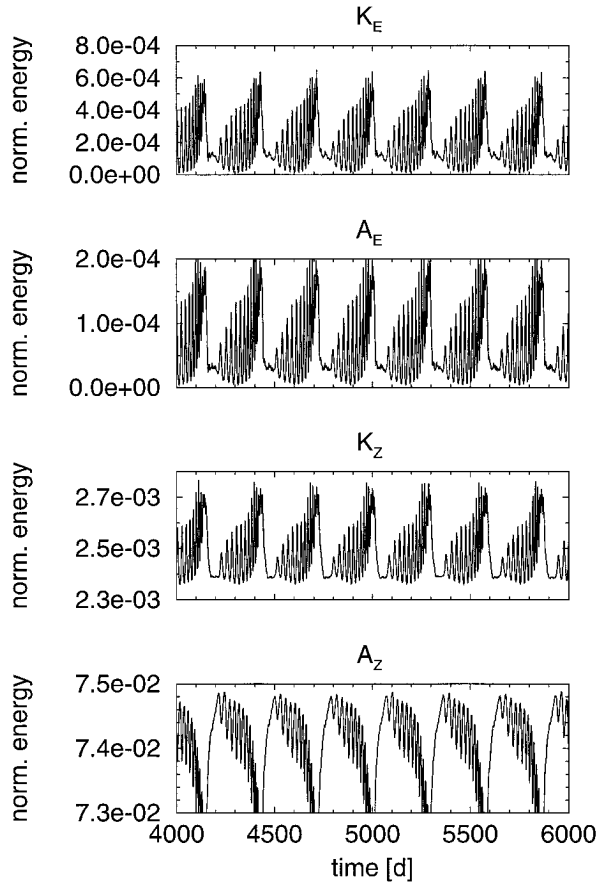
FIG. 11. Same as Fig. 8 but for the LC model (a nonlinearly closed model in which all closure terms observe energy conservation constraints).

much variance oriented as an EOF projection but more intended to reproduce the actual *dynamics* faithfully.

b. A PIP model

What is now the PIP model we can use in our context? First of all, zonal symmetry of forcing and boundary conditions can help us once more in excluding a large set of possibilities. It has been pointed out in the previous section that EOFs reproduce this symmetry in that just by using their own degrees of freedom they can be shifted arbitrarily in longitude. This is only possible because each EOF has only one zonal wavenumber so that multiplication of its expansion coefficient with a phase factor produces this shift. There is no reason to assume PIPs to be different from EOFs with this respect. We therefore demand that there is a set of basis vectors of PIP subspace with the property that each has only one zonal wavenumber. In fact, the EOF basis of PIP subspace can be shown to do this in the same manner as described above. In complete analogy to the EOF case, we can therefore write for the PIP expansion of the deviation of the state vector from its time mean

$$\delta \mathbf{\Phi}^{\mathrm{pr}} = \sum_{\nu} \tilde{z}_{\nu} \mathbf{p}^{\mathbf{z}}_{\nu} + \sum_{\nu} \tilde{w}_{\nu} \mathbf{p}^{\mathbf{w}}_{\nu}. \qquad (54)$$

Here $\mathbf{p}^{\mathbf{z}}_{\nu}$ is a zonally symmetric basis vector (zonally symmetric PIP), $\mathbf{p}^{\mathbf{w}}_{\nu}$ a basis vector with a zonal wavenumber $m_{\nu} > 0$ (wave PIP); $\tilde{z}_{\nu}$ and $\tilde{w}_{\nu}$ are the corresponding expansion coefficients. If we assume for the PIP model tendencies of the latter a quadratic form, symmetry considerations lead us to Eqs. (15) and (16) with only tensors A–G being nonzero and fulfilling the coupling conditions (17)–(21). It is understood that the EOF expansion coefficients have to be replaced by their PIP counterparts.

Also the closure of our PIP model can be found in close analogy to the EOF case. Based on the gauge invariance demonstrated by Kwasniok (1996), we assume from now on that, without loss of generality, all PIPs are orthonormal, that is,

$$\sum_{\nu} p^{z}_{i\nu} \tilde{M}_i p^{z}_{i\mu} = \delta_{\nu\mu}, \qquad (55)$$

$$\sum_{\nu} \overline{p^{w}_{i\nu}} \tilde{M}_i p^{w}_{i\mu} = \delta_{\nu\mu}, \qquad (56)$$

and

$$\sum_{\nu} p^{z}_{i\nu} \tilde{M}_i p^{w}_{i\mu} = \sum_{\nu} \overline{p^{w}_{i\nu}} \tilde{M}_i p^{z}_{i\mu} = 0. \qquad (57)$$

Projecting the two-layer model onto the patterns, we once more arrive at (28) and (29). As before, the nonlinear terms of this projected model conserve turbulent energy. Turbulent energy conservation in the nonlinearly closed case is generally ensured by the constraints (36)–(43). Drawing from the experience gathered with EOF models we will in the following use as the PIP model the analogue of the LC model; that is, we will use (38), (42), (43), (49) (i.e., $A^{\mathrm{cl}}_{\nu\mu} = 0$), and (48) ($G_{\nu\mu\rho} = 0$) as constraints on the closure.

c. The reductions and their results

In the practical optimization we assumed that the PIP subspace is contained within a suitably large subspace of leading EOFs. This reduces computation time considerably without loss of precision. The PIP expansions in terms of the EOFs considered are

$$\mathbf{p}^{\mathbf{z}}_{\nu} = \sum_{\mu} p^{z}_{\mu\nu} \mathbf{e}^{\mathbf{z}}_{\mu} \qquad (58)$$

and

$$\mathbf{p}^{\mathbf{w}}_{\nu} = \sum_{\mu} p^{w}_{\mu\nu} \mathbf{e}^{\mathbf{w}}_{\mu}, \qquad (59)$$

where for each EOF the associated expansion coefficient is only nonzero if its zonal wavenumber agrees with that of the PIP. The objective function of the minimization is then

$$\varepsilon_t^{\mathrm{PIP}}[p_{\mu\nu}^z, p_{\mu\nu}^w, C_{\nu\mu}^{\mathrm{cl}}(\mu \geq \nu), D^{\mathrm{cl}} \, F^{\mathrm{cl}}]$$

$$= \frac{1}{\sigma_t^{\mathrm{PIP}}} \sum_t \left[\sum_\nu (\dot{\tilde{z}}_\nu - \dot{\tilde{z}}_\nu^{\mathrm{pr}})^2 + \sum_\nu |\dot{\tilde{w}}_\nu - \dot{\tilde{w}}_\nu^{\mathrm{pr}}|^2 \right]. \quad (60)$$

with

$$\sigma_t^{\mathrm{PIP}} = \sum_t \left[\sum_\nu (\dot{\tilde{z}}_\nu^{\mathrm{pr}})^2 + \sum_\nu |\dot{\tilde{w}}_\nu^{\mathrm{pr}}|^2 \right], \quad (61)$$

$$\dot{\tilde{z}}_\nu^{\mathrm{pr}} = \sum_\mu p_{\mu\nu}^z \dot{z}_\mu^{\mathrm{pr}}, \quad (62)$$

and

$$\dot{\tilde{w}}_\nu^{\mathrm{pr}} = \sum_\mu \overline{p_{\mu\nu}^w} \dot{w}_\mu^{\mathrm{pr}}. \quad (63)$$

Equations (55)–(57) are the associated constraints. For comparison with previous work, we have also determined a linearly closed PIP model without conservation constraints. There optimization variables are all $p_{\mu\nu}^z$, $p_{\mu\nu}^w$, $A_{\nu\mu}^{\mathrm{cl}}$, and $C_{\nu\mu}^{\mathrm{cl}}$, while all nonlinear closure coefficients are zero.

The optimization problem is a truly nonlinear one, with a rather large number of free variables (e.g., counting real and imaginary parts separately, for 15 patterns, in addition to 744 real closure coefficients, 437 PIP expansion coefficients have to be determined if a basis of 60 EOFs is chosen for the subspace containing all PIPs). For its solution the same sequential quadratic programming algorithm has been used as in the EOF case. For each number of PIPs the minimization has been initialized with patterns being identical to the same number of leading EOFs and the closure corrections of the corresponding LC model. Since discontinuous jumps are not allowed in the optimization, the number of PIPs at each zonal wavenumber is thereby fixed from the very beginning. It cannot be excluded that playing with this number could further improve our results. For the present we have abstained from such an analysis.

