Is There Evidence of Multiple Equilibria in Planetary Wave Amplitude Statistics?

GREGOR NITSCHE AND JOHN M. WALLACE

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

CHARLES KOOPERBERG

Department of Statistics, University of Washington, Seattle, Washington

(Manuscript received 5 April 1993, in final form 9 July 1993)

ABSTRACT

Results obtained by Hansen and Sutera concerning the occurrence of bimodal probability density functions (PDFs) in a wave amplitude index (WAI) calculated from large-scale atmospheric flow data are reexamined. The PDFs are found to be highly sensitive to changes in the parameters used to calculate the WAI. The excessive sensitivity is suggestive of an insufficient number of degrees of freedom in the PDFs.

The Monte Carlo test used by Hansen and Sutera to establish the statistical significance of their PDFs is reexamined, with emphasis on their attempt to compensate for the interdependence between neighboring data points in their time series of the WAI. Their random samples contained only one (independent) data point for each 4.5 data points in the WAI time series. It is shown that in order to generate PDFs with the same degree of smoothing as the WAI PDF, they should have simultaneously reduced the smoothing parameter in the maximum penalized likelihood (MPL) algorithm by the same factor. When this scaling factor is properly taken into account, more than half of the randomly generated samples exhibit multimodality: hence, the occurrence of bimodality in the PDFs calculated from the WAI data does not appear to be statistically significant. It is estimated that in order to distinguish between samples drawn from populations with a degree of bimodality comparable to that reported in the WAI data and samples drawn from a Gaussian population, a period of record of at least 150 years would be needed.

1. Introduction

The question of whether large-scale atmospheric circulation possesses multiple equilibria has intrigued many researchers during the last 15 years and has led to a number of publications of both theoretical and observational results. Theoretical results indicating that the atmosphere might possess more than one equilibrium flow regime were obtained by Lorenz (1963). A seminal work, centered around the issue of multiple equilibria for the planetary waves, was Charney and DeVore's (1979) investigation based on a simple one-layer free surface barotropic channel model that included the β effect and a frictional momentum source and that was perturbed by bottom topography and by a barotropic analog of thermal driving. After truncating the model by considering only the first six eigenfunctions of the Laplacian, they were able to show that for certain choices of parameters, the system possesses three equilibria, two of which are stable. One of the stable equilibria corresponds to a strong zonal flow, the other to a weak zonal flow with strong wavelike features, analogous to the blocking patterns frequently observed in the atmosphere.

The applicability of Charney and DeVore's results to the real atmosphere has been disputed in a number of articles. Tung and Rosenthal (1985) found that the parameter ranges for which Charney and DeVore obtained multiple equilibria are not consistent with the real atmosphere; for realistic parameter values, the existence of multiple equilibria cannot be shown. Furthermore, their existence depends on the truncation applied to the model; for less severely truncated models, they showed that no multiple equilibria exist. However, it can be argued that any model for which the existence or nonexistence of multiple equilibria can be proven theoretically is probably too inexact a representation of the real atmosphere to allow conclusions on the existence or nonexistence of these equilibria in the actual large-scale circulation. For this reason, numerous attempts have been made to find evidence of multiple equilibria in observational data (Sutera 1986; Hansen 1986; Hansen and Sutera 1986; Molteni et al. 1988; Hansen 1988; Hansen and Sutera 1988; Mo and Ghil 1988; Molteni et al. 1990) and in the output of general

Corresponding author address: Gregor Nitsche, Department of Atmospheric Sciences, GJ-40, University of Washington, Seattle, WA 98195.

circulation models (GCMs) (Hansen and Sutera 1990, 1992). If such equilibria exist, then they might give rise to multiple peaks or, as they are referred to by statisticians, modes in the probability density functions (PDFs) for certain observable atmospheric variables. Since we expect the two modes to be associated with different intensity in planetary-scale wave activity, which is usually best defined in the middle and upper troposphere, it seems appropriate to analyze 500-mb height data. Furthermore, we expect any effects associated with such waves to be most distinct during winter, which is why most analyses are based on wintertime data.

The high dimensionality of hemispheric 500-mb height data poses a problem concerning the estimation of PDFs. The European Centre for Medium-Range Weather Forecasts (ECMWF) dataset, for example, represents the 500-mb heights using 7500 grid points for each hemisphere, making a PDF estimate for the raw data practically unfeasible (particularly considering the relatively short length of the observational record). However, if the multiple equilibria indeed correspond to different intensities in wave activity, they might be reflected as two modes in the PDF estimate for a one-dimensional variable that somehow measures this intensity.

