JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 101, NO. D21, PAGES 26,309–26,318, NOVEMBER 27, 1996

Empirical orthogonal functions of monthly precipitation and temperature over the United States and homogeneous stochastic models

Robert F. Cahalan, Lawrence E. Wharton and Man-Li Wu

Goddard Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland

Abstract. The monthly mean precipitation and temperature at $p = 62$ stations over the United States and Canada for $N = 91$ years (1900–1990) are analyzed in terms of empirical orthogonal functions (EOFs) and their variances. The eigenvalues and eigenfunctions are compared with a succession of stochastic noise models: (1) uncorrelated noise, having eigenvalues depending on the ratio p/n, with $n = N - 1$; (2) homogeneous noise having spatial correlations which are fit to the observations; and (3) homogeneous noise having both spatial and temporal correlations fit to the observations. Individual monthly data for January and July were analyzed as well as a combined data set of all months. The eigenvalue spectra of the homogeneous noise models are found to be in close agreement with the observed spectra even when time correlation is excluded from the model. Time correlations only slightly affect the results for temperature and have less impact for precipitation. The EOF patterns of the noise models contain inhomogeneities due only to the distribution of stations, the common correlation length, and the limited sample but are nevertheless in good agreement with the observed patterns, whose inhomogeneities may also be affected by secular trends and physical inhomogeneities such as orography. The observed EOF eigenvectors also show identifiable deviations from the homogeneous EOFs. Further work will be needed to see if these deviations can be convincingly associated with true physical inhomogeneities.

1. Introduction

Empirical orthogonal functions (EOFs) were introduced to meteorology over 40 years ago [*Wadsworth et al.*, 1948; *Lorenz*, 1956]. EOFs are the eigenvectors of the covariance matrix of a set of data. The form of the EOFs and the associated eigenvalue spectra are capable of describing coherent variability in large data sets and, consequently, are a most useful tool in climate research. Physical interpretation of EOF patterns requires careful determination of the signal in the presence of background natural variability or "noise" and also sampling errors [*North et al.*, 1982]. Here we attempt to model the noise by a succession of stochastic models having a small number of parameters. The EOF modes estimated from the observations will be assumed to be described by the stochastic model unless the estimated eigenvalues (i.e., mode variances) are larger than the model eigenvalues by an amount that exceeds the expected statistical error. The data set is assumed to consist of instantaneous snapshots of a geophysical field sampled at N equally spaced intervals of time, and at p locations in space, or stations. The time average at each station is removed, leaving $n = N - 1$ samples in time, which may be correlated. For simplicity, it is assumed that $p \leq n$. It is essential to have an adequate distribution of the p points to resolve the spatial variations of the geophysical field, and a sufficiently large number of samples, n, to reduce the statistical fluctuations of the eigenvalue spectra so that estimated eigenvalues may be distinguished from each other and from the noise.

Paper number 96JD01611.

In the pioneering work of *Kutzbach* [1967], precipitation, surface temperature, and sea level pressure over the United States were considered for 23 regions of roughly equal area and for January 25 months. In that case none of the precipitation eigenvalues could be distinguished from uncorrelated noise [*Cahalan*, 1983], though some modes could be distinguished for the other two fields. *Walsh and Mostek* [1980] were able to obtain improved sampling by employing 78 years of data at 61 points over the United States. In that case, the first four or five modes of precipitation become distinguishable from uncorrelated noise in January. However, in July, because of the shorter convective length scales in summer, only the first two eigenvalues are distinguishable from uncorrelated noise, and these are effectively degenerate [*Cahalan*, 1983].

For uncorrelated noise the eigenvalue spectrum is known as an analytic function of the ratio p/n [*Cahalan*, 1993], and this result is used as a check on the homogeneous noise model, by taking the limit of small correlation length. It should be emphasized that we are considering "noise" to include not only measurement error but also any actual physical variability of the temperature or precipitation consistent with the model assumptions of homogeneity and isotropy. In fact, most of this noise is likely due to natural variability. Typical standard deviations in this monthly data set are about 40 mm for the precipitation and 2°C for the temperature. These values exceed the measurement error as well as secular trends in the data. This natural variability will be modeled with homogeneous isotropic models without any secular trend. Aspects of the data not consistent with these assumptions are considered to be "signal," including any trends as well as inhomogeneous spatial and temporal correlations.

The goal here is to model the homogeneous noise as accu-

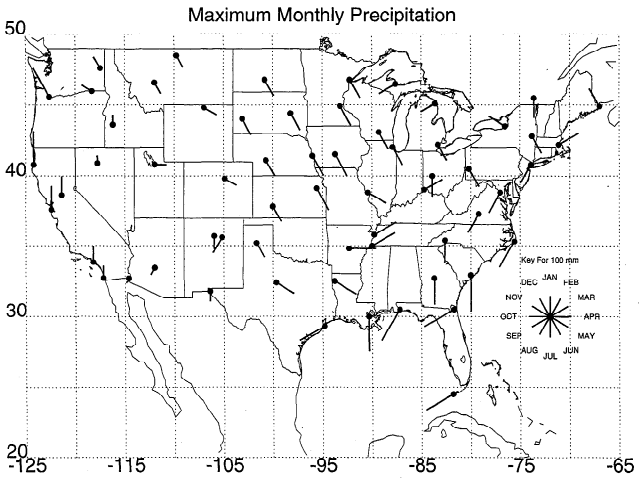

Figure 1. Location of the 62 stations used in the study along with a vector with length indicating the maximum precipitation and direction indicating the month of the maximum precipitation. Note the clear distinction between maximum precipitation in the various physical regions of the United States: the large winter maxima in the northwestern Pacific coast region, changing to early summer maxima east of the Rockies and throughout the Great Plains, switching to late summer in the southeast and Florida, with an intervening corridor of early spring maxima in "tornado alley."

rately as possible in order to increase our confidence that deviations from the noise will represent true physical inhomogeneities and not merely limitations in the model, or in the limited data sample. The approach is: (1) fit the spatial average correlations, ignoring regional differences, to determine parameters of the homogeneous model; (2) use the model to simulate data having the same station locations and time sampling as the observations; (3) apply the same EOF analysis to both observed and modeled data sets and identify any significant differences, considered to be "signal." We defer the fourth and final step, which is modeling the signal in terms of inhomogeneous and anisotropic processes. The inhomogeneous "signals" could involve both correlations and secular trends. Possible secular trends in the signal, which have been omitted from our analysis, are of particular interest in view of recent work suggesting that detection of possible global warming (increasing secular trend in temperature) may be formulated in terms of EOF analysis [e.g., *North et al.* 1995; *Shen et al.* 1994].

