Linear and Nonlinear Signatures in the Planetary Wave Dynamics of an AGCM: Probability Density Functions

JUDITH BERNER
NCAR, Boulder, Colorado, and Meteorologisches Institut, Universität Bonn, Bonn, Germany

GRANT BRANSTATOR
NCAR, Boulder, Colorado

(Manuscript received 2 September 2005, in final form 25 April 2006)

ABSTRACT

To identify and quantify indications of linear and nonlinear planetary wave behavior and their impact on the distribution of atmospheric states, characteristics of a very long integration of an atmospheric general circulation model (GCM) in a four-dimensional phase space are examined. The phase space is defined by the leading four empirical orthogonal functions of 500-hPa geopotential heights.

First it is established that nonlinear tendencies similar to those reported in an earlier study of the phase space behavior in this GCM have the potential to lead to non-Gaussian features in the probability density function (PDF) of planetary waves. Then using objective measures it is demonstrated that the model’s distribution of states has distinctive non-Gaussian features. These features are characterized in various subspaces of dimension as high as four. A key feature is the presence of three radial ridges of enhanced probability emanating from the mode, which is shifted away from the climatological mean. There is no evidence of multiple maxima in the full PDF, but the radial ridges lead to three distinct modes in the distribution of circulation patterns.

It is demonstrated that these key aspects of non-Gaussianity are captured by a two-Gaussian mixture model fitted in four dimensions. The two circulation states at the centroids of the component Gaussians are very similar to those associated with two nonlinear features identified by Branstator and Berner in their analysis of the trajectories of the GCM. These two dynamical features are locally linear, so it is concluded that the behavior of planetary waves can be conceptualized as being approximately piecewise-linear, leading to a two-Gaussian mixture with three preferred patterns.

1. Introduction

a. Background

Nonlinearities in the internal dynamics of the atmosphere have the potential to influence the behavior of planetary waves in key respects. For example they can produce highly predictable states, and they can affect the way planetary waves react to external forcing (Palmer 1999). Although effects of nonlinearity are obvious in highly truncated models, as for example in the formation of multiple equilibria in Charney and DeVore’s (1979) model of planetary waves in the presence of orography, it is not apparent what the effect of non-linearity is in nature or in comprehensive atmospheric models. The one consequence that has often been reported is non-Gaussianity in the distribution of planetary wave states, sometimes in the form of multiple modes\(^1\) (e.g., Hansen and Sutera 1986; Kimoto and Ghil 1993a; Hannachi 1997; Cheng and Wallace 1993; Corti et al. 1999). But there is reason to doubt the statistical reliability of some of these findings (Hsu and Zwiers 2001; Stephenson et al. 2004). This is especially true for studies of state distributions in more than one dimension, where strong non-Gaussianity and even multimodality have often been reported. Hence it is not clear whether the prominence sometimes given to non-

\(^1\) Here and throughout the paper the word “mode” is used exclusively in the statistical sense of referring to a local maximum of probability density.
linear interpretations of various large-scale circulation properties [e.g., in the Third Assessment Report of the Intergovernmental Panel of Climate Change (Stocker et al. 2001)] is justified.

A recent body of work has begun to change this situation. Several studies have found signatures of planetary wave nonlinearities by investigating properties of trajectories rather than distributions of states. These include studies by Kimoto and Ghil (1993b), Itoh and Kimoto (1997), Berner (2003), Crommelin (2004), Selten and Branstator (2004) and Branstator and Berner (2005, hereafter BB05), each of which considered trajectories in highly reduced state spaces of high-dimensional systems. An example of results from BB05 is displayed in Fig. 1, which shows mean 24-h tendencies in six planes whose coordinates are defined by the leading four EOFs of 500-hPa geopotential height in an atmospheric general circulation model (AGCM). As BB05 explain, if planetary wave dynamics were well approximated by the standard, rudimentary model of planetary wave behavior, namely a linear model forced by additive white noise, the vectors in this plot would be characterized by elliptical motion that is damped toward the origin, much like Fig. 1f. A strong departure from such a signature, like that in Fig. 1b, can only result from either deterministic nonlinearities or state-dependent noise, which is just another indication of nonlinearities.

Interestingly, not only do the mean tendencies in Fig. 1 serve as evidence that planetary wave nonlinearities have tangible effects on atmospheric systems with many degrees of freedom, they also imply that those nonlinearities should produce marked departures from Gaussianity in the distribution of planetary wave states. That one can draw these conclusions stems from the stochastic modeling study of Berner (2005), which demonstrated that the nonlinear component of the tendencies in Fig. 1 can produce nontrivial non-Gaussianity in PDFs of planetary wave states. This point is further made in the simple example of the following subsection.

Given these results we hypothesize that past failures to detect statistically significant non-Gaussianity are not an indication that planetary wave nonlinearities are inconsequential but rather a result of the short observational record. With this in mind we have undertaken
a study of the distribution of planetary wave states in the same AGCM examined by BB05 with the intent of quantifying and characterizing its non-Gaussianities. We have reasoned that this endeavor should be informative because 1) the AGCM has more complete dynamics than the toy and intermediate models often used to study the influence of planetary wave nonlinearities, 2) a much larger population of states is available than has been used in corresponding studies of data from nature, and 3) the results from BB05 serve as a useful background for guiding the analysis.

b. Simple example

Before describing our study we further motivate it by presenting a simple example to demonstrate that nonlinearities of the kind displayed in Fig. 1 have the potential to produce substantial departures from Gaussianity in the distribution of planetary wave states. The example is based on the fact that if—as in BB05—the planetary wave behavior of the AGCM is projected onto two-dimensional subspaces and if the planetary wave dynamics in that plane can be approximated as a first-order Markov process driven by additive white noise, then the deterministic dynamics of the stochastic model equals the mean tendencies in Fig. 1. Berner (2005) has demonstrated that for the EOF1–4 plane (Fig. 1c) these assumptions lead to a stochastic model that reproduces many properties of the AGCM planetary waves in that plane.

