Hetero- and homoclinic orbits were shown to play an important role in the interannual and interdecadal LFV of the midlatitude oceanÕs wind-driven circulation (Meacham 2000; Chang et al. 2001; Nadiga and Luce 2001; Simonnet et al. 2003a, b, 2005). Crommelin (2002, 2003) explored in detail the role of hetero- and homoclinic orbits in a low-order and an intermediate-order model of the extratropical atmosphere, respectively. These coarse-grained features of the LFV may be better understood by using reduced models, which have considerably fewer degrees of freedom. Such models can be derived from either observational or simulated datasets (Majda et al. 2003; Franzke et al. 2005; Kravtsov et al. 2005), and have been shown to accurately represent both linear and nonlinear aspects of LFV (Kondrashov et al. 2005, 2006), provided the effect of higher-frequency, synoptic transients is suitably parameterized as stochastic forcing.

DÕAndrea and Vautard (2001, 2002) found a good correspondence between their low-order modelÕs quasi-stationary states and the regimes of the quasi-geostrophic, three-layer (QG3) model of Marshall and Molteni (1993). Kondrashov et al. (2006) showed that one of the QG3 modelÕs regimes is indeed associated with the unstable Þxed point of the reduced model obtained by the multilevel regression method of Kravtsov et al. (2005).

Despite better observational and theoretical understanding of weather regimes, the forecasting of regime breaks and subsequent onsets still presents a challenge for general circulation models (GCMs) and operational systems. Kimoto et al. (1992) examined the temporal variability of the skill of operational medium-range forecasts in three operational models: the Japan Meteorological AgencyÕs (JMA) global spectral model, and those of the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Meteorological Center (NMC). During the period under study, the forecast skill of the three models exhibited considerable LFV and showed a pronounced temporal minimum during transitions from zonal to blocked ßow over the North Paciþc. DÕAndrea et al. (1998) have reviewed the performance of 15 GCMs and found them to generally underestimate the number and duration of atmospheric blocking episodes, in agreement with previous work.

More recently, Pelly and Hoskins (2003) emphasized the importance of transitions from and to a blocked state in operational forecasts over Europe, and found that such forecasts in the ECMWF Ensemble Prediction System still have limited skill beyond a lead time of 6 days. Palmer et al. (2007) highlighted the substantial and continuing under-prediction of blocking episodes in leading operational models (their Fig. 3) and emphasized that the number and severity of such episodes play a key role in climate change and its socioeconomic effects.

Vannitsem (2001) suggested a hybrid clustering method that combines hierarchical clustering of phase-space position with predictability considerations using Lyapunov exponents; see Legras and Ghil (1985) and Ghil and Childress (1987, Sect. 6.4) for the distinction between regime predictability and pointwise predictability, and Table 1 in Ghil and Robertson (2002) for clustering methods based on either position or velocity in phase space. VannitsemÕs hybrid method produces well-separated clusters with distinct instability properties in the QG3 model, and medium-range predictability of the clusters beyond that obtained by hierarchical clustering alone (Trevisan 1995).

The present work builds on results of Deloncle et al. (2007), who showed considerable skill in forecasting regime breaks and subsequent onsets from the output of the QG3 model, by using a statistical learning method called random forests (Breiman 2001). If at least part of this skill carried over to actual atmospheric regimes, the method could help improve the toolbox of the medium-range forecaster.

Kondrashov et al. (2004) analyzed in detail a long-time integration of this model, and showed that several regime transitions were characterized by preferential directions, which fall within narrow solid angles in a low-dimensional phase space. These preferred paths are tangent to the vectors pointing from a regime centroid to the exit points, and can be identiÞed as maxima in the joint probability density function (PDF) of two angles on the unit sphere around the centroid. By applying the same analysis to the Lorenz (1963) model, Kondrashov et al. (2004) concluded that the observed narrow maxima in the PDF of the regime exit angles in the QG3 model are likely to be determined by the dynamical properties of the underlying nonlinear system of equations (Guckenheimer and Holmes 1983; Ghil and Childress 1987, Sect. 5.4). These properties include the linearly stable and unstable directions associated with the steady states.

Selten and Branstator (2004) used a different approach to Þnd preferred transition directions by analyzing the QG3 modelÕs mean tendencies in a low-dimensional subspace. These authors related their three weather regimes to the slow phases of their 20-day QG3 oscillatory mode, and also showed that preferred transition paths between regimes do exist. Ghil and Robertson (2002) had already highlighted the connection between regimes and intraseasonal oscillations and its potential importance for medium-range forecasting. The results of Selten and Branstator (2004) differ in this respect somewhat from those of Kondrashov et al. (2004), where two of the four weather regimes were found to correspond to the slow phases of their 37-day oscillatory mode; the latter authors also found a 19-day oscillation with a very similar spatial pattern, although these two oscillations were not phase-locked. Kondrashov et al.Õs

(2006) reduced-model analysis showed that the 37-day oscillation is associated with the least-damped eigenmodes of this model, linearized either about its climatological state or about its unique steady state, while the 20-day mode is strongly damped.

Following Kimoto and Ghil (1993b), Crommelin (2004) used an angular PDF to determine ßow regimes in the reanalysis dataset of the National Centers of Enviromental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) for the Northern Hemisphere (NH) circulation (1948Ð2000). This methodology is based on angular distance in a reduced phase space as a measure of similarity between ßow patterns, as opposed to the Euclidian distance used by Kimoto and Ghil (1993a), Kondrashov et al. (2004), and in this study. Crommelin (2004) exploited asymmetries in the transition probability matrix of the Markov chain associated with a partition of the reduced phase space to detect a preferred, closed path on the unit sphere in this space. This preferential cycle of transitions alternates between states with zonal symmetry and meridionally oriented anomalies, in agreement with the oscillatory mode found by Kondrashov et al. (2004) in the QG3 model.

In another study of preferred regimes in actual atmospheric data, Monahan et al. (2001) applied a nonlinear generalization of principal component (PC) analysis to the NCEP-NCAR Reanalysis dataset. They obtained three regimes, labeled A, G and R, which resemble well those of Cheng and Wallace (1993) and of Smyth et al. (1999). The preferred transition route between these regimes follows a curved path in their reduced phase space, which involves certain similarities with CrommelinÕs (2004) most signiÞcant cycle.

Despite the differences in methodology and datasets, Monahan et al. (2001), Vannitsem (2001), Crommelin (2004), Kondrashov et al. (2004), and Selten and Branstator (2004) all provide evidence on preferred transition routes in the atmospheric LFVÕs phase space. This evidence reinforces the conjecture that weather regimes represent a coarse-grained and predictable component of the atmosphere (Mo and Ghil 1988; Ghil et al. 1991). Deloncle et al. (2007) showed that good predictability of regime transitions in the QG3 model can be obtained by utilizing Kondrashov et al.Õs (2004) information on preferred transition paths within the framework of novel statistical learning methods.

We apply the methodology of Deloncle et al. (2007) to obtain the preferred transition paths between weather regimes in the wintertime geopotential heights from the 1948Ð2003 NCEP-NCAR reanalysis (Kalnay et al. 1996). The random-forests method is then applied to study the extended-range predictability of regime transitions (Ghil and Robertson 2002; Deloncle et al. 2007).

The purpose of the paper is twofold: (a) to test on actual atmospheric data the usefulness of extended forecasts conditioned on a subset of initial states; and (b) the role of preferential directions of regime exit in such extended forecasts. The paper is organized as follows. In Sect. 2, we describe the data and identify robust weather regimes by Gaussian mixture modeling (Smyth et al. 1999; Kondrashov et al. 2004), as well as highly signiÞcant transitions in the Markov chain of these regimes. The preferred transition paths (Sect. 3.1) allow us to deÞne an efÞcient set of predictor variables for the study of long-term predictability with statistical learning techniques (Sect. 3.2). The results for regime transition forecasts are described in Sect. 4, and followed by a summary and discussion in Sect. 5.

2 Weather regimes and transitions

2.1 Data

The dataset for this analysis consists of NH wintertime geopotential heights. The data are based on the NCEP-NCAR reanalysis (Kalnay et al. 1996) for the years 1948Ð2003 and cover the NH in space and 55 winters that are 90-day long, from 1 December 1948 till 28 February 2003; including 13 leap years yields a sample of 4,963 maps, deÞned on a 5 × 5 grid. To examine robust features of the observations phase-space structure we reduce the datasetÕs dimensionality: empirical orthogonal function (EOF) analysis is applied to the unÞltered 700-hPa level height anomalies, with data points being weighted by the square root of the cosine of their latitude. The choice of the 700-hPa level, as opposed to 500 hPa in Monahan et al. (2001) and Crommelin (2004), is motivated in part by our concern for actual prediction: as shown by Namias (1953, 1968), this level represents a good compromise between intrinsic variability of the free atmosphere (best captured by the 500-hPa level) and surface inßuences (best reßected by the 1,000-hPa level).

The ten leading EOFs capture 60% of the variance of the dataset: the Þrst mode captures 9.9%, the second one 8.4%, the third one 7.5%, and the tenth one 3.1%; the cumulative variance described by the four leading EOFs equals 32%. The spatial patterns associated with the Þrst and second EOFs (not shown here) are very similar to to those found in Kimoto and Ghil (1993a); see Fig. 6 there. Analysis of the geopotential height Þeld at the 250 and 500-hPa levels (not shown) reveals spatial EOF patterns that resemble well the ones at the 700-hPa level, but with a Þner structure present at the lower pressure levels. This similarity indicates a predominantly barotropic structure of the LFV in the reanalysis. The fact that this structure dominates even the leading EOFs of the unÞltered data is due, as shown by

Kimoto and Ghil (1993a), to the LFVÕs far exceeding the synoptic-scale variability.