2. Data Analysis

The data sets consist of 91 years of monthly temperature and precipitation data from 62 stations in the contiguous United States and Canada, taken from 1900 to 1990. The station locations along with the maximum monthly precipitation are given in Figure 1. These locations were chosen to maximize both the uniformity and the completeness of the data. The effect of the seasonal variation may be treated by studying each set of data for a given month or season separately, or by considering all the months together and subtracting out the seasonal pattern. The starting point in the analysis will be the combined approach. The advantage of the combined approach is improved statistics, with a factor of 12 increase in the number of data samples, and the disadvantage is that seasonal variations in the covariances are averaged out. After sufficient understanding of the seasonally averaged EOF results from the combined data set is achieved, then seasonal differences may be studied with more confidence.

The annual cycle was subtracted from each month at each station by subtracting the 91-year mean for that month at that station. The computed mean annual cycle is summarized in Figure 1, which shows a clock hand at each station pointing in the appropriate direction to indicate the month of maximum precipitation, and with a length indicating the amount of maximum monthly precipitation at that location. Note the clear separation of the well-known climatological regimes of precipitation in the United States: wintertime maxima along the west coastal Cascades region; switching to a wide area of May–June maxima east of the Rocky Mountains; then a corridor of March–April–May maxima in the "tornado alley" from Texas up into southwestern Ohio; followed by July–August maxima east of the Allegheny mountains; and finally some large maxima during the September hurricane season in gulf coastal areas. Although these mean values (and the analogous temperature means) contain well-known inhomogeneities, we will attempt to model the monthly anomalies in terms of purely homogeneous stochastic processes. Comparison of the EOF analysis of the simple homogeneous models to the EOF analysis of the data will indicate the necessity of including inhomogeneities in modeling any "signal" in the observed anomalies.

We will write the meteorological field, F, here either monthly precipitation or temperature, as

$$F_{ik} = F_i(t_k) \tag{1}$$

where F is the observed quantity, i is the station index, and k is the observation time index. The data are normalized so that the average is zero and the standard deviation is unity, or

$$\hat{F}_{ik} = \frac{F_{ik} - \bar{F}_i}{\sigma_i} \tag{2}$$

with

$$\bar{F}_i = \frac{\sum_{k\in m(i)} F_{ik}}{\sum_{k\in m(i)} 1} \tag{3}$$

$$\sigma_i = \sqrt{\frac{\sum_{k\in m(i)} (F_{ik} - \bar{F}_i)^2}{\sum_{k\in m(i)} 1}} \tag{4}$$

and $m(i)$ equal to the set of all measurements which are not missing at station i. The covariance matrix will then be written as

$$C_{ij} = \frac{\sum_{k\in m(i)\cup k\in m(j)} \hat{F}_{ik}\hat{F}_{jk}}{N_{ij}} \tag{5}$$

The eigenvalue distribution and the eigenvectors of this matrix are then evaluated and compared to corresponding quantities computed from homogeneous stochastic models. The denominator in (5), N_{ij}, is the total number of observations for which both data points i and j are not missing.

The effect of missing data in this analysis causes two conflicting problems: If we use N_{ij} in (5), the positive definite property will be lost, but the statistically correct form of the correlation will be used. If we replace each missing F_{ik} by the mean, $\bar{F}_i$, and average, which is equivalent to using N, the total number of observations, in the denominator of (5), with no change in the numerator, then a statistically incorrect form of the correlation is used, but the positive definite property is

restored. The use of N has the effect of slightly underestimating the magnitude of the covariance, depending upon the amount of missing data. The use of N_{ij} has the unfortunate effect, from the loss of the positive definite property, of allowing some of the smallest eigenvalues to become negative, so that they cannot be meaningfully compared to the variances of the simulated EOFs. The data analysis was done both ways (using N_{ij} and N in (5)) and was found to have no effect on all but a few of the smallest eigenvalues in the individual monthly data. The larger and more significant eigenvalues were unchanged. There was no effect at all on the statistically more significant combined monthly data. In most EOF analysis results the lowest valued eigenvalues are not shown because they are not thought to be significant. Here we choose to show the positive definite results using uniform N, so that we can show all the eigenvalues for completeness. Disagreement between the model and the data spectrum, in excess of the expected statistical error, as the smallest eigenvalues are approached should be ignored. Alternate approaches are to throw out all data points for which a point is missing for any other station, or to replace missing data by spatial interpolation. The first was rejected because it would greatly increase the sampling error, and the second because it would tend to minimize inhomogeneities in the data, precisely what we are trying to identify by searching for places where homogeneous models fail.

3. A Stochastic Model With Spatial Correlation

Although the effect of uncorrelated noise alone is not sufficient to simulate the data, for simplicity we first consider the uncorrelated case and compare spectra determined from simulation to the analytic result of *Cahalan* [1993]. This provides a check on the accuracy of the simulations for a case in which the exact result is known as a function of p and N. As an example, $N = 100$ measurements, consisting of only noise with a Gaussian distribution of unit standard deviation, were generated independently at each of $p = 30$ stations. Then the eigenvalues of the covariance matrix were evaluated. This procedure was repeated 40 times and the results averaged to form the cumulative distribution shown in Figure 2. The analytic result is also plotted in Figure 2 and is completely hidden by the simulated eigenvalue distribution, which is thus shown to be quite accurate. The standard deviation of the simulation is of the order of the width of the curve.

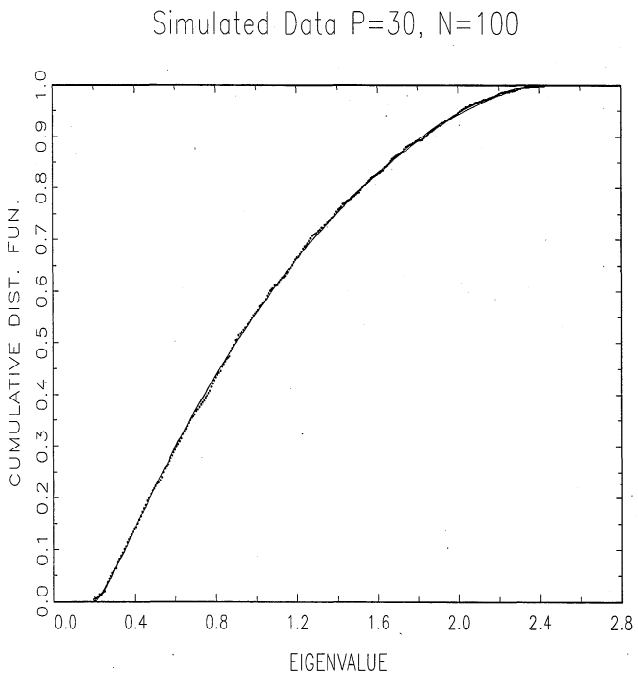

Figure 2. The cumulative distribution for the empirical orthogonal function (EOF) eigenvalues of a simulation with uncorrelated Gaussian noise. The dots are from the simulation and the solid line is the analytic spectrum given by *Cahalan* [1983, 1993]. The agreement is so thorough that the two curves cannot be distinguished.

