Circulation Regimes: Chaotic Variability versus SST-Forced Predictability

DAVID M. STRAUS

George Mason University, Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

SUSANNA CORTI

Institute of Atmospheric Sciences and Climate, National Research Council, Bologna, Italy

FRANCO MOLTENI*

International Centre for Theoretical Physics, Trieste, Italy

ABSTRACT

The circulation regimes in the Pacific–North American region are studied using the NCEP–NCAR reanalyses for the 18-winter period (1981/82–1998/99; NCEP18) and for the 54-winter period (1948/49–2001/02; NCEP54). The sampling properties of the regimes are estimated using very large ensembles (of size 55) of winter simulations made for the NCEP18 period with the atmospheric general circulation model of the Center for Ocean–Land–Atmosphere Studies, forced by observed SST and sea ice.

The regimes are identified using a modified version of the k-means method. From the NCEP54 dataset a set of four clusters was found [i.e., the Alaskan ridge (AR), Arctic low (AL), Pacific trough (PT), and the Arctic high (AH)], which are significant (vis-à-vis a multinormal background), and more reproducible (within randomly chosen half-length samples) than would be expected from a multinormal process. The frequency of occurrence of the PT (AH) has increased (decreased) significantly during the past two decades.

The PT cluster obtained from NCEP18 dataset more closely resembles the El Niño–forced seasonal mean pattern of recent decades than it does the traditional PNA.

The GCM simulates the AR, AL, and PT clusters (but not the AH). The simulated AR and PT patterns have errors (cf. the NCEP18 results), which are outside the range of internal variability. The simulated frequency of occurrence agrees with the NCEP18 results within sampling variability.

The differences in cluster properties of the PT and AR regimes between the NCEP18 and NCEP54 datasets are due to changes in SST forcing, not sampling error.

Year-to-year changes in the frequency of occurrence of the PT, AL, and AR clusters in the simulations and the NCEP18 dataset are generally consistent with each other.

1. Introduction

The notion that the large-scale circulation of the extratropical atmosphere can be described by the alternation of circulation regimes (or weather regimes), in which anomalies in the amplitude and phase of planetary waves are dynamically equilibrated by variations in diabatic energy sources and nonlinear interactions with synoptic-scale eddies, can be traced back to a number of seminal studies in the 1980s (e.g., Reinhold and Pierrehumbert 1982; Vautard and Legras 1988). Since then, circulation regimes have been simulated and diagnosed using a variety of models, ranging from quasigeostrophic (e.g., Marshall and Molteni 1993; D'Andrea and Vautard 2001; Kondroshov et al. 2004) to balance equation models (Itoh and Kimoto, 1999) to primitive equation general circulation models (GCMs) (e.g., Hansen and Sutera 1990; Haines and Hannachi 1995; Monahan et al. 2000). Although regime identifi-

* Current affiliation: European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom.

Corresponding author address: David Straus, George Mason University/COLA, 4041 Powder Mill Rd., Suite 302, Calverton, MD 20705.
E-mail: straus@cola.iges.org

cation in the actual observed record remains a non-trivial statistical problem (see the review by Molteni et al. 2006), recent studies have obtained results that are consistent across different methodologies (Alhamed et al. 2002; Arnott et al. 2004; Cassou et al. 2004; Robertson and Mechoso 2003).

The methods used to establish the existence and significance of regimes have, however, been questioned by a number of authors. Recently Stephenson et al. (2004) and Hsu and Zwiers (2001) have criticized the specific methodology of Corti et al. (1999) who sought evidence for regimes in the two-dimensional probability density function (PDF) of winter monthly mean height fields based on reanalyses. Stephenson et al. (2004) argue in more general terms that rigorous hypothesis testing is essential and that a null hypothesis of a multinormal distribution for large-scale atmospheric fields may be very hard to reject. This was also the conclusion of Toth (1991), although only in a phase-averaged sense.

In this paper, regimes are identified from the mathematical technique of cluster analysis. Since the concept of regimes is still somewhat controversial, we refer to specific results obtained in this paper in terms of clusters. The issue of careful hypothesis testing in our analysis will be taken up again in sections 3 and 4. The difficulty in determining the number of clusters (Christiansen 2006) will also be discussed.

In the last decade, a number of studies have investigated the possible dependence of atmospheric regime properties on the state of slowly varying components of the climate system, such as sea surface temperature (SST). By analogy with the behavior of low-order chaotic systems, it was argued that the first-order response to a moderate variation of boundary forcing should be manifested only by a change in the regime's frequency of occurrence (Molteni et al. 1993; Palmer 1993, 1999). It was also suggested that the observed interdecadal variability in Northern Hemisphere regime frequency may be one aspect of the atmospheric response to increased greenhouse forcing (Corti et al. 1999; Shindell et al. 1999; Hsu and Zwiers 2001). However, recent studies have pointed out that large variations in the distribution of diabatic forcing (such as those associated with strong El Niño events) may lead to more substantial changes of regime properties, including the spatial pattern of the cluster "centroids" or even the number of regimes (e.g., Molteni and Corti 1998; Straus and Molteni 2004).

It is now widely recognized that the variation of circulation regime (cluster) properties may be an important issue for climate predictions, from seasonal forecasts to future climate scenarios. However, either when validating model results against observations or when comparing regime properties in different periods, one is faced with the problem of estimating the significance of the products of complex statistical tools. Specifically, one needs to quantify the uncertainty in regime properties arising from the internal, chaotic dynamics of the atmosphere and use such information to assess the predictability of such properties as a response to external variations in forcing terms.

Given the limitations of the observed record of upper-air fields, it is very difficult to address these issues from reanalysis data. However, ensembles of GCM simulations can provide reliable estimates of the effects of internal versus external variability in cluster properties. In this paper, we make use of the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data (Kalnay et al. 1996) and of a large set of seasonal ensemble simulations with the Center for Ocean–Land–Atmosphere Studies (COLA) atmospheric GCM, previously analyzed by Straus and Molteni (2004) to investigate three specific topics:

1) To what extent are the properties of model-simulated wintertime clusters over the Pacific–North American (PNA) region consistent with those obtained from reanalysis data (for the common period of 1981/82–1998/99)?

2) Are the differences between the observed cluster properties for the recent period (as above) and those for the extended period of 1948/49–2001/02 within the range of internal atmospheric variability?

3) Are interannual variations in cluster frequencies (at least partially) reproducible as a function of SST anomalies?

In principle it would be more consistent to pose question 2 as a comparison between the clusters occurring during two equal-length, nonoverlapping, periods. However, the shortness of the observed record dictates that we use as many winters as possible for one of the periods; the choice of the other (recent) period is dictated by the availability of the simulations.

Although the basic k-means clustering procedure adopted here (following Michelangeli et al. 1995, hereafter MVL) is the same as used by Straus and Molteni (2004, hereafter SM04), two important methodological differences between this study and SM04 should be pointed out.

1) Only quasi-stationary periods are examined, as in Toth (1991) and Toth (1993). The physical motivation is that, in order for circulation regimes (or preferred patterns) to also correspond to "weather re-

gimes" during which the character of the synoptic disturbances is unusually persistent, the large-scale flow must be quasi-stationary (Vautard 1990). The association of persistent large-scale patterns with such weather regimes is quite explicit in earlier works (see, e.g., Rex 1950; Namias 1964; Reinhold and Pierrehumbert 1982). The use of quasi-stationary prefiltering of the data in cluster analysis has been discussed by Toth (1992), Kimoto and Ghil (1993b), and Itoh and Kimoto (1999). To accomplish this filtering, only fields with a slow time evolution in the state space have been retained, a point to which we return later.

2) Only the seasonal cycle averaged over all years for a given period has been removed, followed by the definition of a single set of regimes (clusters). This is true for both reanalysis and model-simulated datasets. This contrasts with the methodology of SM04, in which the entire analysis, encompassing the removal of the seasonal cycle and the subsequent definition of the clusters, was carried out in a completely independent manner for each winter. Thus in particular the interannual circulation anomalies associated with strong El Niño–Southern Oscillation (ENSO) events play a direct role in the current analysis, whereas by design they played only an indirect role in SM04.

Section 2 of the paper gives a description of the datasets, experimental settings, and basic data reduction employed in the paper. The clusters from reanalysis data are described in section 3, where a detailed description of the cluster analysis procedure is given, the impact of the introduction of the "quasi-stationarity" criterion is discussed, the clusters obtained from reanalysis data both for the 54- and the 18-yr period corresponding to that of the simulations are presented, and the dependence of the cluster probability (or occupancy) on tropical Pacific SST is estimated.

Issues of significance and reproducibility for the clusters obtained from the reanalysis datasets are described in section 4, which can be skipped without loss of context by readers not interested in these more technical matters.

The regimes from the large ensemble of simulations and detailed comparisons of cluster statistics for modeled and observed regimes, designed to address the issues outlined in questions 1 and 2 above, are presented in section 5, while section 6 addresses question 3. In these sections, we estimate the contribution of atmospheric internal variability to fluctuations of cluster properties (such as centroid patterns and frequencies of occurrence) by comparing results from different sub-

samples extracted from the full 55-member ensemble dataset. Section 7 discusses the sensitivity of our results to the choice of the number of clusters.

Our conclusions, which are summarized and discussed in section 8, offer a rather optimistic view on the prospect of exploiting regime predictability in GCM simulations and suggest that internal atmospheric dynamics alone cannot account for all observed variations in regime statistics between the last half and the last two decades of the twentieth century.

2. Data and analysis methods

a. NCEP–NCAR reanalysis

The daily wintertime 200-hPa geopotential height and zonal wind fields[1] were accessed from the NCEP–NCAR reanalysis for two periods: 18 winters (1981/82–1998/99) and 54 winters (1948/49–2001/02). These datasets are referred to as NCEP18 and NCEP54, respectively. The shorter, more recent period was chosen to match the very large set of GCM ensemble simulations described in section 2b, while the longer period covers almost the whole available data record. For each winter, the 110-day period starting on 12 December was accessed. To ensure consistency with the simulated data only 0000 UTC data were used. The original fields, available on a 2.5° × 2.5° latitude–longitude grid, were interpolated to a Gaussian grid corresponding to triangular 42 (T42) truncation (approximately 2.8° × 2.8° latitude–longitude grid) in order that the reanalysis fields have the same horizontal resolution as the GCM fields.

b. GCM "Grand Ensemble" dataset

The atmospheric GCM used is version 2.2 of the COLA GCM, run at horizontal spectral resolution of triangular T63, with 18 sigma levels. It uses the dynamical core of the NCAR Community Climate Model version 3 (CCM3) described in Kiehl et al. (1998); otherwise, it is as described in Schneider (2002). The dependent variables of the model are spectrally treated, except moisture which is advected using the semi-Lagrangian technique. The land surface model, which is coupled to the atmospheric model, is the Simplified Biosphere Model (SSiB) documented in Xue et al. (1991). The parameterization of deep convection is the relaxed Arakawa–Schubert scheme (Moorthi and

---

[1] The equivalent barotropic nature of midlatitude intraseasonal variability, with amplitudes increasing with height, motivated the use of upper-tropospheric fields for the analysis.

Suarez 1992). For further details consult Schneider (2002).

For each of the 18 winters (1981/82–1998/99) we analyze an ensemble of 55 members. We label the initial conditions (ICs) for each winter as follows: ICs 1–5 are from the NCEP–NCAR reanalyses for 0000 UTC of 26–30 November. Initial conditions 6–10 are obtained from ICs 1–5 by adding a perturbation that consists of the difference between the analyses 12 h prior and 12 h subsequent. The remaining 45 ICs were obtained by adding nine further perturbations each to ICs 1–5, chosen randomly from a larger set of perturbations consisting of 24-, 48-, and 72-h differences of 0600, 1200, and 1800 UTC reanalysis states. In each case, the two reanalysis states used to form the perturbation bracket the original IC in time.

The observed SST and sea ice fields used as boundary conditions for the GCM simulations are taken from Reynolds (see Reynolds and Smith 1994); these fields vary weekly, so both intraseasonal and interannual variations in boundary forcing are used.

