Weather regime prediction using statistical learning

Axel Deloncle, R. Berk, F. d'Andrea, M. Ghil

Weather Regime Prediction Using Statistical Learning

A. DELONCLE, R. BERK,* F. D'ANDREA, AND M. GHIL+

Département Terre–Atmosphère–Océan, and Laboratoire de Météorologie Dynamique du CNRS/IPSL, Ecole Normale Supérieure, Paris, France

(Manuscript received 7 October 2005, in final form 11 July 2006)

ABSTRACT

Two novel statistical methods are applied to the prediction of transitions between weather regimes. The methods are tested using a long, 6000-day simulation of a three-layer, quasigeostrophic (QG3) model on the sphere at T21 resolution.

The two methods are the *k* nearest neighbor classifier and the random forest method. Both methods are widely used in statistical classification and machine learning; they are applied here to forecast the break of a regime and subsequent onset of another one. The QG3 model has been previously shown to possess realistic weather regimes in its northern hemisphere and preferred transitions between these have been determined. The two methods are applied to the three more robust transitions; they both demonstrate a skill of 35%–40% better than random and are thus encouraging for use on real data. Moreover, the random forest method allows one, while keeping the overall skill unchanged, to efficiently adjust the ratio of correctly predicted transitions to false alarms.

A long-standing conjecture has associated regime breaks and preferred transitions with distinct directions in the reduced model phase space spanned by a few leading empirical orthogonal functions of its variability. Sensitivity studies for several predictors confirm the crucial influence of the exit angle on a preferred transition path. The present results thus support the paradigm of multiple weather regimes and their association with unstable fixed points of atmospheric dynamics.

1. Introduction and motivation

The low-frequency intraseasonal variability of the extratropical atmosphere involves phenomena with time scales that are longer than the baroclinic-eddy life cycles and shorter than the change of seasons, that is, 10 to 100 days. This variability is characterized by the existence of large-scale persistent and recurrent flow patterns called weather regimes (Ghil and Robertson 2002; Molteni 2002). Several regimes have been identified in a consistent way by using diverse statistical and combined stochastic-dynamical methods.

These studies generally use advanced multivariate statistical methods to identify significant deviations from Gaussianity in the probability density function (PDF) of relevant dynamical variables in a reduced phase space. Such studies have been carried out using observed atmospheric data, as well as output from numerical models. The results do vary to a certain extent, as summarized by Ghil and Robertson (2002), according to the nature and length of the dataset, as well as to its preparation. For instance, when using monthly mean data, Stephenson et al. (2004) find that the existence of separate *climate* regimes is elusive. This is not too surprising, given the small number of months in their dataset.

In spite of these difficulties, agreement on at least a minimal set of weather regimes—extracted from daily, rather than monthly data—has emerged in the community (Cheng and Wallace 1993; Smyth et al. 1999). A review of classification methods and results is included, for example, in Ghil and Robertson (2002) and Molteni (2002).

The concept of weather regimes has been used successfully in different fields of the atmospheric sciences, from predictability through the downscaling of general

* Additional affiliation: Department of Statistics, University of California, Los Angeles, Los Angeles, California.

+ Additional affiliation: Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California.

Corresponding author address: Axel Deloncle, LadHyX, CNRS, Ecole Polytechnique, F-91128 Palaiseau CEDEX, France. E-mail: axel.deloncle@ladhyx.polytechnique.fr

DOI: 10.1175/JAS3918.1

circulation model (GCM) results to climate change impact assessment. In this paper, we examine the possibility that, because of their persistence, weather regimes provide a coarse-grained, predictable component of the atmosphere (Mo and Ghil 1988; Ghil et al. 1991) capable of circumventing the deterministic predictability barrier of 10 to 15 days (Lorenz 1969).

Markov chains of multiple regimes have been shown to provide extended predictability, at the cost of less detail in the predicted variables (Fraedrich and Klauss 1983; Ghil and Robertson 2002). Moreover, the most advanced numerical weather prediction models still have problems at forecasting regime transitions. This shortcoming has been investigated, for example, in the context of atmospheric blocking inception. Tibaldi and Molteni (1990) showed that much of the forecast error of the European Centre for Medium-Range Weather Forecasts (ECMWF) forecast model was due to its inability to enter a blocked state 3–4 days into the forecast. This difficulty reflected a general underestimation of blocking frequency in GCMs (D'Andrea et al. 1998). Although much progress has been made since, forecasts of blocking inception still have no skill starting from a lead time of 6 days (Pelly and Hoskins 2003).

The purpose of this article is to present a novel strategy, based on advanced statistical methods, to forecast regime breaks and subsequent onsets. Using weather regimes as a predictable component of the flow relies on theoretical considerations. Weather regimes are often explained as the manifestation of nonlinear equilibria in the slow manifold of the flow, and high-frequency transients can be seen as a stochastic perturbation of this underlying slow movement. For a discussion see Ghil and Robertson (2002), Branstator and Berner (2005), and references therein.

An alternative theory for the non-Gaussianity of atmospheric PDFs relies on the hypothesis that the stochastic forcing due to high-frequency transients depends, in fact, on the large-scale flow (Sura et al. 2005). This state dependence may also be connected to the often discussed eddy feedback on the large-scale flow (Robinson 2000; Kravtsov et al. 2005). Non-Gaussianity could then arise from the interaction of multiplicative stochastic noise with linear, or quasi-linear, large-scale dynamics. The multiplicative-noise paradigm and the multiple-equilibria one have distinct, and almost opposite, implications in terms of the system's predictability. The latter paradigm postulates the existence of nonlinear, large-scale dynamics with low or intermediate dimensionality, and enhanced predictability of certain major features of the flow; instabilities of the large-scale equilibria appear to be associated with preferential directions of system evolution (Legras and Ghil 1985). In the multiplicative-noise paradigm, there are no multiple equilibria, only an enhancement of the noise near the unique equilibrium; consequently, no preferential directions of evolution exist. Sura et al. (2005) provide a very clear discussion of the predictability properties associated with the two paradigms.

The goal of this work is to show the applicability and promise of regime transition forecasts. Aside from their potentially practical utility, these forecasts enhance the credibility of the multiple-equilibria paradigm. It is not our goal here to establish an operational forecast system. For this reason, we work with the output of an intermediate-complexity, quasigeostrophic, three-layer (QG3) model introduced by Marshall and Molteni (1993). This model has been widely used to investigate the Northern Hemisphere atmosphere's low-frequency variability and has been shown to possess a very reasonable, fairly realistic climatology, as well as multiple equilibrium states of the large-scale flow (D'Andrea and Vautard 2001; D'Andrea 2002).

More important, the QG3 model has been recently shown to have interesting regime-transition dynamics. Kondrashov et al. (2004) carried out a long-time integration of this model and studied its properties in a phase space spanned by its three leading empirical orthogonal functions (EOFs). Using two distinct clustering procedures, these authors obtained four statistically significant weather regimes: the two phases of the North Atlantic Oscillation (NAO^+, NAO^-) and the two phases of a more hemispheric and zonally symmetric mode, which they identified with the Arctic Oscillation (AO^+, AO^-). They found that these four regimes were in good agreement with previous results (Kimoto and Ghil 1993a,b; Michelangeli et al. 1995; Corti et al. 1997; Smyth et al. 1999). By studying the Markov chain of transitions between regimes, they identified five highly significant transitions that could be organized into two cycles: $NAO^- \rightarrow NAO^+ \rightarrow AO^+ \rightarrow NAO^-$ and $AO^+ \leftrightarrow NAO^+$.

They also showed that several specific transitions were characterized by preferential directions in phase space. To do so, they defined for every transition an exit point on the regime boundary; the exit vector, pointing from the regime centroid to the exit point, could then be described by two angles on the unit sphere around the centroid. The joint PDF of these two angles for the five highly significant transitions exhibited one or two sharp maxima. The directions in the reduced phase space associated with these angular-PDF maxima pointed away from the straight line passing through the centroid of the regime being exited and that of the target regime that would be visited next by the trajectory.

The existence of such preferential directions, along which the system's trajectory leaves a regime, has been conjectured by Legras and Ghil (1985), based on the nonlinear dynamics of their barotropic model on the sphere. In this model, certain regimes were associated with slowing down of the trajectories in the neighborhood of unstable fixed points. These trajectories were then ejected along the small number of unstable directions. Finding traces of similar behavior in the much more realistic, baroclinic QG3 model used here renders its investigation even more interesting in the present context.

In this article, we make use of the same clustering methodology as Kondrashov et al. (2004) to define weather regimes and the preferred transition paths between them. Statistical learning techniques are then applied to exploit this knowledge for forecasting purposes.

