Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability

S. KRAVTSOV, D. KONDRASHOV, AND M. GHIL*

Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California

ABSTRACT

Predictive models are constructed to best describe an observed field's statistics within a given class of nonlinear dynamics driven by a spatially coherent noise that is white in time. For linear dynamics, such inverse stochastic models are obtained by multiple linear regression (MLR). Nonlinear dynamics, when more appropriate, is accommodated by applying multiple polynomial regression (MPR) instead; the resulting model uses polynomial predictors, but the dependence on the regression parameters is linear in both MPR and MLR.

The basic concepts are illustrated using the Lorenz convection model, the classical double-well problem, and a three-well problem in two space dimensions. Given a data sample that is long enough, MPR successfully reconstructs the model coefficients in the former two cases, while the resulting inverse model captures the three-regime structure of the system's probability density function (PDF) in the latter case.

A novel multilevel generalization of the classic regression procedure is introduced next. In this generalization, the residual stochastic forcing at a given level is subsequently modeled as a function of variables at this level and all the preceding ones. The number of levels is determined so that the lag-0 covariance of the residual forcing converges to a constant matrix, while its lag-1 covariance vanishes.

This method has been applied to the output of a three-layer, quasigeostrophic model and to the analysis of Northern Hemisphere wintertime geopotential height anomalies. In both cases, the inverse model simulations reproduce well the multiregime structure of the PDF constructed in the subspace spanned by the dataset's leading empirical orthogonal functions, as well as the detailed spectrum of the dataset's temporal evolution. These encouraging results are interpreted in terms of the modeled low-frequency flow's feedback on the statistics of the subgrid-scale processes.

1. Introduction and motivation

Comprehensive general circulation models (GCMs) currently used to understand and predict climate variations (Randall 2000), as well as their simpler conceptual counterparts (Ghil and Robertson 2000), share a common difficulty of having to parameterize unresolved processes in terms of dynamical variables of interest.

---

* Additional affiliation: Départment Terre–Atmosphère–Océan and Laboratoire de Météorologie Dynamique/IPSL, Ecole Normale Supérieure, Paris, France.

Corresponding author address: Dr. Sergey Kravtsov, Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095-1565.
E-mail: sergey@atmos.ucla.edu

Some progress in achieving this goal has recently been made in the realm of so-called empirical climate models, by relaxing the requirement of a strict closure and assuming that errors may be treated as spatially coherent noise, that is, white in time (Penland 1989; Winkler et al. 2001). In this paper, we develop a data analysis approach that builds upon this progress and leads to the construction of a hierarchy of stochastically forced dynamical models that are based on observed climate statistics.

a. Statement of the problem

If $\mathbf{X}$ is the climate-state vector, $\overline{\mathbf{X}}$ its time mean, and $\mathbf{x} = \mathbf{X} - \overline{\mathbf{X}}$ the vector of anomalies, then the evolution of $\mathbf{x}$ is expressed as

$$\dot{\mathbf{x}} = \mathbf{L}\mathbf{x} + \mathbf{N}(\mathbf{x}). \qquad (1)$$

Here the dot denotes time derivative, $\mathbf{L}$ is a linear operator, and $\mathbf{N}$ represents nonlinear terms; both $\mathbf{L}$ and $\mathbf{N}$

may be functions of $\overline{\mathbf{X}}$, but this dependence is suppressed here. Even if the exact form of Eq. (1) were known, it would contain a very large number of degrees of freedom, so that its direct numerical integration would not be feasible due to insufficient computer power.

A common approach to solving Eq. (1) in practice is based on assuming scale separation. In this case, the full climate-variable vector $\mathbf{x}$ is represented as the sum of a climate "signal" $\mathbf{x}_S$ and a "noise" $\mathbf{x}'_N$:

$$\mathbf{x} = \mathbf{x}_S + \mathbf{x}'_N, \tag{2}$$

where the noise field is typically characterized by smaller scales in both space and time. Upon substituting the decomposition (2) into Eq. (1) and omitting the subscripts, the latter becomes

$$\dot{\mathbf{x}} = \mathbf{L}\mathbf{x} + \mathbf{N}(\mathbf{x}) + \mathbf{R}(\mathbf{x}, \mathbf{x}'). \tag{3}$$

To obtain a closed form of the reduced dynamics in Eq. (3), one has to make assumptions about the term $\mathbf{R}(\mathbf{x}, \mathbf{x}')$. A closure of this Reynolds stress term is used in many climate GCMs: one assumes that small-scale, high-frequency transients—due to instabilities of the large-scale, low-frequency flow—act on the latter as a linear diffusion that merely flattens, on long time scales, spatial gradients of the large-scale field; the corresponding eddy diffusivities are estimated from available data by trial and error. It is widely recognized, however, that the underlying assumption in this "eddy diffusion" closure does not generally hold.

If the full dynamical model in (1) is available and its integration is feasible, one can derive closed forms of the reduced model in (3) using a statistical–dynamical approach, by combining statistical properties of the modeled data and Eq. (1) that governs the flow dynamics. For example, Eq. (1) can be linearized about the numerically computed climatological state by setting $\mathbf{N} \equiv \mathbf{0}$, while the term $\mathbf{R}(\mathbf{x}, \mathbf{x}')$ in Eq. (3) may be treated as flow independent, spatially correlated noise, that is, white in time (Farrell and Ioannou 1993, 1995). One may also add information about climate variability to the knowledge about the time-mean climate.

To do so, one can define the large-scale flow as the one represented by a few leading empirical orthogonal functions (EOFs; Preisendorfer 1988). The nonlinear reduced dynamics model in (3) is then obtained by rewriting Eq. (1) in the truncated EOF basis (Rinne and Karhila 1975; Schubert 1985; Sirovich and Rodriguez 1987; Mundt and Hart 1994; Selten 1995, 1997), while treating the residual forcing as random. Alternatively, one can or by develop a deterministic, flow-dependent parameterization of unresolved processes $\mathbf{R}(\mathbf{x}, \mathbf{x}')$,

based on the library of differences between the tendency of the full and truncated models (D'Andrea and Vautard 2001; D'Andrea 2002). Yet another approach to this closure problem, which is mathematically rigorous in the limit of significant scale separation, has been developed by Majda et al. (1999, 2001, 2002, 2003). Franzke et al. (2005) have recently applied this approach to a barotropic flow model on the sphere, with a T21 resolution, while C. Franzke and A. Majda (2005, personal communication) have applied it to the Marshall and Molteni's (1993) baroclinic model.

b. Inverse stochastic models

The closure problem above can be effectively addressed in a data-driven, rather than model-driven approach, by using inverse stochastic models; these models rely almost entirely on the dataset's information content, while making only minimal assumptions about the underlying dynamics. The simplest type of inverse stochastic model is the so-called linear inverse model (LIM; Penland 1989, 1996). These models are obtained by assuming that $\mathbf{N}(\mathbf{x}) \, dt \approx \mathbf{T}\mathbf{x} \, dt + d\mathbf{r}^{(0)}$ in Eq. (1), where $\mathbf{T}$ is the matrix that describes linear feedbacks of unresolved processes on $\mathbf{x}$, and $d\mathbf{r}^{(0)}$ is a white noise process that can be spatially correlated:

$$d\mathbf{x} = \mathbf{B}^{(0)}\mathbf{x} \, dt + d\mathbf{r}^{(0)}, \quad \mathbf{B}^{(0)} = \mathbf{L} + \mathbf{T}. \tag{4}$$

The matrix $\mathbf{B}^{(0)}$ and the covariance matrix of the noise $\mathbf{Q} \equiv \langle \mathbf{r}^{(0)}\mathbf{r}^{(0)\mathrm{T}} \rangle$ can be directly estimated from the observed statistics of $\mathbf{x}$ by multiple linear regression (MLR; Wetherill 1986). LIMs have shown some success in predicting ENSO (Penland and Sardeshmukh 1995; Johnson et al. 2000), tropical Atlantic SST variability (Penland and Matrosova 1998), as well as extratropical atmospheric variability (Winkler et al. 2001). These models are typically constructed in the phase space of the system's leading EOFs. The state vector $\mathbf{x}$, or predictor-variable vector, consists of amplitudes of the corresponding principal components (PCs), while the vector of response variables is that of their time derivatives $\dot{\mathbf{x}}$. One should bear in mind, however, that the choice of predictor and response variables is not unique: to focus on the phenomena and time scales of interest, one has to choose the most appropriate subspaces of the climate system's full phase space.

c. This paper

In most geophysical situations, the assumptions of linear, stable dynamics, and white noise forcing used to construct LIMs are only valid to a certain degree of approximation. In particular, the stochastic forcing $\mathbf{r}^{(0)}$

in Eq. (4) typically involves serial correlations. In addition, when the nonlinearity is strong, the matrices $\mathbf{B}^{(0)}$ and $\mathbf{Q}$ obtained from the data can exhibit substantial dependence on the time scales considered (Penland and Ghil 1993).

In the present paper, we consider generalizations of LIMs that use additional statistical information to account for nonlinearity, among other things, and apply these generalized models to the analysis of climatic time series. One major modification of LIMs is obtained by assuming a polynomial, rather than linear form of $\mathbf{N}(\mathbf{x})$ in Eq. (1), in particular, a quadratic dependence.

The ith component $N_i(\mathbf{x})$ of $\mathbf{N}$ can then be written as

$$N_i(\mathbf{x}) \, dt \approx (\mathbf{x}^{\mathrm{T}} \mathbf{A}_i \mathbf{x} + \mathbf{t}_i \mathbf{x} + c_i^{(0)}) \, dt + dr_i^{(0)}. \quad (5)$$

The matrices $\mathbf{A}_i$ are the blocks of a third-order tensor, and the vectors $\mathbf{b}_i^{(0)} = \mathbf{l}_i + \mathbf{t}_i$ are the rows of the matrix $\mathbf{B}^{(0)} = \mathbf{L} + \mathbf{T}$ [cf. Eq. (4)]. These objects, as well as the components of the intercept vector $\mathbf{c}^{(0)}$, are estimated here by multiple polynomial regression (MPR; McCullagh and Nelder 1989), rather than by the MLR used to construct LIMs. The two methods (MLR and MPR) are algorithmically similar, though, inasmuch as their dependence on the regression parameters is linear in both.

The other major modification is to consider an iterative process of model construction, in which the residual noise $\mathbf{r}^{(0)}$ is tested for whiteness. If this test fails, $\mathbf{r}^{(0)}$ is modeled in turn by the same regression approach, and so on, until $\mathbf{r}^{(L)}$ satisfies the white noise test.

We introduce the MPR methodology in section 2 using data generated by simple nonlinear models of geophysical flows. The appendix describes in some detail MPR algorithms and the ways to solve the collinearity problems that often arise when the number of regression parameters to be determined is large. Section 3 deals with our multilevel generalization of the standard regression procedure, which addresses the problem of serial correlations in $\mathbf{r}^{(0)}$.

We apply the resulting multilevel, nonlinear method to a long simulation of the three-layer, quasigeostrophic (QG) atmospheric model of Marshall and Molteni (1993) and to a set of Northern Hemisphere (NH) geopotential height data (Smyth et al. 1999) in section 4. A summary of the results and a brief discussion of practical applications of the method follow in section 5. The use of the same methodology to quantify and predict seasonal-to-interannual climate variability based on a set of global sea surface temperatures is described in a companion paper (Kondrashov et al. 2005).

2. Didactic examples

a. The Lorenz model

1) DYNAMICAL MODEL AND EXPERIMENTAL DESIGN

The Lorenz (1963a) model (see also Ghil and Childress 1987, their section 5.4) is derived via spectral truncation of the full system of partial differential equations that describes Rayleigh–Bénard convection. The resulting system of three ordinary differential equations for the nondimensional variables $\mathbf{x} \equiv (x_1, x_2, x_3)^{\mathrm{T}}$ is

$$\dot{x}_1 = -sx_1 + sx_2, \quad (6a)$$

$$\dot{x}_2 = -x_1 x_3 + rx_1 - x_2, \quad (6b)$$

$$\dot{x}_3 = x_1 x_2 - bx_3, \quad (6c)$$

where a dot denotes the derivative with respect to nondimensional time t, and $(\ldots)^{\mathrm{T}}$ denotes the transpose. For the parameters, we choose the usual values $s = 10$, $r = 28$, and $b = 8/3$, known to produce chaotic behavior, and perform a long numerical intergration of the model. The time series of model variables from a sample subset of length $T = 20$ nondimensional units of this integration are shown in Fig. 1. The shortest characteristic time scale of the model's chaotic variability can be estimated by counting the number of oscillations of x_3 over the interval of integration; it is equal, approximately, to $T_{\mathrm{ch}} = 0.75$.

