Modified "Rule N" Procedure for Principal Component (EOF) Truncation

DANIEL S. WILKS

Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York

(Manuscript received 13 November 2015, in final form 3 February 2016)

ABSTRACT

Principal component analysis (PCA), also known as empirical orthogonal function (EOF) analysis, is widely used for compression of high-dimensional datasets in such applications as climate diagnostics and seasonal forecasting. A critical question when using this method is the number of modes, representing meaningful signal, to retain. The resampling-based "Rule N" method attempts to address the question of PCA truncation in a statistically principled manner. However, it is only valid for the leading (largest) eigenvalue, because it fails to condition the hypothesis tests for subsequent (smaller) eigenvalues on the results of previous tests. This paper draws on several relatively recent statistical results to construct a hypothesis-test-based truncation rule that accounts at each stage for the magnitudes of the larger eigenvalues. The performance of the method is demonstrated in an artificial data setting and illustrated with a real-data example.

1. Introduction

Principal component analysis (PCA), also known as empirical orthogonal function (EOF) analysis, has a long history of use in meteorology, climatology, and oceanography since its introduction into this literature by Obukhov (1947), Lorenz (1956), and Davis (1976). The method often is very effective at compressing high-dimensional datasets by producing a relatively small number of uncorrelated derived variables, while preserving much of the original variance. The computations involve extracting eigenvalue–eigenvector pairs from the covariance matrix of the data under analysis, and projecting those data onto the eigenvectors corresponding to the largest eigenvalues.

A key issue in the use of PCA is choice of the number of these derived variables to retain in the compressed dataset, with this truncation point generally chosen on the basis of the magnitudes of the sample eigenvalues. A wide variety of methods have been proposed to guide this process (Preisendorfer and Mobley 1988; Jolliffe 2002; Wilks 2011), although these methods often disagree and no consensus has developed regarding the most appropriate approach. Some authors recommend against use of these selection rules at all (von Storch and Zwiers 1999).

Typically the choice of the truncation point is framed in terms of separating hypothesized "signal" and "noise" subspaces of the data space. It is usually assumed that the noise is relatively weak and uncorrelated across the original variables, and so will be represented by a sequence of higher-order generating-process eigenvalues of equal magnitude. Although this is a reasonable viewpoint for many applications, and will be assumed in this paper, the definition of what is considered as signal may depend on the scientific context. For example, if the scientific interest focuses on one or a few large-scale processes, some portion of the noise may be spatially correlated, and represented by intermediate eigenvalues. Alternatively, real physical processes occurring at unresolved scales may be aliased as noise onto the resolved scales, and make contributions across the low-variance eigenvalues.

Rule N (Preisendorfer et al. 1981; Overland and Preisendorfer 1982) is a popular method for PCA truncation that assumes the noise subspace is represented by a sequence of trailing eigenvalues of equal magnitude. It is based on statistical hypothesis testing ideas, and it continues to be used widely in climatology and oceanography (e.g., Pérez-Hernández and Joyce 2014; Ortega et al. 2015; Wang et al. 2015; Feng et al. 2016). However, Rule N is only strictly valid for the hypothesis test involving the leading eigenvalue, which amounts to

Corresponding author address: Daniel S. Wilks, Department of Earth and Atmospheric Sciences, Cornell University, 1104 Bradfield Hall, Ithaca, NY 14853.
E-mail: dsw5@cornell.edu

DOI: 10.1175/JCLI-D-15-0812.1

© 2016 American Meteorological Society

testing the null hypothesis that the signal subspace is null. The result is that Rule N is usually conservative, meaning that it retains too few components.

This paper proposes a modification of Rule N for PCA truncation that is based on several relatively recent results from the statistics literature. Section 2 reviews Rule N and describes the proposed modification, section 3 compares the performance of the two methods in synthetic-data settings where the correct truncation points are known, section 4 illustrates the methods in a small real-data setting, and section 5 concludes the paper.