Homogeneous stochastic simulations having a given spatial correlation will be generated following *North and Cahalan* [1981]. That technique expands the EOF eigenvectors in terms of spherical harmonics with the form of the expansion constrained in a way that allows imposition of any given spatial covariance function. The exact spherical harmonic expansion is too computationally intensive for our purposes, so we follow a simplified approach, using an expansion in terms of the EOF eigenvectors evaluated only at the given station locations. Although this calculation is greatly simplified, it must be emphasized that the main drawback is that it computes the model values only at the individual station locations or grid points. Should a model EOF field be required at other locations in space, the full spherical harmonic expansion or some appropriate analogous calculation must be done. A model series is required that simulates the homogeneous properties of the measured quantity $F(r_i, t_j)$, where r_i is the spatial position of the ith station and t_j is the time of the jth measurement. The simulated data, $F_s(r_i, t_j)$, will consist of a series of random numbers with the property that the homogeneous correlation for a given spatial separation will be the same as that of the actual data, $F(r_i, t_j)$. The variance of the simulated data is assumed to be the same for all stations, consistent with the treatment of the actual data, but this constraint may be relaxed, as discussed in the next section.

We will guess an expansion for $F_s(r_i, t_j)$ and check to see if the correct spatial correlation may be imposed on it. Consider then the ansatz

$$F_s(r_i, t_j) = \sum_n a_n(t_j)\, g_n(r_i) \tag{6}$$

where $a_n(t_j)$ is a series of random numbers with the index j taking the role of the time index. The vector $g_n(r_i)$ is the nth eigenvector, of the zero error or infinite N covariance matrix, evaluated at spatial point, or station index, i. Since the series of random numbers, $a_n(t_j)$, is assumed independent for different n, the only degree of freedom affecting the covarience of F will be the variances $\sigma^2(a_n)$. The simulated infinite N covariance matrix is then

$$A_{ik} = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} F_s(r_i, t_j) F_s(r_k, t_j) \tag{7}$$

In the case of precipitation the spatial correlation will be assumed to be in the form of an exponential, decreasing with spatial separation. The goal will then be to force the simulated covariance matrix to be in the form

$$A_{ik} = \exp(-\beta r_{ik}) \tag{8}$$

where β is the inverse of the spatial correlation length and r_{ij} is the great circle distance between stations i and k. The value

of β is determined by fitting to the data covariance matrix. For temperature the simple exponential form gives a poor fit and a third-order polynomial expansion in r_{ij} performs much better. So for temperature, the exponential in (8) is replaced by a cubic polynomial. The one degree of freedom in (6) is fixed by selecting the normalization

$$\sigma^2(a_n) = \lambda_n \tag{9}$$

with λ_n the nth eigenvalue. The simulated covariance matrix, (7), may now be evaluated by inserting the simulated data expansion, (6), giving

$$A_{ik} = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \left(\sum_{n=1}^{p} a_n(t_j)\, g_n(r_i) \sum_{m=1}^{p} a_m(t_j)\, g_m(r_k) \right) \tag{10}$$

Since the random numbers $a_n(t_j)$ are independent we have, after combining together the sum over j

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} (a_n(t_j) a_m(t_j)) = \delta_{nm} \sigma^2(a_n) \tag{11}$$

This result may be placed into (10) and the matrix A is then

$$A_{ij} = \sum_{n=1}^{p} \sigma^2(a_n)\, g_n(r_i)\, g_n(r_j) = \sum_{n=1}^{p} \lambda_n g_n(r_i)\, g_n(r_j) \tag{12}$$

where the normalization (9) has been used. This last expression is simply the diagonal form of the matrix given by (8), as may be verified using the matrix projection operator. The same result holds when (8) is replaced by the cubic form used for temperature, as long as the appropriate temperature eigenvectors are also used. Thus the desired goal of forcing the proper spatial correlation on our simulated data is achieved. Time correlation may be included by imposing a time correlation on the random numbers $a_n(t_j)$. Then both the desired spatial and the time correlation would be contained in the model. The drawback of this procedure is that the time correlation would be the same for all the stations, whereas the actual time correlation varies from station to station.

4. Results

A comparison of the spatial correlation for monthly precipitation versus the simulated series is shown in Figure 3. First the actual correlation length was obtained by fitting to the data, then this correlation length was put into the stochastic model through the exponential form in (8). We see that the average form of the spatial correlation agrees well between the model and the data but that the value of the data correlation varies over a wider range than that of the stochastic model. This effect is likely due to the fact the spatial correlation has a significant variation because it is not homogeneous and isotropic and these variations are not included in the model. In Figure 4 the eigenvalue spectrum of the data and model is shown. The ratio of the individual eigenvalues to the sum of all eigenvalues are shown and expressed as the individual percent variances. The effects of homogeneous and isotropic spatial correlation as well as noise are given by the N finite curve. The model with spatial correlation alone, without noise, is given by the N infinite curve. This model was generated by simply not adding in the effects of noise and is called the N infinite model because with an infinite number of measurements the noise would have no effect and the covariance matrix would be given by the spatial correlation matrix (8) exactly. The spectrum of the model with a correlation length of 1% of the correct value is also shown. This length is so low that the effects of spatial correlation vanish and only a noise model remains. We see good agreement between the data and the model with spatial correlation but poor agreement with the noise-only model. With the high value of N resulting from combining the months together, the effect of noise is small as indicated by the small difference between the N finite and N infinite model. There is a slight irregularity in these two curves which is due to the actual spatial locations of the stations. The N infinite model contains only the smooth exponential spatial correlation form and no noise. The only remaining explanation for the stochasticlike irregular appearance is the quasi-random spatial positioning of the stations. The first four EOFs for the monthly precipitation are shown in Figure 5, and the first four EOFs for the simulation are shown in Figure 6. There is good qualitative agreement with the data with the differences increasing with increasing mode number.

For temperature the simple exponential law is not valid and a third-order polynomial fit, constrained to equal unity at zero spatial lag, was used. The model and data spatial correlation are shown in Figure 7. As in the case of precipitation, the model simulates the average value of the spatial correlation well but the variation is greater in the data. The spectrum of the data, model, and noise only are shown in Figure 8. Here the agreement is not so good at the lower value eigenvalues with deviations exceeding the standard error. The lower values of the eigenvalues at which there is a deviation, in excess of the standard error, contain only about 1% of the variance and so, this deviation is not significant in describing the variations in the data field. As in the precipitation analysis, the effects of homogeneous and isotropic spatial correlation as well as noise are given by the N finite curve, and the model with spatial correlation alone without noise is given by the N infinite curve. The first four EOFs of the monthly temperature are shown in Figure 9 and the first four EOFs of the simulation are shown in Figure 10. As in the case of precipitation, there is good qualitative agreement with the data.