The paper is organized as follows. In section 2, the atmospheric model and the preprocessing performed to obtain the weather regimes and the transition paths are briefly described; some details on the model appear in appendix A. In section 3, we present the two main statistical tools of this study: the k nearest neighbor classifier and the random forest technique. Further details about the latter are given in appendix B.

Section 4 is devoted to the main results of this study, in two cases of increasing complexity. In section 4a, we forecast the three specific regime breaks that constitute the first transition cycle identified by Kondrashov et al. (2004). In section 4b, we extend our study to any possible transitions starting from the NAO^- regime. In both situations, we show that our statistical methods have verifiable predictive skill. The performance of the random forest algorithm can also be modulated according to the different weights one gives for different type of error: false alarms versus failure to predict. A sensitivity study of the forecast skill to the predictors demonstrates the critical influence of preferred transition directions. A summary and discussion of the results follow in section 5.

2. The QG3 model and its weather regimes

a. The QG3 model

The model used in this study was first proposed and investigated by Marshall and Molteni (1993). It consists in the quasigeostrophic (QG) potential vorticity (PV) equations, integrated on the sphere; the horizontal discretization is spectral, with a T21 truncation, and there are three levels in the vertical (200, 500, and 800 hPa); hence the QG3 abbreviation. At each vertical level, the prognostic equations for PV read

$$\frac{\partial q}{\partial t} = -J(\psi, q) - D(\psi) + S, \tag{1}$$

where q is the potential vorticity, ψ the streamfunction and J the Jacobian operator on a pair of two-dimensional fields. The term $D(\psi)$ is a linear operator representing the effects of Newtonian relaxation of temperature, linear drag on the lower level (with drag coefficients depending on the nature of the underlying surface), and horizontal diffusion. The spatially varying, time-independent forcing S is designed to represent PV sources that result from processes not explicitly included in the model. This source term is constructed empirically, as in Marshall and Molteni (1993), to keep the model's mean state close to that of an observed wintertime climatology; see appendix A.

Despite its simplicity, the model has a remarkably good climatology and low-frequency variability, with a plausible stationary-wave pattern, Pacific and Atlantic storm tracks, and maxima in low-frequency activity at the end of the storm tracks. The model also produces wintertime weather regimes that are very similar to the observed ones (Corti et al. 1997; D'Andrea and Vautard 2000; Kondrashov et al. 2004).

b. The weather regimes

The main steps to calculate the weather regimes are only summarized here; further details are given by Kondrashov et al. (2004). A 54 000-day-long, perpetual-winter integration of the QG3 model is first carried out. To reduce the dimension of the phase space in which the coarse graining will be carried out, we perform an EOF analysis on the unfiltered, daily 500-hPa streamfunction anomalies over the model's Northern Hemisphere. We keep the first three EOFs, thus capturing 27% of the total day-to-day variance. The coordinates are normalized in this three-dimensional phase space spanned by EOFs 1, 2, and 3, so that each EOF has unit length.

Weather regimes are then identified as areas of higher probability density in this three-dimensional phase space by applying the Gaussian mixture classification method of Smyth et al. (1999). To do so, we assume that every weather regime (or cluster) is described by a Gaussian density function. The total PDF is then modeled by a weighted linear combination of the individual weather regime density functions. With the QG3 output data, we obtain four regimes that we call, following Kondrashov et al. (2004): NAO^+, NAO^-, AO^+, and AO^-.

The next step is to determine the Markov chain of transitions between regimes. Each weather regime is defined in phase space as an ellipsoid whose centroid

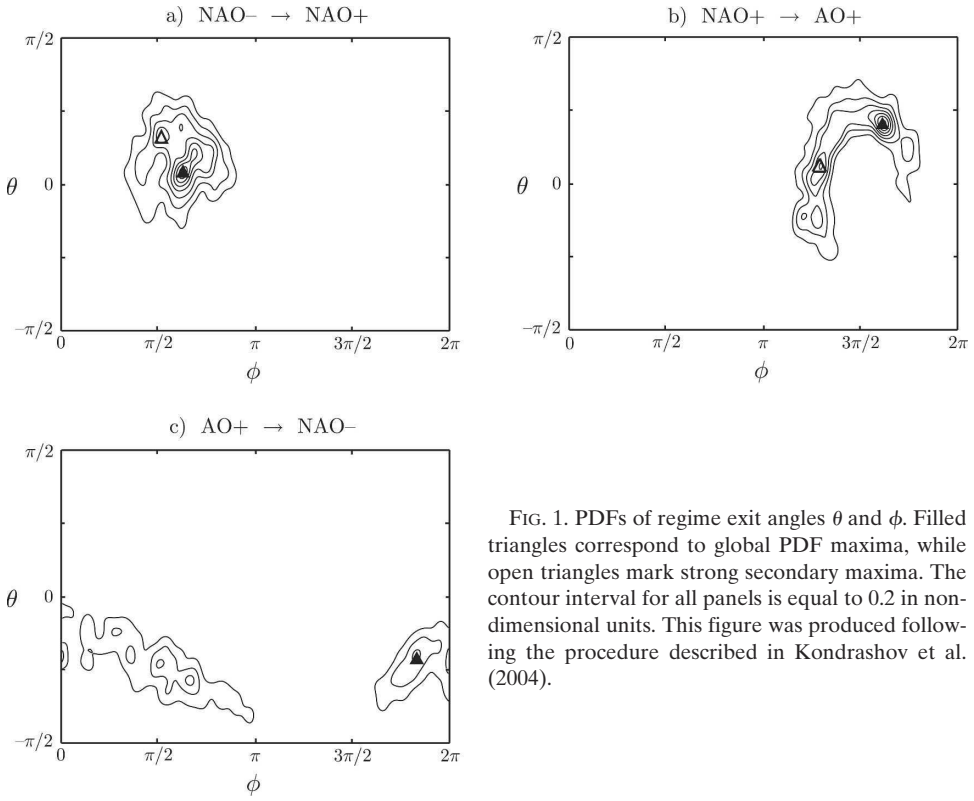

FIG. 1. PDFs of regime exit angles θ and ϕ. Filled triangles correspond to global PDF maxima, while open triangles mark strong secondary maxima. The contour interval for all panels is equal to 0.2 in nondimensional units. This figure was produced following the procedure described in Kondrashov et al. (2004).

and semiaxes are given by the mean and the covariance matrix of the corresponding Gaussian density component. The exact volume of every cluster is fixed by a scaling factor $\sigma = 1.25$ along each axis of the ellipsoid; the axes are the principal directions of the covariance matrix, and $\sigma = 1$ corresponds to the associated standard deviations.

A data point is assigned to a weather regime if it lies within the corresponding ellipsoid. When a data point belongs to several ellipsoids, we assign it to a regime according to the maximum probability value found. With this classification, about 11% of the points are in the NAO^+ regime, 13% in NAO^-, 15% in AO^+, and 9% in AO^-; the remaining 52% of the points do not belong to any cluster.

c. The preferred transition paths

Each transition is characterized by an exit point. The exit point is the midpoint between two consecutive trajectory points that lie on the opposite side of the cluster boundary, as defined in section 2b. The exit vector is then defined as the vector pointing from the cluster centroid to the exit point. In the three-dimensional phase space spanned by EOFs 1, 2, and 3, the coordinates of an exit point are (x, y, z) and the unit vector in its direction can be fully described by two angles θ and ϕ with

$$\tan\theta = \frac{z}{\sqrt{x^2 + y^2}}, -\frac{\pi}{2} < \theta < \frac{\pi}{2}, \qquad (2)$$

$$\tan\phi = \frac{y}{x}, 0 < \phi < 2\pi,$$

the positive pole being aligned with EOF-3. Computing the two-dimensional PDF of these two angles using a Gaussian kernel estimator (Silverman 1986), we obtain the preferred exit directions as the maxima of this PDF.

