On the predictability of flow-regime properties on interannual to interdecadal timescales

Franco Molteni, Fred Kucharski and Susanna Corti*

Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

***Institute of Atmospheric Sciences and Climate ISAC-CNR, Bologna, Italy*

1.   Introduction

Atmospheric flow regimes are usually defined as large-scale circulation patterns associated with statistical equilibria in phase space, in which the dynamical tendencies of the large-scale flow are balanced by tendencies due to non-linear interactions of high-frequency transients. The existence of states with such properties can be verified in a rigorous way in numerical simulations with simplified numerical models (as in the pioneering study of Reinhold and Pierrehumbert (1982), or in the experiments by Vautard and Legras (1988)). On the other hand, the existence of flow regimes in the real atmosphere has been strongly debated. The detection of regimes in the observational record of upper-air field is indeed a complex task, which has been approached by a number of research groups with a variety of sophisticated statistical methods (see Section 3).

Although the regime classifications provided by the different observational studies were not identical, a "core" number of regimes were consistently detected in most studies devoted to a specific spatial domain. For example, the three Northern-Hemisphere clusters found by Cheng and Wallace (1993) were also identified by Kimoto and Ghil (1993a), Corti et al. (1999) and Smyth et al. (1999). However, consistency does not necessarily imply statistical significance, and one may question whether the level of confidence attached to these regime classifications is sufficiently high.

The search for regimes in the real atmosphere is also made complex by the fact that, unlike in simple dynamical models, the sources of energy and momentum at the lower boundary display variations on seasonal, interannual and interdecadal timescales. Therefore, one is dealing with a dynamical system subject to a continuously variable forcing, which may alter the statistical properties of flow regimes. Seasonal forcing variations are usually accounted for by analysing anomalies defined with respect to an estimated yearly cycle, and belonging to separate seasons. But how to deal with interannual and interdecadal forcing anomalies is not a trivial problem. Should one, for example, stratify the observed sample according to the El Niño – Southern Oscillation (ENSO) phase, or consider that the radiative forcing of the atmosphere has been modified by the long term increase in greenhouse gases (GHGs)?

If regimes statistics are dependent on boundary conditions and other forcing parameters, the consequent inhomogeneity of the observed record makes the detection of atmospheric flow regimes more difficult. On the other hand, if we were able to determine the statistical properties of regimes as a function of the forcing anomalies, this would imply some degree of predictability of the atmospheric conditions on interannual to interdecadal timescales, which we can refer to as *regime predictability*. Given the limited size of the observed record, it is difficult to find statistically significant results on such an issue from observed data. In order to understand the dynamical meaning of interdecadal differences in regime distributions such as those shown by Corti et al. (1999), one has to resort to ensemble simulations made with general circulation models (GCMs), in which multiple realizations of the atmospheric flow for the same boundary (or GHG) forcing can be obtained.

Is regime predictability a sound concept? Is it feasible to obtain reliable information about the structure and frequency of regimes from observations and GCM simulations? Is the level of significance of these results acceptable? In this paper, we will argue for positive answers to these questions, focussing on two specific issues. Firstly, we will address the issue of the statistical significance of regime estimates from the current upper-air observational record. Secondly, we will demonstrate the feasibility of regime predictions as a function of SST conditions, investigating the effects of ENSO on extratropical regime statistics in ensembles of GCM simulations, and their impact on the predictability of inter-decadal variations of the wintertime Northern-Hemisphere circulation.

The concept of regime predictability will be given a more formal definition in the next section. Since confidence about regimes in the real atmosphere is needed before addressing the problem of regime predictability in GCM simulations, section 3 will be devoted to a brief review of relevant observational results, and will discuss what tests and what levels of statistical significance should be used in the assessment of observed statistics. In Sect. 4, results from numerical simulations performed with an intermediate-complexity atmospheric GCM will be presented, with emphasis on the model climatology during the boreal winter. Observed and modelled effects of ENSO on the northern extratropical circulation on interannual and interdecadal scales will be compared, using traditional statistics such as composite anomaly maps and their spatial correlations. In Section 5, the impact of ENSO will be addressed from the viewpoint of regime properties. Conclusions will be presented in Section 6.

2.   A definition of regime predictability

If one regards the atmosphere as a multi-dimensional dynamical system, the time evolution of the atmospheric state can be described by a set of differential equations of the form:

$$d_t \mathbf{X} = \mathbf{\Phi} \left(\mathbf{X}(t), \mathbf{F} \right) \tag{1}$$

where $X(t)$ is the time-dependent atmospheric state vector and F a set of forcing parameters and/or boundary conditions which can be considered either as constant, or as varying on very long time scales.

In a deterministic framework, the problem of atmospheric predictability can be seen as the study of the uncertainty in our knowledge of $X(t)$ given an "accurate" estimate of $X(t_0)$ at a previous time. In a probabilistic approach, the knowledge about the atmospheric state at any time t is expressed through the probability density function (PDF) in phase space, $\rho(X, t)$.

Given the dynamical system in Eq. (1), the time evolution of the PDF is formally related to the time derivative of the state vector through to the Liouville equation:

$$\partial_t \rho \left(\mathbf{X}, t \right) + \text{div} \left[\rho \left(\mathbf{X}, t \right) \cdot \mathbf{\Phi} \left(\mathbf{X}, \mathbf{F} \right) \right] = 0 \tag{2}$$

(Gleeson, 1970; Ehrendorfer 1994)

In weather forecasting, $\rho(X, t)$ has to be evaluated starting from an estimate $\rho(X, t_0)$ at the initial time t_0. Lorenz (1975) referred to such initial-value problems as "predictions of the first kind". After a sufficiently long time, independently from its initial value, the PDF will approach an asymptotic value $\rho^*(X)$, which is representative of the "climate" of the system.

On the other hand (again following Lorenz 1975), a second kind of predictions can be considered, in which one wants to estimate the variations of the "climate PDF" $\rho^*(X)$ as a function of anomalies in the forcing and boundary parameters of the system (represented by the F vector in Eq. (1)). Estimates of future climate scenarios as a function of the concentration of atmospheric greenhouse gases (see e.g. Chapter X in IPCC 2001) are typical examples of predictions of the second kind. As far as seasonal predictability is concerned, numerical experiments using prescribed SST to estimate the average impact of (say) El Niño events can

again be regarded as predictions of the second kind, while actual seasonal forecasts with coupled ocean-atmosphere models are actually initial-value problems in a complex, multi-scale environment.

Whatever the time scale of the prediction, one is faced with the problem of estimating the properties of a PDF in a multi-dimensional space with many degrees of freedom. If the only required information is either the mean state or the variance of the state variables, these can be easily computed and displayed in physical (i.e. geographical) space. However, if more sophisticated probabilistic estimates are needed (typically, when the PDF is non-normal), a suitable methodology is needed to condense the large amount of information provided by a multi-dimensional PDF.

Let us now assume that a "meaningful" partition of the phase space into N distinct regions could be found, and let $\{X_j, j=1,..,N\}$ be the set of the mean states (or centroids) of such partitions. Let us also assume that the variability within each of these partitions is smaller than the variability between the centroids. Then, the continuous, multi-dimensional PDF $\rho(X, t)$ can be approximated by a discrete probability distribution $P(X_j, t)$, which gives the probability of the atmospheric state vector belonging to each of the N partitions at a given time.

For probabilistic predictions of the first kind (typically provided by ensemble forecasting systems such as those described by Toth and Kalnay 1993, Molteni et al. 1996, Houtekamer et al. 1996), the phase space partitions can be defined a priori based on the properties of the climatological PDF $\rho*$, or can be redefined as a function of the initial conditions (examples of both kinds can be found in Molteni et al. 1996, while the former approach was adopted by Chessa and Lalaurette 2001).

For predictions of the second kind, which by definition are concerned with the nature of the climatological PDF, the choice of the optimal partition should reflect the statistical and dynamical characteristics of the system's attractor. Specifically, if dynamical non-linearities are strong enough to generate significant anisotropic features in the climatological PDF, the system is said to display a "regime" behaviour, where the regimes are defined by the most densely populated regions of the phase space (as in the seminal study by Reinhold and Pierrehumbert 1982). In such a case, the regime structure provides a natural way to discretize the climatological PDF in predictability studies.

When one investigates changes in regime properties as a function of anomalies in forcing and boundary conditions, two possibilities should be considered. If the forcing variations are relatively small, then the number of distinct regimes is likely to remain the same, with minor variations of the positions of the regime centroids in phase space. In this case, one should mainly be concerned with changes in the frequency of a given set of regimes (as argued by Molteni et al. 1993, Palmer 1993, 1999; Corti et al. 1999). Strong forcing variations, on the other hand, can take non-linear systems through bifurcation points, and the number of flow regimes can be altered (e.g. Ghil 1987). Variations in the number of regimes caused by strong ENSO events have been reported by Molteni and Corti (1998), Straus and Molteni (2003) and Molteni et al. (2003) in modelling studies of the extratropical and tropical circulation, respectively. In such a case, predictions of the second kind should also provide information on the regime number and their phase space position.