Based on these facts we construct a two-dimensional Markov model that is likely to act rather like the AGCM. This is the model

\[ x_i = h_i(x) + \sigma_i \epsilon_i, \quad i = [1, 2], \]

where the idealized deterministic drift \( h_i(x) \) for \( x = (x_1, x_2) \) is chosen to have most of the key features of mean tendencies in the EOF1–4 plane of the AGCM. In (1) each \( \epsilon_i \) is centered Gaussian noise that is uncorrelated in time and the constant noise amplitudes \( \sigma_i \) are specified in such a way that the standard deviation of each \( x_i \) is one. As described more fully in appendix A, \( h \) is composed of a linear damping defined at all locations in the plane (Fig. 2a) and two linear oscillations (Fig. 2b), each defined in a half plane and corresponding to oppositely sensed rotations about two points of equal distance from the origin on the \( x_1 \) axis. These components are intended to mimic similar components that BB05 found were good approximations to the mean phase space tendencies of the AGCM (Fig. 1). When combined (Fig. 2c) they generate a piecewise-linear function that has a V-shaped distribution of small deterministic velocities and two regions with opposing swirling motions, much like the EOF1–4 AGCM plane.

Integrating (1) forward in time and estimating the PDF of the resulting population of states we find that if we use only the linear damping component (Fig. 2a) for \( h \) then the distribution of states is purely Gaussian (Fig. 2d). This is a well-known result, for in this case (1) is simply an Ornstein–Uhlenbeck process. On the other hand when the complete, piecewise-linear form of \( h \) (Fig. 2c) is used, a distinctly non-Gaussian PDF results (Fig. 2e). In this case the mode of the distribution shifts and ridges of probability radiating from the mode toward the first and fourth quadrants are produced, so that there is enhanced probability in three directions. In this simple system the reason for this non-Gaussian distribution is easy to understand. Because \( h \) is piecewise-linear, in each half domain the equation defining the system is identical to the equation of an Ornstein–Uhlenbeck process. Hence, to the extent that the effect of the boundary between the subdomains can be ignored the distribution in each half domain will be locally Gaussian. Using a method described later in the paper we can verify this fact by approximating the PDF as a mixture of two Gaussians. These are shown in Fig. 2f from which it can be seen that away from the boundary the complete PDF (Fig. 2e) is well approximated by a single Gaussian in each half domain and that those Gaussians have orientations and structures as one would expect from the locally linear deterministic dynamics. On the other hand, for points near the ordinate the influence of both subdomains is felt, and as a result no finite number of Gaussians can completely match the entire distribution.

c. Issues

Based on properties of this simple example, as well as the findings of earlier studies of planetary wave PDFs, we have selected several issues to address concerning the distribution of states in the same four-dimensional AGCM subspace investigated by BB05.

- Is the distribution of planetary wave states Gaussian? We consider this by plotting histograms and calculating objective measures of Gaussianity. As a step toward alleviating the masking of important features that can potentially happen when states are projected onto a one- or two-dimensional subspace, we consider PDFs in as many as four dimensions.

- Is there any correspondence between the structure of PDFs in various EOF planes and the structure of the mean tendencies in those planes (Fig. 1)? Such a relationship is anticipated by our simple example’s sensitivity to \( h \) and by Berner’s (2005) success at modeling planetary wave PDFs with two-dimensional
models whose deterministic tendencies match those in Fig. 1.

• Are there multiple modes in the PDFs? Earlier studies of planetary waves have reported their existence, and the simple stochastic example above makes it apparent that there is the potential for them to exist. For if the noise used in the solution of Fig. 2f is weakened then the solutions have modes in each of the linear subdomains.

• Rather than determining state distributions, is it more appropriate to find pattern distributions in which amplitude information is ignored? Some studies [principally Kimoto and Ghil (1993a) and Crommelin (2004)] have found it is advantageous to investigate the distribution of planetary wave patterns.

• Given the simple example’s demonstration that the approximate piecewise-linearity of mean tendencies in the AGCM can produce distributions that are approximately mixtures of a few Gaussian components, are the AGCM’s PDFs a mixture of Gaussians? We pay special attention to a four-dimensional, two-component Gaussian mixture. This is the distribution anticipated by a four-dimensional generalization of the simple example together with BB05’s finding that the AGCM four-dimensional mean tendencies are well-approximated by a two-segment piecewise-linear function.

Underlying each of these questions is our curiosity as to what degree planetary wave behavior matches that expected from linear dynamical considerations and to what degree nonlinearities (like those in Fig. 1) leave a distinguishable imprint.

2. Data

Data for our investigation come from the same AGCM used in BB05 and Berner (2005), namely the model known as Community Climate Model version 0 (CCM0), which was developed at the National Center for Atmospheric Research. It has been run in perpetual January mode; that is, with temporally constant bound-
ary conditions and no diurnal cycle, for a total of 14 million days and was sampled twice a day. Further details about the AGCM and its characteristics are available in BB05 and the references therein. All of the results described in this paper are derived from the same 7 million days used in BB05. To assure the robustness of our findings we have repeated each calculation for the second 7 million days and found that features we discuss are unchanged. In some instances we have performed further tests of sampling adequacy as described in the text.

For data reduction purposes the AGCM states are expressed in terms of 500-hPa geopotential height, and their dimensionality is further reduced by projecting the departures of this field from its long-term mean onto their leading EOFs (Fig. 2 of BB05). The EOFs are calculated from area-weighted, global, temporally unfiltered data. Since these leading EOFs are very weak in the Southern Hemisphere, our analysis primarily concerns Northern Hemispheric variability. The variances of 12 hourly data explained by the first four temporal EOF coefficients, also called principal components (PCs), are 8%, 5%, 4%, and 4%, respectively. (For monthly means these same patterns explain 38% of the variance.) Subsequently, we have standardized the PCs to have a standard deviation of one. For brevity, we introduce the symbol $\varphi_{i,j}$ to denote the direction spanned by the $i$th EOF. Analogously, the plane spanned by the $i$th and $j$th EOF is denoted by $\varphi_{i,j,k,l}$ and $\varphi_{i,j,k}$ and $\varphi_{i,j,k,l}$ denote three- and four-dimensional subspaces.

3. Methods

For visual inspection of the distribution of states we use histograms to estimate various PDFs of PCs. These histograms are based on state counts in bins formed by dividing each state space direction into 30 segments within the range $[-3, 3]$. Contrary to common practice, except where noted in the text, our large dataset makes it unnecessary to apply smoothing or function fitting to the raw counts to estimate PDFs.

To objectively determine whether a particular AGCM PDF departs from being Gaussian we use four measures that compare its properties to characteristics of a multivariate Gaussian distribution. The reference distribution used in this comparison is the multivariate Gaussian whose parameters maximize the likelihood that the AGCM data were drawn from it. This distribution is

$$f_{\text{ref}}(x) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} \exp[-1/2(x - \mu)^T C^{-1}(x - \mu)],$$

where $d$ is the dimension of the state vector $x$, and $\mu$ and $C$ are the mean and covariance-matrix of realizations of $x$.