2. The wave amplitude index

As a measure of the amplitude of the quasi-stationary planetary waves, Sutera (1986) introduced a quantity that he referred to as the wave amplitude index (WAI). The procedure for calculating the WAI from any record of daily hemispheric 500-mb height data is as follows. In the first step, the data for each day are averaged over a prescribed latitude belt so that an average geopotential height as a function of longitude only is obtained. The resulting data are then Fourier decomposed with respect to latitude, yielding one complex number Z_i representing each wavenumber i. The wave amplitude index is calculated by adding up the absolute values of the squared amplitudes of wavenumbers 2 through 4 and then taking the square root:

$$[Z_{2-4}] = \left(\sum_{i=2}^{4} |2Z_i^2|\right)^{1/2}. \qquad (2.1)$$

To remove the spurious influence of the annual cycle and the noise associated with high-frequency variability, the time series of $[Z_{2-4}]$ is then filtered in time; for this purpose, it is Fourier transformed, the Fourier components corresponding to the undesired frequencies are removed, and the signal is resynthesized. (The average value of $[Z_{2-4}]$ is removed as well.) Hansen and Sutera (1986, hereafter referred to as HS86) point out that the removal of the interannual variability from the time series could actually conceal bimodality that might be present in the data; for this reason, the interannual variability was not removed in the present study. The resulting filtered series will be referred to as the WAI. Obviously, the values of the WAI depend on both the latitude belt over which the averaging is carried out and the exact frequency ranges removed by the filtering. The wintertime WAI data obtained this way can now be interpreted as a sample taken from the probability density function (PDF). Using the WAI data we can estimate this PDF. The algorithm used by both HS86 and us to perform this estimation is the maximum penalized likelihood (MPL) algorithm that will be discussed later in more detail.

3. Probability density function estimates for the WAI

Hansen and Sutera (1986) investigated the Northern Hemisphere WAI for the winters 1964/65 through 1979/80. In their study, they averaged geopotential height data over a latitude belt extending from 45° to 70°N and removed the variability in the high-frequency range, corresponding to periods of less than five days, and in the range surrounding the annual cycle, corresponding to periods of 9 to 22 months. In their PDF estimate, two modes are clearly visible.

We first verified that we were actually able to reproduce the time series used by Hansen and Sutera. In this effort, we were aided by A. Hansen, who supplied a number of technical details concerning their calculations of the WAI and provided data for comparison. For our study, we used once-daily 500-mb heights obtained from the U.S. National Meteorological Center (NMC) Final Analysis from 1 January 1946 through 31 May 1991 with missing analyses supplied by linear interpolation. For further details of the preprocessing, see Kushnir and Wallace (1989). The data were linearly interpolated onto a 5° latitude by 8° longitude grid. For an example, we compared our time series segment for the winter 1964/65 with Fig. 11a in HS86. The two time series are shown together in Fig. 1; evidently, they are virtually identical.

We then investigated the dependence of the PDF estimate on the latitude belts over which the data are averaged in computing the index, on the frequencies that are removed to eliminate the annual cycle, and on the period of record on which the estimates are based. A variety of different latitude belts and frequency bands have been used in previous studies: as noted above, in HS86 a 45° to 70°N latitude belt was used and the frequency band corresponding to periods from 9 to 22 months was filtered out. Hansen (1986), by comparison, averaged over the 20° to 80° latitude belt and removed all signals with periods greater than six months. The period of record 1964–80 was used by HS86; periods of record spanning parts of the 1980s have been used in the studies of Hansen (1986, 1988). Using the

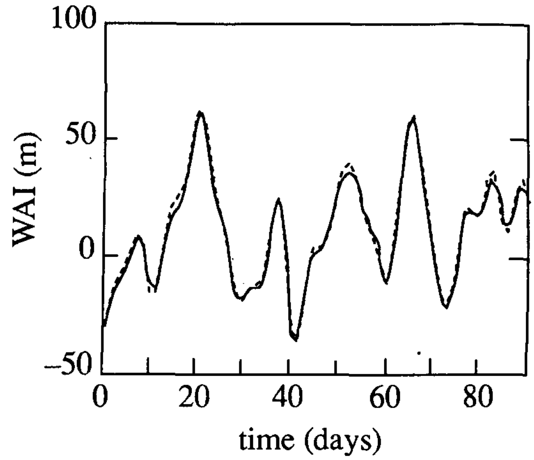

FIG. 1. Wave amplitude indicator for the winter 1964/65. Horizontal axis: time in days (day 1 = 1 December 1964). Vertical axis: WAI in m. The data shown by the dashed line were adapted graphically from Fig. 11a of Hansen and Sutera (1986); the data shown by the solid line were obtained by us, using the same parameters.

same latitude belt and the same temporal filtering parameters as HS86, we obtain the PDF estimate shown in Fig. 2a. As expected, it possesses two modes with maxima at about the same values as found in HS86.

Our PDF estimates for different combinations of parameters are summarized in Fig. 2. The shaded graphs indicate PDF estimates that obviously exhibit two or more modes, where modes occurring in the tails of the PDFs (with maximum values less than one-half the main maximum) are not counted. As the figure shows, the occurrence or nonoccurrence of bimodality depends sensitively on all the parameters mentioned, and no systematic dependence can be discerned unambiguously. In general, it seems that the PDFs for the longer periods of record exhibit less pronounced departures from a smooth, unimodal shape than those for the shorter records. Multimodal distributions, when they occur, exhibit a variety of shapes.