We sample the model-generated data every Δt time units, and vary the sampling interval Δt, as well as the length T of the time series. The coefficients of the Lorenz model are then reconstructed for each dataset, characterized by a pair of $(T, \Delta t)$, using multiple quadratic regression (see below). The response variables (y_1, y_2, y_3) are the time derivatives $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$ of the predictor variables (x_1, x_2, x_3). These time derivatives are estimated by finite differences as

$$y_i^j \approx \frac{1}{2\Delta t}(-3x_i^j + 4x_i^{j+1} - x_i^{j+2}), \quad i = 1, 2, 3, \quad (7)$$

where j is the time index; concentrated differencing is used in Eq. (7) to avoid the computational mode inherent to centered "leapfrog" differencing. The standard error σ_{y_i} of our "observations" can be estimated from a time series of the second derivatives $y_i''^j$ as

$$\sigma_{yi} \approx \frac{1}{3} < y_i''^j > \Delta t^2. \quad (8)$$

Here, each time series $y_i''^j$, $i = 1, 2, 3$ is constructed by using the standard second-order-accurate second time derivative of y_i^j, $i = 1, 2, 3$, while the angle brackets denote the standard deviation of this time series.
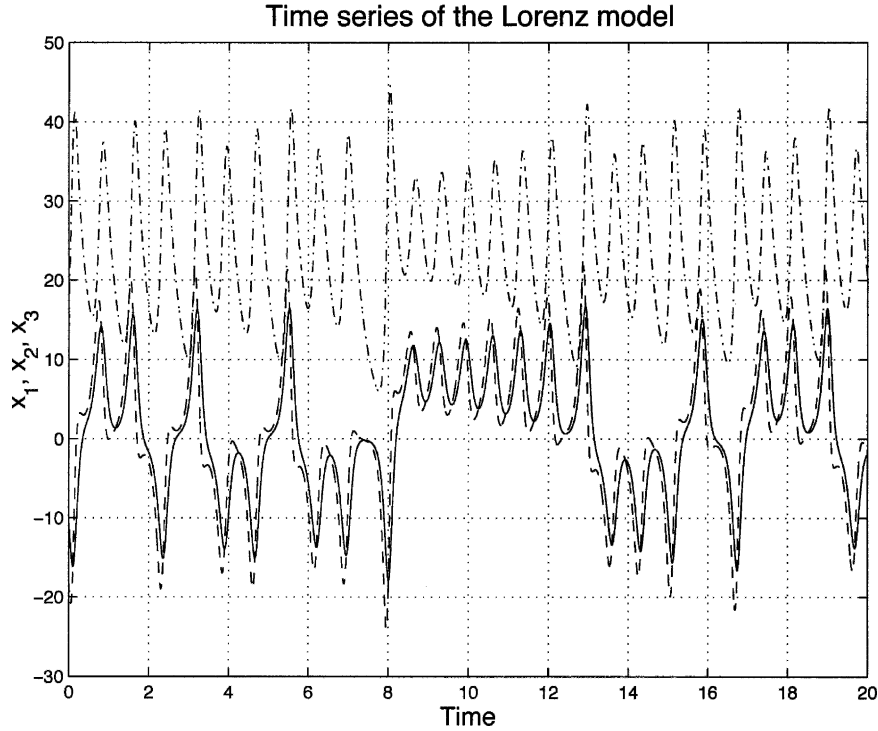
Time series of the Lorenz model

FIG. 1. Sample time series from the integration of the Lorenz model; solid line $x_1(t)$, dashed line $x_2(t)$, dash–dotted line $x_3(t)$.

2) REGRESSION MODEL: FORMULATION AND PERFORMANCE

We assume that $y_i = f_i(\mathbf{x})$ and that the right-hand sides f_i have the general quadratic form

$$f_i = a_{0,i} + \sum_{n=1}^{3} a_{n,i} x_n + \sum_{n=1}^{3} \sum_{p \geq n} \tilde{a}_{np,i} x_n x_p. \quad (9)$$

This requires us to estimate $N = 1 + 3 + 3(3 + 1)/2 = 10$ regression parameters, as well as their standard errors, by a linear least squares fit; see Eq. (A2) in the appendix. If the standard deviation σ_{r_i} of the residual time series

$$r_i \equiv y_i^j - f_i^j \quad (10)$$

is small compared to σ_{y_i}, the regression fit is considered to be successful.

The summary of the fits to the Lorenz model is presented in Table 1. We illustrate changes in the properties of the fit for the different pairs of $(T, \Delta t)$ by displaying the estimated values of the parameters s, r, b, along with their standard errors σ_s, σ_r, and σ_b. The parameter s is defined here as the coefficient in front of x_1 in Eq. (6a). Other parameters that we do not list, including those in front of the nonlinear terms, are estimated equally well. We also show in Table 1 the values of σ_{y_i} and σ_{r_i} whose comparison gives a measure of the goodness of fit. In general, for a given pair of $(T, \Delta t)$, $\sigma_{r_i} < \sigma_{y_i}$, meaning that in all the cases we present, the fit is statistically significant.

Inspection of Table 1 shows that the coefficients of the Lorenz model are recovered with a good accuracy by the quadratic regression fit for large enough T and small enough Δt. The fit is expected to give meaningful results provided the record is long enough to capture the internal periods of the system and the data points sample this variability with sufficient frequency. Other combinations than those shown in the table are possible, trading off the values of T against Δt, provided $T > T_{ch}$ and $\Delta t < T_{ch}/2^6$.

Of the two dependencies, the one on Δt appears to be more crucial: the shortest record with the finest resolution (0.992, 0.001) still recovers regression coefficients that are quite close to the true coefficients, while the longest record with the coarsest resolution (20, 0.016) produces mediocre results. The accuracy of the results for such a short record length $T \approx 1$, which barely exceeds $T_{ch} \approx 0.75$, is actually pretty surprising, given the very long records required to produce a stable probability density function (PDF) for the Lorenz model. The statistical significance of the result in this case also ensures its stability with respect to choice of sample.

TABLE 1. Quadratic regression fit of the Lorenz model (see text for details).

T	Δt	σ_{y1}	σ_{y2}	σ_{y3}	s	σ_s	r	σ_r	b	σ_b	σ_{r1}	σ_{r2}	σ_{r3}
	0.001	0.01	0.02	0.02	10.02	0.0001	28.02	0.0003	2.66	0.0002	0.001	0.02	0.01
	0.002	0.03	0.09	0.09	10.01	0.0006	28.04	0.002	2.67	0.001	0.005	0.03	0.01
20	0.004	0.14	0.35	0.35	9.98	0.004	28.10	0.009	2.69	0.007	0.02	0.08	0.05
	0.008	0.56	1.40	1.41	9.85	0.02	28.31	0.05	2.75	0.04	0.09	0.29	0.21
	0.016	2.30	5.78	5.81	9.36	0.12	28.88	0.31	2.87	0.23	0.36	1.08	0.88
	0.001	0.01	0.02	0.02	10.02	0.0002	28.02	0.0004	2.66	0.0003	0.001	0.02	0.01
	0.002	0.04	0.09	0.09	10.01	0.001	28.04	0.002	2.67	0.002	0.006	0.03	0.01
10	0.004	0.15	0.37	0.37	9.98	0.005	28.10	0.01	2.69	0.01	0.02	0.09	0.05
	0.008	.059	1.49	1.49	9.86	0.03	28.31	0.08	2.75	0.05	0.09	0.30	0.23
	0.016	2.41	6.13	6.16	9.38	0.18	28.91	0.45	2.87	0.30	0.38	1.12	0.94
	0.001	0.01	0.02	0.02	10.02	0.0003	28.04	0.0007	2.66	0.0007	0.001	0.02	0.01
	0.002	0.04	0.09	0.09	10.01	0.002	28.06	0.004	2.67	0.004	0.004	0.03	0.01
4.8	0.004	0.15	0.37	0.38	9.96	0.01	28.13	0.02	2.68	0.02	0.02	0.08	0.05
	0.008	0.61	1.50	1.52	9.80	0.05	28.40	0.13	2.72	0.13	0.07	0.29	0.21
	0.016	2.51	6.18	6.25	9.11	0.30	29.00	0.75	2.66	0.74	0.31	1.09	0.87
	0.001	0.01	0.02	0.02	10.02	0.0005	28.05	0.001	2.66	0.001	0.001	0.01	0.01
	0.002	0.04	0.10	0.10	10.00	0.003	28.06	0.007	2.67	0.008	0.003	0.02	0.01
2	0.004	0.16	0.39	0.39	9.95	0.02	28.12	0.04	2.68	0.04	0.01	0.06	0.05
	0.008	0.64	1.57	1.58	9.76	0.09	28.29	0.22	2.71	0.26	0.05	0.21	0.19
	0.016	2.62	6.44	6.53	9.00	0.53	28.60	1.30	2.60	1.48	0.21	0.75	0.80
	0.001	0.01	0.03	0.03	10.02	0.001	28.05	0.002	2.67	0.004	0.002	0.01	0.002
	0.002	0.04	0.11	0.11	10.00	0.006	28.07	0.01	2.66	0.02	0.01	0.01	0.01
0.992	0.004	0.18	0.44	0.44	9.95	0.03	28.16	0.08	2.62	0.13	0.004	0.05	0.04
	0.008	0.72	1.76	1.79	9.76	0.18	28.44	0.45	2.43	0.71	0.02	0.18	0.18
	0.016	2.97	7.24	7.40	8.94	1.08	29.04	2.64	1.27	4.08	0.12	0.68	0.70
	0.016*	2.97	7.24	7.40	8.93	1.02	29.11	2.49	2.96	0.45	0.12	0.69	0.81
	0.016**	2.97	7.24	7.40	8.92	1.02	29.10	2.49	2.94	0.44	0.12	0.69	0.80

* Assuming that $a_{0,i} = 0$ in Eq. (9).
** Using PCR.

3) THE PROBLEM OF COLLINEARITY AND REGULARIZATION

Even though the estimated standard errors of the regression parameters are large at coarse sampling rates, their expected values turn out to be close to the true values of the parameters, except for the case of the shortest record and coarsest resolution $(T, \Delta t) = (0.992, 0.016)$, where the regression fit fails to reproduce a correct value of $b = 8/3$ even approximately, giving $b = 1.27 \pm 4.08$. Closer inspection of the fit shows that this happens due to very large errors in determining $a_{0,i}$, which introduces large linear trends in the regression fit when T is dangerously close to T_{ch}. Reducing the number of regression coefficients to 9 by setting $a_{0,i} = 0$ results in a much more acceptable value, $b = 2.96 \pm 0.45$, as shown in the row marked by an asterisk in Table 1. Note that the standard error has been reduced considerably as well. The standard error for the parameters s and r is also reduced, although the reduction is not as drastic as that for b. The large reduction of σ_b, as well as the smaller reductions in σ_s and σ_r, occur at the expense of a small increase in σ_{r_i}.

This behavior is due to the phenomenon of collinearity, or ill conditioning, in which the vectors of predictor variables are close to linear dependence (Wetherill 1986). Very large errors in determining the coefficient b in the above example are due to a common linear trend in the short time series of model variables. Discarding $a_{0,i}$ eliminated this trend from consideration, and resulted in an improved parameter estimation.

An objective way of screening out the coefficients that do not contribute much to improving the accuracy of the regression fit is known as principal component regression (PCR; see the appendix). This approach relies on the singular value decomposition (SVD) of the so-called design matrix, whose N columns are the time series of the N predictor variables. The idea behind editing the regression fit coefficients is to define the reciprocal of the design matrix's singular value w_n to be zero once w_n is small enough; a small w_n means that the associated singular vector does not contribute much to the reduction of σ_{r_i} (Press et al. 1994). Setting to zero all singular values that are smaller than $W = \epsilon \max\{w_n|_{n=1}^{N}\}$ turns out, in the problem at hand, to be equivalent to discarding the $a_{0,i}$, as can be seen from the results presented in the last row of Table 1; ε is a preset tolerance, taken here as $\varepsilon = 0.001$.