In the more physically relevant case of data for a specific month, patterns unique to the selected season will be present in the data. The spatial correlation model still is fairly accurate in describing the EOF analysis with the relative effects of noise having an increased effect, as would be expected. In the case of precipitation for a given month the standard error of the eigenvalues increases by $\sqrt{12}$ with the model still agreeing approximately with the data. The January and July precipitation spectrum is shown in Figure 11 and the January and July temperature spectrum is shown in Figure 12.

Here we see that the spatial correlation model with sampling errors (N finite) agrees quite well with the observed spectrum, while the exact spatial correlation model (N infinite) agrees less well with the observed. This difference is greater than in the case of the combined monthly precipitation, thus showing the increased effect of noise on the individual monthly analyses. As with the case of combined monthly precipitation, the sampling-only curve, generated by selecting a correlation length too low to allow any spatial correlation effects, does not agree with the data curve. However, the disagreement is not so great as in the combined monthly case, because of the increased importance of sampling error in generating the eigenvalue spectrum for a reduced (by a factor of 12) number of observations.

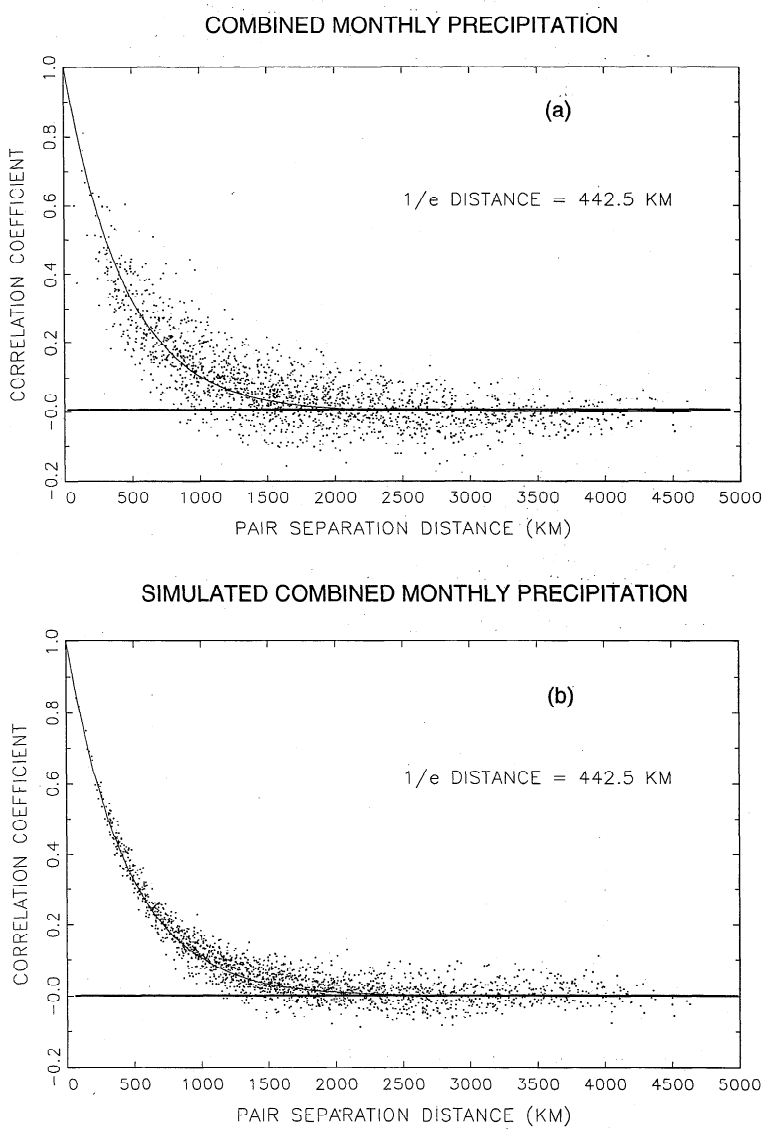

Figure 3. (a) The spatial correlation of the monthly precipitation, with the solid line an exponential fit, with correlation length 443 km. (b) This fit is used to generate the simulated values.

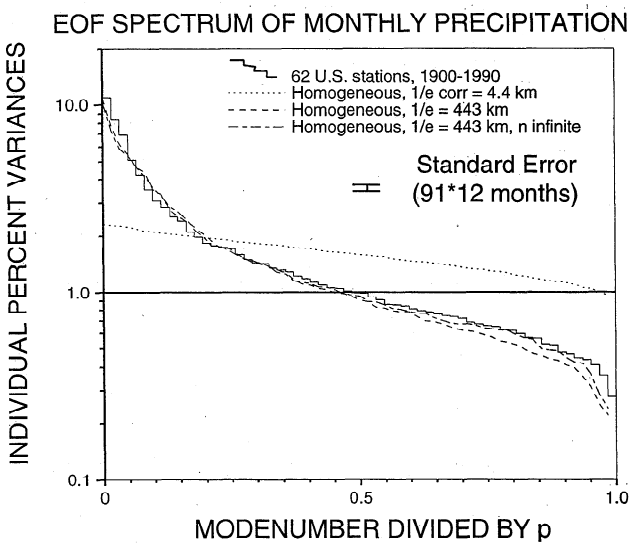

Figure 4. The spectrum of the combined monthly precipitation with eigenvalues given as a percentage of the total. The stair step line is the actual data and the dashed lines are from the model. The N finite model is determined by diagonalizing the correlation matrix determined from the same amount of data as for the actual precipitation and thus includes the effects of both homogeneous spatial correlation and sampling error. The N infinite model results from diagonalizing the exact exponential spatial correlation matrix. The model with correlation length 1% of the observed (i.e., 4.4 km) is indistinguishable from the analytic form for zero correlation length derived by *Cahalan* [1993].