As with the reanalysis data, for each winter we access the 110-day records of 200-hPa height and zonal wind (once daily, at 0000 UTC), starting on 12 December. We interpolate these fields, originally on a T63 Gaussian (approximately $1.9° \times 1.9°$ latitude–longitude) grid, to the common T42 Gaussian grid. The dataset consisting of all 55 simulations for each of the 18 winters is referred to as the "Grand Ensemble," GE18, or GE in this paper.

c. Time-scale filtering

A "seasonal cycle" was obtained by retaining the lowest three components of a Legendre expansion in time of the 110-day time series at each grid point for each winter simulation (Straus 1983), followed by averaging the parabolic series over all ensemble simulations for all years. This mean seasonal cycle was removed from each individual simulation. For the observational datasets (NCEP54 and NCEP18), the seasonal cycle was obtained by averaging the parabolic time series at each grid point over all available years (54 or 18) and removing this seasonal cycle from the data for all years. Following many other studies on regimes, we filter out baroclinic, synoptic-scale fluctuations with the use of temporal filtering. The filtering was applied to the 110-day time series for each winter (after removal of the seasonal cycle) by the application of a 15-point digital Lanczos filer (Duchon 1979) so that fluctuations with periods greater (less) than about 10 days were retained (filtered out). After filtering, 96 days are available for further analysis. Thus, the time series have only

TABLE 1. Accumulated space–time variance (in %) explained by the leading 6 EOFs and the leading 10 EOFs: N54 refers to the EOFs of the NCEP54 record, N18 to the NCEP18 record, and GE for the Grand Ensemble in the domain 20°–80°N, 150°E–30°W. Here Z refers to 200-hPa height and U, the 200-hPa zonal wind.

	N54 Z	N54 U	N18 Z	N18 U	GE Z
EOF 1–6	68.9	62.6	69.5	63.0	76.1
EOF 1–10	83.8	76.7	84.3	77.6	88.9

the mean seasonal cycle and components with periods less than about 10 days filtered out. We then use only every fourth point of this filtered dataset, giving 24 points in each seasonal time series (see section 2g).

d. Reduction of dimensionality

We use an empirical orthogonal function (EOF) analysis in order to reduce the dimensionality of the datasets. The EOFs (spatial patterns) of 200-hPa height and zonal wind were computed from the T42 grids as the normalized eigenvectors of the covariance matrix[2] calculated over the Pacific–North American region 20°–80°N, 150°E–30°W for the NCEP54, NCEP18, and GE18 datasets. The dimensional principle components (PCs), which are the time series of the EOFs, form the coordinates of the reanalysis or model state in a state space. The total space–time variance explained jointly by EOF 1–6 and EOF 1–10 are given in Table 1 for the various datasets. For geopotential height (u wind), the leading 6 EOFs explain at least ~69% (~63%) of the regional variance, and the leading 10 EOFs explain over ~80% (~77%) of the variance.

e. GCM one-member "EXP18" samples

From the GE18 200-hPa height PCs, we extracted 55 GCM one-member samples. A single sample consists of PCs from one winter simulation for each of the 18 winters. Each sample contains simulations with the same IC label. Since even the simulations with the same IC label are intialized independently for each calendar year, they may be considered to be independent. Note also that each sample of integrations is influenced by the identical record of winter SST. We subsequently refer to these samples as the EXP18 dataset.

---

[2] To assure appropriate area weighting, the fields are multiplied by the square root of the cosine of latitude prior to entering the covariance matrix calculation. The final centroid maps presented here, which are synthesized from the EOFs, have this square root weighting divided out before plotting.

f. GCM three-member "EXP54" samples

We also make use of 55 GCM three-member samples, each consisting of PCs from *three* winter simulations for each of the 18 winters. For each winter, the PCs belonging to the three-member sample *n* are obtained by taking the PCs of the simulation with IC*n* from the one-member sample EXP18, and augmenting it by the PCs from two randomly chosen simulations from among the other 54 members for the same winter. Here *n* runs from 1 to 55. We subsequently refer to this as the EXP54 dataset. It should be noted that the EXP54 samples include data from the same number (54) of winters as NCEP54, but associated with the same range of SST variability as in NCEP18 and EXP18. Therefore, any differences in statistics derived from the EXP18 and EXP54 datasets should only be ascribed to internal atmospheric variability.

g. Synthetic Markov datasets

In assessing whether the null hypothesis of multinormality can be rejected, it is necessary to perform Monte Carlo simulations using a large number of synthetic datasets, which are described here. (The actual statistical testing is described in sections 3 and 4). A large pool (from 100 to 500) of synthetic datasets is created separately to be comparable to each of the NCEP54, NCEP18, and GE18 datasets. In each case, each synthetic dataset has precisely the same length as the observed or simulated dataset against which it is compared, and is generated from a series of *n* Markov processes, whose variance and first-order autocorrelation are obtained from the first *n* PCs of the appropriate observed or simulated dataset. A requirement for using the surrogate synthetic datasets for hypothesis testing is that the low-frequency PC time series (sampled every 4 days) is well approximated by a Markov process.

This is the case if the residuals of the Markov process are close to white noise (see Straus and Halem 1981). In a very similar context to that of this paper, Penland and Ghil (1993) made a similar choice of sampling interval in Markov modeling of temporally filtered height PCs.

3. Regimes from NCEP–NCAR reanalysis

a. Cluster analysis of quasi-stationary states

Following SM04, the *k*-means method (as described by MVL) was used. Recent studies using observationally based datasets have found that this method gives results consistent with those of hierarchical and mixture model methods (Alhamed et al. 2002; Cassou et al.

2004; Robertson and Mechoso 2003). Using the 6- or 10-dimensional state space spanned by the EOFs (as described in section 2), a partition into *k* clusters, or categories, is sought. This partition is accomplished in such a way that the ratio of the variance between cluster *centroids* (a centroid being the average over all states in a cluster) to the average intracluster variance is maximized. This is equivalent to minimizing the intracluster variance since the total variance of the dataset is fixed. The variances are computed using an Eulerian distance metric in the state space of the PCs.

Before the cluster analysis is carried out, the temporally filtered time series (see section 2c) are screened to focus on quasi-stationary periods, a procedure recommended by Toth (1992), Kimoto and Ghil (1993a), and Itoh and Kimoto (1999) in order to ensure that periods of persistent synoptic weather are considered. This quasi-stationary selection was accomplished by first computing a histogram of the distance in state space between successive points to give a measure of the difference between low-frequency states four days apart. (Fifty bins were used to compute the histogram, based on the computed upper limit of the distribution.) For $m = 1, \ldots, 50$, a cumulative distribution is computed as the fraction of states that fall into the lowest m bins; once that fraction exceeds α, the states in the remaining bins ($m + 1$ and higher) are excluded from the cluster calculation. Since the choice of α is arbitrary, we will explore the sensitivity of some of our results to the value of α, which ranges between 0.40 (strong filtering) and 1.00 (no filtering).

Methods used to verify the significance and stability of the clusters include significance vis-à-vis a null hypothesis of a multinormal background distribution, reproducibility using randomly chosen half-length datasets, and consistency between clusters obtained from different variables (see MVL and SM04). Since the null hypothesis of a multinormal distribution may be rejected not only because of multimodality but also because of skewness (Stephenson et al. 2004; Christiansen 2007, hereafter C07), we have carried out statistical analyses (similar to those of C07) that indicate the modest degree of skewness of the PCs is not sufficient to explain the high significance levels that we obtain. See section 4 and the appendix for details.

Significance vis-à-vis a multinormal distribution is computed here as in SM04, using a Monte Carlo procedure in which a large number of synthetic datasets are created, each the same length as the dataset analyzed. The synthetic datasets are described in detail in section 2g. The cluster analysis is repeated (and the variance ratio calculated) for each set of synthetic data,
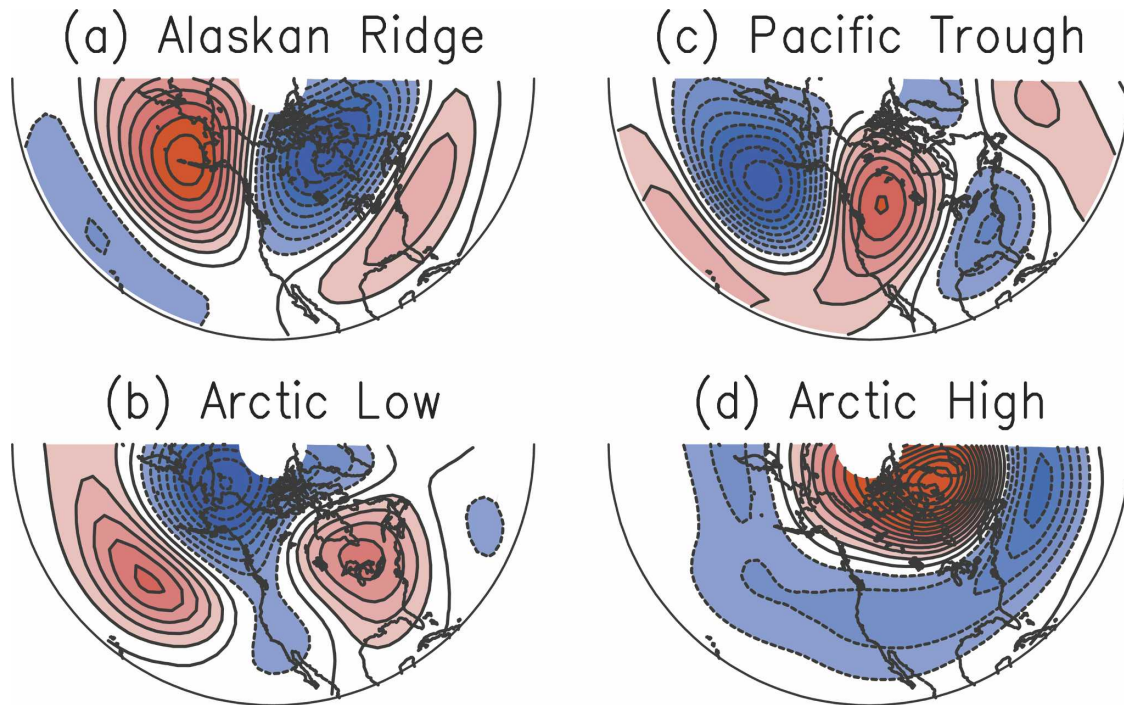
FIG. 1. Cluster centroid maps of 200-hPa height for $k=4$ from the NCEP54 record. The cluster analysis was carried out using six EOFs and a value of $\alpha=0.5$ (see text for explanation). Contour interval is 20 m with the 0 contour omitted.

and the significance level p is given as the fraction of cases in which the variance ratio of the Monte Carlo data fall below that computed from the real data.

For those cases in which quasi-stationary filtering was invoked ($\alpha < 1.0$), the Markov process variance and lag-1 autocorrelation were obtained from the unfiltered PCs, and the quasi-stationary filtering was applied prior to the cluster analysis of each synthetic dataset to ensure consistency.

In the context of determining the dependence of the cluster properties on external forcing (in this case SST), the reproducibility tests are carried out by comparing randomly chosen half-length subsets within each simulation in order to avoid mixing different sets of SST forcings. To have a measure of reproducibility expected in the context of the multinormal null hypothesis, we have also carried out equivalent reproducibility tests using a number of synthetic datasets (as above). By comparing the clusters of the 200-hPa u-wind field to the cluster centroid maps of the geostrophic u wind calculated from the height clusters, we have tried to assess the consistency of the clusters across variables. Since geostrophic balance is expected to work well in midlatitudes, this comparison is equivalent to a sensitivity test of the cluster analysis to the choice of the

distance metric in state space. Results of these tests are discussed in section 4.

b. Regimes in the observed 54-winter record

One of the most difficult problems in cluster analysis is to uniquely define the number (k) of clusters (as discussed in C07). In the next section and the appendix we present evidence that k should be *at least 3* based on the strength of the clustering given by the variance ratio, but also that k should be *less than 5* based on reproducibility results. Although we focus on four clusters, we give some results for $k = 3$ in section 7 to indicate the sensitivity of our results to the cluster choice.

