The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field

CHRISTOPHER S. BRETHERTON, MARTIN WIDMANN, VALENTIN P. DYMNIKOV,*
JOHN M. WALLACE, AND ILEANA BLADÉ

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

(Manuscript received 11 March 1998, in final form 14 August 1998)

ABSTRACT

The authors systematically investigate two easily computed measures of the effective number of spatial degrees of freedom (ESDOF), or number of independently varying spatial patterns, of a time-varying field of data. The first measure is based on matching the mean and variance of the time series of the spatially integrated squared anomaly of the field to a chi-squared distribution. The second measure, which is equivalent to the first for a long time sample of normally distributed field values, is based on the partitioning of variance between the EOFs. Although these measures were proposed almost 30 years ago, this paper aims to provide a comprehensive discussion of them that may help promote their more widespread use.

The authors summarize the theoretical basis of the two measures and considerations when estimating them with a limited time sample or from nonnormally distributed data. It is shown that standard statistical significance tests for the difference or correlation between two realizations of a field (e.g., a forecast and an observation) are approximately valid if the number of degrees of freedom is chosen using an appropriate combination of the two ESDOF measures. Also described is a method involving ESDOF for deciding whether two time-varying fields are significantly correlated to each other.

A discussion of the parallels between ESDOF and the effective sample size of an autocorrelated time series is given, and the authors review how an appropriate measure of effective sample size can be computed for assessing the significance of correlations between two time series.

1. Introduction

The number of effective spatial degrees of freedom (ESDOF) is among the most fundamental properties of a time-varying spatial field such as 500-mb height. It can serve as a basis for (i) estimating the number of stations or grid points required to represent the field (Bagrov 1969; TerMegreditchian 1969, 1990), (ii) assessing field significance (e.g., Lorenz 1969; Livezey and Chen 1983; van den Dool and Chervin 1986; Wallace et al. 1991), (iii) characterizing and comparing observed and simulated fields (Fraedrich et al. 1995), and (iv) estimating the dimension of an underlying attractor that may describe the evolution of the field (Toth 1995; Fraedrich et al. 1998, manuscript submitted to *J. Atmos. Sci.*; Dymnikov and Gritsun 1997). It could also be useful for assessing the feasibility and robustness of matrix inversions such as those used in statistical prediction.

Taylor (1921) and Keller (1935) related the complexity of random, isotropic turbulence to the spatial autocorrelation function $r(\tau)$, where τ is the separation distance. Comparing the temporal variance of the mean value of a random field at N contiguous, equally spaced grid points versus the mean of n independent values of the field, they showed that

$$n = \frac{N}{2L},$$

where L is the correlation radius, as estimated from

$$L = \int_0^\infty r(\tau)\, d\tau.$$

Hence, $2L$ is, in effect, the distance between independent observations. This formalism can be generalized to two or three dimensions, but it is not well suited for dealing with the wavelike or modal structures that often dominate large-scale meteorological and climatological fields.

An alternative approach, first suggested and justified by Bagrov (1969) and TerMegreditchian (1969), is to match the probability density function (PDF) of a specified functional of the field such as the spatially integrated variance to the χ^2 distribution. The matching has been done on the basis of the first two statistical mo-

* Current affiliation: Institute of Numerical Mathematics of the Russian Academy of Sciences, Moscow, Russia.

Corresponding author address: Dr. C. S. Bretherton, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640.
E-mail: breth@atmos.washington.edu

© 1999 American Meteorological Society

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

ments (for the variance functional, these are determined by the spatially integrated variance and kurtosis of the field itself); see, for example, Bagrov (1969) and TerMegreditchian (1969); visual inspection (Wallace et al. 1991); and the interquartile range (Toth 1995; Bagrov 1969). For a given functional, these approaches will give almost identical answers. An attractive feature of this measure of ESDOF is that if the fluctuations in the field variable are normally distributed, the ESDOF can be easily estimated from the eigenvalue spectrum of the temporal covariance matrix (TerMegreditchian 1969; Fraedrich et al. 1995).

Other measures of ESDOF have been developed for other problems. Livezey and Chen (1983) discussed how to test whether the correlation between two spatial fields, such as 500-mb height and sea surface temperature, is statistically significant. Their approach is to determine the fraction of the grid points at which the local temporal correlation is significant at some level, then to test whether this fraction is significantly larger than one would expect if the fields were independent. The critical fraction for field significance depends considerably on the spatial correlation structure of the fields and can be higher than if all grid points were independent. Livezey and Chen defined an ESDOF for their problem using a Monte Carlo approach, as did van den Dool (1987) for a similar type of problem.

The concept of ESDOF is certainly no less clearly defined (in a χ^2 context) and of no less practical value than the number of significant modes in an EOF expansion, yet it seems to be much less widely understood and appreciated in the research community. Its usage thus far has tended to be restricted to those who have, in effect, reinvented it and customized it for their own purposes.

This paper is intended for researchers who wish to use the ESDOF methodology without reinventing it. In section 2, we review Bagrov's and TerMegreditchian's formulas for ESDOF. The remaining sections go beyond prior work to systematically investigate the performance of these ESDOF formulas in contexts relevant to geophysical data analysis. Section 3 shows that for a typical application the ESDOF formulas give results similar to those of past investigators with much less computation. It also explores how long a time record is needed to accurately determine ESDOF, and how ESDOF can be affected by nonnormality of the data. In section 4 we demonstrate that these measures of ESDOF are appropriate for tests of field significance, for example, whether a realization of a field deviates significantly from its mean, or whether two realizations of a field are significantly correlated, even when the data are nonnormally distributed. Section 5 shows the connection of ESDOF to the parallel concept of effective sample size for temporal autocorrelated time series and briefly reviews the application of effective sample size to statistical significance testing of the correlation coefficient between two temporally autocorrelated time series. Section 6 presents our conclusions.

2. Effective number of degrees of freedom of a field

Following Bagrov (1969) and TerMegreditchian (1969), let us consider a time-varying spatial field defined at N stations[1] and over some finite time interval, within which the mean value is identically equal to zero at all stations, and let $\psi_i(t)$ be its value at the ith station at time t. We choose some quadratic functional of the $\psi_i(t)$, such as the variance summed over all N stations,

$$E(t) := [\psi(t), \psi(t)] := \sum_{i=1}^{N} \psi_i^2(t),$$

where $\psi(t)$ denotes the vector of the $\psi_i(t)$ and (,) denotes an inner product. We use this to construct a measure of the effective number of spatial degrees of freedom. This measure, N^*, is defined as the number of uncorrelated random normal variables a_k, each having zero mean and the same population variance $\langle a^2 \rangle$, for which the χ^2 distribution for the specified functional most closely matches the PDF of the functional of $\psi(t)$. This approach has been used by Lorenz (1969) and Wallace et al. (1991) as a basis for estimating the statistical significance of analogs. In order to increase the number of samples for estimating the PDF they used as their functional squared Euclidean distance $\Sigma_i (\psi_{im} - \psi_{in})^2$ between all possible pairs of maps (m, n), excepting those separated in time by an interval less than the correlation time.

Rather than determining the best match visually, Bagrov (1969) required that the χ^2 distribution match the observed distribution's ensemble mean value $\langle E \rangle$ and its temporal variance of E about the ensemble mean,

$$\mathrm{var}(E) = \langle E'^2 \rangle = \langle (E - \langle E \rangle)^2 \rangle$$

(i.e., the first two moments). Unless otherwise noted, ensemble means $\langle ( ) \rangle$ are estimated on the basis of time averages of all available data. For the χ^2 distribution, $\langle E \rangle = N^* \langle a^2 \rangle$ and $\mathrm{var}(E) = 2N^* \langle a^2 \rangle^2$. Solving for N^*, we obtain a ''moment matching'' estimate of ESDOF,

$$N^*_{mm} = \frac{2\langle E \rangle^2}{\mathrm{var}(E)}, \qquad \langle a^2 \rangle_{mm} = \frac{\mathrm{var}(E)}{2\langle E \rangle}. \tag{1}$$

Using this relationship it is possible to estimate N^* directly from the time series of E without explicit calculation of its PDF. There is no guarantee that N^*_{mm} remains within the anticipated range $1 \le N^*_{mm} \le N$ unless the number of independent sampling times is large and the ψ_i's are normally distributed.

An alternative way of estimating N^*, also discussed

[1] We will use the term ''station'' to denote the locations in three-dimensional space at which the field is defined.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

by Bagrov (1969) and TerMegreditchian (1969), is based on the $N \times N$ covariance matrix $\mathbf{C}$ of the ψ_i's. This approach is based on two assumptions: (i) that the ψ_i's are normally distributed and (ii) that $\mathbf{C}$ is known with sufficient accuracy (i.e., that it is based on a relatively large number of independent samples). We assume that $\mathbf{C}$ has eigenvalues λ_k and normalized principal components $z_k(t)$ formed by projecting the field vector $\psi(t)$ onto the kth normalized eigenvector (EOF) $\mathbf{e}_k$.

We may now calculate N^* in terms of the λ_k from (1):

$$E(t) = \sum_{k=1}^{N} \lambda_k z_k^2,$$

$$\langle E \rangle = \sum_{k=1}^{N} \lambda_k, \tag{2}$$

and

$$\mathrm{var}(E) = \sum_{k=1}^{N} \lambda_k^2 \,\mathrm{var}(z_k^2) = \sum_{k=1}^{N} \lambda_k^2 \langle (z_k^2 - \langle z_k^2 \rangle)^2 \rangle$$

$$= \sum_{k=1}^{N} \lambda_k^2 (\langle z_k^4 - \langle z_k^2 \rangle^2 \rangle).$$

Since the z_k are normalized and normally distributed, it follows that $\langle z_k^2 \rangle = 1$, $\langle z_k^4 \rangle = 3$, and

$$\mathrm{var}(E) = 2 \sum_{k=1}^{N} \lambda_k^2.$$

Note that for a time record of finite length or nonnormally distributed observations $\psi_i(t)$, this estimate of var(E) will be different from that obtained directly from the time series of E. Matching the mean and variance of a χ^2 distribution with N^* degrees of freedom from these estimates as in (1), we obtain an ''eigenvalue formula'' estimate of ESDOF,

$$N_{ef}^* = \frac{\left(\sum_{k=1}^{N} \lambda_k \right)^2}{\sum_{k=1}^{N} \lambda_k^2}. \tag{3}$$

In fact, N_{ef}^* depends only on the partitioning of the variance between the EOFs. If f_k is the fraction of the overall variance explained by EOF k, (3) can be written

$$N_{ef}^* = \left(\sum_{k=1}^{N} f_k^2 \right)^{-1}. \tag{4}$$

It is readily verified that $1 \le N_{ef}^* \le N$, so that in contrast to N_{mm}^*, N_{ef}^* does remain within the anticipated range. In the limiting case of equal λ_k, $N_{ef}^* = N$, whereas if a single EOF accounts for all the variance of ψ, $N_{ef}^* = 1$.

The N_{ef}^* can be estimated directly from $\mathbf{C}$ without performing eigenvector analysis by rewriting (3) in the form

$$N_{ef}^* = \frac{(\mathrm{tr}\mathbf{C})^2}{\mathrm{tr}(\mathbf{C}^2)} = \frac{\left(\sum_{i=1}^{N} C_{ii} \right)^2}{\sum_{i,j=1}^{N} C_{ij}^2}. \tag{5}$$

Here, tr($\mathbf{C}^2$) is also the square of the Frobenius norm of $\mathbf{C}$.

The estimates N_{mm}^* and N_{ef}^* need not be identical unless the number of independent observations of the spatial field is large and the principal components are normally distributed.

In order to obtain some insight into the interpretation of N_{ef}^*, suppose that the eigenvalues of the covariance matrix drop off geometrically with k, that is,

$$\lambda_{k+1}/\lambda_k = \exp(-\mu), \qquad (\mu > 0). \tag{6}$$

If μ is large, just a few EOFs dominate the variance, while if μ is small, the variance will be more evenly distributed among the EOFs. This type of variance partitioning frequently occurs in real geophysical applications (see Fig. 1 of section 3 for an example). Estimating N_{ef}^* based on (3) yields

$$N_{ef}^* = \frac{\left[\sum_{k=0}^{\infty} \exp(-k\mu) \right]^2}{\sum_{k=0}^{\infty} \exp(-2k\mu)} = \frac{1 + \exp(-\mu)}{1 - \exp(-\mu)}. \tag{7}$$

As we would expect, N_{ef}^* is large if μ is small and many modes contribute significantly to the variance. When μ is fairly small, $N_{ef}^* \approx 2/\mu$ and the fraction of the overall variance explained by the first N_{ef}^* EOFs is

$$\frac{\sum_{k=0}^{N_{ef}^*-1} \exp(-k\mu)}{\sum_{k=0}^{\infty} \exp(-k\mu)} = 1 - \exp(-N_{ef}^*\mu)$$

$$\approx 1 - e^{-2} = 86\%, \tag{8}$$

regardless of the rate at which the eigenvalues decrease. Huang et al. (1996) mentioned that in their experience, roughly 90% of the variance was explained by the first N_{ef}^* EOFs. They were computing ESDOF using a different method in a different context and did not present an argument why this should be so, but the agreement is reassuring.