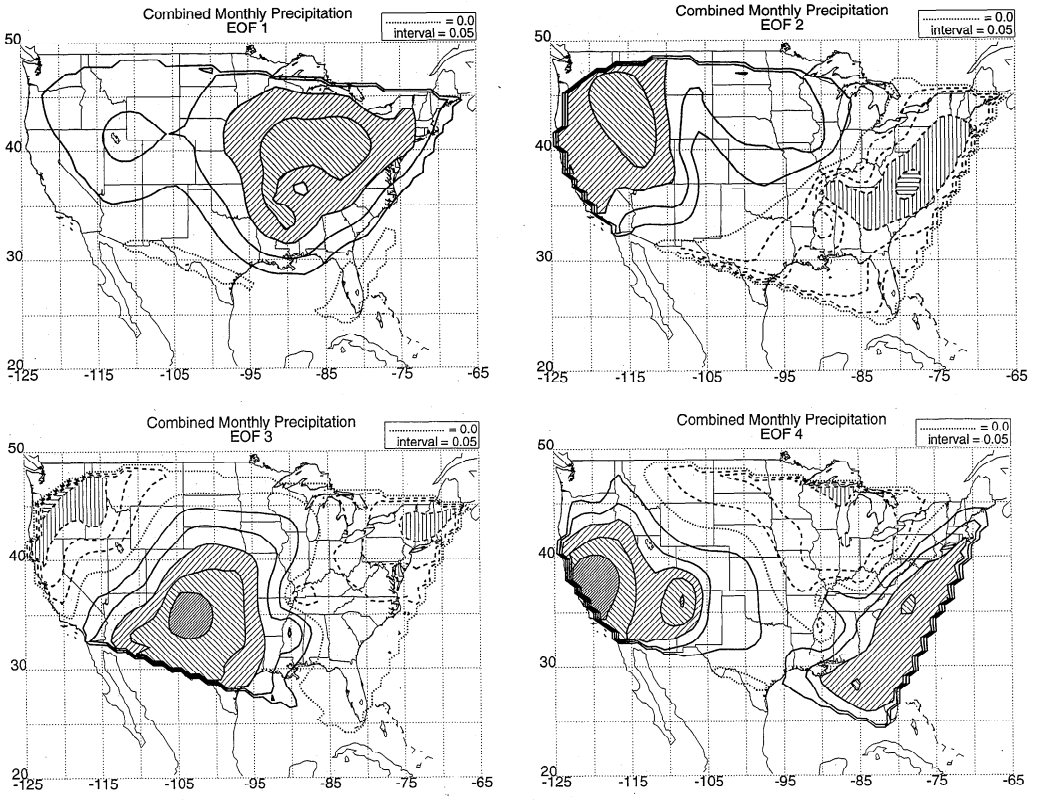

Figure 5. The first four EOFs for the combined monthly precipitation. Note in EOFs 1 and 2 the minimum in New Mexico and northward along the Rockies and the elongation of the positive region in the East along the southwest-northeast direction, parallel to the Allegheny mountain range.

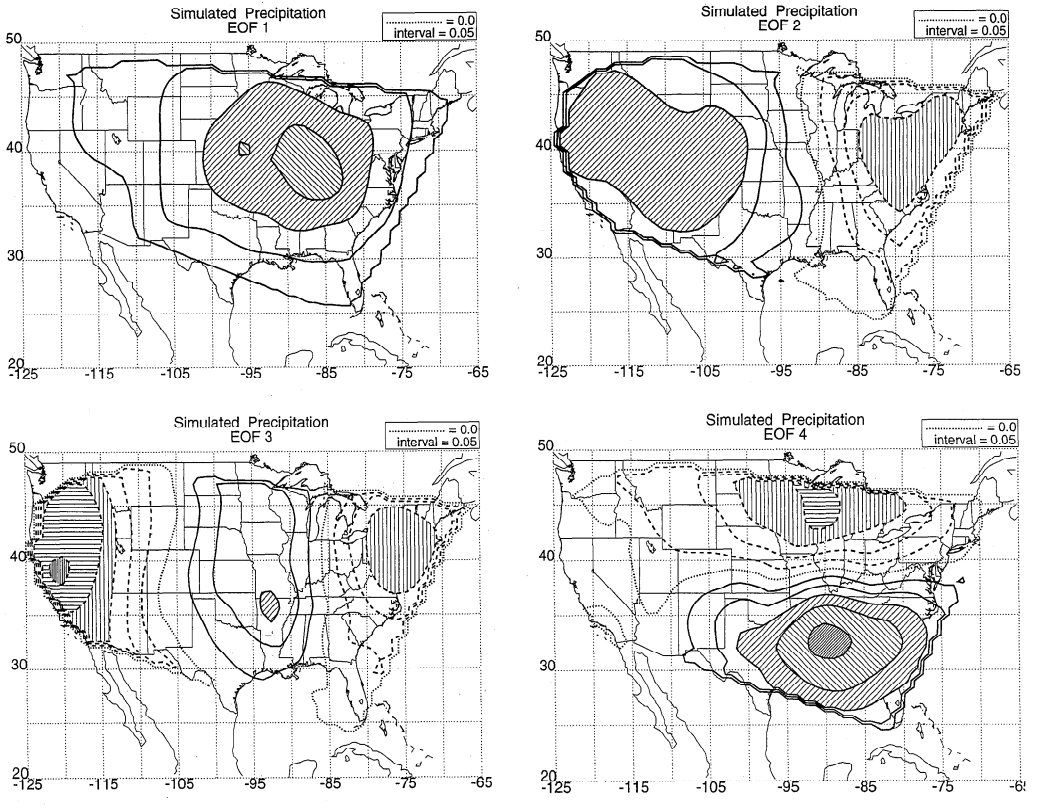

Figure 6. The first four EOFs for the simulated monthly precipitation. The same general pattern of one, two, and three east-west oriented extrema are seen in these first three simulated EOFs as occurs in the first three observed EOFs shown in Figure 5. However, the minimum along the Rockies and the elongation along the Alleghenies are not captured in the simulation. Also, the north-south pattern seen in simulated EOF 4 is not evident in the first four observed EOFs.

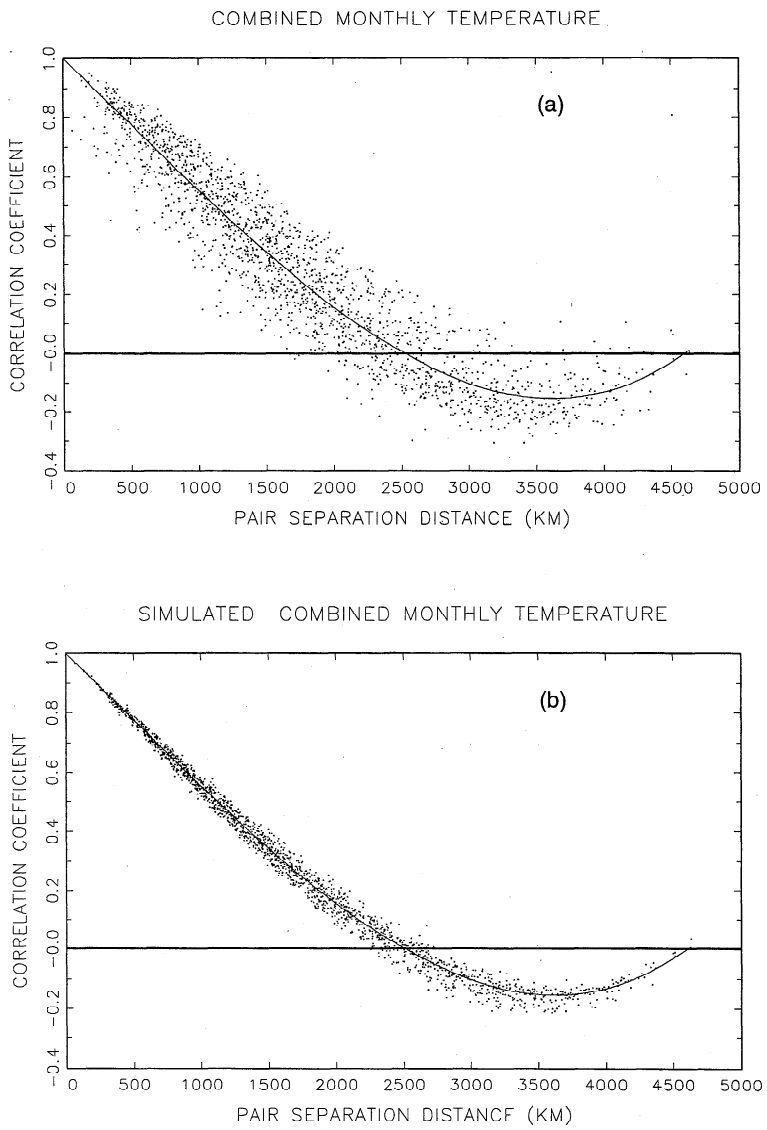

Figure 7. (a) The spatial correlation of the monthly temperature, with the solid line a cubic polynomial fit. (b) Simulated values determined from the cubic fit.