Figure 1 shows the height maps of the four centroids from the NCEP54 record, computed using six EOFs and $\alpha = 0.50$. We refer to cluster 1 (Fig. 1a) as the Alaskan ridge (AR) pattern, in agreement with the classification of SM04 and Renwick and Wallace (1996). It is similar to the P6 Pacific cluster reported by Kimoto and Ghil (1993b, hereafter KG), to the Pacific cluster reported in Fig. 11b of Smyth et al. (1999, hereafter SIG), to the A global cluster of Cheng and Wallace (1993, hereafter CW) in their Fig. 4, and to the A Pacific cluster shown in Fig. 8 of CW. This pattern has
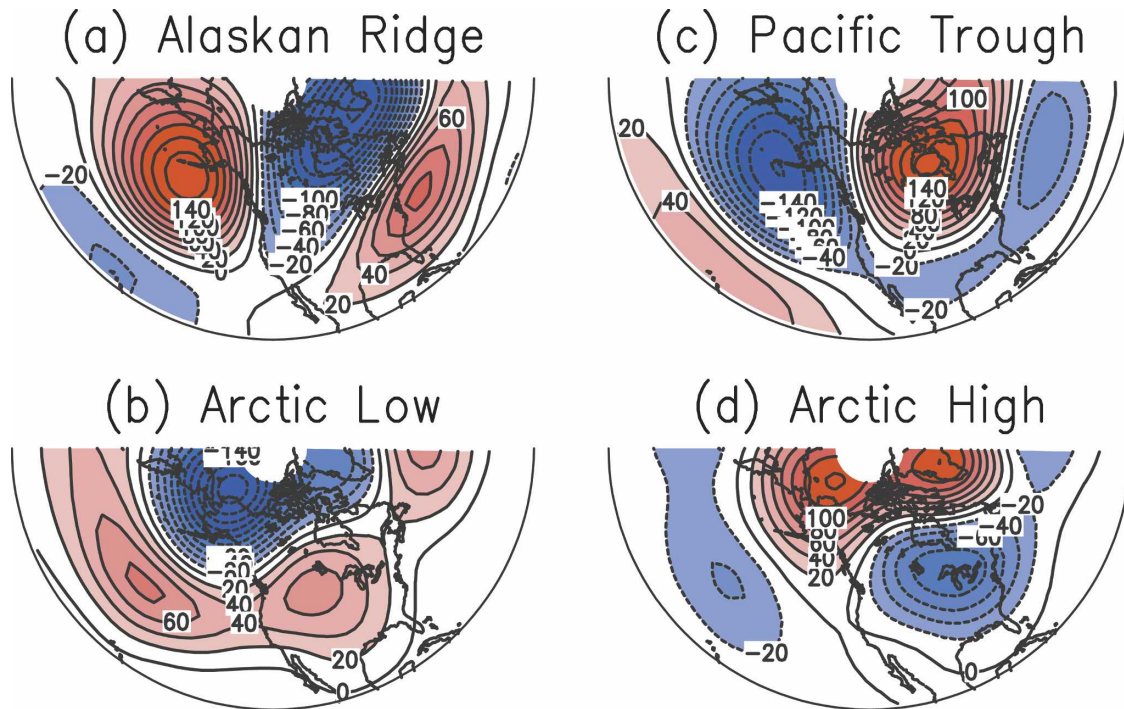
FIG. 2. As in Fig. 1 but from the NCEP18 record.

the appearance of a wave train originating in the subtropical Pacific, and is associated with Alaskan blocking (Renwick and Wallace 1996). Cluster 3 (Fig. 1c) consists of a wave train nearly out of phase with the AR but with the centers shifted, particularly the ones over the continent. This cluster, referred to as the Pacific trough (PT), resembles the P1 Pacific pattern of KG, the Pacific cluster of SIG shown in their Fig. 11a, and the R Pacific cluster of CW (their Fig. 8). This regime pattern is referred to as PNA-like, after the well-known PNA pattern of Wallace and Gutzler (1981).

Cluster 4 (Fig. 1d) is termed the Arctic high (AH), and is strongly suggestive of the North Atlantic Oscillation (NAO) and/or the Northern Hemisphere annular mode. It strongly resembles the global cluster G of SIG (their Fig. 9c), and the global cluster G of CW (their Fig. 4). Finally, cluster 2 (Fig. 1b), termed the Arctic low (AL), is similar to the Pacific P4 cluster of KG (their Fig. 5).

c. NCEP–NCAR 18-winter record

We chose to analyze the 18-winter record so that the tropical SST forcing would be the same as that of the GCM simulations to which we compare. However, with only 18 winters (and hence a rather small sample size of $18 \times 24 = 432$) it is not possible to demonstrate that the clusters have high formal significance.

The cluster centroids are shown in Fig. 2 for the same parameters used in Fig. 1. One can easily establish a visual one-to-one correspondence of these patterns with those shown in Fig. 1 for the NCEP54 dataset with, for example, the AR regime having a very strong high over the Gulf of Alaska in both cases and the PT regime a strong low. The associated pattern correlations of the AR, AL, PT, and AH clusters are 0.96, 0.91, 0.62, and 0.66, respectively. (The corresponding pattern correlations for the cluster analysis with no quasi-stationary filtering are 0.97, 0.60, 0.67, and 0.67, respectively. Thus only the AL changes dramatically using the pattern correlation as a measure of correspondence. Another measure of correspondence of two cluster centroids will be introduced later in the paper.)

The modifications in the PT pattern between the two periods are worth noting: the continental high has been shifted from northwestern North America in NCEP54 to Hudson Bay in NCEP18. The degree of mixing between the PT and AH also differs in the two periods. The latter period PT pattern is quite similar to the seasonal mean response forced by tropical Pacific SST in more recent times (Barnston and Livezey 1987; Straus and Shukla 2002, and many studies quoted therein), whereas the configuration of the PT pattern in NCEP54 is closer to the classical PNA pattern of Wallace and Gutzler (1981). This suggests that changes in the cluster

maps may be associated with different tropical Pacific SST patterns. This hypothesis will be tested in section 6.

d. Evolution of cluster frequency

To assess the degree to which we can associate certain regimes with SST forcing, it is helpful to examine the winter-by-winter time series of the residence frequency of the quasi-stationary low-frequency atmosphere flow in each of the four NCEP54 clusters in Fig. 1. Such a series is shown for each of the clusters in Figs. 3a–d along with scatterplots of the standardized frequency against the standardized (seasonally averaged January–March) Niño-3.4 SST index[3] shown in Figs. 3e–h. The frequency for a given cluster for a given winter gives the fraction of time the circulation resides in that cluster. For the $\alpha = 0.50$ results, the frequencies add up to approximately 0.50 considering the whole dataset, but this is not true for each year since there could be a different proportion of quasi-stationary states in any given year.

While the Alaskan ridge and Arctic low frequencies do not display any obvious trend, the Pacific trough frequency clearly increases in the last half of the record while the Arctic high frequency decreases. The linear trend of the PT is 0.15 $(54 \text{ yr})^{-1}$, that of the AH is -0.16 $(54 \text{ yr})^{-1}$. Both trends are significant at the 98% level with respect to trends seen in 100 equivalent synthetic datasets. (The AH negative trend is very stable with respect to changes in the quasi-stationary filtering parameter α, while the PT trend increases gradually from 0.15 to 0.29 as α is increased from 0.5 to 1.0, i.e., as the filtering is removed.) The PT frequency also seems to undergo a qualitative change in the late 1970s; a link to the regime shift in the stratosphere found by Christiansen (2003) is suggested.

Both the AR and PT are correlated with Niño-3.4, (-0.4 and 0.4, respectively)[4] based on a 54-yr time series, supporting the notion that the AR is more active in cold (La Niña) events (as in SM04 and Renwick and Wallace 1996) and also the association of the PT with warm events. (One can clearly pick out the strong warm El Niño winters of 1982/83, 1986/87, and 1997/98 as having a higher frequency of occurrence of the PT regime.)

Table 2 shows the cluster frequencies averaged over the three 18-yr periods within the 54 yr, as well as the 54-yr average, and the frequencies for the NCEP18 regimes. The most common regime in both the NCEP54 and NCEP18 calculations (the PT) becomes more common from the first to the third period, while the least frequented regime (AH) becomes less common. The AR and AL do not show obvious trends.

It is interesting to note that the positive trend in the North Atlantic Oscillation index as shown for example by Hurrell (1995) is expressed here as a suppression of the Arctic high.

4. Significance, consistency, and reproducibility of NCEP54 regimes

The significance of the NCEP54 height clusters vis-à-vis a suitably defined multinormal synthetic dataset (as in section 2g) is presented in Table 3. Results are given for both 6-EOF and 10-EOF truncations, and for a wide range of α. (The smaller the value of α, the greater the degree of prefiltering.) Significance values greater than 90% are shown in bold. Also shown is the variance ratio (ratio of variance among centroid clusters to average intracluster variance) for various values of α for the six-EOF truncation.

Following the approach of C07, we apply the cluster analysis to a synthetic dataset consisting of independent processes, each one having the same lag-1 autocorrelation and skewness as the corresponding PC of NCEP54. The results show that skewness alone is not enough to generate the high levels of significance in the clustering procedure[5] reported in Table 3. For details, consult the appendix.

---

[3] The Niño-3.4 SST index is defined as the average SST over the region 5°S–5°N, 190°–240°.

[4] The lag-1 year autocorrelation of the Niño-3.4 time series is quite small (-0.031). Thus, any correlation whose absolute value is greater than 0.22 is statistically significant at the 95% level.

[5] C07 show that for skewness ≥ 0.63 in a unimodal one-dimensional time series, the k-means method will indicate a preferred choice of more than cluster. The magnitude of the skewness of the leading four PCs for NCEP54 is 0.24, 0.23, 0.15, and 0.27, respectively.

$\rightarrow$

Fig. 3. (a)–(d) Fraction of occurrence for each of four NCEP54 clusters (six EOFs, $\alpha = 0.5$) as a function of winter. Significant trends given in brackets, in units of fraction $(54 \text{ yr})^{-1}$. (e)–(h) Scatterplots of standardized regime frequency vs the (standardized) Niño-3.4 index for the four NCEP54 clusters with the correlation indicated. Each scatterplot contains 54 points: one per winter.
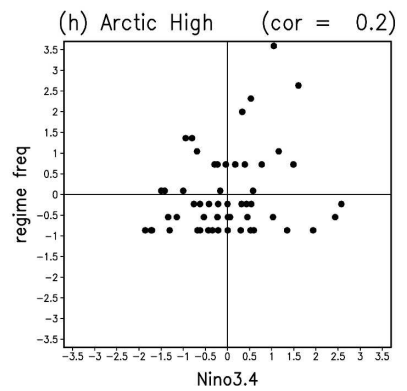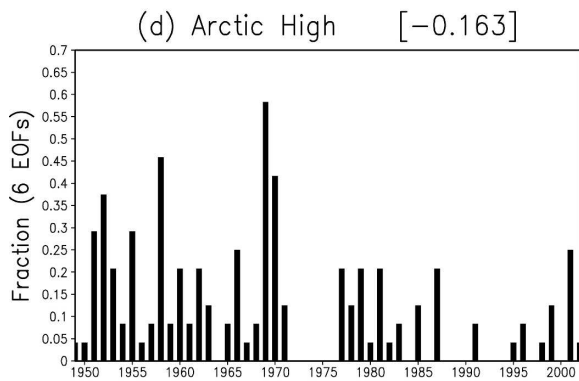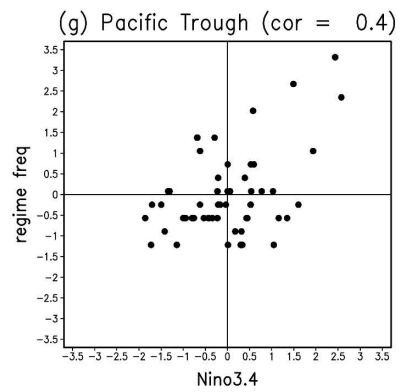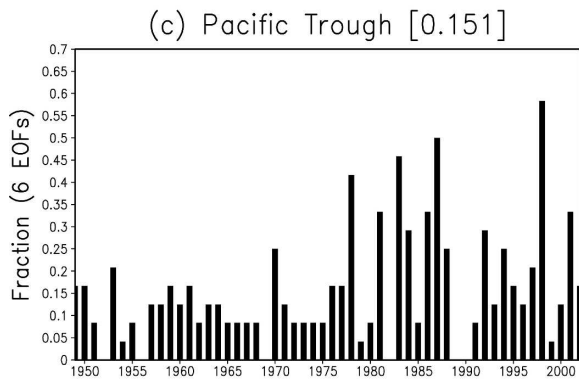
(a) Alaskan Ridge   NCEP 54 winters

(b) Arctic Low   NCEP 54 winters

(c) Pacific Trough [0.151]

(d) Arctic High   [−0.163]

(e) Alaskan Ridge (cor = −0.4)

(f) Arctic Low   (cor = −0.1)

(g) Pacific Trough (cor = 0.4)