3. Regimes in the real atmosphere: what significance?

3.1. The search for regimes: a brief review

The concept of weather regimes was introduced by Reinhold and Pierrehumbert (RP, 1982) to explain the preferred circulation patterns produced by the interaction of planetary-scale and synoptic-scale waves in a highly truncated two-level quasi-geostrophic model forced by large-scale topography. The RP model was an extension to synoptic scales of the model used by Charney and Straus (1980) to investigate multiple equilibria (i.e. multiple steady states) of orographically-forced waves in a baroclinic framework. However, it

was clear from the RP study that there was no one-to-one correspondence between steady states and weather regimes. In fact, weather regimes or flow regimes should be regarded as statistical-dynamical equilibria, which are defined by averaging the dynamical tendencies on a time scale longer than the typical period of baroclinic transients.
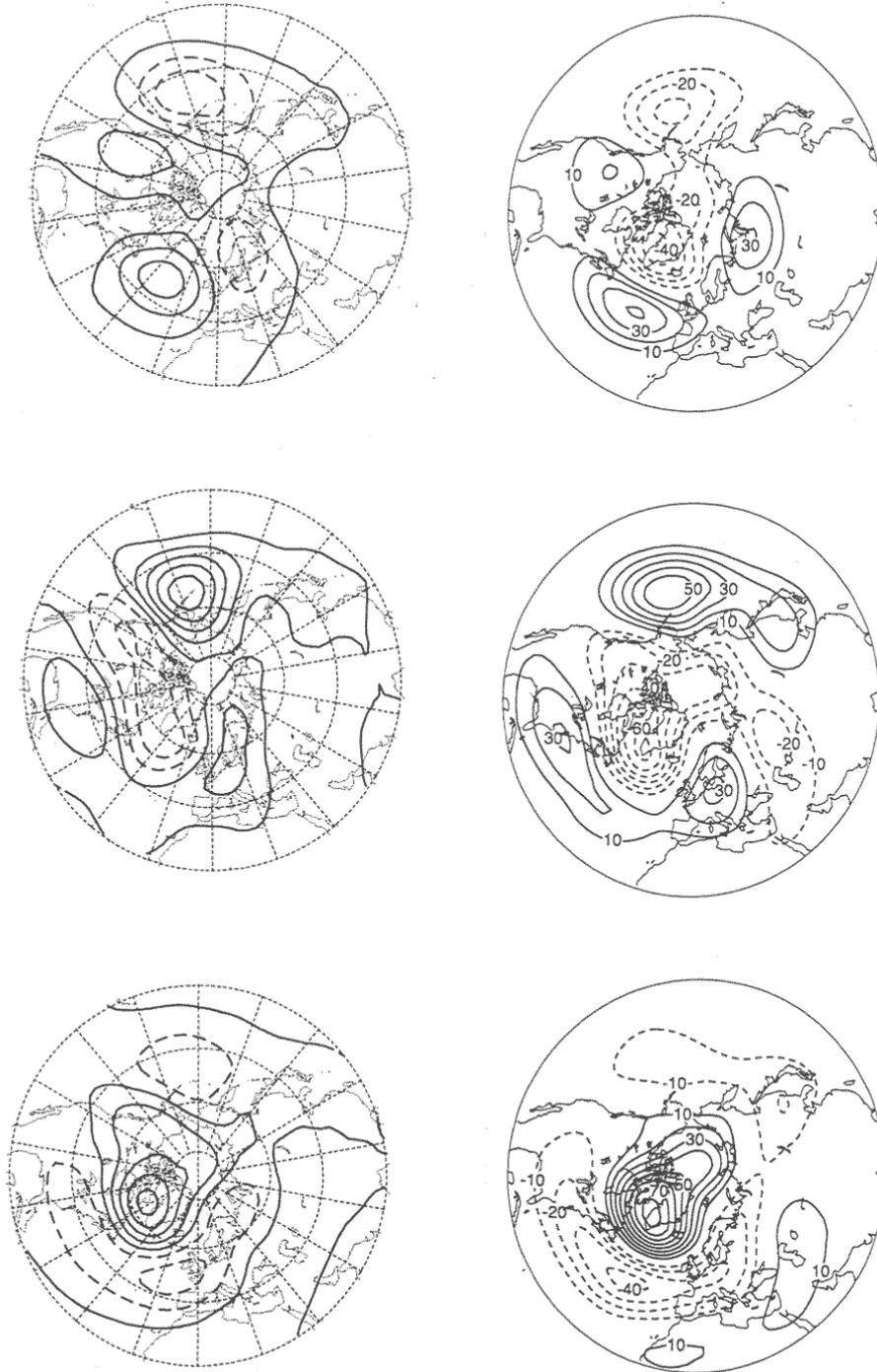
Figure 1 Left column: Anomalies corresponding to the three cluster centroids of 500-hPa height found by Cheng and Wallace (1993) using a hierarchical clustering method. Right column: Anomalies corresponding to three of the (four) density maxima in the PDF of the two leading PCs of 500-hPa monthly-mean anomalies, computed by Corti et al. (1999). Data are for December to February, 1946 to 1985 in Cheng and Wallace, November to April, 1949 to 1994 in Corti et al.

Although regime behaviour was detected in a number of simplified dynamical models (e.g. Vautard and Legras 1988; Mo and Ghil 1987; Marshall and Molteni 1993), the relevance of such results to the real atmospheric flow was often questioned (e.g. Cehelsky and Tung 1987). The search for observational evidence of flow regimes gave rise to even more controversy: for example, the significance of the bimodality in the distribution of the planetary-wave amplitude index found by Hansen and Sutera (1986) was strongly questioned by Nitsche et al. 1994 (see also the "revisitation" by Hansen and Sutera 1995). Further debate focussed on whether regimes were best defined on a hemispheric domain (as suggested by works on planetary wave dynamics), or rather within sectors including one of the two storm tracks located over the northern oceans (e.g. Vautard 1990; Kimoto and Ghil 1993b, Michelangeli et al. 1995; D'Andrea and Vautard 2001).

Methodological differences in the search for regimes also contributed to the complexity of the picture. While model simulations allowed the use of methods based on the equilibration of dynamical tendencies (e.g. Vautard and Legras 1988; Haynes and Hannachi 1995), methods to detect densely populated regions of phase space have been mostly applied to the observational record. These included univariate PDF estimation (Hansen and Sutera 1986, 1995); multivariate PDF estimation (Kimoto and Ghil 1993a; Corti et al. 1999, Hsu and Zwiers 2001); and different methods of cluster analysis (Mo and Ghil 1988; Molteni et al. 1990; Cheng and Wallace 1993; Michelangeli et al. 1995; Smyth et al. 1999).

Although the regime classifications provided by the different methodologies were (obviously) not identical, a few "common" regimes were detected in most studies devoted to a specific spatial domain. As mentioned in the Introduction, the three clusters identified by Cheng and Wallace (1993) for the whole Northern Hemisphere were also identified by Kimoto and Ghil (1993a), Corti et al. (1999) and Smyth et al. (1999). Of these clusters (shown in Fig. 1), two show nearly opposite anomalies over the Pacific-North American region, while the third one corresponds to the negative phase of the North Atlantic Oscillation (NAO).

Among the analyses devoted to either the Atlantic or the Pacific sector of the Northern Hemisphere, a robust partitioning method and a detailed analysis of statistical significance were employed by Michelangeli et al. (1995). Their analysis yielded 3 and 4 clusters for the Pacific and the Atlantic sectors respectively; these clusters were also identified in earlier studies on regional regimes (e.g. Vautard 1990; Kimoto and Ghil 1993b).

3.2.   The estimation of significance

In order to evaluate the significance of regimes found either by a PDF estimate or by cluster analysis, a common methodology has been adopted in a number of the studies cited above. Assuming that the times series of the selected phase-space coordinates are uncorrelated (as in the case of principal components), this procedure can be summarised by the following steps:

1. Define a quantity q which can be taken as a (positively-oriented) measure of the likelihood of the existence of multiple modes or clusters in a given data sample.

2. Perform the PDF estimate or the cluster analysis on the selected data sample by varying the algorithm's parameters, in such a way to obtain regime partitions with an increasing number of modes or clusters, and define $q*_m$ as the value of q corresponding to the m-regime partition.

3. Generate a large number (N_s) of samples of pseudo-random red-noise data, with the same size and the same mean, variance and lag-1 auto-correlation of the actual data sample.

4. For each red-noise sample, repeat the PDF estimate or cluster analysis with the same parameters yielding m regimes in the actual data, and compute q_m in order to obtain a sample of N_s values, say $\{q_{mk}, k = 1, ..N_s\}$.

5. Since the red-noise data are assumed to have a unimodal distribution, the proportion P_m of red-noise samples for which $q_{mk} > q*_m$ is an inverse measure of the significance of the m-regime partition of the actual data, and $1 - P_m$ is the corresponding confidence level for the existence of m regimes.

In cluster analysis, the total variance of the data sample can be divided into a fraction accounted for by the cluster means (centroids) and an intra-cluster part representing the mean-squared distance from the appropriate cluster centroid. The ratio of centroid variance to intra-cluster variance is a measure of the separation between clusters, and therefore provides a suitable definition of q (see Straus and Molteni 2003). Other definitions used in the literature include quantities which indicate the level of reproducibility of a given cluster set using different subsets of the full data sample, or starting iterative aggregations of data from different random "seeds" (e.g. Michelangeli et al. 1995).

When regimes are studied using non-parametric PDF estimation (see Silverman 1986), the number of regimes may be directly related to the number of local maxima (modes) in the PDF. However, such a number depends on the degree of smoothing used in the estimation, which is controlled by a disposable parameter. A widely used methodology is represented by the kernel estimation, which provides an estimate of the PDF of a (vector) variable X as the sum of elementary, unimodal functions centred around each data point:

$$\rho(X) = N_d^{-1} \sum_i K [h^{-1} \| X - X_i \|] \qquad (3)$$

where K is the (normalised) kernel function, X_i are the N_d input data, $\| . \|$ represents a norm in phase space and h is a parameter called the kernel width, which defines the level of smoothing in the PDF estimate.

The relationship between the kernel width h and the number of modes m in the PDF is clearly defined when a multi-normal (Gaussian) form is assumed for the kernel function. In this case, it can be demonstrated that, for any data sample, m is a monotonically decreasing function of h (Silverman 1981). The *largest* value of h for which m modes are found is called the critical kernel width for m modes ($h*_m$), and can be used to estimate the significance of multimodality against red-noise data as outlined above for a generic variable q.

Note that, although many PDF estimates with increasing h are needed to determine $h*_m$ from the actual data sample, just one estimate is needed for each red-noise sample if a multi-normal kernel is used. Because of the monotonic relationship between h and m, it is sufficient to compute the PDF using $h*_m$ as kernel width. If the number of modes in the PDF of k-th random sample is less than m, h_{mk} must be less than $h*_m$. For example, if $h*_2$ is the critical width for bimodality in the actual data, the proportion of unimodal PDFs obtained from red-noise samples setting $h = h*_2$ gives a confidence level for bimodality.