In Fig. 1 the PDFs of θ and ϕ are shown for the three transitions that will be analyzed in section 4a: $NAO^- \rightarrow NAO^+ \rightarrow AO^+ \rightarrow NAO^-$. For two of them, $NAO^- \rightarrow NAO^+$ and $NAO^+ \rightarrow AO^+$, the PDF has two sharp maxima close to each other; the regime break consequently occurs along either one of two paths. In the third case, $AO^+ \rightarrow NAO^-$, there is only one maximum, which is much less pronounced. Kondrashov et al. (2004) described these three transitions as the first cycle of significant transitions; they provide good examples of the two kinds of regime breaks that these authors observed on a larger set of highly significant transitions: on the one hand, sharp and pronounced maxima, on the other, less peaked angular PDFs. The first type of transition was found to be more frequent on the whole. We chose to study this transition cycle because it allowed us to compare the results here with those of Kondrashov

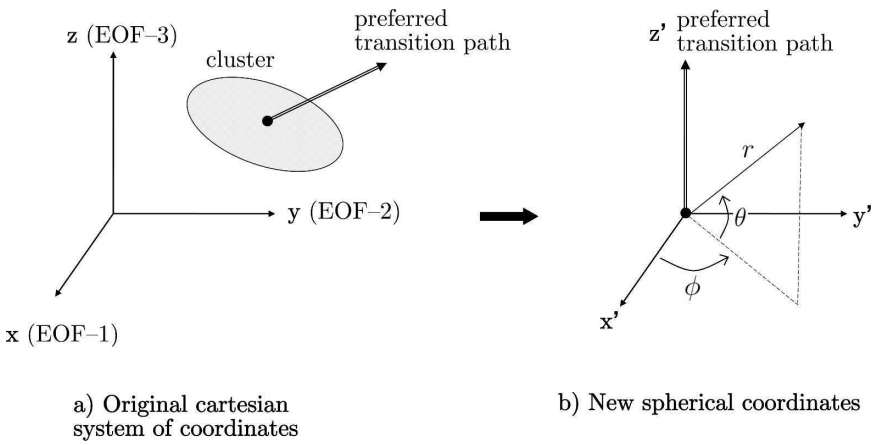

FIG. 2. Change of coordinate system to take into account the existence of a preferred direction of transition. In the new coordinate system, θ is related to the angle formed by the state vector with the preferred direction of transition.

et al. (2004), and also because they illustrate rather well the different situations in terms of exit-angle PDFs.

3. Methodology

a. Predictands and predictors

For each individual transition we are trying to forecast, we define a data point as an event or a nonevent. Let us consider the transition $NAO^- \rightarrow NAO^+$. For this transition, a point belonging to NAO^- is considered as an *event* if it is going to exit the NAO^- cluster the following day, and to enter the destination cluster NAO^+ at some moment in the future, after possibly having spent one or several days outside any regime boundary. Any other point of the NAO^- regime is considered as a *nonevent*. Nonevents can be points not leaving the NAO^- regime the next day (staying longer in the regime) or leaving NAO^- to reach a different regime than NAO^+. Forecasting the $NAO^- \rightarrow NAO^+$ regime break means to classify NAO^- points into one of the two possible outcomes: event or nonevent.

Our predictors are based on the position and the velocity of a data point. To exploit the preferential directions of regime breaks identified by Kondrashov et al. (2004), and in section 2c here, we use the spherical coordinates (r, θ, ϕ) centered on the regime centroid and with the polar axis aligned with the preferred transition path, rather than with EOF-3. When the transition under consideration has two local maxima ($NAO^- \rightarrow NAO^+$ and $NAO^+ \rightarrow AO^+$), we use the global maximum as the pole. Figure 2 illustrates this change of coordinates.

In these modified spherical coordinates, the deviation angle formed by the current state vector and the preferred transition direction is given by a single variable, θ. A value $\theta = \pi/2$ means the state vector is perfectly aligned with the preferred exit vector, while a value of $\theta = -\pi/2$ 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 QG3 model, are also called tendencies and will be expressed in the spherical coordinate system by (v_r, v_θ, v_ϕ). In summary, our predictors are daily data points in these modified, data-adaptive spherical coordinates and their tendencies $(r, \theta, \phi, v_r, v_\theta, v_\phi)$.

The choice of modified spherical coordinates emphasizes the crucial role played in regime breaks by the preferential directions identified by Kondrashov et al. (2004). Indeed, the main statistical tool of this study, random forests, allows us to estimate the relative importance of each predictor used in the forecast. A detailed discussion [see section 4a(3)] will focus on the role of the key variable θ that indicates whether a transition follows the preferred transition path or not. This deeper insight into the dynamical properties of regime breaks would have been impossible, had we kept the Cartesian coordinates.

We have at our disposal a very long model simulation of 54 000 days, but wish to evaluate our method in a manner that is consistent with the amount of data one can obtain from a reanalysis dataset. To do so, we will keep in the following only 6000 days of the simulation and thus obtain a fair estimate of our method's forecast performance when using a realistic number of training data. Although the QG3 model was run in a perpetual winter mode, these 6000 days can be thought to correspond to 50 winters of 120 days (mid-November to mid-

March). As we shall see later [see section 4a(2)], this sample length suffices for a robust estimation of the method's forecast skill.

b. *k* nearest neighbor classifier

We have used two forecast methods to classify events or nonevents from the six predictors described in section 3a. The first is a classical analog procedure. We dispose of a library of 6000 days that correspond to past observed data and that constitutes a training dataset. From these, we can build a lookup table of predictors, classified into events and nonevents.

We now consider a new point that is in NAO^- at initial forecast time and we want to determine if it is an event or not. We first search for its k nearest neighbors in the lookup table in terms of Euclidean distance in the space of the six predictors (r, θ, ϕ, v_r, v_θ, v_ϕ). Once the k nearest neighbors are identified, we count the number of events and nonevents in these k table members. The forecast then assigns the new point to the category that is the best represented among its k nearest neighbors. It is easy to check if the forecast was correct by looking at the simulated days that follow in time the point that we just classified. The number k of analogs kept in the procedure is not fixed and several values, from 1 to 20, were tested to determine the one that gives the highest probability of correct forecasts.

c. Random forests

Random forests is a more advanced classification procedure, introduced in the past fifteen years; it is based on a generalization of classification and regression trees (CART). To the best of our knowledge, the present work is the first use of random forests to forecast meteorological phenomena. As in section 3b, the key idea is to assign a given point to a class based on information contained in a set of predictors. Random forests is largely based on recursive partitioning of a training dataset by logical splits that permit accurate classifications.

Classical classification trees use successive if–then conditions to obtain a unique deterministic tree. A random forest is constructed from a set of K such deterministic trees, each based on a random sample of training data and on using at each split within a given tree, a random sample of predictors. Data points are then classified through a majority vote over all of the trees in the forest. Classification trees and their extension, random forests, are usually very effective statistical methods for classifying complex data structures when no simple relationship (e.g., linear) between predictands and predictors is apparent. Random forests is described in greater detail in appendix B here and in Breiman (2001), who also provides a convergence proof as the number of trees goes to infinity.

4. Forecast results

For the sake of simplicity, we first concentrate on one specific transition. This will also allow us to introduce contingency tables and the forecast score used. The transition chosen is $NAO^- \rightarrow NAO^+$, as anticipated in section 3a. For given points belonging to the NAO^- cluster, we forecast their regime transition to NAO^+ with the two methods above. There are only two outcomes possible in this case: either there is a transition to NAO^+ or not; these two outcomes are classified as an event or a nonevent. We then briefly compare the results obtained for two other single transitions, $NAO^+ \rightarrow AO^+$ and $AO^+ \rightarrow NAO^-$, with the ones we got in the $NAO^- \rightarrow NAO^-$ case.

In section 4b, we forecast all the possible transitions from cluster NAO^-. In this case, there are five possible outcomes: (i) no transition: the point does not leave NAO^- in the next 24 h, (ii)–(iv) transition to one of the three other clusters, and (v) reentry, with the trajectory exiting the NAO^- cluster and then returning to it.

a. Single-transition forecasts

1) *k* NEAREST NEIGHBOR CLASSIFIER

We apply this classifier with our data library of 6000 days and then test it on 1000 independent points belonging to the NAO^- weather regime. The results are summed up in a 2×2 contingency table that gives the discrete joint sample distribution of forecasts and validating observations. Table 1 summarizes the definition of contingency tables, and of user and model errors. As their name indicates, the former errors provide mainly information to the user of the forecast model, the latter mainly to the modeler.

The contingency table found with this dataset for the $NAO^- \rightarrow NAO^+$ transition is presented in Table 2. A basic difficulty of any regime-based forecast method is that a transition from a given regime A to a given regime B is essentially a rare event. We immediately see in the table that the event points are much less numerous than the nonevent points: the former represent only $11\% \approx 7.5\% + 3.3\%$ of the total. This is not surprising because we consider as events only the points that are about to leave their original weather regime in the next 24 h. As we will see later, this makes the forecast much more difficult.

To estimate the skill of this statistical predictor com-

TABLE 1. Definition of a 2 × 2 contingency table. The observations (actual category) of the points are in the rows and the forecasts in the columns. Here, a, b, c, d are the percentages of each case obtained on the assessment dataset ($a + b + c + d = 100$). Thus, true forecasts are on the diagonal and correspond to true negatives a and true positives d. The misclassified points are off the diagonal and consist in the false positives (false alarms) b and the false negatives (misses) c; the overall user error is the sum of the off-diagonal elements $b + c$.