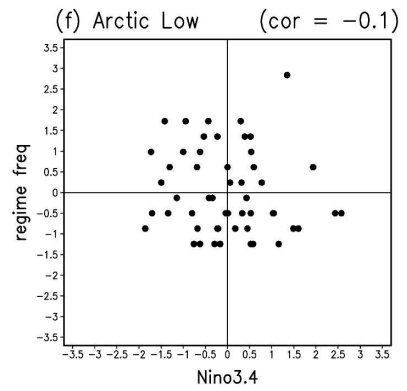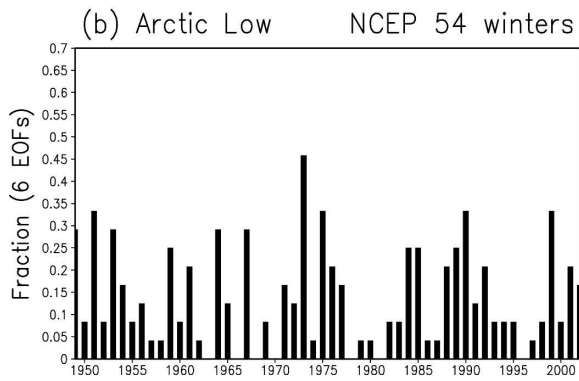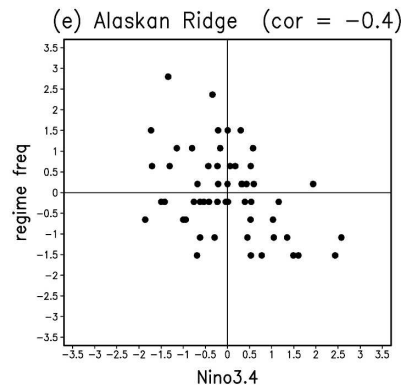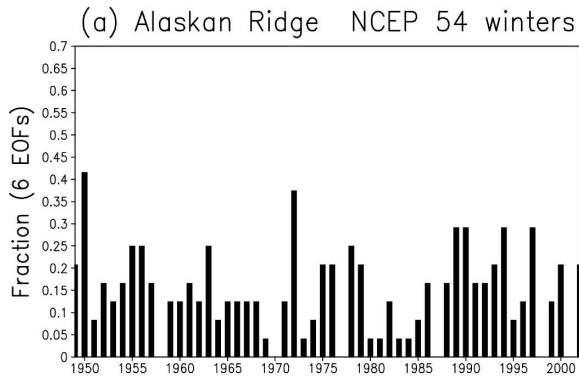(h) Arctic High   (cor = 0.2)

TABLE 2. Relative fraction of occurrence of four regimes in the NCEP54 dataset, the three consecutive 18-yr periods of the NCEP54 dataset [labeled NCEP(1), NCEP(2), and NCEP(3)], the NCEP18 dataset, and the GE (GrEn18) dataset. In each case the relative fractions sum to approximately 0.5 since only quasi-stationary states are included in the analysis. The regimes were calculated using six EOFs with $\alpha = 0.5$ (see text for explanation). The last column (std dev) gives the standard deviation of the frequency of clusters of GE18 among the 55 members of the ensemble.

Regime	NCEP54	NCEP(l)	NCEP(2)	NCEP(3)	NCEP18	GrEn18	Std dev
AR	0.147	0.164	0.116	0.160	0.104	0.083	0.023
AL	0.140	0.141	0.132	0.146	0.102	0.132	0.024
PT	0.157	0.109	0.157	0.204	0.180	0.184	0.035
AH	0.113	0.164	0.120	0.056	0.114	0.132	0.020

For six EOFs, the significance values of Table 3 show relatively little variation for $k = 3$–6 and $\alpha \leq 0.8$ (with values generally above 87%). The significance is clearly lower for $k = 2$ (<73%), and is somewhat lower for the unfiltered data ($\alpha = 1.0$). The behavior of the significance is less systematic for 10 EOFs, indicating that the NCEP54 dataset is too short to adequately sample the larger state space.

TABLE 3. Significance p (in %) and variance ratio for the height clusters of the NCEP54 dataset. Significance is assessed against synthetic Markov model datasets constructed using same variance and lag-1 correlation for the indicated number of PCs, as a function of cluster number k. Here p is given as the percentage of synthetic datasets having a variance ratio less than the NCEP54 dataset. (Values greater than 90% are bold.) The variance ratio is the ratio of the variance between cluster centroids to the mean, weighted intracluster variance. See text for details.

			$\alpha =$		
k	1.0	0.8	0.6	0.5	0.4
		Significance for 6 EOFs			
2	62	73	66	58	72
3	81	**90**	87	88	**96**
4	74	87	89	**92**	**93**
5	67	83	**92**	**90**	**94**
6	81	**90**	**95**	**95**	**96**
		Significance for 10 EOFs			
2	73	73	71	85	61
3	83	84	84	**90**	52
4	77	76	83	89	88
5	74	81	76	87	74
6	85	86	86	**93**	88
		Variance ratio for 6 EOFs			
2	0.264	0.274	0.267	0.272	0.277
3	0.465	0.480	0.481	0.491	0.496
4	0.608	0.624	0.638	0.653	0.661
5	0.733	0.754	0.782	0.792	0.810
6	0.861	0.874	0.904	0.920	0.938
		Variance ratio for 10 EOFs			
2	0.212	0.215	0.214	0.222	0.210
3	0.354	0.365	0.365	0.375	0.351
4	0.453	0.462	0.471	0.468	0.484
5	0.537	0.552	0.560	0.578	0.570
6	0.620	0.633	0.644	0.663	0.664

The reproducibility is calculated by dividing the low-frequency PC time series of length 24 for each winter into two randomly chosen time series of length 12 and carrying out the cluster analysis for the two 54-winter datasets, each of length $12 \times 54 = 648$. (This procedure insures that all randomly chosen samples feel approximate the same SST forcing.) This procedure is repeated for 100 trials, and for each trial the two sets of clusters obtained (for each value of k) are matched. The matching is done in two ways: by maximizing the mean pattern correlation over every possible pairing of the k centroids between the two half-length samples and by minimizing the mean squared error. (See the appendix for more details. All averaging of correlations is carried out using the Fisher Z transform.)

The pairing giving the maximum average correlation, or minimum mean squared error (over k centroids), is then archived for each trial, yielding a distribution for each measure. For the pattern correlation, the average of these distributions is 0.95, 0.85, and 0.78 for $k = 3$, $k = 4$, and $k = 5$, respectively, while the median correlation is 0.96, 0.86, and 0.65. The centroid patterns for the average of all 200 half-length samples (matched on the basis of pattern correlation) are consistent with the patterns shown in Fig. 1. (For $k = 4$, the pattern correlations for the AR, AL, PT, and AH patterns are 0.98, 0.69, 0.87, and 0.98, respectively.)

The average of the mean squared error distribution of half-length samples was compared with that computed using a large number (100) realizations of synthetic Markov processes (see section 2g) for the NCEP54 dataset. The ratio of the NCEP54 error to that obtained from the synthetic data is shown in Table 4 for values of k from 3 to 5. Here we report the reproducibility of each cluster individually. It is clear that for $k > 4$ all cluster centroids are not on average as reproducible as the multinormal data; hence, $k = 5$ and larger can be rejected. It is noteworthy that for $k = 4$ the Arctic low is not reproducible.

Since considerations of the variance ratio suggest that k should be at least 3, but reproducibility consid-

TABLE 4. Ratio of the mean-squared error of best matched cluster centroids from 100 pairs of randomly chosen half-length datasets derived from the full NCEP54 dataset to that computed from multinormal Markov synthetic data. The calculation of the mean-squared errors from the latter dataset was repeated 100 times and the grand average taken. Regimes PT, AH, AR, and AL as in Fig. 1.

k	PT	AH	AR	AL	
3	0.50	0.60	0.41		
4	0.84	0.84	0.92	2.06	
5	1.33	1.29	1.72	1.96	2.28

TABLE 5. Average pattern correlation between height clusters computed using six EOFs and $\alpha = 0.5$ and those using three other configurations of the number of EOFs and the quasi-stationary filtering parameter α (see text for details). The value of α is given in parentheses.

k	6 EOFs (1.0)	10 EOFs (1.0)	10 EOFs (0.5)
2	0.99	0.99	0.99
3	0.99	0.99	0.99
4	0.93	0.93	0.96
5	0.84	0.84	0.79

erations suggest that $k < 5$, we have chosen four clusters for many of the results in this paper, with some results repeated for $k = 3$.

We show that $k = 4$ is a reasonable choice in two further ways, based on the consistency of centroids with respect to variable (height versus zonal wind) and with respect to domain (regional versus hemispheric).

The consistency with respect to variable is assessed by comparing the maps of the geostrophic zonal wind obtained from the NCEP54 height cluster centroids with the maps of the centroids computed from the cluster analysis of the zonal wind. For six PCs and $\alpha = 0.50$ these patterns (not shown) match quite well, with pattern correlations of 0.93, 0.84, 0.72, and 0.82 for the AR, AL, PT, and AH, respectively. (Note that, since the filtering is done independently for height and zonal wind clusters, the set of states going into the zonal wind and height cluster analyses may be somewhat different.)

We have compared the centroids shown in Fig. 1 to those computed using a full-hemispheric domain (north of 20°N) in order to provide a test of the sensitivity to the choice of lateral boundary. The two sets of patterns (not shown) are nearly identical over the hemispheric region 150°–30°W. The clusters that we have called the AH and AL have elements of both the Northern Hemisphere annular mode (Thompson and Wallace 2000) and the NAO (Hurrell 1995).

As in SM04, consistency with respect to variable and region also hold for values of k other than 4; however, it is important to note that they do hold for $k = 4$.

Finally, we show that the patterns given in Fig. 1 for NCEP54 are consistent with the cluster centroids obtained using other combinations of PCs retained and quasi-stationary filtering parameter α. Table 5 reports the average pattern correlation between the clusters of the standard NCEP54 calculation (six PCs, $\alpha = 0.50$) and a number of other configurations. For each pair of configurations and for each k all possible pairwise matchings are considered, and the matching with the maximum average correlation is chosen.

5. Clusters of the Grand Ensemble

The cluster centroids of geopotential height for $k = 4$ are shown for the "Grand Ensemble" with the standard parameters (six EOFs, $\alpha = 0.50$) in Fig. 4. Compared to Fig. 2, we are able to easily recognize the counterpart of the NCEP18 Alaskan ridge, Arctic low, and the Pacific trough from the simulations (Figs. 6a–c). (The corresponding pattern correlations are 0.74, 0.76, and 0.84.) The AH cluster in nature is replaced in the GCM results with a distinctive zonally oriented wave train pattern (Fig. 6d).

Despite the overall pattern agreement for the AR, AL, and PT regimes, there are some differences. To assess these differences, we compute the pattern correlation for each regime from the cluster analyses based on each of the 55 members of the EXP18 samples, and the cluster analysis based on the entire GE18. From these values we construct a probability distribution function, which is shown for each regime in Fig. 5. We match the regime centroids in each of the EXP18 sample clusters with those in the GE18 as before, by considering all possible pairings of k centroids and retaining the pairing with the largest average correlation. The width of the PDFs for the AL and AH patterns indicates that these clusters, although significant within the entire GE18, are less reproducible across samples than the AR and PT. In fact, if only two clusters are retained in the GE18 analysis ($k = 2$), the centroids obtained strongly resemble the AR and PT (not shown).