2. Testing a sequence of sample eigenvalues

The underlying assumption in the following is that the multivariate data space can be partitioned into a κ-dimensional signal subspace and a remaining noise subspace. The eigenvalues corresponding to eigenvectors spanning these subspaces are

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\kappa > \lambda_{\kappa+1} = \lambda_{\kappa+2} = \cdots = \lambda_M = \sigma^2_{\text{noise}}, \quad (1)$$

where M is the dimension of the data space. These data-generating-process ("population") eigenvalues are estimated in PCA by extracting the sample eigenvalues of a covariance matrix, which has been calculated from a sample of n data vectors of dimension M. In a typical climatological application, M is the number of spatial locations (often grid points), and n is the number of temporal observations.

The goal of Rule N (Preisendorfer et al. 1981, Overland and Preisendorfer 1982) is to estimate the dimension of the signal subspace, κ, by comparing the members of an observed sample eigenvalue sequence $\hat{\lambda}_k$, $k = 1, 2, 3, \ldots$, to a high quantile (e.g., the 95th percentile) of their respective sampling distributions. These distributions are estimated from a large number of PCAs for covariance matrices that have been computed from uncorrelated (usually Gaussian) random data vectors having the same sample size and dimension. In order that the test and synthetic sample eigenvalues are scaled comparably, both are normalized by the average of the nonzero sample eigenvalues:

$$\tilde{\ell}_k = \frac{\hat{\lambda}_k}{\frac{1}{N_{\text{rank}}} \sum_{i=1}^{N_{\text{rank}}} \hat{\lambda}_i}, \quad (2)$$

where $N_{\text{rank}} = \min(n - 1, M)$ is the number of nonzero sample eigenvalues. The name "Rule N" refers to the normalization of the sample eigenvalue estimates by the average eigenvalue in the denominator of Eq. (2) (Preisendorfer et al. 1981).

Using Rule N involves examining a sequence of hypothesis tests with null hypotheses H_k, specifying the propositions

$$H_k\colon \{\lambda_k = \lambda_{k+1} = \cdots = \lambda_{N_{\text{rank}}}\}, \quad k = 1, \ldots, N_{\text{rank}} - 1, \quad (3)$$

that is, that following the assumption specified by Eq. (1) the trailing $N_{\text{rank}} - k + 1$ eigenvalues represent noise. The Rule N test statistic for each H_k is $\tilde{\ell}_k$, which is compared to the collection of Monte Carlo realizations for the corresponding normalized eigenvalue. The null hypothesis H_k represents the proposition that $\kappa < k$, so that rejecting H_k indicates $\kappa \geq k$. The estimated truncation point is then taken to be the eigenvalue index k that is one less than that of the first sample eigenvalue not regarded as statistically significant:

$$\hat{\kappa} = \min(k \in \{1, 2, \ldots, N_{\text{rank}} - 1\}\colon p_k > \alpha_{\text{crit}}) - 1, \quad (4)$$

where α_{crit} is the chosen test level (often 0.05). For Rule N, p_k is estimated as the fraction of the scaled [according to Eq. (2)] synthetic noise eigenvalues with index k that are larger than the current test statistic $\tilde{\ell}_k$. Thus, for example, if the first null hypothesis H_1 is not rejected because $p_1 > \alpha_{\text{crit}}$, Eq. (4) specifies that $\hat{\kappa} = 0$, with the corresponding inference being that the underlying dataset is comprised entirely of noise.

Although it is conceptually attractive, the primary problem with Rule N is that, having rejected the first null hypothesis H_1 that λ_1 corresponds to noise (because $\tilde{\ell}_1$ is unusually large compared to its Monte Carlo critical value), the remaining tests (for H_2, H_3 . . .) are incorrect. The reason is that the total remaining variance in the observed sample eigenvalues $\hat{\lambda}_k$ for $k > 1$ will be too small relative to the corresponding Monte Carlo distributions to which they are being compared. That is, the Rule N hypotheses tests beyond the first null hypothesis H_1 do not account for previous significant results, because those results will imply a greater variance fraction represented by the leading eigenvalues than is assumed by the subsequent Monte Carlo distributions. The consequence is that the Rule N test is overly conservative, tending to conclude that $\hat{\kappa} < \kappa$ (Preisendorfer and Mobley 1988).

An improved hypothesis-testing-based approach to estimating the dimension of the signal subspace in PCA can be constructed by combining several relatively recent statistical results. Johnstone (2001) has shown that the sampling distribution of the leading eigenvalue of a sample covariance matrix derived from uncorrelated random variates, when appropriately scaled, follows a distribution proposed by Tracy and Widom (1996), which will be defined later. For example, over many Monte Carlo simulations, a histogram of appropriately

scaled multiple random realizations for $\tilde{\ell}_1$ [Eq. (2)] from uncorrelated noise data will estimate the Tracy–Widom distribution.