Removing the sensitivity of results to small changes in the data, as in the case of collinearity, is called regularization. Using regularization techniques such as the PCR is essential in most geophysically relevant examples (section 4), in which a large number of coefficients needs to be estimated from a limited amount of data. In the latter situations, optimal regularization is achieved via the so-called partial least squares (PLS) procedure. PLS is analogous to PCR, but is based on computing those linear combinations of the predictor variables that are well correlated with the predictands, while accounting for a large amount of variation in the predictor field itself; the procedure is outlined in the appendix.

b. The double-well potential

In the example in the previous subsection, we have assumed that the standard errors of the "observed" time derivatives (i.e., of our response variables) were known, due to the deterministic nature of the Lorenz model. We next consider the one-dimensional (1D) double-well potential problem with stochastic forcing (Ghil and Childress 1987, their section 10.3; Miller et al. 1994) as a simple example of a system with bimodal variability, where the deterministic low-frequency signal is weak compared to the noise. This low signal-to-noise ratio characterizes many geophysical applications. In performing the regression, one needs to find, therefore, not only the coefficients of the deterministic part of the equations, but also the major characteristics of the stochastic forcing. This task leads us to introduce the cross-validation method, in which the regression analysis is performed on subsamples of the full time series available.

The deterministic part of the dynamics in the double well is described by

$$\dot{x} = -\frac{dF}{dx} \equiv -4x(x^2 - 1), \qquad (11)$$

where

$$F(x) = x^2(x^2 - 2) \qquad (12)$$

is the potential function with wells that appear at $x = 1$ and $x = -1$, and a relative maximum at $x = 0$. Thus $x = 1$ and $x = -1$ are the stable equilibria of Eq. (11), while $x = 0$ is an unstable one. When stochastic forcing is applied, the trajectory will move from the basin of attraction of one stable equilibrium to the other, provided the amplitude of the forcing or the waiting time is large enough. The stochastically forced double-well system is described by

$$dx = -\frac{dF}{dx} dt + \sigma \, db, \qquad (13)$$

where b is a Wiener process whose independent increments have mean zero and unit variance.

We numerically integrate a discretized version of Eq. (13)

$$\frac{y^j}{\Delta t} \equiv \frac{x^{j+1} - x^j}{\Delta t} = -4x^j[(x^j)^2 - 1] + \frac{1}{\Delta t} \sigma \Delta b, \quad (14)$$

with $\Delta t = 0.01$ and $\sigma = 0.5$ (Miller et al. 1994) for $T_* = 3000$ time units. A sample record of length $T = 50$ from the resulting time series is shown in Fig. 2. The amplitude of the stochastic forcing is large enough for the system to irregularly switch from one stable equilibrium to the other on a time scale comparable to or shorter than the characteristic time of the deterministic dynamics. Moreover, the trajectory often dwells for quite some time in the vicinity of the unstable fixed point $x = 0$ before falling into one of the two stable wells (Miller et al. 1994).

We now fit a cubic polynomial in x to the response variable y:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3. \qquad (15)$$

Numerical experimentation shows that a reasonably good fit can be obtained with sample records as short as $T = 50$, although the regression coefficients are subject to large sampling variation that will be quantified below. Our goal is, once again, to estimate both the regression coefficients $\{a_n\}|_{n=0}^3$ and the amplitude of the stochastic forcing σ_r. If the data record is long enough, a way to do this is to divide the time series into shorter samples and find the regression coefficients, their standard errors, and the residual-forcing amplitude for each sample. The expected value of each of these quantities can then be found as an average over all samples, while their respective errors will be the standard deviations of each quantity from its expected value over all samples.

The results of such a regression fit are presented as experiment 1 in Table 2. We have divided a time series of total length $T_* = 3000$ units into segments of length $T = 50$ that overlap by half of their length to increase the number of samples.

The first of the two rows that summarize experiment 1 results contains the ensemble mean coefficients and their (ensemble mean) standard errors (see the appendix) from the multisample regression fit, while the spread of these values over the whole ensemble is presented as standard deviations in the second row. The errors in determining the regression parameters for each sample are computed a posteriori, by using the estimated σ_r for this sample. Both the regression coef-
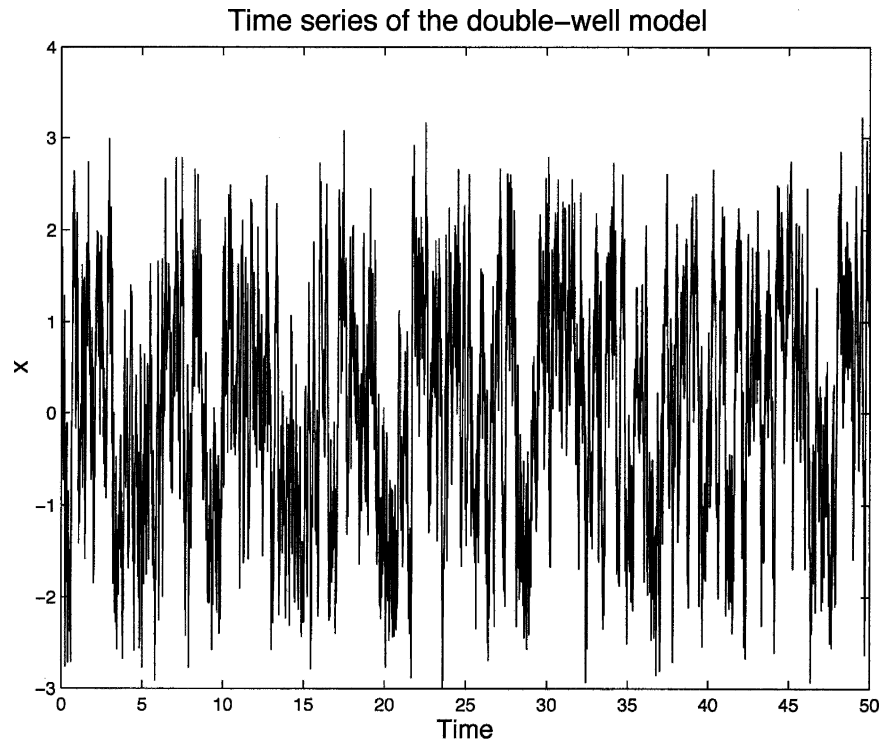
Time series of the double-well model

FIG. 2. Sample time series from the integration of the double-well stochastic model.

ficients and the noise amplitude are reconstructed well by the fit. All of the quantities σ_r and $\sigma_{a_n}|_{n=0}^{3}$ change very little from sample to sample (see second row). In contrast to their standard errors, the values of the coefficients vary considerably from sample to sample, with an associated standard error that is larger than the estimated standard error of each sample. For example, the value of 1.11 in the first column, second row of the table should be compared with $\sigma_{a_1} = 0.75$ of experiment 1.

Since we have verified that σ_b is estimated with a good accuracy $\sigma_b = 0.5 \pm 0.005$, it is possible to narrow the error bars on model coefficients by using the entire time series to estimate regression coefficients. The results are shown as experiment 2 of Table 2, and show overall improvement in estimating the reconstructed

coefficients, as well as tight error bars. Experiments 3 and 4 correspond to experiments 1 and 2, but with the coefficients a_0 and a_2 set to zero, as they are in the true model, given by Eqs. (12) and (13). The quality of the fit is similar to that of the full regression model, although experiment 3 gives the best estimates of the true model parameters.

In the procedure above, we have stopped one step short of the full cross-validation procedure, in which the available dataset is generally split into two parts: the training set, which is used to construct the regression model, and the validation set, whose statistical properties are compared with those of the regression model. The splitting can be performed in several different ways and the results of the statistical comparison between the actual and regression-generated data are ensemble

TABLE 2. Regression fit of the double-well stochastic model (see text for details).

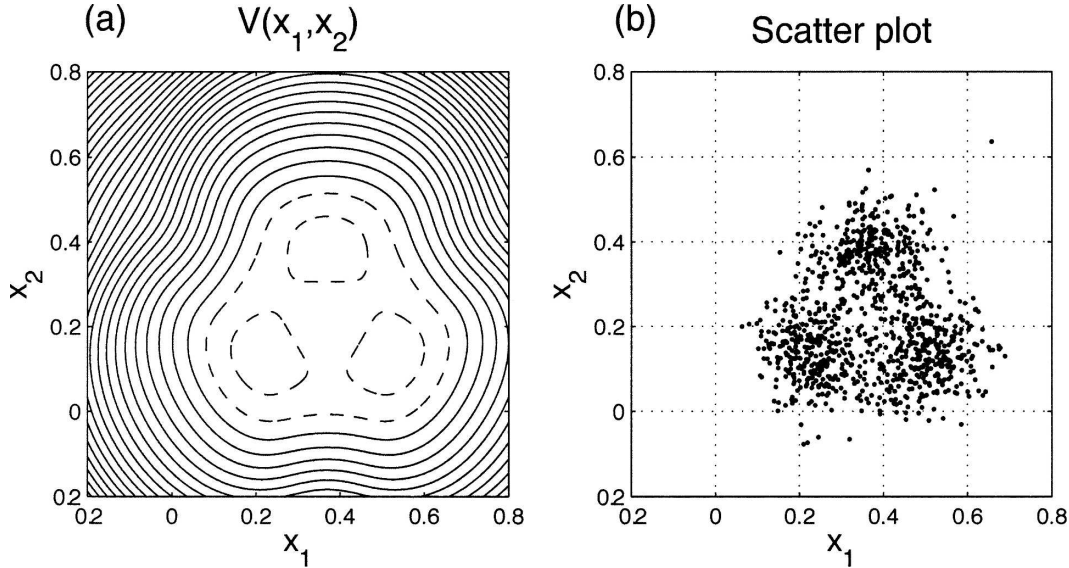| Expt No.		a_0	σ_{a_0}	a_1	σ_{a_1}	a_2	σ_{a_2}	A_3	σ_{a_3}	σ_b
1	qty	−0.14	0.75	3.94	0.85	0.1	0.34	−4.01	0.22	0.5
	σ_{qty}	1.11	0.01	1.08	0.01	0.49	0.01	0.27	0.01	0.005
2		−0.07	0.117	4.07	0.132	0.07	0.05	−4.02	0.03	0.5
3	qty	0	0	3.99	0.84	0	0	−4.01	0.22	0.5
	σ_{qty}	0	0	1.08	0.01	0	0	0.27	0.01	0.005
4		0	0	4.07	0.132	0	0	−4.02	0.03	0.5

FIG. 3. Triple-well model. (a) Contour plot of the potential function; contour interval $CI = 0.003$, with negative contours dashed. (b) Scatterplot of a 50 000-unit-long integration; every 50th point is plotted.

averaged over these different ways. This approach is routinely applied to determine the number of factors used in PCR and PLS (see the appendix). The statistical measure one uses for comparing the observed and model time series is application dependent. An example of comparing the true and simulated statistical distributions of the model variables will be given in the following subsection.

c. A triple-well system

It is often speculated that large-scale atmospheric flows, or global climate variability, bear some resemblance to the behavior of a stochastically driven particle moving between several potential wells (Hasselmann 1999). Hannachi and O'Neill (2001) have thus considered a two-dimensional (2D) generalization of the double-well problem. The stochastic system in this case is governed by

$$d\mathbf{x}(t) = -\nabla V(\mathbf{x})\,dt + \sigma\,d\mathbf{b}, \quad (16)$$

where $\mathbf{b} = (b_1, b_2)$ is a random walk, and the potential function V is given by

$$
\begin{aligned}
V(x_1, x_2) = &-v(x_1^2 + x_2^2) - v\{(x_1 - 2a)^2 + x_2^2\} \\
&- v\{(x_1 - a)^2 + (x_2 - a\sqrt{3})^2\} \\
&+ b[(x_1 - a)^2 + (x_2 - a/\sqrt{3})^2], \quad (17)
\end{aligned}
$$

with $b = 0.12$ and $a = 0.37$. The well shape is defined by

$$v(x) = \alpha \exp\left(\frac{1}{x^2 - a^2}\right) \quad \text{if} \quad |x| < a,$$

$$\text{and} \quad 0 \quad \text{otherwise}, \quad (18)$$

with the parameter $a = 21$. The resulting triple-well structure of V is shown in Fig. 3a. If the value of σ is chosen to be 0.05 and dt is replaced by $\Delta t = 1$, the model produces low-frequency behavior that consists of jumps between the three potential wells, as shown in Fig. 3b for the results of a sample integration of Eqs. (16)–(18) having length $T_* = 50000$.