One of the four measures is simply the normalized root-mean-square difference (RMSD, shown as $R$ below) between an AGCM histogram and the reference distribution:

$$R = \left \{ \frac{\sum_{n:m|f(x_m) > 0} [f(x_m) - f_{\text{ref}}(x_m)]^2 \sum_{m:f(x_m) = 0} f_{\text{ref}}(x_m)^2} \right \}^{1/2}.$$  

Here $f(x_m)$ is the AGCM density in the $m$th bin of a histogram. Though not commonly used, we find $R$ to be an easily understood quantifier of non-Gaussianity. A second measure is Pearson’s statistic ($P$) (e.g., Wilks 1995), which is frequently used in the $\chi^2$-goodness-of-fit test to determine if two samples are drawn from the same underlying distribution. It is given by

$$P = n_{\text{eff}} \sum_m \left[ f(x_m) - f_{\text{ref}}(x_m) \right] \frac{f_{\text{ref}}(x_m)}$$

where $n_{\text{eff}}$ is the effective sample size and $f^*$ the distribution of the reduced sample consisting of $n_{\text{eff}}$ statistically independent states. Based on the PC decorrelation times reported in Berner (2005) we conservatively assume that every hundredth day of our dataset represents an independent sample, so that $n_{\text{eff}} = 3.5 \times 10^4$. Note that if $f$ is exactly Gaussian, then $R$ and $P$ both have the value zero.

While $R$ and $P$ are general indicators of non-Gaussianity, they give no indication of the form of the non-Gaussianity they measure. By contrast the other two measures are indicators of specific types of non-Gaussianity. These two are the standard measures of third and fourth moment quantities, namely skewness $S$ and kurtosis $K$, as generalized to multidimensions (Mardia et al. 1979) by the expressions

$$S = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [(x_i - \mu)^T C^{-1}(x_j - \mu)]^3,$$

$$K = \frac{1}{n} \sum_{i=1}^n [(x_i - \mu)^T C^{-1}(x_i - \mu)]^4,$$

where $n$ denotes the sample size. The values of skewness and kurtosis in the Gaussian reference distribution are zero and $d(d+2)$, respectively, where $d$ is the dimensionality of $x$, so departures from these values are indications of non-Gaussianity; $S$ from (4) is positive.
definite, but its more familiar one-dimensional definition,

\[ S = \frac{1}{n} \sum_{i=1}^{n} \frac{(x - \mu)^3}{C}, \]

(7)
can take either sign with a positive value corresponding to a distribution with an enhanced positive tail and a negative value meaning a strong negative tail. Because of this useful interpretation we use (7) when calculating skewness for one-dimensional data. In the multidimensional case we take nonzero values of skewness to be an indication of a lack of symmetry. A highly negative value of \( K - d(d + 2) \) corresponds to a flat distribution with weak tails.

The values of the four measures of departures from Gaussianity are necessarily estimated based on a finite number of samples. To ensure that the values we calculate are significantly different from those obtained from randomly drawing from a Gaussian distribution, we calculate their \( p \) values (e.g., Wilks 1995). This is done analytically for skewness, kurtosis, and Pearson statistic because the distributions of these measures for samples from Gaussian populations are known (e.g., Mardia et al. 1979) as a function of the effective sample size. In the case of the RMSD no analytical expressions exist and we use the Monte Carlo technique applied to draws from a Gaussian population with the same covariance and lagged-covariance structure as our dataset.

As a further means of characterizing non-Gaussianity, we fit a sequence of Gaussian mixture models of ever greater complexity to the data. These models consist of weighted sums of multivariate Gaussians:

\[ h^{(k)}(x) = \sum_{j=1}^{k} \alpha_j g_j(x), \]

(8)

where

\[ g_j(x) = \frac{1}{(2\pi)^{d/2}|C_j|^{1/2}} \exp\left[-1/2(x - \mu_j)^T C_j^{-1}(x - \mu_j)\right] \]

(9)

and \( \alpha_j, \mu_j, \) and \( C_j \) denote the weight, mean and covariance of the \( j \)-th Gaussian component. As \( k \) is increased, \( h^{(k)} \) is necessarily a better and better approximation to the sampled distribution as measured by the log-likelihood

\[ L = \sum_{j=1}^{n} \log g(x_j) \]

(10)

for some distribution \( g \). Since \( h^{(1)} \) equals \( f_{\text{eff}} \), comparison of the likelihoods for various values of \( k \) to the likelihood of \( h^{(1)} \) provides another means of measuring departures from Gaussianity. Furthermore, as the simple example of the introduction suggests, if our AGCM PDFs can be approximated by mixtures for small values of \( k > 1 \), this will be consistent with the tendencies of Fig. 1 having an imprint on the distribution of states.

As described in Titterington et al. (1985) and carried out by Smyth et al. (1999) and Hannachi and O’Neill (2001) for planetary wave datasets, fitting the \( h^{(k)} \) mixtures can be accomplished by using the estimation-maximization (EM) algorithm. We have used this algorithm according to the implementation strategy described in appendix B. When evaluating the resulting mixtures, we calculate \( L \) for a given \( k \), not for the data used to find the parameters \( \alpha_j, \mu_j, \) and \( C_j \), but rather for independent data. This is especially important when we wish to decide how many components constitute the best mixture approximation to a dataset. If we did not cross-validate we would always find that increasing \( k \) led to a better fit.

4. Probability density functions

a. Non-Gaussianity

We begin our analysis of planetary wave state distributions by searching for indications of non-Gaussianity in subspaces of higher and higher dimension. Histogram estimates of the one-dimensional PDFs of each of the leading four PCs together with the reference Gaussian distribution (gray shading) are shown in Fig. 3.

According to the values of skewness, kurtosis, Pearson statistic, and RMSD (Table 1a) and an effective sample size of \( n_{\text{eff}} = 3.5 \times 10^4 \) we can say with at least 98% confidence that in all four directions the distributions are non-Gaussian. But to the eye, departures from Gaussianity are only evident in PC1 and possibly PC2. The \( p \) values for the measured kurtosis in the first four directions are \( 7 \times 10^{-3}, 2 \times 10^{-14}, 2 \times 10^{-5}, \) and \( 1.5 \times 10^{-2} \), while they are smaller than machine precision for the measured values of skewness. Thus we find that the non-Gaussianity manifests itself more in terms of skewness than of kurtosis. From Table 1a we notice that when kurtosis is ignored, the objective measures support our initial impression that PC1 has by far the strongest departures from the reference distribution.