The best choice of the number of underlying EOFs has been determined by performing a reduction of the system to a nonlinearly closed model for 10 PIPs with 20, 40, and 80 EOFs. The analyzed dataset is the same as used in the determination of the closed EOF models. It comprises days 2001 to 4000, data taken once every day. The resulting relative tendency variance errors are 5.336×10^{-5}, 1.963×10^{-5}, and 1.882×10^{-5}, respectively. From this we conclude that four times the number of PIPs seems to be a good dimension of the considered EOF subspace. This has been adopted for all analyses described here.

As in the EOF case the statistical sufficiency of the analyzed dataset has been tested by recalculating the relative tendency variance error as defined in (52), after the minimization, using the patterns and model coefficients obtained for days 4001 to 6000. The result is shown in Fig. 12. No relevant difference between the two datasets can be observed so that for the PIP analysis
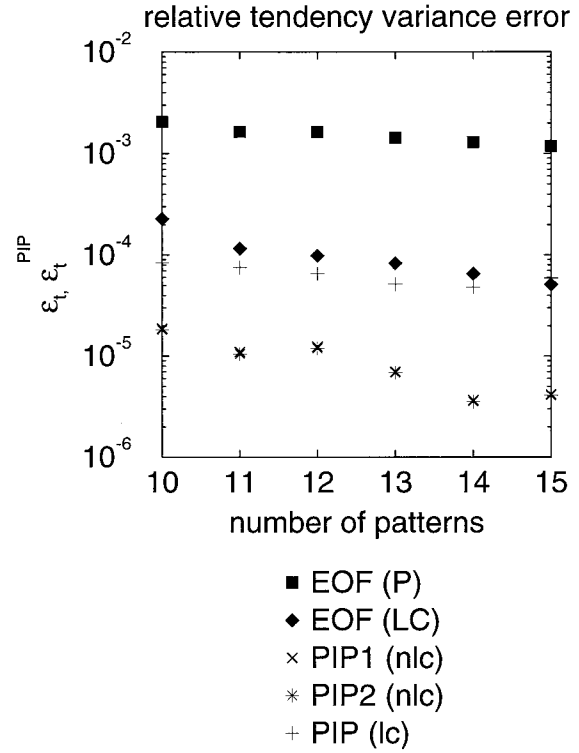

relative tendency variance error

- ■ EOF (P)
- ◆ EOF (LC)
- × PIP1 (nlc)
- ✳ PIP2 (nlc)
- + PIP (lc)

FIG. 12. In dependence of the number of determined PIPs, the relative tendency variance error, $\varepsilon_t^{\mathrm{PIP}}$ for the nonlinearly closed PIP model (nlc) as determined from days 2001–4000, recalculated for the latter time span (1), and days 4001–6000 (2). Also given for the same number of EOFs is the corresponding error ε_t of the purely projected (i.e., nonclosed) (P) and the best closed (LC, i.e., all closure terms observe energy conservation constraints) EOF model and a linearly closed PIP model without conservation constraints (lc), all for days 2001–4000.

also the analyzed dataset seems to be sufficiently large from a statistical point of view.

As for an evaluation of the results, a comparison with the relative tendency errors of the best EOF model (LC) for same number of patterns (also Fig. 12) indicates that by using a PIP analysis with a nonlinear closure much can be gained with respect to local predictive capabilities of the reduced model. The relative tendency variance error is in all cases examined about an order of magnitude smaller than in the case of EOF models with a nonlinear closure and of linearly closed PIP models. This is further borne out by the same prediction tests as those described for the EOF case. Once again we have calculated for the two 11-PIP models a mean anomaly correlation and a root-mean-squared prediction error as defined by (50) and (51). Here EOF expansion coefficients must be substituted by their PIP counterparts. Two thousand integrations over 100 days have been taken into account, starting from day 4001 up to 6000. Figure 13 shows the results, together with corresponding values for projected and best closed EOF model (LC) and for a persistence assumption for the PIPs. The advantage of the nonlinearly closed PIP model
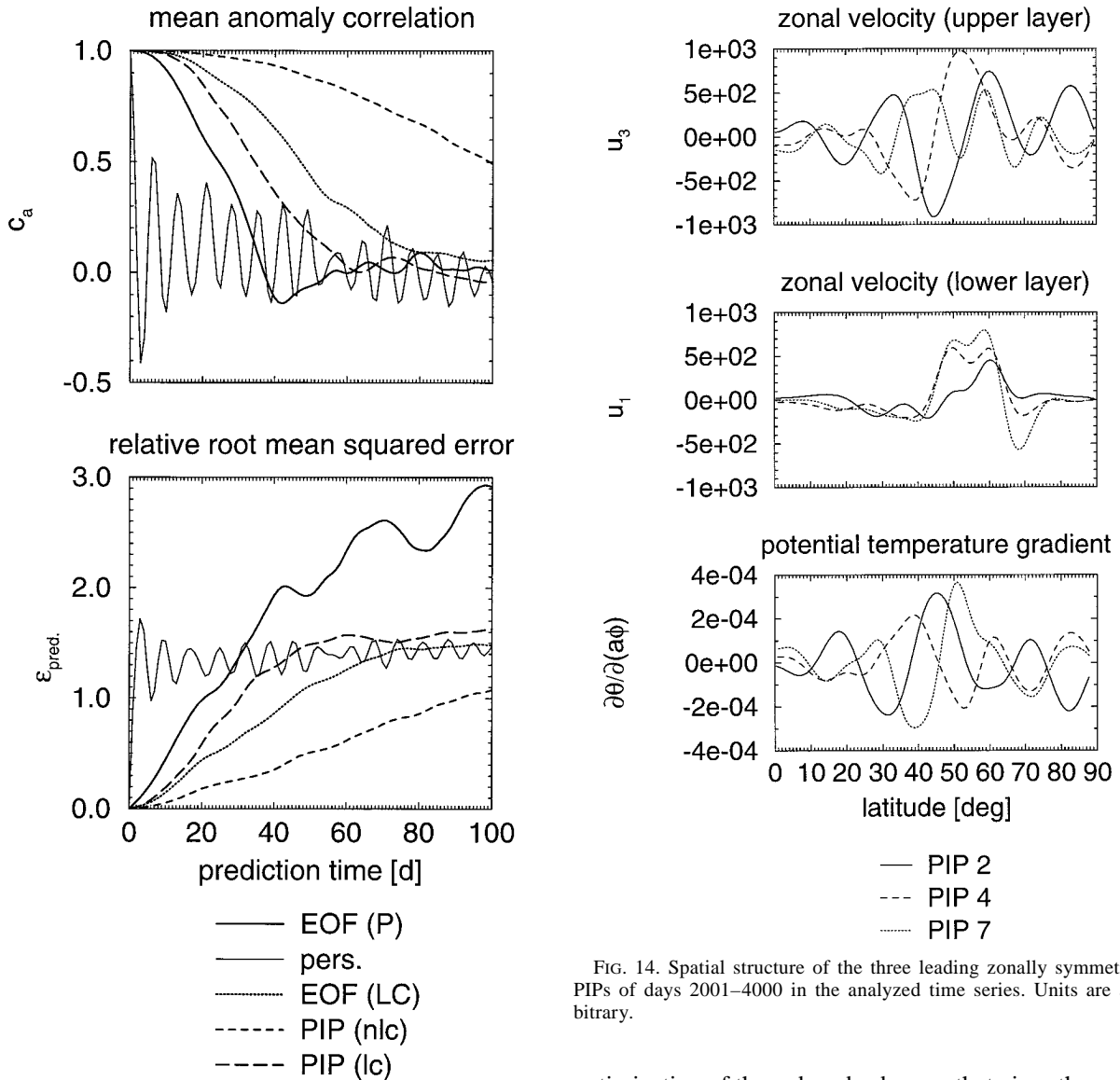
FIG. 13. Mean anomaly correlation and relative root-mean-square error of the purely projected and truncated (i.e., nonclosed) EOF model (P), the best closed EOF model (LC, i.e., all closure terms observe energy conservation constraints), the linearly (lc) and nonlinearly (nlc) closed PIP model, and a persistence assumption for EOFs, for 11 patterns.