		Forecast		
		Nonevent	Event	Model error
Observed	Nonevent	a (true negatives)	b (false alarms)	$b/(a + b)$
	Event	c (misses)	d (true positives)	$c/(c + d)$
User error		$c/(a + c)$	$b/(b + d)$	

pared to a random guess, we use the Heidke skill score (HSS; Von Storch and Zwiers 1999) H:

$$H = \frac{S - S_r}{N - S_r}, \tag{3}$$

with S the number of correct forecasts, S_r the number of correct forecasts that a random predictor would give, and N the number of assessment points. A perfect predictor would get a score of 1, whereas a value of 0 means the evaluated predictor demonstrates no skill over a random guess.

Another convenient definition of H can be given in terms of the numbers a, b, c, d introduced in Table 1:

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

In the case of our regime transition forecast in Table 2, we find $H = 0.40$, meaning that the k nearest neighbor classifier is 40% better than a random guess. This result shows that the variables we used as predictors do contain useful information for the break of the NAO^- regime and subsequent transition to NAO^+.

To better understand how this score is obtained, we must study more closely the contingency table. The user error is especially useful in practical applications of a forecasting system. When the model forecasts a nonevent, it is wrong in 7.8% of the cases; this percentage becomes 25% when a transition to NAO^+ is forecast to occur. Both of these scores are very encouraging and the overall user error rate is low, only 8.6% = 7.5% + 1.1%.

TABLE 2. Contingency table with k nearest neighbor classifier for the transition $NAO^- \rightarrow NAO^+$. A value of $k = 9$ nearest neighbors was found to give the best results.

		Forecast		
		Nonevent	Event	Model error
Observed	Nonevent	88.1	1.1	1.2
	Event	7.5	3.3	69.4
User error		7.8	25.0	

But these user errors must be taken with caution. The complementary point of view is to consider the model error, which indicates how well the statistical model performs: respectively 1.2% and 69% of the nonevent and event points are forecast incorrectly. Thus, in spite of its very low rate of false alarms, the k nearest neighbor predictor is handicapped by a relatively low detection rate: only about one-third of the transitions are forecast.

How can we explain these apparently contradictory results? In the k nearest neighbor classifier, we do not assign any particular cost to the two possible types of error, false negative versus false positive. More precisely, we implicitly consider them to be equal when we choose to classify a point in the category best represented among its k nearest neighbors. The ratio of false negatives to false positives is actually imposed, in this algorithm, by the data; that is, by the underlying dynamics and the variables used to forecast it. In the case of a rare event like an $NAO^- \rightarrow NAO^+$ transition, the overall error is dominated by the misses compared to the false alarms, with a ratio of about 7:1. One implication of this shortcoming is the relatively low detection rate of events, which may not be acceptable for a practical user. Random forests may be a good way to address this issue, as we shall see forthwith.

2) RANDOM FORESTS

In our first run of the random forest algorithm, we let the data determine the default ratio of false negatives to false positives. As described in appendix B, a contingency table is built with data points not used to construct the classifier. The results, presented in the "Data ratio" columns in Table 3, are qualitatively similar to the previous ones and the HSS, $H = 0.36$, is also quite comparable. In this case, neither statistical classifier demonstrates a significant advantage over the other.

An interesting property of random forests, though

TABLE 3. Contingency table with random forests algorithm for the transition $NAO^- \rightarrow NAO^+$; 500 trees were used and two variables were tried at each split. Results are shown for three different ratios of false negatives to false positives: the default ratio imposed by the data and two other ratios, approximately 1:4 and 1:8.

		Forecast								
		Nonevent			Event			Model error		
		Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio
Observed	Nonevent	88.3	78.5	74.4	1.5	11.3	15.4	1.7	12.6	17.2
	Event	7.3	2.9	1.9	2.9	7.3	8.3	72	*28.0*	18.3
User error		7.7	3.5	2.4	34.3	60.7	64.9			

(see again appendix B), is the algorithm's ability to impose unequal cost weights on false negatives and false positives, and yield therewith different ratios between the two types of outcomes. One way of achieving this is by allowing the bootstrap samples used in generating each random tree to overrepresent transition events versus the nonevents.

In the previous experiment, the data gave a default ratio of about 7:1, with many more false negatives than false positives. The results so far suggest two additional experiments, in which we give a much greater weight to the misses than to the false alarms. The ratio of the two types of error is now inverted; more precisely, we tried to get them as close as possible to 1:4 and 1:8. The results of these two experiments are also shown in Table 3.

The detection rate increases considerably as greater weight is given to the misses: it was initially only 28% = 100%–72% in the default case and it is now 72% = 100%–28% in the 1:4 ratio case and 82% = 100%–18% in the 1:8 case. The classifier is now much better at correctly predicting transitions, which was our initial goal.

However, this improvement of detection rate comes at the detriment of the number of false alarms. It was only 1.7% in the default-ratio case and it rises to respectively 13% and 17% in the two new experiments. This modification of the detection and false-alarm rates have of course direct consequences for the errors that a user would expect. For a given forecast that indicates a transition, the probability to be wrong rises from 34%–61% and 65%, respectively.

Note, finally, that the HSS remains of comparable size: it is now 0.43 and 0.40, in the two unequal-weight cases. It means the general skill of the classifier is not modified, what is modified is only the distribution of the error.

To assess the robustness of these results, another experiment is performed, using a dataset whose length is doubled to 12 000 days; this corresponds to 100 winters, rather than the 50 winters used so far. The results of this experiment are shown in Table 4 and are practically indistinguishable from those already discussed in Table 3. Using more data does not improve the forecast and thus the dataset of realistic length, which could be obtained from reanalysis, already contains enough statistical information.

Another robustness test uses the same dataset of 50 winters as before, but a different definition of events. This definition takes into account the residence time in the target regime, and we carried out the test for the same $NAO^- \rightarrow NAO^+$ transition as in Tables 3 and 4. A regime break is now considered as an event only if the trajectory dwells for at least three days in the destination cluster NAO^+, which corresponds to the average residence time in that cluster. As a consequence, the number of events in the dataset is reduced by approximately a factor of 2, which should make the transitions more difficult to forecast.

We found in this test that $H = 0.24$ when imposing a 1:8 weight ratio, which clearly falls short of the $H = 0.40$ obtained when using the original event definition. When using the longer, 100-winter dataset of Table 4, we get about the same number of events as in the 50-winter dataset with the original definition of events; that is, 80 events versus 82 events. In this case, the HSS skill was improved to $H = 0.29$, but a further increase in the length of the datasets did not give better results. It appears therefore desirable to keep the original event definition in forecasting the transition to a given regime and then rely on the duration versus number plot for that regime in forecasting subsequent evolution of the trajectory.

3) OPTIMIZING PREDICTOR CHOICE

In the experiments of Table 3, a subset of the predictors is sampled at random for each split within each tree (see section 3c and appendix B). This increases the flexibility of the fitting algorithm by allowing predictors that are important for small fractions of the data to enter the model. To evaluate the relative impact of each predictor on the forecasts, we present in Fig. 3 a plot of

TABLE 4. Contingency table with random forests. Same transition $NAO^- \rightarrow NAO^+$; algorithm and presentation as in Table 3 but a dataset twice as long is used: 12 000 days instead of 6000 days.

		Forecast								
		Nonevent			Event			Model error		
		Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio
Observed	Nonevent	87.8	79.3	75.0	1.9	10.4	14.7	2.1	11.6	16.4
	Event	6.9	2.6	1.9	3.3	7.7	8.4	67.6	74.8	18.4
User error		7.3	3.2	2.5	36.5	57.4	63.6			

forecast sensitivity to the predictors. Using modified spherical coordinates makes this sensitivity study all the more interesting because the impact of each predictor, especially the angle θ, gives information on the dynamical role played by the preferred transition direction in the regime break.

This is an "importance plot" that shows the decrease of detection rate when using the random forest algorithm, as each one of the six variables (r, θ, ϕ, v_r, v_θ, v_ϕ) is rendered irrelevant to the forecasting process. More precisely, when forecasts are made, we keep the values of five predictors unchanged, while randomly shuffling all the values of the sixth variable, namely the one whose importance is being evaluated. The predictor is not removed but the shuffling randomizes its values, making them uncorrelated on the average with the class to which the point is supposed to belong, event or nonevent. This process is repeated for each predictor. When each predictor is shuffled in turn, we expect a decrease in the detection rate for each, because information is lost in the shuffling. The larger the drop in the detection rate, the more critical for the forecast is the shuffled variable.