In geophysical applications, the observations are often not normally distributed, and N^* must be estimated based on a limited number of sampling times. In the next two subsections, we discuss these complications, which affect N_{ef}^* and N_{mm}^* in different ways. We consider both N_{ef}^* and N_{mm}^*, since we will see later that each can be useful for different purposes.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

a. Differences between N^*_{ef} and N^*_{mm} due to non-normally distributed data

The moment-matching estimate N^*_{mm} of ESDOF is not based on any assumptions about the data at individual stations, although it should be expected that when the data are far from normally distributed, the actual distribution of $E(t)$ should not be well approximated by a χ^2 distribution, even if the two distributions have the same mean and variance. However, the assumption of normality did enter the derivation of the formula for N^*_{ef} in the calculation of var(E), where it was assumed that the fourth moment of the normalized principal components, also known as their kurtosis, is equal to 3. For observed distributions, the kurtosis can vary considerably. For a distribution with long tails, the kurtosis κ can be much larger than 3. For example, the exponential distribution has $\kappa = 9$. For a compact distribution, κ can be small, for example, $\kappa = 1.8$ for a uniform distribution.

To gain some sense how nonnormality affects N^*_{ef}/N^*_{mm}, we assume that all the principal components (PCs) have the same kurtosis. Repeating the derivation of (3), we obtain

$$\frac{N^*_{ef}}{N^*_{mm}} = \frac{\kappa - 1}{2}. \tag{9}$$

This ratio can vary considerably; it is 4 for PCs with an exponential distribution and 0.4 for PCs with a uniform distribution.

b. Effect of sample length on estimated N^*_{ef} and N^*_{mm}

Estimates of both N^*_{mm} and N^*_{ef} based on a limited set of sampling times may differ considerably from their ''true'' (long sample record) values. To quantify these errors, we let a hat above a quantity denote an estimate of that quantity based on observations at a random set of T independent sampling times. Such an estimate is a random variable; we denote its mean by angle brackets and its standard deviation by δ placed in front of it. We will use $\langle\widehat{N^*_{mm}}\rangle$ and $\delta\widehat{N^*_{mm}}$ to characterize the bias and scatter in $\widehat{N^*_{mm}}$ compared to the true value $\widehat{N^*_{mm}}$, and similarly for $\widehat{N^*_{ef}}$.

We are not aware of any prior analysis of the bias and scatter of $\widehat{N^*_{mm}}$. Fraedrich et al. (1995) estimated the scatter of $\widehat{N^*_{ef}}$ by using North et al.'s (1982) rule of thumb for the typical error in eigenvalues of a covariance matrix determined from a limited set to observations:

$$\delta\widehat{N^*_{ef}}/\widehat{N^*_{ef}} = 4(2/T)^{1/2}. \tag{10}$$

Fraedrich et al. (1998, manuscript submitted to *J. Atmos. Sci.*) also found empirically that $\widehat{N^*_{ef}}$ tend to be smaller than the true $\widehat{N^*_{ef}}$ but did not propose a formula for this low bias. One can understand how temporal undersampling can lead to a low bias of $\widehat{N^*_{ef}}$ by considering the limiting case of a single sample of a zero mean process, $T = 1$. If the one measured value at station i is ϕ_i, the elements of the covariance matrix are just $\hat{C}_{ij} = \phi_i\phi_j$. The covariance matrix can be written

$$\hat{\mathbf{C}} = \phi\phi^T,$$

where ϕ is the vector with components ϕ_i. From the spectral theorem (Strang 1988, p. 296), such a matrix has rank 1, that is, one nonzero eigenvalue, with corresponding eigenvector ϕ. Hence, the estimated $\widehat{N^*_{ef}} = 1$, which is always an underestimate of N^*_{ef}. Similarly, an estimate $\widehat{N^*_{ef}}$ based on T observations can never exceed T.

The same principle can be expected to hold for other estimates of ESDOF. To distinguish N^* independent spatial modes, we must be able to differentiate between their temporal variability. We cannot expect to do this accurately unless there are many more observing times than modes, that is, unless $T \gg N^*$.

In the appendix, we derive formulas for the bias and scatter of estimates of $\widehat{N^*_{mm}}$ and $\widehat{N^*_{ef}}$ in the parameter regime

$$T \gg N^* \gg 1. \tag{11}$$

This excludes samples so short ($T \leq N^*$) that one cannot expect a useful estimate of N^* and also excludes fields with a small number (in practice, 5 or less) of ESDOF. If N^* is small, our formulas are still qualitatively useful, but may be in error by a factor of 2 or more.

For $\widehat{N^*_{mm}}$ we find

$$\langle\widehat{N^*_{mm}}\rangle/N^*_{mm} \approx 1 + 2/T. \tag{12}$$

$$\delta\widehat{N^*_{mm}}/N^*_{mm} \approx (2/T)^{1/2}, \tag{13}$$

For short samples, $\widehat{N^*_{mm}}$ has a slight positive bias, but the error is dominated by the scatter.

For $\widehat{N^*_{ef}}$ we find

$$\langle\widehat{N^*_{ef}}\rangle/N^*_{ef} \approx \frac{1}{1 + N^*_{ef}/T}, \tag{14}$$

$$\delta\widehat{N^*_{ef}}/N^*_{ef} \approx cT^{-1/2}, \tag{15}$$

Unlike Fraedrich's formula (10), which was obtained using a scaling analysis, these formulas are rigorously derived. The constant c, given in (A26), depends upon the eigenvalue spectrum. It is typically less than 0.5, much less than predicted by Fraedrich's formula, so the scatter of $\widehat{N^*_{ef}}$ is very small. In fact, if all EOFs explained the same fraction of the variance, or if one EOF explained the entire variance, c would be zero. The N^*_{ef} has a considerable low bias unless T is many times as large as N^*_{ef}. The predicted bias is 50% for $T \approx N^*_{ef}$ independent samples and 11% for $T \approx 10N^*_{ef}$ independent samples.

Let us consider a specific example. Consider a field observed at a large number of stations with a normal distribution of observed values at each station, and $N^* = 20$ ESDOF. With an infinitely long time record, we would correctly conclude that $N^*_{mm} = N^*_{ef} = 20$. With

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

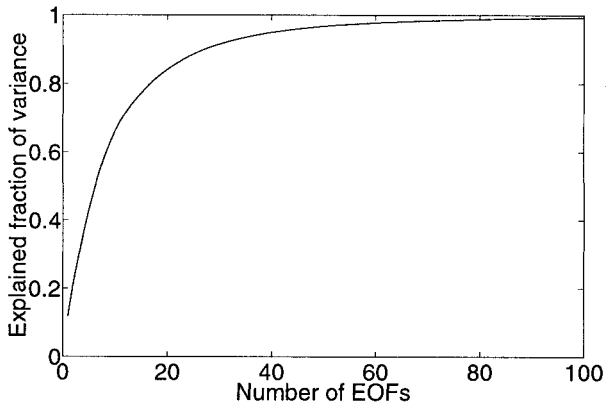

FIG. 1. Cumulative fraction of the variance explained by the leading EOFs of the 500-mb height pentads.

$T = 100$ independent (serially uncorrelated) observation times, (12) and (13) imply $\widehat{N^*_{mm}} = 20.4 \pm 2.8$, an estimate with little bias but moderate scatter. The distribution of $\widehat{N^*_{ef}}$ depends on the eigenvalue spectrum of the covariance matrix. For this example, we will suppose that the eigenvalues form a geometric series with a constant ratio r. For this distribution of eigenvalues to give $N^*_{ef} = 20$, (7) implies that the ratio $r = \exp(-\mu) = 0.904$. To find the scatter of $\widehat{N^*_{ef}}$, we must compute c from (A26). Value of c depends upon N_4 and R, whose dependence on the eigenvalue spectrum is found from (A18) and (A23), respectively. We deduce that $N_4 = (1 - r^4)/(1 - r^2)^2 \approx 10$, $R = (1 - r^2)^{1/2}(1 - r^4)^{1/2}/(1 - r^3) = 0.94$, and hence $c = 0.37$. From (14) and (15), we deduce that $\widehat{N^*_{ef}} = 16.7 \pm 0.7$. This estimate has a moderate bias but a small scatter. For comparison, Fraedrich's formula (10) predicts a much larger scatter in N^*_{ef} of ± 12 and makes no prediction about the bias.

We have verified that our formulas for the bias and scatter of both N^*_{mm} and N^*_{ef} work well over a wide range of T for the two examples in this paper for which $N^* \gg 1$. These examples are (i) the first application discussed in section 3 (Fig. 3 shows the comparison), and (ii) Monte Carlo simulations of the idealized no-dominant-mode case of section 4.

c. Spatial degrees of freedom based on other metrics

From the foregoing, it is clear that if we had defined $E(t)$ based on some other norm $(\psi, \mathbf{D}\psi)$, where $\mathbf{D}$ is some positive definite symmetric matrix, we would have obtained a different value of N^*. Note that this is equivalent to retaining the standard metric and using the field $\mathbf{D}^{1/2}\psi$ instead of ψ. For example, it is often desirable to area-weight observations, in which case D would be a diagonal matrix with elements $D_{ii} = w_i$, the area represented by the ith station. The formulas for N^*_{mm} and N^*_{ef} are still applicable as long as the covariance matrix is similarly weighted so that $C_{ij} = \langle \psi_i(\mathbf{D}\psi)_j \rangle$. Bagrov (1969) suggested one can usually increase the ESDOF by normalizing the time series at each station to have the same (e.g., unit) variance, which may be regarded as an implicit form of weighting. This can slightly improve the power of statistical tests of field significance. In this case, (3) assumes the form

$$N^*_{ef} = \frac{N^2}{\sum_{ij} r_{ij}^2},$$

where r_{ij} is the correlation coefficient between ψ_i and ψ_j. This formula was also derived by Fraedrich et al. (1995).

If the covariance matrix $\mathbf{C}$ is accurately known, one can define a ''prewhitening'' norm,

$$E(t) = [\psi(t), \mathbf{C}^{-1}\psi(t)] = \sum_{k=1}^{N} z_k^2(t).$$

The second equality shows that this is equivalent to defining $E(t)$ using the standard metric, but using the normalized principal components $z_k(t)$ in place of the individual station data. The $z_k(t)$ are uncorrelated and all have the same variance of 1. Hence one obtains $N_{ef} = N$ degrees of freedom, the maximum possible. One might regard this norm as attractive for maximally discriminating between two realizations of a process. However, it should be used with great caution, as it weights PCs that have very small variance (and that are probably subject to large measurement uncertainty) as highly as the dominant PCs.

3. Application to real datasets

In this section, we discuss the estimation of ESDOF for two meteorological datasets. Our focus is on the following issues.

1) How similar are N^*_{mm} and N^*_{ef} in typical applications?
2) How robust are estimates of N^*_{mm} and N^*_{ef} from a limited set of times and locations?
3) Can we identify from the data situations in which we should be particularly cautious in interpreting and using ESDOF?

a. Northern Hemisphere wintertime 500-mb heights

Our primary example is a dataset of Northern Hemisphere wintertime 500-mb heights averaged over 5-day intervals for 48 winter seasons (DJF of 1946–93), based on National Meteorological Center (now the National Centers for Environmental Prediction) daily analyses. The data are given on a $5° \times 7.5°$ lat–long grid northward of 20°N, at a total of 672 grid points. Gridpoint values are area-weighted when calculating spatial averages, and at all grid points, data are converted into anomalies by subtracting the mean annual cycle at that grid point. This dataset is the same as was used by Wallace et al. (1991) but extended for 9 yr.

Figure 1 shows the fraction of the variance explained

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

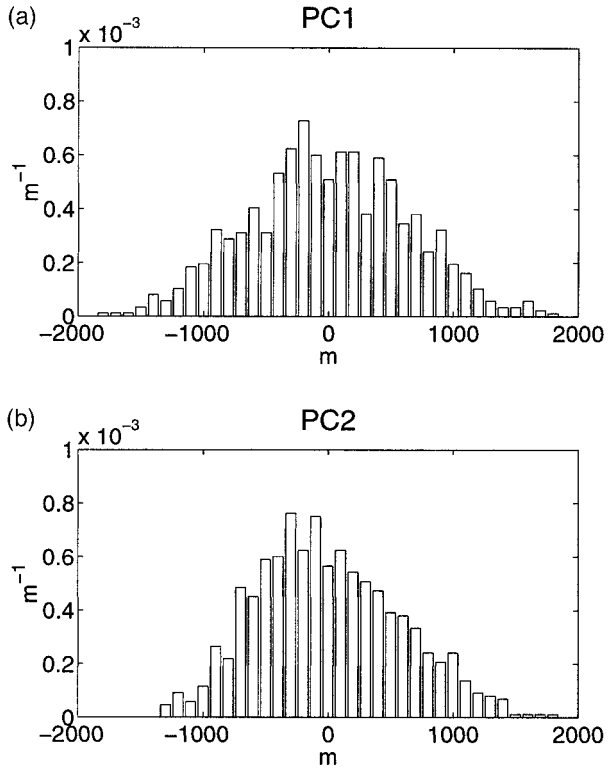

FIG. 2. Frequency distribution of the two leading PCs of 500-mb height pentads.