2.2 Weather regimes

Weather regimes have been associated with regions of higher PDF on the systemÕs attractor. In order to objectively identify weather regimes in our data, we apply the Gaussian mixture model of Smyth et al. (1999) and Kondrashov et al. (2004). The Gaussian mixture model uses a linear combination of k Gaussian density functions to approximate the systemÕs PDF. For a given number d of leading EOFs, it provides a number of clusters k and the cluster centroids in a d-dimensional subspace of the modelÕs phase space. We associate each cluster with a weather regime of our dataset.

The mixture model has a built-in criterion for determining the optimal number of clusters supported by the data. This criterion is based on the cross-validated log likelihood: the higher its algebraic value for a given dimension d, the more likely it is that k is the correct number of clusters for that d. Our results (not shown) indicate there is a clear maximum at $k = 3$ for various numbers d of leading EOFs; this result is the same as that of Smyth et al. (1999) for the 44 winters from 1 December 1949 to 31 March 1993. Therefore we choose $k = 3$ as the optimal number of clusters for our dataset.

The PDF based on the mixture model for $d = 3$ and $k = 3$ is shown in Fig. 1a, in the phase plane spanned by PCs 1 and 2, while the anomaly maps of the clusters are shown in Fig. 1bÐd. The left column of Fig. 1 shows results for the full dataset (1948Ð2003). We conducted, furthermore, the same analysis for the two half-data sets, years 1948Ð1975 and 1976Ð2003, to examine the robustness of the cluster analysis. The results for the latter data set are presented in the right column of Fig. 1; the three cluster centroids and their covariance structures are very similar between the left and right columns.

All PCs are normalized by the standard deviation of PC-1. This normalization is chosen merely so that the system's trajectory in reduced phase space scale properly, since the variance captured by each EOF is different (see previous subsection). None of our results are sensitive to this choice of normalization factor.

The size of the clusters is determined by choosing the length of the semi-axes of the covariance ellipsoid around each cluster centroid; in Fig. 1a this length corresponds to 1.5 of the standard deviation in each direction. A data point is assigned to the weather regime if it lies within the corresponding ellipsoid. However, each data point can have a degree of membership in several clusters, depending on its position with respect to the centroid and the weight of the cluster (Smyth et al. 1999). If a data point belongs to several ellipsoids, we assign it to a single cluster, according to the maximum probability value given by the mixture model.

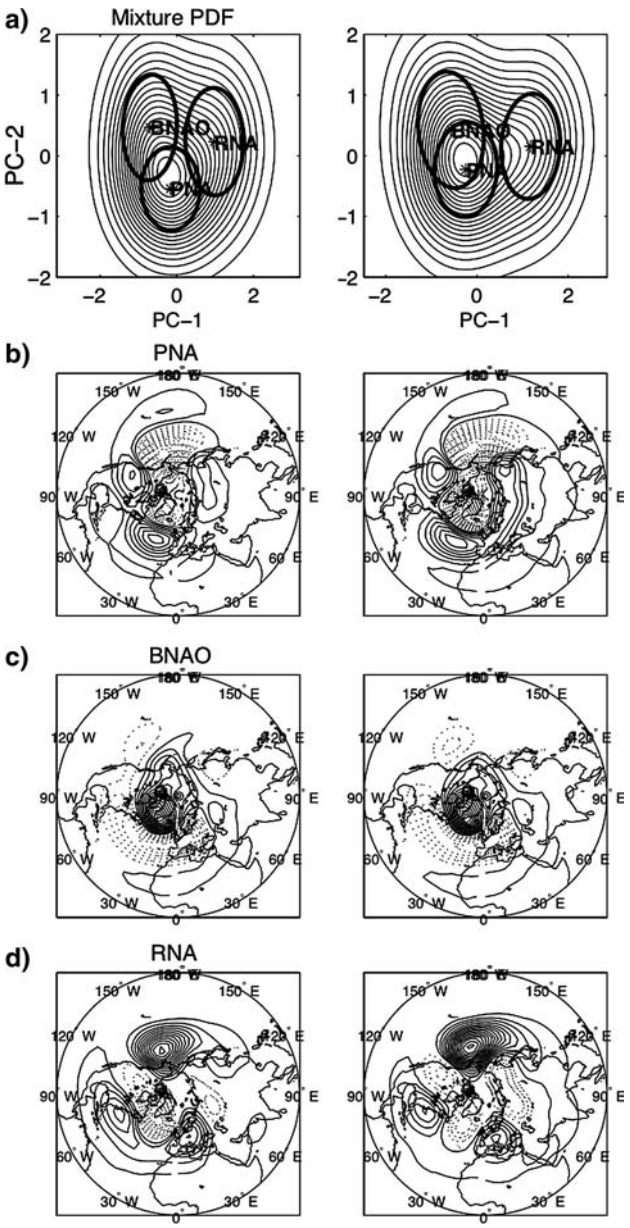

Fig. 1 Probability density function (PDF) of the 700-hPa geopotential height anomaly field and cluster centroids: left panels for the years 1948Ð2003, and right panels for the years 1976Ð2003. PDF estimated by the mixture model for $d = 3$ and $k = 3$ and projected onto the plane spanned by PC-1 and PC-2; 20 contour levels are used and the heavy ellipses correspond to semi-axes equal to 1.5 each principal direction. Mixture model centroids showing geopotential height anomaly maps: b PNA; c BNAO; and d RNA. Negative contours are dotted and land masses are contoured; 20 contour levels between maximum and minimum values are used, with the following intervals (in m): b 4.6, c 6.2, and d 6.1

The anomaly maps of the clusters are shown in Fig. 1bÐd; they are very similar to those obtained by Cheng and

Wallace (1993), Smyth et al. (1999), and Monahan et al. (2001). We denoted these three regimes, therefore, by (b) PNA, (c) BNAO and (d) RNA, following the labels given by Smyth et al. (1999). These spatial patterns have several prominent and easily identiÞable features. Thus the BNAO mapÑblocked phase of the North Atlantic Oscillation (NAO)Ñexhibits a strong high over southern Greenland, while the PNA or PaciÞcÐNorth-American pattern is characterized by an intensiÞcation of the PaciÞc jet stream and an enhancement of the climatological mean ridge over the Rockies; the latter lead to below-average heights in the Central North PaciÞc, and above-average heights over the Northwest of the United States and western Canada. The RNA, or (approximately) reversed PNA map, possesses a distinctive ridge over the Gulf of Alaska. In the North Atlantic region, both the PNA and the RNA pattern contain features of a zonal NAO (ZNAO) cluster, found usually in analyses of the Atlantic sector alone, but not in hemispheric regimes.

2.3 SigniÞcant transitions

Given the clustering results of Sect. 2.2, we can determine the Markov chain of transitions between the three regimes. We deÞne a regime episode as the subset of consecutive sample points (days) along the model trajectory that fall within a given cluster (Mo and Ghil 1988; Kimoto and Ghil 1993b; Kondrashov et al. 2004). Table 1 shows the number of distinct episodes in each regime and the total number of days spent in the regime for scaling factors of 1.5 and 1.75. A total number of 3,490 days (70%) out of the entire 4,963 days belong to one of the large-size clusters (1.75σ), while for the smaller clusters (1.5σ) the total regime population is of about 2,800 days (57%).

Table 1 Regime statistics for two cluster sizes

	PNA	BNAO	RNA	Total
(a) Size				
1.5σ				
Episodes	288	231	250	769
Days	1,209	799	799	2,807
1.75σ				
Episodes	303	241	267	811
Days	1,516	959	1,015	3,490
(b) Duration (days)				
PNA	4.19	1.72	2.02	
BNAO	2.95	3.42	2.08	
RNA	2.58	3.27	3.20	

(a) Number of episodes for, and days in, each cluster; and (b) mean residence time in each cluster and mean duration of transition between clusters. The cluster size is determined by the scaling factor of the standard deviation along each semi-major axis

We deÞne regime transition simply as the passage from one regime episode to another. The number of unclassiÞed days that is allowed between the regime episodes varies according to cluster size, but is not constrained by our methodology. For the smaller-size clusters the mean transition times are shown in Table 1, and there are 96 PNA $\rightarrow$ BNAO, 107 PNA $\rightarrow$ RNA, and 112 BNAO $\rightarrow$ PNA transitions, respectively.

Table 2 shows transition probabilities between the clusters using the sequence of regime episodes along the trajectory, counting also self-transitions. Monte Carlo simulation was applied to provide a statistical signiÞcance test for the elements of the transition matrix, following Vautard et al. (1990). The test is designed to take into account the difference in size between regimes and uses random shuffling of the sequence of regime episodes in the model simulation, subject to the constraint of the number of episodes in each regime being Þxed and equal to the one in Table 1. The transition probabilities between regimes that are higher or lower than the Monte Carlo result at the 95% level appear in Table 2 in bold or italics, respectively. Note that each row in the table sums to unity, for both cluster sizes being reported (whether 1.5σ or 1.75σ). The transitions that are highly signiÞcant for both cluster sizes are shown as a double arrow in Fig. 2a, while the transition that is only signiÞcant for the larger cluster is shown as a single arrow.

The transition probabilities and their signiÞcance are fairly similar for the small and large clusters. We Þnd that three transitions are signiÞcantly higher against random shuffling for both cluster sizes: PNA $\rightarrow$ BNAO, PNA $\rightarrow$ RNA, and BNAO $\rightarrow$ PNA. It is interesting to note that both transitions from PNA are signiÞcant, and that there are no signiÞcant transitions between BNAO and RNA. Transition RNA $\rightarrow$ PNA is signiÞcant only for the larger clusters, which could be due to the smaller dataset for 1.5σ.

Monahan et al. (2001) obtained similar results: their nonlinear PC analysis yielded three preferred regimes, A, B and G, which closely resemble those of Smyth et al. (1999).