We now fit to this dataset the general 2D polynomial regression model, while changing the order m of the polynomial $P = P_m(x_1, x_2)$. For $m = 1$, $P_1(x_1, x_2) = ax_1 + bx_2 + c$ has three coefficients, the second-order polynomial P_2 will involve the same linear part (three coefficients), as well as a quadratic part $dx_1^2 + ex_2^2 + fx_1x_2$, the cubic polynomial P_3 will have an additional cubic part $gx_1^3 + hx_2^3 + kx_1^2x_2 + lx_1x_2^2$, and so on. The true underlying function that generated the data is not polynomial in this case, since $V(x_1, x_2)$ is based on exponential functions. We therefore assess the goodness of fit by simulating the data using the regression model we have constructed in each case, and comparing the 2D PDFs of the true and simulated data. The results of such a comparison are presented in Fig. 4.

A Gaussian mixture model is used to estimate each PDF as a sum of k Gaussians (Hannachi 1997; Smyth et al. 1999). The natural choice k for the triple well is $k = 3$. In addition to the contour plot of the PDF, we plot the estimated centroids and covariance ellipses for all
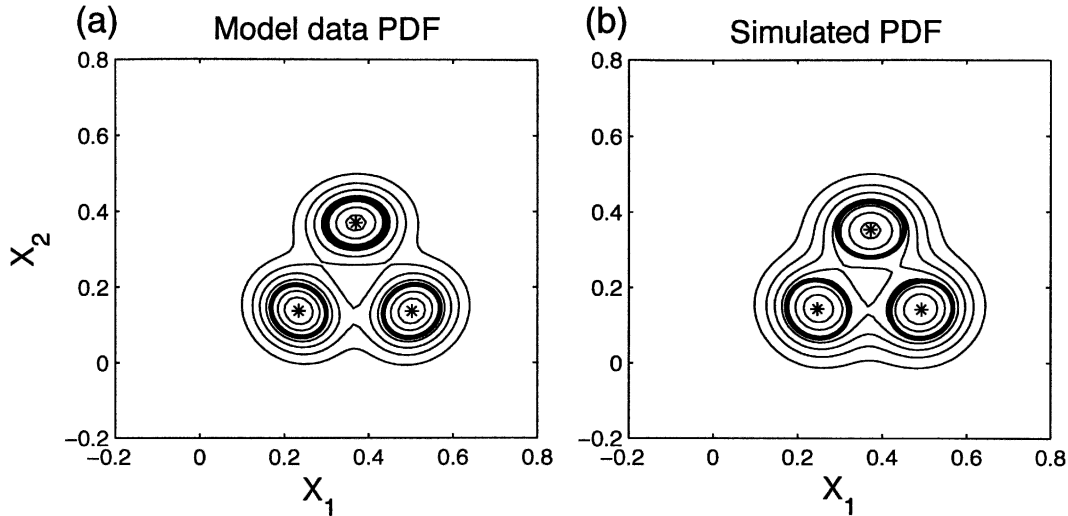
FIG. 4. Mixture model PDFs and clusters for the triple-well model. Cluster centroids are plotted as asterisks, while the cluster boundaries are shown as projections of corresponding ellipses. (a) PDF of the data generated by the triple-well model; (b) PDF of the data simulated by the inverse stochastic model based on fitting a cubic polynomial of the predictor variables to the data.

three clusters. The results for the data shown in Fig. 3b are plotted in Fig. 4a. The mixture model captures the structure of the potential function (Fig. 3a) fairly well, with cluster centroids that are visually indistinguishable from the potential-well bottoms. In Fig. 4b, we plot the analogous results for the data simulated by our regression model with a polynomial $P_m(x_1, x_2)$ of the order $m = 3$. The regression model simulation reproduces very well the triple-well structure of the true data.

Similar results are obtained by using higher-order regression polynomials. We found, however, no significant improvement of the fit's accuracy over that of the cubic regression model, as shown in Table 3.

As expected, the regression model is often numerically unstable when high-order polynomials are used, due to the collinearity problem that arises in determin-

ing its coefficients. We have thus applied PCR for $m \geq 4$, as described in section 2a and the appendix. For $m = 3$, the number N of regression coefficients is $N = 10$, and thus no editing is necessary, while the tolerance ε has to be increased to 5×10^{-4} for $m = 11$, $N = 78$. In spite of the editing, the standard deviation of the estimated residual stochastic forcing does not tend to the true value of $\sigma = 0.05$ with increasing m, although it does stay very close to this value. The distance between the true and estimated centroids Δd also remains small for all cases, although it increases slightly with m. The fit with $m = 5$ produces marginally better results than that with $m = 3$. In general, the parsimony principle requires choosing the lowest-order fit, which has the most stable coefficients (Smyth et al. 1999; Hand et al. 2001); accordingly, the third-order fit is the optimal one here.

d. Summary

The examples presented in this section show that inverse stochastic modeling that uses a polynomial right-hand side to fit the data generated by a nonlinear process can be quite effective, provided the dataset is well sampled. The latter requirement is often satisfied to a reasonable extent in model-generated datasets and in observed data. Both types of data are typically characterized, however, by a large number of degrees of freedom, so that the direct application of MPR is not feasible and regularization techniques such as the PCR and PLS should be used (see sections 2a, 2c, and the appendix).

TABLE 3. Triple-well model fit; N is the number of coefficients in the polynomial used for regression fitting and ε the tolerance used in PCR to edit the singular values of the design matrix (see text for details). The last three columns list the standard deviations of the residual forcing σ_{r_1} and σ_{r_2}, as well as the sum Δd of the distances between the true and estimated cluster centroids.

Expt No.	FUNC	N	ε	σ_{r_1}	σ_{r_2}	Δd
0	$\nabla V(x_1, x_2)$	—	—	0.0500	0.0500	0
1	P_3	10	0	0.0516	0.0516	0.0137
2	P_5	21	0.5×10^{-4}	0.0513	0.0511	0.0129
3	P_7	36	10^{-4}	0.0513	0.0510	0.0132
4	P_9	55	4×10^{-4}	0.0512	0.0511	0.0148
5	P_{11}	78	5×10^{-4}	0.0513	0.0512	0.0160

The next section develops a general strategy for nonlinear inverse stochastic modeling of multivariate datasets.

3. Multilevel inverse models

In this section and the next one, we construct inverse stochastic models in the phase space of the leading EOFs of the fields considered. The quadratic model that we will use below has the general form

$$dx_i = (\mathbf{x}^T\mathbf{A}_i\mathbf{x} + \mathbf{b}_i^{(0)}\mathbf{x} + c_i^{(0)})\, dt + dr_i^{(0)}; \qquad 1 \le i \le I,$$
(19)

where $\mathbf{x} = \{x_i\}$ is the state vector of dimension I. The matrices $\mathbf{A}_i$, the rows $\mathbf{b}_i^{(0)}$ of the matrix $\mathbf{B}^{(0)}$, and the components $c_i^{(0)}$ of the vector $\mathbf{c}^{(0)}$, as well as the components $r_i^{(0)}$ of the residual forcing $\mathbf{r}^{(0)}$, are determined by least squares. If the inverse model contains a large number of variables, the statistical distribution of $\mathbf{r}^{(0)}$ at a given instant is nearly Gaussian, according to the central limit theorem (Von Mises 1964).

However, the stochastic forcing $\mathbf{r}^{(0)}$ in Eq. (19) typically involves serial correlations and might also depend on the modeled process $\mathbf{x}$. We include, therefore, an additional model level to express the time increments $d\mathbf{r}^{(0)}$ [equivalent, in numerical practice, to the time derivative of the residual forcing $\mathbf{r}^{(0)}$] as a linear function of an extended state vector $[\mathbf{x}, \mathbf{r}^{(0)}] \equiv [\mathbf{x}^T, \mathbf{r}^{(0)T}]^T$, and estimate this level's residual forcing $\mathbf{r}^{(1)}$. The linear dependence is used since the non-Gaussian statistics of the data have already been captured by the first nonlinear level. More (linear) levels are being added in the same way, until the $(L + 1)$th level's residual $\mathbf{r}^{(L+1)}$ becomes white in time, and its lag-0 correlation matrix converges to a constant matrix:

$$dx_i = (\mathbf{x}^T\mathbf{A}_i\mathbf{x} + \mathbf{b}_i^{(0)}\mathbf{x} + c_i^{(0)})\, dt + r_i^{(0)}\, dt,$$

$$dr_i^{(0)} = \mathbf{b}_i^{(1)}[\mathbf{x}, \mathbf{r}^{(0)}]\, dt + r_i^{(1)}\, dt,$$

$$dr_i^{(1)} = \mathbf{b}_i^{(2)}[\mathbf{x}, \mathbf{r}^{(0)}, \mathbf{r}^{(1)}]\, dt + r_i^{(2)}\, dt,$$

$$\cdots$$

$$dr_i^{(L)} = \mathbf{b}_i^{(L)}[\mathbf{x}, \mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(L)}]\, dt$$

$$+ dr_i^{(L+1)}; \qquad 1 \le i \le I.$$
(20)

The convergence of this procedure is guaranteed since, with each additional level $l \ge 1$, we are accounting for additional time-lag information, thereby squeezing out any time correlations from the residual forcing. Section 4a formulates a simple and convenient convergence criterion.

In practice, we approximate the increments dx_i, $dr_i^{(l)}$ as

$$dx_i = x_i^{j+1} - x_i^j, \qquad dr_i^{(l)} = r_i^{(l),j+1} - r_i^{(l),j}, \qquad 1 \le l \le L,$$
(21)

where j is the time index, while dt is assumed to be equal to the dataset's sampling interval; without loss of generality, we use $dt = 1$. The last-level residual's $dr_i^{(L+1)}$ covariance matrix is estimated directly from its multivariate time series; in subsequent integrations of the inverse model, this forcing is approximated as a spatially correlated white noise.

One can in principle rewrite the multilevel system in (20) as a single equation that involves time-lagged values of x_i and $r_i^{(l)}$; the resulting construct is equivalent to a multivariate version of autoregressive-moving average (ARMA) model (Box et al. 1994), except for the nonlinear dependence on x_i that we allow here, and that is not present in standard ARMA models. Even for a standard, linear model, though, the way we estimate the coefficients of this model by successive introduction of additional levels is novel; the main advantages of our method are its algorithmic simplicity, numerical efficiency, and dynamical interpretability.

The system in (20) describes a wide class of nonlinear, non-Gaussian processes in a fashion that explicitly accounts for the modeled process $\mathbf{x}$ feeding back on the noise statistics. A multilevel representation similar to Eq. (20) above has been used by Berloff and McWilliams (2002) in a slightly different form and context to model tracer paths in a numerical model of wind-driven ocean gyres [see their Eqs. (3), (15), (36), and (53)]. DelSole (1996, 2000) considered a special, linear and autoregressive case of the system in (20) in order to investigate the suitability of Markov models for representing quasigeostrophic turbulence.

The optimal number of state-vector components in Eq. (20) is assessed in practice using Monte Carlo simulations (see section 2b and the appendix): in these cross-validation tests, the inverse model is trained on one segment of the available data and is then used to estimate the properties of the model evolution during the validation interval. The measure used to assess the statistical model's performance depends on the purpose at hand: If the model is to be used for prediction, the forecast skill, quantified by the correlation between the forecast and observed fields or the root-mean-square (rms) distance between the two is an appropriate measure of model performance; in the more theoretical applications below, it is the statistical characteristics of the observed and modeled evolution, such as PDFs of

model variables (see section 2c) and their power spectra.