FIG. 14. Spatial structure of the three leading zonally symmetric PIPs of days 2001–4000 in the analyzed time series. Units are arbitrary.

over all EOF models is quite obvious. On the other hand, the linearly closed PIP model lies in predictive capability somewhere between projected EOF model and nonlinearly closed EOF model. This is not in contradiction to the determined relative tendency errors shown above, since a closer inspection of the first few days of Fig. 13 shows that the linearly closed PIP model performs slightly better than the nonlinearly closed EOF model for the first 4 days. Thereafter, however, it quickly falls behind the EOF model. Thus, it seems to be indeed the nonlinear closure scheme in combination with the

optimization of the reduced subspace that gives the nonlinearly closed PIP model its advantage. The same can be stated for all other numbers of patterns examined so that nonlinearly closed PIP models seem to be the best alternative for a reduced model with strong short-term predictive capabilities.

Similar statements can be made about longtime climate simulations, however with a caveat. Because a PIP analysis is not variance oriented like EOF analysis, the identified PIP subspace can describe only part of the full variance of the system. For example, 15 PIPs explain only 86.45% of the variance, whereas for 15 EOFs the corresponding number is 99.81%. The different spatial structures of PIPs that lead to this effect can be observed for the first three zonally symmetric and wave patterns, respectively, from the 15-PIP optimization in Figs. 14–16. The basis patterns have been made unique by choosing them to be the eigenvectors of the scatter matrix for days 2001–4000 in PIP space (i.e., EOFs in
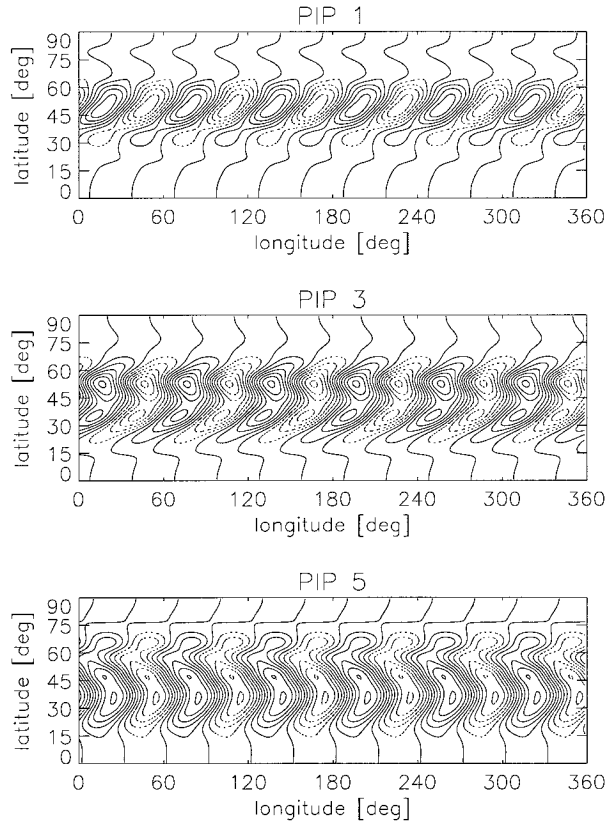
FIG. 15. Lower-layer streamfunction of the three leading wave PIPs of days 2001–4000 in the analyzed time series. Dashed lines indicate negative values, solid lines positive ones. Contour intervals are constant but in arbitrary units.

FIG. 16. Same as Fig. 15 but for the upper layer. Contour intervals are as in Fig. 15.

PIP space; they have as a consequence also been ordered with respect to the variance in PIP space thus being found to be associated with them). Therefore, one either should compare PIP integrations with the behavior of the full system's projection onto PIP subspace or use some additional tool for parameterizing the error between full system and PIP subspace as a function of PIP coefficients. Such a "downscaling" scheme can be readily formulated. If

$$\delta\mathbf{\Phi}^{\text{EOF}} = \sum_\nu z_\nu \mathbf{e}_\nu^z + \sum_\nu w_\nu \mathbf{e}_\nu^w$$

is the expansion of the deviation of the state vector from the time mean in all EOFs taken into account in the PIP reduction, what is needed is a function $\mathbf{S}(\tilde{z}_\nu, \tilde{w}_\nu)$ such that, at good accuracy,

$$\delta\mathbf{\Phi}^{\text{EOF}} = \delta\mathbf{\Phi}^{\text{pr}} + \mathbf{S}. \tag{64}$$

If S_ν^z and S_ν^w are expansion coefficients of this function for zonally symmetric and wave EOFs, respectively, this is equivalent to
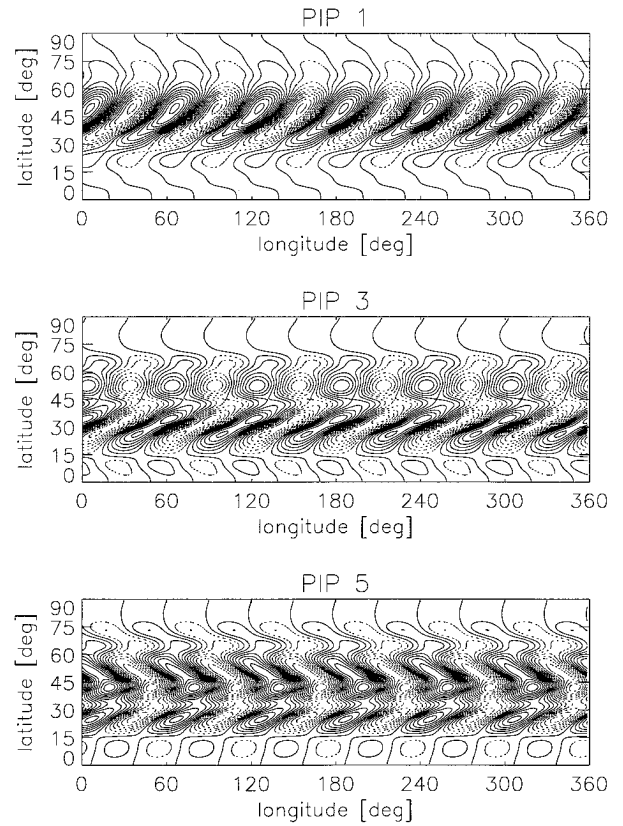
$$S_\nu^z = z_\nu - \sum_\mu \tilde{z}_\mu p_{\nu\mu}^z \tag{65}$$

and

$$S_\nu^w = w_\nu - \sum_\mu \tilde{w}_\mu p_{\nu\mu}^w. \tag{66}$$

As additional properties, we require that the error modeling scheme does not violate translational invariance with respect to longitude (because of the zonal symmetry in the whole system) and that the zonally symmetric equilibrium state with error modeling is the same as without. All of these conditions are fulfilled by the rather general functions

$$S_\nu^z = \sum_\mu \alpha_{\nu\mu}(\tilde{z}_\mu - \tilde{Z}_\mu) + \sum_{\mu,\rho} \beta_{\nu\mu\rho}\tilde{w}_\mu\overline{\tilde{w}_\rho} \tag{67}$$