Table 2 Transition probabilities estimated using the mixture model, with row and column corresponding to origin and destination, respectively

1.5σ/1.75σ	PNA	BNAO	RNA
PNA	0.30/0.25	0.34/0.36	0.37/0.39
BNAO	0.49/0.48	0.26/0.24	0.25/0.28
RNA	0.36/0.43	0.30/0.27	0.34/0.30

Transitions that are signiÞcantly higher at the 95% level with respect to 10,000 random shuffles of the sequence of regime episodes are in bold, while entries that are italicized are signiÞcantly lower at the 95% level for the same test. The entries in the table are for clusters of size 1.5σ and 1.75σ, in this order. Not all rows sum exactly to 1.00 because of round-off effects

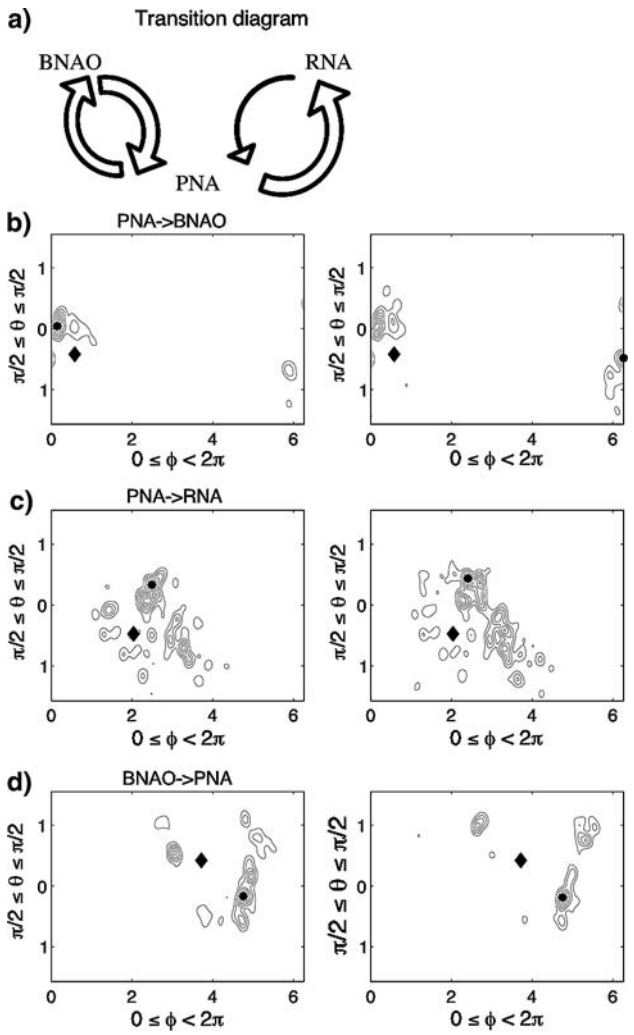

Fig. 2 Transition diagram and preferred-transition directions for our NH weather regimes: a Transition diagram and 2-D PDF of regime exit angles for the 1.5 (left column) and 1.75 (right column) clusters: b PNA → BNAO, c PNA → RNA, and d BNAO → PNA. The contour interval for all panels is equal to 0.2 in nondimensional units, black asterisks indicate preferred transition path, and black diamonds correspond to the straight line that connects the two cluster centroids in question

and correspond to RNA, PNA and BNAO in this study, respectively. Monahan et al. (2001) also found that direct transitions between regimes A and G are rare, and that transitions between these regimes occur via a regime that resembles the PNA.

Crommelin (2004) applied yet another methodology to examine preferred paths of regime transitions in NH winter geopotential height data. He projected the system's trajectory onto the unit sphere in the reduced phase space spanned by the three leading EOFs, and utilized angular distance, which is equivalent to pattern correlation, to classify regimes as regions of angular PDF maxima on this sphere. Among five flow regimes examined by Crommelin (2004), reasonable agreement exists between his regime A and our PNA, his E and our BNAO, and (in the Pacific sector) a combination of his C and D and our RNA. Using various regular partitions on the unit sphere, Crommelin (2004) calculated then the associated Markov chain transition matrix, following the work of Pasmanter and Timmermann (2003). The antisymmetric part of this matrix is associated with preferred transition cycles, as opposed to the symmetric part of this matrix, presumably corresponding to diffusive flow in the reduced phase space. The best agreement between the Crommelin's results and ours is found for a particular partition of the unit sphere into four cells (see Fig. 6d in his paper), for which the cycle 4 → 2 → 1 → 3 is highly significant and can be associated with the path BNAO → PNA → RNA of this study, while lower significance was obtained for the path between his cells 3 and 4 (RNA → BNAO).

2.4 Preferred transition paths

Kondrashov et al. (2004) refined the Markov chain representation of regime transitions for the QG3 model by finding that most of the highly significant regime breaks (see again Table 2) occur within one or two narrow solid angles in phase space—the preferred exit paths. These solid angles are formed by the regime exit vectors in a three-dimensional (3-D) subspace of the model's phase space. The exit vector is defined to point from the cluster centroid to the exit point, which is the mid-point between two consecutive trajectory points that lie on the opposite sides of the cluster boundary. In the 3-D subspace spanned by EOFs 1–3, with coordinates (x,y,z), each exit vector is uniquely defined by two angles: $\tan\phi = y/x$, $0 < \phi < 2\pi$, and $\tan\theta = z/\sqrt{x^2 + y^2}$, $-\pi/2 < \theta < \pi/2$.

Following Kondrashov et al. (2004), we compute 2-D PDF distributions of the exit angles in the (ϕ,θ) coordinates for all of the regime episodes using a Gaussian kernel estimator (Silverman 1986). This estimator approximates a PDF by assigning a localized kernel density function of a given shape (Gaussian in our case) to each data point. The PDF estimate at any point in phase space is then given by the sum of these kernel density functions. We use the data-adaptive version of the kernel estimator (Kimoto and Ghil 1993a) and modify it to account for the periodic nature of the PDF in the ϕ direction.

Figure 2b–d show the exit-angle PDFs of the three significant transitions (thick arrows in Fig. 2a) for the smaller-size clusters (1.5, left column) and the larger-size clusters (1.75, right column), obtained with an adaptively estimated kernel window width $\pi/6$ for all cases. Black asterisks indicate PDF maxima and correspond to the preferred exit direction, while black diamonds correspond to the straight line that connects the two cluster centroids in

question. The preferred exit directions and shape of exit-angle PDFs are similar for the smaller (1.5σ) and larger (1.75σ) clusters. Thus, these features of the preferred transition paths appear to be statistically robust.

Note that the preferred directions of transition lean away from the line connecting origin and destination clusters, which is in general agreement with Kondrashov et al. (2004) results for the QG3 model. By applying the same analysis to the Lorenz (1963) model, Kondrashov et al. (2004) showed that such feature of exit-angle PDF can be explained by the geometry of the model's attractor and is related to its unstable and stable manifolds.

Inspection of Fig. 2 shows that the PDF for the PNA $-$ BNAO transition has one sharp maximum; in contrast, BNAO $-$ PNA is more bimodal in nature and less compact, while PNA $-$ RNA is quite diffuse and the noisiest one among the three highly significant transitions. These differences in the angular PDFs for pairwise transitions will have important consequences for long-range prediction, as we will see in Sect. 4.

3 Forecasting methodology

3.1 Predictants and predictors

For the purpose of studying transitions between weather regimes it is useful to consider any data point in terms of individual "events." For a particular transition between origin and destination regimes, the "event" is defined as the system's trajectory exiting the origin cluster on the next day, and entering the destination cluster sometime in the future, without entering any other regime. Any other data point will be a "non-event," which includes points not leaving the origin cluster next day, or reaching another regime than its specified destination. Thus, forecasting the transition means to classify all points in the origin cluster between "events" and "non-events." Our predictant is the assignment of a particular data point to one of two possible classes, using a set of predictors.

Following Deloncle et al. (2007), we base our predictors on the position and the velocity of a trajectory at a given data point. In order to exploit the preferential paths of regime transitions, as defined by Kondrashov et al. (2004) and in Sect. 2.4 here, we use the spherical coordinates (r, ψ,φ) centered on the origin regime's centroid and with the polar axis aligned with the preferred transition path, as defined by the PDF maxima for regime exit angles; see Fig. 2 here and in Deloncle et al. (2007). In these modified coordinates, a value of $\psi = \pi/2$ means the state vector is perfectly aligned with the preferred exit vector (marked by asterisks in Fig. 2), while a value of $\psi = -\pi/2$ indicates that it is in the opposite direction. The coordinate r is the distance to the center of the regime centroid. The cartesian velocity components (dx/dt, dy/dt, dz/dt) given by the model are expressed in the spherical coordinate system by (v_r, v_ψ,v_φ). In summary, our predictors are the six scalars associated, in the modified spherical coordinates, with the daily data points and their velocities: (r,ψ,φ,v_r,v_ψ,v_φ).

3.2 Random forests

Random forests is an extension of classification and regression trees (CART); it is described in greater detail in Breiman (2001). Deloncle et al. (2007) applied random forests to forecasting weather regime transitions in a simulated dataset of Kondrashov et al. (2004), obtained by running the QG3 model (Marshall and Molteni 1993; D'Andrea and Vautard 2001). We build on this work by considering weather regimes and transitions obtained from our NH observational dataset, described in Sect. 2.1, 2.2 and 2.3. The goal is to assign a given data point to a class based on information contained in a set of predictors. In our case, there are only two classes of binary response variable: either there is a designated regime transition next day or not, classified as an event or a non-event.

A random forest is an ensemble of CARTs. Each tree is based on a recursive partitioning of the data set. The partitioning of the data at each step is determined by minimizing a loss function that captures the amount of heterogeneity remaining after a partition is constructed; the sum of errors squared at that step is a special case of such a loss function.

With a categorical outcome, there are two popular loss functions. Both take as their arguments the proportions of categorical variable, computed for each node as the tree is grown. Because each partition of the data is chosen so that the loss function is minimized, the overall tree is a product of a node-by-node, stagewise minimization.