3.3.   On the significance of Northern Hemisphere regimes computed from monthly-mean anomalies

We will now apply these concepts to the estimation of the significance of multimodality in PDFs derived from a sample of monthly-mean anomalies of 500-hPa height for the northern winter (following Corti et al. 1999; CMP hereafter). For the sake of comparison with the model results presented in Sect. 5, the analysis will cover the 44 winters from 1954/55 to 1997/98, using months from December to March (DJFM). The 500-hPa height monthly-mean data are taken from the National Centers for Environmental Predictions /National Center for Atmospheric Research (NCEP/NCAR) re-analysis (Kalnay et al. 1996). Although the main results of this analysis do not differ substantially from those of CMP and other related studies (see references above), here the goal is to illustrate and discuss the significance test in greater detail.

As in many previous studies, the leading variability patterns are first identified through a principal component (PC) analysis of the 500-hPa height anomalies. The empirical orthogonal functions (EOFs) associated with the first two PCs are shown in Fig. 2; choosing the same sign as in CMP, they are quite similar to the negative phase of the Arctic Oscillation and to the so-called Cold Ocean – Warm Land pattern

respectively (Thompson and Wallace 1998; Wallace et al. 1996). The first two PCs (normalised by their respective standard deviations) are used as the coordinates of a bi-dimensional phase space in which PDFs will be estimated. An iterative version of the Gaussian kernel estimator (see Eq. (3)) is used to compute the bi-dimensional PDFs, in order to avoid spurious local maxima in scarcely populated regions of phase space (Silverman 1986; Kimoto and Ghil 1993a).
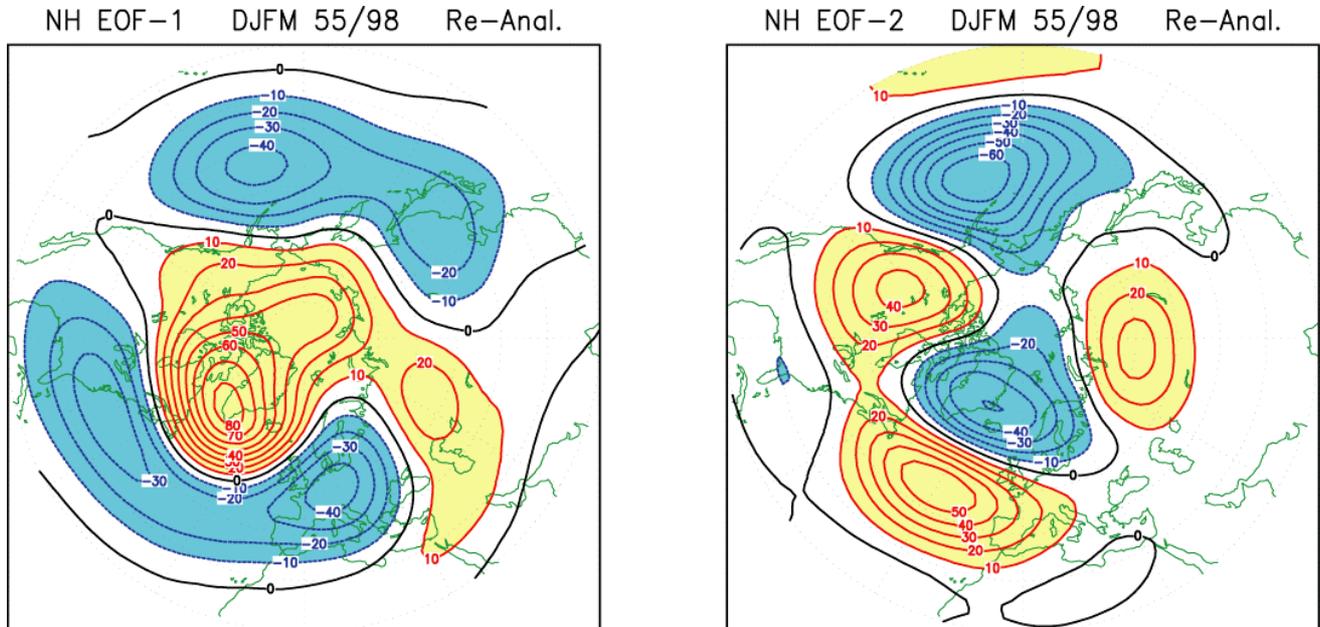
Figure 2 First two EOFs of 500-hPa height monthly-mean anomalies over the Northern Hemisphere (20-90 N), computed from the NCEP/NCAR re-analysis in the 44 winters (DJFM) 1954/55 to 1997/98. The EOFs are scaled to represent the anomaly patterns corresponding to one standard deviation of the associated PCs. Contour interval 10 m..

The estimated PDF in the PC1-PC2 plane is shown in Fig. 3 for three different values of the kernel width h (which becomes a non-dimensional parameter when coordinates are normalised), namely 0.3, 0.4 and 0.5. For $h = 0.3$, four local maxima are evident in the PDF. Three of them are also present in the estimate with $h = 0.4$; when the PC coordinates of these modes are multiplied by the corresponding EOFs to obtain a 500-hPa height anomaly, such patterns are (as expected) almost identical to the three local maxima of CMP shown on the right-hand column of Fig. 1. On the other hand, the PDF estimated with $h = 0.5$ is unimodal; further calculations show that the critical kernel width for bimodality is $h*_2 = 0.48$.
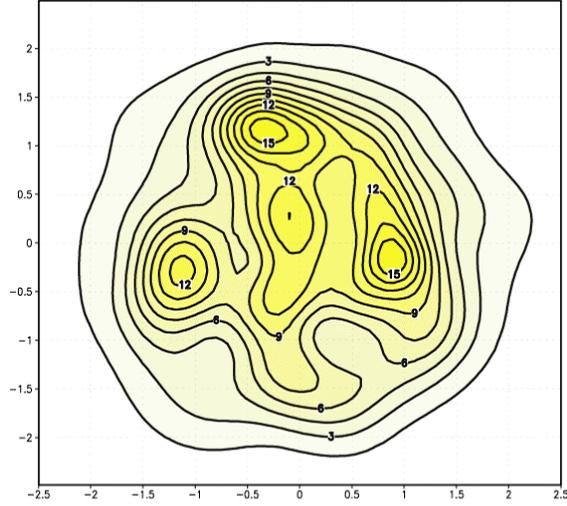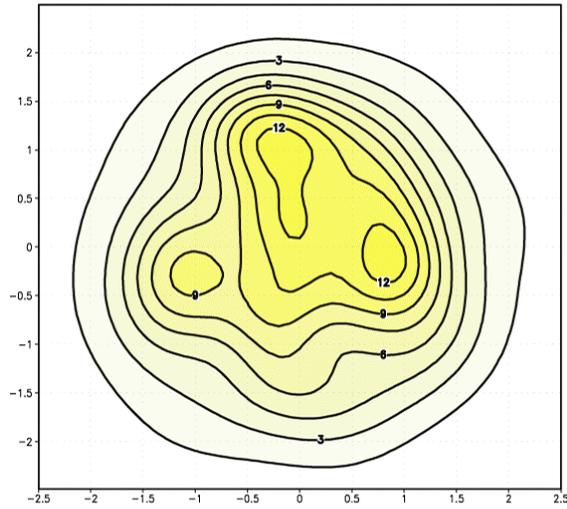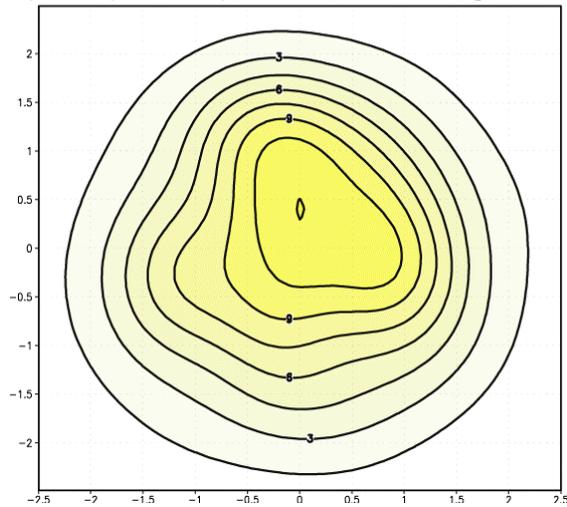
Figure 3 Bi-dimensional probability density function (PDF) of the first two PCs of 500-hPa height anomalies in winter (for the EOFs shown in Fig. 2), computed with different values of the Gaussian kernel width h (see text). a) : h = 0.3 ; b) : h = 0.4 ; c) : h = 0.5.

PDFs from 500 samples of pseudo-random red-noise data with the same number of data and the same mean, variance and lag-1 auto-correlation of the actual PCs were then computed, using increasing kernel width. The histograms in Fig. 4 show the frequency of PDFs with different number of modes for a few selected values of h. For $h = 0.3$, a very large proportion of the red-noise PDFs are multimodal, with more than half of them displaying at least four modes. Therefore, no significance statement can be made about the four modes shown in the top panel of Fig. 3. With $h = 0.4$, most of the red-noise PDFs are multimodal; however, a three-modal PDF (as estimated from actual PCs with this kernel width) is found in just less than 10% of the cases. If we are simply concerned with the existence of more than one mode, we found that 88% of the red-noise PDFs are unimodal when the critical width for bimodality of the PC sample was used. We can therefore attach an 88% level of confidence to the multimodality of the observed PC sample.