4. Geophysical examples

a. Analysis of an atmospheric model

1) CONSTRUCTION OF THE MULTILEVEL INVERSE MODEL

We analyze first a 54 000-day-long simulation of the three-layer QG (QG3) model of Marshall and Molteni (1993) on the sphere; this particular simulation was carried out at a T21 spatial resolution, sampled once a day, and analyzed by Kondrashov et al. (2004). The model's low-frequency variability is equivalent barotropic. Our inverse stochastic models are constructed, therefore, in the phase space of the leading EOFs of the middle-layer streamfunction, and we use quadratic regression.

The governing equations of the QG3 model—like those of many purely fluid dynamical models of atmospheric and oceanic flows (Lorenz 1963a,b; Ghil and Childress 1987, their section 5.4; Dijkstra and Ghil 2005)—have only quadratic nonlinearities. Still, the mode-reduction strategy of Majda et al. (1999, 2001, 2002, 2003) argues for the presence of cubic terms in the optimal reduced model equations. Franzke et al. (2005) have shown that these cubic terms introduce only slight modifications into the barotropic reduced model dynamics, but C. Franzke and A. Majda (2005, personal communication) have found them to be much more important for the QG3 model's dynamics. Our choice of a quadratic model is dictated by the trade-off between the number of regression coefficients to be estimated and the expected, weakly nonlinear dynamics of the QG3 model's behavior. In fact, we expect the quadratic model to be quite adequate for a wide class of geophysical problems, including those in which the inverse modeling is based on observational data alone, rather than on mode reduction for a known dynamic model (see section 4b here and Kondrashov et al. 2005).

Our inverse stochastic model has 15 primary variables, that is $\mathbf{x}$ in Eqs. (19) and (20) has 15 components. It can be verified directly that the statistical distribution of the first-level residual in this model is indistinguishable from Gaussian, so that this level adequately captured the nonlinearities present in the QG3 model.

Once we have chosen the number of primary variables, we still need to determine the number of additional linear levels in (20). To do so, we computed lag-0 and lag-1 covariance matrices of the residual forcing at each level and found that the third-level residual forcing is indeed white (i.e., its lag-1 covariance vanishes)
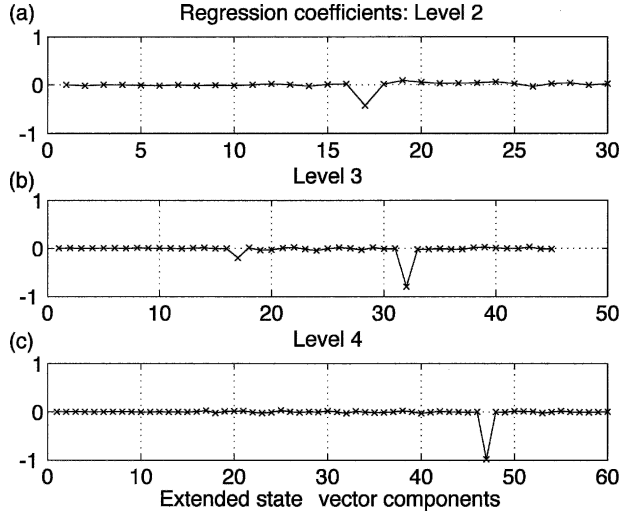

FIG. 5. Multiple linear regression coefficients for the first components $dr_1^{(0)}$, $dr_1^{(1)}$, and $dr_1^{(2)}$ of the second, third, and fourth level of the inverse model given by Eq. (20) and constructed in the phase space spanned by the 15 leading EOFs of the long QG3 model integration (see text): (a) level-2 (30 coefficients); (b) level-3 (45 coefficients); and (c) level-4 (60 coefficients). In (a), (b), and (c), the largest negative coefficient corresponds to the one multiplying $r_1^{(0)}$, $r_1^{(1)}$, and $r_1^{(2)}$ in Eq. (20), respectively.

and that its spatial covariance does not change when adding more levels. This means that if we estimate the coefficients of the next, fourth-level equation for each $r_i^{(3)}$ ($1 \leq i \leq 15$), we get, to a very good approximation,

$$dr_i^{(3)} \equiv r_i^{(3),j+1} - r_i^{(3),j} = -r_i^{(3),j} + r_i^{(4),j}, \qquad (22)$$

where $dt = 1$ and j is the time index: the only nontrivial regression coefficient that multiples $r_i^{(3)}$ in the equation for $dr_i^{(3)}$ should thus be equal to -1. In this case, $r_i^{(4),j} = r_i^{(3),j+1}$, so that the residual $r_i^{(4)}$, which is uncorrelated with $r_i^{(3)}$ by construction, is a 1-day-lagged copy of $r_i^{(3)}$; it thus follows that the lag-1 autocorrelation of both $r_i^{(3)}$ and $r_i^{(4)}$ vanishes, while the third- and fourth-level residual's lag-0 covariance matrices are identical.

The criterion above is thus a simple and convenient way to determine the number of levels in a multilevel regression model: the procedure is stopped when at some level L the coefficient that multiplies each $r_i^{(L-1)}$ is close to -1 and all the other coefficients are close to zero. Figure 5 illustrates this convergence for the first-variable residuals $r_1^{(1)}$, $r_1^{(2)}$, and $r_1^{(3)}$ at the second third and fourth levels of the 15-variable QG3 fit, respectively. Similar convergence is achieved for all of the other 14 variables (not shown).

The total number of variables in our inverse model is, therefore, 45 (15 variables at each of the three levels). By comparison, each of the three levels of the QG3 model in the Northern Hemisphere has 1024 state-

vector variables. Although the number of coefficients N that need to be estimated in the inverse stochastic model, $N = \{[15 \times (15 + 1)/2 + 1 + 15] + 30 + 45\} \times 15 = 3165$, is large, this model is much more numerically efficient than the QG3 model: the regression coefficients are determined offline, once and for all, while the dimension of the state vector is almost an order of magnitude smaller in the former model. The QG3-generated time series of the leading PCs are compared below with those produced by a 54 000-day-long integration of the inverse stochastic model; we also test the sensitivity of the results to the number of variables and of levels.

2) PERFORMANCE OF THE MULTILEVEL INVERSE MODEL

The Gaussian-mixture PDFs of the datasets produced by the QG3 and the inverse model are compared in Fig. 6 in the subspace of the QG3 model's three leading EOFs. The clusters were found using mixtures of $k = 4$ Gaussian components in a phase subspace of four leading EOFs, which capture 29% of the total variance. The optimal number of clusters is $k = 4$ for both the QG3 simulation and for the dataset generated by the inverse model, as determined by the cross-validation procedure of Smyth et al. (1999); see also Kondrashov et al. (2004). The locations, shapes and sizes of clusters, and hence the general shape of the PDF, are well reproduced by the inverse model simulation in Fig. 6.

The composites over the data points that belong to each of the ellipses in Fig. 6 represent, in physical space, the patterns of four planetary flow regimes (Legras and Ghil 1985; Ghil and Childress 1987, see chapter 6; Mo and Ghil 1988; Cheng and Wallace 1993; Kimoto and Ghil 1993a,b; Hannachi 1997; Smyth et al. 1999; Hannachi and O'Neill 2001; Molteni 2002). In Fig. 6a, cluster AO$^-$ occupies a distinctive region on the PDF ridge that stretches along PC-1. It corresponds to the low-index phase of the well-known Arctic Oscillation (AO; Deser 2000; Wallace 2000). The clusters AO$^+$, (North Atlantic Oscillation) NAO$^-$, and NAO$^+$ are located around the global PDF maximum, with the centroid of AO$^+$ to the left and below, NAO$^+$ above, and NAO$^-$ slightly to the right of this maximum, respectively. These four regimes are not identical to but in fairly good agreement with the observational results of Cheng and Wallace (1993) and Smyth et al. (1999); see also Ghil and Robertson (2000) and Kondrashov et al. (2004).

The streamfunction anomalies associated with each regime centroid of the QG3 model are plotted in Fig. 7. The spatial correlations between these anomaly pat-

terns and those obtained from the inverse model (not shown) all exceed 0.9. They are thus much higher than the correlations obtained by D'Andrea and Vautard (2001) and D'Andrea (2002), who used a reduced deterministic model obtained by a statistical–dynamical approach to reproduce the behavior of the largest scales in the QG3 model.

Given that our inverse model captures well the location of the QG3 model clusters, the former can be used to relate these statistical features to the dynamical properties of the latter. Preliminary results indicate, in particular, that the AO$^-$ cluster corresponds to a steady state of the unperturbed, deterministic part of the inverse model (not shown).

To examine how well the inverse model reproduces the low-frequency variability of the QG3 model, we applied singular spectrum analysis (SSA; Vautard and Ghil 1989; Dettinger et al. 1995; Ghil et al. 2002) to the time series produced by the two models, as shown in Fig. 8. No significant pairs stand out in the QG or the inverse model spectra. The general shape of the QG3 spectrum is reproduced quite well by the inverse model simulation.

3) COMPARISON WITH ONE-LEVEL INVERSE MODEL

In contrast to the multilevel inverse model, the spectra from the one-level inverse model simulation are statistically indistinguishable from a red spectrum (not shown). Despite this, the one-level fit turns out to be sufficient to reproduce the clusters in Fig. 6, as well as the general shape of the PDF there (not shown). The cluster locations in this case are, however, sensitive to the particular realization of the white noise forcing and change significantly from sample to sample; moreover, some realizations of a given stochastic model, with coefficients obtained by a one-level regression, are unstable, in the sense that their trajectories blow up in finite time. The cluster locations obtained with the three-level MPR model are not sensitive to the choice of the random forcing sample and the inverse model simulations never result in unphysically large values of its variables.