One loss function is the "entropy":

$$\phi(p) = p\log(p) + (1-p)\log(1-p), \qquad (1)$$

where p is the proportion of the categorical response in a given tree node. For example, if there are 20 events and 30 non-events in a node, p for events is 0.40. The other option is the Gini index:

$$\phi(p) = p(1-p). \qquad (2)$$

Both functions are concave, having minima at $p = 0$ and $p = 1$ and a maximum at $p = 0.5$. Entropy and the Gini index generally give very similar results when there are only two response categories. The Gini index is preferred in this case and we use it here.

Random forests is based on an ensemble of trees, and thus involves many passes through a given dataset. The method is motived in part by the well-known tendency of CARTs to overfit the data and it uses a bootstrap procedure to avoid this overfitting. Imagine there is a data set of 1,000 observations. Before a classification tree is grown, random forests takes a random sample of 1,000 with replacement from the data set. Included in the sample of 1,000 will likely be some duplicates and even triplicates. On average, about a third of the original 1,000 observations will by chance not be included in this sample. The classification tree is built with this sample, and the observations not included serve as a test sample for this tree.

To enhance the quality of the test samples, random forests alters the CART algorithm in one additional manner. When each partition of the data is determined, only a random sample (without replacement) of predictors is used. For example, if there are 50 predictors, any given partition might be determined by the best performing predictor in a random sample of 5. By growing trees that rely on random samples of predictors, independence across the trees is increased.

The sampling process is repeated for each classification tree in the random forest. Each time, a new random sample with replacement from the original data is drawn and serves as the training data. Each time, the data not chosen serves as the test data. The classifications of observations in the test data are used to arrive at fitted values. Each time an observation is a member of a test sample, the assigned class is recorded. Then, for each observation, a ``vote'' is taken over these results, and the class with the most votes is the class assigned, and become the fitted value for that observation.

This procedure represents a forecasting scheme, since classification votes are based only on the test data, and thus the final classification error does represent the actual forecasting error. The most credible measures of model performance (discussed in Sect. 4.2), such as forecasting errors in the contingency table and importance of predictors, are assessed from held-out test data as well.

Overfitting can be damaging in two ways. First, it can lead to overly optimistic measures of how the model fits the data. In addition, the statistical model itself may not generalize well to other datasets. Averaging across test sample results through the voting procedure mitigates overfitting. The averaging is made more effective because of the random sampling of predictors from tree to tree. Such averaging can effectively cancel out the idiosyncractic features of particular trees. Indeed, if the number of trees were to increase without limit, Breiman et al. (1984) have proved that the estimate of the classification error is statistically consistent in a rigorous mathematical sense: thus, random forests does not overfit.

Random forests provides several ways in which the relative costs of false negatives and false positives can be introduced. Perhaps the simplest way is to require more or less than a majority vote when averaging for the purposes of classification. For example, if false negatives are less costly than false positives, one might require a ``veto-resistant majority'' of two-thirds before a case is classified as a positive. We adopt an alternative method by oversampling the true positives when the training sample is drawn: the fitting process will then respond more to these observations. This is a better approach then altering the vote threshold because each tree can then adjust to the differential costs directly, rather than trying to adjust the results at the end.

The implications of using differential costs in random forests for forecasting regime transitions are discussed in the next section. In the meteorological literature, Roulston and Smith (2004) have recently studied the issues of cost/loss ratio, tolerance to false alarms, and the relative value of a probabilistic weather forecast. In short, classifying points by a simple-majority vote assumes that the costs of false positives and false negatives are the same; in practice, this may not be the best choice. The approach used in this paper systematically gives more weight to observations in which a transition does occur, so that if these events are improperly classified as non-events the costs are larger.

In summary, random forests have several features that make the algorithm attractive for our purposes. First, for the kinds of highly nonlinear and noisy relationships analyzed in this paper, there are no classifiers to date that consistently classify and forecast more accurately (Breiman 2001). Second, since random forests does not overfit, measures of forecasting performance and features of the model itself will generalize well to new random samples from the same population. Third, as we shall see in the next section, random forests provides instructive plots of the relationships between inputs and outputs, i.e. predictors and responses.

4 Forecasting results

4.1 Transition forecast skill

The random-forests method is now applied to forecast the three significant transitions identified in Sect. 2.3: PNA $\rightarrow$ BNAO, PNA $\rightarrow$ RNA, and BNAO $\rightarrow$ PNA (see again Fig. 2). For any given transition and point belonging to the origin cluster, we forecast the regime transition to the specified destination cluster, by classifying on each given day as an event or a non-event. As discussed in Sect. 3.2, random forests classifies each point

based on the held-out data, and so its results are cross-validated. To evaluate such forecasts it is useful to consider a contingency table, which provides information about user and model classiÞcation errors made in the process.

The rows of Table 3 condition on observed outcomes, and the columns condition on forecasted outcomes. The cells a and d on the diagonal correspond to the true forecasts; see also Table 1 in Deloncle et al. (2007). The off-diagonal cells correspond to one of two kinds of forecasting errors: the false positives and the false negatives. Forecasts falling in cell b, above the diagonal, are false positives: a transition is predicted but does not occur. Forecasts falling in cell c are false negatives: a transition is not predicted but it does occur. For each row, the proportions of errors shown in the far right column capture mistakes made by the model: given the actual outcome, what did the model forecast? For each column, the proportions of errors shown in the bottom row capture errors made when the model is put to use: given a forecast, what was the actual outcome?

The contingency Tables 4aÐc correspond to forecasting results for PNA $\times$ BNAO, PNA $\times$ RNA and BNAO $\times$ PNA transitions, respectively. In each cell, the Þrst number results when the cost ratio of false negatives to false positives is 1 to 1. This is the default ratio; it means that falsely forecasting a transition has the same cost as failing to forecast when a transition does occur.

The default ratio of 1:1 is arbitrary and may not be the most appropriate one should results such as the present ones be used in actual decision making. Cost ratios should be determined by the individuals or organizations intending to use the results, and random forests can use any prescribed ratio.

Arguably, false negatives in forecasting rare weather events are more costly than false positives. A good example of this cost was the very strong December 2003 storm over Western Europe: failure to correctly forecast transition between zonal and blocked ßow over Northern Europe resulted in very large economic losses.

Table 3 DeÞnition of a 2 $\times$ 2 contingency table

	Non-event forecasted	Event forecasted	Model error
Non-event observed	a (true negatives)	b (false alarms)	b/(a + b)
Event observed	c (misses)	d (true postives)	c/(c + d)
User Error	c/(a + c)	b/(b + d)	

The observations (actual category) of the points are in the rows and the forecasts in the columns. The values a, b, c, d represent the numbers of each case obtained on the assessment data set. Thus, true forecasts are on the diagonal and correspond to true negatives (a) and true positives (d). The incorrectly classiÞed points are off the diagonal and consist in the false positives (false alarms) (b) and the false negatives (misses) (c)

The second number in each cell of Table 4 corresponds to the cost ratio of 1:8. It implies that falsely forecasting a transition has one-eighth the cost of failing to forecast when a transition occurs, i.e. ÔÔmissingÕÕ a transition is 8 times more costly then a ÔÔfalse alarm.ÕÕ The cost ratio is introduced into the algorithm by an appropriate oversampling of the observed transitions when drawing the random sample to construct each tree (see Sect. 3.2). This ratio is imposed only as a target and, given the random sampling, the actual ratio of false negatives to false positives will only approximate the cost ratio speciÞed.

We turn Þrst to the PNA $\times$ BNAO case, which will illustrate patterns across Table 4aÐc. Transitions between the two regimes are fairly rare: they occur approximately 10% of the time. Hence forecasting such transitions is likely to be difÞcult. Indeed, when the cost ratio is 1:1, the model produces false positives (transitions that did not occur) only about 1% of the time, but false negatives (missed transitions) about 55% of the time. The result for user error, though, is larger for forecasts of no-transitions but smaller for forecasts of a transition occurring, 5 and 19%, respectively. Even more unbalanced results can be seen for the other two transitions (Table 4b, c).

Table 4 Contingency table with random-forests algorithm for the transition (a) PNA $\times$ BNAO, (b) PNA $\times$ RNA, and (c) BNAO $\times$ PNA; 3,000 trees were used with two variables selected at random evaluated for each split

	No transition forecasted	Transition forecasted	Model Error
(a) PNA $\times$ BNAO 1:1 (1:8) cost ratio			
No transition observed	1,103 (1025)	10 (88)	0.01 (0.08)
Transition observed	53 (11)	43 (85)	0.55 (0.11)
User error	0.05 (0.01)	0.19 (0.51)	
(b) PNA $\times$ RNA 1:1 (1:8) cost ratio			
No transition observed	1,095 (966)	7 (136)	0.01 (0.12)
Transition observed	80 (18)	27 (89)	0.75 (0.17)
User error	0.07 (0.02)	0.21 (0.60)	
(c) BNAO $\times$ PNA 1:1 (1:8) cost ratio			
No transition observed	679 (551)	8 (136)	0.01 (0.20)
Transition observed	79 (17)	33 (95)	0.71 (0.15)
User error	0.1 (0.03)	0.20 (0.59)	

Results are shown for two different ratios of false negatives to false positives: Þrst number is for the 1:1 default ratio, while the second number (in parentheses) is for a 1:8 cost ratio

An overall measure of forecasting skill is given by the Heidke Skill Score (HSS) (Von Storch and Zwiers 1999):

$$H = \frac{S - S_r}{N - S_r} \quad (3)$$

S is the number of correct forecasts, S_r the number of correct forecasts that would follow from just using the mode of the marginal distribution of the response, and N the number of observations. The score H can be used to compare improvement of forecasting skill with the available predictors over the skill expected by pure chance. A flawless model with given predictors would achieve a score of 1, whereas a value of 0 means the model predictors have no forecasting skill whatsoever.