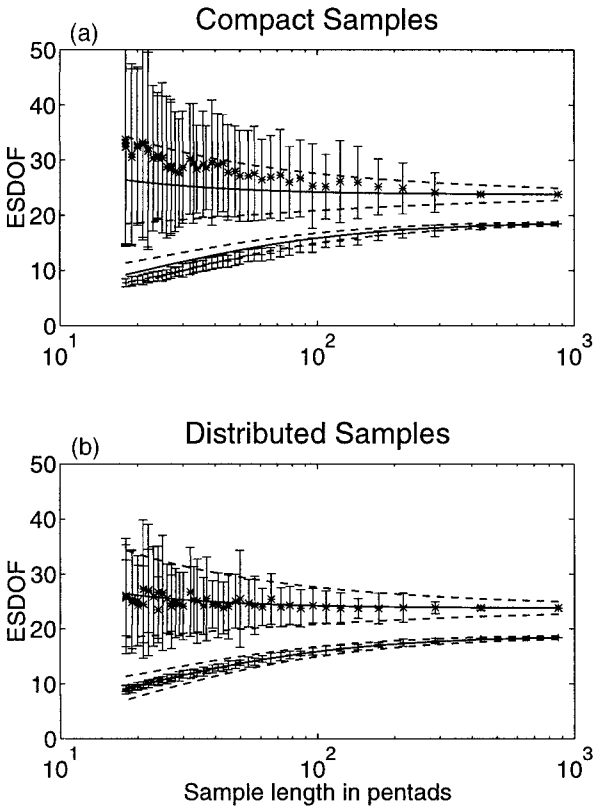

FIG. 3. $\widehat{N^*_{mm}}$ (*) and $\widehat{N^*_{ef}}$ (+) vs sampling periods of different lengths. The crosses and error bars denote ensemble means, bracketed by one standard deviation, of the estimates from different sampling periods of the given length. The solid and dashed lines show the theoretical estimates of the means, bracketed by one standard deviation, of $\widehat{N^*_{mm}}$ and $\widehat{N^*_{ef}}$. Upper panel: Each sample consists of consecutive pentads. Lower panel: The elements of each sample are separated by a constant time interval.

by the leading EOFs. Many EOFs contribute significantly to the variance; the leading 10 modes explain only about 68% of the overall variance and 27 modes are required to explain 90%. The frequency distribution of the two leading PCs are shown in Fig. 2. They exhibit noticeable positive skewness and have kurtosis of 2.6 and 2.7, respectively, both slightly less than the kurtosis of the normal distribution. This is characteristic of all 10 leading PCs. Hence, based on (9), we should expect that the moment matching estimate N^*_{mm} of ESDOF will be slightly larger than the eigenvalue formula N^*_{ef}.

Figure 3 shows the two estimates of ESDOF, $\widehat{N^*_{mm}}$ and $\widehat{N^*_{ef}}$, computed based on data from samples of different lengths. Recall that we use a hat to distinguish an estimate made on the basis of a finite set of sampling times. In the top panel, the samples consist of consecutive intervals. In the bottom panel, they are nonconsecutive and uniformly spaced throughout the entire data record to better ensure independence of the samples. For each possible period of a given sampling length, for example, one season = 18 pentads, the ensemble means and standard deviations of the estimates from all possible samples (each of the 48 winters, in this case) are shown. The theoretical estimates from section 2b of the mean and standard deviation of the estimates are also shown for comparison.

We see that in this dataset, for large sample sizes, $N^*_{mm} \approx 24$ is indeed systematically larger than $N^*_{ef} \approx 19$, due to the slight nonnormality of the leading PCs. We have also performed the same calculation as shown in Fig. 3 with the actual time series of the normalized PCs replaced by independent samples drawn from a standard normal distribution. This does not change the distribution of variance between the EOFs, so it does not change N^*_{ef}. However, it reduces N^*_{mm} so that it is indistinguishable from N^*_{ef} for large sample sizes as expected. The small sample bias and scatter of N^*_{mm} and N^*_{ef} are unchanged. Figure 4 shows χ^2 distributions fit to the observed frequency distribution of spatially averaged variance $E(t)$ with $N^*_{mm} \approx 24$ and $N^*_{ef} \approx 19$ degrees of freedom. The distribution based on the moment-matching estimate gives a better fit, reflecting the bias in N^*_{ef} deriving from the assumption of normality. However, Wallace et al. (1991), who computed the frequency distribution of the covariance and correlation coefficient between all pairs of maps for the same dataset (over a slightly smaller number of years), found that fitted distributions with 20 ($\approx N^*_{ef}$) degrees of freedom (DOF) best matched the observed distributions. This result also held for our longer data record. In the next section, we

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

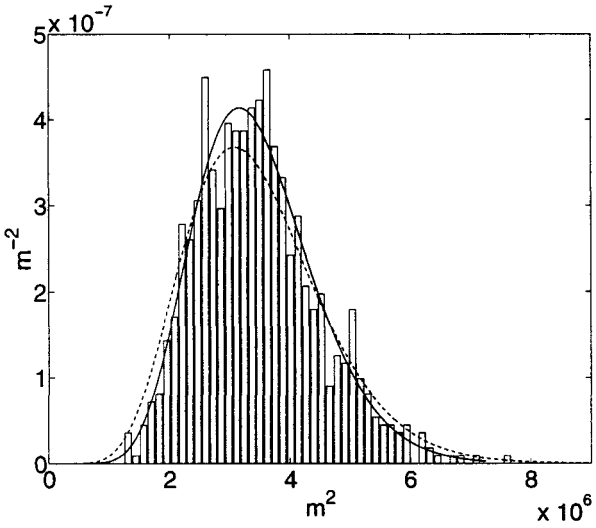

FIG. 4. χ^2 distributions with $N^*_{mm} = 23.8$ (solid) and $N^*_{ef} = 18.5$ (dashed) degrees of freedom superposed on the observed frequency distribution of spatially averaged 500-mb height variance.

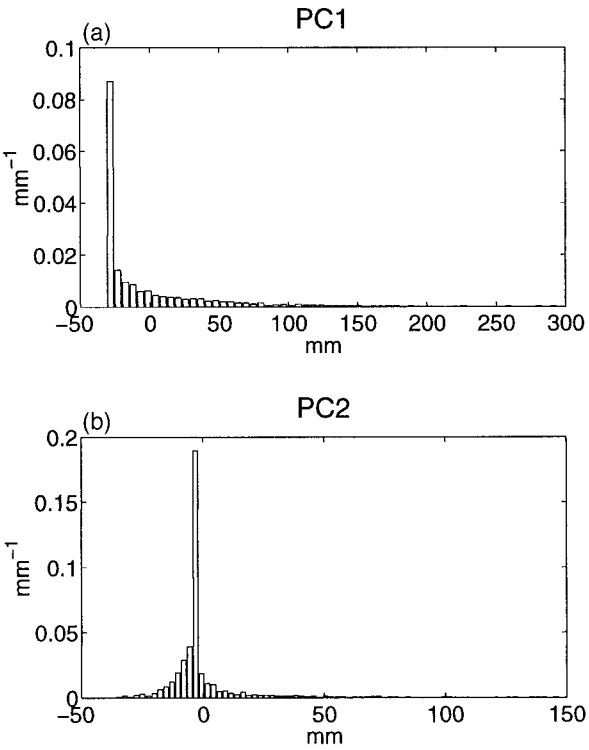

FIG. 5. Frequency distributions of the leading two PCs of daily Swiss wintertime precipitation.

will see that in general, N^*_{mm} ESDOF gives a good fit to variance distributions, while N^*_{ef} gives a good fit to covariance and correlation distributions; that is, the two measures of ESDOF are useful in different contexts.

It is also interesting to note that the first N^*_{ef} EOFs explain about 82% of the variance, which is quite close to the prediction (8) of 86% that was derived assuming that the variance fraction associated with the EOFs decreases geometrically with EOF index. For the dominant EOFs, the assumption of geometric falloff is quite good for this example, though this approximation underestimates the variance fraction in the long tail of small-variance EOFs.

For the distributed samples, the theoretical formulas of section 2b for the bias and scatter of $\widehat{N^*_{mm}}$ and $\widehat{N^*_{ef}}$ agree very well with the bias and scatter calculated directly from the data. For moderate sampling lengths, $\widehat{N^*_{mm}}$ exhibits considerable scatter, increasing with decreasing sample size T proportional to $T^{-1/2}$. When consecutive samples are used, $\widehat{N^*_{mm}}$ also tends to have a high bias for short samples. Since this effect nearly disappears when the samples are distributed through the entire time record, we attribute the high bias to temporal correlation of the 500-mb height fields between pentads. For a short sample, temporal correlation (e.g., a tendency toward anomalously low or high 500-mb heights over the North Pacific may persist over periods of weeks, months, or even years) causes the sample variance of $E(t)$ to underestimate the true variance, which from (1) leads to an overestimate of N^*_{mm}. There are also smaller, but noticeable, changes in the bias of $\widehat{N^*_{ef}}$ when consecutive samples are used.

Fraedrich et al. (1995) considered the effect of spatial resolution on N^*_{ef}. They found that as long as the spacing between grid points is 500 km or less, N^*_{ef} computed for daily 500-mb height data is insensitive to resolution since the leading EOFs vary only on much larger length scales. We repeated this type of analysis for our pentad data (which is spatially smoother) and found that N^*_{ef} and N^*_{mm} are insensitive to spatial resolution even with a grid spacing as large as 1500 km.

b. A precipitation dataset

As our second example, we consider a dataset of daily precipitation at 113 rain gauge locations throughout Switzerland, for the winter seasons DJF 1961–90 (Widmann and Schär 1997). In this dataset, most (80%) of the variance is concentrated in the three leading EOFs. The frequency distribution of the PCs for the leading two EOFs is shown in Fig. 5. PC1 is highly positively skewed with a long tail and far from normal, as would be the rainfall distribution at any individual rain gauge. PC2 is not so skewed, but it also has very long tails. Over the entire data record, the two estimators of ESDOF are $N^*_{mm} = 0.24$ and $N^*_{ef} = 2.4$. Here N^*_{mm} is much smaller than N^*_{ef} due to the large kurtosis (long tails) of the distribution of the leading PCs. PDFs and cumulative distribution functions (CDFs) of the corresponding χ^2 distributions are compared with the observed frequency distribution of spatially integrated daily rainfall variance in Fig. 6. Neither χ^2 distribution fits the observed frequency distribution that well. However, as we can see by looking at the CDFs, the χ^2 distribution based on

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

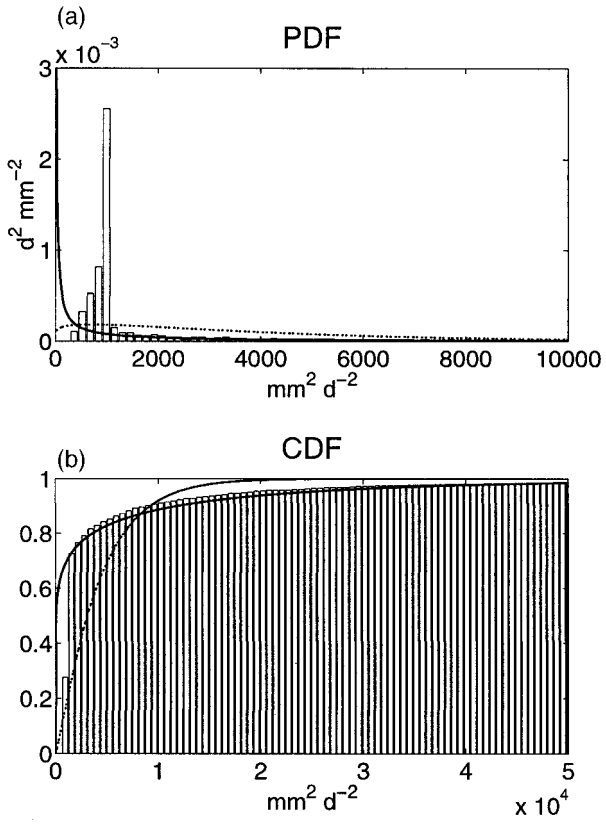

FIG. 6. Frequency distribution and CDF (bin size is different) for spatially integrated variance of daily Swiss wintertime precipitation (bars) along with the PDF and CDF of χ^2 distributions with $N^*_{mm} = 0.24$ (solid lines) and $N^*_{ef} = 2.4$ (dotted lines) degrees of freedom.

N^*_{mm} DOF does satisfactorily fit the upper tail of the variance distribution. Since N^*_{mm} is less than one, it is obviously not literally interpretable as a number of degrees of freedom. The value of N^*_{ef} is a better indicator of the number of dominant spatial modes in the system.

If monthly rainfall totals are analyzed instead, N^*_{mm} rises to 0.9, while N^*_{ef} drops to 1.7. The monthly average rainfall frequency distribution is quite skewed, but less so than the daily data. As longer averaging intervals are used, we would expect the frequency distribution of rainfall to tend toward a normal distribution and the two estimates of ESDOF to converge.

4. Use of ESDOF for significance testing

An attractive application of ESDOF is to assess whether two realizations of a field are either significantly similar or significantly different in a statistical sense. For example, we may wish to compare the field with some prediction of that field at that time. Alternatively, we might want to compare the spatial structure of the field at two different times or a particular realization of a field with its mean value.

Wallace et al. (1991) [and earlier, Lorenz (1969), van den Dool and Chervin (1986), and others] have discussed ways of measuring the similarity or difference between two realizations of a field. We might compute the rms difference between the realizations, averaged over all grid points. Alternatively, we might compute the spatial correlation between the realizations. We expect that if the field has only a few ESDOF, then two randomly chosen realizations of the field will have a much broader distribution of spatial correlations and rms differences than if the field has many ESDOF. In particular, we might hope based on the examples shown in the previous section that the statistics of the field could be approximately modeled with a combination of N^* independent random variables with standard normal distributions. This would permit classic statistical tests based on the chi-squared or Student's t-distributions to be applied to these problems. To justify this approach, we must determine whether the resulting PDF provides a conservative estimate of the frequency of outliers; this will be the goal of this section.