For the AR and PT regimes, the PDF of pattern correlation has most of its weight at values higher than the pattern correlation between the NCEP18 and GE18 centroid (shown as a vertical line in Fig. 5), indicating that the subtle differences between simulated and observed patterns are in fact significant vis-à-vis expected sampling variations. The AL correlation of 0.76, however, lies within the (wide) range indicated by the sampling PDF, so we can say that this simulated regime pattern has statistically significant agreement with the reanalyses.
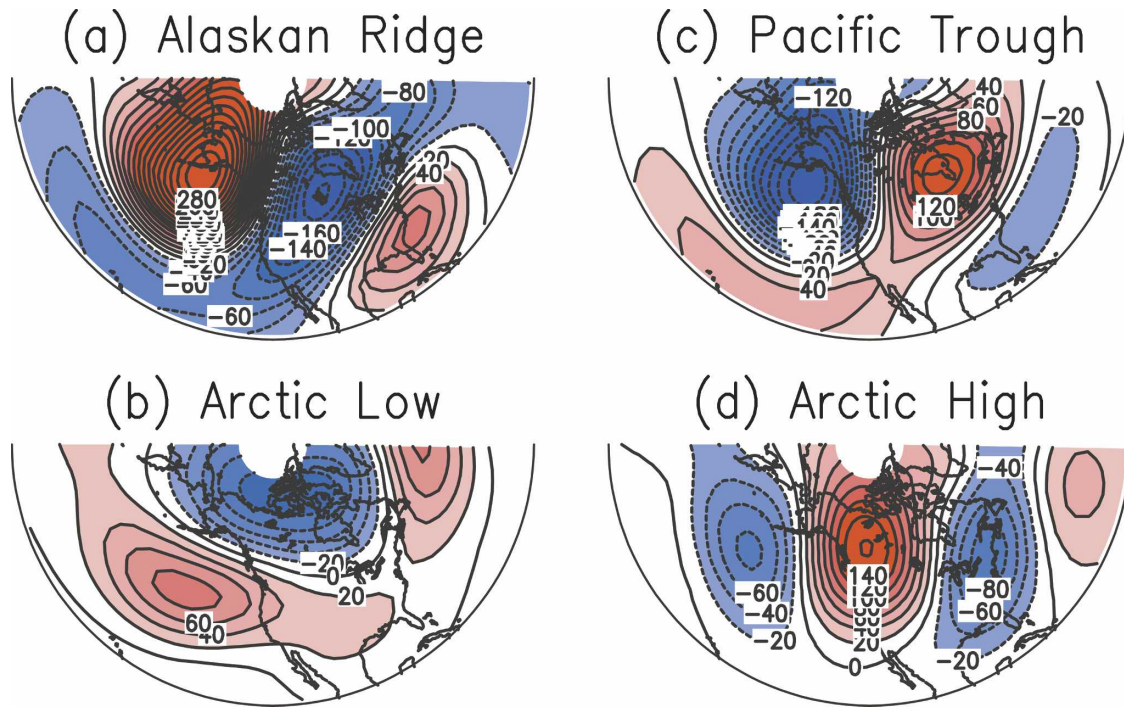
FIG. 4. As in Fig. 1 but from the Grand Ensemble dataset.

The formal significance tests of the GE cluster analysis indicates 100% significance for every value of k tested, from 2 to 6. This is in part due to the size of the dataset (24 independent states for each of 55 ensemble members for each of 18 winters yields 23 760 states): even a modest degree of clustering is very rarely reproduced with a multinormal synthetic dataset of this size.

The frequency of occurrence of the AR, PT, and AL for the GE are compared to those observed (NCEP18) for the same winters in Table 2. While the PT is visited as frequently as observed, the AR is somewhat less common, while the AL is somewhat more common. The GCM underestimate of the AR frequency is consistent with the general underprediction of blocking in most models (see Watson and Colucci 2002 and references therein).

We evaluate whether the differences in the cluster occupation for the AL and AR in the simulations compared to the reanalyses could be due to sampling error by again utilizing the EXP18 samples. The PDFs in Fig. 6 indicate the distribution of frequencies for the 55 cluster calculations carried out on the EXP18 dataset. In contrast to the differences in pattern evaluated in Fig. 5, here we see that the modest differences in cluster population between the GE and NCEP18 (both indicated by vertical lines) are well within the spread of the samples. Thus, we can conclude that while the simulated circulation regime patterns show some significant differ-

ences from our reanalysis estimates, the cluster frequencies do not.

6. Cluster frequencies and SST forcing

As mentiond in the introduction, the analyses presented in this paper for simulated and reanalysis datasets do not remove the seasonal cycles of individual winters prior to computing the clusters, but only the climatological seasonal cycle. This means that SST-forced interannual variability may play a role in determining the regimes. The purpose of this section is to explore that role. Specifically, we ask three questions: Are the differences in regime patterns seen from the NCEP54 and NCEP18 analyses presented in section 3 only due to sampling error, or does the change in SST forcing play a role? Are the decadal shifts in cluster occupancy seen within the NCEP54 analysis (shown in Table 2) larger than the estimated sampling error? Finally, is the interannual variability of cluster frequencies within the GE consistent with that of NCEP18?

Again, we use the sampling properties of the large set of EXP18 cluster analyses to try to answer these questions. Figure 7 presents the PDF of pattern correlation (for each regime) between the clusters in each of the EXP18 sample members with the matching cluster in the corresponding EXP54 dataset. (Remember, each of the 54-yr EXP54 datasets was obtained by starting with
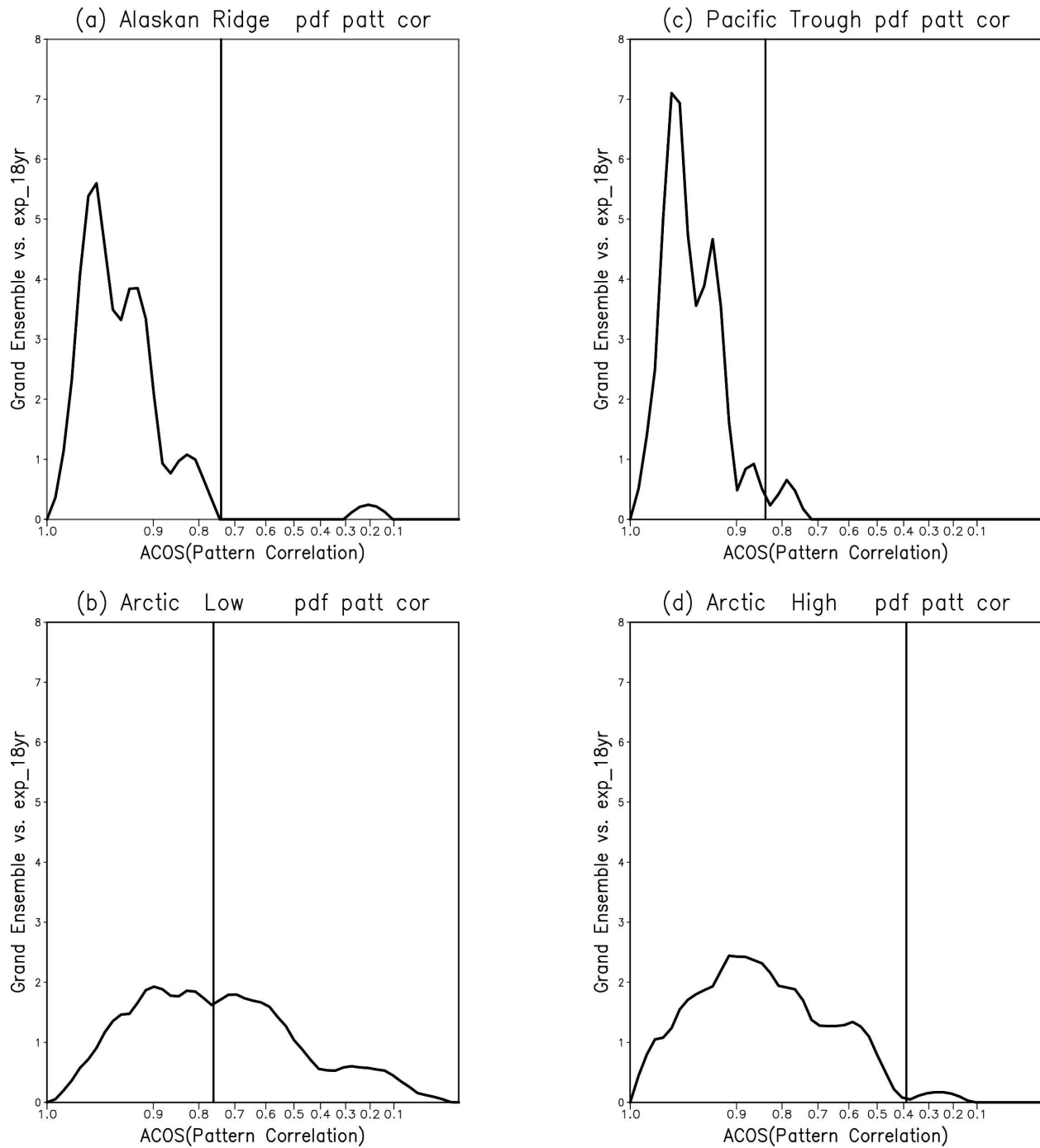
FIG. 5. PDFs of the pattern correlation of the GE18 height centroids and each of the 55 corresponding height centroids from the EXP18 dataset. In all cases six EOFs and a value of $\alpha = 0.5$ were used. The vertical lines show the pattern correlations between the GE18 height centroids and the corresponding NCEP18 height cluster centroid. The abscissa of each plot is the arccosine of pattern correlation, with the pattern correlation values indicated.

one EXP18 sample and augmenting it with two other randomly chosen 18-yr time series from other initial conditions.) The clusters obtained from each EXP18 and EXP54 sample were identified by matching with

the full GE cluster centroids using the pattern correlation obtained as before. The PDFs in Fig. 7 thus give an estimate of the difference in regime patterns that might be expected from 18-yr versus 54-yr time series based
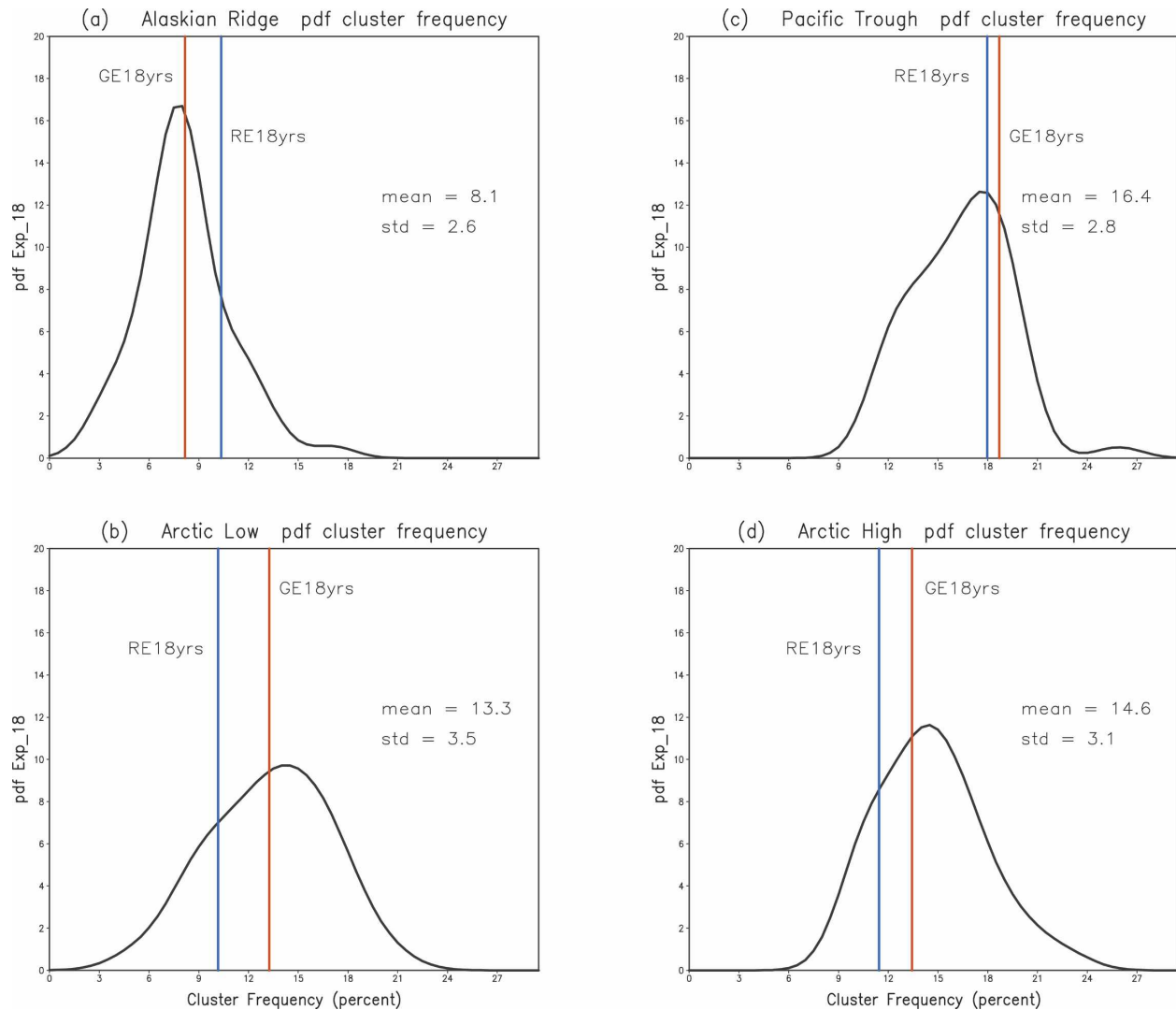
FIG. 6. PDFs (for each of the $k = 4$ clusters) of the cluster frequency (in %) computed from the clusters of the EXP18 dataset. In all cases six EOFs and a value of $\alpha = 0.5$ were used. Red (blue) vertical lines show the corresponding cluster frequency from the GE (NCEP18) datasets.

on sampling error alone. The vertical lines indicate the actual pattern correlations between the NCEP54 and NCEP18 cluster patterns. While the AR and AL pattern differences certainly lie within the range of sampling error (as might be expected since these differences are small), the PT shows significant differences, while the significance of the change in AH is marginal. The PT and AH showed a distinct trend in frequency of occurrence during the 54 yr (shown in Fig. 3).