The non-Gaussian shape of the PDF in Fig. 6, and hence the presence of multiple regimes, is due to nonlinearities in the main level of the inverse model in (20), while the statistics of the regimes is sensitive to the way the residual $\mathbf{r}^{(0)}$ depends on the large-scale flow $\mathbf{x}$. Molteni (2002) has reviewed semiempirical parameterizations of such an eddy feedback based on long libraries of residual tendencies, estimated as a running mismatch between an assumed large-scale model and a detailed model or observational data (Vautard and Legras
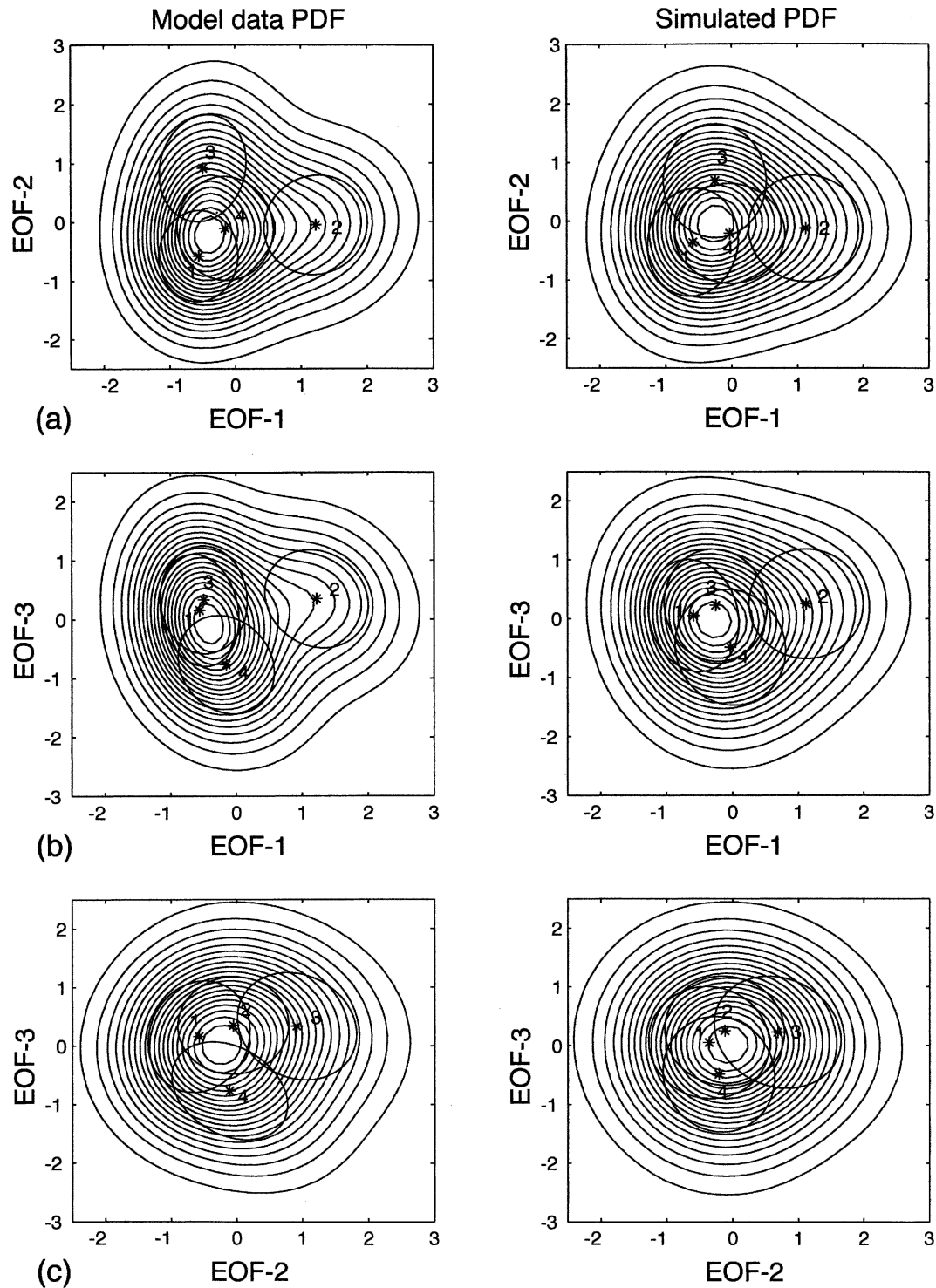
FIG. 6. Mixture model PDFs and clusters for the (left) QG3 model and (right) an inverse stochastic model. Same symbols and conventions as in Fig. 4. (a), (b), (c) Projections onto pairs of EOFs, as indicated on the axes. The cluster centoid indices correspond to $1 - AO^+$, $2 - AO^-$, $3 - NAO^+$, and $4 - NAO^-$ (see also Fig. 7).
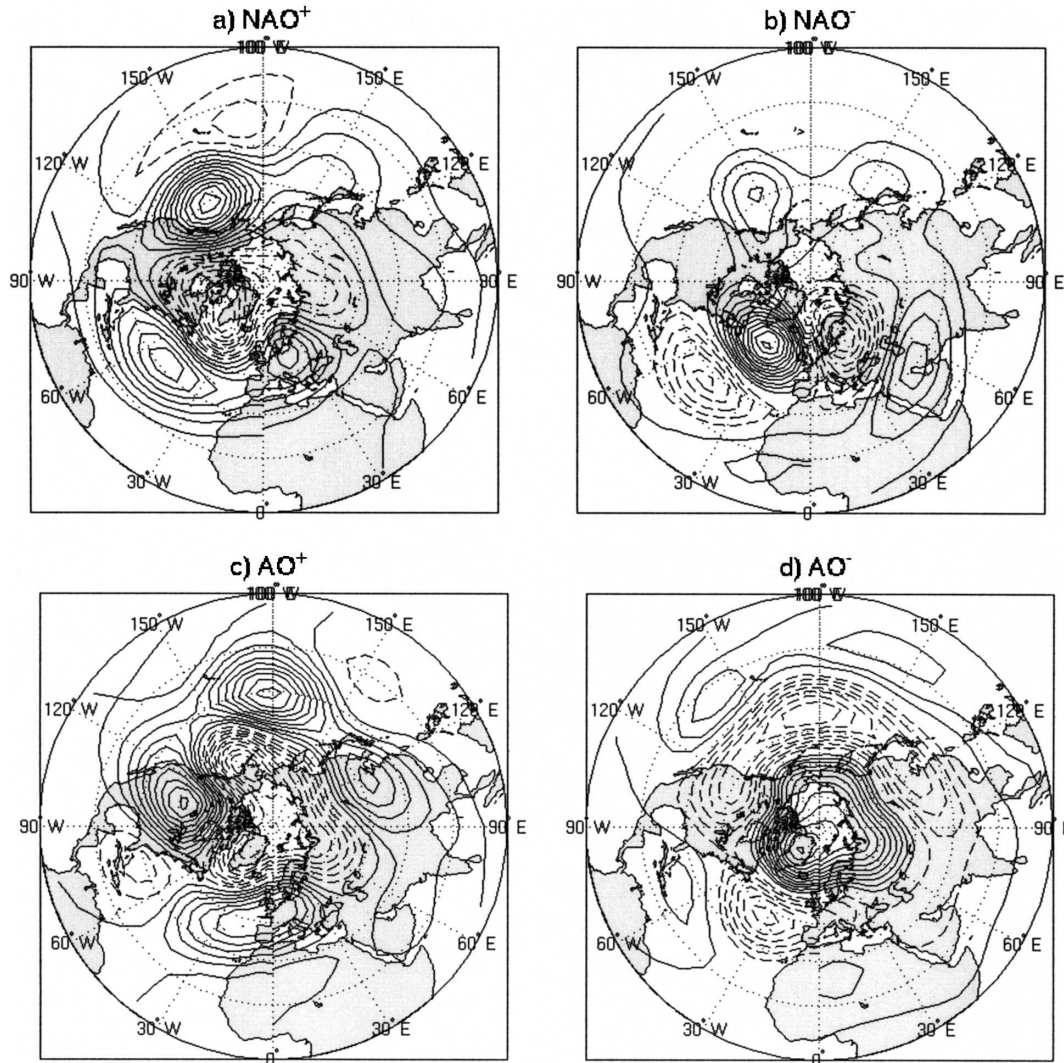
FIG. 7. Mixture model centroids, showing streamfunction anomaly maps at 500 hPa, for the QG3 model: (a) NAO$^+$; (b) NAO$^-$; (c) AO$^+$; and (d) AO$^-$. Negative contours are dashed and landmasses are shaded; 20 contour levels between maximum and minimum values are used, with the following intervals (in 10^6 m^2 s^{-1}): (a) 1.1; (b) 0.8; (c) 0.8; and (d) 1.1. Reproduced from Kondrashov et al. (2004), with the permission of the AMS.

1988; Molteni 1996a,b; Da Costa and Vautard 1997). D'Andrea and Vautard (2001) and D'Andrea (2002) have recently applied this technique to a truncated version of the QG3 model. The main disadvantage of this approach is that the reduced model formulation with the eddy feedback is not analytical, and is also fairly difficult to analyze numerically. In contrast, our purely empirical modeling method provides a simple and efficient way to quantify and analyze the eddy feedbacks.

4) SENSITIVITY TO THE NUMBER OF MODEL VARIABLES

Much of the QG3 model behavior above can be modeled with as few as 4 predictor variables (not shown)

and the results for 10 variables (not shown either) are very similar to those for 15 variables (Figs. 6–8). However, when using fewer predictor variables (such as 4 or 10), the location of cluster centroids is much more sensitive to the particular random forcing samples, that is, two different realizations of the inverse model simulation can be characterized by one or more clusters being significantly shifted from their position in the QG3 model simulation. No such sensitivity is observed in either the QG3 model or the 15-variable inverse model. The choice of 15 predictor variables seems to be optimal: the inverse model fit to the data becomes worse again for 20 variables (not shown).

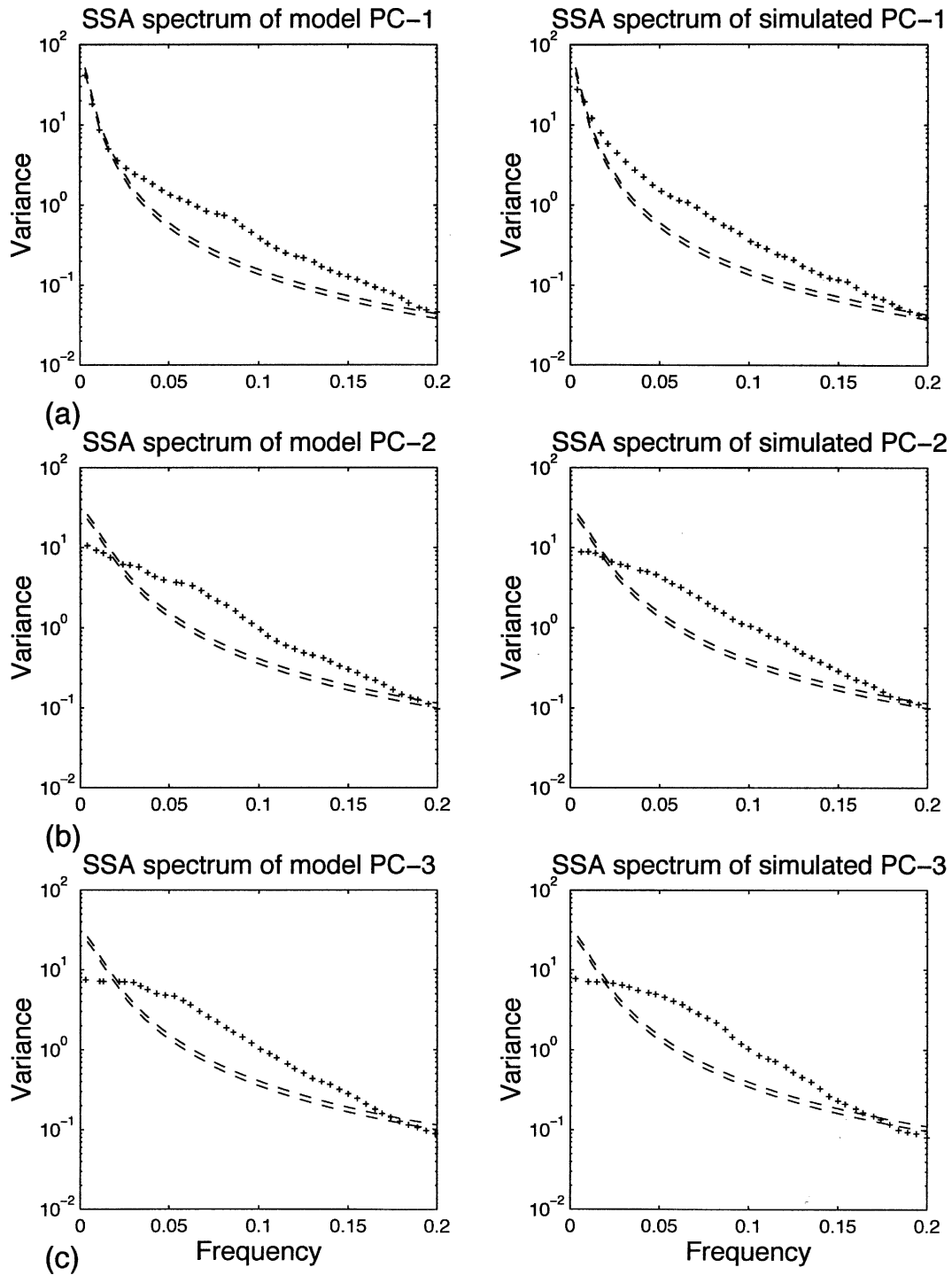The optimality of the fit for an intermediate, and

FIG. 8. Singular spectrum of the PCs used in Fig. 6: (left) QG3 model and (right) an inverse stochastic model. The size of plus signs that are used to plot the spectrum roughly corresponds to the size of the ad hoc error bars of Ghil and Mo (1991). Light dashed curves represent the 2d and 97th percentile of the chi-square red noise test of Allen and Smith (1996).

fairly low, number of regression model variables can be understood in the following way. Too few variables clearly cannot capture both the non-Gaussian structure of the phase-space PDF and the deviations from a purely red spectrum of time dependence that the non-linear QG3 dynamics engenders. If, on the other hand, too high a number of degrees of freedom is used, we start to resolve explicitly some of the high-frequency dynamics, but not all of it. It is highly plausible, therefore, that such a semiresolved model would produce unphysical and unstable results.

b. Analysis of NH geopotential height anomalies

We analyze next 44 yr of NH winter geopotential height anomalies (1 December 1949–31 March 1993). The dataset of 700-mb heights was compiled by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center; see Smyth et al. (1999) for details. A shorter, 37-winter sample of the same dataset had been used by Kimoto and Ghil (1993a,b). All analyses of this section were performed on the $44 \times 90 = 3960$ daily maps of winter data, defined as 90-day sequences starting on 1 December of each year.