An assumption of most of this section is that the dominant principal components of the EOF decomposition of the field ψ are approximately normally distributed, so that both the moment-matching and eigenvalue formulas for ESDOF are equal to a common value N^*. This will be true if the gridpoint data are normally distributed, but not, for example, for the Swiss rainfall dataset. At the end of the section, we consider examples with highly nonnormal principal components and find that our statistical tests derived for normal PCs are still useful, but that the appropriate estimate of ESDOF is a combination of N^*_{mm} and N^*_{ef} that depends on the test.

a. Difference of a realization from the mean

As discussed in section 2, the PDF of the spatially integrated variance of a field, $E(t)$, is the generalized χ^2 distribution (2), that is, a variance-weighted sum of squares of unit normal random variables. In that section, we matched the mean and variance of this distribution to those of a standard χ^2 distribution with N^* degrees of freedom, scaled by a variance factor $\langle a^2 \rangle$, hoping that the PDFs of the resulting distributions would also approximately match:

$$E(t)/\langle E \rangle \sim \chi^2(N^*)/N^*. \tag{16}$$

Here, the notation $\sim$ means "having approximately the same PDF as." If these PDFs match reasonably well, we can use the 95th percentile of $\chi^2(N^*)$ to derive a confidence interval for $E(t)$.

We now investigate how well the PDFs of $E(t)$ and the χ^2 distribution match in three diverse examples.

1) MANY EOFS WITH COMPARABLE VARIANCE

We might expect the PDFs to match quite well if many of the PCs explain comparable fractions of the overall variance. An idealized example is shown in Fig. 7. Sup-

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

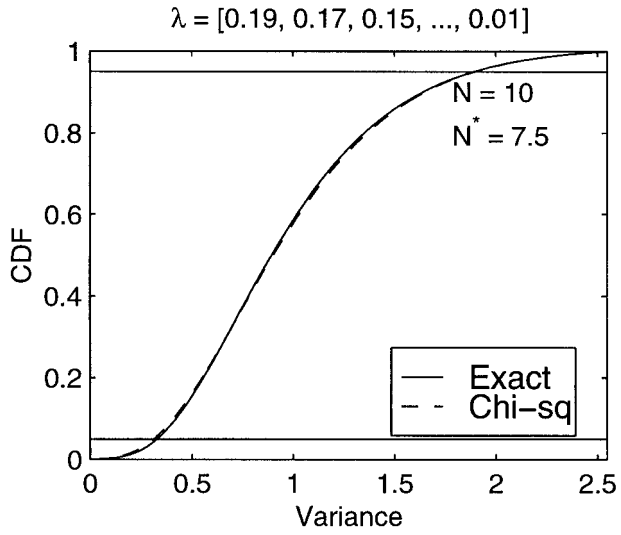

FIG. 7. Comparison of the exact CDF of a generalized χ^2 distribution (with $N = 10$ principal components with variance fractions λ) with a scaled χ^2 random variable with N^* DOF and the same mean and variance. Horizontal lines indicate 5th and 95th percentiles.

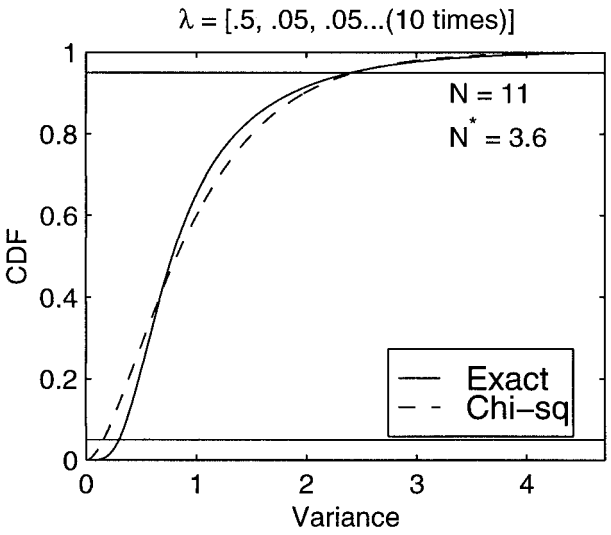

FIG. 8. As in Fig. 7, for a field with 11 grid points and one dominant EOF.

pose that an EOF analysis of a time series of a field observed at 10 grid points yields 10 principal components with variances $\lambda_k = 0.19, 0.17, \ldots, 0.01$ (since these variances sum to 1, they are also the variance fractions f_k associated with the 10 modes). No subset of the PCs is dominant. If each principal component is normally distributed, $E(t)$ has a generalized χ^2 distribution given by (2). From (3), we find that there are $N^* = 7.5$ ESDOF.

Rather than comparing the PDFs directly, we compare the CDF, for which threshold values corresponding to, for example, the 5th and 95th percentiles of the distributions, can be directly compared. In Fig. 7, the CDF of the spatial variance E, given by the generalized χ^2 distribution (2), is compared to the CDF of the matched standard χ^2 distribution. The CDFs of the generalized χ^2 distribution are computed as integrals of the PDF, which is found in turn as a discrete Fourier transform of the moment generating function (Hogg and Craig 1978, p. 107) of the two distributions (the CDF can also easily be computed using a Monte Carlo approach). The 5th and 95th percentiles of each distribution are the intersection of the two horizontal lines with the corresponding CDF in Fig. 7.

The exact (generalized χ^2) and matched $\chi^2(N^*)$ distributions have virtually identical CDFs. Close examination shows that the $\chi^2(N^*)$ distribution has a very slightly lower probability of extremely large values, and a very slightly higher probability of extremely small values than the exact distribution, but these differences are so small that a statistical test can be confidently based on the matched χ^2 distribution with N^* ESDOF.

2) A FEW DOMINANT EOFS

An opposite extreme is if one EOF is dominant, such as in the example shown in Fig. 8. In this example, we consider a field measured at 11 points for which one EOF explains half the variance, and the other 10 EOFs each explain only 5% of the variance. This is a particularly difficult case for matching to an effective χ^2 distribution, since the tails of the exact distribution will be heavily dominated by contributions from the one large EOF, while the effective chi-squared distribution must weight all its contributing normal random variables equally.

The match is less good, but still reasonable, especially in the crucial upper tail. The exact and $\chi^2(N^*)$ distributions have almost identical 95th percentiles and agree closely even farther out into the upper tail. The discrepancies in the lower tail, though larger, are not that important for significance testing, since we would be unlikely to want to test whether a given realization was significantly closer than expected to the mean.

We also examined a case with a few dominant EOFs, as commonly found in climate statistics applications. We took $\lambda_k = 0.5, 0.25, \ldots, 2^{-k}, \ldots$, truncating after 10 EOFs. The results (not shown) are similar to the case of one dominant EOF, but with considerably closer agreement between the exact and $\chi^2(N^*)$ distributions.

Our conclusion is that the matched $\chi^2(N^*)$ distribution is a remarkably accurate and convenient tool for testing the significance of departures of a field from its mean, as long as the leading principal components of the field are approximately normally distributed.

b. Difference of two realizations

Are two realizations of a field significantly more or less different than would be two randomly chosen re-

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

alizations, that is, are they ''analogs'' or ''antilogs''? One measure of this is the spatially integrated, squared difference between the two realizations, normalized by dividing by $\langle E \rangle$. If their normalized principal components are y_k and z_k, with difference δ_k, this difference can be written

$$\Delta = \sum_{k=1}^{N} f_k \delta_k^2, \tag{17}$$

where, as before, f_k is the variance fraction explained by EOF k. To test this for statistical significance, we can compare Δ to the percentiles of the distribution of the difference between two independent, randomly chosen instances of the field, for which δ_k would have mean zero and variance equal to 2 (the sum of the variances of y_k and z_k). Thus $2^{-1/2}\Delta$, like the normalized spatially integrated variance E, is distributed like a generalized chi-squared random variable with parameters f_k, and its PDF is similar to that of $\chi^2(N^*)/N^*$.

The results of the previous subsection imply that a significantly larger than expected Δ can be tested for using a one-sided chi-squared test with N^* DOF. To test for a smaller than expected Δ, one can also use a one-sided test based, for instance, on the 5th percentile of a chi-squared distribution with N^* DOF. Since the 5th percentile of the exact distribution is slightly farther from zero than for the equivalent chi-squared distribution, a test based on $\chi^2(N^*)$ gives a conservative significance threshold for Δ.

c. Field covariance and correlation

Another measure of the similarity of two realizations of a field is their ''anomaly correlation,'' that is, the spatial correlation of their departures from the time mean of the field. Unfortunately, we have not found a simple way to calculate the PDF of this correlation. We can, however, argue that this PDF should be quite similar to that based on a t test with N^* DOF, so that again an approximate significance test can be based on a standard test statistic.

Our argument is based on a calculation of the PDF of the covariance of two independent, randomly chosen realizations of the field. We normalize the covariance by the variance of the field. With notation as before, this covariance is

$$C = \sum_{k=1}^{N} f_k y_k z_k. \tag{18}$$

We can write $y_k z_k$ as follows:

$$y_k z_k = [(y_k + z_k)^2 - (y_k - z_k)^2]/4. \tag{19}$$

If y_k and z_k are independent standard normal random variables, their normalized sum $r_k = 2^{-1/2}(y_k + z_k)$ and difference $s_k = 2^{-1/2}(y_k - z_k)$ are also independent standard normal random variables. Hence, we can write C as a generalized chi-squared random variable,

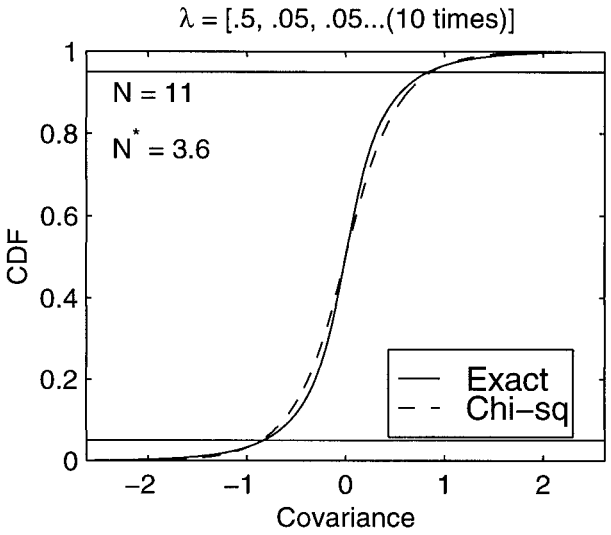

FIG. 9. The CDF of the covariance between two randomly chosen realizations of the field used for Fig. 8, with one dominant EOF. Approximation based on N^* ESDOF is dashed.

$$C = \sum_{k=1}^{N} (f_k/2)(r_k^2 - s_k^2). \tag{20}$$

Even though the coefficients in this quadratic form that multiply the s_k^2 are negative, the PDF can be computed using the same numerical method as before. Note, too, that C is half of the difference between two generalized chi-squared random variables with the same distribution as the spatially integrated variance. Since the CDF of the matched $\chi^2(N^*)$ distribution is very similar to that of the exact CDF of the spatially integrated variance, we can expect that the CDF of the difference of two $\chi^2(N^*)$ variables to be similarly close to the exact difference,

$$C \sim C^* = (X_1 - X_2)/(2N^*), \tag{21}$$

where X_1 and X_2 are independent and both distributed as $\chi^2(N_*)$. Figure 9 compares the distributions of C and C^* in the ''worst case'' of Fig. 8 in which there is one dominant EOF. Even in this case, the CDFs of the covariance are quite similar and, for more geophysically likely distributions of variance among the EOFs, the distribution based on N^* ESDOF is extremely close to the exact distribution. Figure 1 of Wallace et al. (1991) shows a similar comparison for the 500-mb height dataset of section 3. They also found close agreement between the actual PDF and a fitted distribution based on Monte Carlo simulation with 20 ($\approx N^*_{mm}$) = ESDOF.

We could test the similarity between two realizations of a field using the covariance as a test statistic. The main complication is that percentiles of the distribution of the difference of two chi-squared random variables is not tabulated in books or standard software packages. It is more straightforward to use a t test based on the spatial correlation coefficient R between the realizations.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

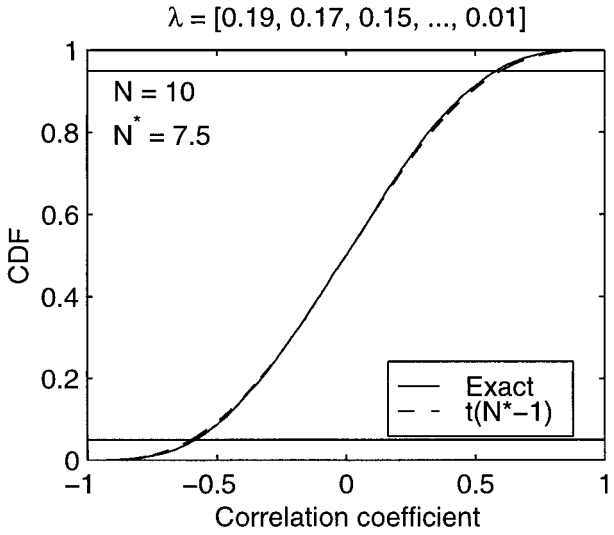

FIG. 10. Exact CDF of the correlation between two randomly chosen realizations of the field used for Fig. 7, with the variance partitioned among many dominant EOFs.