The change in PT from a pattern showing some resemblance to the wave-train-like PNA of Wallace and Gutzler (1981) in the NCEP54 results to the Tropical Northern Hemisphere (TNH) pattern of Barnston and Livezey (1987) in the more recent years is certainly consistent with the work of Straus and Shukla (2002),

who showed that the seasonal mean response to ENSO during the more recent decades closely resembles the NCEP18 PT regime.

The year-by-year variation in cluster frequency from the GE18 and NCEP18 cluster calculations are shown in Fig. 8 as the red and blue lines, respectively) along with a band (shown in light gray) representing the one standard deviation uncertainty in the frequency within each given winter due to sampling error. (This was estimated from the GE18 directly by evaluating the cluster frequency for each of the 55 ensemble members for each year, based on the cluster classification for the entire GE dataset.)

There are individual years in which statistically significant discrepancies can be seen between NCEP18
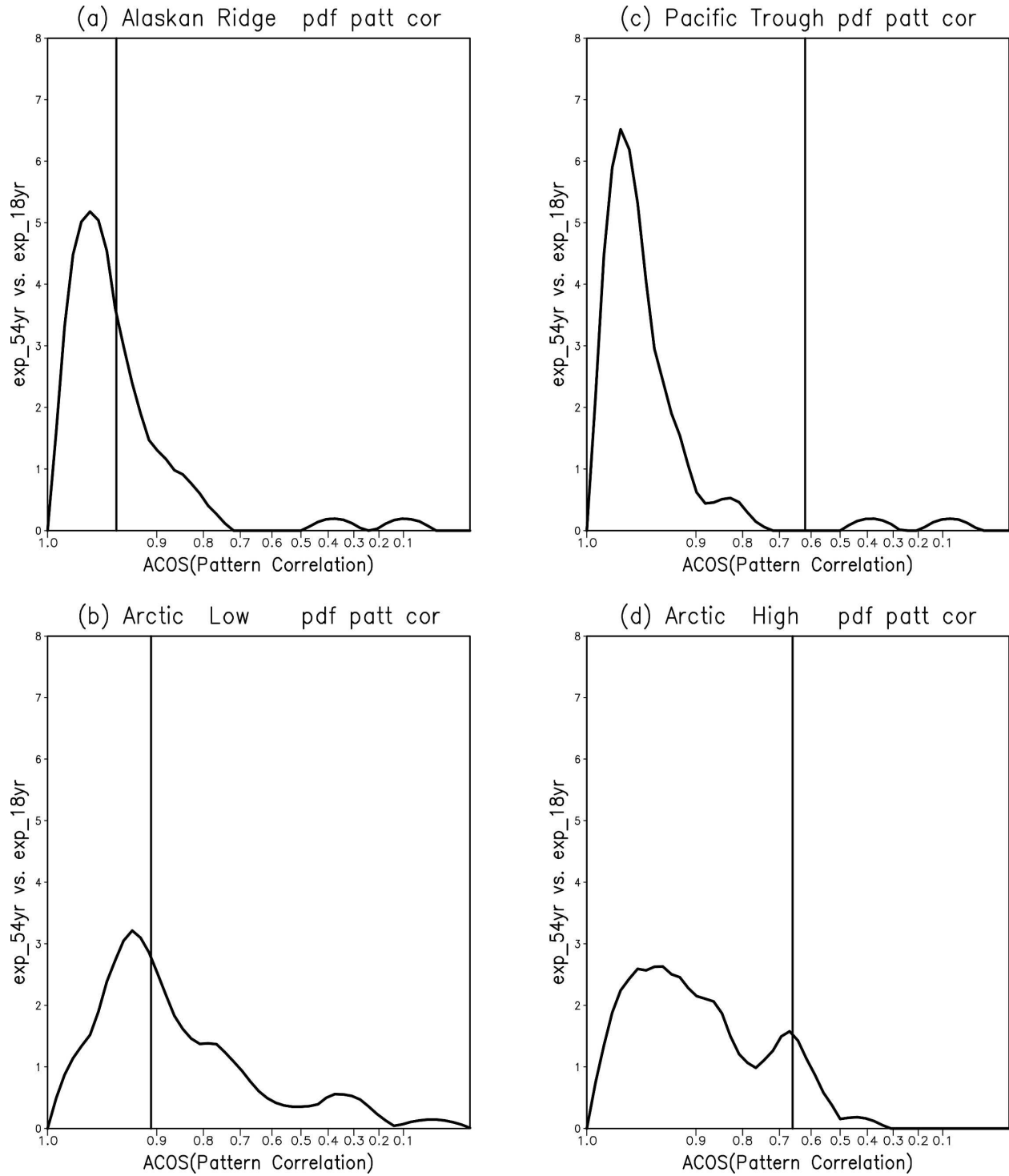
FIG. 7. PDFs of the pattern correlation of the 55 height centroids from the EXP18 dataset with the corresponding height centroid from the EXP54 dataset. In all cases six EOFs and a value of $\alpha = 0.5$ were used. The vertical line shows the pattern correlation between the NCEP54 height centroids and the corresponding NCEP18 height centroid. The abscissa of each plot is the arccosine of pattern correlation, with the pattern correlation values indicated.
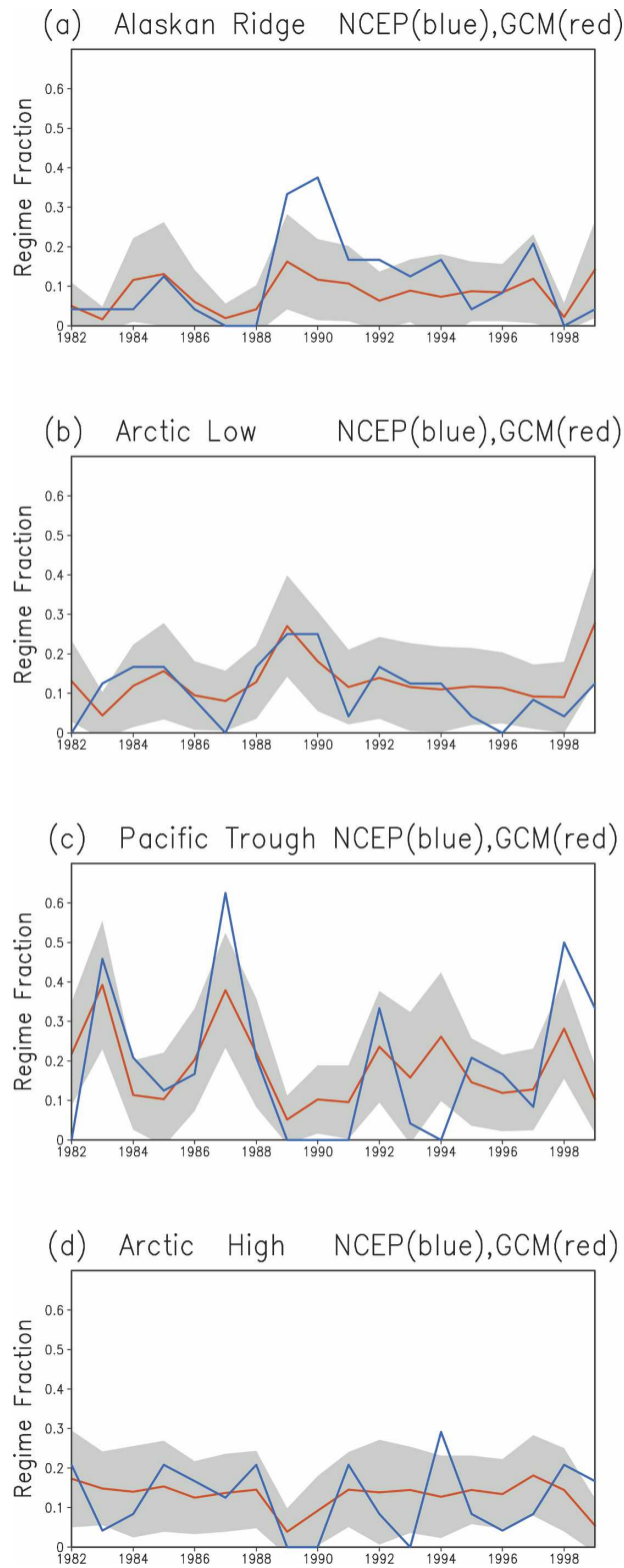
FIG. 8. Interannual time series of the fractional occurrence for each regime from the NCEP18 (GE) datasets are given in blue (red). All clusters correspond to six EOFs, $k = 4$, and $\alpha = 0.5$. The green band gives plus and minus one standard deviation about the red line, where the standard deviation is computed from the 55 ensemble members in the GE (see text for details).

and GE. For example, the simulations underestimate the occurrence of the Alaskan ridge during the winters of 1988/89 and 1989/90, and also underestimate the occurrence of the Pacific trough during the warm ENSO years of 1986/87 and 1997/98. For the PT, AL, and AR, the blue and red lines track each other reasonably well except during the mid-1990s, while, in general, the differences between red and blue lines tend to stay within the gray band.

The overall intermember standard deviation σ of the frequency of occurrence of each regime from the GE18 results is given in the last column of Table 2. Comparing these to the observed changes in frequency also shown in Table 2 for the three 18-yr periods spanning 1948/49–2001/02 and with the time series in Fig. 3, we see that the observed decrease in the AR from the first to second period is significant ($\sim 2\sigma$), as is the increase from the second to the third period. The increase in PT occurrence in period 3 compared to period 1 in the table is highly significant ($\sim 3\sigma$). It is curious that the decrease in what we label the AH regime between period 1 and period 3 in NCEP54 is also nominally significant ($\sim -5\sigma$) based on the GE18 estimates of variability although, since this circulation regime is misrepresented in the GCM, the estimate of variability may not be very good.

7. Results for three clusters

As we have indicated in sections 3 and 4, partitions of the NCEP54 dataset into more than four clusters are less reproducible on average than would be expected from the null hypothesis of a multinormal PDF (see Table 4), while the strength of the clustering (measured by the variance ratio) for $k = 2$ is rather low (see Table 3). But one cannot rule out $k = 3$ as a viable alternative for the NCEP54 data.

The centroids for $k = 3$ are shown in Fig. 9, in a form analogous to the $k = 4$ maps of Fig. 1. (Note that there is no panel b on purpose.) The AR and AH patterns carry over from $k = 4$ to $k = 3$ nearly unaltered, while the $k = 4$ PT and AL are clearly merged into a single pattern shown in Fig. 9c.