and

$$S_\nu^w = \sum_\mu \gamma_{\nu\mu}\tilde{w}_\mu + \sum_{\mu,\rho} \delta_{\nu\mu\rho}\tilde{z}_\mu\tilde{w}_\rho$$
$$+ \sum_{\mu,\rho} \varepsilon_{\nu\mu\rho}\tilde{w}_\mu\tilde{w}_\rho + \sum_{\mu,\rho} \zeta_{\nu\mu\rho}\tilde{w}_\mu\overline{\tilde{w}_\rho}, \tag{68}$$
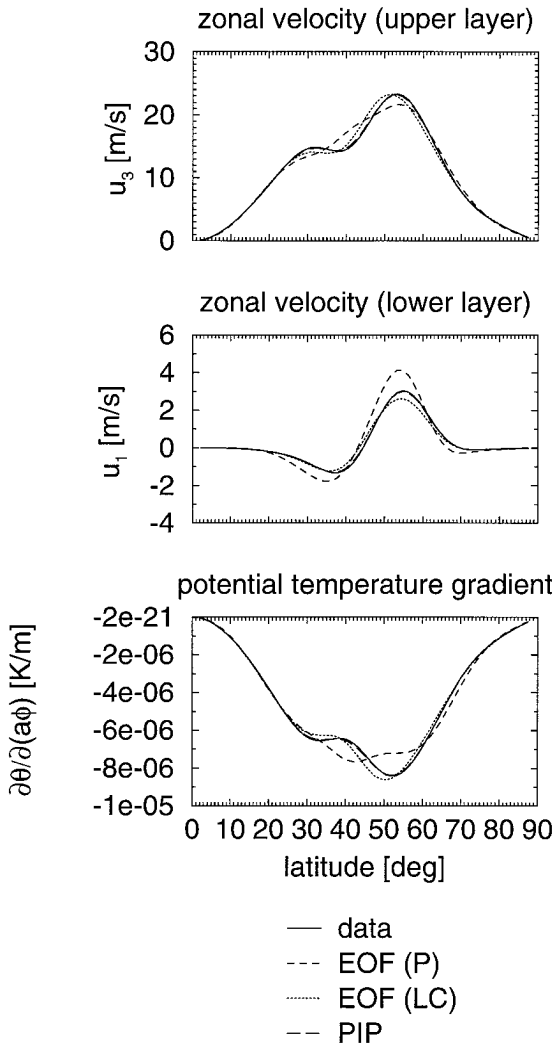
where

FIG. 17. Time and zonal means of the zonal wind in both layers and of the meridional potential temperature gradient, as obtained by a nonlinearly closed PIP model with 11 PIPs with error modeling, integrated from day 4001 to day 6000, in comparison with the corresponding values in the two-layer model data, the nonclosed, that is, purely projected (P), and the closed (LC) EOF model (in which all closure terms observe energy conservation constraints) with the same number of patterns. The PIP result is virtually identical to the corresponding values in the original two-layer model data.
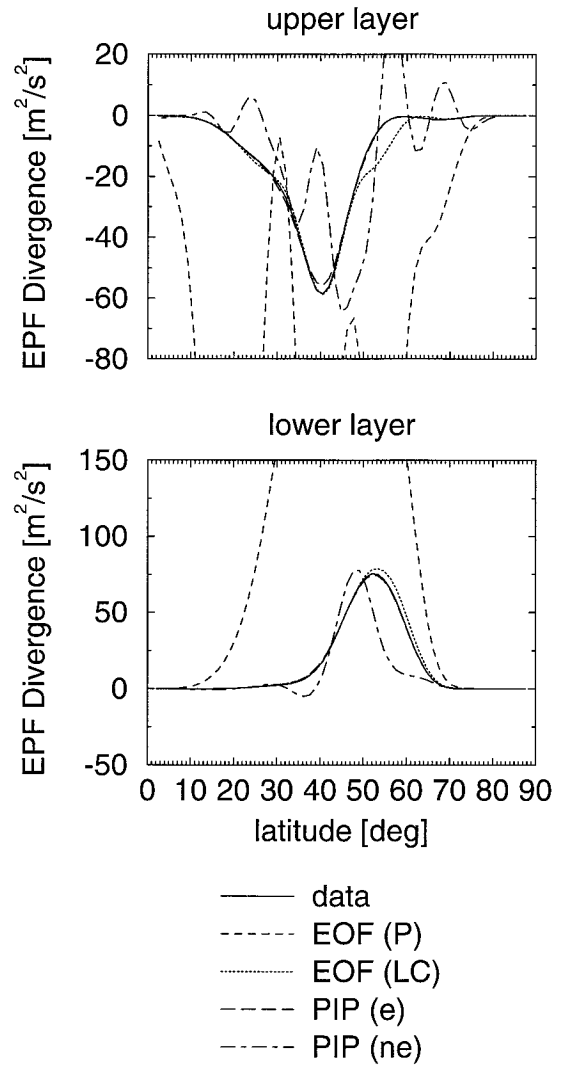
FIG. 18. Same as Fig. 12 but for the Eliassen–Palm Flux (EPF) divergencies in both layers. PIP results are given both with (e) and without error modeling (ne). The PIP result (with error model) is virtually identical to the corresponding values in the original two-layer model data.

$$\beta_{\nu\mu\rho} = 0 \quad \text{if } m_\mu \neq m_\rho, \tag{69}$$

$$\gamma_{\nu\mu} = 0 \quad \text{if } m_\nu \neq m_\mu, \tag{70}$$

$$\delta_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\rho, \tag{71}$$

$$\varepsilon_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\mu + m_\rho, \tag{72}$$

and

$$\zeta_{\nu\mu\rho} = 0 \quad \text{if } m_\nu \neq m_\mu - m_\rho. \tag{73}$$

For each number of PIPs the tensors α, β, γ, δ, ε, and ζ (e.g., counting real and imaginary parts separately, 4507 coefficients for the case of 15 PIPs and 60 un-

derlying EOFs) have been determined, from (65)–(73) and the dataset used for the PIP analysis, by linear regression.

For the case of 11 patterns Fig. 17 shows the zonal-mean zonal wind in both layers and the zonal-mean meridional potential temperature gradient, averaged over day 4001–6000, as obtained by integrations of the projected (nonclosed) EOF model, a closed EOF model (LC), and the nonlinearly closed PIP model with error modeling, in comparison to the same values in the original data. Figure 18 shows the corresponding time and zonal-mean EPF divergences. Here the PIP result is given both with and without error modeling. While the closed EOF model can simulate all climatological means considered rather accurately, the PIP model yields near-
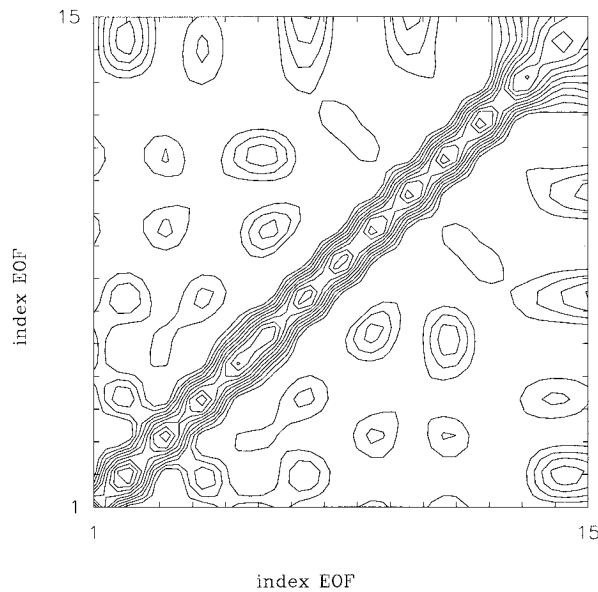
FIG. 19. Matrix of absolute values of the coefficients of the correlation matrix between expansion coefficients of the leading 15 EOFs, calculated from the result of an integration of the nonlinearly closed EOF model with 15 patterns, in which all closure terms observe energy conservation constraints (LC), from day 4001 to day 6000. Contours are at 0.1, 0.2, . . . , 1.0.

ly perfect agreement with the full system if error modeling is applied. This advantage of the nonlinearly closed PIP model in simulating time and zonal means has been observed in all other cases besides the one with 14 patterns. Here PIPs and EOFs perform about equally well. A more general impression than from the
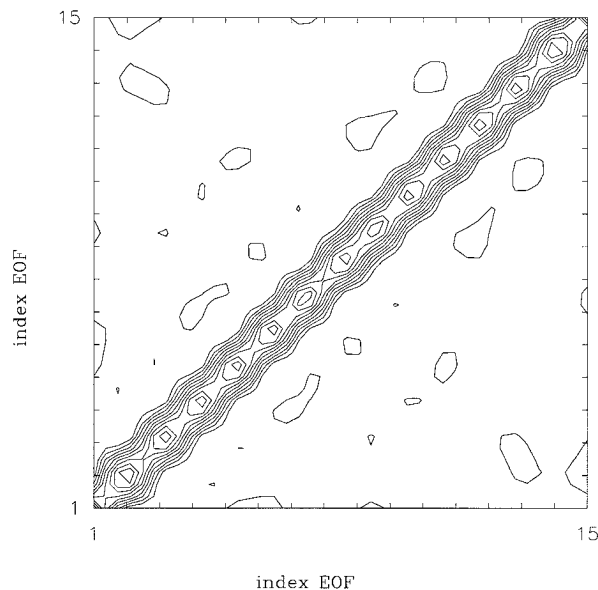