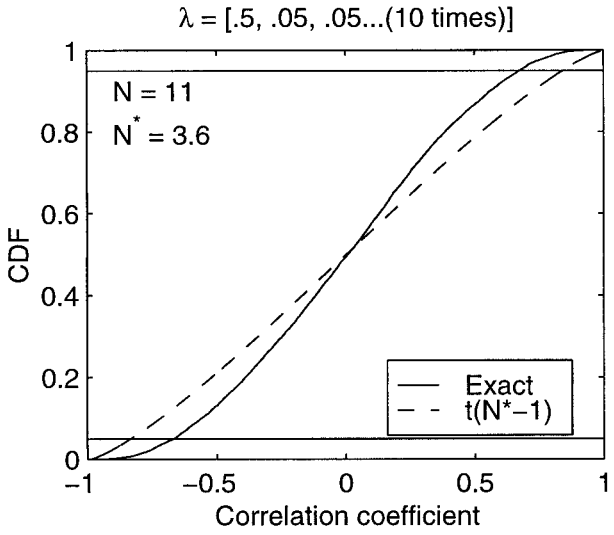

FIG. 11. Exact CDF of the correlation between two randomly chosen realizations of a field with 11 grid points and the variance partitioned among the EOFs as in Fig. 8, with one dominant EOF.

The correlation coefficient is the spatial covariance normalized by dividing by the square root of the product of the variances of the two realizations. Note that unlike in time series analysis, the covariance is computed without subtracting the spatial mean anomaly from each realization. This is because the null hypothesis is that both realizations are random cases of a field whose mean value at each grid point, and hence whose true spatial mean, is accurately known from our long time record of realizations of that field. Consider the PDF of the R between two randomly chosen realizations of a field. Since both the variances and the covariance are well represented using distributions based on N^* ESDOF, the same should hold for R. Hence, we anticipate that the PDF of $\tau = R(N^* - 1)^{1/2}/(1 - R^2)^{1/2}$ has an approximate t distribution with $N^* - 1$ DOF (we use $N^* - 1$, rather than $N^* - 2$, because the time mean of the field is known, adding one DOF to the classic test in which the means of each of the two correlated fields must be estimated as a sample mean).

Wallace et al. (1991) showed that for the 500-mb height dataset, a PDF of correlation generated from a t distribution based on 20 ($\approx N^* - 2$) ESDOF matched their observed PDF of R very closely. One would deduce from this that two realizations of a 500-mb height field (or a forecast and an analysis of 500-mb height) are positively correlated at a 95% (one sided) significance level if $\tau > 1.73$, that is, if their pattern correlation $R > 0.37$.

Here we examine the two examples presented earlier in this section. We use 10 000 trials of Monte Carlo simulation of the principal component amplitudes to generate the ''exact'' distribution of R. For the example with many dominant modes (Fig. 10), the t distribution matches the Monte Carlo simulation nearly perfectly. For the example with one dominant mode, the t distribution shows a considerably higher probability of large correlations than the Monte Carlo simulation (Fig. 11). The t test would produce conservative estimates of significance levels. In no example that we looked at did the t test based on $N^* - 1$ DOF overestimate the significance of a particular correlation, so it is safe, and in most cases quite accurate, to use this test as long as time series of the dominant PCs are approximately normally distributed.

d. Nonnormally distributed data

How well do these significance tests based on N^* degrees of freedom work if the grid points, and hence the PCs, are not normally distributed? The two extreme examples discussed below show that even in this case, the tests can still give good significance bounds, as long as N^* is chosen appropriately for the test, as summarized in Table 1. For all of the tests, the appropriate N^* lies somewhere in the range between N^*_{mm} and N^*_{ef}.

First, we consider the test discussed in section 4a for significance of the difference of a realization from the mean. This test is based on the CDF of the spatially integrated variance $E(t)$. We examine this CDF in the two extreme cases that the PCs are (a) exponentially distributed and (b) uniformly distributed, rather than normally distributed. These are examples of distribu-

TABLE 1. ESDOF appropriate for various statistical tests.

Significance test	Appropriate N^*
Difference of realization from mean	N^*_{mm}
Difference of two realizations	$(2/N^*_{mm} + 2/N^*_{ef})^{-1}$
Covariance and correlation	N^*_{ef}

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

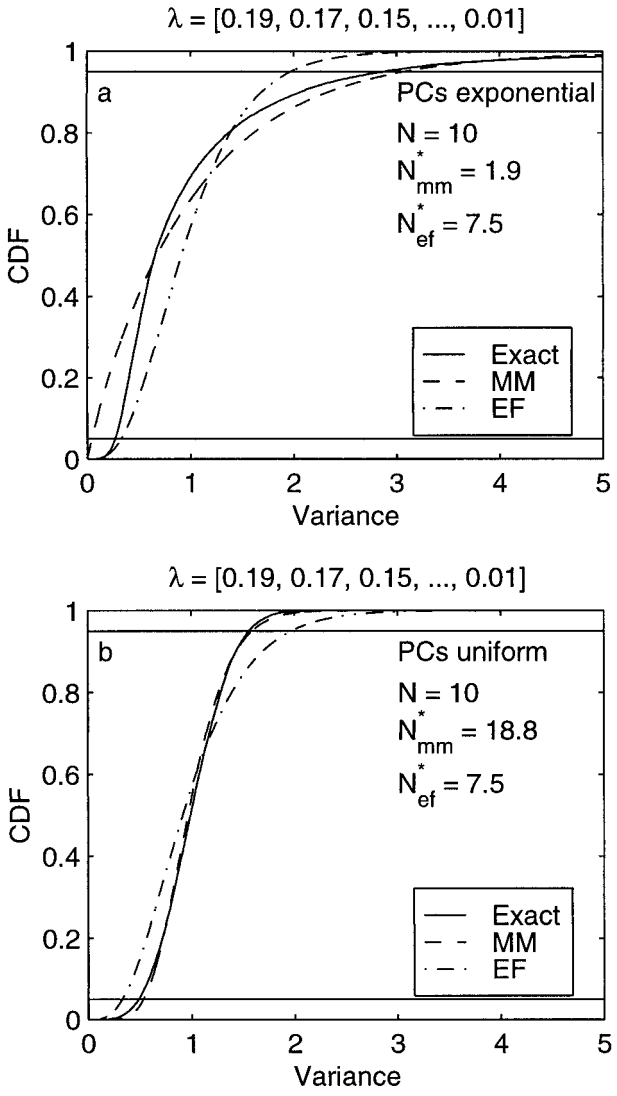

FIG. 12. As in Fig. 7, but with the individual principal components (a) exponentially distributed and (b) uniformly distributed.

tions with particularly large and small kurtosis, respectively, compared to a normal distribution. A distribution with large kurtosis (long tails) can be expected to lead to a broader distribution of $E(t)$, given a fixed partitioning of variance between the PCs. In both cases, the PCs are normalized to zero mean and unit variance, then we use 10 000 Monte Carlo trials to compute the CDF of $E(t)$, which is the sum of the squares of the PCs, weighted by their variances λ_k.

We consider one of the examples discussed earlier in this section, that of 10 principal components with variances 0.19, 0.17, . . . , 0.01. With normally distributed PCs, Fig. 7 showed that the exact CDF of $E(t)$ was quite well approximated by a χ^2 distribution with $N^* = 7.5$ ESDOF. Figure 12 shows the corresponding results for exponentially and uniformly distributed PCs. This does not affect N^*_{ef}, which depends only on the partitioning of the variance between the EOFs. However, the moment-matching estimate N^*_{mm} is much lower for the exponentially distributed PCs and much higher for the uniformly distributed PCs. This is because, as discussed in section 2, N^*_{mm} differs from the eigenvalue estimate of 7.5 by a factor related to the kurtosis. Hence, the χ^2 distribution that is matched to the exact distribution of $E(t)$ is quite different in these two cases, with a much broader variation around its mean for the exponentially distributed case. For the exponentially distributed PCs, the fit of $\chi^2(N^*_{mm})$ is only fair for small values of variance but is very good in the upper tail. The 95th percentile is very well estimated. The overall fit is nearly as good for the uniformly distributed PCs as for normally distributed PCs throughout the distribution. In both cases, the χ^2 distribution based on the eigenvalue formula is an unsatisfactory fit to the true distribution. Both of these cases are quite extreme compared to most geophysical datasets, but the fit based on N^*_{mm} ESDOF is good enough so that an approximate significance test may safely be based on $\chi^2(N^*_{mm})$.

Second, we consider the test discussed in section 4b for significance of the normalized difference Δ of two realizations. The PDF of Δ is determined by the PDFs of the differences $\delta_k = y_k - z_k$ between the PCs in the two realizations. These differences have a PDF that is intermediate between the PCs of one realization and a normal distribution. In fact, if y_k and z_k are independently chosen from a distribution with kurtosis κ, then the kurtosis κ_δ of δ_k can easily be shown to be $(3 + \kappa)/2$, which is halfway between the kurtosis of the individual distributions and the kurtosis of a normal distribution. Hence, applying the moment-matching arguments of section 2 to estimate the ESDOF of Δ, we find from (9) that

$$N^*_\Delta = 4N^*_{ef}/(1 + \kappa_\delta) = (2/N^*_{ef} + 2/N^*_{mm})^{-1} \qquad (22)$$

is the harmonic mean of N^*_{ef} and N^*_{mm}.

Last, we consider the normalized covariance C. Equation (21) gave the approximate distribution of C in terms of the difference of two $\chi^2(N^*)$ random variables. We can use moment matching to calibrate this formula for non-normally distributed PCs. The mean of C is zero, so we match the variance. From (18), noting that y_k and z_k are mutually independent and have unit variance, we find

$$\mathrm{var}(C) = \mathrm{var}\left(\sum_{k=1}^{N} f_k y_k z_k\right) = \sum_{k=1}^{N} f_k^2 \,\mathrm{var}(y_k)\,\mathrm{var}(z_k)$$
$$= \sum_{k=1}^{N} f_k^2 = 1/N^*_{ef}, \qquad (23)$$

that is, that now the *eigenvalue* estimate of N^* may be the appropriate one. We might anticipate that the same would hold for the distribution of the correlation coefficient. Indeed, Fig. 13 plots the CDF of the correlation coefficient between two independent realizations of a

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

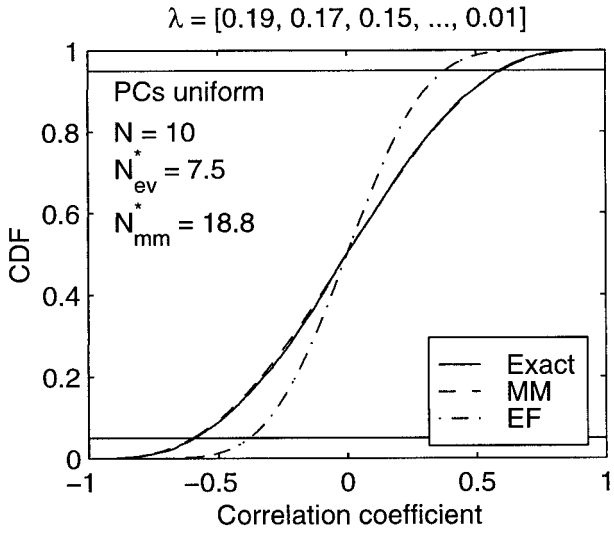

FIG. 13. CDF of the correlation coefficient and of t distributions corresponding to $N^*_{mm} - 1$ and $N^*_{ef} - 1$ DOFs, when the individual principal components are uniformly distributed and the variance is partitioned as in Fig. 7.

field in which the gridpoint data are uniformly distributed and the variance is partitioned among the 10 EOFs as before. This CDF is generated from 10 000 Monte Carlo trials. Superimposed upon this exact CDF are t distributions based on $N^*_{mm} - 1$ and $N^*_{ef} - 1$ DOF. We see that the t distribution based on $N^*_{ef} - 1$ DOF fits the exact distribution very well, while that based on $N^*_{mm} - 1$ DOF is not a good fit.

e. Significance of temporal correlations between two time-varying fields of different spatial structure

Livezey and Chen (1983) discussed the concept of field significance of temporal correlations between two fields, for example, gridded 500-mb height and sea surface temperature. They looked at the temporal correlation coefficients between the 500-mb height at each grid point and the SST at the same grid point. Even if the fields are in fact independent, some of these correlation coefficients are likely to be large enough to test positive for significance at a given significance level (e.g., 95%). Livezey and Chen proposed comparing the fraction of correlation coefficients that are significant to a probability distribution of the same fraction computed for two fields that have the same spatial correlation structure as the given fields but are not related to each other. A Monte Carlo method for determining this distribution is to correlate realizations of the first field with realizations of the second field from different randomly chosen years. They called the overall correlation between the fields ''significant'' if the fraction of significant gridpoint correlations exceeded the 95th percentile of the Monte Carlo distribution. They suggested that the Monte Carlo distribution was broader if the fields had fewer spatial degrees of freedom but did not suggest a formal procedure for determining an appropriate ESDOF.

We propose a related procedure with the same aim that does not require a Monte Carlo simulation; instead it uses the ESDOF of the individual fields to determine the distribution of a test statistic. Unlike Livezey and Chen's approach, our procedure does not require that the two fields be tabulated at the same grid points. Denote the two fields by $\phi(t)$ and $\psi(t)$, and assume they are observed at grid points $i = 1, \ldots, M$ and $j = 1, \ldots, N$ respectively, at times $t = 1, \ldots, T$. Assume further that their gridpoint values have been normalized to zero mean and unit standard deviation. Let

$$\hat{C}^{\phi\psi}_{ij} = \widehat{\phi_i \psi_j}$$

be the empirical matrix of correlation coefficients between the two fields at all possible pairs of grid points, based on the T observations. A measure of the field correlation between ϕ and ψ is their integrated squared cross covariance (ISCC):

$$\text{ISCC} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (\hat{C}^{\phi\psi}_{ij})^2. \tag{24}$$

The ISCC lies between 0 and 1. If a lot of cross-correlation coefficients are large, then the ISCC will also be large. If the ISCC is close to one, then the covariances between ϕ and ψ explain almost all the variance of the individual fields. The ISCC plays a similar role to Livezey and Chen's fraction of significant correlation coefficients. However, it also includes correlation of ϕ at each location with ψ at other locations, that is, teleconnections between the two fields.