4. Monte Carlo test for the statistical significance of bimodal spectra

The fact that the PDFs exhibit so much sensitivity to parameter changes raises the question of whether the PDF estimates presented here are actually statistically significant, or whether the multimodality that can be found in some of them is merely a result of insufficient sample size. To show that the bimodal density they observed in the 1964 to 1980 WAI data could not be interpreted as resulting from the sampling variability of an underlying unimodal PDF, Hansen and Sutera (1986) used a Monte Carlo test. For this purpose, they drew samples from a similar unimodal density (which was obtained from the 1964 to 1980 WAI data by applying more smoothing). The sample size was chosen equal to the number of independent observations in the 1964 to 1980 record; according to their estimate, one winter contributes roughly 20 independent observations so that the appropriate sample size is 320. MPL density estimates were then made for each of these samples. Out of 100 samples, only 14 showed any evidence of bimodality; following HS86, this leads to the conclusion that the true WAI PDF is most likely bimodal.

Comparing the first plot in Fig. 5b of HS86, which shows the PDF estimate for the 1964 to 1980 WAI data, with their PDF estimates for random samples shown in their Fig. 5a, one may notice that the latter appear in general smoother than the former. This raises the question of whether the results of the Monte Carlo simulation are actually comparable with the 1964–80 PDF estimate, or whether adjustments have to be made in the PDF estimation algorithm. To address this question, we first need to discuss this algorithm.

The MPL algorithm used by HS86 is discussed by Scott et al. (1977) and Good and Gaskins (1980). [For an overview of density estimation methods, see Silverman (1986).]

To motivate this algorithm, we first consider the log-likelihood function for an independent sample $\{x_i\}$, $i = 1, \ldots, N$,

$$\log L(f; \{x_i\}) = \sum_{i=1}^{n} \log f(x_i). \tag{4.1}$$

Here $L(f)$ measures the likelihood that a sample $\{x_i\}$ was generated from the PDF $f(x)$. One could try to maximize $\log L(f)$ with respect to f in order to obtain a "most likely" density; however, the maximization of this expression (under the constraint that $\int f(x)dx = 1$) would be achieved for an f that is a sum of Dirac δ functions at each data point. To obtain a more likely smooth density estimate, we add a roughness penalty term to the log-likelihood $\log L(f)$,

$$\log L_p(f) = \sum_{i=1}^{n} \log f(x_i) - \alpha\left[\int \left(\frac{d^2 f(x)}{dx^2}\right)^2 dx\right] \tag{4.2}$$

with $\alpha > 0$. Obviously, the penalty term will be close to zero for smooth density estimates and it will assume very large negative values for ragged PDF estimates. The degree of smoothness of the PDF estimate depends on the value of the parameter α that is used. It can be shown that when α is very large, the density estimate is smooth and unimodal whereas for sufficiently small α, the estimate will generally possess as many peaks as there are different points in the sample. Since we are concerned with the detection of bimodality, the choice of α is of critical importance: if we choose α small enough, the density estimate will certainly be multimodal, whereas if α is large enough, the density esti-