Is 88% confidence high enough to state that multiple regimes exist in the Northern Hemisphere circulation? Many statisticians would not be comfortable with such a claim unless 95% confidence would be achieved. From the red-noise PDF estimates (see again Fig. 4), one finds that a critical kernel width of 0.51 would be required in the PC sample to exceed the 95% confidence level. But should one reject the multimodality hypothesis because the actual value $h*_2 = 0.48$ gives it a 12% probability to be incorrect for this specific dataset?
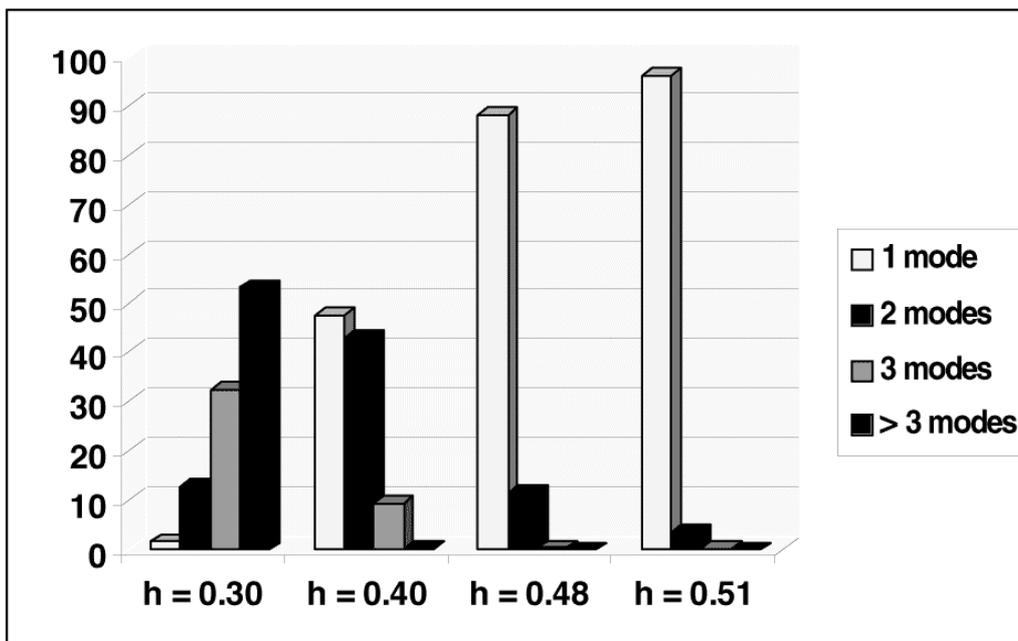
Figure 4 Histograms showing the proportion of PDFs with different number of modes (i.e. local maxima), computed from red-noise samples with increasing values of the Gaussian kernel width h .

There are different ways to answer this question. One may point out that the typical duration of extratropical flow regimes is of the order of $1 \sim 2$ weeks; therefore a monthly time averaging is bound to weaken the signature of multiple regimes in the PDF. (The PDFs computed by e.g. Hansen and Sutera (1986), Kimoto and Ghil (1993a), Marshall and Molteni (1993) were derived from low-pass-filtered daily data or 5-day means). On the other hand, if one wants to stick to monthly means because of the wider data availability (for example, in comparisons with GCM simulations), one should evaluate if the required level of significance actually allows a distinction between a unimodal and a multimodal distribution, *given the size of the observed record.*

To address this issue, an *analytical* 3-modal PDF was defined by a superposition of multi-normal functions centred in different points of the PC plane, close to the local maxima of the actual PDF of PC data. The analytical form of this distribution (plotted in Fig. 5) was used to generate 500 samples of pseudo-random

data with the same size as the observed monthly-mean record. From each of these samples, the PDF was recomputed using the Gaussian kernel estimator, and the critical kernel width for bimodality was recorded. It was found that 58% of the samples from the 3-modal distribution had a larger critical width than the actual PC data; however, only in 38% of the cases the critical width for multimodality was larger than the value (0.51) providing a 95% confidence level in the test against a unimodal red-noise.
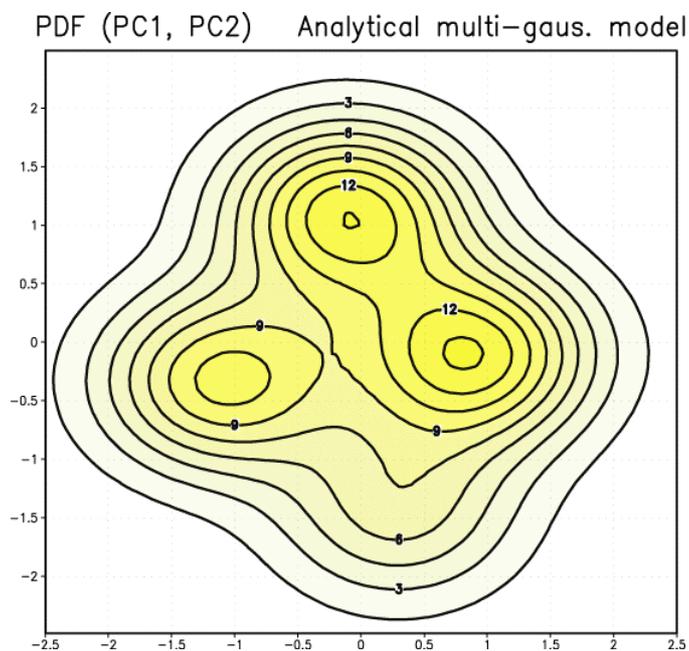
Figure 5 PDF corresponding to an analytical 3-modal distribution in the PC1-PC2 plane, used for the estimation of "useful" significance levels for the monthly-mean record.

These results show that, if we require a 95% confidence limit for multimodality in a sample as large as the 44-winter record of monthly means, we have a large chance (62%) that Silverman's (1981) test would *fail* to recognize a significant multimodality in a PDF derived from the analytical distribution in Fig. 5. Therefore, the fact that the multimodality of the monthly-mean PC distribution does not pass a 5% significance test can hardly be used to refuse the regime hypothesis. One may wish to analyse the observed record with a different time frequency, higher dimensionality or more sophisticated clustering techniques to get a larger confidence. Here, we will accept the existence of multiple flow regimes as a reasonable working hypothesis, and proceed to explore the issue of regime predictability in the context of GCM simulation.

4. Model climatology and variability

The numerical experiments used to investigate regime predictability in the following section were run with a 7-level version of the intermediate-complexity atmospheric GCM developed at the International Centre for Theoretical Physics (Molteni 2003). The model (nicknamed SPEEDY, for Simplified Parametrization, primitivE-Equation DYnamics) is based on a hydrostatic, spectral dynamical core developed at the Geophysical Fluid Dynamics Laboratory (see Held and Suarez 1994), using the vorticity-divergence form described by Bourke (1974). A set of physical parametrization schemes has been developed from the same basic principles used in more complex GCMs, with a number of simplifying assumptions which are suited to a model with a coarse vertical resolution (details can be found in the on-line appendix to Molteni (2003) available at http://www.ictp.trieste.it/~moltenif/speedy-doc.html). These include short- and long-wave radiation, large scale condensation, convection, surface fluxes of momentum, heat and moisture, and vertical diffusion. The model is currently run with a T30 spectral truncation in the horizontal; at this resolution, one year of integration takes 23 minutes on a single Xeon 2.4 GHz processor.

The experiments consist of an ensemble of 8 simulations for the period 1954-1999, using the EOF-reconstructed observed SST by NCEP (Smith et al 1996) as boundary conditions. The model climatology of upper-air fields and its variability on multi-decadal time-scales are validated here using the NCEP/NCAR re-analysis (Kalnay et al. 1996). A number of verification maps for the winter and summer climatology of these experiments can be found on-line at the web site above.

With respect to the 5-level version described by Molteni (2003), the wintertime climatology of the 7-level model used here shows a much improved stationary wave pattern in the Pacific sector of the Northern Hemisphere, while it still suffers from an underestimation of the stationary wave amplitude over the Atlantic. These features are illustrated by the comparison of the December-to-February (DJF) mean of 500-hPa height for the NCEP/NCAR re-analysis and the ensemble-mean model simulation in the 1979-1998 period (Fig. 6a-b). The systematic error of the model (defined by the ensemble-mean minus re-analysis difference in Fig. 6c) reaches a large negative amplitude (over 150 m) close to the British Isles, but it just exceeds 60 m (in absolute value) in the North Pacific.
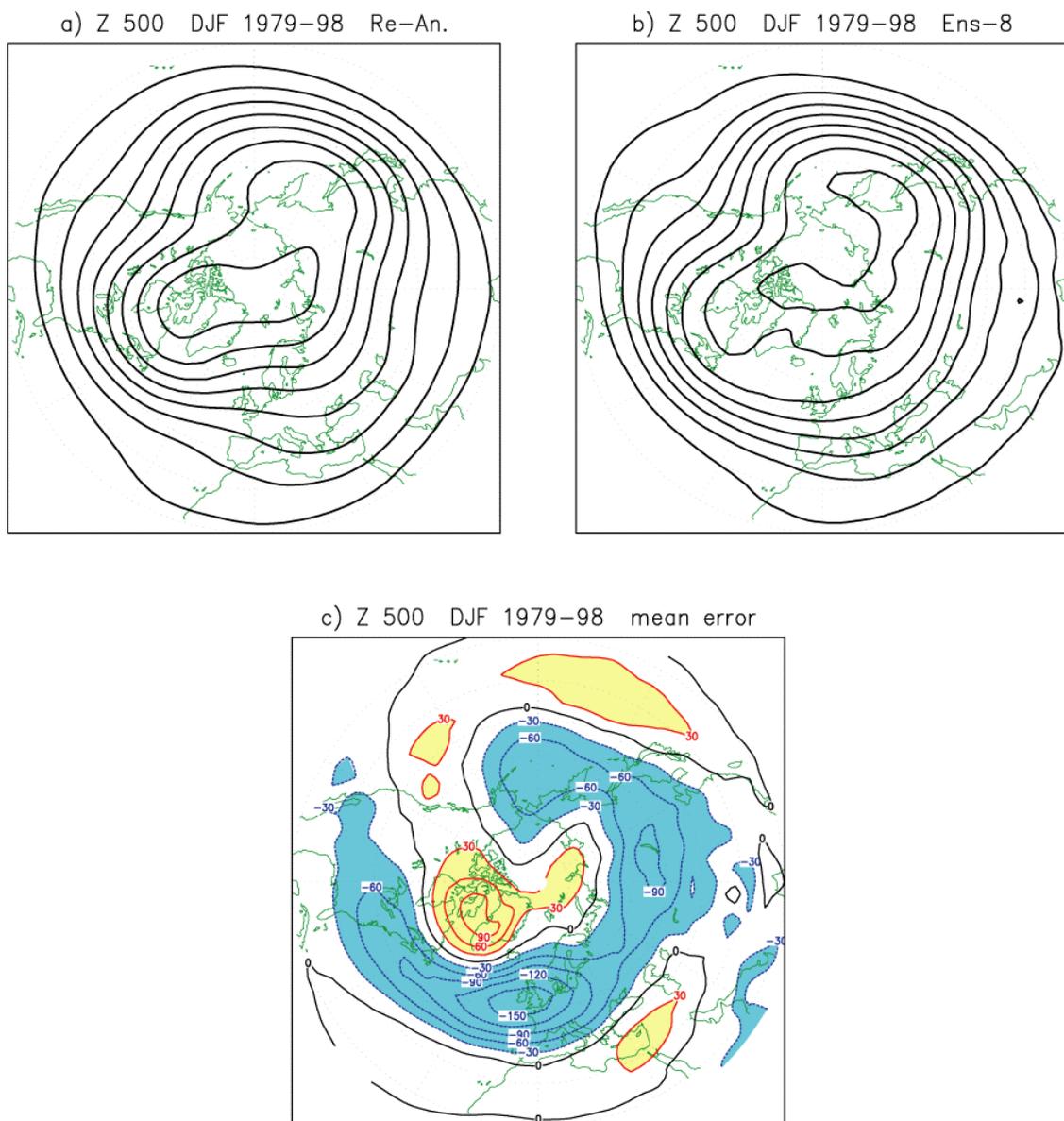
Figure 6a) : Mean field of 500-hPa height for the DJF season, from the NCEP/NCAR re-analysis in winters 1978/79 to 1997/98. b) : as in a), but from an ensemble of 8 simulations with the Int'l. Centre for Theoretical Physics (ICTP) AGCM forced by observed SST. c) : mean model error (ensemble-mean minus re-analysis). Contour interval 100 m in (a) and (b), 30 m in (c).