Consider two fields, ϕ and ψ, that are unrelated. For any finite sampling period, their empirical cross correlations will still be nonzero, so their ISCC will be positive. If their ISCC is computed from T independent samples, one may easily show that its expected value is

$$\langle \text{ISCC} \rangle = T^{-1}. \tag{25}$$

Routine but lengthy computation shows that for large T,

$$\text{var(ISCC)} = 2/(N^*_{\phi\psi} T^2), \tag{26}$$

where

$$N^*_{\phi\psi} = 2N^*_{\phi} N^*_{\psi}, \tag{27}$$

and N^*_{ϕ} and N^*_{ψ} are the ESDOF for the two fields computed using the eigenvalue formula. By moment-matching arguments, the distribution of the ISCC can be approximated by $\chi^2(N^*_{\phi\psi})/(N^*_{\phi\psi} T)$, allowing one to compute approximate confidence intervals for testing whether the ISCC is significantly larger than one would expect by chance, that is, whether the two fields are significantly correlated. Since the DOF will typically be very large, the χ^2 distribution will have a very narrow peak, and an ISCC only fractionally larger than T^{-1} will be statistically significant. Hence, a key issue in practical

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

use of this significance test (as with Livezey and Chen's) is deciding whether observations at successive times are really independent and, if not, how best to estimate the true effective number of independent observation times.

5. Relation to effective sample size for serial correlation

There is a close connection between formulas for the ESDOF of a field and notions of effective sample size (ESS) for temporally correlated time series. This section has two purposes. First, we will show the connection between ESS and ESDOF. Second, we will mention some formulas in the literature for ESS for testing significance of correlations between time series. This is a common problem in atmospheric science, yet the relevant formulas and issues are mainly buried in old articles in statistical journals and are not widely appreciated.

The connection between ESS and ESDOF can be seen by applying the arguments of section 2 to T successive observations from a zero-mean unit variance stationary Gaussian random process X_i, rather than to a realization of a spatial field. Different ''realizations'' would now correspond to independent sets of T observations. For two realizations to be independent, each set of observations would have to be taken at a sufficiently long time from other sets so that temporal correlations between observations from different sets would be negligible. In developing the theory of ESS below, the temporal correlation structure is initially treated as a known. However, in practice it must be estimated. When we considered the ESDOF of spatial fields, the spatial covariance matrix could be accurately estimated from a few hundred independent observations of the spatial field. In contrast, we have only the one time series from which to estimate its correlation structure and ESS. This would be impossible were it not for the stationarity of the time series, which drastically simplifies its correlation structure. The estimation uncertainty in ESS may be considerable, particularly if the time series has substantial spectral power at periods of T and longer.

Consider the distribution of the sample variance of the T observations,

$$E := \sum_{i=1}^{T} X_i^2.$$

The value of E will depend on which T observations were taken, so it has a PDF, whose mean and variance can be computed as in section 2. Let λ_k be the eigenvalues of the lag-correlation matrix, whose elements are the lag correlations

$$C_{ij} = \langle X_i X_j \rangle = \rho_{|i-j|}. \tag{28}$$

Then

$$\langle E \rangle = \sum_{k=1}^{T} \lambda_k,$$

$$\mathrm{var}(E) = 2 \sum_{k=1}^{T} \lambda_k^2.$$

Moment-matching implies that $E/\langle E \rangle$ has an approximate distribution $\chi^2(T^*)/T^*$ where T^* is defined by the eigenvalue formula (3) and can be regarded as the effective sample size of the T observations after accounting for their temporal correlation.

The λ_k are sometimes called the singular spectrum for window length T (Vautard and Ghil 1989). They are approximately equal to the power spectrum of the time series at the frequencies $f_k = k/(2T\Delta t)$, where Δt is the time between observations (Vautard and Ghil 1989). Thus (3) implies that a random process with most of its spectral power concentrated in a localized frequency band (e.g., bandpass or long-memory processes) will have a effective sample size much smaller than T.

To see the equivalence to more commonly used formulas for T^*, we can substitute (28) into (5), which expresses T^* in terms of the C_{ij} instead of the eigenvalues:

$$T^* = \frac{\left(\sum_{i=1}^{T} C_{ii} \right)^2}{\sum_{i,j=1}^{T} C_{ij}^2}, \qquad = \frac{T}{\sum_{\tau=-(T-1)}^{T-1} \left(1 - \frac{|\tau|}{T} \right) \rho_\tau^2}. \tag{29}$$

This formula dates back in the statistical literature at least to Bayley and Hammersley (1946) and is mentioned in some current texts on spectral analysis, for example, Percival and Walden (1993, p. 254).

Following the approach of section 4b of looking at the distribution of the sample covariance between two unrelated time series, one can derive an analogous formula for the ESS appropriate for significance tests of the correlation between two times series X_i and Y_i with different autocorrelation sequences ρ_τ^X and ρ_τ^Y:

$$T_{XY}^* = \frac{T}{\sum_{\tau=-(T-1)}^{(T-1)} (1 - |\tau|/T) \rho_\tau^X \rho_\tau^Y}. \tag{30}$$

Orcutt and James [1948, Eq. (4)] quote an equivalent formula. Davis (1976, p. 252) rederived this formula in a climate statistics application. The special case of this formula for red-noise (Markov) processes was derived even earlier by Bartlett (1935). If r_X is the lag-one autocorrelation of X_i and similarly for Y, and if r_X^T, $r_Y^T \ll 1$, then

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

$$T^*_{XY} = T\frac{1 - r_1 r_2}{1 + r_1 r_2}. \quad (31)$$

Formulas (29) and (30) are sometimes said to define an ESS of order 2 (Kikkawa and Ishida 1988), as they are appropriate for significance tests involving second-order moments of the original time series (variance, covariance, and correlation). Formulas for computing the variance of the sample mean (i.e., the first-order moment) of T observations from a time series can be expressed in terms of a different ESS (of order 1); see, for example, Jones (1975) and Thiébaux and Zwiers (1984). For instance, for red noise with lag-one autocorrelation r, the ESS of order 1 is $T^* = T(1 - r)/(1 + r)$ [see Jones (1975), which gives the discrete time version of Leith's (1973) commonly used ESS formula], while (31) shows that the ESS of order 2 is $T^* = T(1 - r^2)/(1 + r^2)$, which is up to twice as large as the Jones–Leith ESS when r approaches 1, that is, if the autocorrelation is strong.

The above definitions of ESS involves the actual and usually unknown covariance structure of the random process. Now we discuss estimation of the ESS from an observed time series. An approximation of ESS to within 10%–20% will suffice for most practical purposes. Several approaches have been proposed, all requiring care on the part of the user. First, we can use the empirical autocorrelations to estimate the true autocorrelations in the ESS formulas. The ''biased'' estimator of autocorrelation (Percival and Walden 1993, p. 191) should be used. To reduce the noise in this estimator if necessary, one can multiply the resulting autocorrelation sequence by a lag window (Percival and Walden 1993, p. 238). Alternatively, if the autocorrelation in the time series is primarily on timescales much shorter than T, one can divide the time series into blocks and treat each block as an independent realization of the time series. We could estimate the temporal covariance matrix, and hence its eigenvalues, as in the spatial case by averaging over the blocks. This is singular spectral analysis (Vautard and Ghil 1989). Last, we could fit a low-order autoregressive process (e.g., red noise) to the empirical autocorrelation sequence and estimate the ESS from the autocorrelation sequence of the fitted process (Orcutt and James 1948). The latter approach can minimize the random error in our estimate of ESS at the expense of introducing a possible bias if the fitted process is not a good approximation to the true random process.

6. Conclusions

We have comprehensively investigated two measures of the effective number of spatial degrees of freedom in a time-varying field of data, calculated respectively from moment-matching (N^*_{mm}) and from an eigenvalue formula (N^*_{ef}) based on EOF analysis, specifically the partitioning of variance between the EOFs. The two measures are the same if the data at each station are normally distributed and the field has been sampled at a large number of independent (serially uncorrelated) times. Each has both strengths and weaknesses, depending on the application.

We find that for short samples (of length less than a few hundred independent sampling times) estimates of N^*_{mm} exhibit large scatter. Estimates of N^*_{ef} tend to be biased low unless the number of samples is many times as large as N^*_{ef}. This bias is due to insufficient sampling to distinguish the temporal variability of the different spatial modes. In general, our experience is that if the number of independent observation times is much larger than either of these measures of N^* (in practice, $10N^*$ or more), then the sample statistics are sufficient to define all of the dominant spatial modes (EOFs) accurately, but if the number of observation times is comparable to N^*, then some leading EOFs will not be accurately determined.

If the field exhibits substantial autocorrelation from one sampling time to the next, the estimated N^*_{mm} is biased high for short samples. If the frequency distribution of gridpoint data has long tails, such as for precipitation data, N^*_{mm} is considerably smaller than N^*_{ef}, while if the tails are truncated compared to a normal distribution, N^*_{ef} is smaller than N^*_{mm}; we quantify this in terms of the kurtosis of the principal components.

We find that both of these measures are useful for statistical significance testing. For testing whether a given realization of a field is significantly different from the mean, we find that a chi-squared test with N^*_{mm} degrees of freedom works surprisingly satisfactorily regardless of how the variance is partitioned between the EOFs and the probability distribution of the station data. For testing the significance of the correlation between two realizations of the field, such as a forecast and a verification, we show that a t test with $N^*_{ef} - 1$ degrees of freedom works well. A method is also proposed for testing whether two fields with different spatial correlation structure have a significant simultaneous correlation to eachother. This leads to a chi-squared test with DOF dependent on the product of the N^*_{ef}'s of the two fields.

Last, we have briefly discussed the close analogy between effective spatial degrees of freedom of a time-varying field and effective sample size (ESS) or temporal degrees of freedom of a time series, that is, treating time instead of space as the dimension over which correlations occur. Formulas for ESS appropriate for assessing the significance of the correlation coefficient between two time series already exist and are in fact equivalent to our eigenvalue formula applied to the temporal autocorrelation matrix. In fact, in real applications involving fields of data, temporal and spatial autocorrelation both occur simultaneously. As long as the temporal autocorrelation structures of all dominant spatial modes are comparable, we can separate the two issues,

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

computing an ESDOF relevant to spatial correlation and an ESS for temporal correlation.

Acknowledgments. The first author was supported by the National Science Foundation (NSF) under Grant DMS-9524770, the second by a grant from the Swiss National Science Foundation, and the third by the NSF Climate Dynamics Program under Grant ATM 9215512. Participation of the fourth author was sponsored jointly by NOAA (Office of Atmospheric Research and Office of Global Programs) and the University of Washington through the Joint Institute for the Study of the Atmosphere and Ocean. We are indebted to the Swiss Meteorological Institute–MeteoSwiss for providing access to their precipitation data. This work was partly motivated by stimulating discussions with George Gruza, who helped introduce us to the relevant Russian literature, Antonio Navarra, and Doug Martinson. Don Percival provided invaluable references and advice about temporal correlation. Hans von Storch, Ian Jolliffe, and an anonymous reviewer also provided valuable advice.

APPENDIX

Bias and Scatter of ESDOF Estimates Based on Short Records

In this appendix, we derive formulas for the bias and scatter of the estimators $\widehat{N^*_{mm}}$ and $\widehat{N^*_{ef}}$, which are based on a finite number T of independent observation times, in the regime $T \gg N^* \gg 1$.

a. Bias and scatter in $\widehat{N^*_{\mathrm{mm}}}$

We start from the definition (1) of $\widehat{N^*_{mm}}$. In this definition, we estimate the time mean and the variance of E based on T independent observation times:

$$\hat{E} = \frac{1}{T}\sum_{i=1}^{T} E(t_i),$$

$$\widehat{\mathrm{var}(E)} = \frac{1}{T-1}\sum_{i=1}^{T} [E(t_i) - \langle E\rangle]^2. \tag{A1}$$

While $E(t_i)$ will not in general have an exactly normal distribution, the central limit theorem states that the mean of T independent samples of E (and of its squared deviation from the mean) will have a distribution that is much closer to normal, since $T \gg 1$. Thus we regard $\hat{E}$ and $\widehat{\mathrm{var}(E)}$ as being normally distributed and entirely characterized by their means (denoted by angle brackets) and standard deviations (denoted by the prefix δ). These are easily calculated:

$$\langle\hat{E}\rangle = \langle E\rangle, \tag{A2}$$

$$\delta\hat{E} = (\mathrm{var}(E)/T)^{1/2}, \tag{A3}$$

$$\langle\widehat{\mathrm{var}(E)}\rangle = \mathrm{var}(E). \tag{A4}$$

$$\delta[\widehat{\mathrm{var}(E)}] \approx \mathrm{var}(E)(2/T)^{1/2}. \tag{A5}$$

Since T is large, we have ignored the small relative difference between $T - 1$ and T in computing the standard deviation of the variance. The relative standard deviations of $\hat{E}$ and $\widehat{\mathrm{var}(E)}$ can be found from the above:

$$a = \delta\hat{E}/\langle\hat{E}\rangle = \frac{[\mathrm{var}(E)]^{1/2}}{\langle E\rangle T^{1/2}} = (N^*_{mm}T)^{-1/2}, \tag{A6}$$

$$b = \delta[\widehat{\mathrm{var}(E)}]/\langle\widehat{\mathrm{var}(E)}\rangle \tag{A7}$$

$$= (2/T)^{1/2}, \tag{A8}$$

Note that both relative standard deviations are $O(T^{-1/2})$, which is small since T is large.