An estimate of H can be given in terms of the numbers a,b,c,d introduced in Table 3:

$$H = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)}. \quad (4)$$

For our regime transition forecast in Table 4a, we find $H = 0.54$. Thus, random forests using a 1:1 cost ratio does 54% better when it takes the information contained in the six predictors into account than forecasts derived only from the marginal distribution of the PNA → BNAO transition.

The second number in each cell of Table 4 is obtained when using the target 1:8 cost ratio. The ratio of misses to false alarms is actually quite close to the intended value: 11:88 for PNA → BNAO, 18:136 for PNA → RNA, and 17:136 for BNAO → PNA. Using this ratio, random forests forecasts transitions far more successfully: the forecasts are correct 89% of the time for the PNA → BNAO transition, 83% of the time for the PNA → RNA transition, and 85% of the time for the BNAO → PNA transition.

The necessary tradeoff is that random forests now forecasts the absence of a transition somewhat less well than for the 1:1 cost ratio: forecasting skill for no transition has declined to 92, 88 and 80% for Table 4a–c, respectively, which is still quite high. One thus obtains precisely the tradeoff desired when the cost of false negatives is several times the cost of false positives.

4.2 Relative importance of predictors

To learn which predictors contribute most to forecasting skill, we use the importance plots in Fig. 3. For each plot the horizontal axis represents the decrease in the proportion

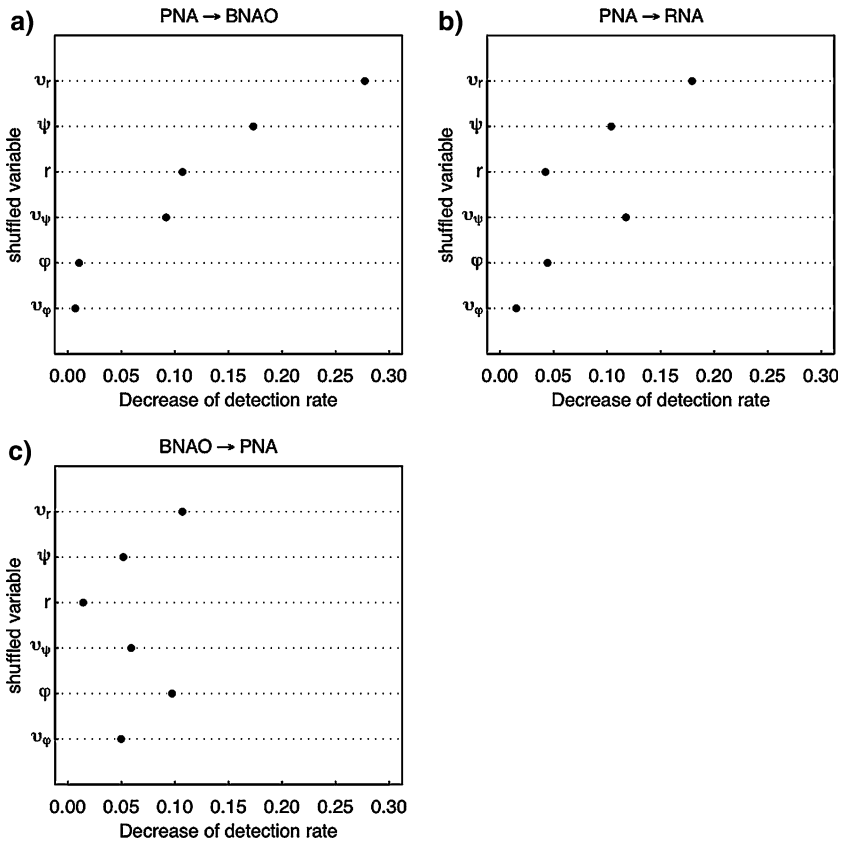

Fig. 3 Relative importance of the predictors: a PNA → BNAO, b PNA → RNA, and c BNAO → PNA. The plot shows the decrease in detection rate when a variable is shuffled; this decline measures the importance of each variable in the forecasting process

of transitions correctly forecasted when the variable on the vertical axis is prevented from participating in the forecasting process. More specifically, the classification errors are still based on majority vote over the trees in held-out data, but now with the values of each predictor being shuffled in a random order in those data. The overall increase in forecasting errors (more specifically model error in Tables 3, 4) when each predictor is shuffled shows how much forecasting skill is lost. If a particular predictor is not helpful at all, there is no loss in forecasting skill and vice-versa.

The results are shown in Fig. 3 for the forecasts that used the target cost ratio of 1:8 (see above); similar results were obtained with the default cost ratio of 1:1 (not shown). For all transitions two or three predictors are much more important than the rest, and the radial velocity v_r is the most important one of all. Deloncle et al. (2007) showed, for a QG3 model simulation, that the importance of v_r is due to transition points moving out of the cluster with a radial velocity larger that the radial velocity of the other points. The angular predictor ψ is the second most important one for PNA → BNAO and the third one for PNA → RNA; we recall that ψ measures alignment with the preferred exit angle. The PDF peak of regime exit angles for the PNA → BNAO transition is much narrower than for the other two; see again Fig. 2. This may explain why ψ is somewhat less important for the PNA → RNA and BNAO → PNA transitions, with φ being the second most important predictor for BNAO → PNA. These two transitions are more complex, which leads to more predictors being of some importance than for the PNA → BNAO transition. We will return to the question of difference in predictor importance when discussing the partial-dependence plots below.

Partial-dependence plots (Hastie et al. 2001) show the relationship between a given predictor and the response averaged over the joint values of the other predictors. In effect, the role of the other predictors is integrated out over the tree structure. Let us say predictor variable v has N distinct values. We construct N new datasets, each of which has the same values for all predictors except v, while the value of v is fixed and characterizes the set under consideration. Then we forecast our response variable with random forests for each of these N datasets, and determine proportions of trees that respectively forecast event and non-event. It is common to plot the results of this calculation against the N values of v, in log-odds units or "logits". The log-odds equal the logarithm of the ratio between a probability p and its complement $q = 1 - p$, in the present case between that of a transition having occurred for each dataset and its complement.

The partial dependence plots for the two leading predictors in each transition are plotted in Fig. 4. On the vertical axis is the likelihood of the transition in question, in logits, while on the horizontal axis are values of the predictor. The dependency curves are rather smooth, with a clear maximum at which a transition is most likely to occur. There is a consistently rapid increase in the chances of

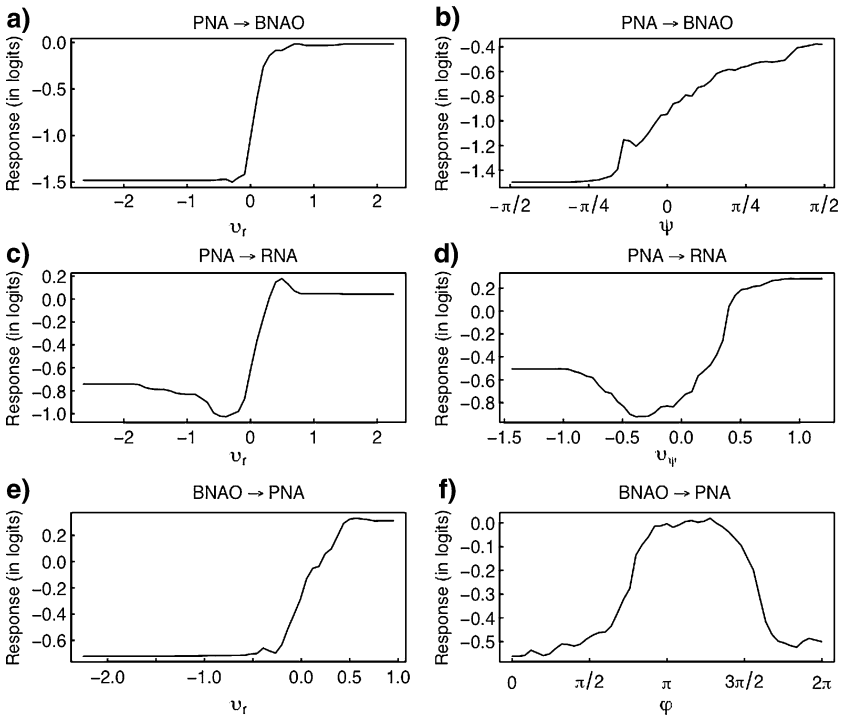

Fig. 4 Partial-dependence plots for the two most important predictors: a v_r and b ψ for PNA → BNAO, c v_r and d v_ψ for PNA → RNA, and e v_r and f φ for BNAO → PNA. The ordinate measures the marginal probability of event detection with respect to the high-impact variable; the velocities are in normalized units per day

a transition as the predictor v_r crosses the origin, that is as the point starts moving outwards from the regime. Such behavior of v_r is in agreement with results of Deloncle et al. (2007) for the QG3 model and with its importance (see Fig. 3). Likewise, there are rapid increases in the chances of a transition when v_ψ (shown for PNA → RNA in Fig. 4d) crosses its zero, indicating that the point starts moving towards the preferred exit direction.

The response function for ψ (shown for PNA → BNAO in Fig. 4b, but common for other transitions as well) changes far less dramatically, but reaches its maximum value at $\pi/2$ for the preferred exit direction, as expected. Among all the transitions, PNA → BNAO has the narrowest range of important values of ψ. This is consistent with having a narrower PDF peak of regime exit angles than for PNA → RNA and BNAO → PNA, and with the latter two having a lower detection rate of transitions, as shown in Table 3.

Taken together, these results indicate the velocity vector along the trajectory has to be directed outwards for transition to occur, and to lie close to the preferred transition path, located by the maximum in the PDF of regime exit angles. These results are consistent with the conclusions of Kondrashov et al. (2004, 2006) and Deloncle et al. (2007); they confirm the inhomogeneity of the transitions in phase space and the crucial role of a preferred direction as a useful predictor.