For two-dimensional PDFs, the values of \( S, K, R, \) and \( P \) in Table 1b give much the same general impression seen in the one-dimensional results. Non-Gaussianity is highly significant in all planes, with the lowest confidence level of 94% occurring in \( \varphi_{1,4} \) for kurtosis, while it is more than 99% for all other planes and measures.

Non-Gaussian skewness, RMSD, and Pearson statis-
tic are most pronounced in planes that include PC1. Examining plots of the PDFs that these measures are derived from (Fig. 4) we see that considering a higher dimension has added to our knowledge of properties that are contributing to the non-Gaussianity. These diagrams show that in addition to modes being shifted toward negative PC1 and positive PC2, as can be inferred from the one-dimensional results, in some planes non-Gaussianity takes the form of three radial ridges of enhanced probability emanating from the modes. In the next section we objectively find directions of enhanced probability in the 4D PDF and their projections are indicated in Fig. 4.

To see if additional information about departures from Gaussianity can be learned by considering still higher-dimensional PDFs, in Fig. 5 we plot three-dimensional PDFs for each subspace spanned by our four PCs. These results are qualitatively like those for two dimensions. In subspaces that include PC1 there are ridges (now bulges) of unusually high probability but the number of such features remains at three in any one subspace. Moreover we see that, just as in lower-dimensional PDFs, there is no indication of multiple modes. Turning to Table 1c’s objective measures of non-Gaussianity in these three-dimensional PDFs we again find that qualitatively the conclusions drawn from the lower dimensional analysis continue to hold. The $p$ values in all subspaces are smaller than machine precision, so that we can be more than 99% sure that the PDFs are non-Gaussian.

For four dimensions a plot of the PDF is impractical. But Table 1d indicates that our conclusions concerning $S, K, R$, and $P$ are unchanged; by all four measures the PDF is non-Gaussian with at least 99% confidence.

b. Multimodality

In light of the emphasis that earlier studies have put on multimodality we look more carefully at the distri-

![Fig. 3. Histogram of 12-hourly CCM0 500-hPa geopotential heights projected onto the leading four principal components together with their respective maximum-likelihood Gaussians (shading). Measures of non-Gaussianity are denoted in the upper right corner. RMSD and Pearson statistic are given in units $10^{-1}$ and $10^{-3}$, respectively. Conventions and units are the same as in Figs. 1 and 2.](image)
Table 1. Measures for assessing non-Gaussianity of (a) one-, (b) two-, (c) three-, and (d) four-dimensional PDFs. Each estimated value, which is based on $7 \times 10^6$ days of data, has an accuracy of at least $\pm 0.01$. This has been confirmed by comparing them with estimates using an independent sample of the same size as well as with four estimates based on $3.5 \times 10^6$ days and with an estimate based on $14 \times 10^6$ days. In all cases differences between each reported value and all other estimates had magnitude less than 0.01.

<table>
<thead>
<tr>
<th></th>
<th>(a) 1D</th>
<th>(b) 2D</th>
<th>(c) 3D</th>
<th>(d) 4D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varphi_1$</td>
<td>$\varphi_2$</td>
<td>$\varphi_3$</td>
<td>$\varphi_4$</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.36</td>
<td>-0.21</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.90</td>
<td>2.86</td>
<td>2.92</td>
<td>2.96</td>
</tr>
<tr>
<td>RMSD (in $10^{-1}$)</td>
<td>1.12</td>
<td>0.68</td>
<td>0.28</td>
<td>0.18</td>
</tr>
<tr>
<td>Pearson (in $10^4$)</td>
<td>4.32</td>
<td>1.54</td>
<td>0.28</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(b) 2D</th>
<th>(c) 3D</th>
<th>(d) 4D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varphi_{1,2}$</td>
<td>$\varphi_{1,3}$</td>
<td>$\varphi_{1,4}$</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.19</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.62</td>
<td>7.76</td>
<td>7.95</td>
</tr>
<tr>
<td>RMSD (in $10^{-1}$)</td>
<td>1.47</td>
<td>1.43</td>
<td>1.45</td>
</tr>
<tr>
<td>Pearson (in $10^4$)</td>
<td>3.26</td>
<td>3.74</td>
<td>3.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(c) 3D</th>
<th>(d) 4D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varphi_{1,2,3}$</td>
<td>$\varphi_{1,2,4}$</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.28</td>
<td>0.32</td>
</tr>
<tr>
<td>RMSD (in $10^{-1}$)</td>
<td>1.86</td>
<td>1.85</td>
</tr>
<tr>
<td>Pearson (in $10^4$)</td>
<td>2.69</td>
<td>2.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(d) 4D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varphi_{1,2,3,4}$</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.44</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>23.33</td>
</tr>
<tr>
<td>RMSD (in $10^{-1}$)</td>
<td>2.38</td>
</tr>
<tr>
<td>Pearson (in $10^4$)</td>
<td>3.10</td>
</tr>
</tbody>
</table>

For each slice we calculate $S$, $K$, $R$, and $P$ and find in some planes they indicate departures from Gaussianity that are larger than for any of the 2D projections we studied earlier. This suggests that some non-Gaussian features are indeed masked by considering projected states.

Two examples of especially non-Gaussian slices are displayed in Fig. 6. The PDF in Fig. 6a pertains to $\varphi_{1,2}$ under the constraints, that $2.6 < PC3 < 2.8$ and $0.8 < PC4 < 1.0$. It has the largest skewness (15.0) and largest kurtosis (18.4) of any planes considered. The corresponding $p$ values for a very conservative effective sample size of $n_{\text{eff}} = 1000/200 = 5$ yields a $p$ value of $1.4 \times 10^{-2}$ for skewness and $3.6 \times 10^{-3}$ for kurtosis, so that we have more than 98% confidence that the slice is non-Gaussian. Interestingly, the slice has two distinct local maxima. One of the maxima is located at $PC1 = -1.0$, $PC2 = -0.1$ while the other is at $PC1 = 1.2$, $PC2 = -0.3$. Figure 6b shows the slice with the largest $R$ (0.32) and $P$ (118.4). It occurs for $\varphi_{1,3}$ when $0.6 < PC2 < 0.8$ and $2.8 < PC4 < 3.0$. This distribution is unimodal but has strong non-Gaussian signatures in the form of two distinct ridges. The PDFs obtained from the second half of the data (not shown) show the same bimodality as Fig. 6a and have the same ridges as Fig. 6b, so that we deem these non-Gaussian features robust.