Figure 3 was built with the 1:8 weight ratio between false positives and false negatives, but other choices of the weights (not shown) produce only very slight differences in the results and lead to the same conclusions. Namely, for the $NAO^- \rightarrow NAO^+$ transition (Fig. 3a), two variables, v_r and θ, are much more important that the four others. This result is consistent with and expands upon the conclusions of Kondrashov et al. (2004): it confirms the inhomogeneity of the transitions in phase space and the crucial influence of a preferred direction.

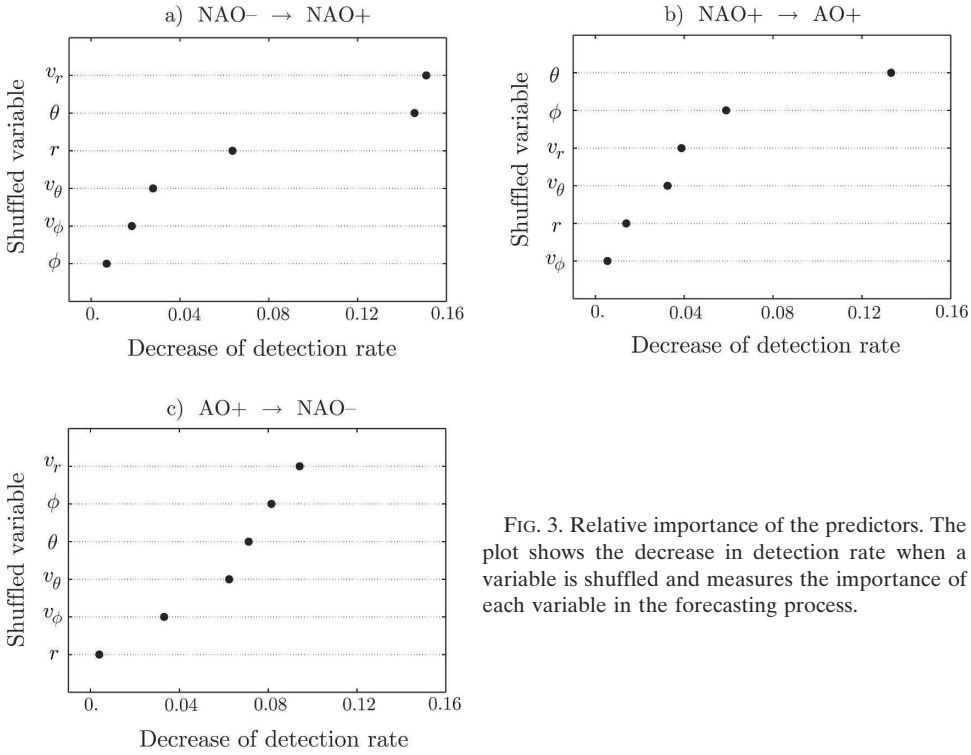

FIG. 3. Relative importance of the predictors. The plot shows the decrease in detection rate when a variable is shuffled and measures the importance of each variable in the forecasting process.

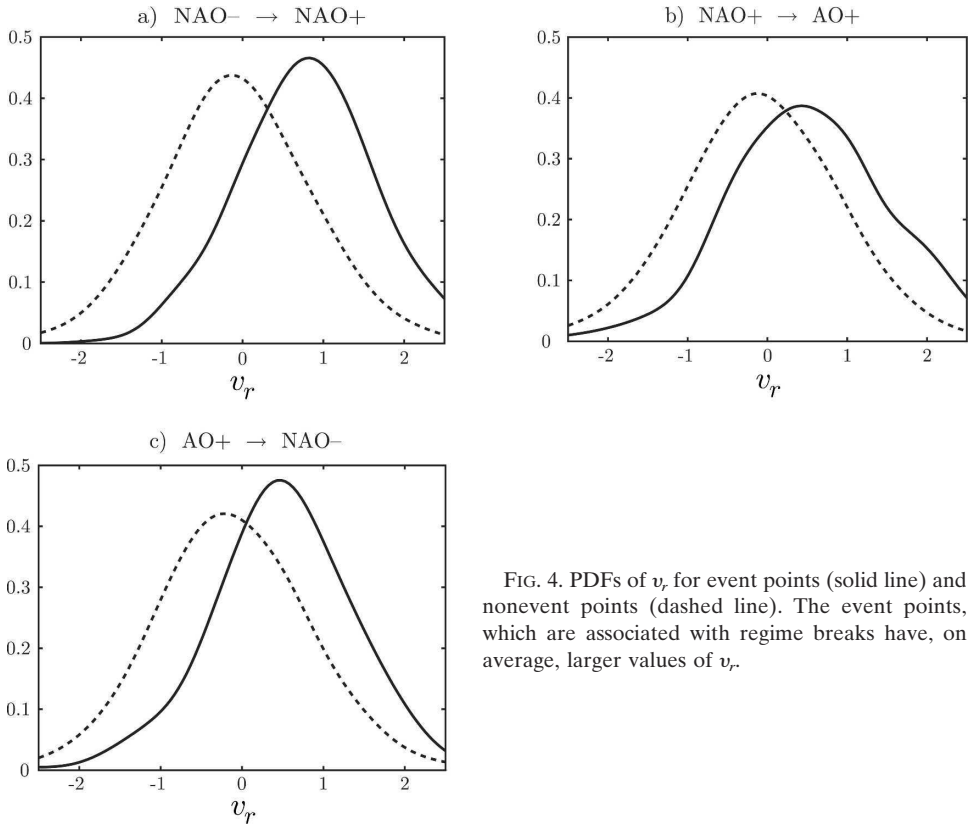

FIG. 4. PDFs of v_r for event points (solid line) and nonevent points (dashed line). The event points, which are associated with regime breaks have, on average, larger values of v_r.

The importance of v_r may indicate that the points that are moving out of the cluster and thus away from the centroid are characterized by specific radial velocities that are presumably larger that the radial velocities of the other points. To assess this hypothesis, we built the PDFs of v_r for the two groups of interest, events and nonevents, by using a Gaussian kernel estimator (Silverman 1986). We present the results in Fig. 4a. As expected, the transition points show, on average, larger values of v_r than the nonevent points.

The sensitivity of the classifier to the high-impact variable θ is investigated by producing the "partial-dependence plot" in Fig. 5. This plot provides an estimate of the conditional probability of the forecast (in Log-Odds Units or logits) with respect to the angular variable θ.

In general, the impact on classifier results of one particular variable depends on the values of the other predictor variables as well and cannot, therefore, be represented in a simple plot. The partial-dependence plot in Fig. 5 isolates the dependence of correctly forecasting an event on the value of θ, by averaging over the values of the other predictors. In effect, the other predictors are held constant. The algorithm for computing the results in Fig. 5 is given in appendix C.

Of the two sensitivity plots, the importance plot (Fig. 3) indicates that θ is a critical predictor in the forecasting process, while the partial-dependence plot (Fig. 5) tells which values of θ are most likely to yield a transition forecast. The curve in Fig. 5a shows a fairly sharp peak for θ around $\pi/2$. It means that, as expected, a transition is more likely to be forecast for vectors that point in the direction of the preferred transition path. This result is quite consistent with Legras and Ghil (1985) in attributing a key role in regime predictability to preferred directions of instability.

4) OTHER TRANSITIONS

We carried out a similar study for the two other transitions of the Kondrashov et al. (2004) cycle (see section 2c above): $\mathrm{NAO}^+ \rightarrow \mathrm{AO}^+$ and $\mathrm{AO}^+ \rightarrow \mathrm{NAO}^-$. We used only the random forest algorithm, since the results when using the k nearest neighbor classifier (not shown) were quite similar to those obtained when allowing the weights to be determined by the data in the random forest case.