The NCEP18 centroid patterns for $k = 3$ are shown in Figs. 10a–c. The AR and AL patterns appear nearly unchanged from their four-cluster versions, while the AH and PT are merged into a pattern that, however, still strongly resembles the PT (cf. Figs. 2c and 10c). Figures 10d–f show the $k = 3$ patterns for the GE18, and can be compared to the maps in Fig. 4. Here the AR and PT, as well as the wave-train-like AH carry over quite well from Figs. 4 to 10, while the AL regime in GE18 is lost.
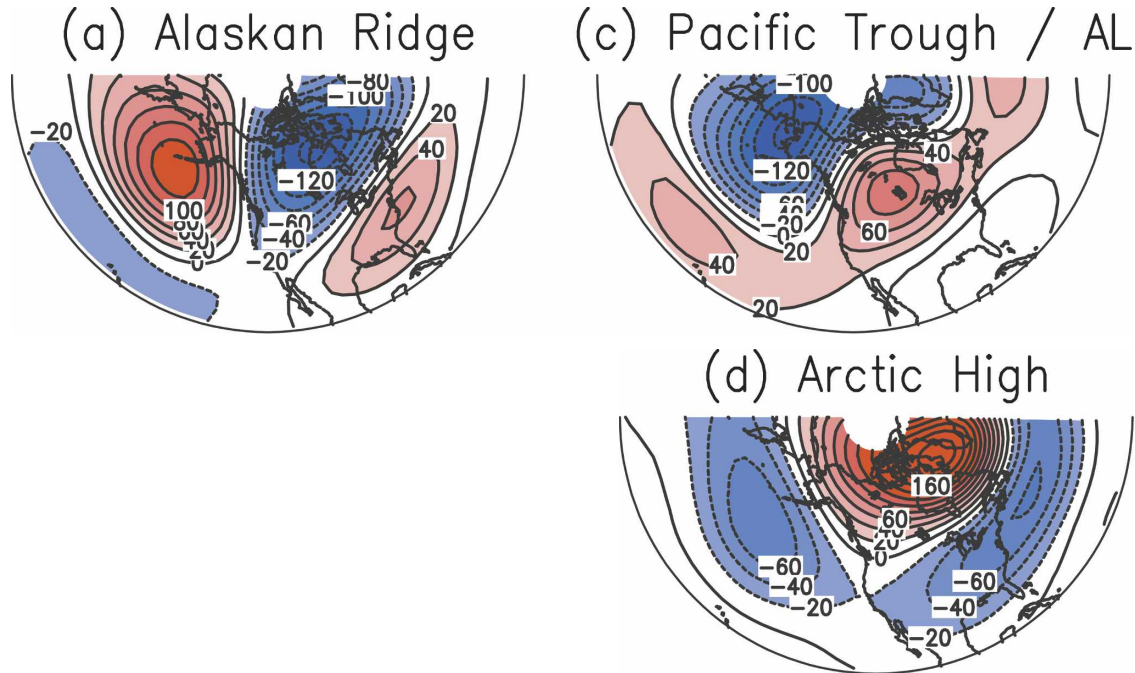
FIG. 9. Cluster centroid maps of 200-hPa height for $k = 3$ from the NCEP54 record, plotted for easy comparison to Fig. 1. (Note that no panel b is present.) The cluster analysis was carried out using six EOFs and a value of $\alpha = 0.5$ (see text for explanation). Contour interval is 20 m with the 0 contour omitted.

The year-to-year variability in the frequency of occurrence of the AR and PT patterns, which are the only ones now in common between GE18 and NCEP18, is presented in Fig. 11 in the same form as the $k = 4$ results in Fig. 8, with the gray band denoting plus or minus one standard deviation in frequency, computed from among the 55 members of GE18.

The AR results for both the GE18 and NCEP18 datasets are remarkably robust. Overall the frequencies of both increase very little in Fig. 11 (cf. Fig. 8) in spite of the fact that there are only three possible clusters. The peaks in 1988/89, 1989/90, and 1996/97 for the NCEP18 dataset are well reproduced, and the peak in 1993/94 is stronger. The overall impression that the simulated (GE18) AR does not occur often enough, but yet tracks the NCEP18 frequency except during the mid-1990s still holds.

Comparison of Figs. 8c and 11b shows that the PT frequency increases somewhat when only three clusters are used, but the year-to-year variation in both simulated and reanalysis datasets is very well preserved. The NCEP18 frequency departs from the gray GE18 band (on the positive side) only during the warm events of 1986/87 and 1997/98, and on the negative side during the winters of 1988/89 and 1989/90. (While the PT does not occur in NCEP18 for $k = 4$ at all during the winter of 1993/94, for $k = 3$ a small fraction is seen, putting the

NCEP18 result just in the gray band.) Also note that the simulations track the reanalysis reasonably well except during the mid-1990s, and this holds for three or four clusters.

8. Discussion and conclusions

While there has been a great deal of interest in better understanding the properties of circulation regimes in the real atmosphere, the short record lengths available for low-frequency boreal wintertime flow present a formidable technical challenge to verifying the significance of deviations of the circulation PDF from Gaussian behavior. This difficulty is only compounded by the realization that the regime properties may be sensitive to forcing, for example, tropical Pacific SST as discussed in this paper or global radiative forcing in a future climate.

To assess whether differences between regime characteristics from reanalysis or simulated datasets are "significant," and whether changes from one period to another are related to forcing, or are simply due to sampling error, it is essential to have some estimates of the relevant sampling properties. Numerical models of the atmosphere or atmosphere–ocean system play an important role in this regard since very large samples (generated with the same forcing) can be constructed.

NCEP18

(a) Alaskan Ridge

(b) Arctic Low

(c) Pacific Trough

GRAND ENSEMBLE

(d) Alaskan Ridge
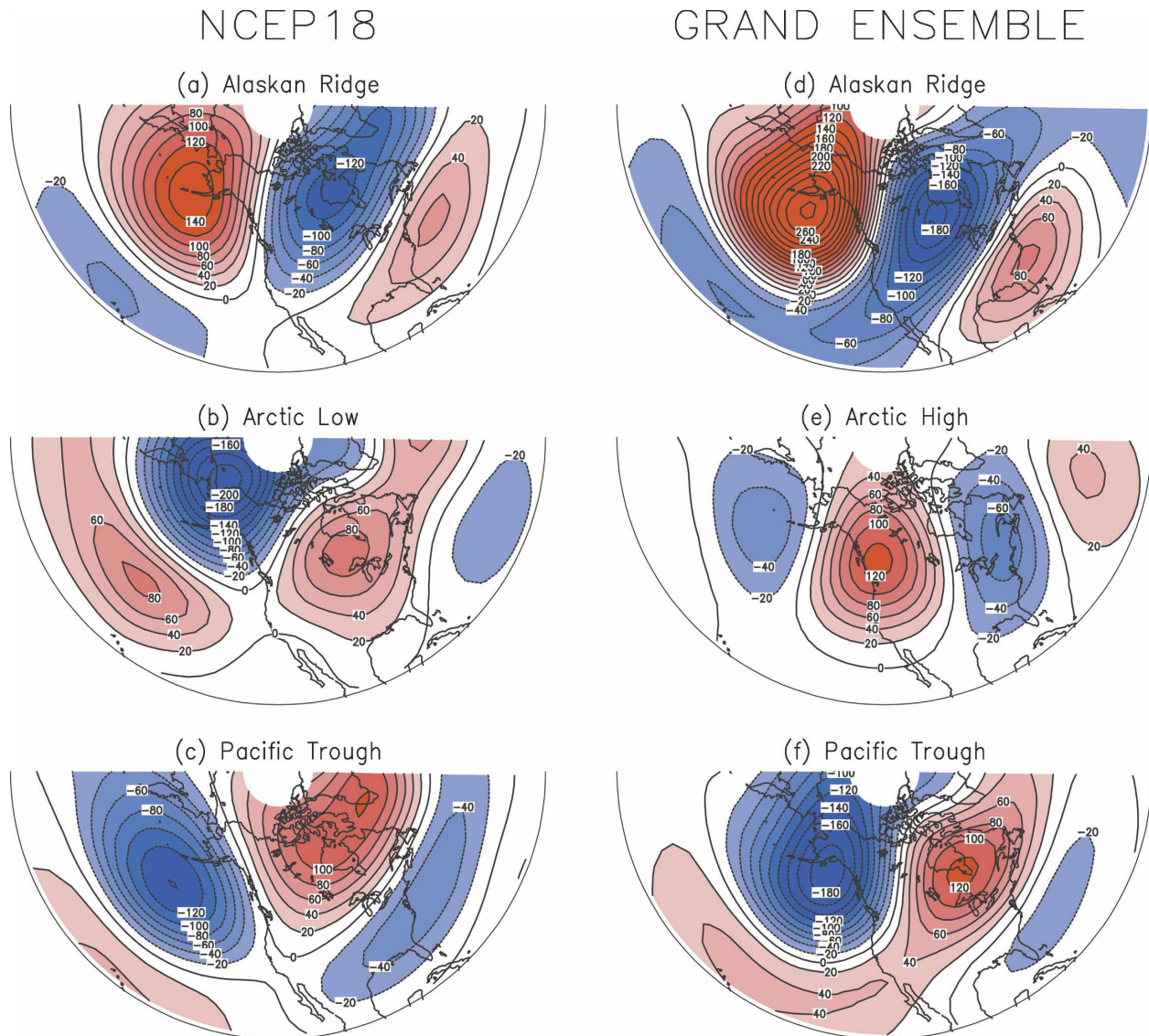
(e) Arctic High

(f) Pacific Trough

FIG. 10. Cluster centroid maps of 200-hPa height for $k = 3$ from the (a)–(c) NCEP18 record and the (d)–(f) GE18 dataset. The NCEP18 clusters are named to be consistent with the NCEP18 in Fig. 2. The GE18 clusters are named to be consistent with GE18 clusters in Fig. 4. The cluster analysis was carried out using six EOFs and a value of $\alpha = 0.5$ (see text for explanation). Contour interval is 20 m with the 0 contour omitted.

In this paper we utilize 55-member ensembles of seasonal integrations of the COLA AGCM utilizing observed SST for 18 winters to provide cluster sampling properties. The simulated clusters shown in Fig. 4 are associated with a variance ratio that is achieved only by far fewer than 1% of multivariate normal Markovian synthetic time series.

The reliance on an imperfect atmosphere-only model to provide the sampling properties motivates the question of the consistency of the properties of model-simulated clusters with those obtained from reanalysis data. Figures 2 and 4 (which compare the regime cen-

troids of the NCEP18 and simulated datasets) clearly show that three out of the four observed clusters (the Alaskan ridge, Arctic low, and Pacific trough) have identifiable counterparts in the reanalysis, while the Arctic high is not simulated. Yet the pattern correlation between the simulated and observed AR and PT clusters falls below the range implied by sampling error, indicating some error in their simulation. The relative frequency of occurrence is better simulated, however, with the small differences between NCEP18 and simulated frequencies lying well within the expected spread due to sampling error.

(a) Alaskan Ridge NCEP(blue), GCM(red)

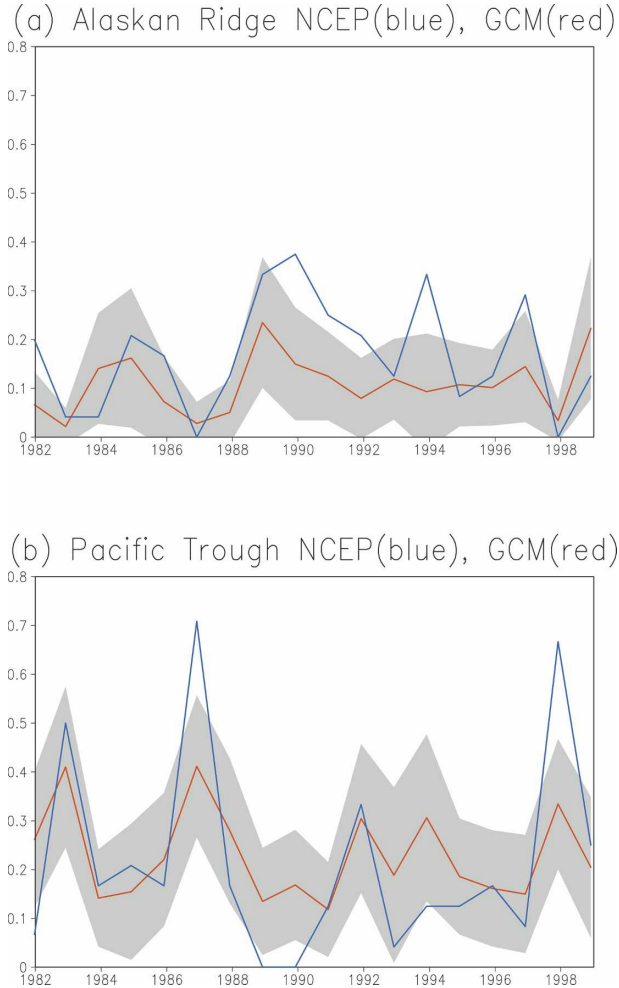## (b) Pacific Trough NCEP(blue), GCM(red)

FIG. 11. Interannual time series of the fractional occurrence for the PT and AR regimes from the NCEP18 (GE) datasets are given in blue (red). (Cluster calculations used six EOFs, $k = 3$, and $\alpha = 0.5$.) The gray band gives plus and minus one standard deviation about the red line, where the standard deviation is computed from the 55 ensemble members in the GE (see text for details).