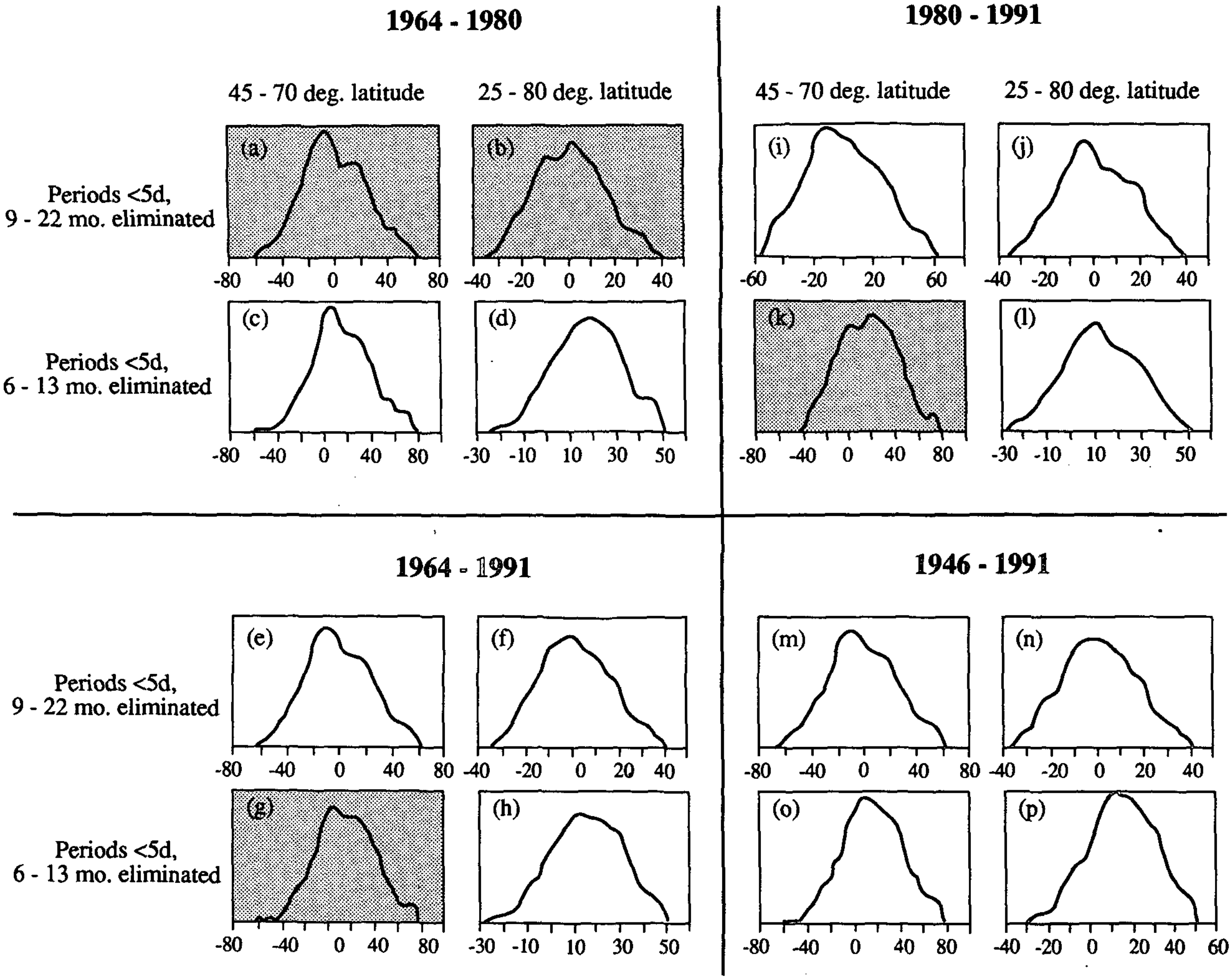

FIG. 2. PDF estimates for the WAI, using different periods of record, latitude bands, and time filters. Horizontal axis: WAI in m. Vertical axes: probability density. Shaded graphs indicate multimodal estimates.

mate will always be unimodal! Of course, we do not know a priori the "correct" value for α; however, we want to make a choice that at least ensures that PDF estimates for different datasets will be comparable with regard to the effective importance of the penalty term. It can be shown (see appendix A) that if the data are scaled by a factor c, α has to be scaled as c^5 in order to produce a scaled version of the same density estimate. We therefore normalize α by σ^5, where σ^2 is the sample variance of the data, and define $\tilde{\alpha} \equiv \alpha/\sigma^5$, which we will refer to as the equivalent penalty weight.[1]

Although in our situation the data are obviously not independent and therefore (4.1) is no longer correct, it is still reasonable to estimate the density by minimizing $\log L_p(f)$. For the results shown in Figs. 2 and 3, we have used a value of $\tilde{\alpha} = 1.0$; in the case of Fig. 2a, this corresponds to $\alpha = 7.75 \times 10^6$, which is in between the two values used in HS86.

When HS86 performed their Monte Carlo test, they used a value of $\alpha = 10^7$, which was the largest value for which they obtained a detectable bimodality in the 1964–80 WAI data. Now let us consider the question: Is it appropriate to use the same value of α for the samples of 320 independent data as for the original WAI dataset that contains 1440 dependent data points equivalent to 320 degrees of freedom, or should the different degrees of dependence in the samples be taken into account by appropriately scaling α? Formally, if

[1] HS86 do not use this normalization; since they do not compare estimates for samples with different variance, it does not appear necessary in their case.

we have 320 independent observations $y_1, \ldots, y_{320}$ replacing the 1440 dependent ones, expression (4.2) becomes

$$\log L_p(f) = \sum_{i=1}^{320} \log f(y_i) - \alpha\left[\int \left(\frac{d^2 f(x)}{dx^2}\right)^2 dx\right]. \quad (4.3)$$

Evidently, since we are only adding 320 terms instead of 1440 terms, while $f(x)$ is approximately the same, we will have to rescale $\tilde{\alpha}$ by dividing it by the factor $1440/320 = 4.5$.

As a heuristic argument in support of this scaling law, let us consider a similar (though not equivalent) situation. We start out with the PDF estimate for the 1964–80 WAI data. In the first step, we sort the sample points in ascending order and divide them into groups, each consisting of four consecutive points. Then we replace the value of each point in the sample by the average value of the group to which it has been assigned. This sample (which we will refer to as S2) still contains 1440 data points; however, there are obviously not more than $1440/4 = 360$ degrees of freedom in this sample because it consists of 360 groups of four identical values. Yet we do not expect that we have deleted much information from the sample by using this procedure (since the data were not independently sampled in the first place), and indeed, a PDF estimate for the new sample, using $\tilde{\alpha} = 1.0$, is virtually identical to the one for the original data obtained with the same $\tilde{\alpha}$ (Fig. 3a).