The broad peak of ρ near π for the BNAO → PNA transition (Fig. 4f) and the high importance of this predictor (Fig. 3c), indicate how the exit-angle PDF's shape can influence the prediction process. This feature is probably due to the complex pattern of the exit-angle PDF, with its pronounced bimodality and the extensive ridge of the main peak seen in Fig. 2d. How the unique features of exit-angle PDF affect the importance of ρ and ψ as predictors is revealed further by examining the scatter plot of transition angles, shown in Fig. 5. The PNA → BNAO transition has a very uniform distribution in ρ, but is extremely one-sided with respect to ψ, in agreement with having peak at $\pi/2$ in Fig. 4a. The PNA → RNA transition is less compact in ψ, which leads to a lower importance of this predictor, as seen in Fig. 3b. The transition BNAO → PNA stands on its own: its angle distribution is broader in ψ, but quite compact in ρ, with an apparent peak around $\rho = \pi$, in agreement with Figs. 3c and 4f.

4.3 Regime-based medium-range forecasts

We now use the random forests to construct medium-range weather forecasts based on anomalies in the subspace spanned by the three leading EOFs. For a transition from Regime A to Regime B, we will consider only Regime A episodes after which such a transition does take place and we initiate forecasts on all days of any such episode. The current study focuses on forecasting single-regime outcomes and not multiple transitions. When using such a subsample of initial states, our statistical forecasts will still be penalized for both misses and false alarms, but rewarded

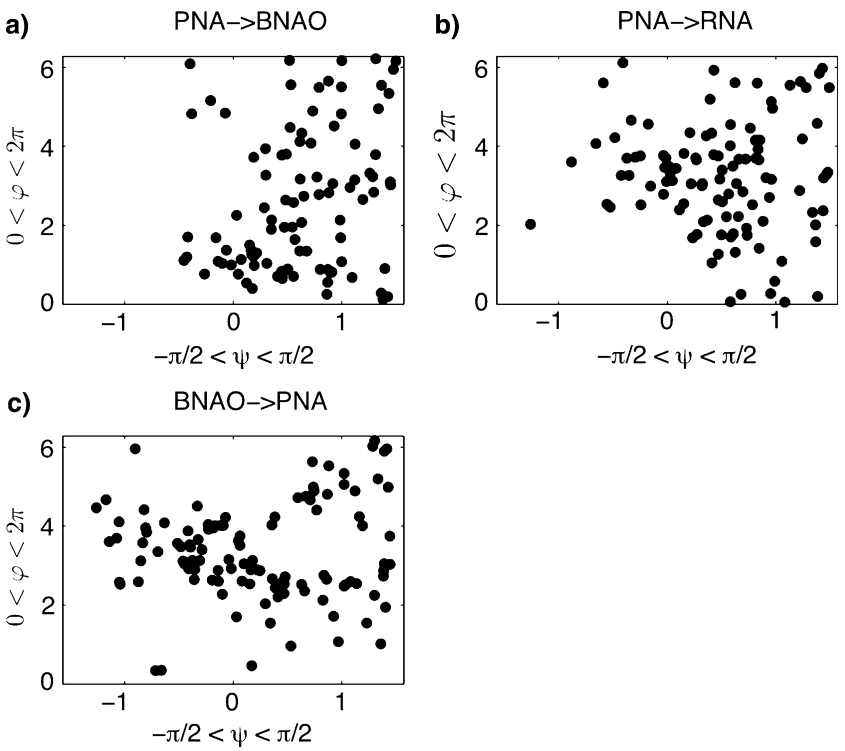

Fig. 5 Scatter plot of regime exit angles in transformed coordinates for the 1–5 clusters: a PNA → BNAO, b PNA → RNA, and c BNAO → PNA. The value of $\psi = \pi/2$ corresponds to the exit vector being aligned with the preferred direction path (the PDF maxima shown in Fig. 2b–d)

for pinpointing correctly transition days. Including the rest of the regime episodes in the forecast would penalize the method for false alarms only. Still, Deloncle et al. (2007) did carry out multiple-transition predictions for the QG3 model with promising results, and we intend to evaluate such predictions for observational data in future studies.

Given all regime episodes (i.e. passages through Regime A) for which an actual transition to Regime B did occur, we count all the days during such an episode as part of our sample for the transition A → B. There are thus a total of 405 days for 96 transitions PNA → BNAO, 461 days for 107 transitions PNA → RNA, and 383 days for 112 BNAO → PNA transitions, respectively (see Sect. 2.3). The total number of forecasted transitions that are included in this study sample are 43(114) for PNA → BNAO, 28(121) for PNA → RNA, and 33(133) for BNAO → PNA transition, where the numbers correspond to the 1:1(1:8), cost ratio respectively; compare these numbers with those in Table 4, for the number of transitions forecasted for all regime episodes.

Recall from Sects. 3.2 and 4.1 that random forests provide us a cross-validated forecast on whether the transition will take place on any given day within Regime A episodes or not. In addition, we will use statistical information on the mean residence time, τ_A, in the origin Regime A and τ_B in the destination Regime B, as well as mean transition time τ_{AB} between these regimes (see Table 1). These time scales, along with the transition forecasts, provide us information on where we expect the trajectory of the system to be at any given lead time, as sketched in Fig. 6.

At initial forecast time $t = t_0$, assume we have been in Regime A for a time τ_A^0. The standard Markov chain forecast (Ghil 1987; Mo and Ghil 1987; Fraedrich 1988; Kimoto and Ghil 1993b) shown in Fig. 6a, assumes the trajectory will persist in Regime A for a time $\tau_A - \tau_A^0$, where τ_A^0 is the time already spent in Regime A at epoch $t = t_0$, and will then transition to Regime B with probability p_{AB} or to Regime C with probability p_{AC}. The expected trajectory is shown (double solid line) as a weighted

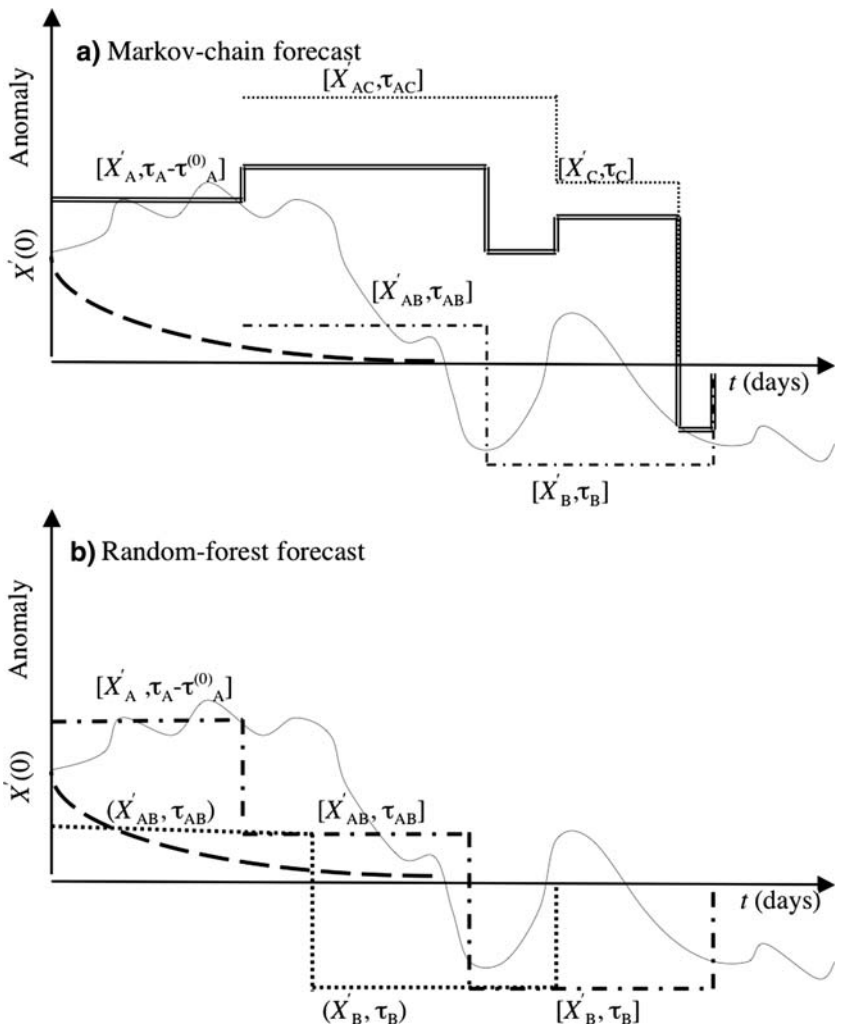

Fig. 6 Sketch of statistical forecast scheme: a Markov chain, and b random forests. Here X'_A, X'_B and so on indicate mean anomalies in regimes A, B, etc., while τ_A, τ_B and so on indicate expected residence times in these regimes; indices AB and AC indicate the corresponding transitions. Markov chain: double solid line is for anomaly forecast originating in Regime A; dotted and dash-dotted lines are for Regime C and Regime B transitions, respectively. Random forests: dotted line is anomaly trajectory when transition to Regime B is forecast, while dash-dotted line is a "no-transition" forecast. In both panels a) and b), the damped-persistence forecast is shown by the dashed line, and the actual trajectory is the light solid line

average with these two weights, p_{AB} and p_{AC}, between the mean anomalies X^{ϕ}_{AB} for time τ_{AB} (dash-dotted line) and of X^{ϕ}_{B} for time τ_B, on the one hand, and of X^{ϕ}_{AC} for time τ_{AC} (dashed line) and X^{ϕ}_{C} for time τ_C, on the other; both the trajectory passing through Regime B and that passing through Regime C revert to climatology after time $\tau_{AB} + \tau_B$ or $\tau_{AC} + \tau_C$, respectively. The ÔÔmeanÕÕ transition patterns X^{ϕ}_{AB} and X^{ϕ}_{AC} between regimes, like τ_{AB} and τ_{AC}, are calculated from the observed history of transitions. The light solid line shows the actual evolution of the anomalies in the NH reanalysis.