As in section 4a(2), we first let the data determine the ratio of false positives to false negatives and then we prescribe the relative weights of false outcomes so that this ratio equal about 1:4 and 1:8, respectively. Tables 5

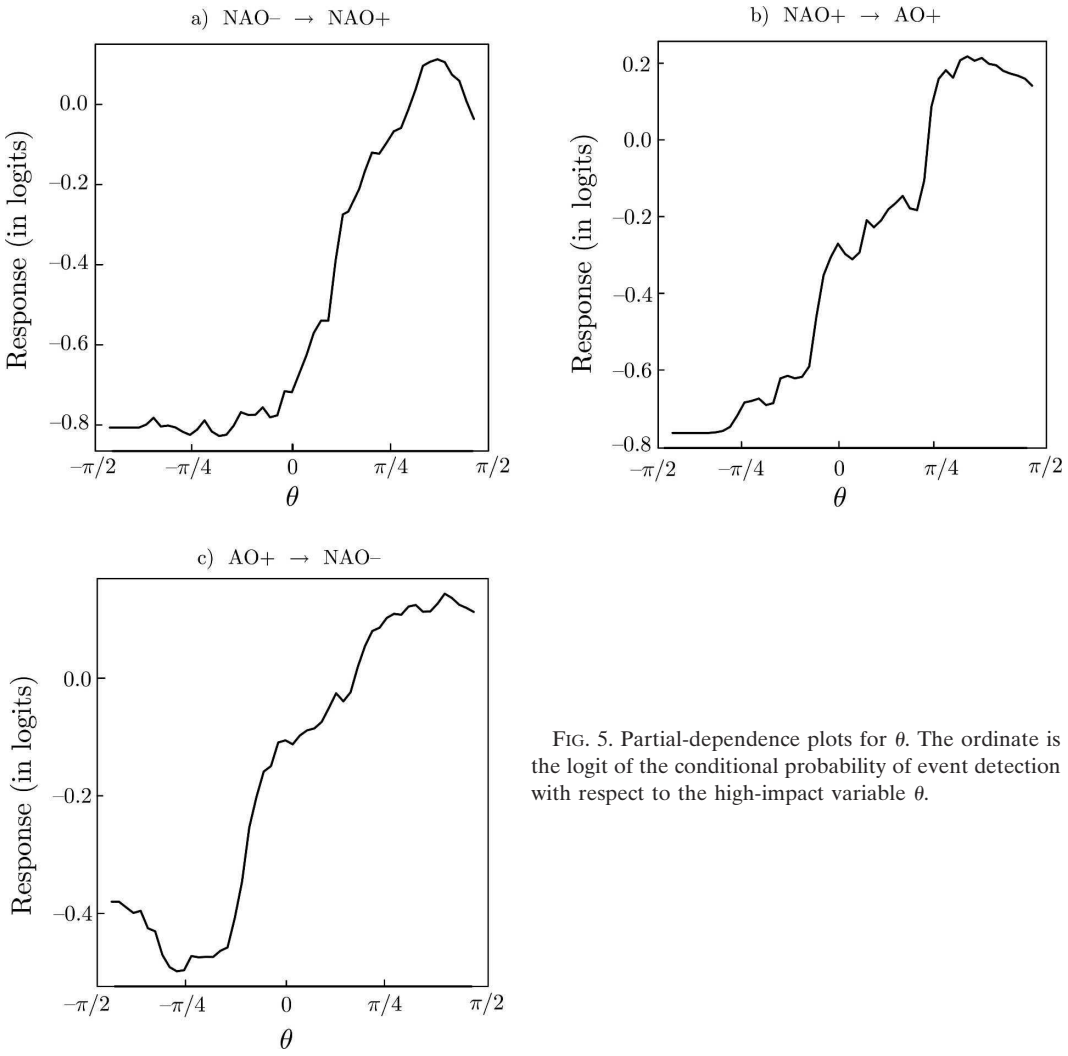

FIG. 5. Partial-dependence plots for θ. The ordinate is the logit of the conditional probability of event detection with respect to the high-impact variable θ.

and 6 are the contingency tables for these two transitions and they are both quite similar to the one already discussed. We can expect approximately the same performance in forecasting these two transitions as in Table 3 and the issue of detection rate is still critical.

The sensitivity plots in Figs. 3b,c differ more substantially from Fig. 3a than Tables 5 and 6 from Table 3. In the case of the $NAO^+ \rightarrow AO^+$ transition, the angle θ is clearly more important than all the other variables. The situation is very close to the one presented in the previous section. As seen in section 2c, this regime break is characterized by a sharp peak in the angular PDF of exits (Fig. 1b), which explains the importance of the angle θ, but the variable v_r is less important than in Fig. 3a. This state of affairs is confirmed by Fig. 4b, which shows that the values of v_r associated with the regime breaks are less well separated, in this case, from those of the nonevents than in Fig. 4a.

The $AO^+ \rightarrow NAO^-$ transition has different properties still: a group of four variables has larger importance than the other two, with v_r still the first and θ being only the third in order of importance. As discussed in section 2c, the preferred exits are not confined in this case to a narrow solid angle but are much more widely spread out (Fig. 1c). The dynamics of this transition probably has a degree of complexity that requires several predictors, rather than just one or two.

We have also plotted in Figs. 5b,c the partial dependence plots for these two additional transitions. In spite of the differences noted between the three panels in Fig. 3 and those in Fig. 4, the results in these two panels resemble quite well those obtained for the first transition we studied, namely a large, albeit broader peak for large values of θ with a maximum close to $\theta = \pi/2$. Transitions are thus more likely to be forecast when the state vector is aligned with the preferred transition path, in all three cases.

TABLE 5. Contingency table with random forests for the transition $NAO^+ \rightarrow AO^+$. (Same algorithm and presentation as in Table 3.)

		Forecast								
		Nonevent			Event			Model error		
		Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio
Observed	Nonevent	83.3	67.9	63.7	2.0	13.3	21.6	2.3	20.3	25.4
	Event	11.5	4.0	2.8	3.2	10.7	12.0	78.1	27.1	18.8
User error		12.1	5.5	4.2	38.2	61.7	64.4			

b. *Multiple-transition forecasts*

We study here all the possible transitions of a point belonging to a given cluster. This leads to distinguishing five categories, or outcomes, for the forecast. On the one hand, when a transition does occur, the point leaves the cluster within the next 24 h to reach one of the four clusters, including reentry; this gives four possible outcomes, one per cluster. On the other hand, when the point remains in its cluster for at least 24 h more, we classify it into the fifth category called the nonevent.

The random forest method is applied to the points that are in the NAO^- cluster at initial time. The same number of data, 6000 days, is used as in the previous chapter. In the present situation, the state vector cannot be expressed in the same system of coordinates as above. Since there are four possible transitions for a given point, with four different preferred exit directions, it makes no sense to choose one or another of these directions as the pole of the coordinate system. Thus, the spherical coordinates were computed with the pole being aligned with EOF-3.

The results are shown in Table 7, which is a generalized contingency table that allows five possible outcomes. The rows still contain the observations and the columns the forecasts. The cells on the diagonal thus still correspond to forecasts that are correct. Although the different possible errors and their interpretation become more complex, we can define all the same two important types of errors: the false positives and the false negatives. The first type corresponds to the points that are actually nonevents and that are forecast as transitions. They are located in the first row of the contingency table. The second type includes the points that are transitions and that are forecast as nonevents. These correspond to the first column of the contingency table. In addition, we have now a new type of error that did not exist in the two-outcome case: a transition point whose destination cluster is not correctly forecast. A point that is going to cluster AO^+ and that has been classified in the AO^- transition group would fall into this category.

We performed only two multiple-outcome experiments with different ratios of false positives to false negatives. One is the control experiment, which lets the dataset the weights, and the second is an experiment that assigns a higher cost to false negatives, so as to achieve a higher detection rate. The control experiment yields the same result as in the two-outcome case: the false negatives are much more numerous than the false positives and the detection rate is low.

In this more general case, the overall user error is the complement of the correct forecasts; that is, the complement of the diagonal elements. This error equals $26\% \approx 100\% - (62\% + 4.7\% + 4.9\% + 2.2\% + 0.1\%)$ and it is much higher in Table 7 than in Tables 2–6, where it does not exceed 16%. Indeed, the forecast of multiple outcomes is much more difficult than for only two outcomes, especially when each type of transition is a relatively rare event.

In the other experiment, with a higher weight on false negatives, we get a better rate of detection, and thus succeed in forecasting about half the transitions. The accuracy of the forecasts differs from transition to transition: the best results are obtained for the AO^- desti-

TABLE 6. Contingency table with random forests for the transition $AO^+ \rightarrow NAO^-$. (Same algorithm and presentation as in Table 3.)

		Forecast								
		Nonevent			Event			Model error		
		Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio	Data ratio	1:4 ratio	1:8 ratio
Observed	Nonevent	77.0	65.9	59.9	4.1	15.2	21.2	5.0	18.7	26.2
	Event	12.0	3.7	2.6	6.9	15.2	16.3	63.5	19.8	13.8
User error		13.5	5.4	4.2	37.1	50.0	56.5			

TABLE 7. Contingency table with random forests for every possible transition starting from the NAO^- cluster. Two different experiments are presented: in the first one (first number of each pair), we let the data impose the detection rate; in the second one (second number of pair), we tried to get a higher detection rate.