The good simulation of the Pacific stationary waves is reflected in a realistic reproduction of the extratropical response to tropical SST anomalies in opposite ENSO phases. Composites of DJF 500-hPa height in 8 warm and 8 cold ENSO events during the 1954-1999 period are shown in Fig. 7 for the model ensemble mean and the NCEP/NCAR re-analysis. In order to eliminate possible effects of inter-decadal variability on such composites, the anomaly for each event has been defined with respect to a 10-winter mean centred around the selected winter. Both the pattern and the amplitude of the model responses compare well with the observations in the Pacific/North American region. For the cold-event composite, even the features on the eastern Atlantic are well simulated, yielding a large spatial correlation (81%) between the model and re-analysis composites on the full hemispheric domain.
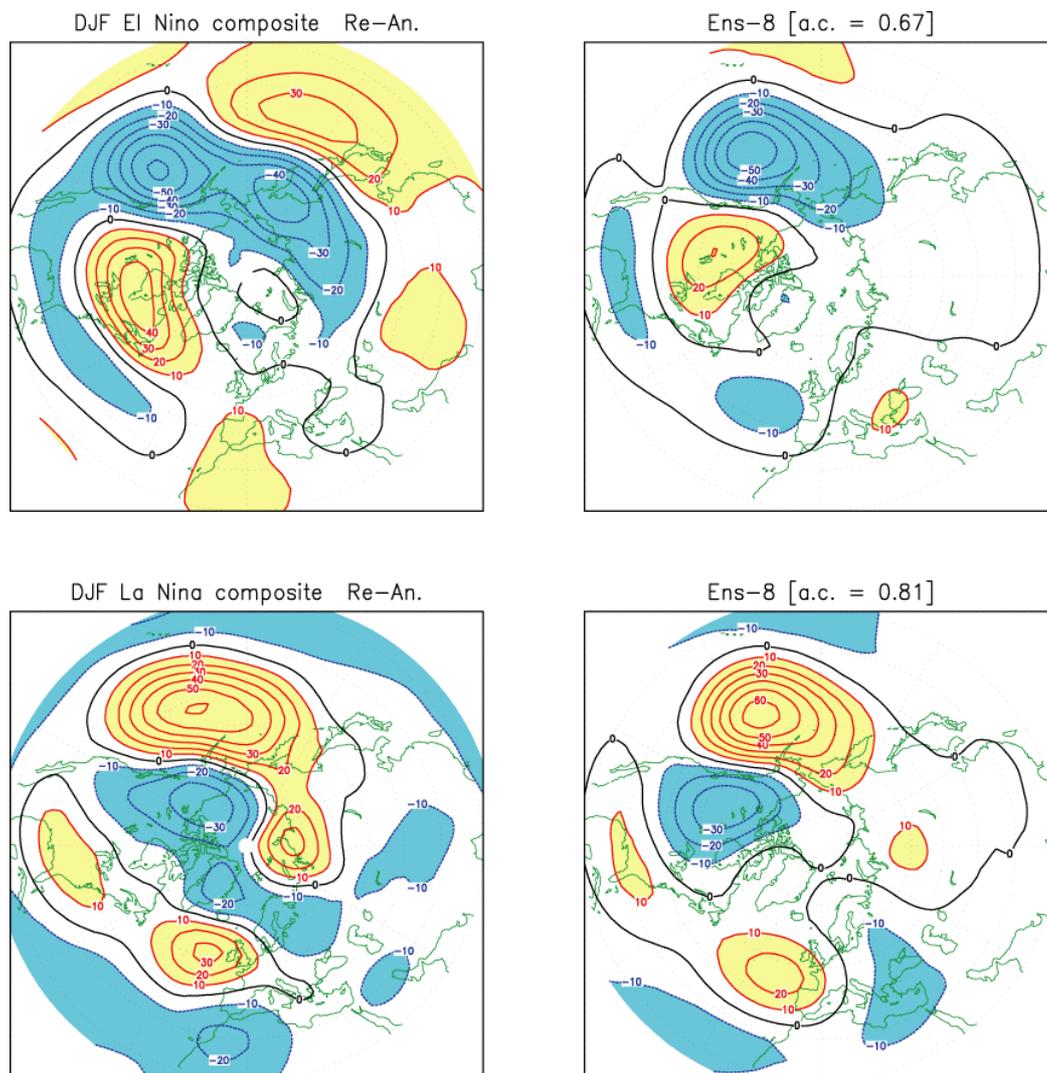
Figure 7a) : Composites of 500-hPa height anomalies in DJF for 8 warm (top row) and 8 cold (bottom row) ENSO events in the 1955-to-98 period, from the NCEP/NCAR re-analysis (left column) and from an ensemble of 8 AGCM simulations forced by observed SST (right column). Contour interval 10 m. The anomaly correlation (a.c.) between observed and modelled patterns is shown in brackets.

Since model reproduces ENSO-related interannual variability in a realistic way, interdecadal differences related to the well documented SST trends in the tropical oceans should also be detectable in the simulations. Significant differences in Pacific SST before and after 1976 have been documented in many studies (e.g.

Zhang et al. 1997; Trenberth and Hoar 1997); therefore, we will focus on the differences in geopotential height between the mean of 22 winters from 1976/77 to 1997/98 and the preceding 22-winter period. The 500-hPa height interdecadal differences for the 8-member ensemble-mean and the re-analysis are shown in the top panels of Fig. 8. The spatial correlation between the modelled and observed patterns is 43% over the Northern Hemisphere, while the average amplitude for the model is about half of the observed amplitude (as in Hoerling et al. 1999). Looking at different regions, one finds a better correlation over the Pacific/North American sector, while the modelled negative anomalies over the North Atlantic and Siberia are located in a more southerly position than the corresponding observed features.
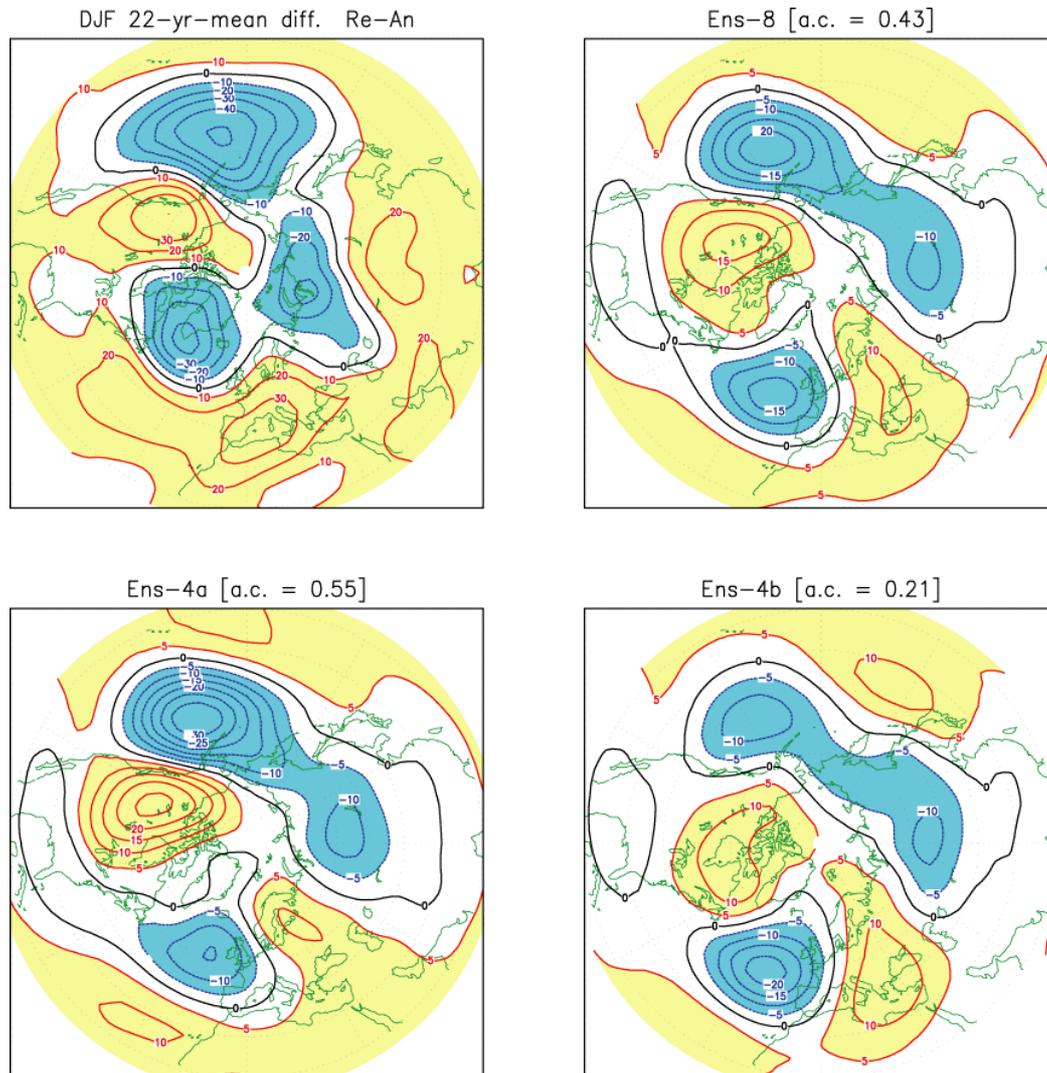
Figure 8a) : Mean difference of 500-hPa height in DJF between two consecutive 22-year periods (1977/98 minus 1955/76), from the NCEP/NCAR re-analysis. b) : as in a), but from an ensemble of 8 AGCM simulations forced by observed SST (Ens-8). c) and d) : as in (b) but from two ensembles (Ens-4a and Ens-4b) of 4 members each, selected according to the anomaly correlation between the observed and the simulated inter-decadal difference. Contour interval 10 m in (a), 5 m in (b) to (d). The correlation between observed and modelled patterns is shown in brackets.