We now remove three out of the four identical data points from S2, creating a sample S3 that consists of 360 data points, but still possesses as many degrees of freedom as S2 (because the operation that created S3 from S2 obviously did not remove any information from the data). If we now perform a PDF estimate, again using $\tilde{\alpha} = 1.0$, we obtain an estimate that is more smoothed than the one obtained for S2 (or the original data) (Fig. 3b). This fact is easy to understand if we consider once again the penalized log-likelihood function that is minimized in the MPL algorithm [Eq. (4.2)]. Assume we had obtained an MPL estimate for the sample S2. If we now use the same α to obtain an estimate for S3, the first term on the right-hand side will become smaller by a factor of 4 (since every data point is now only considered once instead of four times). The penalty term, on the other hand, does not depend directly on the sample points and is therefore not going to change. It follows that the penalty term will become four times more important than in the S2 estimate, leading to stronger smoothing. However, we can recover the S2 estimate from S3 data if we correspondingly decrease α by a factor of 4. In this case, the penalized likelihood function for S3 is just proportional to what it was for S2 and hence is minimized by the same f, as is illustrated in Fig. 3c. If we want to compare PDF estimates for samples with the same number of degrees of freedom, but different numbers of sample points, we should therefore scale α (or $\tilde{\alpha}$) to make it directly proportional to the number of sample points.

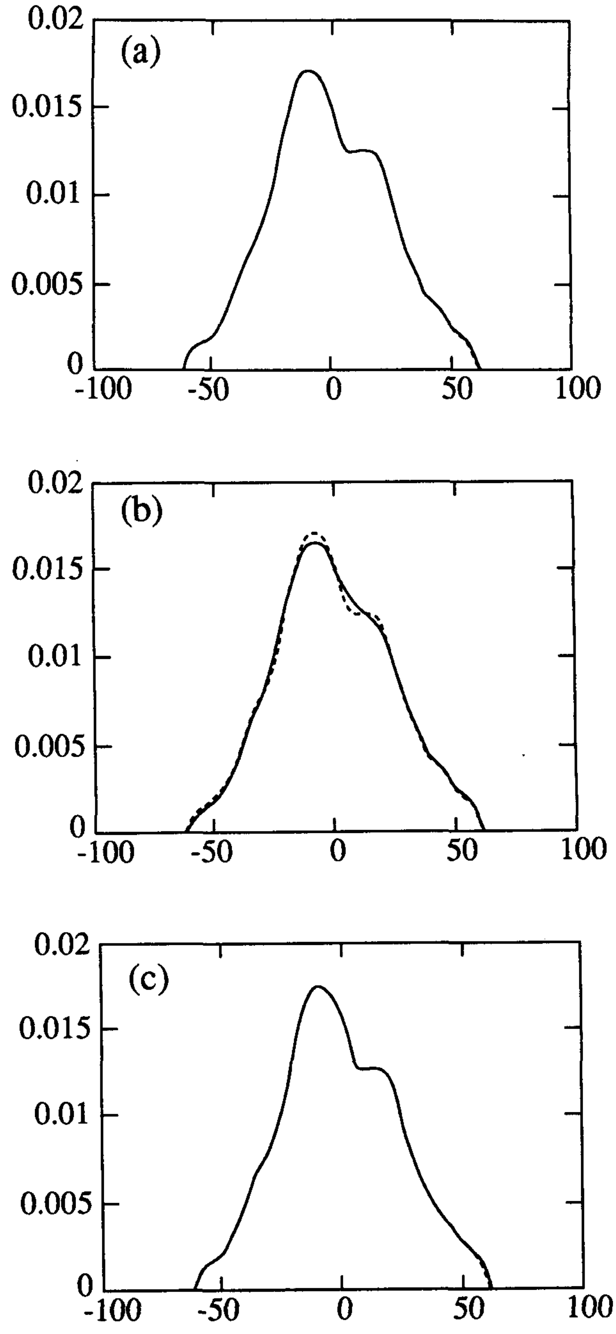

FIG. 3. PDF estimates for grouped and averaged sample points. Horizontal axes: WAI in m. Vertical axes: Probability density. (a) All data retained, $\tilde{\alpha} = 1.0$. (b) Only one data point from each group is considered, $\tilde{\alpha} = 1.0$. (c) Only one data point from each group is considered, $\tilde{\alpha} = 0.25$. Dashed lines in all three graphs represent the estimate obtained from the unmanipulated dataset for comparison.

We repeated the bootstrap test suggested by HS86 (section 4b). Since we wished to examine a much larger number of Monte Carlo realizations than they did, we replaced their subjective criterion for classifying PDFs as unimodal or multimodal by the automated algorithm described in appendix B. In general, results of such tests will obviously depend on the choice of criteria. However, the testing scheme can be "tuned" to increase or decrease the incidence of multimodality in the random samples by lowering or raising the value of $\tilde{\alpha}$. We drew 500 samples, consisting of 320 realizations each, from both a Gaussian PDF (Fig. A3.1d) and a skewed unimodal PDF that was obtained by applying the MPL algorithm on the WAI 1964–1980 data with $\tilde{\alpha} = 10$ (Fig. A3.1c). The testing scheme was tuned by setting our $\tilde{\alpha} = 1.67$: the value that yielded the same percentage of occurrence of multimodal PDF estimates (14%) as HS86 obtained using their subjective criterion.