		Forecast											
		Nonevent		NAO^+		AO^+		AO^-		NAO^-		Model error$^-$	
		Data ratio	Higher ratio	Data ratio	Higher ratio	Data ratio	Higher ratio	Data ratio	Higher ratio	Data ratio	Higher ratio	Data ratio	Higher ratio
Observed	Nonevent	61.7	39.7	0.6	2.5	1.0	8.6	1.0	6.5	0.2	7.5	4.6	38.7
	NAO^+	4.1	1.1	4.7	4.4	1.0	1.1	0.2	0.4	0.1	3.2	53.7	57.3
	AO^+	6.1	2.4	0.6	0.6	4.9	7.1	0.0	0.5	0.6	1.6	60.2	41.8
	AO^-	4.4	0.4	0.7	0.4	0.2	0.2	2.2	5.3	0.1	1.5	71.4	31.7
	NAO^-	2.9	0.7	1.2	1.2	0.4	0.0	0.5	0.7	0.1	2.4	97.6	53.7
User error		22.1	10.4	40.6	52.1	35.0	58.4	45.5	60.2	90.0	85.4		

nation cluster, with a model error of only 32%, while the worst results are for NAO^+, with a model error of 57%. Once again, the results are considerably worse than in the two-outcome case, where the model error at predicting a transition was about 20%. The practical interest of a multi-outcome statistical forecast is therefore more limited than for a simpler case.

5. Concluding remarks

In this article, we have studied the predictability of the Northern Hemisphere's low-frequency variability in an intermediate-complexity model: the quasigeostrophic, three-layer (QG3) model of Marshall and Molteni (1993). This model (section 2a) exhibits four significant weather regimes in a low-dimensional subspace spanned by the three leading EOFs of its variability. Kondrashov et al. (2004) showed that certain regime transitions in the QG3 model are characterized by preferred-direction paths in this phase space (see Fig. 1).

Our goal here was to use these specific features in phase space to forecast the regime breaks in advance. To do so, we used two statistical tools: the classical k nearest neighbor classifier (section 3b) and the novel random forest method (section 3c). The application of both methods to medium-to-long-range prediction of large-scale flow patterns appears to be new.

The model's EOFs and weather regimes (section 2b) were computed using a 54 000-day, perpetual-winter simulation. To put the statistical forecast methods under study to a more severe test, we used only a 6000-day segment of this simulation as a learning set; this corresponds to 50 winters, each 120 days long, which could be obtained from the existing reanalysis of atmospheric observations.

We first focused on forecasting single transitions and obtained surprisingly good predictability, even with this short learning set. We considered the cycle of three transitions $NAO^- \rightarrow NAO^+ \rightarrow AO^+ \rightarrow NAO^-$ and, for each of the three, the statistical prediction is about 35% to 40% better than random (see Tables 3–6).

A major obstacle in correctly predicting regime transitions is the fact that these are fairly rare events. In practical situations, though, misses and false alarms may be given different weights, in particular when the two types of forecast outcomes are qualitatively different. The random forest method allows one to easily assign distinct costs to false positives versus false negatives. Of course, any improvement in the detection rate of transitions is inevitably associated with a larger number of false alarms and vice versa. Eventually it is the end user's choice to define precisely what risk is acceptable according to the prospective application of the forecast. Since the transitions of interest are rare events, we were able to obtain higher detection rates by assigning higher weights to the misses than to the false alarms, while keeping the overall skill unchanged.

The preferred transition paths identified by Kondrashov et al. (2004) were found to carry predictive information on regime transitions. Sensitivity studies to different predictors, through importance (Fig. 3) and partial-dependence (Fig. 5) plots showed the key role of the deviation angle θ formed by the exit vector with the preferred exit direction. These studies indicate that a transition is more likely to be forecast for points aligned with the preferred transition direction. We also found that the influence of θ is more crucial when the preferred transition path is confined within a fairly sharp solid angle: underlying exit dynamics seems to be largely dependent on θ in this case, although the velocity component v_r along the preferred exit direction also plays a role. The role of θ decreases when the exit-vector PDF is not limited to a narrow angle but is more spread out.

The results for the single-transition case are encouraging in view of a practical use of statistical methods in

medium-to-long-range forecasting. These results provide further support for the Legras and Ghil (1985) conjecture that (i) certain atmospheric flow regimes are associated with unstable fixed points in the flows' phase space; and, hence, (ii) exit from such regimes and subsequent transitions to other regimes originate along preferred directions of unstable growth of perturbations. Our results do not appear to be consistent with other theories for the origin and maintenance of weather regimes, as reviewed by Ghil and Robertson (2002), Molteni (2002), and Sura et al. (2005).

A natural development of the present work would be to study in greater detail the physical nature of the instabilities associated with the preferential directions of regime breaks. Another development, currently in progress, is to apply the present approach to observed data, where preferred transition paths were also hypothesized by Kimoto and Ghil (1993a,b). This will make it possible to compare the skill of statistical and dynamical models on specific transitions like those between zonal and blocked states. Such transitions are of real meteorological interest and remain a problem for numerical weather prediction models.

Acknowledgments. It is a pleasure to thank D. Kondrashov for providing the code to preprocess the data used in this article and for his advice in using it. This work grew up out of a DEA diploma project carried out by AD under the guidance of MG and O. Talagrand; both AD and MG appreciate O. Talagrand's input. RB would like to thank the Ecole Normale Supérieure and the Ecole Polytechnique for their hospitality and support during a sabbatical term in Paris. Support for this work was provided by NSF Grants SES04-37169 (RB) and ATM00-82131 (MG), as well as by the European Commission's Project 12975 (NEST) "Extreme events: Causes and consequences (E2-C2)" (FD and MG).

APPENDIX A

Average Source Term

In the atmospheric model governed by Eq. (1), the time-independent forcing S represents sources of potential vorticity that result from processes not explicitly included in the equations: radiative forcing, other diabatic heat fluxes (linked, for example, to precipitation), and the effect of divergent flow. In addition, the forcing implicitly contains the effects of subgrid-scale processes. The forcing term has been estimated here empirically, following Marshall and Molteni (1993), as follows.

From a long series of wintertime analyzed states, one can substitute $\hat{q}$ and $\hat{\psi}$ into Eq. (1), for every day of observed fields available; the hat indicates observed fields. Equation (1) holds for observed fields and gives a value of S for that day. Taking then the time average, represented by the overbar, an equation for a mean field S is obtained:

$$S = \overline{J(\hat{\psi}, \hat{q})} + D(\overline{\hat{\psi}}). \tag{A1}$$

Daily streamfunction fields were obtained from the ECMWF operational analysis for the months of January and February of the years 1984–92.

APPENDIX B

Random Forests as Classification Tools

With categorical predictands, such as those used in this paper, random forests provides a classification method. The intent is to assign classes to observations using information contained in a set of predictors. A random forest is constructed from a large number of classification trees, each tree based on a random sample (with replacement) of the data, and for each partitioning of the data for each tree, a random sample of predictors. Classification trees will be described briefly, before explaining random forests. For ease of exposition, and with no major loss of generality, we consider in the following only a binary response variable: only two outcomes are possible, for instance, "event" and "nonevent."

a. Classification trees

Each classification tree provides a recursive partitioning of a training dataset. The goal is to construct contiguous subsets within the space defined by the predictors that are less heterogeneous than the data before the partitioning. All possible predictors are screened before a potential partitioning of the data is selected; the predictor eventually used at each step is the one that decreases heterogeneity the most. Two popular measures of heterogeneity are entropy E, defined in the binary outcome case as $E = -p \log p - (1 - p) \log(1 - p)$, and the Gini index G, defined as $G = p(1 - p)$. In section 4a here, p is for instance the proportion of event points in a data partition, with $1 - p$ the proportion of nonevents.

Figure B1 represents a simple example. There is a binary response coded A or B, and two predictors x and y. The single vertical line at $x = 3$, say, produces the first partition. The double horizontal line at $y = 6$ produces the second partition. The triple horizontal line at $y = -4$ produces the third partition. Partition boundaries must be straight lines perpendicular to the predictor axes.

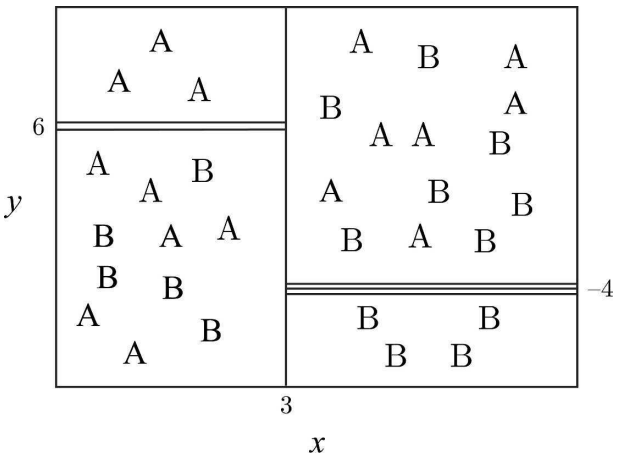

FIG. B1. Recursive partitioning used in classification trees. There is a binary response coded A or B and two predictors x and y.