FIG. 20. Same as Fig. 19 but calculated from the result of an integration of the nonlinearly closed PIP model with 15 patterns. Contours are at 0.1, 0.2, . . . , 1.0.
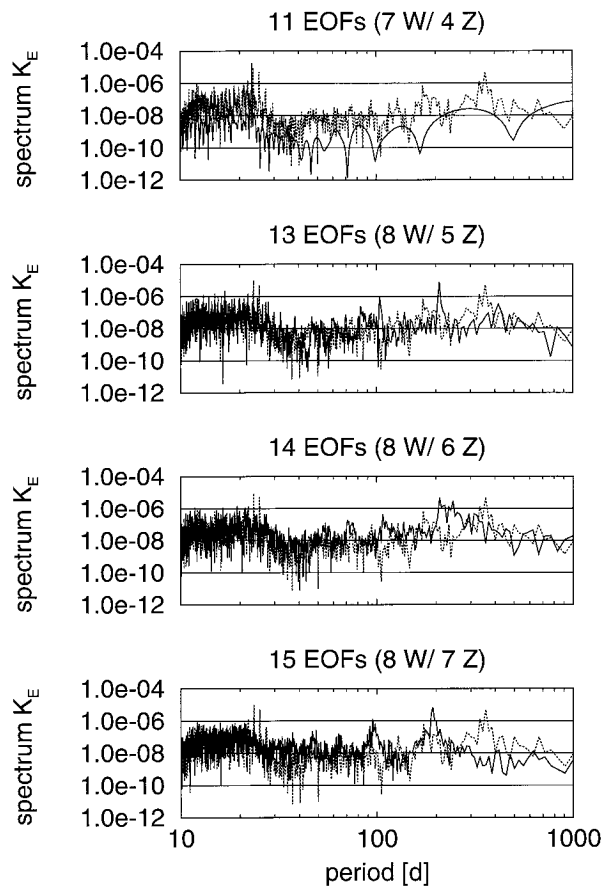

FIG. 21. Power spectrum of eddy kinetic energy in simulations by the closed EOF model in which all closure terms observe energy conservation constraints (LC) for 11 and 13–15 patterns (solid line) in comparison to the original data (dotted). The simulation time is from day 4001 to day 14 000. The brackets indicate how many of the considered EOFs are wave patterns and how many are zonally symmetric

EPF divergences in the second-moment simulation capabilities of the two respective models is given, for the case of 15 patterns, in Figs. 19 and 20. There the matrices of absolute values of relative correlation coefficients between expansion coefficients of the leading 15 EOFs from simulations by the two respective models—nonlinearly closed EOF (LC) and PIP model—are shown. The covariance structure of the original time series, that is, a diagonal correlation matrix, is better reproduced by the PIP model. Long-time simulation results for the linearly closed PIP model are not shown since it turned out to perform systematically worse than the nonlinearly closed EOF model. For 10–13 patterns the projected EOF model yields better agreement with the data.

As for simulations of the longtime behavior of energy reader can conclude from the power spectrum of eddy kinetic energy in simulations by the closed EOF model (LC) and by the nonlinearly closed PIP model from day 4001 up to 14 000, for 11 and 13–15 patterns, in Figs.
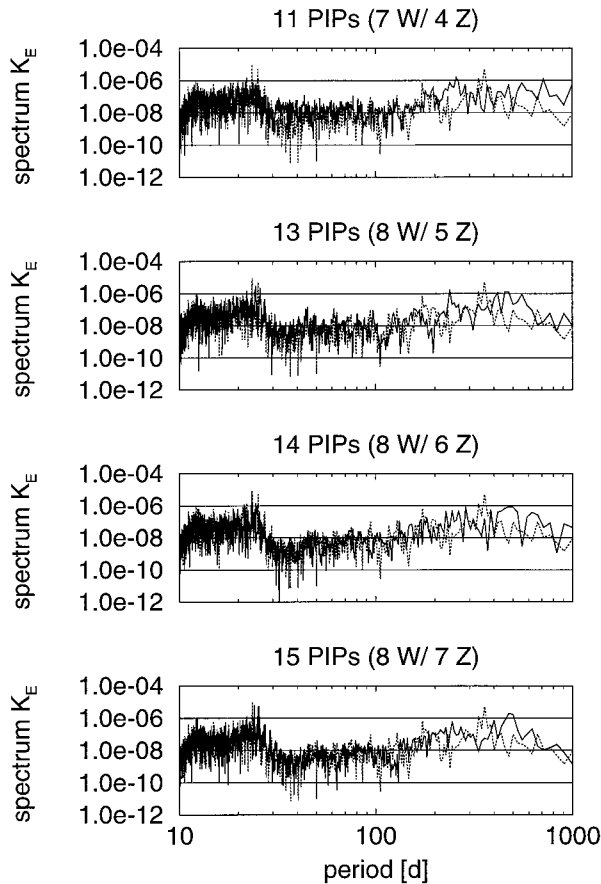
FIG. 22. Same as Fig. 21 but for the nonlinearly closed PIP model with error modeling.

FIG. 23. Same as Figs. 10 and 11 but for a nonlinearly closed PIP model with error modeling, with 15 PIPs.

21 and 22, that in all cases PIPs are better able to simulate the full system. In the two other cases considered (10 and 12 patterns) neither model is able to reproduce the energy behavior at any acceptable quality. In all cases shown the characteristic double peak at lower periods (between 20 and 30 days) due to the life cycles is mimicked much better by PIPs than by EOFs. A problem inherent in both reductions is however also visible: both lead to simulations with some kind of longtime modulation. Unfortunately, the corresponding timescale is too short in the EOF case and too long for PIPs. Nonetheless, the difference is smaller in PIP simulations. This can also be observed directly in the time-dependent energies as given in Fig. 23 for the case of 15 PIPs. The modulation decay phase especially is better reproduced by this model than even by the above-discussed 17-EOF (LC) model (see Fig. 11).

5. Summary and conclusions

So far, purely projected and truncated models (without parameterization of the feedback from unresolved modes) on the basis of EOFs or PIPs had been demonstrated to have some potential for reductions of stan-
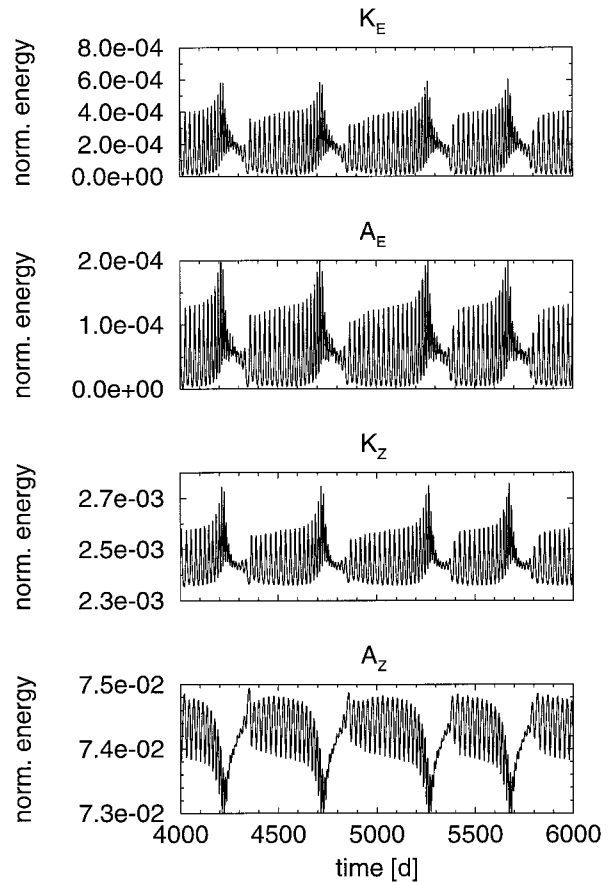
dard models of atmospheric behavior. A problem that had not yet been well solved is the related closure problem, that is, how the influence of patterns not explicitly resolved in the reduced model could be parameterized as a function of explicitly considered patterns. Thus, in reduced models described in the literature the number of degrees of freedom is still much higher than theoretically necessary. Nonetheless, first steps toward an improvement of this situation have been taken. In their analyses of the potentials of EOF and PIP models Selten (1995a) and Kwasniok (1996) were successful in including empirical linear closure schemes. An important attempt at using a nonlinear closure was made by Selten (1995b). However, in spite of progress with respect to local prediction, longtime integrations of his quadratically closed EOF model are not possible because of unbounded explosive growth of energy in the model after some short integration time. As a way out of this dilemma he proposed, among two possibilities, to impose some kind of energy conservation constraints on the closure parameters, which are themselves determined empirically from the full system examined. We have followed this suggestion and thereby have been able to make substantial progress.