It is convenient to normalize $\hat{E}$ and $\widehat{\mathrm{var}(E)}$ into the unit normal random variables u and v, in terms of which we can write:

$$\hat{E} = \langle E\rangle(1 + au),$$
$$\widehat{\mathrm{var}(E)} = \mathrm{var}(E)(1 + bv).$$

Substituting these expressions into the definition (1) of $\widehat{N^*_{mm}}$, we obtain

$$\widehat{N^*_{mm}} = \frac{\langle E\rangle^2}{\mathrm{var}(E)}\frac{(1+au)^2}{1+bv} = N^*_{mm}\frac{(1+au)^2}{1+bv}. \tag{A9}$$

To calculate the mean and variance of $\widehat{N^*_{mm}}$ from (A9), one must account for any correlation between the numerator and denominator. We have shown that E has an approximate scaled $\chi^2(N^*_{mm})$ distribution. In our parameter regime, $N^*_{mm} \gg 1$, and this distribution becomes approximately normal. Laha and Rohatgi (1979) showed that the standard estimators of the mean and variance of a finite number of independent samples from a normal distribution are uncorrelated. We conclude that $\langle E\rangle$ and $\mathrm{var}(E)$, and thus u and v, are approximately uncorrelated.

With this result, we calculate the mean and variance of $\widehat{N^*_{mm}}$, including all terms at least as large as $O(T^{-1})$ (i.e., second order in a, b). Substituting (A6) and (A7) into (A9), and noting that $\langle u^2\rangle = \langle v^2\rangle = 1$ and $\langle u\rangle = \langle v\rangle = \langle uv\rangle = 0$, we obtain (12):

$$\langle\widehat{N^*_{mm}}\rangle/N^*_{mm} = \left\langle\frac{(1+au)^2}{1+bv}\right\rangle \approx 1 + 2a\langle u\rangle + a^2\langle u^2\rangle - b\langle v\rangle - 2ab\langle uv\rangle + b^2\langle v^2\rangle = 1 + a^2 + b^2 \approx 1 + \frac{2}{T}. \tag{A10}$$

In deriving the final line above, we note that $a^2/b^2 = (2N^*_{mm})^{-1} \ll 1$.

Similarly,

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

$$\begin{aligned}\mathrm{var}[\widehat{N^*_{mm}}] &= N^{*2}_{mm}\left\{\left\langle\left[\frac{(1+au)^2}{1+bv}\right]^2\right\rangle - \left\langle\left[\frac{(1+au)^2}{1+bv}\right]\right\rangle^2\right\}\\ &\approx N^{*2}_{mm}\{(1+4a\langle u\rangle + 6a^2\langle u^2\rangle - 2b\langle v\rangle\\ &\qquad - 8ab\langle uv\rangle + 3b^2\langle v^2\rangle)\\ &\qquad - (1+a^2+b^2)^2\}\\ &\approx \{(1+6a^2+3b^2) - (1+2a^2+2b^2)\}\\ &= 4a^2 + b^2. \end{aligned} \tag{A11}$$

This expression could have been more simply derived through a standard error propagation analysis that assumes that errors in u and v are uncorrelated; we have chosen an approach that can be generalized to the case of correlated u and v, as this will be needed to make a similar analysis for N^*_{ef}. Substituting for a and b in (A11) using (A6) and (A7), and converting from variance to standard deviations, we obtain (13):

$$\delta\widehat{N^*_{mm}}/N^*_{mm} = \left(\frac{4}{N^*_{mm}T} + \frac{2}{T}\right)^{1/2} \approx (2/T)^{1/2}.$$

b. Bias and scatter in $\widehat{N^*_{ef}}$

Let $\mathbf{C}$ be the true covariance matrix. From (3) applied to the eigenvalues of $\mathbf{C}$, we could calculate the true $N^*_{ef} = (\Sigma_i \lambda_i)^2/\Sigma_i\lambda_i^2$. Let $\hat{\mathbf{C}}$ be an estimate of $\mathbf{C}$ based on T independent sampling times. From (3) applied to the eigenvalues of $\hat{\mathbf{C}}$, we obtain an estimate $\widehat{N^*_{ef}}$ of the true N^*_{ef}. We regard $\widehat{N^*_{ef}}$ as a random variable and estimate its mean and standard deviation. Our estimates assume that the field values observed at each grid point (and hence the true principal components) are normally distributed.

As with $\widehat{N^*_{mm}}$, $\widehat{N^*_{ef}}$ is the quotient of two functions (in this case, functions of the eigenvalues of the covariance matrix), each having a mean and standard deviation that can be computed. As for $\widehat{N^*_{mm}}$, the relative standard deviation of each of these quantities is $O(T^{-1/2})$, which is small in our parameter regime. Unlike for $\widehat{N^*_{mm}}$, these two functions are strongly correlated, and this must be accounted for in our analysis.

We must first compute the mean, standard deviation, and correlation coefficient of these two functions. To simplify our analysis, we work in the basis of the eigenvectors of the true covariance matrix. Let $\mathbf{E}$ be the orthogonal matrix whose columns are these normalized eigenvectors, and let $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, where the λ_i are the eigenvalues of $\mathbf{C}$. Then

$$\mathbf{\Lambda} = \mathbf{E}^T\mathbf{C}\mathbf{E}.$$

We can apply the same transformation to $\hat{\mathbf{C}}$ to obtain a matrix,

$$\hat{\mathbf{L}} = \mathbf{E}^T\hat{\mathbf{C}}\mathbf{E}.$$

The matrix $\hat{\mathbf{L}}$ is no longer exactly diagonal, but it has the same eigenvalues as $\hat{\mathbf{C}}$, and its elements can be computed in terms of the true normalized principal components $z_i(t)$:

$$\hat{L}_{ij} = \frac{\lambda_i^{1/2}\lambda_j^{1/2}}{T}\sum_{n=1}^{T} z_i(t_n)z_j(t_n). \tag{A12}$$

To compute $\widehat{N^*_{ef}}$, we do not have to find the eigenvalues of $\hat{\mathbf{C}}$ or $\hat{\mathbf{L}}$. Instead, we can apply (5) to $\hat{\mathbf{L}}$:

$$\widehat{N^*_{ef}} = \frac{(\mathrm{tr}\hat{\mathbf{L}})^2}{\mathrm{tr}(\hat{\mathbf{L}}^2)} = \frac{\left(\sum_{i=1}^{N}\hat{L}_{ii}\right)^2}{\sum_{i,j=1}^{N}\hat{L}_{ij}^2}. \tag{A13}$$

The numerator $[\mathrm{tr}(\hat{\mathbf{L}})]^2 = (\Sigma_i\,\hat{\lambda}_i)^2$, and the denominator $\mathrm{tr}(\hat{\mathbf{L}}^2) = \Sigma_i\,\hat{\lambda}_i^2$.

It is straightforward to show that

$$\langle\hat{L}_{ij}\rangle = \begin{cases}\lambda_i & j = i\\ 0 & j \neq i\end{cases}$$

and

$$\delta\hat{L}_{ij} = \lambda_i^{1/2}\lambda_j^{1/2}\begin{cases}(2/T)^{1/2} & j = i\\ (1/T)^{1/2} & j \neq i.\end{cases}$$

Furthermore, each element of $\hat{\mathbf{L}}$ is uncorrelated with all other elements of $\hat{\mathbf{L}}$. In the rest of this analysis, we will restrict ourself to the case in which T is large, so each element $\hat{L}_{ij}$, which is a sum of T independent and identically distributed random variables, must have an approximately normal distribution by the central limit theorem. Thus the statistics of $\hat{\mathbf{L}}$ can be determined by noting that

$$\hat{L}_{ij} \sim \lambda_i^{1/2}\lambda_j^{1/2}\begin{cases}1 + w_{ij}(2/T)^{1/2} & j = i\\ w_{ij}/T^{1/2} & j \neq i,\end{cases}$$

where the w_{ij} are uncorrelated unit normal random variables and $\sim$ indicates that the two sides have the same PDF.

We are now ready to compute the mean and variance of $\mathrm{tr}(\hat{\mathbf{L}})^2$ and $\mathrm{tr}(\hat{\mathbf{L}}^2)$. To make our results more compact, we define

$$S_m = \sum_{i=1}^{N}\lambda_i^m.$$

We will also make use of some moments of the unit normal distribution: $\langle w_{ij}^2\rangle = 1$, $\langle w_{ij}^4\rangle = 3$, and $\mathrm{var}(w_{ij}^2) = 2$. The statistics of $\mathrm{tr}(\hat{\mathbf{L}})$ are as follows:

$$\langle\mathrm{tr}(\hat{\mathbf{L}})\rangle = \sum_{i=1}^{N}\langle\hat{L}_{ii}\rangle = \sum_{i=1}^{N}\lambda_i = S_1, \tag{A14}$$

$$\begin{aligned}\mathrm{var}[\mathrm{tr}(\hat{\mathbf{L}})] &= \sum_{i=1}^{N}\mathrm{var}[\hat{L}_{ii}] = \sum_{i=1}^{N}\lambda_i^2\,\mathrm{var}[1 + w_{ii}(2/T)^{1/2}]\\ &= 2S_2/T,\\ \delta[\mathrm{tr}(\hat{\mathbf{L}})] &= (2S_2/T)^{1/2}. \end{aligned} \tag{A15}$$

Note that since each of the $\hat{L}_{ii}$ has a normal distribution, so will $\mathrm{tr}(\hat{\mathbf{L}})$.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

The statistics of $\mathrm{tr}(\hat{\mathbf{L}}^2)$ are computed as follows:

$$\begin{aligned}\langle \mathrm{tr}(\hat{\mathbf{L}}^2)\rangle &= \sum_{i,j=1}^{N} \langle \hat{L}_{ij}^2\rangle \\ &= \sum_{i\neq j} \lambda_i\lambda_j\langle w_{ij}^2\rangle/T + \sum_{i=1}^{N} \lambda_i^2\langle(1 + w_{ij}(2/T)^{1/2})^2\rangle \\ &= \sum_{i\neq j} \lambda_i\lambda_j/T + \sum_{i=1}^{N} \lambda_i^2(1 + 2/T) \\ &= \sum_{i,j=1}^{N} \lambda_i\lambda_j/T + \sum_{i=1}^{N} \lambda_i^2(1 + 1/T) \\ &= S_2 + (S_1^2 + S_2)/T, \end{aligned} \tag{A16}$$

$$\begin{aligned}\mathrm{var}[\mathrm{tr}(\hat{\mathbf{L}}^2)] &= \sum_{i,j=1}^{N} \mathrm{var}[\hat{L}_{ij}^2] \\ &= \sum_{i\neq j} \lambda_i^2\lambda_j^2\ \mathrm{var}(w_{ij}^2)/T^2 \\ &\quad + \sum_{i=1}^{N} \lambda_i^4\ \mathrm{var}[(1 + w_{ij}(2/T)^{1/2})^2].\end{aligned}$$

To compute $\mathrm{var}[(1 + w_{ij}(2/T)^{1/2})^2]$, we proceed as follows:

$$\begin{aligned}&\mathrm{var}[(1 + w_{ij}(2/T)^{1/2})^2] \\ &\quad = \langle(1 + w_{ij}(2/T)^{1/2})^4\rangle - \langle(1 + w_{ij}(2/T)^{1/2})^2\rangle^2 \\ &\quad = 1 + 12\langle w_{ij}^2\rangle/T + 4\langle w_{ij}^4\rangle/T^2 - (1 + 2\langle w_{ij}^2\rangle/T)^2 \\ &\quad = 8/T + 8/T^2.\end{aligned}$$

We deduce that

$$\begin{aligned}\mathrm{var}[\mathrm{tr}(\hat{\mathbf{L}}^2)] &= \sum_{i\neq j} \lambda_i^2\lambda_j^2/T^2 + \sum_{i=1}^{N} \lambda_i^4(8/T + 8/T^2) \\ &= 8S_4/T + (S_2^2 + 7S_4)/T^2. \end{aligned} \tag{A17}$$

$$\delta[\mathrm{tr}(\hat{\mathbf{L}}^2)] \approx (8S_4/T)^{1/2}, \qquad T \gg N_4 = S_2^2/S_4. \tag{A18}$$

Note that N_4 is analogous to N_{ef}^*, except with λ_i^2 used in place of λ_i. Since this accentuates the relative importance of the highest-variance PCs, $1 < N_4 < N_{ef}^*$. Thus it is safe to replace the condition $T \gg N_4$ by the more conservative condition $T \gg N_{ef}^*$, which is the parameter regime assumed throughout this analysis.

The above analysis shows that the dominant contribution to the variability of $\mathrm{tr}(\hat{\mathbf{L}}^2)$ comes from the sum of the diagonal terms $\hat{L}_{ii}^2$, and in particular from the sum of the subterms $w_{ii}(8/T)^{1/2}$ of those diagonal terms. Since w_{ii} is normally distributed, this sum, and hence $\mathrm{tr}(\hat{\mathbf{L}}^2)$, should also be normally distributed.