Computing ENSO composites and interdecadal differences from individual ensemble members, one finds that the former are quite consistent among individual members, while larger differences are found in the simulation of inter-decadal variability. If the difference patterns of 500-hPa height between the post-1976 and pre-1976 periods for individual experiments are correlated with the observed difference, the correlation coefficients over the Northern Hemisphere range from slightly negative to greater than 60%. To illustrate the effects of internal atmospheric variability, the bottom panels of Fig. 8 show the interdecadal differences for

the ensemble means of two sub-ensembles, one including the 4 members with the best interdecadal correlations, the other one including the 4 least-correlated members (hereafter referred to as Ens-4a and Ens-4b respectively, while Ens-8 will indicate the full ensemble). With respect to Ens-4b, the difference pattern for Ens-4a shows much stronger amplitude over the Pacific and North America, and a northward shift of the Atlantic anomalies. As a result, the correlation with the observed pattern (top left) is raised to 55% for Ens-4a, while it drops to 21% for Ens-4b.

5.   ENSO-related variability in flow regime statistics

In this section, the impact of ENSO on the interannual and interdecadal variability of flow regimes will be investigated through a PDF analysis of the wintertime monthly-mean anomalies of 500-hPa height simulated by the SPEEDY model. Since the ensemble simulation gives 8 times more data than the observed record, the issue of multimodality can be addressed with much stricter levels of statistical significance.

Following the same procedure used in Sec. 3c, we first computed the two leading EOFs of monthly-mean anomalies in DJFM for the Northern-Hemisphere extratropics (shown in Fig. 9). When compared with its counterpart in the re-analysis (Fig. 2), the first EOF of the model shows a more zonally symmetric pattern, with features of comparable amplitude in the Atlantic and the Pacific sectors. Taking the same sign as in CMP, the first model EOF can be interpreted as the negative phase of the Arctic Oscillation pattern, similarly to the first re-analysis EOF. On the other hand, stronger differences between model and re-analysis are found for the second EOF. The two patterns resemble each other in the Pacific sector, but the model EOF-2 has a much smaller amplitude in the Atlantic and Eurasian region. In the model, the second EOF may be described as the manifestation of the Pacific/North American teleconnection pattern, rather than a COWL-related pattern (Wallace et al. 1996).
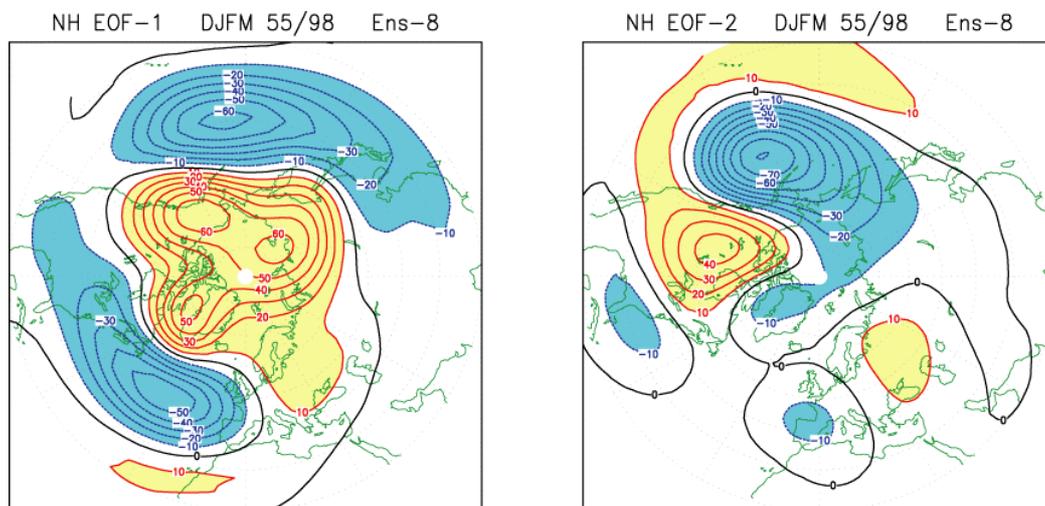
Figure 9 First two EOFs of 500-hPa height monthly-mean anomalies over the Northern Hemisphere (20-90 N) in the 44 winters (DJFM) 1954/55 to 1997/98, computed from an ensemble of 8 AGCM simulations forced by observed SST. The EOFs are scaled to represent the anomaly patterns corresponding to one standard deviation of the associated PCs. Contour interval 10 m.

Because of the differences in the EOF patterns, the subspace spanned by the first two PCs does not have exactly the same meaning in the model as in the observation. Still, it is interesting to see whether any variability in the model PDFs associated with different periods or ENSO phases have a counterpart in re-analysis data. Therefore, in the following we shall compare model and observed PDFs for the same subsamples, bearing in mind that only for the model simulations we have enough data to estimate the significance of multimodality using just a fraction of the total sample. As for the observed data, bi-

dimensional PDFs of the model PCs have been computed with an iterative Gaussian kernel estimator (Silverman 1986).
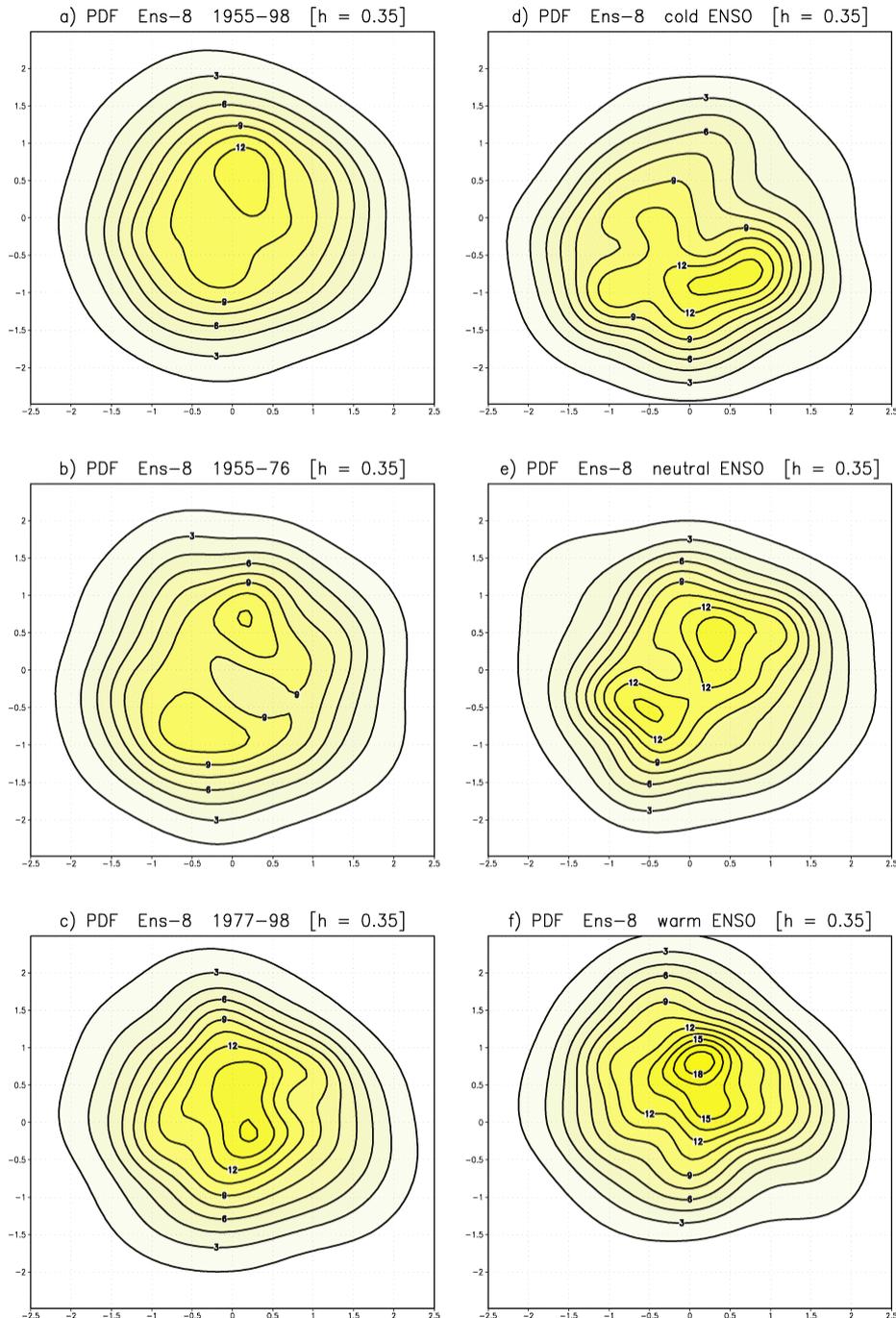
Figure 10 Bi-dimensional PDFs of the first two PCs of 500-hPa height anomalies in winter over the Northern Hemisphere, from an ensemble of 8 GCM simulations forced by observed SST in the period 1954 to 1999. a) : PDF for the full 44-winter period. b) : PDF for the 22-winter period 1955 to 1976. c) : PDF for the 22-winter period 1977 to 1998. d) to f) : PDFs for samples of 15 winters with cold (d), neutral (e) and warm (f) ENSO phase.