The length of the interval of available data in this case is thus 15 times shorter than in the QG3 model simulation of the previous subsection. We need, therefore, to apply a regularized version of the least squares procedure, namely the PLS method (see sections 2a, 2c, and the appendix). For the QG3 model simulation, on the other hand, the PCR and PLS procedures produce the same coefficients as the usual regression, without any regularization: this happens because the length of the dataset is sufficient to resolve well the angles between different predictor variables and the problem of collinearity does not occur.

For the case of this section, an inverse stochastic model with nine predictor variables at the first level produces the best results. The necessary number of inverse model levels is determined by the procedure described in the previous subsection and turns out to be equal to 3.

The observational data and simulated time series are compared in Fig. 9 using again both PDFs and singular spectra. The analysis is carried out in the subspace spanned by the three leading EOFs and the PDFs of both the data and the inverse model are approximated using three Gaussian components (see Smyth et al. 1999). In Fig. 9a, the PDFs are plotted in the EOF-1–EOF-2 plane. Like in the previous subsection, the general shape of the PDF, as well as the locations and overall shapes of the clusters are reproduced fairly well

by the inverse model, while finer details of the clusters are not captured equally well. The spatial correlations between the anomaly patterns associated with the cluster centroids in the data (Fig. 9a, left panel; see Smyth et al. 1999 for the description of these patterns) and the inverse model simulation (Fig. 9a, right panel) are higher than 0.95. The general shape of the SSA spectra is, once again, remarkably similar in the data and our inverse model simulation; compare left and right panels in Figs. 9b,c.

If we use a smaller or larger number of predictor variables to construct a three-level inverse model, the results are not as good as in the nine-variable case; the two-level model results with nine primary variables, on the other hand, are similar to those of the three-level model. Most of the model realizations produce time series whose properties replicate those of the observed one (see Figs. 9b,c); still, in a few cases the inverse model simulates unphysically large values of the variables, pointing to a conditional instability in the model's dynamical operator. Using a different number of model variables and levels failed to suppress this instability. This property is due to an insufficient sample of predictor-variable fields used to construct our inverse model.

The above instability occurs, however, quite rarely: on average, the system tends to wander to a region with unrealistically large values of the variables once per 30 000 days of the inverse model integration. We can, in fact, avoid such situations altogether, by tracking the instantaneous norm of the modeled-state vector; if values of this norm that exceed a given threshold occur, we "rewind" the modeled time series by a few time steps and restart the model with a different realization of the random forcing. Such "sanity checks" are quite common in the engineering practice of nonlinear estimation and control theory (Miller et al. 1994).

The PDF and spectra of the long modeled time series so obtained are very close to the ones shown in Fig. 9. We have thus constructed a nonlinear model that describes well the statistical properties of the observed data.

5. Summary and discussion

We have presented a methodology for constructing data-based inverse stochastic models that allow one to isolate nonlinear dynamical processes that govern the observed variability of climatic fields. These models can also be used for climate prediction (see Kondrashov et al. 2005).

The simplest such model is the so-called linear inverse model (LIM; Penland 1989, 1996; Penland and
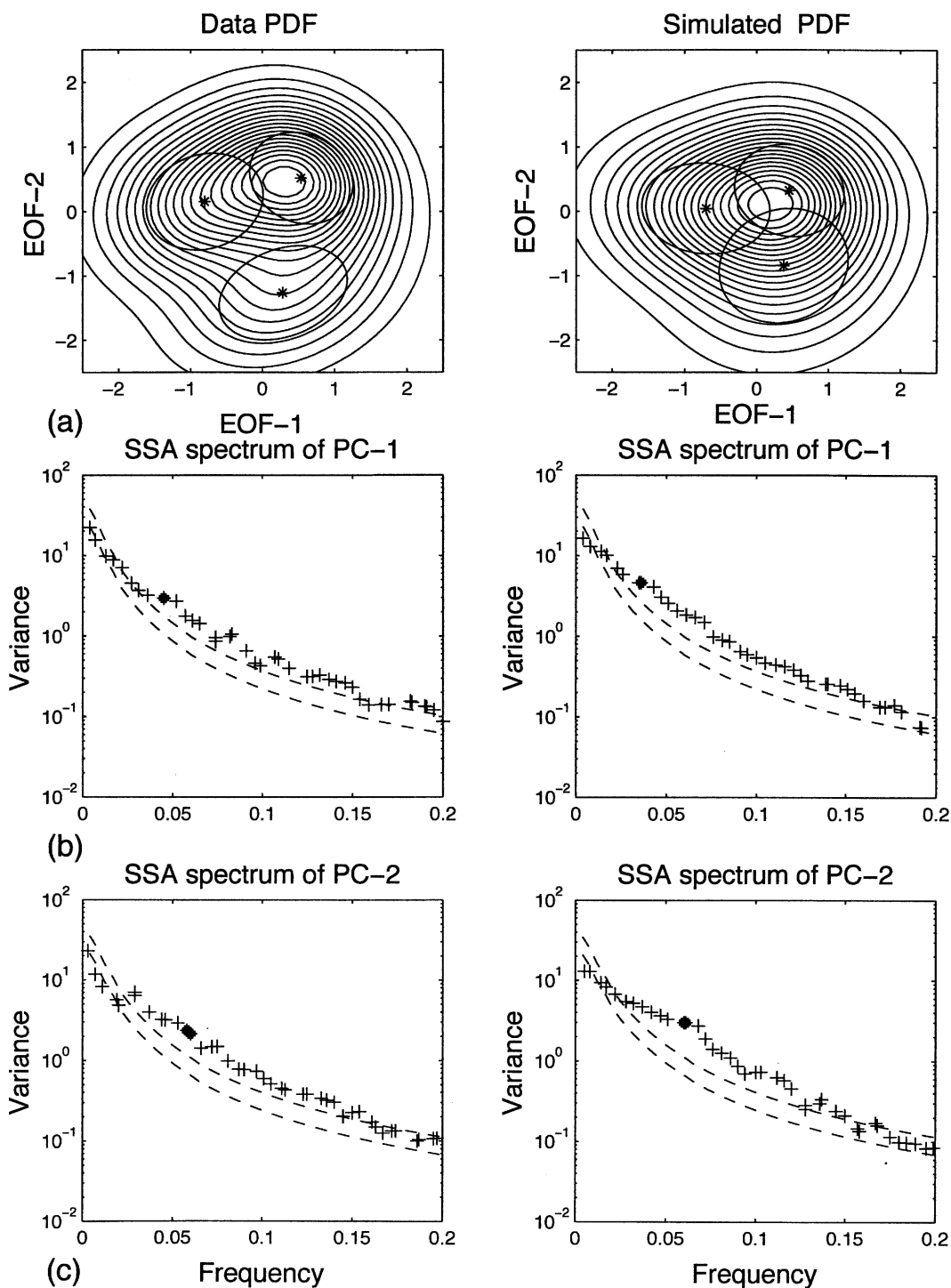
FIG. 9. PDFs and singular spectra of NH wintertime geopotential heights. (right) Data; (left) inverse stochastic
model. Same symbols and conventions as in Figs. 4 and 8.

Ghil 1993; Penland and Sardeshmukh 1995; Penland
and Matrosova 1998; Johnson et al. 2000; Winkler et al.
2001). A LIM considers the dynamics to be linear,
stable, and stochastically forced. The linear determin-

istic propagator, as well as the structure of the stochas-
tic forcing, are estimated directly from observations by
multiple linear regression (MLR), while assuming the
forcing to be white in time. The modifications we in-

troduce to this methodology deal with relaxing the assumptions of linearity and white noise.

To this end, we assume the dynamical propagator of the inverse stochastic model to be a polynomial function of the climate-state predictor variables, and use multiple polynomial regression (MPR), rather than MLR, to estimate the parameters of this function. In section 2 we have shown that, provided a long enough dataset, MPR successfully reconstructs the coefficients of two "toy" models: the deterministic Lorenz (1963a) model (Fig. 1; Table 1) and the stochastically perturbed double-well (Ghil and Childress 1987, their section 10.3; Miller et al. 1994) model (Fig. 2; Table 2). Moreover, the inverse stochastic model based on a third-order polynomial captures the three-regime probability density function (PDF) of the 2D, triple-well model of Hannachi and O'Neill (2001) (Figs. 3, 4; Table 3).

In geophysical applications to phenomena with many degrees of freedom, robust inverse models can only be constructed in the phase space of the leading empirical orthogonal functions (EOFs). The number of predictor variables to include in the model is best determined by cross validation to optimize the model's performance. Depending on the motivation, this optimization can be carried out in terms of the forecast skill or in terms of other statistical properties, such as the structure of the phase-space PDF and power spectra of the observed and modeled fields. In typical situations, the estimated stochastic forcing in an inverse model is not white in time and involves serial correlations that arise, among other things, due to the dependence of the stochastic forcing on the modeled flow. In section 3, we have formulated inverse stochastic models that involve additional levels simulating this dependence. The number of levels is chosen so that the residual forcing at the last level be white in time (Fig. 5).

The major technical difficulty that arises in formulating nonlinear inverse models is associated with the large number of regression parameters that need to be estimated. If the data record is short, direct application of MPR might result in meaningless and unstable values for the estimated parameters. This problem can be solved, fully or at least partially, by regularization techniques (see sections 2a, c and the appendix), such as principal component regression (PCR; Wetherill 1986; Press et al. 1994) or the partial least squares (PLS) approach (Wold et al. 1984; Höskuldsson 1996). We recommend the latter method for general geophysical applications.

In section 4, we have applied our MPR methods to the analysis of Marshall and Molteni's (1993) three-layer QG (QG3) model and to NH geopotential height anomalies; in the latter case, PLS regularization was also used. For both QG3-generated (Figs. 6–8) and observed height data (Fig. 9), the optimal inverse model captures well the non-Gaussian structure of the PDF, as well as the detailed spectra of the data. The reduced models we have built can thus be used to further explore the macroscopic, coarse-grained processes behind the QG3 model's dynamical behavior and its connection to the observed data statistics.

The multiple regime structure of the QG3 model's PDF can be captured by a single-level, quadratic inverse model. In this case, however, the inverse model's PDF is very sensitive to the sampling, and its trajectories may diverge in time for some of the stochastic-forcing realizations; in contrast, the multilevel versions of the quadratic inverse model are more robust. Moreover, the spectra of the single-level inverse model are statistically indistinguishable from a red spectrum, while those of the multilevel model resemble the spectrum of the QG3 model's principal components. The multilevel model is stable and much less sensitive to sampling. This smaller sensitivity and greater stability is due to the stabilizing feedback between high- and low-frequency components of the flow, as expressed in the mathematical formulation (20) of this model's additional levels. In the present context, we speculate that this feedback is the counterpart of the much-discussed synoptic eddy–large-scale flow feedback.

A different way to quantify this feedback has been suggested in semiempirical studies that assume a reduced-dynamics model, while the unresolved processes are being parameterized statistically by accumulating long libraries of residual flow dependence on the large-scale flow (Vautard and Legras 1988; Molteni 1996a,b; Da Costa and Vautard 1997; Molteni 2002). D'Andrea and Vautard (2001) and D'Andrea (2002) have recently applied this technique to construct a large-scale model by projecting Marshall and Molteni's (1993) QG3 model on its leading EOFs. Unlike these studies, our multilevel approach does not assume any explicit dynamical model, deducing it instead from the data. Despite that, the correlation between full and reduced model clusters in our case is higher than in D'Andrea (2002), while our mathematical representation of the eddy feedback is simpler and easier to interpret.

Majda et al. (1999, 2001, 2002, 2003) have developed a strategy for systematic mode reduction in deterministic models that govern geophysical flows. This strategy has recently been applied to the analysis of a barotropic atmospheric model by Franzke et al. (2005) and to the QG3 model (C. Franzke and A. Majda 2005, personal communication). In particular, the optimal set of reduced equations that approximate the underlying

model's quadratic nonlinearities has been shown to involve cubic nonlinearities and additive, as well as multiplicative white noise. In contrast, our regression procedure explicitly involves only quadratic nonlinearities and an additive noise.

The number of variables in our reduced model is much less than the number of degrees of freedom in the full QG3 model, but it is still larger than the number of modes that contain most of the model's low-frequency variance (EOFs 1–4). The higher-ranked EOFS 5–15 can thus be associated with intermediate-scale modes, while the additive noise in our reduced model's highest level represents the smallest scales, captured by EOFs 16 and higher in the full QG3 model.