In this simple illustration, the upper-left set and the lower-right set are fully homogeneous. There remains considerable heterogeneity in the other two sets and, in principle, their partitioning could continue. When there are no longer any ways to further partition the data to make them more homogeneous, the algorithm stops. Each final set is then assigned a class, based on a majority vote of the observations in that set. Here either class A or class B would be assigned to a set according to which has a greater proportion of observations in that set. The classification of a new point not included in the training dataset requires only to determine in which set the observation lies and the associated class.

b. Random forests

Random forests generalizes classification trees by considering a large set of trees generated by a process that introduces random factors. Let n be the number of training observations on hand. The random forest method then operates with the following steps:

1) Take a random sample of size n with replacement from the total dataset on hand.
2) Take a random sample without replacement of all the possible choices of predictors included in the data.
3) Construct the first data partition of a classification tree.
4) Repeat steps 2 and 3 for each subsequent split, until the classification tree is as deep as desired. Do not prune the tree.
5) Drop the data not included in the sample from step 1 down the tree. Store the class assigned to each observation along with each observation's predictor values.
6) Repeat steps 1–5 a large number of times (we used 500 trees in this paper), so that there is a large number of trees, which constitute a random forest.
7) Using only the class assigned to each observation when that observation is not used to build a tree, count the number of times over trees that the observation is classified in one outcome category and the number of times over trees it is classified in the other outcome category.
8) Assign each observation to one of the two outcome classes by a majority vote over the set of trees.

Random forests has five demonstrable assets. First, for the kinds of data analyzed in this paper, there are no classifiers to date that will consistently classify and forecast more accurately. Most will do worse, especially when the true relationships with the response are highly nonlinear and noisy. Second, one can prove (Breiman 2001) convergence of the algorithm in measure ("almost surely") as the number of trees goes to infinity. An important practical consequence of this theoretical result is that the algorithm does not overfit. This is very useful because it implies that the results will be robust when drawing new random samples from the same population (i.e., data with the same characteristics except for random sampling error).

Third, because performance is determined by a contingency table computed from observations not used to construct a given tree (i.e., observations not selected in step 1), performance rests on real forecasting skill. Fourth, random forests provides a means by which the relationships between inputs and outputs can be represented in an instructive way, using importance plots and partial-dependence plots.

Finally, there are several systematic ways in which the relative costs of false negatives and false positives can be taken into account. The approach used in this paper gives more weight to observations in which a transition does occur, so that if such observations are misclassified, the consequences are greater. This is accomplished by oversampling transition events when bootstrap samples are drawn for each tree; in other words, transition events are made more common in the analysis than they are in the data. The presence of such a random element in determining the weights of events versus nonevents is the reason for achieving a targeted weight ratio only approximately in Tables 3–6.

This is an improvement over the more classical k nearest neighbor algorithm. Indeed, we tried to force different cost weights with the latter method by modifying the classification process. A point was forecast as an event if $k_{\text{event}}/k > a$ with k_{event} the number of event points in its k nearest neighbors and a as a parameter

setting the relative cost of false positives to false negatives. Equal costs corresponds to $a = 0.5$, while $a < 0.5$ gives more weight to the false negatives compared to the false positives. We could not get the same results than random forests especially for the larger ratio 1:8; the overall skill dropped making the forecast of no practical interest. This is probably a consequence of the limited size of the dataset that imposes to choose a small value of k; hence one cannot fine-tune the value of a without giving rise to critical sampling problems. This is not an issue for random forests because the classification is made through a majority vote over a large number of trees (we used 500 trees in this study) and not among a small number of k nearest neighbors.

In practice, it is the robustness of random forests that matters more than the above-mentioned rigorous convergence proof: given a reasonable choice of parameters (number of trees, number of predictors to try at each split, etc.), one wants to get pretty much the same results when running the algorithm several times. The number of trees is indeed an important parameter and as it gets large, the results are found to be increasingly replicable, as expected from the theory. The algorithm runs so quickly that thousands of trees can be run in a minute or so on a laptop in real time. Actually, even with several hundred trees only, the results here are stable and, therefore, replicable. We typically used in this paper 500 trees and this appears to be far more than one probably needs in most cases.

Breiman (2001) gives a formal exposition of classification and regression trees, while Breiman et al. (1984) provides a full presentation of random forests. An excellent reference to statistical learning in general is Hastie et al. (2001), which contains many examples of random forests.

APPENDIX C

Partial-Dependence Plots

Partial-dependence plots display in logits how the probability of a particular event (here, a transition) is related to a given predictor, the values of all other predictors being fixed. A partial-dependence plot is constructed in the following manner:

1) Grow a forest.
2) Suppose x has υ distinct values in the training dataset. Construct υ datasets as follows. For each of the υ values of x, make up a new dataset where x only takes on that value, leaving all other variables untouched.
3) For each of the υ datasets thus obtained, predict the response using random forests.
4) For each of the υ datasets, average these predictions determining the proportions p and $1 - p$ of trees that respectively forecast an event and a nonevent. Compute in logits the ratio of these proportions, $R = 0.5 \log[p/(1 - p)]$.
5) Finally, plot this ratio R (expressed in logits) for each of the υ values of x.

Thus, partial-dependence plots show the relationship between a given predictor x and the response averaged over the joint values of the other predictors as they are represented in the tree structure. In this way, the other predictors are being held constant by matching, so that no assumptions are being made about how the predictors are related to one another or to the response variable. More details about partial-dependence plots can be found in Hastie et al. (2001).

REFERENCES

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

Breiman, L., 2001: Random forests. *Mach. Learn.,* **45,** 5–32.

——, J. Friedman, R. Olshen, and C. Stone, 1984: *Classification and Regression Trees.* Wadsworth Press, 368 pp.

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

Corti, S., A. Giannini, S. Tibaldi, and F. Molteni, 1997: Patterns of low-frequency variability in a three-level quasi-geostrophic model. *Climate Dyn.,* **13,** 883–904.

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, 2000: Reducing systematic errors by empirically correcting model errors. *Tellus,* **52A,** 21–41.

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

——, and Coauthors, 1998: Northern Hemisphere atmospheric blocking as simulated by 15 general circulation models in the period 1979–1988. *Climate Dyn.,* **14,** 385–407.

Fraedrich, K., and M. Klauss, 1983: On single station forecasting: Sunshine and rainfall Markov chains. *Beitr. Phys. Atmos.,* **56,** 108–134.

Ghil, M., and A. W. Robertson, 2002: "Waves" vs "particles" in the atmosphere's phase space: A pathway to long-range forecasting? *Proc. Natl. Acad. Sci. USA,* **99,** 2493–2500.

——, M. Kimoto, and J. D. Neelin, 1991: Nonlinear dynamics and predictability in the atmospheric sciences. *Rev. Geophys., Suppl., U.S. National Report to the International Union of Geodesy and Geophysics,* **29,** 46–55.

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning.* Springer-Verlag, 552 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 quasigeostrophic model. *J. Atmos. Sci.,* **61,** 568–587.

Kravtsov, S., A. W. Robertson, and M. Ghil, 2005: Bimodal behavior in the zonal mean flow of a baroclinic β-channel model. *J. Atmos. Sci.,* **62,** 1746–1769.

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

Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc.,* **50,** 345–349.

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

Michelangeli, P. A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.,* **52,** 1237–1256.

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

Molteni, F., 2002: Weather regimes and multiple equilibria. *Encyclopedia of Atmospheric Sciences,* J. Holton, J. Cury, and J. Pyle, Eds., Vol. 6, Academic Press, 2577–2586.

Pelly, J., and B. Hoskins, 2003: How well does the ECMWF ensemble prediction system predict blocking? *Quart. J. Roy. Meteor. Soc.,* **129,** 1683–1703.

Robinson, W. A., 2000: A baroclinic mechanism for the eddy feedback on the zonal index. *J. Atmos. Sci.,* **57,** 415–422.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.,* **56,** 729–752.

Stephenson, D. B., A. Hannachi, and A. O. Neill, 2004: On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc.,* **130,** 583–605.

Sura, P., M. Newman, C. Penland, and P. D. Sardeshmukh, 2005: Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes? *J. Atmos. Sci.,* **62,** 1391–1409.

Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus,* **42A,** 343–365.

Von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.