Fig. 10 shows PDFs in the PC1-PC2 plane computed from model anomalies in the 44-year period (DJFM 1955-to-1998, where the year refers to JFM), in its first and second halves (1955-1976 and 1977-1998 respectively), and in three 15-winter subsamples defined by sorting the available winters into cold, neutral and warm ENSO phases according to the bivariate index by Smith and Sardeshmukh (2000) (the cold-ENSO winter of 1998/99 was also included here). The PDF of model PCs is unimodal in the full sample and in its second half, but it is bimodal in its first half. The separation between the two modes occurs mainly along the

PC-2 axis, with a smaller difference along the PC-1 axis. When the significance of bimodality in the 1955-1976 period is tested using Silverman's (1981) procedure as described in Sec. 3a, the confidence level exceeds 99.5%.

Looking at the PDFs for different ENSO phases, one finds that bimodality in the model phase space is primarily generated during neutral (i.e. near-normal) ENSO winters. The confidence level estimated from the neutral-ENSO sample still exceeds 96%, despite the reduction in the number of data. During the cold and warm ENSO phases, the PDF appears to be shifted along the PC-2 axis (corresponding to a PNA-like pattern), but evidence of multiple regimes is either weak (cold phase) or totally absent (warm phase). Comparing the PDF for the second half of the record with the PDF of warm ENSO events, it appears that the unimodal nature of the distribution in the 1977-1998 period is closely related to higher frequency of warm events in that period. This is consistent with the modelling results of Molteni and Corti (1998) and Straus and Molteni (2003), who found that evidence of multiple regimes in the Pacific/North American circulation was suppressed during warm ENSO events.

In order to verify whether the model behaviour has any correspondence in the observations, Fig. 11 shows PDFs from the re-analysis record for the 1955-1998 and 1977-1998 periods, and for the two 15-winter subsamples of neutral and warm ENSO cases. Consistently with model simulation, the multimodal nature of the 44-year PDF is made more evident by the selection of neutral ENSO years. Also, the PDFs for the second half of the record and for warm events resemble each other, and are dominated by one mode, corresponding to the positive phase of the COWL pattern (i.e. the regime in the top row of Fig. 1).
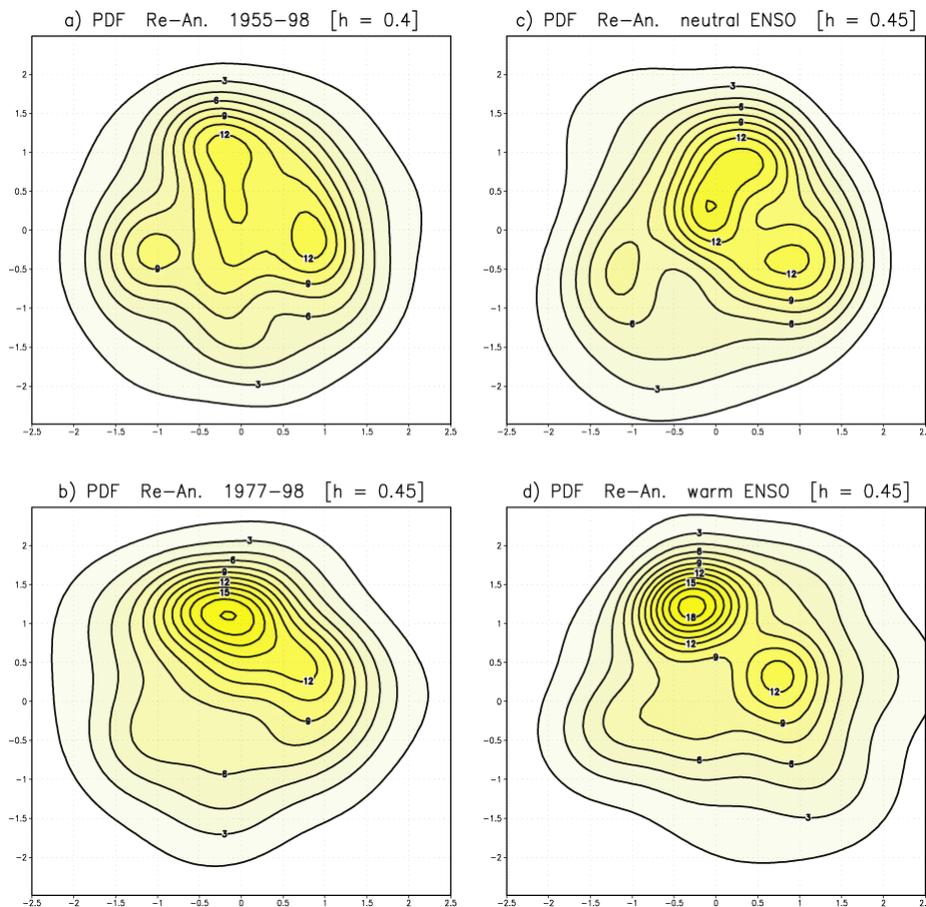
Figure 11 Bi-dimensional PDFs of the first two PCs of 500-hPa height anomalies in winter over the Northern Hemisphere, from the NCEP/NCAR re-analysis in the period 1954 to 1999. a) : PDF for the full 44-winter period. b) : PDF for the 22-winter period 1977 to 1998. c) and d) : PDFs for samples of 15 winters with neutral (c) and warm (d) ENSO phase.

As a further test of the robustness of the modelled inter-decadal variations in flow regime properties, Fig. 12 compares the PDFs for the 1955-1976 and 1977-1998 periods, evaluated from two 4-member subsets of the ensemble, namely the Ens-4a and Ens-4b subsets with "good" and "bad" simulations of interdecadal variability (see the bottom row of Fig. 8). It is found that the bimodal vs. unimodal nature of the PDF in the two periods is reproduced in both 4-member ensembles, although the difference is more evident in the ensemble (Ens-4a) that better reproduced the observed spatial pattern of interdecadal difference.
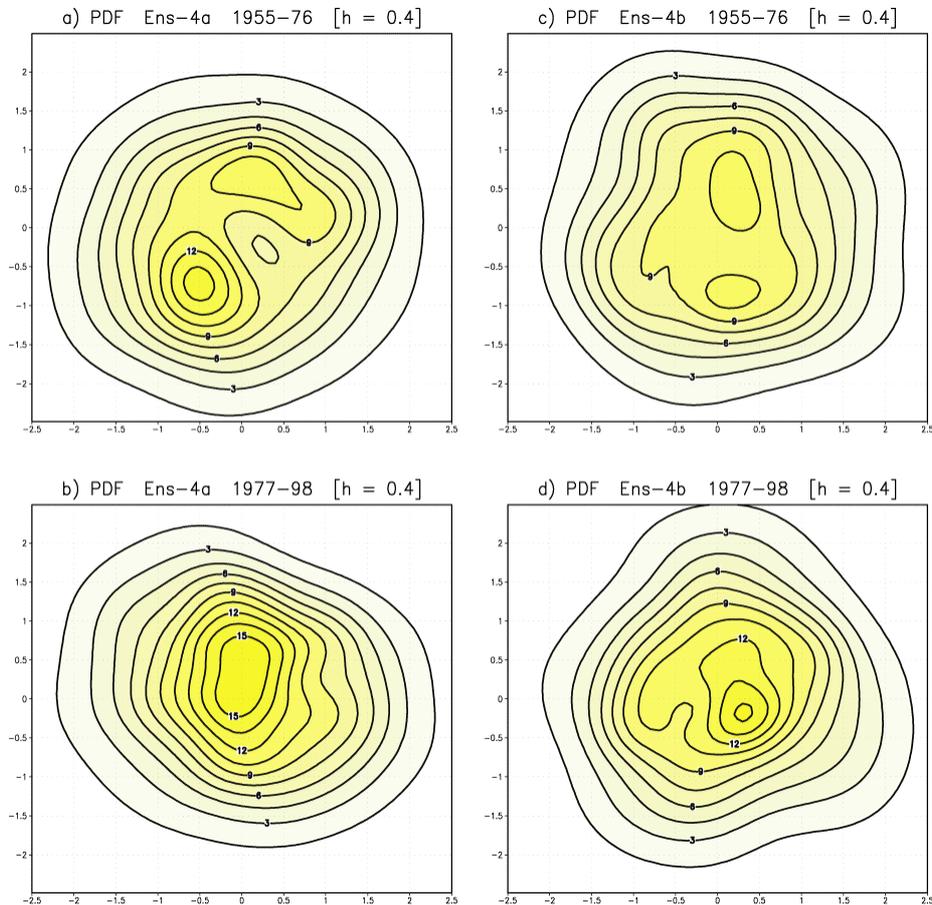
Figure 12 Bi-dimensional PDFs of the first two PCs of 500-hPa height anomalies in winter over the Northern Hemisphere, from two ensembles (Ens-4a and Ens-4b) of 4 members each. Top row: PDF for the 22-winter period 1955 to 1976, from Ens-4a (a) and Ens-4b (c). Bottom row: PDF for the 22-winter period 1977 to 1998, from Ens-4a (b) and Ens-4b (d).