The random-forests forecast shown in Fig. 6b, like the Markov-chain forecast in Fig. 6a, uses the actual anomaly $X(t_0)$ at initial forecast time; afterwards, though, the predicted trajectories differ. If there is a forecast of transition to Regime B, the expected trajectory follows, from this epoch on, X^{ϕ}_{AB} for a time τ_{AB} and, afterwards, X^{ϕ}_{B} for a time τ_B (dotted line). This trajectory, too, reverts to climatology after time $\tau_{AB} + \tau_B$. If no transition to Regime B is forecast at $t = t_0$, the forecast assumes that the trajectory (dash-dotted line) will remain in Regime A for a time $\tau_A - \tau^0_A$, followed by X^{ϕ}_{AB} for a time τ_{AB} and, afterwards, X^{ϕ}_{B} for a time τ_B, Þnally returning to climatology. The damped-persistence forecast at time $t = t_0$ is shown by the dashed line in both panels.

For baseline comparison, we use damped-persistence forecasts based on Þtting an AR(1) model to the time series of each of the three leading PCs; this model damps initial anomalies exponentially to climatology, with the corresponding decorrelation time scales, which are 15 days for all three leading PCs. Damped persistence is the optimal persistence forecast in least-square error and is quite difÞcult to beat for medium-to-long-range weather forecasts.

In Fig. 7 we present the error variance of forecasts, normalized by the climatological variance, in the subspace of the three leading EOFs. Since we focus here on predicting the state of the atmosphere being in a particular regime, but not its detailed behavior, both forecast and veriÞcation time series are averaged over a time interval equal to a given lead time, i.e., we concentrate on the mean evolution of errors. For short lead times damped persistence provides a more accurate forecast, since the systemÕs trajectory ÔÔremembersÕÕ the initial data, albeit less and less precisely, while the random-forests prediction relies on average patterns.

The random-forests forecast, however, becomes more accurate than damped persistence at lead times of roughly 7–10 days, when information about the detailed initial state has been lost; this behavior holds for all three transitions. The mean-square error of the damped-persistence forecast can exceed the climatological variance in the medium range but, as expected, all forecast errors slowly converge to the climatological value as the lead time exceeds roughly one month. The quadratic forecast error depends but little on the cost ratio considered, although for the default ratio of 1:1 (dash-dotted line) it is slightly higher than for the 1:8 cost ratio (heavy solid line in Fig. 7) over a short lead time, of up to 5 days.

The improvement of our random-forests forecasts over both climatology and damped persistenceÑbeyond

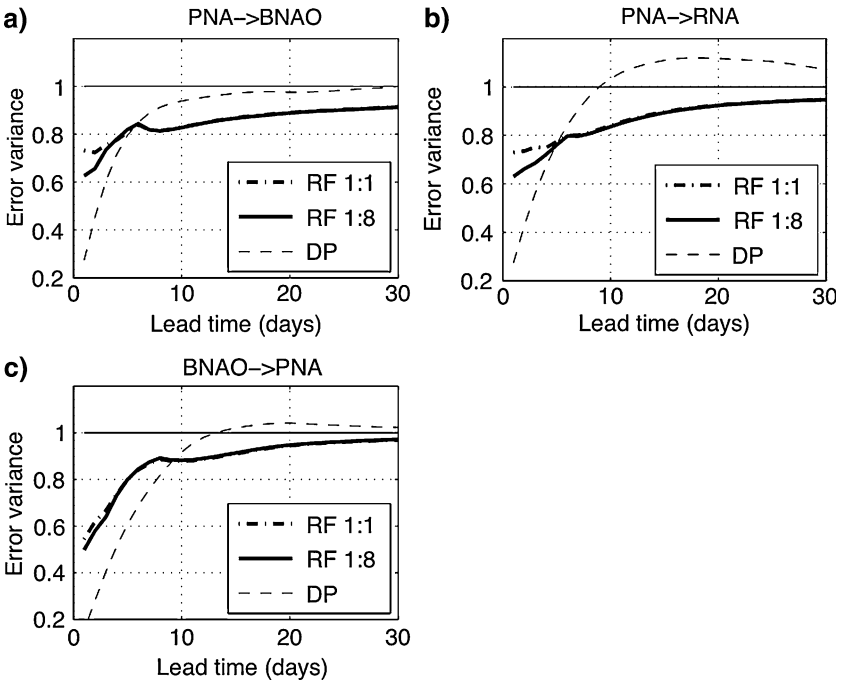

Fig. 7 Normalized error variance of forecasts as a function of lead time: a PNA “ BNAO, b PNA “ RNA, and c BNAO “ PNA. Random forests: heavy dash-dotted line for default cost ratio 1:1 and heavy solid line for 1:8 cost ratio; damped persistence (DP) forecasts (light dashed line) and climatology forecasts (light solid)

10 days and out to a month, especially for the PNA " BNAO transitionÑis quite noticeable. Winkler et al. (2001) compared the skill of their linear inverse model (LIM) forecasts, suitably smoothed, with those of NCEPÕs medium-range forecast model and found it to be competitive. At the same time, our statistical forecasts seem to be at least as good when considering a particular, statistically signiÞcant transition, conditioned on a subset of initial states within a given regime; when started in the PNA regime, our forecast are better than theirs: compare our Fig. 7 with their Fig. 8, which does not include damped persistence among the skill scores to beat.

5 Summary and discussion

Kondrashov et al. (2004) showed that certain weather regime transitions in the QG3 model are characterized by preferred exit directions in the modelÕs phase space. Deloncle et al. (2007) in turn used predictors connected to these preferred directions to forecast the regime breaks extracted from a long model simulation by relying on two statistical methods. Encouraged by their results, we studied medium- to long-range prediction of large-scale, mid-latitude ßow patterns in NH atmospheric data using the NCEP-NCAR reanalysis dataset for the 55 three-month winters of 1948Ð2003 (Sect. 2.1). In Sect. 2.2 we showed that there are three signiÞcant weather regimes in a low-dimensional subspace spanned by the three leading EOFs of the datasetÕs variability. The three clusters (see Fig. 1 and Table 1) are PNAÑthe PNA pattern, BNAOÑthe blocked phase of the NAO, and RNAÑthe (approximately) reversed PNA pattern. These three regimes are in good agreement with the observational results of Cheng and Wallace (1993), using hierarchical clustering, and those of Smyth et al. (1999), using a Gaussian mixture model for a NH dataset of 44 winters (1949Ð1993).

In Sect. 2.3 we identiÞed three signiÞcant transitions in the Markov chain between these weather regimes (Ghil 1987; Mo and Ghil 1988; Molteni et al. 1990; Kimoto and Ghil 1993b): PNA " BNAO, PNA " RNA and BNAO " PNA; see Table 2. Following Kondrashov et al. (2004), we determined preferential directions in which the systemÕs trajectory in phase space leaves the regimes by computing PDFs of the exit angles on the unit sphere centered on each clusterÕs centroid (Sect. 2.4; cf. Fig. 2). Based on the results of Deloncle et al. (2007), we used this information about preferred exit directions to construct a set of predictors in Sect. 3.1, and applied the random-forests method (Sect. 3.2) to study predictability of signiÞcant regime transitions in Sect. 4.

Predicting these regime transitions presents a challenge because they are relatively rare events; see Tables 2 and 3. For such events, the detection rate is likely to be relatively low or, in other words, the statistical model error can be quite high. However, when the relative costs of failing to detect rare events is high, random forests can improve forecasting skill for those events. One consequence is a decrease in forecasting skill for the common events, but this is quite tolerable when errors in forecasting common events are less costly.

The choice of which cost ratio to use, whether 1:1 for failures to predict vs. false alarms or different, depends on the context in which the forecasting results will be used; see also Roulston and Smith (2004). It is often the case that a failure to correctly forecast rare and extreme weather events, such as the December 2003 storm over Western Europe, can lead to very large costs. Considerable forecast skill for rare transitions was obtained even when the cost ratio was taken equal to 1:1; this skill was further increased when using a cost ratio of 1:8, cf. Table 4.

Computing importance (Fig. 3) and partial-dependence (Fig. 4) plots conÞrmed that a transition is more likely to occur for trajectory segments that are aligned with the preferred regime exit directions. In partial agreement with the results of Deloncle et al. (2007), the radial velocity component v_r along such a direction is the dominant predictor for all the highly signiÞcant transitions in our NH data. We also found that the inßuence of ψ, the angle formed by the exit vector with the preferred exit direction, is greater when the preferred transition paths are conÞned within a fairly sharp solid angle, like for the PNA " BNAO transition, and smaller for more diffuse PDFs of regime exit angles (Figs. 2, 5). Unique features of the exit-angle PDFs, such as bimodality and an extended ridge for the main peak in the case of the BNAO " PNA transition, can lead to an increased significance of the angle ψ, complementary to ψ in our spherical coordinates.

Given the practical interest of medium- to long-range forecasts, we went further than Deloncle et al. (2007) in formulating an algorithm that produces not only categorical forecasts of speciÞc regime transitions (event) or not (non-event), but actual expected trajectories up to 30 days. This algorithm takes into accountÑbesides the categorical, random-forests based forecast of a transitionÑalso the anomalies and residence times associated with the regimes and the transitions between them (Fig. 6). Comparing the skill of this novel algorithm with the baseline standard of damped persistence (Fig. 7) yields very encouraging results for the single-transition case.