The time series analyzed for a test of the formulated closure approach describes multiple baroclinic wave life cycles at zonal wavenumber 6 and all higher harmonics in a quasigeostrophic two-layer model on the sphere with surface friction and Newtonian cooling, integrated at horizontal resolution T42. The choice of this time series was motivated by the general importance of baroclinic wave life cycles and by the quasiperiodic character of the phenomenon, which yields rather clear criteria about what aspects of the time series a reduced model should be able to simulate. Additionally, sampling problems can already be excluded at a modest length of the time series.

Because, in any case, the empirical identification of an acceptable nonlinear closure has, at its heart, a multidimensional regression problem that can be numerically demanding, it is quite helpful to minimize the number of free variables in a given number of patterns and for some predefined class of closure schemes. It is therefore, from a theoretical point of view, attractive to exclude by use of first principles beforehand as many conceivable candidates for reduced closed models as possible. In this respect much can already be gained from the fact that our two-layer model has no orography and that all physical processes included are independent of longitude. From this it can be deduced that each EOF must be completely contained within the subspace of one zonal wavenumber. Also, for PIPs it can be shown that the identified PIP subspace must have an orthonormal set of basis vectors that have the same property.

We assume the tendency equations of the reduced model to be at most quadratic in the expansion coefficients of the state vector in all patterns, as also projected and truncated (nonclosed) models are. Zonal symmetry in the boundary conditions and physical processes in the two-layer model, together with the associability of each basic pattern with a single zonal wavenumber can then be further used to greatly diminish the number of terms allowed in the tendency equations. A lot of helpful coupling conditions arise that are by no means due to the quadratic nature of the model. Generalizations are easily possible for all models with tendencies given by polynomials of the expansion coefficients.

As experience tells us, purely projected and truncated (i.e., nonclosed) models never exhibit catastrophic growth of energy at any time. The crucial point now is to ensure that the difference of the reduced model from the purely projected and truncated version, that is, the closure terms, does not lead to such an effect either. This can be done by demanding that all closure terms do not change the time derivative of turbulent energy as given by the truncated model. Following such an approach, we have derived a set of constraints on the closure corrections, the use of which we could show in all cases examined to facilitate avoidance of the above-described stability problems.

The best closure corrections have been determined both in the EOF and PIP case by minimizing for our

dataset (days 2001–4000 of the time series) the squared sum of differences between observed and model tendencies. In spite of the fact that unconstrained linear closures have so far not been found to destroy, in the absence of nonlinear closure terms, a reduced model's longtime simulation capability, we observe the opposite for the general case. Unconstrained determination of linear closure terms for nonlinearly closed EOF models leads to reduced models that are much less well behaved than the ones with linear closure terms not violating turbulent energy conservation. In simulations with the first, wave energy tends to become dramatically too large. Thus, in nonlinear closure schemes it seems rather helpful to ensure that even the linear closure terms do not lead to time derivatives of turbulent energy that are different from the one in truncated models. This means that at any time there is absolutely no net exchange of energy between resolved and unresolved modes. It is clear that, in reality, only on time average, that is, in a statistical sense, should net energy loss by the resolved modes vanish because of a balance between energy influx at predominantly large, and outflow at mainly small, scales. The progress made possible by the scheme justifies the approach. Nonetheless, improvements that take this discrepancy into account might still be possible. This notwithstanding, nonlinearly closed models using our approach are already superior to nonclosed models, both with respect to short-time prediction (as checked by standard prediction tests) and to longtime climate simulations. Both time and zonal climate means and the typical frequency behavior of turbulent energy are much better reproduced with than without a closure. As for the reductive capabilities, it is found that a closed model with 17 EOFs (7 zonally symmetric, 10 in wave subspace) is rather well able to simulate most important aspects of the original time series. It can be expected that a reduced model with higher nonlinearity than quadratic (as used here) could suffice with even less patterns.

PIP analyses have been conducted for the same dataset with a PIP model analogous to the best closed EOF model, that is, with energy-conserving linear and nonlinear closure corrections. In addition to the model closure coefficients the patterns also were variables of the associated minimization problem. The minimized relative tendency error can be reduced further by about an order of magnitude in comparison to closed EOF models. As a consequence nonlinearly closed PIP models are strongly superior to closed EOF models with respect to short-time prediction. Similar advantages are observed for long-time climate simulations, provided an additional error modeling scheme is utilized, after integrating the PIP model, for parameterizing the difference between the state vector and its projection onto the PIP subspace as a function of all PIP expansion coefficients. At the same number of patterns PIPs are at least as good as EOFs with a closed model at simulating important time and zonal means. In long-time simula-

TABLE 1. Relative tendency variance error made in the explanation of the tendencies in days 2001–4000 from runs of the two-layer model as the one analyzed so far in this work, but with different Newtonian cooling h_N, by various reduced models with 15 patterns, with patterns and closure terms as determined from the previously analyzed time series with $h_N = 0.1$ d^{-1}. EOF (P) denotes projected (nonclosed) EOF model, EOF (LC) the nonlinearly closed EOF model with energy conserving closure terms, and PIP (nlc) the nonlinearly closed PIP model. The corresponding number for the projected (nonclosed) EOF model using for each time series its own (new) leading 15 EOFs is also given.

h_N (d^{-1})	EOF (P, new patterns)	EOF (P)	EOF (LC)	PIP (nlc)
0.09	1.902×10^{-3}	1.627×10^{-3}	4.748×10^{-4}	1.474×10^{-4}
0.08	2.392×10^{-3}	2.302×10^{-3}	1.726×10^{-3}	7.131×10^{-4}
0.07	3.372×10^{-3}	3.917×10^{-3}	5.133×10^{-3}	2.249×10^{-3}
0.06	4.736×10^{-3}	5.970×10^{-3}	1.199×10^{-2}	5.472×10^{-3}
0.05	4.449×10^{-3}	8.090×10^{-3}	1.960×10^{-2}	9.475×10^{-3}

tions they are clearly better at reproducing most important frequencies of energy in the data. Reductions with a linearly closed PIP model without conservation constraints have also been performed. For these no advantage over nonlinearly closed EOF models was found.

In a nutshell, we have outlined a viable approach for extracting from data of complex atmospheric models reduced models on the basis of EOFs or PIPs with a total energy metric that includes a nonlinear closure scheme parameterizing the impact of not explicitly resolved components of the dynamics. Closed models determined by the approach are much more powerful than purely projected (nonclosed) EOF models. Given that in long-time simulations EOF models also tend to have reasonable merits, the choice of the kind of patterns (EOFs or PIPs) to be used will in practice depend strongly on the stringency of the requirements for the reduced model. Here the numerical effort can also play its part. As an example, the extraction of a reduced model on the basis of 15 EOFs took 14 h 30 min on an HP 735/125 workstation, whereas a PIP reduction for the same number of patterns needed about 960 h on the same machine. It should be noted that there might still be one possibility of considerably reducing this time. Since with respect to the closure coefficients the optimization problem is a linear one, it would be possible to solve for every set of patterns this part by standard methods and use the (analytic) result in the relative tendency variance error. The minimization problem might thus be reduced to one with respect to the pattern coefficients.