6.   Conclusions

In this paper, we have addressed the issue of regime predictability by focussing on two aspects:

▪   the significance of regimes detected in the observational record;

▪   the impact of interannual and interdecadal variability of the ENSO phenomenon on the statistical properties of extratropical flow regimes as simulated by a GCM.

On the first issue, we applied Silverman's (1981) methodology to test the significance of multimodality in the distribution of the two leading PCs of monthly-mean anomalies of 500-hPa height in winter, using data from a 44-year sample of NCEP/NCAR re-analyses. We estimated the confidence level for the existence of *at least* two regimes to be close to 90%. We also showed that asking for a stricter significance level has little meaning in the case of the observed monthly-mean record, since the required level of PDF smoothing would prevent the detection of multimodality even for samples generated by an analytical tri-modal distribution.

With regard to the second issue, the results of our model simulations indicate that predictability "of the second kind" (Lorenz 1975) is indeed present in regimes statistics, as a result of their dependence on anomalous boundary forcing related to the ENSO phenomenon. In addition to a shift of the PDF along a PNA-like axis, ENSO variations change the number of regimes detected in the phase space spanned by the two leading PCs. For both model and observed data, multimodality is best revealed in years with weak ENSO forcing. On the other hand, our intermediate-complexity AGCM shows no evidence of multiple regimes during the warm ENSO phase. Taking into account that the two leading PCs of model data tend to emphasise low-frequency variability in the Pacific/North American region, the latter result is consistent with the statistics of Pacific regimes from the quasi-geostrophic simulations by Molteni and Corti (1998) and the GCM ensembles analysed by Straus and Molteni (2003).

Such findings give a more complex picture of the ENSO impact on regimes than the one advocated by Molteni et al. (1993) and Palmer (1993), which was based on a change of frequencies in quasi-invariant regimes. While the frequency-change hypothesis is likely to be appropriate for relatively small variations of the atmospheric forcing terms (such as those related to greenhouse gas and aerosol concentrations in the 20th century), strong ENSO events alter the forcing of stationary waves in a substantial way, possibly leading the atmospheric system through bifurcation points. Interestingly, a variation in the number of regimes as a function of the ENSO phase has also been detected in the Asian monsoon simulations described by Molteni et al. (2003).

The goal of this paper was to show that the concept of regime predictability has robust statistical and dynamical foundations. We have not addressed the issue in such a detail to quantify, for example, its practical relevance in the case of seasonal forecasting. However, cluster analysis has been increasingly used in recent years to condense the information provided by ensemble forecasts, so that practical examples are already available in the literature (e.g. Brankovic and Molteni 1997). We are well aware that many aspects of atmospheric predictability may be addressed without the need of introducing flow regimes. On the other hand, regimes provide a valuable framework to investigate the non-linear aspects of the relationship between anomalous forcing and atmospheric response, and may play an important role in understanding the mechanism of climate changes originated by the interactions of "internal" climate dynamics and anthropogenic perturbations of the Earth's energy balance.

References

Bourke W, 1974: A multilevel spectral model. I. Formulation and hemispheric integrations. *Mon. Wea. Rev.*, **102**, 687-701.

Brankovic, C. and Molteni F., 1997: Sensitivity of the ECMWF model northern winter climate to model formulation. *Climate Dyn.*, **13**, 75-101

Cehelsky, P. and Tung K.K., 1987: Theories of multiple equilibria and weather regimes—A critical re-examination. Part II: Baroclinic two-layer models . *J. Atmos. Sci,* **44**, 3282–3303.

Charney, J.G, and Straus D.M., 1980: Form-drag instability, multiple equilibria and propagating planetary waves in baroclinic, orographically forced, planetary wave systems. *J. Atmos. Sci.,***37***,* 1157-1176.

Cheng, X. and Wallace J.M., 1993. Cluster analysis on the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50**, 2674-2695.

Chessa, P.A. and Lalaurette F., 2001: Verification of the ECMWF Ensemble Prediction System forecasts: A study of large-scale patterns**.** *Weather and Forecasting,* **5**, 611–619.

Corti, S., Molteni F., and Palmer T.N., 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799 - 802.

D'Andrea, F. and Vautard R., 2001: Extratropical low-frequency variability as a low dimensional problem. Part I: a simplified model. *Q. J. R. Meteorol. Soc*, **127**, 1357-1374.

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703-713.

Ghil, M., 1987: Dynamics, statistics and predictability of planetary flow regimes, *Irreversible Phenomena and Dynamical Systems Analysis in the Geosciences*, C. Nicolis and G. Nicolis (Eds.), D. Reidel, Dordrecht/Boston/Lancaster, pp. 241-283

Gleeson, T.A., 1970: Statistical-dynamical predictions. J. Appl. Meteor., 9, 333-344.

Haines, K., and Hannachi A., 1995: Weather regimes in the Pacific from a GCM. *J. Atmos. Sci.*, **52**, 2444-2462

Hansen, A.R. and Sutera A., 1986: On the probability density distribution of large-scale atmospheric wave amplitude. *J.Atmos.Sci.*, **43**, 3250-3265.

Hansen, A.R. and Sutera A., 1995: The probability density distribution of planetary-scale atmospheric wave amplitude revisited. *J.Atmos.Sci.*, **52**, 2463-2472.

Held, I.M. and Suarez M.J., 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Am. Meteorol. Soc,* **75**, 1825–1830.

Hoerling M.P., Hurrell J.W. and Xu T., 2001: Tropical origins for recent North Atlantic climatic change. *Science*, **292**, 90-92.

Houtekamer, P.L., Lefaivre L., Derome J., Ritchie H., Mitchell H.L., 1996: A system simulation approach to ensemble prediction**.** *Mon. Wea. Rev*, **124**, 1225–1242.

Hsu, C.J., and Zwiers F., 2001: Climate change in recurrent regimes and modes of Northern Hemisphere atmospheric variability. *J Geophy.Res.-Atmos* , **106**, 20145-20159.

IPCC, 2001: *Climate Change 2001: The scientific basis*. Eds: Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J., Dai X., Maskell K. and Johnson C.A. Cambridge University Press, 881 pp.

Kalnay, E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Chelliah M., Ebisuzaki W.,.Higgins W, Janowiak J., Mo K.C., Ropelewski C., Wang J., Leetmaa A., Reynolds R., Jenne R., and Joseph D., 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, **77**, 437-471

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

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

Lorenz, E.N., 1975 : Climate predictability : The physical basis of climate modeling. WMO, *GARP Publ.Series*, **16**, 132-136.

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

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

Mo, K.C. and Ghil M., 1987: Statistics and dynamics of persistent anomalies. *J.Atmos.Sci.*, **44**, 877-901.

Mo, K.C. and Ghil M., 1988: Cluster analysis of multiple planetary flow regimes. *J.Geophys.Res.*, **93D**, 10927-10952.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, *20, 175-191*

Molteni F. and Corti S., 1998: Long term fluctuations in the statistical properties of low-frequency variability: dynamical origin and predictability. *Q. J. R. Meteorol. Soc.* **124**, 495-526

Molteni, F., S. Tibaldi and T.N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Q.J.R.Meteorol.Soc.*, **116**, 31-67.

Molteni, F., Ferranti L., Palmer T.N. and Viterbo P., 1993: A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *J.Clim.*, **6**, 777-795.

Molteni, F., Buizza R., Palmer T.N. and Petroliagis T., 1996: The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119

Molteni, F., Corti S., Ferranti L. and Slingo J.M., 2003: Predictability Experiments for the Asian Summer Monsoon: Impact of SST Anomalies on Interannual and Intraseasonal Variability. *J. Climate.*, in press.

Nitsche, G., Wallace J.M. and Kooperberg C., 1994: Is there evidence of multiple equilibria in planetary wave amplitude statistics? *J.Atmos.Sci.*, **51**, 314-322.

Palmer, T.N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Am. Meteorol. Soc.*, **74**, 49-66.

Palmer, T.N., 1999. A nonlinear dynamical perspective on climate prediction. *J. Climate*, 12, 575-591.

Reinhold, B., and Pierrehumbert R.T., 1982: Dynamics of weather regimes: quasi-stationary waves and blocking. *Mon.Wea.Rev.*, **110**, 1105-1145.

Silverman, B.W., 1981: Using kernel-density estimators to investigate multimodality. *J.Roy.Stat.Soc.*, **B43**, 97-99.

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

Smith, C.A. and Sardeshmukh P., 2000: The effect of ENSO on the intraseasonal variance of surface temperature in winter. *Int. J. Climatology*, **20**, 1543-1557.

Smith, T.M., Reynolds R.W., Livezey R.E. and Stokes D.C., 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403-1420.

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

Straus, D. and Molteni F., 2003: Flow regimes, SST forcing, and climate noise: Results from large GCM ensembles. *J. Climate*, submitted.

Thompson, D.W.J., and J.M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys.Res.Lett*., 25, 1297-1300.

Toth, Z., and Kalnay E., 1993: Ensemble Forecasting at NMC: The Generation of Perturbations**.** *Bull. Am. Meteorol. Soc.*, **74**, 2317–2330

Trenberth, K.E., and. Hoar T.J, 1997: El Niño and climate change. *Geophys. Res. Lett*., **24**, 3057-3060.

Vautard, R. 1990. Multiple weather regimes over the North Atlantic. Analysis of precursor and successors. *Mon. Wea. Rev.* **118**, 2056-2081

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

Wallace, J.M., Zhang Y. and Bajuk L., 1996: Interpretation of interdecadal trends in Northern Hemisphere surface air temperature. *J.Clim.*, 9, 249-259.

Zhang, Y., Wallace J.M. and Battisti D.S., 1997: ENSO-like Interdecadal Variability: 1900–93. J.*Clim*: **10**, 1004–1020.