To determine whether slices like those in Fig. 6a are manifestations of multiple four-dimensional modes that our earlier analysis missed, we expand the analysis of conditional samples to three dimensions. When we do this we find that all planes that have multiple maxima are cuts through distributions with only a single maximum. As an example Fig. 7a shows the three-dimensional hyperslice in $\varphi_{1,2,3}$ for which $0.8 < PC4 < 1.0$. Note that the slice of Fig. 6a corresponds to a cut through this PDF at $PC3 = 2.7$. From this diagram and

$$f(PC_i, PC_j | u_m - \delta < PC_k < u_m + \delta; \nu_n - \delta < PC_i < \nu_n + \delta)$$ (11)

and examine each such slice when $\delta = 0.1$, $u_m = -3 + (2m - 1)\delta$, $\nu_n = -3 + (2n - 1)\delta$ and $m$ and $n$ are integers between 1 and 30. To ensure the robustness of our results we consider only slices for which there are at least 1000 samples, but even so this means estimating PDFs using far smaller samples than elsewhere in our study. Therefore we employ an Epanechnikov kernel (Silverman 1986). The smoothing parameter in the kernel is adjusted for each slice in such a way that the non-Gaussian features found in one-half of the data are also present in the second half. Thus our smoothing can be considered conservative, even if only a small number of states fall into a slice.

On the other hand, by modifying the space in which we search for multiple modes, we can find instances of multiple local density maxima that help to characterize non-Gaussian features. For example, if we conditionally sample the dataset by restricting ourselves to two-dimensional slices that are parallel to two of the EOF directions, then in some planes there is more than one density maximum. To locate these special planes we consider conditional distributions
others drawn to show additional density surfaces, we find that the two maxima of Fig. 6a are connected by density ridges to a global maximum at PC1 = 0.3, PC2 = 0.3, PC3 = 0.2. Hence a distribution like that in Fig. 6a, though certainly a distinctly non-Gaussian conditional PDF, does not correspond to bimodality in three or four dimensions. Rather it is actually an artifact of taking a two-dimensional slice through a V-shaped region of enhanced density in a space of higher dimension.

A second approach that we find will produce multiple local maxima in plots of probability density is to use Kimoto and Ghil’s (1993b) suggestion of considering the probability of patterns rather than the probability of states. Stephenson et al. (2004) refer to these PDFs in which amplitude information is discarded as angular PDFs. They are found for a space of dimension $d$ by projecting the states onto the unit sphere thus producing a marginal PDF of dimension $d - 1$. Since this method reduces the dimensionality by one it is often used as a way to increase the statistical significance of results. But we use it because several studies, including Kimoto and Ghil (1993b), Crommelin (2004), and Selten and Branstator (2004) found angular PDFs of planetary waves have multiple maxima.

When we create angular PDFs for each of the 3D subspaces of Fig. 5, the distributions of Fig. 8 result. In these plots, if the PDF for $\varphi_{i,k}$ is being analyzed, then states with Cartesian coordinates $x_i, x_j, x_k$ have been transformed to spherical coordinates via

$$r = (x_i^2 + x_j^2 + x_k^2)^{1/2}$$

$$\theta = \tan^{-1} \frac{x_j}{x_i}$$

$$\phi = \cos^{-1} \frac{x_k}{r},$$

and state density is displayed as a function of angles $\theta$ and $\phi$. The first three panels of Fig. 8 all pertain to subspaces that include EOF1 and they are very similar. In each there are three maxima, near points denoted M1, M2, and M3. The M1 maxima are isolated while the M2 and M3 features are always connected by a ridge that is situated along a great circle. By contrast, the fourth panel of the figure, which concerns $\varphi_{2,3,4}$, has a
single maximum, and it is weaker than the maxima in the other subspaces.

The points we have marked M1, M2, and M3 in the various panels of Fig. 8 result from consideration of an additional angular PDF. This is the PDF that is produced by projecting four-dimensional states \((x_i, x_j, x_k, x_l)\) onto the unit hypersphere by transforming to generalized spherical coordinates \((r, \psi, \theta, \phi)\) such that

\[
    r = \left( x_i^2 + x_j^2 + x_k^2 + x_l^2 \right)^{1/2} 
\]

\[
    \psi = \cos^{-1} \frac{x_l}{r} 
\]

\[
    \theta = \tan^{-1} \frac{x_j}{x_i} 
\]

\[
    \phi = \cos^{-1} \frac{x_k}{r \sin \psi} 
\]

As seen in Fig. 9a the resulting distribution of densities is complex. It becomes easier to decipher when, in Fig. 9b, we look at a second, higher-valued density surface for the same angular PDF. Combining information from these two plots it is apparent that there are three local maxima, one isolated, labeled M1, and two connected, labeled M2 and M3. It is the projections of the positions of these three features onto the angular PDFs of Fig. 8 that we have given the same labels to in those plots. Since they coincide with the local maxima in each of the first three panels of Fig. 8, we conclude that the features in Fig. 8 with the same label have a common higher-dimensional counterpart.

The three maxima of Fig. 9b also have counterparts in the state PDFs of Figs. 4 and 5. In those figures the three straight lines emanating from the origin indicate the three patterns. In planes and cubes that include EOF1, the patterns M1, M2, and M3 coincide with the...
radial ridges we have already noticed. In the remaining subspaces there is no clear correspondence with high density features, but the three patterns have smaller projections onto these subspaces so we have not indicated their positions in the figure.

5. Gaussian mixtures

a. Two-dimensional analysis

As section 3 points out, other studies have used mixtures of Gaussian distributions to identify structure in planetary wave PDFs, and this appears to be an especially appropriate approach for our data given the approximately piecewise-linear functionality of its associated mean tendency field. Considering the strong directional dependence of the PDFs that we found when planetary states are projected onto different planes, we use Gaussian mixtures to differentiate the structure of the PDFs in these planes.

When we apply the EM algorithm in the manner described in section 3 and appendix B to each of our six planes for up to eight Gaussian components $k$ and evaluate the resulting mean cross-validated log-likelihoods, the values in Fig. 10a result. In every plane the log-likelihood values increase with increasing $k$ indicating that the more component Gaussians are fitted the better the mixture model is able to capture the details of the GCM PDF. This clearly indicates that the mixture rejects a one-Gaussian fit. The non-Gaussianity is of a complexity that cannot be represented fully by a mixture model of less than eight Gaussians. This is consistent with the results of Christiansen (2007), who fitted mixture models to idealized data from skewed distributions and found that the more data were used the more Gaussians were supported by the cross-validated likelihoods. On the other hand, the mean cross-validated log-likelihood increases greatly for fits of more than one Gaussian, with the greatest relative and absolute increase occurring for a mixture of two Gaussians. Indeed, in every plane the total increase from including the last six components is less than the increase from just including a second Gaussian.