We then performed our actual test, taking into account the scaling considerations discussed earlier in this section. To account for the fact that the samples are viewed as representing independent points 4.5 days apart, an equivalent penalty weight $\tilde{\alpha} = 1.67/4.5 = 0.371$ was used. In this case, 49% (245 out of 500) of the estimates for Gaussian samples and 69% (345 out of 500) of the estimates for the skewed unimodal samples were found to be multimodal. We do not claim that HS86 would have obtained exactly the same rate of incidence of multimodality based on their subjective criterion, had they lowered their α by a corresponding amount, but they certainly would have obtained a much higher rate of incidence than the 14% that they reported. The observation of bimodality in 16-winter samples of WAI data thus does not support the conclusion that the WAI PDF actually possesses two modes.

5. Conclusions

The results presented seem to indicate that the bimodality reported in the WAI PDF is not statistically significant. However, this does not imply that we can say with certainty that the true PDF possesses only one mode; in fact, the results presented do not support either conclusion with even moderate statistical significance.

To investigate whether statistically significant results concerning the existence or nonexistence of bimodality could be obtained by means of the PDF algorithm using different values of $\tilde{\alpha}$, a more comprehensive Monte Carlo simulation was performed (see appendix C). It turns out that in order to distinguish between a weakly bimodal PDF (as shown in Fig. 2a) and a unimodal one (as shown in Fig. C1c) with even moderate confidence, a minimum of 3000 independent sample points is needed; if we assume that the estimate of 20 independent sample points per winter is valid, it follows that we would need at least a 150-year record of data. For this reason, a definitive resolution of whether the actual WAI PDF is bimodal cannot be expected in the near future.

Finally, it should be pointed out that the existence of multiple equilibria in the large-scale general circulation is not necessarily coupled to the occurrence of bimodality in the WAI PDF. First of all, we do not know whether the WAI is the most appropriate variable to examine in search of multiple equilibria (even though it seems a plausible candidate). An alternative would be indexes related to the patterns that have been identified in cluster analysis as in the works of Kimoto and Ghil (1993) and Cheng and Wallace (1993). Also, using the theory of nonlinear dynamic systems, it can be shown that for noise-driven systems (and in treating the large-scale circulation as a separate system, we implicitly treat it as driven by noise corresponding to smaller-scale effects), multiple equilibria are not necessarily reflected in bimodality in the PDF for any observable quantity. For example, if the amplitude of the driving noise were too high, it could conceal the existence of the two separate equilibria, and if it were too low, the system might not be able to overcome the potential maximum between the equilibria.[2]

Acknowledgments. This work was supported by the National Science Foundation through the Climate Dynamics Program Office under Grant 92/55/2. One of us (GN) was supported in part by the German Academic Exchange Service.

APPENDIX A

Scaling of α with the Standard Deviation of the Sample

Consider a sample $\{x_i\}$ with unit standard deviation; for a given α, we calculate a PDF estimate f by minimizing Eq. (4.2), which is repeated here for reference:

$$\log L_p(f) = \sum_{i=1}^{n} \log f(x_i) - \alpha\left[\int \left(\frac{d^2 f(x)}{dx^2}\right)^2 dx\right]. \tag{A.1}$$

Now let us assume we scale the sample by multiplying each sample point by an arbitrary factor σ. The scaled version of f would then be given by

$$f_\sigma(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right), \tag{A.2}$$

where the factor $1/\sigma$ is included to maintain the normalization. Forming the second derivative yields

[2] As a reviewer pointed out, a different question that cannot be addressed by the methods used in this study is the nature of the atmospheric flow if the background parameters were to change slightly, e.g., due to volcanic eruptions or changes in the earth's orbital parameters. It is conceivable that equilibria corresponding to quite different flow regimes might exist within the range of naturally occurring variations of these parameters.