For climatological applications another point might be interesting to note. There can be no doubt that all of the closure corrections (and the optimal patterns) are in reality functions of the parameters of the complex system. Changes in these do, in principle, necessitate new determinations of all patterns and closure terms from a time series produced by a model with the new parameters. On the other hand, a short analysis has shown that, for not too dramatic changes, all can at least approximately be assumed to be constant. Table 1 gives the result obtained by using the above-determined closure corrections and patterns for both an EOF model (LC, that is, with energy conservation in all closure terms) and a PIP model with 15 patterns in recalculating the relative tendency variance error in explaining days 2001–4000 of time series produced by the same two-

layer model as before, but with Newtonian cooling h_N relaxed by varying degrees. For modest changes of this parameter the old closure corrections and patterns still give smaller errors than a newly determined nonclosed EOF model (with patterns determined from the respective new time series). Larger parameter changes alter the dynamical character of the new time series so much (it becomes more irregular) that the old closure corrections lose their usefulness. One might be tempted to think that the positive result for moderate parameter changes might be due to the elements of the closure tensors being much smaller than their projected counterparts. Inspection of this has, however, shown that they are indeed considerably larger than the latter.

Finally, it should be stressed that the approach taken here is by no way limited to the special case considered. Although zonal symmetry in boundary conditions and physics of the examined system helps in reducing the numerical effort and also permits theoretically nice beforehand results on the class of allowed reduced models, it is not a prerequisite for the deduction of constraints on the closure terms ensuring energy conservation. Also, the quasigeostrophy of the examined system or the number of layers is no precondition for the applicability of the approach. The numerical effort for the reduction of multilayer primitive equation models would be greater and possibly only tolerable on high-performance computers. Nonetheless, given the prospect that a reduced model of atmospheric flow could be much more transparent to conceptual understanding of simulation results than standard complex models, we are optimistic that the admittedly still long way toward such an aim will eventually be successfully taken.

APPENDIX

Derivation of the Tendency Equations for the Projected EOF Model

Together with the reality conditions (13) and (14) the spherical harmonic expansions (4) and (5) can be rewritten as

$$\psi = \sum_n \psi_n^0 Y_n^0 + \sum_{m>0,n} \left(\psi_n^m Y_n^m + \overline{\psi_n^m}\,\overline{Y_n^m} \right) \quad \text{(A1)}$$

and

$$\tau = \sum_n \tau_n^0 Y_n^0 + \sum_{m>0,n} \left(\tau_n^m Y_n^m + \overline{\tau_n^m}\,\overline{Y_n^m} \right). \quad \text{(A2)}$$

The truncated EOF expansions of these are

$$\psi_{\mathrm{tr}} = \hat{\psi} + \sum_n (\psi_{\mathrm{tr}})_n^0 Y_n^0$$
$$+ \sum_{m>0,n} \left[(\psi_{\mathrm{tr}})_n^m Y_n^m + \overline{(\psi_{\mathrm{tr}})_n^m}\,\overline{Y_n^m} \right] \quad \text{(A3)}$$

and

$$\tau_{\mathrm{tr}} = \hat{\tau} + \sum_n (\tau_{\mathrm{tr}})_n^0 Y_n^0$$
$$+ \sum_{m>0,n} \left[(\tau_{\mathrm{tr}})_n^m Y_n^m + \overline{(\tau_{\mathrm{tr}})_n^m}\,\overline{Y_n^m} \right] \quad \text{(A4)}$$

with

$$(\psi_{\mathrm{tr}})_n^m = \begin{cases} \sum_\nu z_\nu (\psi_\nu^z)_n^0 & \text{if } m = 0 \\ \sum_\nu w_\nu (\psi_\nu^w)_n^m & \text{if } m > 0 \end{cases} \quad \text{(A5)}$$

and

$$(\tau_{\mathrm{tr}})_n^m = \begin{cases} \sum_\nu z_\nu (\tau_\nu^z)_n^0 & \text{if } m = 0 \\ \sum_\nu w_\nu (\tau_\nu^w)_n^m & \text{if } m > 0. \end{cases} \quad \text{(A6)}$$

Each $(\psi_\nu^z)_n^0$ and $(\psi_\nu^w)_n^m$ $(m > 0)$ is identical to one $e_{i\nu}^z$ or $e_{i\nu}^w$, respectively. The same holds for the equivalent coefficients of the baroclinic streamfunction. Inserting (A5) into (A3) and (A6) into (A4) yields as expressions for the truncated streamfunctions

$$\psi_{\mathrm{tr}} = \hat{\psi} + \sum_\nu z_\nu \psi_\nu^z + \sum_\nu \left(w_\nu \psi_\nu^w + \overline{w_\nu}\,\overline{\psi_\nu^w} \right) \quad \text{(A7)}$$

and

$$\tau_{\mathrm{tr}} = \hat{\tau} + \sum_\nu z_\nu \tau_\nu^z + \sum_\nu \left(w_\nu \tau_\nu^w + \overline{w_\nu}\,\overline{\tau_\nu^w} \right), \quad \text{(A8)}$$

where

$$\psi_\nu^z = \sum_n (\psi_\nu^z)_n^0 Y_n^0, \quad \text{(A9)}$$

$$\tau_\nu^z = \sum_n (\tau_\nu^z)_n^0 Y_n^0, \quad \text{(A10)}$$

$$\psi_\nu^w = \sum_n (\psi_\nu^w)_n^{m_\nu} Y_n^{m_\nu}, \quad \text{(A11)}$$

and

$$\tau_\nu^w = \sum_n (\tau_\nu^w)_n^{m_\nu} Y_n^{m_\nu}. \quad \text{(A12)}$$

In the last two equations we have used the fact that each wave EOF has only one zonal wavenumber.

Now we are ready for calculating the projected tendency equations. Using in (25) and (26) the metric as defined in (23) together with the facts that because of the orthonormality of spherical harmonics, that is,

$$\oint_{4\pi} d\Omega\, \overline{Y_{n'}^{m'}} Y_n^m = \delta_{mm'} \delta_{nn'}, \quad \text{(A13)}$$

for any two horizontal functions

$$f = \sum_{m,n} f_n^m Y_n^m \quad \text{and} \quad g = \sum_{m,n} g_n^m Y_n^m$$

$$\sum_{m,n} \overline{f_n^m} g_n^m = \oint_{4\pi} d\Omega\, fg, \quad \text{(A14)}$$

and that

$$\nabla^2 Y_n^m = -n(n+1) Y_n^m,$$

we obtain

$$\dot{z}_\nu = -\frac{1}{2} \oint_{4\pi} d\Omega \left[\psi_\nu^z \nabla^2 \frac{\partial \psi}{\partial t}(\psi_{\mathrm{tr}}, \tau_{\mathrm{tr}}) \right.$$
$$\left. + \tau_\nu^z (\nabla^2 - r^2) \frac{\partial \tau}{\partial t}(\psi_{\mathrm{tr}}, \tau_{\mathrm{tr}}) \right] \quad \text{(A15)}$$

and

$$\dot{w}_\nu = -\oint_{4\pi} d\Omega \left[\overline{\psi_\nu^w} \nabla^2 \frac{\partial \psi}{\partial t}(\psi_{\mathrm{tr}}, \tau_{\mathrm{tr}}) \right.$$
$$\left. + \overline{\tau_\nu^w} (\nabla^2 - r^2) \frac{\partial \tau}{\partial t}(\psi_{\mathrm{tr}}, \tau_{\mathrm{tr}}) \right], \quad \text{(A16)}$$

which, together with (A7) and (A8), spherical harmonic orthonormality, and the coupling conditions that

$$\oint_{4\pi} d\Omega \left[\overline{Y_{n_1}^{m_1}} J(Y_{n_2}^{m_2}, Y_{n_3}^{m_3}) \right]$$

is only nonzero if $m_2^2 + m_3^2 > 0$ and $m_1 = m_2 + m_3$ (Silberman 1954), gives the desired equations (28) and (29) where

$$A_{\nu\mu}^{\mathrm{pr}} = -\frac{1}{2} \oint_{4\pi} d\Omega \left\{ \psi_\nu^z \{ k_s \nabla^2 (\tau_\mu^z - \psi_\mu^z) - \kappa \nabla^2 [\nabla^2(\nabla^2 \psi_\mu^z)] \} - \tau_\nu^z \{ k_s \nabla^2 (\tau_\mu^z - \psi_\mu^z) - h_N r^2 \tau_\mu^z \right.$$
$$\left. + \kappa(\nabla^2 - r^2)[\nabla^2(\nabla^2 \tau_\mu^z)] \} \right\},$$
$$\text{(A17)}$$