Improvements in log-likelihood are difficult to understand in a quantitative fashion, so we use another indicator, called RMSE, of the similarity of mixture models $h(k)$ and histograms $f$ of the data they are based on:

$$\text{RMSE} = \left( \frac{\sum_{m,f(x_m)<0} [h^{(k)}(x_m) - f(x_m)]^2}{\sum_{m,f(x_m)>0} f(x_m)^2} \right)^{1/2}. \quad (19)$$

One drawback of basing results on log-likelihood that RMSE does not suffer from is that log-likelihood cannot be used to compare how well models fit different observed distributions. Note, for example, that all values of log-likelihood in Fig. 10a are the same for $k = 1$. (This happens because for $k = 1$ likelihood depends only on the mean and covariance of the data, which are the same for all planes in our calculations with standardized data.)
When the RMSE is calculated for each of the mixture models in each of the planes we find (Fig. 10b) that by this measure, too, the more components in the mixture the better the model fits the independent data. And again much of the improvement over a single Gaussian can be attained simply by using two Gaussian components. Squaring the values in this figure gives quantities that are easy to interpret. Consider the three most non-Gaussian planes. For a single-Gaussian model the mean squared errors are more than twice as large as mean squared density values, but for a two-component model mean, square errors are only about 25% of mean squared densities.

Our results indicate that though no small number of

![Fig. 7. (a) Three-dimensional cut P(PC1, PC2, PC3|0.8 < PC4 < 1.0) through the 4D GCM histogram. This conditional PDF contains the bimodal slice shown in Fig. 6a. (b) Same as (a), but for a 3D cut through the two-Gaussian mixture model fitted in 4D.](image)

![Fig. 8. Histogram estimates of the 2D angular PDFs embedded in the 3D phase spaces (a) ρ1,2,3, (b) ρ1,2,4, (c) ρ1,3,4, and (d) ρ2,3,4. Superimposed are the locations of the pattern modes Mi, i = 1, . . . 3 taken from the 3D angular PDF depicted in Fig. 9b.](image)
components can completely describe the GCM distribution, in each plane two components are a very good approximation. Figure 11 portrays these individual components together with the resulting mixture PDF. When compared to the AGCM histograms of Fig. 4, it is seen that adding a second Gaussian enables the model to capture the main non-Gaussian features evident in the AGCM PDFs. Not only are the modes and overall skewness modeled well, but also the positions of the three probability ridges in each of the planes containing PC1 are represented. Generally speaking these ridges result because the two component Gaussians have major axes that are at roughly 60° angles to each other so that at one end they overlap to produce a third region of enhanced probability.

b. Four-dimensional synthesis

The two-dimensional Gaussian mixture results demonstrate that much of the structure of our planetary wave distributions can be captured by a few parameters, but we wonder if the distributions can be approximated even more efficiently if all four dimensions are considered simultaneously. This should be true if the Gaussian components we have identified in different planes correspond to common, higher-dimensional Gaussians.

The solid line of Fig. 12 displays the mean cross-validated log-likelihood of Gaussian mixtures with various numbers of components when the EM algorithm is applied to four-dimensional state vectors. As with the two-dimensional results, the more components in the mixture model the more likely the data came from that model. And by far the largest increase in cross-validated likelihood comes when the number of components is increased from one to two. Adding a third component also produces a marked improvement in the model but components beyond that make only marginal contributions.

When we look at the structure of the 4D two-component PDF we find that many of the features of the 2D two-component mixtures in Fig. 11 planes can be traced to it. Figure 13 shows the 4D two-component PDF, and its individual components, when they are projected onto various planes. In planes containing PC1, the two 4D components are sufficient to represent the centroids and covariance structure of each of the two-dimensional components. On the other hand, in the other three planes, there are large departures between the two- and four-dimensional components though both have the property that they produce distributions whose non-Gaussianity is weak compared to the non-Gaussianity of the PC1-containing planes.

Interestingly the 4D two-component mixture model approximates most of the other important features of the AGCM PDFs as well. For example, a slice with PC3 = 2.7 and PC4 = 0.9 (not shown) has two maxima, located in nearly the same positions as those in Fig. 6a. But just as with the AGCM PDF this is not an indication of multiple modes; in the 3D PDF derived from the 4D two-component model by setting PC4 = 0.9 (Fig. 7b), these maxima are actually found on two lobes of an apricot-shaped distribution whose single maximum occurs where PC3 is negative. Likewise, the 4D two-
component distribution corresponds to a distribution of pattern amplitudes that is surprisingly similar to that described in section 4b. When its three-dimensional PDFs are projected on the unit sphere, they (Fig. 14) have all the important features of Fig. 8 including the three local maxima in subspaces that include PC1 and the great circle ridges that connect M2 and M3. Only the single prominent maximum of \( \rho_{2,3,4} \) is missing. This correspondence carries over to patterns in four dimensions as Fig. 9c demonstrates. This plot of a density surface on the hypersphere embedded in four dimensions contains much of the complex structure of its AGCM counterpart in Fig. 9a. For the higher-valued density surface we find three distinct pattern modes.

**Fig. 10.** (a) Mean cross-validated log-likelihoods, \( \mathcal{L}^{(cv)} \), of a mixture model of \( k = 1, \ldots, 8 \) Gaussians fitted to the two-dimensional PDFs in all planes \( \rho_{i,j} \), spanned by the leading four EOFs. (b) RMSE between mixture model and GCM histograms.

**Fig. 11.** Two-Gaussian mixture PDFs in the planes spanned by the leading four EOFs. Contour intervals are 0.02 starting with the 0.02 contour. The means (filled circles) and the 0.02 contour (shading) of the two component Gaussians are superimposed. The RMSE between mixture model PDFs and GCM histograms is denoted in the right upper corner.
that are in similar locations as those of the GCM PDFs in Fig. 9b. When these modes are projected onto the 2D-coordinate planes (Fig. 13), we see that the three probability ridges of the 4D mixture model correspond remarkably well to those in the PDF of the GCM (Fig. 4).