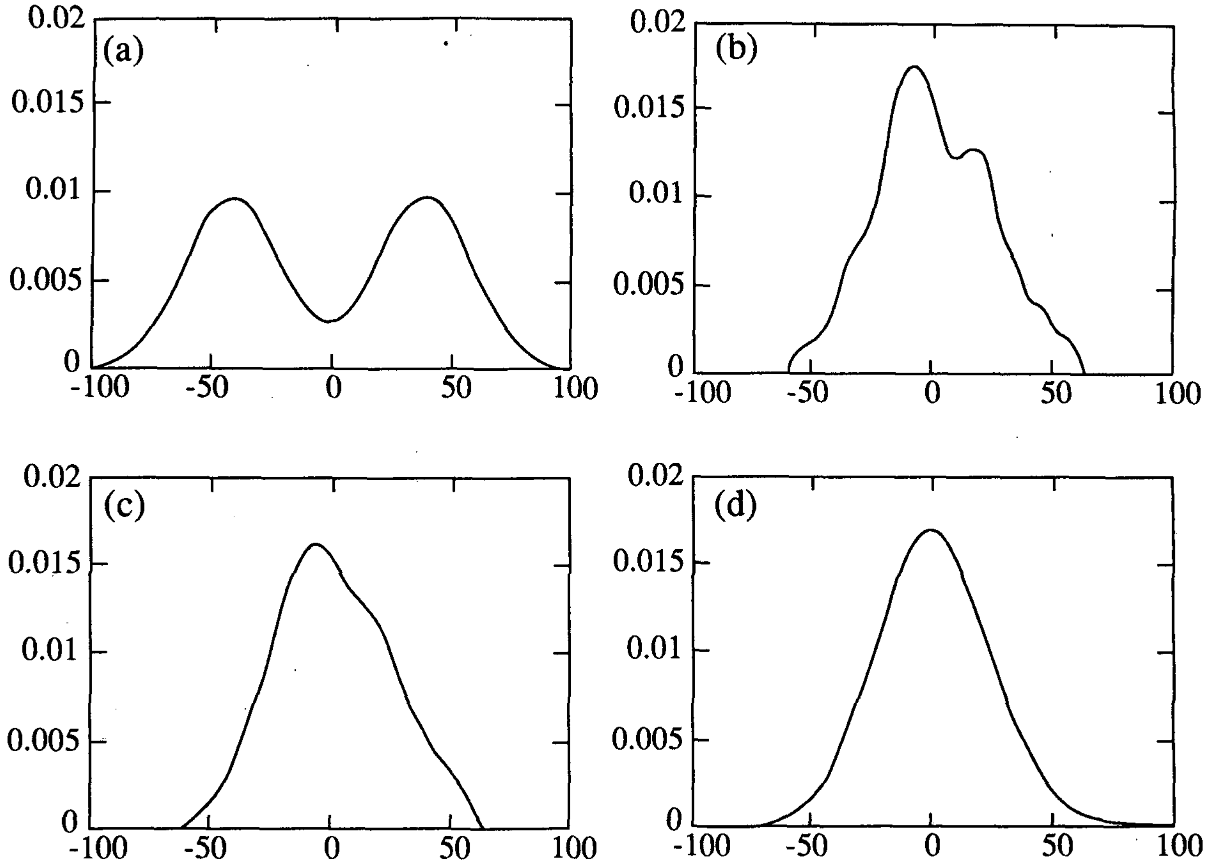

FIG. C1. The four PDFs used for the Monte Carlo test described in appendix C. Horizontal axes: WAI in m. Vertical axes: probability density.

$$\left(\frac{d^2 f_\sigma(x)}{dx^2}\right)^2 = \frac{1}{\sigma^6}\left(\frac{d^2 f(u)}{du^2}\right)^2 \quad (\text{where } u = x/\sigma). \tag{A.3}$$

Integrating, we obtain

$$\int \left(\frac{d^2 f_\sigma(x)}{dx^2}\right)^2 dx = \frac{1}{\sigma^5}\int \left(\frac{d^2 f(u)}{du^2}\right)^2 du. \tag{A.4}$$

At the same time, the first term on the rhs of (A.1) does not need to be scaled when each sample point is multiplied by σ because

$$\sum_{i=1}^{n} \log f_\sigma(\sigma x_i) = \sum_{i=1}^{n} \log f(x_i) - n \log\sigma \tag{A.5}$$

and the constant term $- n \log\sigma$ does not influence the maximization. Therefore, it is necessary to scale α with σ^5 to obtain f_σ as the density estimate. In order to quantify the degree of smoothing that one is applying in calculating the PDF estimate independently of the sample's standard deviation, we define the *equivalent penalty weight* $\tilde{\alpha}$ by letting $\alpha \equiv \sigma^5\tilde{\alpha}$, where σ is the standard deviation of the sample.

APPENDIX B

Detection of Multimodality

In classifying a PDF estimate as multimodal in the computerized evaluation of the Monte Carlo experiment, we applied the following criteria: in the form of the MPL algorithm that we used, the PDF is approximated by its values at 101 equidistant mesh points that are linearly interpolated.

- A PDF estimate was counted as multimodal if it possessed two or more distinct maxima.
- Maxima were regarded as distinct if they were ten mesh points, equivalent to one-tenth the total width of the PDF, apart.
- For a maximum to be counted, the values on either side had to be decreasing over at least two neighboring mesh points.
- Maxima occurring in the tails of the PDF with

values of less than one-half the value of the main maximum were not counted.

APPENDIX C

A Monte Carlo Test for the Ability of the MPL Algorithm to Distinguish between Unimodal and Bimodal PDFs

The Monte Carlo test in section 4 has shown that the application of the MPL algorithm with a penalty weight $\tilde{\alpha} = 1.0$ (corresponding to $\tilde{\alpha} = 0.22$ for independently sampled data) leads in most cases to bimodal PDFs even if the PDF from which the sample was taken is unimodal. In contrast, excessively large values of $\tilde{\alpha}$ will certainly lead to unimodal PDF estimates even if the sample is taken from a bimodal PDF. Therefore, we designed a Monte Carlo test to find out if there is *any* value of $\tilde{\alpha}$ for which the algorithm can reliably distinguish between samples from unimodal and bimodal PDFs. Obviously, the answer to this question will depend on the sample size; we therefore considered several different sample sizes. We proceeded as follows. We drew 200 samples each out of four different PDFs shown in Fig. C1. The first one is a distinctly bimodal PDF constructed by adding two Gaussian PDFs. The second is a PDF estimate for the 1964–1980 WAI with the parameters from HS86, but an even smaller value of $\tilde{\alpha}$ ($\tilde{\alpha} = 0.5$) than used for Fig. 2a, resulting in a more distinct second mode. The third one is a more strongly smoothed PDF for the same sample ($\tilde{\alpha} = 10.0$). The fourth one is a Gaussian.