$$B^{\mathrm{pr}}_{\nu\mu\rho} = \frac{1}{2}\oint_{4\pi} d\Omega \big\{ \psi^{z}_{\nu}[J(\psi^{w}_{\mu}, \nabla^2\overline{\psi^{w}_{\rho}}) + J(\tau^{w}_{\mu}, \nabla^2\overline{\tau^{w}_{\rho}}) + J(\overline{\psi^{w}_{\rho}}, \nabla^2\psi^{w}_{\mu}) + J(\overline{\tau^{w}_{\rho}}, \nabla^2\tau^{w}_{\mu})]$$
$$ + \tau^{z}_{\nu}\{J[\psi^{w}_{\mu}, (\nabla^2 - r^2)\overline{\tau^{w}_{\rho}}] + J(\tau^{w}_{\mu}, \nabla^2\overline{\psi^{w}_{\rho}}) + J[\overline{\psi^{w}_{\rho}}, (\nabla^2 - r^2)\tau^{w}_{\mu}] + J(\overline{\tau^{w}_{\rho}}, \nabla^2\psi^{w}_{\mu})\}\big\}, \quad (A18)$$

$$C^{\mathrm{pr}}_{\nu\mu} = \oint_{4\pi} d\Omega \Big\{ 2\Big(\overline{\psi^{w}_{\nu}}\frac{\partial\psi^{w}_{\mu}}{\partial\lambda} + \overline{\tau^{w}_{\nu}}\frac{\partial\tau^{w}_{\mu}}{\partial\lambda}\Big) - \overline{\psi^{w}_{\nu}}\{k_s\nabla^2(\tau^{w}_{\mu} - \psi^{w}_{\mu}) - \kappa[\nabla^2(\nabla^2\psi^{w}_{\mu})]\}$$
$$ + \overline{\tau^{w}_{\nu}}\{k_s\nabla^2(\tau^{w}_{\mu} - \psi^{w}_{\mu}) - h_N r^2\tau^{w}_{\mu} + \kappa(\nabla^2 - r^2)[\nabla^2(\nabla^2\tau^{w}_{\mu})]\}$$
$$ + \overline{\psi^{w}_{\nu}}[J(\psi^{w}_{\mu}, \nabla^2\hat{\psi}) + J(\tau^{w}_{\mu}, \nabla^2\hat{\tau}) + J(\hat{\psi}, \nabla^2\psi^{w}_{\mu}) + J(\hat{\tau}, \nabla^2\tau^{w}_{\mu})]$$
$$ + \overline{\tau^{w}_{\nu}}\{J[\psi^{w}_{\mu}, (\nabla^2 - r^2)\hat{\tau}] + J(\tau^{w}_{\mu}, \nabla^2\hat{\psi}) + J[\hat{\psi}, (\nabla^2 - r^2)\tau^{w}_{\mu}] + J(\hat{\tau}, \nabla^2\psi^{w}_{\mu})\}\Big\}, \quad (A19)$$

$$D^{\mathrm{pr}}_{\nu\mu\rho} = \oint_{4\pi} d\Omega \big\{ \overline{\psi^{w}_{\nu}}[J(\psi^{w}_{\rho}, \nabla^2\psi^{z}_{\mu}) + J(\tau^{w}_{\rho}, \nabla^2\tau^{z}_{\mu}) + J(\psi^{z}_{\mu}, \nabla^2\psi^{w}_{\rho}) + J(\tau^{z}_{\mu}, \nabla^2\tau^{w}_{\rho})]$$
$$ + \overline{\tau^{w}_{\nu}}\{J[\psi^{w}_{\rho}, (\nabla^2 - r^2)\tau^{z}_{\mu}] + J(\tau^{w}_{\rho}, \nabla^2\psi^{z}_{\mu}) + J[\psi^{z}_{\mu}, (\nabla^2 - r^2)\tau^{w}_{\rho}] + J(\tau^{z}_{\mu}, \nabla^2\psi^{w}_{\rho})\}\big\}, \quad (A20)$$

$$E^{\mathrm{pr}}_{\nu\mu\rho} = \oint_{4\pi} d\Omega \big\{ \overline{\psi^{w}_{\nu}}[J(\psi^{w}_{\mu}, \nabla^2\psi^{w}_{\rho}) + J(\tau^{w}_{\mu}, \nabla^2\tau^{w}_{\rho})] + \overline{\tau^{w}_{\nu}}\{J[\psi^{w}_{\mu}, (\nabla^2 - r^2)\tau^{w}_{\rho}] + J(\tau^{w}_{\mu}, \nabla^2\psi^{w}_{\rho})\}\big\}, \quad (A21)$$

and

$$F^{\mathrm{pr}}_{\nu\mu\rho} = \oint_{4\pi} d\Omega \big\{ \overline{\psi^{w}_{\nu}}[J(\psi^{w}_{\mu}, \nabla^2\overline{\psi^{w}_{\rho}}) + J(\tau^{w}_{\mu}, \nabla^2\overline{\tau^{w}_{\rho}}) + J(\overline{\psi^{w}_{\rho}}, \nabla^2\psi^{w}_{\mu}) + J(\overline{\tau^{w}_{\rho}}, \nabla^2\tau^{w}_{\mu})]$$
$$ + \overline{\tau^{w}_{\nu}}\{J[\psi^{w}_{\mu}, (\nabla^2 - r^2)\overline{\tau^{w}_{\rho}}] + J(\tau^{w}_{\mu}, \nabla^2\overline{\psi^{w}_{\rho}}) + J[\overline{\psi^{w}_{\rho}}, (\nabla^2 - r^2)\tau^{w}_{\mu}] + J(\overline{\tau^{w}_{\rho}}, \nabla^2\tau^{w}_{\mu})\}\big\}. \quad (A22)$$

REFERENCES

Achatz, U., G. Schmitz, and K.-M. Greisiger, 1995: Principal interaction patterns in baroclinic wave life cycles. *J. Atmos. Sci.,* **52,** 3201–3213.

Baines, P. G., and J. S. Frederiksen, 1978: Baroclinic instability on a sphere in two-layer models. *Quart. J. Roy. Meteor. Soc.,* **104,** 45–68.

Brunet, G., 1994: Empirical normal-mode analysis of atmospheric data. *J. Atmos. Sci.,* **51,** 932–952.

Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.,* **53,** 2025–2040.

Ford, B., and J. C. T. Pool, 1984: The evolving NAG library service. *Sources and Development of Mathematical Software,* W. Cowell, Ed., Prentice Hall, 375–397.

Frederiksen, J. S., 1982: A unified three-dimensional instability theory of the onset of blocking and cyclogenesis. *J. Atmos. Sci.,* **39,** 969–982.

——, 1983: A unified three-dimensional instability theory of the onset of blocking and cyclogenesis. Part II: Teleconnection patterns. *J. Atmos. Sci.,* **40,** 2593–2609.

——, and R. C. Bell, 1987: Teleconnection patterns and the roles of baroclinic, barotropic and topographic instability. *J. Atmos. Sci.,* **44,** 2200–2218.

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.

——, 1997: Optimal Galerkin approximations of partial differential equations using Principal Interaction Patterns. *Phys. Rev. E,* in press.

Lindenberg, K., and B. J. West, 1984: Fluctuations and dissipation in a barotropic flow field. *J. Atmos. Sci.,* **41,** 3021–3031.

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus,* **12,** 364–373.

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

Schnur, R., G. Schmitz, N. Grieger, and H. von Storch, 1993: Normal modes of the atmosphere as estimated by principal oscillation patterns and derived from quasigeostrophic theory. *J. Atmos. Sci.,* **50,** 2386–2400.

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

——, 1986: The structure, energetics and evolution of the dominant frequency-dependent three-dimensional atmospheric modes. *J. Atmos. Sci.,* **43,** 1210–1237.

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

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

——, 1995b: An efficient description of large-scale atmospheric dynamics. Ph.D. thesis, University of Amsterdam, 169 pp.

Silberman, I., 1954: Planetary waves in the atmosphere. *J. Meteor.,* **11,** 27–34.

Simmons, A. J., J. M. Wallace, and G. W. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.,* **40,** 1363–1392.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.