6. Summary, discussion, and conclusions

We set out to see if there is evidence for non-Gaussianity and perhaps even multimodality in the probability density distribution of planetary waves in a 14-million-day integration of an AGCM. We did this in spite of the failure of previous studies to find such features in nature with statistical significance, because we knew the previously reported (BB05) nonlinearities in the AGCM’s mean phase space tendencies had the potential to produce non-Gaussianity. Using a simple stochastic model we found that these nonlinearities by themselves can produce substantial PDF distortions to the Gaussianity of linear dynamics. Therefore, we reasoned that past failures to confirm non-Gaussianity in more than one dimension might be a consequence of nature’s short sample and not necessarily an indication it truly has Gaussian distributed planetary waves. Furthermore, we expected that by being able to analyze PDFs in as many as four dimensions we would reduce the masking of non-Gaussian features that occurs when projections onto lower-dimensional spaces are used.

To gain insight into the four-dimensional PDF structure, we assessed its non-Gaussianity in various directions, planes, and subspaces. As measured by several quantities, we found indications of non-Gaussianity in all phase subspaces spanned by the four patterns but some subspaces were more non-Gaussian than others. We were able to evaluate these contrasting degrees of Gaussianities with great confidence because of our large dataset. Notably, all phase spaces containing EOF1 were characterized by highly non-Gaussian PDFs, while phase spaces spanned by higher EOFs were much more Gaussian. This contrasting behavior was anticipated by demonstrations in the introduction and in Berner (2005) that tendencies like those in Fig. 1, with their distinct nonlinearities in planes containing EOF1 have a substantial impact on the PDFs of planetary wave states. Interestingly, our EOF1 resembles the North American Annular Mode (Thompson and Wallace 1998), which is structurally similar to states that observational studies (e.g., Kimoto and Ghil 1993a; Corti et al., 1999) have suggested may correspond to local density maxima in nature. The most prominent non-Gaussian features in the various subspaces we considered either corresponded to a shift in the mode away from the climatological mean or in probability ridges radiating away from the mode.

Although we found statistically significant non-Gaussianities, we found no evidence to support the idea of multiple modes in the sense of multiple local density maxima even when considering the full four-dimensional space. This is consistent with the results of Hsu and Zwiers (2001), who reported almost no evidence for multiple modes in lengthy AGCM simulations, and Stephenson et al. (2004), who could not establish that observed PDFs departed from a multinormal null hypothesis. Of course our results do not exclude the possibility that there might exist multiple distinct maxima in still higher-dimensional spaces.

Although we could not find any evidence for multimodality in standard PDFs, two modified approaches yielded multiple local maxima. One way to obtain statistically robust PDFs with multiple modes was to discard amplitude information and find the distribution of atmospheric circulation patterns. When we projected three-dimensional states onto the unit sphere we found three distinct local maxima, just as Crommelin (2004) and Selten and Branstator (2004) found distinct pattern maxima in nature and in a quasigeostrophic model, respectively. The other approach that led to multiple density maxima was to investigate conditional PDFs obtained by making two-dimensional slices through our 4D PDF. For some two-dimensional slices the conditional PDFs have two distinct local maxima, but in the

![Fig. 12. Mean cross-validated log-likelihoods, $T^{(k)}_{cv}$, of a mixture model of $k = 1, \ldots, 8$ Gaussians fitted to the four-dimensional PDF of 12-hourly (solid) and 30-day-mean (dashed) data in the subspace spanned by the leading four EOFs. A constant was added to the log-likelihood of the 30-day means for all $k$, so that $T^{(k)}_{cv}$ is the same for 12-hourly and 30-day-mean data.](image-url)
full four-dimensional space these features do not turn out to be local maxima.

Our result that we can find multiple density maxima only when we consider conditional samples or amplitude-independent data may make one wonder whether our use of unfiltered states from an AGCM may have led to this finding. After all, some reports of prominent multimodality have concerned lowpass data. But we have repeated much of our study with 30 day means and find no qualitative differences. Figure 15, for example, shows PDFs of 30-day means in the same coordinates used in Fig. 4. The radial ridges are more prominent than for instantaneous states, but there is only a single mode in each plane. And the utility of a two-component mixture carries over to the distribution of 30-day mean states as suggested by the Fig. 12 dashed line, which depicts the log-likelihood of Gaussian mixtures fit to 30-day means. In additional work that we only have space to briefly refer to here, we have found that the discrepancy between our lack of multiple density maxima and the reporting of such multiple maxima in observational studies is likely to have sampling as its explanation. For when we consider PDFs of AGCM data in segments of the same length as the observational record then multiple maxima do often occur.

Though multiple modes did not turn out to be a useful avenue for characterizing the non-Gaussianities in our planetary wave data, we did find that Gaussian mixture modeling was an effective framework for doing this. To be sure, the complete structure of two- and four-dimensional PDFs could not be represented by a small number of Gaussian components, but most of the important features could be approximated by just two components. These features include the strong planar dependence of PDF structure, the placement of the mode and ridges of probability radiating from that mode, and the location of multiple density maxima in certain two-dimensional slices and in angular PDFs that represent the density of patterns rather than states.

Given these similarities, we believe it is useful to think of the PDFs of our Northern Hemisphere-dominated states as being approximately a mixture of two Gaussians. Beyond its similarity to the AGCM PDF, this distribution is attractive for two reasons. First, it is consistent with the approximately piecewise-linear mean tendencies reported by BB05, whose care-

**Fig. 13.** Projections of a two-Gaussian mixture model fitted in 4D onto the two-dimensional coordinate planes. Contour intervals are 0.02 starting with the 0.02 contour. The means (×) and the 0.02 contour (shading) of the two projected component Gaussians are superimposed. Filled circles denote the component means of the two-Gaussian mixture models fitted in 2D (Fig. 11). The lines show the directions of the three pattern modes taken from the 3D angular PDF of the 4D mixture model (see text).
Fig. 14. Histogram estimates of the 2D angular PDFs of the 4D two-component mixture model embedded in the 3D-phase spaces (a) $\varphi_{1,2,3}$, (b) $\varphi_{1,2,4}$, (c) $\varphi_{1,3,4}$, and (d) $\varphi_{2,3,4}$. Superimposed are the locations of the pattern modes $M_i$, $i = 1, \ldots, 3$ taken from the 3D angular PDF of the GCM depicted in Figs. 8 and 9b.