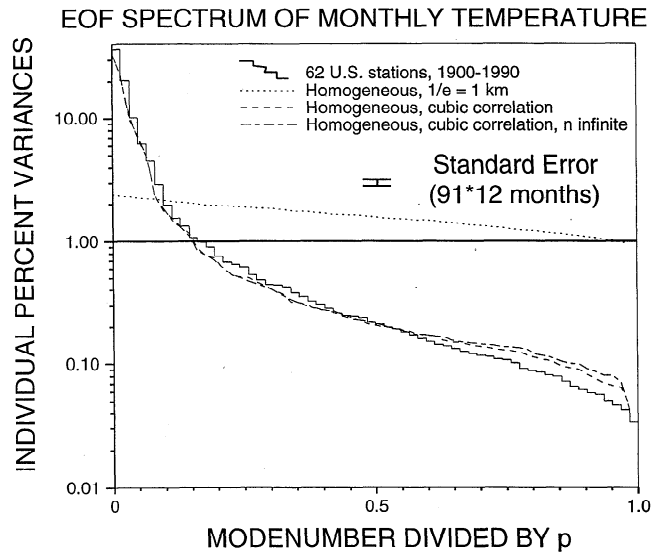

Figure 8. The spectrum of the combined monthly temperature with eigenvalues given as a percentage of the total. The stair step line is the actual data and the dashed lines are from the model. The N finite model is determined by diagonalizing the correlation matrix determined from the same amount of data as for the actual temperature and thus includes the effects of both homogeneous spatial correlation and sampling error. The N infinite model results from diagonalizing the exact cubic polynomial spatial correlation matrix shown in Figure 7. The model with correlation length 1 km is indistinguishable from the analytic form for zero correlation length derived by *Cahalan* [1993].

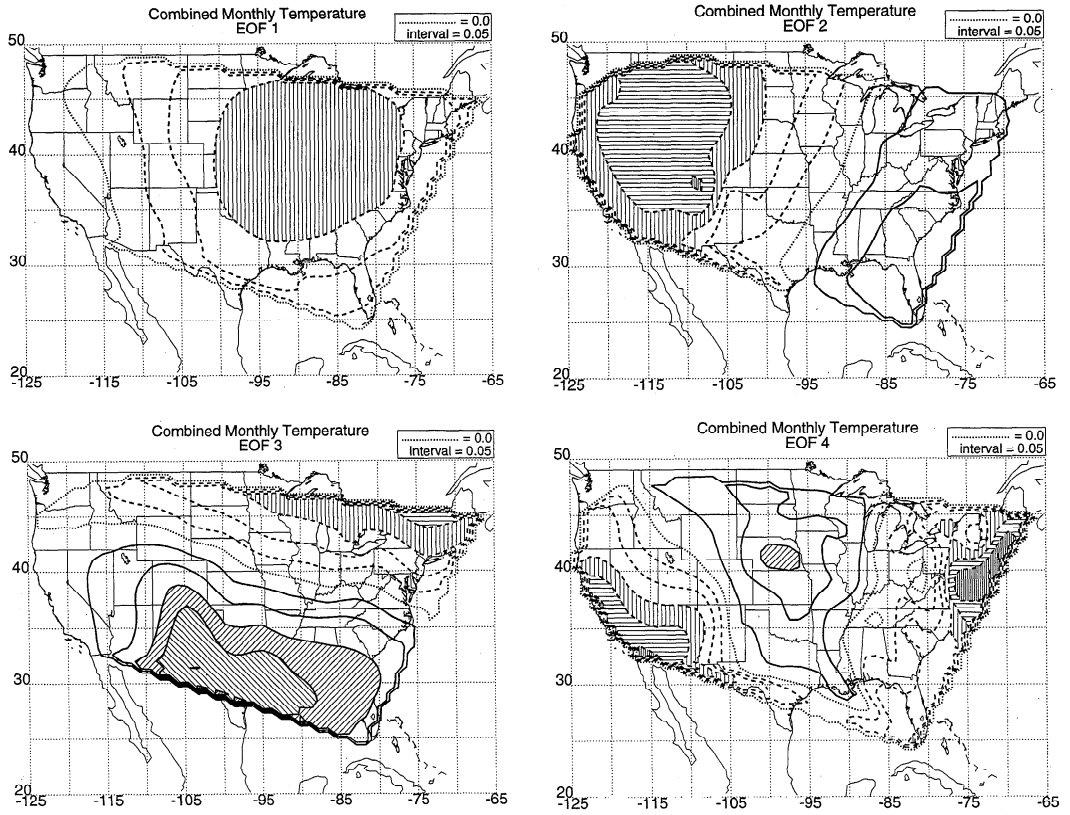

Figure 9. The first four EOFs for the combined monthly temperature. Note the larger spatial scale evident here than in the precipitation patterns shown in Figure 5, which is consistent with the larger scale of the spatial correlation. The influence of the major Rocky and Allegheny Mountain ranges are also less evident here than in the precipitation patterns.