Noting these results, though, is not the same as proposing the present algorithm, as it stands, for operational use. A major restriction for its practical use is the algorithmÕs conditional nature: it beats damped persistence, and probably LIM (Winkler et al. 2001), only for initial states within a given regime, especially PNA. Still, there are

various ways in which such an algorithm can be combined with detailed dynamical forecasts of numerical weather prediction models, and even with LIM predictions, to improve upon the current regime transition skill of the numerical methods. At least one way to do so is to ÔÔassimilateÕÕ the statistical forecast into the numerical model, and thus provide large-scale guidance to the latter (Strong et al. 1995).

These results provide further support for the Legras and Ghil (1985) conjecture that (a) certain atmospheric-ßow regimes are associated with unstable Þxed points in the ßowsÕ phase space; and, hence, (2) exit from such regimes and subsequent transitions to other regimes originate along preferred directions of unstable growth of perturbations.

Acknowledgments We are grateful to Padhraic J. Smyth, who provided the code for the Gaussian mixture model, and to four anonymous reviewers, whose constructive comments greatly improved the presentation. Ann Henderson-Sellers attracted our attention to the work of Palmer et al. (2007), and Tim Palmer provided a preprint thereof. Our study was supported by DOE grant ER63251 (MG and DK) and by NSF grant SES04-37169 (RB and JS).

References

Breiman L (2001) Random forests. Mach Learn 45:5Ð32
Breiman L, Friedman J, Olshen R, Stone C (1984) ClassiÞcation and regression trees. Wadsworth Press, Monterey, 368 pp
Chang KI, Ghil M, Ide K, Lai CCA (2001) Transition to aperiodic variability in a wind-driven double-gyre circulation model. J Phys Oceanogr 31:1260Ð1286
Cheng XH, Wallace JM (1993) Cluster-analysis of the Northern-Hemisphere wintertime 500-hPa height Þeld spatial patterns. J Atmos Sci 50: 2674Ð2696
Crommelin DT (2002) Homoclinic dynamics: a scenario for atmospheric ultra-low-frequency variability. J Atmos Sci 59:1533Ð1549
Crommelin DT (2003) Regime transitions and heteroclinic connections in a barotropic atmosphere. J Atmos Sci 60:229Ð246
Crommelin DT (2004) Observed nondiffusive dynamics in large-scale atmospheric ßow. J Atmos Sci 61:2384Ð2396
DÕAndrea F, Vautard R (2001) Extratropical low-frequency variability as a low-dimensional problem. Part I: A simpliÞed model. Q J R Meteorol Soc 127:1357Ð1374
DÕAndrea F, Vautard R (2002) Extratropical low-frequency variability as a low-dimensional problem. Part II: Stationarity and stability of large-scale equilibria. Q J R Meteorol Soc 128:1059Ð1073
DÕAndrea F, Tibaldi S, Blackburn M, et al. (1998) Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979Ð1988. Clim Dyn 14:385Ð407
Deloncle A, Berk R, DÕAndrea F, Ghil M (2007) Weather regime prediction using statistical learning. J Atmos Sci 64:1619Ð1635
Fraedrich K (1988) El Nin˜o-Southern Oscillation predictability. Mon Wea Rev 116:1001Ð1012
Franzke C, Majda AJ, Vanden-Eijnden E (2005) Low-order stochastic mode reduction for a realistic barotropic model climate. J Atmos Sci 62:1722Ð1745
Ghil M (1987) Dynamics, statistics and predictability of planetary ßow regimes. In: Nicolis C, Nicolis G (eds) Irreversible phenomena and dynamical systems analysis in the geosciences. D. Reidel, Dordrecht, pp 241Ð283
Ghil M, Childress S (1987) Topics in geophysical ßuid dynamics: atmospheric dynamics, dynamo theory and climate dynamics. Springer, New York, 485 pp
Ghil M, Robertson AW (2002) ÔÔWavesÕÕ vs. ÔÔparticlesÕÕ in the atmosphereÕs phase space: a pathway to long-range forecasting? Proc Natl Acad Sci 99(Suppl. 1):2493Ð2500
Ghil, M, Kimoto M, Neelin JD (1991) Nonlinear dynamics and predictability in the atmospheric sciences. Rev Geophys, Supplement (U.S. NatÕl Rept. to IntÕl Union of Geodesy & Geophys 1987Ð1990), 29(S):46Ð55, 10.1029/91RG0071
Guckenheimer J, Holmes P (1983) Nonlinear oscillations, dynamical systems, and bifurcations of vector Þelds. Springer, New York, 453 pp
Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer, New York, 552 pp
Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437Ð471
Kimoto M, Ghil M (1993a) Multiple ßow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. J Atmos Sci 50:2625Ð2643
Kimoto M, Ghil M (1993b) Multiple ßow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions. J Atmos Sci 50:2645Ð2673
Kimoto M, Mukougawa H, Yoden S (1992) Medium-range forecast skill variation and blocking transition: a case study. Mon Weather Rev 120:1616Ð1627
Kondrashov D, Ide K, Ghil M (2004) Weather regimes and preferred transition paths in a three-level quasi-gesotrophic model. J Atmos Sci 61:568Ð587
Kondrashov D, Kravtsov S, Ghil M (2005) A hierarchy of data-based ENSO models. J Clim 18:4425Ð4444
Kondrashov D, Kravtsov S, Ghil M (2006) Empirical mode reduction in a model of extratropical low-frequency variability. J Atmos Sci 63:1859Ð1877
Kravtsov S, Kondrashov D, Ghil M (2005) Multiple regression modeling of nonlinear processes: derivation and applications to climatic variability. J Clim 18:4404Ð4424
Legras B, Ghil M (1985) Persistent anomalies, blocking and variations in atmospheric predictability. J Atmos Sci 42:433Ð471
Lorenz EN (1963) Deterministic nonperiodic ßow. J Atmos Sci 20:130Ð141
Majda AJ, Timofeyev I, Vanden-Eijnden E (2003) Systematic strategies for stochastic mode reduction in climate. J Atmos Sci 60:1705Ð1722
Marshall J, Molteni F (1993) Towards a dynamical understanding of planetary-scale ßow regimes. J Atmos Sci 50:1792Ð1818
Meacham SP (2000) Low-frequency variability in the wind-driven circulation. J Phys Oceanogr 30:269Ð293
Mo K, Ghil M (1987) Statistics and dynamics of persistent anomalies. J Atmos Sci 44:877Ð901
Mo K, Ghil M (1988) Cluster analysis of multiple planetary ßow regimes. J Geophys Res 93D:10927Ð10952
Molteni F, Tibaldi S, Palmer TN (1990) Regimes in the wintertime circulation over northern extratropics. 1. observational evidence. Q J R Meteorol Soc 116:31Ð67
Monahan AH, Pandolfo L, Fyfe JC (2001) The preferred structure of variability of the Northern Hemisphere atmospheric circulation. Geophys Res Lett 28:1019Ð1022
Nadiga BT, Luce BP (2001) Global bifurcation of Shilnikov type in a double-gyre ocean model. J Phys Oceanogr 31:2669Ð2690
Namias J (1953) Thirty-day forecasting: A review of a ten-year experiment. Meteorol Monogr 2(6):83
Namias J (1968) Long-range weather forecasting Ð history, current status and outlook. Bull Am Meteorol Soc 49:438Ð470

Palmer TN, Doblas-Reyes FJ, Weisheimer A, Rodwell M (2007) Towards seamless prediction: calibration of climate-change projections using seasonal forecasts. Bull Am Meteorol Soc (submitted)

Pasmanter RA, Timmermann A (2003) Cyclic Markov chains with an application to an intermediate ENSO model. Nonlinear Proc Geophys 10:197Ð210

Pelly JL, Hoskins BJ (2003) How well does the ECMWF Ensemble Prediction System predict blocking. Q J R Meteorol Soc 129:1683Ð1702

Roulston MS, Smith LA (2004) The boy who cried wolf revisited: the impact of false alarm intolerance on cost-loss scenarios. Weather Forecast 19(2):391Ð397

Selten FM, Branstator G (2004) Preferred regime transition routes and evidence for an unstable periodic orbit in a baroclinic model. J Atmos Sci 61:2267Ð2282

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London, 175 pp

Simonnet E, Ghil M, Ide K, Temam R, Wang S (2003a) Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part I: Steady-state solutions. J Phys Oceanogr 33:712Ð728

Simonnet E, Ghil M, Ide K, Temam R, Wang S (2003b) Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part II: Time-dependent solutions. J Phys Oceanogr 33:729Ð752

Simonnet E, Ghil M, Dijkstra HA (2005) Homoclinic bifurcations in the quasi-geostrophic double-gyre circulation. J Mar Res 63:931Ð956

Smyth P, Ide K, Ghil M (1999) Multiple regimes in Northern Hemisphere height Þelds via mixture model clustering. J Atmos Sci 56:3704Ð3723

Strong CM, Jin Ff, Ghil M (1995) Intraseasonal oscillations in a barotropic model with annual cycle, and their predictability. J Atmos Sci 52:2627Ð2642

Trevisan A (1995) Statistical properties of predictability from atmospheric analogs and the existence of multiple ßow regimes. J Atmos Sci 52:3577Ð3592

Vannitsem S (2001) Toward a phase-space cartography of the short- and medium-range predictability of weather regimes. Tellus 53A:56Ð73

Vautard R, Mo K, Ghil M (1990) Statistical signiÞcance test for transition matrices of atmospheric Markov chains. J Atmos Sci 47:1926Ð1931

Von Storch H, Zwiers S (1999) Statistical analysis in climate research. Cambridge University Press, London, 484 pp

Weeks ER, Tian Y, Urbach JS, Ide K, Swinney HL, Ghil M (1997) Transitions between blocked and zonal ßows in a rotating annulus with topography. Science 278:1598Ð1601

Winkler CR, Newman M, Sardeshmukh P (2001) A linear model of wintertime low-frequency variability. Part I: formulation and forecast skill. J Clim 14:4474Ð4494