Fig. 15. Two-dimensional histograms of 30-day means in the planes spanned by the leading four principal components. Contour intervals are 0.02 starting with the 0.02 contour. Measures of non-Gaussianity are denoted in the panel corners. RMSD and Pearson statistic are given in units of $10^{-1}$ and $10^4$, respectively.
ful analysis of the phase space trajectories revealed two piecewise-linear mean tendencies much like those conceptualized in Fig. 2b. As the example in the introduction demonstrated those tendencies are consistent with a distribution dominated by two Gaussian components. Furthermore, when we plot states at the centroids of the two Gaussian components of our four-dimensional analysis (Fig. 16), we find they correspond to a Pacific zonal and a Pacific blocked state. These are the same states, with somewhat weaker amplitudes, that BB05 showed were at the centers of nonlinear features in the mean tendency fields. The position of these dynamical features is marked by stars on Fig. 1. Their existence deforms the regions of slow mean tendencies into a V in $\varphi_{1,3}$ and $\varphi_{1,4}$, which the introductory example suggests leads to the two Gaussian components being centered near the centers of the nonlinearities and being aligned at approximately 30° angles to the ordinate. Interestingly the structure of these two states is reminiscent of the multiple equilibria first proposed by Charney and DeVore (1979) as being responsible for weather regimes in that they correspond to zonal and blocked states. Of course it will take further analysis to determine whether the dynamical processes responsible for the non-Gaussianities we have found and the nonlinear features reported in BB05 are similar to the processes in the Charney–DeVore model.

A second attraction of the two-Gaussian mixture is that its simplicity helps in the interpretation of certain PDF features. For example, as we have mentioned, though many studies propose that the distribution of planetary wave states is influenced by the existence of special preferred states others have found evidence of preferred patterns. Furthermore these two approaches
often arrive at a different number of preferred circulations. The two-Gaussian mixture approximation helps unravel these contrasting views. As Figs. 11 and 13 show, because of the orientation of the major axes of the two component Gaussians in $p_{1,3}$ and $p_{1,4}$, three ridges of enhanced probability form, even though there are only two preferred states. Hence in analyses that consider patterns, it appears there are three local maxima and thus three preferred patterns, when in fact the more fundamental organizing influence is the presence of two preferred states.

In summary, our findings, together with those in BB05, indicate that though the traditional linear stochastic model of planetary waves captures a great deal of the state distribution of the AGCM we have analyzed (namely the multivariate Gaussian component), nonlinear stochastic models (e.g., Majda et al. 2003; Berner 2005) appear to be needed to capture the non-Gaussian features we have found. Even a model as simple as a two-segment piecewise-linear stochastic model with additive noise and its resulting two-Gaussian mixture of states seems to be a useful improvement when approximating and conceptualizing the behavior of planetary waves.

Acknowledgments. The authors thank C. Tebaldi for numerous beneficial discussions and A. Mai for help with technical issues. Bo Christiansen and two anonymous reviewers provided helpful suggestions and comments. This study was partially funded by NOAA Grant NA17GP1376 and NASA Grant S-44809-G. The support of the first author by NCAR’s Advanced Study Program is gratefully acknowledged.

APPENDIX A

Simple Stochastic Model

The deterministic drifts $h_i(x)$ inserted into the stochastic model (1) to produce the simple examples displayed in Fig. 2 are composed of three linear components. The first component is defined over the entire two-dimensional domain and consists of Rayleigh damping with a time scale of $1/(36d)$ in the $x_1$ direction and $1/(13d)$ in the $x_2$ direction. These values are chosen to be similar to the damping time scales of PC1 and PC4 in CCM0 reported by Berner (2005). Figure 2a displays this component and Fig. 2d shows the PDF of the resulting solution to (1).

The other two components consist of segments of undamped oscillations chosen to be similar to the oppositely rotating features in Fig. 1c. One of these oscillations is centered at $(-1, 0)$ and the other at $(0, 1)$. Each has a period of 75 days and consists of elliptical orbits whose major axes are parallel to the $x_1$ axis and are 4 times as long as the minor axes. The component centered at $(1, 0)$ rotates clockwise and is set to zero for $x_1 > 0$, while the component centered at $(0, 1)$ rotates counterclockwise and is zero for $x_1 \leq 0$. Their sum is shown Fig. 2b. The PDF in Fig. 2e results from solving (1) with $h_i(x)$ set to the sum of all three components, which is displayed in Fig. 2c.

APPENDIX B

EM Algorithm

When applying the EM algorithm, we have found it is important to take several factors into account. First, the algorithm is designed to iteratively find the set of parameters $a$, $\mu$, and $C$ that maximize $L$, but $L$ is strictly defined only for data that are not temporally correlated. Thus following Hannachi and O’Neill’s (2001) suggestion, we apply it to a reduced dataset consisting of every 20th observation (every 10th day) of our data. This spacing is motivated by the decorrelation times of the PCs (Berner 2005). Second, we find the EM results to be sensitive to the initial guess used in the iterative procedure. Hence for each dataset of interest, we find 100 EM solutions, each starting from a different initial guess, and use the solution with largest $L$ as the best fit model. Generation of these guesses must also be done with care to ensure that parameter space is well searched. We find that drawing initial parameters from uniform distributions between $-1.5$ and $1.5$ will find maxima in the log-likelihood function that $k$-means preclustering—as suggested by Smyth et al. (1999)—is unable to find. Third, as for any procedure that fits models to finite samples, we find we must take into account that any given sample will not be completely representative of the true underlying distribution. We address the effects of sampling by partitioning our complete $7 \times 10^6$ day sample into two halves, solving for the parameters $a$, $\mu$, and $C$ by applying the EM algorithm to one-half and then evaluating $L$ by applying (7) to the other half. To reduce the chance that a particular partitioning is unrepresentative we repeat this procedure ten times each for a different partitioning and consider the average of these ten likelihoods as the likelihood of a $k$-component Gaussian mixture for the complete dataset, what we refer to as the mean cross-validated log-likelihood. As one of the reviewer points out, in general it might be favorable to use more than ten partitions, but we are confident that we have used enough partitions, since the cross-validated log-likelihood functions for the different partitions are very similar. This is
most likely caused by the large sample size, so that the features in each partition are very similar.

APPENDIX C

Bump-Hunting

By bump-hunting we mean the search for local maxima in the $d$-dimensional PDF of atmospheric states by an iterative algorithm that uses gradient information in some form to find the directions of maximum probability variation. The challenge in bump-hunting lies in the necessity to find a representation of the PDFs of sufficient smoothness that the maxima found by the algorithm are not the result of insufficient sampling. One option is to employ kernel density estimation, as done by Kimoto and Ghil (1993a). The approach we have used is to approximate our AGCM PDFs by an eight-Gaussian mixture, since the RMSE between the eight-Gaussian mixture and the AGCM PDFs is very small. This approach has the advantage that the derivatives of the PDFs are known analytically, which allows us to use the quasi-Newton algorithm E04KYF, which is part of the NAG-libraries, to locate maxima. We initialize the algorithm from various initial states on a grid of locations to determine if there is more than one local maximum in a given PDF.

REFERENCES