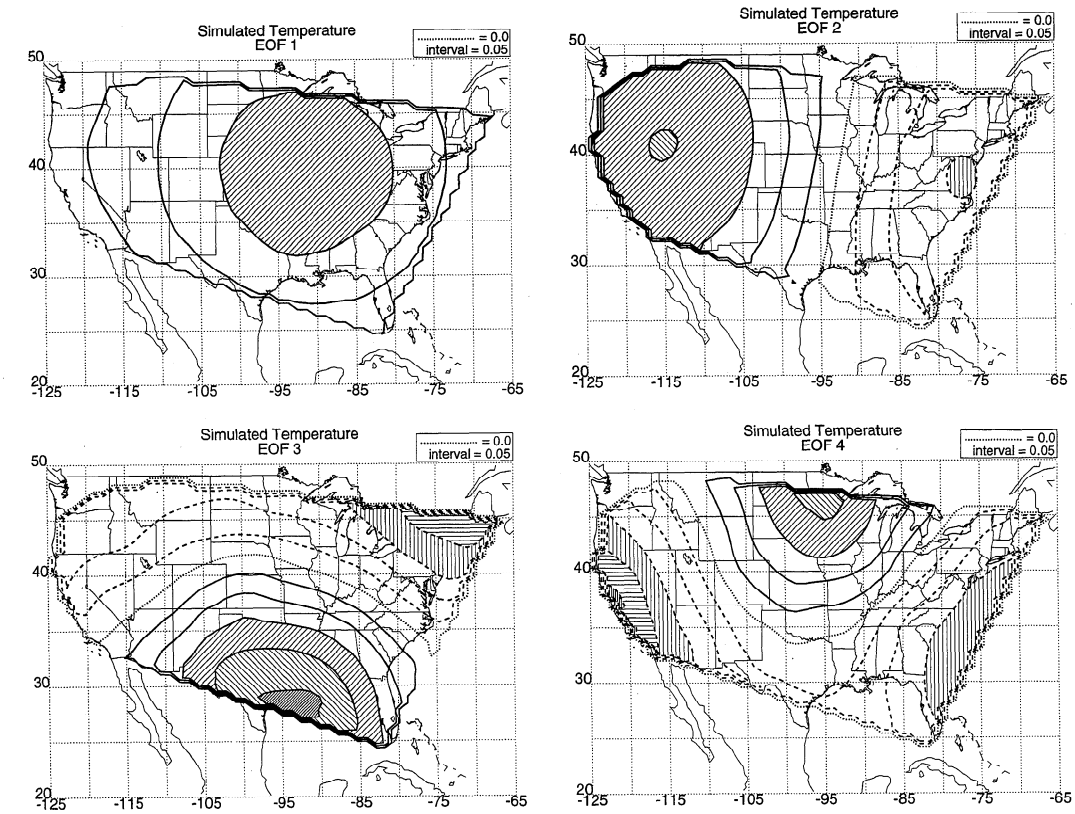

Figure 10. The first four EOFs for the simulated monthly precipitation. These are each remarkably similar to the corresponding observed EOF patterns shown in Figure 9. The main difference is the more localized regions seen in the observed EOF 4.

The effect of time correlation was added to the model for the combined monthly data. The time correlation for a 1 month lag for precipitation and temperature was about 0.06 and 0.2, respectively. There was no effect on the spectra for precipitation and the effect for temperature was small. For the individual monthly data, there also was no effect because each data point was separated by a much longer time (12 months).

5. Summary and Conclusions

An assumed value of EOF analysis, aside from merely compressing the data without losing much of the variance, is that it will reveal useful information or properties of the geophysical field being analyzed. The procedure used here was to construct a homogeneous model of the data, perform EOF analysis of the model, and compare the eigenvalue spectra and EOFs of the simulated and real data. If there are statistically significant deviations between the two spectra, then it is assumed that there exists some additional physical process that should be modeled as an inhomogeneous "signal." The first step of this procedure, comparing the observed spectrum to that of uncorrelated noise, is standard to most EOF analysis. It is usually done by performing a Monte Carlo simulation of the uncorrelated noise, though it is equivalent to use the simple analytic result of *Cahalan* [1983] and *Cahalan* [1993], which gives a better idea of the dependence on the number of stations and observations, p and N. Usual practice is to identify modes having variance greater than that of uncorrelated noise as "signal," which is then interpreted physically, perhaps with the help of a varimax or other rotation, as described, for example,

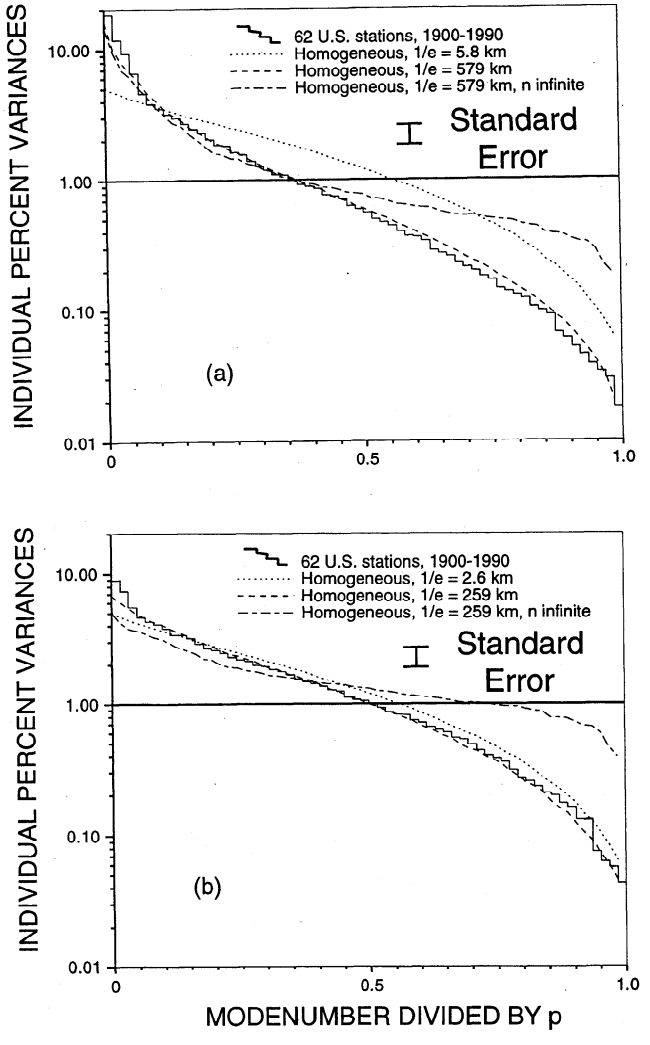

Figure 11. The spectrum of the (a) January and (b) July precipitation with eigenvalues given as the percentage of the total. The stair step line is the actual data and the dashed lines are from the model. The N finite model includes the effects of both homogeneous spatial correlation and sampling, while the N infinite model is the model resulting from spatial correlation alone. The noise-only model is shown as a model with correlation length so low that spatial correlation effects vanish. Note the close agreement of the July precipitation eigenvalue spectrum to the purely uncorrelated noise spectrum for all but the first three eigenvalues, and the second and third are effectively degenerate, being separated by less than a standard error.