Table C1 shows our results. Even for samples with 320 degrees of freedom, the algorithm distinguishes well between the strongly bimodal PDF1 and the others if values of $\tilde{\alpha}$ between 2.0 and 5.0 are used. These values correspond to values of $\tilde{\alpha}$ between 9.0 and 22.5 for the dependent 1964–1980 WAI data; the corresponding PDF estimates for the WAI data are found to be smooth and unimodal (not shown) so that the hypothesis that the WAI PDF is strongly bimodal can be rejected.

The situation is different for samples from the other three PDFs. It turns out that for sample sizes up to 1000, the algorithm cannot reliably distinguish between them. For example, 35% (70 out of 200) of Gaussian samples of 320 independent points yield bimodal PDF estimates using $\tilde{\alpha} = 0.5$; however, if we obtain a unimodal PDF estimate for another 320-point sample with the same value of $\tilde{\alpha}$, we cannot be sure that the underlying PDF is actually unimodal, since roughly 35% of the bimodal PDF samples (71 out of 200) yield unimodal estimates. For a sample of 3000 independent points, in contrast, the unimodal and bimodal PDFs can be distinguished with relatively high significance if values of $\tilde{\alpha}$ between 0.5 and 1.0 are used. If we assume, following HS86, that one winter contributes roughly 20 degrees of freedom to the WAI sample, 3000 independent points correspond to a 150-year record, which means that within the foreseeable future, the observational record will not be long enough to support reliable assessments of bimodality using WAI.

TABLE C1. Number of multimodal PDF estimates (out of 200 each) from samples of various sizes drawn out of the PDFs shown in Fig. A3.1a–d. Samples marked with an asterisk (*) were not evaluated.

$\tilde{\alpha}$	(a) Bimodal PDF1	(b) Bimodal PDF2	(c) Unimodal PDF	(d) Gaussian PDF
		Sample size = 320		
0.1	180	192	193	169
0.25	199	169	162	123
0.5	200	129	106	70
1.0	199	69	50	30
2.0	200	19	16	7
5.0	200	1	1	1
		Sample size = 1000		
0.1	188	197	197	172
0.25	198	176	169	127
0.5	200	152	120	72
1.0	200	120	59	35
2.0	200	56	26	7
5.0	200	5	2	0
		Sample size = 3000		
0.1	*	194	157	81
0.25	*	193	85	39
0.5	*	180	35	16
1.0	*	163	6	6
2.0	*	109	0	2
5.0	*	17	0	1
		Sample size = 10 000		
0.1	*	186	77	11
0.25	*	193	14	4
0.5	*	197	2	1
1.0	*	198	0	0
2.0	*	199	0	0
5.0	*	178	0	0

REFERENCES

Charney, J. G., and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216.

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

Good, I. J., and R. A. Gaskins, 1980: Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *J. Amer. Statist. Assoc.*, **75**, 42–56.

Hansen, A. R., 1986: Observational characteristics of atmospheric planetary waves with bimodal amplitude distributions. *Advances in Geophysics*, Vol. 29, Academic Press, 101–133.

——, 1988: Further observational characteristics of bimodal planetary waves: Mean structure and transitions. *Mon. Wea. Rev.*, **116**, 386–400.

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

——, and ——, 1988: Planetary wave amplitude bimodality in the Southern Hemisphere. *J. Atmos. Sci.,* **45,** 3771–3783.

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

——, and ——, 1992: Structure in the phase space of a general circulation model deduced from empirical orthogonal functions. *J. Atmos. Sci.,* **49,** 320–326.

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

Kushnir, Y., and J. M. Wallace, 1989: Low-frequency variability in the Northern Hemisphere winter: Geographical distribution, structure and time-scale dependence. *J. Atmos. Sci.,* **46,** 3122–3142.

Lorenz, E. N., 1963: The mechanics of vacillation. *J. Atmos. Sci.,* **20,** 448–464.

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

Molteni, F., A. Sutera, and N. Tronci, 1988: The EOFs of the geopotential eddies at 500 mb in winter and their probability density distributions. *J. Atmos. Sci.,* **45,** 3063–3080.

——, S. Tibaldi, and T. N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Quart. J. Roy. Meteor. Soc.,* **116,** 31–67.

Scott, D. W., R. A. Tapia, and J. R. Thompson, 1977: Kernel density estimation revisited. *Nonlinear Anal.: Theor. Methods and Appl.,* **1,** 339–372.

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

Sutera, A., 1986: Probability density distribution of large-scale atmospheric flow. *Adv. Geophys.,* **29,** 227–249.

Tung, K. K., and A. J. Rosenthal, 1985: Theories of multiple equilibria—A critical reexamination. Part I: Barotropic models. *J. Atmos. Sci.,* **42,** 2804–2819.