The direct comparison of simulated and observed clusters for the common recent 18-winter period is somewhat clouded by the fact that formal significance of the cluster analysis in the NCEP18 record is hard to establish. With the longer NCEP54 record, we establish that the four-cluster partition is significant (at a level above 90%) with respect to the null hypothesis of a multinormal dataset, is more reproducible than would be expected on the basis of that null hypothesis, and has as large enough variance ratio to suggest bimodality in the one-dimensional PDF. (Consistency with an independent calculation carried out over the entire hemisphere and agreement with a separate cluster analysis of the zonal wind are also found for a range of cluster numbers.)

Comparing the circulation clusters of the NCEP54 record (Fig. 1) with those of the NCEP18 record (Fig. 2), we ask whether the differences in centroid patterns are simply due to sampling or whether they reflect differences in the external SST forcing between the two periods. Figure 7 suggests that the intrarecord differences for the AR and AL patterns are probably due to sampling error, while the changes in the PT configuration are larger than expected due only to sampling error. The changes seen in this centroid from NCEP54 to NCEP18 resemble the differences between the classic "PNA" pattern and the seasonal mean response of the atmosphere to the strong El Niño events of the last several decades (Straus and Shukla 2002). Our results suggest that the change in the PT is caused by differences in SST anomalies, and also suggest the importance of low-frequency circulation regimes in determining the seasonal mean response to those SST anomalies. An alternate approach to demonstrating the sensitivity of the low-frequency PNA-like cluster to SST forcing is to examine the frequency of occurrence of a fixed regime pattern (that from the NCEP54 record) from year to year; this is shown in Fig. 3.

Figure 3 confirms that both the AR and PT frequency in the NCEP54 record are correlated with seasonal mean tropical Pacific SST, with the former seen more often in cold ENSO winters, and the latter seen more frequently in warm years. The enhancement of the AR pattern during cold winters has been reported previously (e.g., Chen and van den Dool 1997; Renwick and Wallace 1996), but has not before been put into a consistent framework of circulation regimes.

The very strong decadal behavior in the frequency of the AH in the NCEP54 record shown in Fig. 3 (with far fewer occurrences seen in the latter decades) compared to the more consistent behavior of the AL over time suggests an asymmetric behavior of the atmosphere with regard to the expression of the annular-type modes over the Pacific–North American sector. The trend in strengthening of the westerlies over the North Atlantic is seen here to be due to the decreasing importance of the AH pattern.

The regime sensitivity to tropical SST forcing implies a degree of added intraseasonal and interannual potential predictability, and is thus of great interest. Whether there is any hope of realizing this predictability of course depends on the ability of GCMs to capture not only the regimes, but their dependence on SST forcing. Figure 8 shows that the year-to-year variability of simulated frequencies of the AL and PT are close to those observed for the NCEP18 record, and that in general the differences between simulated and observed regime

frequencies stay within the uncertainty band estimated from the simulations.

In the introduction we mention two (competing) hypotheses regarding the change in regime behavior to changes in boundary forcing: a change only in the frequency of occurrence of fixed regimes, versus a change in the regime structure itself. The behavior of the AR is a good example of the former; that of the PT regime of the latter (see Figs. 1, 2, and 8).

Since the sampling properties of the regimes have been estimated with a single uncoupled atmospheric GCM, it is important to investigate the model dependence to these sampling properties, and in particular whether they are strongly changed in atmosphere–ocean coupled models. This work is currently under way.

APPENDIX

Difference Measures and the Effect of Skewness

a. Measures of the differences between maps

In this paper we use two measures of difference between maps. One measure is just the mean squared error; the other the (uncentered) pattern correlation C. These errors are computed in the PC state space. The two maps are described by N-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, where N is the number of PCs retained in the analysis. The mean squared error is proportional to the squared distance (E) between the vectors, which can be written as

$$E = (\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})$$
$$= (x - y)^2 + 2(xy - \mathbf{xy})$$
$$\equiv E_A + E_P, \tag{A1}$$

where x is the length of the vector $\mathbf{x}$, y the length of $\mathbf{y}$; E_A measures the amplitude error, and E_P the phase error. Note that $E_P = 2xy(1 - C)$ and that C is just the cosine of the angle between x and $\mathbf{y}$. When the vectors are aligned in state space, we have $E_P = 0$, $C = 1$, and $E_A = E$. When the vectors have the same length ($x =$

y), we have $E_A = 0$, and $E_P = E$. In all cases discussed in the text, we find that $E \approx E_P$; that is, $E_A \ll E_P$.

b. Effect of skewness on significance estimates

Here we show that the skewness of the NCEP54 PCs does not yield significant clustering against a multinormal background hypothesis using the k-means approach.

We follow the approach of C07, who considers an idealized two-dimensional distribution given by the product of a normal distribution and a Gamma distribution. In C07, clusters are computed for values of the shape parameter $\hat{\alpha} = 50, 20, 10, 5, 3$. The corresponding values of skewness are $s = 0.28, 0.45, 0.63, 0.89, 1.15$, since $s = 2/\sqrt{\hat{\alpha}}$. The k-means method is applied using an approach similar to our significance testing (see section 4) to give the preferred number of clusters. For values of $\hat{\alpha} = 10$ or lower (values of $s = 0.63$ or higher), more than one cluster is preferred, even though the idealized distribution is clearly unimodal.

For the NCP54 dataset, large skewness was confined to the leading four PCs, with (absolute) coefficients of 0.25, 0.23, 0.15, and 0.27, respectively. (Thus from the results of C07, we would not expect the k-means method to falsely report more than one cluster based on skewness alone.) Skewed time series of pseudorandom data using a nonsymmetric Markov model were constructed in a four-dimensional space (see below). In each direction, the skewness, lag-1 autocorrelation, and standard deviation are those of the corresponding PC. (The full multivariate cross-covariance structure is not completely maintained.)

The nonsymmetric Markov model is given by

$$x(t + 1) = x(t) - \beta[1 + \gamma \max(x(t), 0)](x(t) - x_0) + \epsilon, \tag{A2}$$

where β, γ, and x_0 are positive constants, and ϵ is a Gaussian white noise process. Because of the asymmetry in the dissipative term controlled by the parameter γ, large negative values occur more frequently than large positive values. The time series is thus (negatively) skewed with the skewness coefficient increasing with γ. For $\beta = 0.2$ and $x_0 = 1$, values of γ ranging from 0.2 to 1 produce time series with (absolute) skewness coefficients ranging from 0.14 to 0.32.

The number of times in each pseudorandom dataset was set to the length of the time series for $\alpha = 0.5$ in the NCEP54 dataset. For each pseudorandom dataset, a cluster analysis was performed and the significance was assessed as for the NCEP54 dataset. This cluster analysis was repeated for 200 skewed pseudorandom datasets. The average significance of clusters for cluster

number k in the range of 2–4 varied from 50% to 60%, and the number of pseudorandom skewed datasets for which the significance exceeded 90% (95%) was less than 10% (5%). Thus, the k-means method does not indicate significant clustering.

The entire calculation was repeated twice, and the results are the same.

REFERENCES

Alhamed, A., S. Lakshmivarahan, and D. J. Stensrud, 2002: Cluster analysis of multimodel ensemble data from SAMEX. *Mon. Wea. Rev.,* **130,** 226–255.

Arnott, J. M., J. L. Evans, and R. Chiaromonte, 2004: Characterization of extratropical transition using cluster analysis. *Mon. Wea. Rev.,* **132,** 2916–2937.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.,* **115,** 1083–1126.

Cassou, C., L. Terray, J. W. Hurrell, and C. Deser, 2004: North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time and oceanic forcing. *J. Climate,* **17,** 1055–1068.

Chen, W. Y., and H. M. van den Dool, 1997: Asymmetric impact of tropical SST anomalies on atmospheric internal variability over the North Pacific. *J. Atmos. Sci.,* **54,** 725–740.

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

Christiansen, B., 2003: Evidence for nonlinear climate change: Two stratospheric regimes and a regimes shift. *J. Climate,* **16,** 3681–3690.

——, 2007: Atmospheric circulation regimes: Can cluster analysis provide the number? *J. Climate,* **20,** 2229–2250.

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

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

Duchon, C., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.,* **18,** 1016–1022.

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

Hansen, A. R., and A. Sutera, 1990: Weather regimes in a general circulation model. *J. Atmos. Sci.,* **47,** 380–391.

Hsu, C. J., and F. Zwiers, 2001: Climate change in recurrent regimes and modes of Northern Hemisphere atmospheric variability. *J. Geophys. Res.,* **106,** 20 145–20 159.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation. *Science,* **269,** 676–679.

Itoh, H., and M. Kimoto, 1999: Weather regimes, low-frequency oscillations, and principal patterns of variability: A perspective of extratropical low-frequency variability. *J. Atmos. Sci.,* **56,** 2684–2705.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kiehl, J. T., J. J. Hack, G. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate,* **11,** 1131–1149.

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.

Kondroshov, 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.

Marshall, J., and F. Molteni, 1993: Toward 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.

Molteni, F., and S. Corti, 1998: Long-term fluctuations of the statistical properties of low-frequency variability: Dynamical origin and predictability. *Quart. J. Roy. Meteor. Soc.,* **124,** 495–526.

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

——, F. Kucharski, and S. Corti, 2006: On the predictability of flow-regime properties on interannual to interdecadal timescales. *Predictability of Weather and Climate,* T. N. Palmer and R. Hagerdon, Eds., Cambridge University Press, 365–390.

Monahan, A. H., J. C. Fyfe, and G. M. Flato, 2000: A regime view of Northern Hemisphere variability and change under global warming. *Geophys. Res. Lett.,* **27,** 1139–1142.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa–Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.,* **120,** 978–1002.

Namias, J., 1964: Seasonal persistence and recurrence of European blocking during 1958–1960. *Tellus,* **16,** 3–22.

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

——, 1999: A nonlinear dynamical perspective on climate prediction. *J. Climate,* **12,** 575–591.

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

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

Renwick, J. A., and J. M. Wallace, 1996: Relationships between North Pacific wintertime blocking, El Niño, and the PNA. *Mon. Wea. Rev.,* **124,** 2071–2076.

Rex, D. R., 1950: Blocking action in the middle troposphere and its effects upon region climate. II: The climatology of blocking actions. *Tellus,* **2,** 275–302.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimal interpolation. *J. Climate,* **7,** 929–948.

Robertson, A. W., and C. R. Mechoso, 2003: Circulation regimes and low-frequency oscillations in the South Pacific sector. *Mon. Wea. Rev.,* **131,** 1566–1580.

Schneider, E. K., 2002: Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. *J. Climate,* **15,** 449–469.

Shindell, D. T., R. L. Miller, G. A. Schmidt, and L. Pandolfo, 1999: Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature,* **399,** 452–455.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in North-

ern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.,* **56,** 3704–3723.

Stephenson, D. B., A. Hannachi, and A. O'Neill, 2004: On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc.,* **130,** 583–606.

Straus, D. M., 1983: On the role of the seasonal cycle. *J. Atmos. Sci.,* **40,** 303–313.

——, and M. Halem, 1981: A stochastic–dynamic approach to the study of the natural variability of the climate. *Mon. Wea. Rev.,* **109,** 407–421.

——, and J. Shukla, 2002: Does ENSO force the PNA? *J. Climate,* **15,** 2340–2358.

——, and F. Molteni, 2004: Circulation regimes and SST forcing: Results from large GCM ensembles. *J. Climate,* **17,** 1641–1656.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000–1016.

Toth, Z., 1991: Circulation patterns in phase space: A multinormal distribution? *Mon. Wea. Rev.,* **119,** 1501–1511.

——, 1992: Quasi-stationary and transient periods in the Northern Hemisphere circulation series. *J. Climate,* **5,** 1235–1247.

——, 1993: Preferred and unpreferred circulation types in the Northern Hemisphere wintertime phase space. *J. Atmos. Sci.,* **50,** 2868–2888.

Vautard, R., 1990: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.,* **118,** 2057–2081.

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

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

Watson, J. S., and S. J. Colucci, 2002: Evaluation of ensemble predictions of blocking in the NCEP global spectra model. *Mon. Wea. Rev.,* **130,** 3008–3021.

Xue, Y.-K., P. J. Sellers, J. L. Kinter, and J. Shukla, 1991: A simplified biosphere model for global climate studies. *J. Climate,* **4,** 345–364.