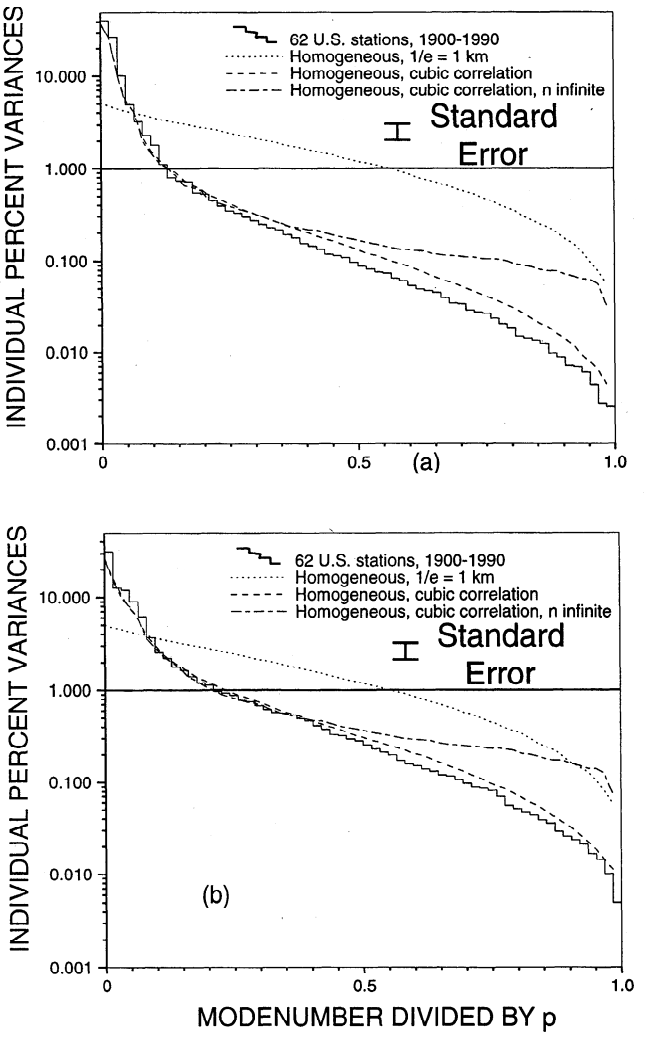

Figure 12. The spectrum of the (a) January and (b) July temperature with eigenvalues given as the percentage of the total. The stair step line is the actual data and the dashed lines are from the model, with the same three models as in Figure 11. Note the concentration of variance in the first 10 or so eigenvalues, giving a spectrum much steeper than that of precipitation. Nevertheless, the homogeneous noise model fits quite well in this region, even without including sampling errors. It is necessary to include sampling error to fit the smaller eigenvalues, however.

by *Richman* [1986]. Here we have seen that some of the features of these large-scale "signal" modes can be understood by the addition of homogeneous spatial correlation to the stochastic model, resulting in a significant improvement in the agreement between the homogeneous model and the data eigenvalues and spatial patterns. The focus then shifts to modeling the deviations from the homogeneous model in terms of physically inhomogeneous processes. If the deviations are sufficiently small, the inhomogeneous signal may lend itself to a perturbative treatment. We recommend that this approach be standard procedure for any EOF analysis of data having significant spatial correlation.

The case of monthly temperature and precipitation analyzed here shows that for such geophysical fields the observed homogeneous spatial correlation has a significant impact on the eigenvalue spectra and thus must be included in any noise model. These correlations, when combined with the limited spatial and temporal sampling of the data, introduce spurious inhomogeneities into the EOFs which may be difficult to distinguish from inhomogeneities in the EOFs that have an inhomogeneous physical origin.

The precipitation spatial correlation is described well with an exponential, while the temperature spatial correlation is better fit with a third-order polynomial. The simultaneous inclusion of noise and spatial correlation yields an eigenvalue spectrum that agrees well with the data spectrum. For both the case of the combined monthly data and the January-only data the deviation of the simulated spectrum from the data spectrum is within the standard error for most of the significant eigenvalues. The greatest deviations occur at eigenvalues that are less than 1% of the largest value and hence are not significant in describing the total variance of the data.

This analysis shows that most of the EOF signal, as measured by the eigenvalue spectrum for the monthly data, arises from homogeneous spatial correlation and sampling error. The good agreement between the homogeneous model and the data is an indication that it may be difficult to infer properties of the data in addition to the effects of noise and spatial correlation. It is, however, clear that inhomogeneous effects are present. There is an excess, of the order of about a factor of 2, in the variation in the covariance matrix from the data over the covariance matrix from the simulation, as evidenced from Figures 3 and 7. This excess must be due to physical effects which are not included in the model. Likely candidates for these effects are spatial anisotropies and inhomogeneities, combined space and time correlation, and secular trends in time. It is not clear if EOF analysis will be useful in diagnosing these phenomena.

The observed and simulated EOFs were in qualitative agreement, but differences are evident in the precipitation patterns in the regions of the major Rocky and Allegheny Mountain ranges and with deviations increasing for larger mode numbers. The requirement of orthogonality causes deviations seen in the first EOF to propagate to other modes. It will be of interest to isolate the effects of physical inhomogeneity from that introduced by the limited station sample by expanding the first few EOFs in components parallel and perpendicular to the simulated EOFs. Studying the perpendicular components alone will be an interesting topic for further research.

References

Cahalan, R. F., EOF spectral estimation in climate analysis, *Proc. Int. Meet. Stat. Clim.*, *2*, 4.5.1–4.5.7, 1983a.

Cahalan, R. F., Analytic variance spectrum for empirical orthogonal functions, in *Precipitation Variability and Climate Change*, edited by B. Sevruk and M. Lapin, pp. 160–162, 1993a.

Kutzbach, J. E., Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America, *J. Appl. Meteorol.*, *6*, 791–802, 1967.

Lorenz, E. N., Empirical orthogonal functions and statistical weather prediction, *Sci. Rep. 1*, 49 pp., Stat. Forecasting Proj., Dep. of Meteorol., Mass. Inst. of Technol., 1956. (Available as *NTIS AD 110268*, Natl. Tech. Inf. Serv., Springfield, Va.)

North, G. R., and R. F. Cahalan, Predictability in a solvable stochastic climate model, *J. Atmos. Sci.*, *38*, 504–513, 1981.

North, G. R., and S. S. Shen, Detection of forced climate signals, I, Filter theory, *J. Clim.*, *8*, 401–408, 1995.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions, *Mon. Weather Rev.*, *110*, 669–706, 1982.

Richman, M. B., Rotation of principal components, *J. Climatol.*, *6*, 293–335, 1986.

Shen, S. S., G. R. North, and K.-Y. Kim, Spectral approach to optimal estimation of the global average temperature anomaly, *J. Clim.*, *7*, 1997–2007, 1994.

Wadsworth, G. P., J. G. Bryan, and C. H. Gordon, Short range and extended forecasting by statistical methods, U.S. Air Force, *Air Weather Serv. Tech. Rep. 105-38*, 186 pp., Washington, D. C., 1948.

Walsh, J. E., and A. Mostek, A quantitative analysis of meteorological anomaly patterns over the United States, *Mon. Weather Rev.*, *108*, 615–630, 1980.

R. F. Cahalan, M.-L. Wu, and L. E. Wharton, Climate and Radiation Branch, Code 913, NASA Goddard Space Flight Center, Greenbelt, MD 20771.

(Received October 18, 1995; revised May 7, 1996;
accepted May 14, 1996.)