Last, the covariance of $\mathrm{tr}(\hat{\mathbf{L}})$ and $\mathrm{tr}(\hat{\mathbf{L}}^2)$ must be computed. This is greatly simplified by noting that the elements of $\hat{\mathbf{L}}$ are uncorrelated with each other:

$$\begin{aligned}\mathrm{cov}[\mathrm{tr}(\hat{\mathbf{L}}), \mathrm{tr}(\hat{\mathbf{L}}^2)] &= \langle[\mathrm{tr}(\hat{\mathbf{L}}) - \langle\mathrm{tr}(\hat{\mathbf{L}})\rangle][\mathrm{tr}(\hat{\mathbf{L}}^2) - \langle\mathrm{tr}(\hat{\mathbf{L}}^2)\rangle]\rangle \\ &= \left\langle\left[\sum_{i=1}^{N} \lambda_i w_{ii}(2/T)^{1/2}\right]\left[\sum_{j,k=1}^{N} (\hat{L}_{jk}^2 - \langle\hat{L}_{jk}^2\rangle)\right]\right\rangle \\ &= \sum_{i=1}^{N} \lambda_i^3\langle w_{ii}(2/T)^{1/2} \\ &\qquad \times [2w_{ii}(2/T)^{1/2} + 2(w_{ii}^2 - 1)/T]\rangle \\ &= \frac{4}{T}\sum_{i=1}^{N} \lambda_i^3\langle w_{ii}^2\rangle = 4S_3/T. \end{aligned} \tag{A19}$$

The relative standard deviation of $\mathrm{tr}(\hat{\mathbf{L}})$ is found from (A14) and (A15):

$$A = \frac{\delta[\mathrm{tr}(\hat{\mathbf{L}})]}{\langle\mathrm{tr}(\hat{\mathbf{L}})\rangle} = \frac{(2S_2/T)^{1/2}}{S_1} = \left(\frac{2}{N_{ef}^*T}\right)^{1/2}. \tag{A20}$$

The relative standard deviation of $\mathrm{tr}(\hat{\mathbf{L}}^2)$ is found from (A16) and (A18), neglecting higher-order terms in T:

$$B = \frac{\delta[\mathrm{tr}(\hat{\mathbf{L}}^2)]}{\langle\mathrm{tr}(\hat{\mathbf{L}}^2)\rangle} \approx \frac{(8S_4/T)^{1/2}}{S_2} = \left(\frac{8}{N_4T}\right)^{1/2}. \tag{A21}$$

We have argued that $\mathrm{tr}(\hat{\mathbf{L}})$ and $\mathrm{tr}(\hat{\mathbf{L}}^2)$ are approximately normally distributed. We can normalize them to unit normal random variables U and V:

$$\begin{aligned}\mathrm{tr}(\hat{\mathbf{L}}) &= \langle\mathrm{tr}(\hat{\mathbf{L}})\rangle(1 + AU), \\ \mathrm{tr}(\hat{\mathbf{L}}^2) &= \langle\mathrm{tr}(\hat{\mathbf{L}}^2)\rangle(1 + BV).\end{aligned}$$

Substituting these expressions into the definition (A13) of $\widehat{N_{ef}^*}$,

$$\begin{aligned}\widehat{N_{ef}^*} &\approx \frac{\langle\mathrm{tr}(L)\rangle^2}{\langle\mathrm{tr}(L^2)\rangle}\frac{(1 + AU)^2}{1 + BV} \\ &= \frac{N_{ef}^*}{1 + N_{ef}^*/T}\frac{(1 + AU)^2}{1 + BV}. \end{aligned} \tag{A22}$$

Following the analogous derivation of the mean and variance of $\widehat{N_{ef}^*}$, we note that A and B are both small $[O(T^{-1/2})]$, and we Taylor-expand the second fraction of (A22) up through quadratic terms in A and B. The new twist is that U and V are now strongly correlated. Their correlation coefficient can be computed from (A19), (A15), and (A18):

$$\begin{aligned}\langle UV\rangle &= \frac{\mathrm{cov}[\mathrm{tr}(\hat{\mathbf{L}}), \mathrm{tr}(\hat{\mathbf{L}}^2)]}{\delta[\mathrm{tr}(\hat{\mathbf{L}})]\delta[\mathrm{tr}(\hat{\mathbf{L}}^2)]} = \frac{4S_3/T}{(2S_2/T)^{1/2}(8S_4/T)^{1/2}} \\ &= R = \frac{S_3}{(S_2S_4)^{1/2}}. \end{aligned} \tag{A23}$$

In the limit of either one dominant eigenvalue, or all eigenvalues equal, this correlation R is unity. For eigenvalues slowly decreasing geometrically like powers of some constant $\exp(-\mu)$, $\mu \ll 1$, R can be shown to be $(8/9)^{1/2} \approx 0.94$. For most realistic eigenvalue spectra, R exceeds 0.9, so the numerator and denominator of $\widehat{N_{ef}^*}$ are highly correlated.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

We now mimic the analysis that led up to Eqs. (A10) and (A11) for the mean and variance of $\widehat{N^*_{mm}}$, but retaining terms proportional to $\langle UV\rangle$. A computation of the mean of $\widehat{N^*_{ef}}$ from (A22) leads to (14):

$$\begin{aligned}\langle\widehat{N^*_{ef}}\rangle/N^*_{ef} &= \frac{1}{1 + N^*_{ef}/T}\left\langle\frac{(1 + AU)^2}{1 + BV}\right\rangle \\ &\approx \frac{1}{1 + N^*_{ef}/T}(1 + 2A\langle U\rangle + A^2\langle U^2\rangle - B\langle V\rangle \\ &\qquad - 2AB\langle UV\rangle + B^2\langle V^2\rangle) \\ &= \frac{1}{1 + N^*_{ef}/T}(1 + A^2 - 2ABR + B^2) \\ &\approx \frac{1}{1 + N^*_{ef}/T}. \qquad \text{(A24)}\end{aligned}$$

In the last step, we have noted that A^2, AB, and B^2 are all proportional to $(N^*_{ef}T)^{-1}$, which is much smaller than N^*_{ef}/T since $N^*_{ef} \gg 1$.

A similar computation of the variance of $\widehat{N^*_{ef}}$ shows that

$$\begin{aligned}\text{var}[\widehat{N^*_{ef}}] &= \left\{\frac{N^*_{ef}}{1 + N^*_{ef}/T}\right\}^2 \\ &\quad\times\left\{\left\langle\left[\frac{(1 + AU)^2}{1 + BV}\right]^2\right\rangle - \left\langle\left[\frac{(1 + AU)^2}{1 + BV}\right]\right\rangle^2\right\} \\ &\approx \left\{\frac{N^*_{ef}}{1 + N^*_{ef}/T}\right\}^2 \\ &\quad\times\{(1 + 4A\langle U\rangle + 6A^2\langle U^2\rangle - 2B\langle V\rangle \\ &\qquad - 8AB\langle UV\rangle + 3B^2\langle V^2\rangle) \\ &\qquad - (1 + A^2 - 2AB\langle UV\rangle + B^2)^2\} \\ &\approx \left\{\frac{N^*_{ef}}{1 + N^*_{ef}/T}\right\}^2\{(1 + 6A^2 - 8ABR + 3B^2) \\ &\qquad - (1 + 2A^2 - 4ABR + 2B^2)\} \\ &\approx N^{*2}_{ef}(4A^2 - 4ABR + B^2). \qquad \text{(A25)}\end{aligned}$$

In the last line, we have neglected the term N^*_{ef}/T as being small compared to one, as it is not a leading-order contributor to the variance.

Substituting for A, B, and R in (A25) using (A20), (A21), and (A23), and converting from variance to standard deviation, we obtain (15):

$$\delta\widehat{N^*_{ef}}/\widehat{N^*_{ef}} = cT^{-1/2},$$

where

$$c^2 = 8\left\{\frac{1}{N^*_{ef}} + \frac{1}{N_4} - 2R\frac{1}{(N^*_{ef}N_4)^{1/2}}\right\}. \qquad \text{(A26)}$$

Because R is close to 1, and N_4 can be nearly as large as N^*_{ef}, the last term in c^2 partially cancels the other two, and the relative standard deviation of $\widehat{N^*_{ef}}$ is quite small. In fact, if all eigenvalues are equal, or if one mode explains all the variance, c is exactly zero, the scatter to $O(T^{-1/2})$ is negligible, and the relative standard deviation must be of higher order in T.

REFERENCES

Bagrov, N. A, 1969: On the equivalent number of independent data (in Russian). *Tr. Gidrometeor. Cent.,* **44,** 3–11.

Bartlett, M. S., 1935: Some aspects of the time-correlation problem in regard to tests of significance. *J. Roy. Stat. Soc.,* **98,** 536–543.

Bayley, G. V., and J. M. Hammersley, 1946: The ''effective'' number of independent observations in autocorrelated time series. *J. Roy. Stat. Soc. Suppl.,* **8,** 184–197.

Blackman, R. B., and J. W. Tukey, 1959: *The Measurement of Power Spectra From the Point of View of Communications Engineering.* Dover, 190 pp.

Da Costa, E., and R. Vautard, 1997: A qualitatively realistic low-order model of the extratropical low-frequency variability built from long records of potential vorticity. *J. Atmos. Sci.*, **54,** 1064–1084.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.,* **6,** 249–266.

Dymnikov, V. P., and A. S. Gritsun, 1996: Lyapunov exponents and the attractor dimension in a two-layer baroclinic model of atmospheric circulation (in Russian). *Dokl. Ross. Akad. Nauk.,* **347** (4), 535–538.

——, and ——, 1997: On the structure of the attractors of the finite-dimensional approximations of the barotropic vorticity equation on a rotating sphere. *Russ. J. Numer. Anal. Math. Model.,* **12,** 13–32.

Fraedrich, K., C. Ziehmann, and F. Sielmann, 1995: Estimates of spatial degrees of freedom. *J. Climate,* **8,** 361–369.

Hogg, R. V., and A. T. Craig, 1978: *Introduction to Mathematical Statistics.* 4th ed. MacMillan Press, 438 pp.

Huang, J., H. M. van den Dool, and K. G. Georgakakos, 1996: Analysis of model-calculated soil moisture over the U.S. (1931–1993) and applications to long range temperature forecasts. *J. Climate,* **9,** 1350–1362.

Jones, R. H., 1975: Estimating the variance of time averages. *J. Appl. Meteor.,* **14,** 159–163.

Keller, L. B., 1935: Expanding of limit theorems of probability theory on functions with continuous arguments (in Russian). *Works Main Geophys. Observ.,* **4,** 5–19.

Kikkawa, S., and M. Ishida, 1988: Number of degrees of freedom, correlation times, and equivalent bandwidths of a random process. *IEEE Trans. Inf. Theory,* **34,** 151–155.

Laha, R. G., and V. K. Rohatgi, 1979: *Probability Theory.* Wiley, 557 pp.

Leith, C. E., 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.,* **12,** 1066–1069.

Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo testing. *Mon. Wea. Rev.,* **111,** 46–59.

Lorenz, E. N., 1969: Atmospheric predictability by naturally occurring analogs. *J. Atmos. Sci.,* **26,** 636–646.

——, 1991: Dimension of weather and climate attractors. *Nature,* **353,** 241–244.

North, G., T. R. Bell, R. Cahalan, and F. Moeng, 1982: Sampling errors in estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Orcutt, G. H., and S. F. James, 1948: Testing the significance of correlation between time series. *Biometrika,* **35,** 397–413.

Percival, D. B., and A. T. Walden, 1993: *Spectral Analysis for Physical Applications.* Cambridge University Press, 583 pp.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020

Strang, G., 1988: *Linear Algebra and Its Applications*. 3d ed. Harcourt Brace Jovanovich, 505 pp.

Taylor, G. I., 1921: Diffusion by continuous movement. *Proc. London Math. Soc.,* **20** (2), 196–212.

TerMegreditchian, M. G., 1969: On the determination of the number of independent stations which are equivalent to prescribed systems of correlated stations (in Russian). *Meteor. Hydrol.,* **2,** 24–36.

——, 1990: Meteorological networks optimization from a statistical point of view. *Comput. Stat. Data Anal.,* **9,** 57–75.

Thiébaux, H. J., and F. W. Zwiers, 1984: The interpretation and estimation of effective sample size. *J. Climate Appl. Meteor.,* **23,** 800–811.

Toth, Z., 1995: Degrees of freedom in Northern Hemisphere circulation data. *Tellus,* **47A,** 457–472.

van den Dool, H. M., 1987: An empirical study on the parameterization of precipitation in a model of the time mean atmosphere. *J. Atmos. Sci.,* **44,** 224–235.

——, and R. W. Chervin, 1986: A comparison of month-to-month persistence of anomalies in a general circulation model and in the Earth's atmosphere. *J. Atmos. Sci.,* **43,** 1454–1466.

Vautard, R., and M. Ghil, 1989: Singular spectum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D,* **35,** 395–424.

Wallace, J. M., X. Cheng, and D. Sun, 1991: Does low-frequency atmospheric variability exhibit regime-like behaviour? *Tellus,* **43AB,** 16–26.

Widmann, M., and C. Schär, 1997: A principal component and long-term trend analysis of daily precipitation in Switzerland. *Int. J. Climatol.,* **17,** 1333–1356.

Yulaeva, E., and J. M. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate,* **7,** 1719–1736.

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2 by guest on 26 September 2020