In this interpretation, the inverse model we construct by MPR from the data will involve quadratic nonlinearities and additive, as well as multiplicative noise; cubic nonlinearities do not explicitly enter the equations. While Franzke et al. (2005) have shown that the effect of cubic nonlinearities on their reduced barotropic model's evolution is weak, C. Franzke and A. Majda (2005, personal communication) argue for a much more important role of such interactions in the QG3 model's behavior. This greater role of cubic nonlinearities might be related to a smaller number of leading modes retained by Franzke and Majda.

Direct comparison of our method with the mode-reduction strategy of Majda and coauthors is difficult at this stage: our strategy emphasizes separation in spatial scales, with large, intermediate, and small scales, theirs emphasizes separation in temporal scales, with slow and fast scales only. The similarities and differences between these two methods, when applying both to a fairly realistic atmospheric model, like QG3, are a matter of considerable interest and further investigation. We note, however, that our regression technique is entirely data driven, and can be used when no dynamical model is explicitly associated with the dataset (see also Kondrashov et al. 2005).

Finally, we recall that the main ingredients of the regression techniques employed in this study (linear least squares, MPR, PCR, PLS) are not new. The primary purpose of this paper is to adapt these techniques to realistic geophysical situations and thus encourage their use for the analysis of complex geophysical phenomena.

APPENDIX

General Linear Least Squares

The general problem posed by statistical regression is to minimize the functional

$$\chi^2 = |\mathbf{D}\mathbf{a} - \mathbf{y}|^2, \tag{A1}$$

where $\mathbf{y} = \{y_q\}|_{q=1}^Q$ is the vector whose components are the observed values of the response variable, $\mathbf{D} = \{X_{qn}(\mathbf{x})\}$ is the so-called design matrix, whose nth column consists of Q-observed values of a specified function $X_n|_{n=1}^N$ of the state vector $\mathbf{x} = \{x_i\}|_{i=1}^I$, and $\mathbf{a}$ is the vector of N regression parameters to be estimated. For example, in the case of MLR, $N = I$ and $X_i = x_i, 1 \leq i \leq I$. For MPR, the number N of basis functions exceeds the dimension I of the state vector and, therefore, additional regression parameters have to be estimated.

The solution of (A1) is

$$\mathbf{a} = \mathbf{C}^{-1}\mathbf{D}^T\mathbf{y}, \tag{A2}$$

where $\mathbf{C} = \mathbf{D}\mathbf{D}^T$ is the covariance of the design matrix. It can be shown that the entries of $\mathbf{C}/\sigma^2$, where σ is the standard error of observations, represent the standard errors of estimated parameters $\mathbf{a}$. If any two or more columns of $\mathbf{D}$ are nearly linearly dependent, the inversion (A2) is ill conditioned. A way to solve this problem of *collinearity* is to regress $\mathbf{y}$ onto the EOFs of $\mathbf{D}$, which are orthogonal and, therefore, well conditioned. This is called PCR (Wetherill 1986).

Let the SVD of $\mathbf{D}$ be

$$\mathbf{D} = \mathbf{U}\mathbf{W}\mathbf{V}^T, \tag{A3}$$

where $\mathbf{U}$ has the dimension of $\mathbf{D}$, $\mathbf{W}$ is the diagonal matrix of N singular values (principal component scores), and $\mathbf{V}$ is an $N \times N$ orthogonal matrix. Then the solution of Eq. (A1) is

$$a = \sum_{n=1}^N \left(\frac{\mathbf{U}_{(n)} \cdot \mathbf{y}}{w_n}\right)\mathbf{V}_{(n)} \pm \frac{1}{w_1}\mathbf{V}_{(1)} \pm \ldots \pm \frac{1}{w_N}\mathbf{V}_{(N)}, \tag{A4}$$

where $\mathbf{U}_{(n)}$ and $\mathbf{V}_{(n)}$ are columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, and each $\pm$ is followed by a standard deviation. The idea of regularization is to edit out small singular values, which do not contribute much to reducing the χ^2 of the dataset (Press et al. 1994). The number of PCs, or factors (as they are often called in the statistical literature), to use is typically determined by cross validation, a procedure where the available data is split between

training and validation sets and the validation is carried out for various numbers of retained PCs in order to choose one number that optimizes the predictive ability of the model.

Yet another regularization method is the PLS procedure (Wold et al. 1984; Höskuldsson 1996), which is analogous to PCR, but in some sense is more powerful, since it not only captures a large fraction of variance in the predictor variables, but also achieves high correlation with the response variables. This is done by rotating PC scores and loadings (or PCs and EOFs, in the meteorological terminology) to maximize covariance between predictor and response variables. The number of rotated PCs (Richman 1986) to use is determined by cross validation, as in PCR.

While conceptually simple, PLS is algorithmically complex and has various formulations; see Jolliffe (2002, chapter 8.4) and references there. We have found the PLS method to be superior to PCR in the geophysical examples used in this study. (A software package containing the PLS routines is available online at http://www.software.eigenvector.com.)

REFERENCES

Allen, R. M., and L. A. Smith, 1996: Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise. *J. Climate,* **9,** 3373–3404.

Berloff, P., and J. C. McWilliams, 2002: Material transport in oceanic gyres. Part II: Hierarchy of stochastic models. *J. Phys. Oceanogr.,* **32,** 797–830.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, 1994: *Time Series Analysis, Forecasting and Control.* 3d ed. Prentice Hall, 592 pp.

Cheng, X. H., and J. M. Wallace, 1993: Analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.,* **50,** 2674–2696.

Da Costa, E., and R. Vautard, 1997: A qualitative realistic low-order model of the extratropical low-frequency variability built from long records of potential vorticity. *J. Atmos. Sci.,* **54,** 1064–1084.

D'Andrea, F., 2002: Extratropical low-frequency variability as a low-dimensional problem. Part II: Stationarity and stability of large-scale equilibria. *Quart. J. Roy. Meteor. Soc.,* **128,** 1059–1073.

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

DelSole, T., 1996: Can quasigeostrophic turbulence be modeled stochastically? *J. Atmos. Sci.,* **53,** 1617–1633.

——, 2000: A fundamental limitation of Markov models. *J. Atmos. Sci.,* **57,** 2158–2168.

Deser, C., 2000: On the teleconnectivity of the "Arctic Oscillation." *Geophys. Res. Lett.,* **27,** 779–782.

Dettinger, M. D., M. Ghil, C. M. Strong, W. Weibel, and P. Yiou, 1995: Software expedites singular-spectrum analysis of noisy time series. *Eos, Trans. Amer. Geophys. Union,* **76,** 12, 14, 21.

Dijkstra, H. A., and M. Ghil, 2005: Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.,* in press.

Farrell, B. F., and P. J. Ioannou, 1993: Stochastic forcing of the linearized Navier–Stokes equations. *Phys. Fluids,* **A5,** 2600–2609.

——, and ——, 1995: Stochastic dynamics of the midlatitude atmospheric jet. *J. Atmos. Sci.,* **52,** 1642–1656.

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

Ghil, M., and S. Childress, 1987: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics.* Springer-Verlag, 485 pp.

——, and K. C. Mo, 1991: Intraseasonal oscillations in the global atmosphere. Part I: Northern Hemisphere and Tropics. *J. Atmos. Sci.,* **48,** 752–779.

——, and A. W. Robertson, 2000: Solving problems with GCMs: General circulation models and their role in the climate modeling hierarchy. *General Circulation Model Development: Past, Present and Future,* D. Randall, Ed., Academic Press, 285–325.

——, and Coauthors, 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.,* **40,** 1003, doi:10.1029/2000RG000092.

Hand, D., H. Mannila, and P. Smyth, 2001: *Principles of Data Mining.* MIT Press, 546 pp.

Hannachi, A., 1997: Low-frequency variability in a GCM: Three-dimensional flow regimes and their dynamics. *J. Climate,* **10,** 1357–1379.

——, and A. O'Neill, 2001: Atmospheric multiple equilibria and non-Gaussian behavior in model simulations. *Quart. J. Roy. Meteor. Soc.,* **127,** 939–958.

Hasselmann, K., 1999: Climate change: Linear and nonlinear signature. *Nature,* **398,** 755–756.

Höskuldsson, A., 1996: *Prediction Methods in Science and Technology.* Thor Publishing.

Johnson, S. D., D. S. Battisti, and E. S. Sarachik, 2000: Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies. *J. Climate,* **13,** 3–17.

Jolliffe, I. T., 2002: *Principal Component Analysis.* 2d ed. Springer, 487 pp.

Kimoto, M., and M. Ghil, 1993a: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.,* **50,** 2625–2643.

——, and ——, 1993b: Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions. *J. Atmos. Sci.,* **50,** 2645–2673.

Kondrashov, D., K. Ide, and M. Ghil, 2004: Weather regimes and preferred transition paths in a three-level quasigesotrophic model. *J. Atmos. Sci.,* **61,** 568–587.

——, S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A hierarchy of data-based ENSO models. *J. Climate,* **18,** 4425–4444.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.,* **42,** 433–471.

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

——, 1963b: The mechanics of vacillation. *J. Atmos. Sci.,* **20,** 448–464.

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

——, ——, and ——, 2001: A mathematical framework for sto-

chastic climate models. *Commun. Pure Appl. Math.,* **54,** 891–974.

——, ——, and ——, 2002: A priori test of a stochastic mode reduction strategy. *Physica D,* **170,** 206–252.

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

Marshall, J., and F. Molteni, 1993: Toward a dynamical understanding of atmospheric weather regimes. *J. Atmos. Sci.,* **50,** 1792–1818.

McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models.* Chapman and Hall, 511 pp.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Mo, K., and M. Ghil, 1988: Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.,* **93D,** 10 927–10 952.

Molteni, F., 1996a: On the dynamics of planetary flow regimes. Part I: The role of high-frequency transients. *J. Atmos. Sci.,* **53,** 1950–1971.

——, 1996b: On the dynamics of planetary flow regimes. Part II: Results from a hierarchy of orographically forced models. *J. Atmos. Sci.,* **53,** 1972–1992.

——, 2002: Weather regimes and multiple equilibria. *Encyclopedia of Atmospheric Science,* J. R. Holton, J. Curry, and J. Pyle, Eds., Academic Press, 2577–2585.

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

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, 1996: A stochastic model of Indo-Pacific sea-surface temperature anomalies. *Physica D,* **98,** 534–558.

——, and M. Ghil, 1993: Forecasting Northern Hemisphere 700-mb geopotential height anomalies using empirical normal modes. *Mon. Wea. Rev.,* **121,** 2355–2372.

——, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

——, and L. Matrosova, 1998: Prediction of tropical Atlantic sea-surface temperatures using linear inverse modeling. *J. Climate,* **11,** 483–496.

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, 425 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1994: *Numerical Recipes.* 2d ed. Cambridge University Press, 994 pp.

Randall, D., Ed., 2000: *General Circulation Model Development: Past, Present and Future.* Academic Press, 781 pp.

Richman, M. B., 1986: Rotation of principal components. *J. Climatol.,* **6,** 293–335.

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

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

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

——, 1997: Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.,* **54,** 2100–2114.

Sirovich, L., and J. D. Rodriguez, 1987: Coherent structures and chaos—A model problem. *Phys. Lett.,* **120,** 211–214.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.,* **56,** 3704–3723.

Vautard, R., and B. Legras, 1988: On the source of midlatitude low-frequency variability. Part II: Nonlinear equilibration of weather regimes. *J. Atmos. Sci.,* **45,** 2845–2867.

——, and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D,* **35,** 395–424.

Von Mises, R., 1964: *Mathematical Theory of Probability and Statistics.* Academic Press, 694 pp.

Wallace, J. M., 2000: North Atlantic Oscillation/annular mode: Two paradigms—One phenomenon. *Quart. J. Roy. Meteor. Soc.,* **126,** 791–805.

Wetherill, G. B., 1986: *Regression Analysis with Applications.* Chapman and Hall, 311 pp.

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

Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III, 1984: The collinearity problem in linear regression: The Partial Least Square approach to generalized inverses. *SIAM J. Sci. Stat. Comput.,* **5,** 735–743.