Kritchman and Nadler (2008) show that if $\kappa = 1$, the (appropriately scaled) sampling distribution of $\hat{\lambda}_2$ also follows the Tracy–Widom distribution, if both n and M are large. Following Faber et al. (1994), this sampling distribution for $\hat{\lambda}_2$ will be the same as the sampling distribution for $\hat{\lambda}_1$ calculated from covariance matrices formed from vectors of uncorrelated noise data having dimension $M_2 = M - 1$, with sample size $n_2 = n - 1$. Kritchman and Nadler (2008) conjecture that their result generalizes for the sampling distributions of $\hat{\lambda}_{k+1}$, $k = 2, 3, \ldots$, when $\kappa = k$, in which case the relevant Tracy–Widom data dimension and sample size would be $M_k = M - k + 1$ and $n_k = n - k + 1$.

Because a sequence of tests based on these results will relate only to those eigenvalues regarded as reflecting noise according to the current null hypothesis [Eq. (3)], the appropriate normalization of a raw sample eigenvalue is with respect to the average over these trailing eigenvalues:

$$\ell_k^* = \frac{\hat{\lambda}_k}{\dfrac{1}{N_{\text{rank}} - k + 1} \displaystyle\sum_{i=k}^{N_{\text{rank}}} \hat{\lambda}_k}. \tag{5}$$

Thus $\tilde{\ell}_1 = \ell_1^*$, but for $k > 1$ the normalization in Eq. (5) is unaffected by the magnitudes of the leading sample eigenvalues $\hat{\lambda}_m$, $m < k$, so that $\tilde{\ell}_k < \ell_k^*$ for $k > 1$.

The Tracy–Widom distribution does not have a closed-form analytic representation, and the numerics required for working with it are elaborate. However, Chiani (2014) shows that the Tracy–Widom distribution can be well approximated by a Pearson-III (i.e., shifted gamma) distribution, the probability density function for which is

$$f(\ell_k^*) = \frac{\left(\dfrac{\ell_k^* - \zeta_k}{\beta_k}\right)^{\alpha-1} \exp\left(-\dfrac{\ell_k^* - \zeta_k}{\beta_k}\right)}{\beta_k \Gamma(\alpha)}, \quad \ell_k^* > \zeta_k, \tag{6}$$

where $\Gamma(\bullet)$ denotes the gamma function. Correspondence with the Tracy–Widom distribution is achieved when the shape, scale, and shift parameters of the Pearson-III distribution are, respectively,

$$\alpha = 46.4, \tag{7a}$$

$$\beta_k = \frac{0.186\sigma_k}{\max(n_k, M_k)}, \tag{7b}$$

and

TABLE 1. Percent error of 95th percentiles of Pearson-III approximations to Tracy–Widom distributions, relative to 10 000-member Monte Carlo counterparts.

	M_k					
n_k	50	100	200	500	1000	2000
10	7.42	4.95	3.70	2.31	1.74	1.22
20	4.19	2.89	2.18	1.38	1.07	0.74
50	2.30	1.70	1.19	0.80	0.53	0.37
100	0.23	1.09	0.64	0.43	0.29	0.33
200	0.39	0.03	0.37	0.33	0.19	0.17
500	0.23	0.14	0.07	0.15	0.14	0.09
1000	0.20	0.11	0.00	0.07	0.05	0.07

$$\zeta_k = \frac{\mu_k - 9.85\sigma_k}{\max(n_k, M_k)}. \tag{7c}$$

Here the Tracy–Widom location and scale parameters are

$$\mu_k = \left(\sqrt{n_k - 1/2} + \sqrt{M_k - 1/2}\right)^2 \tag{8a}$$

and

$$\sigma_k = \left(\sqrt{n_k - 1/2} + \sqrt{M_k - 1/2}\right)\left(\frac{1}{\sqrt{n_k - 1/2}} + \frac{1}{\sqrt{M_k - 1/2}}\right)^{1/3}. \tag{8b}$$

As before, $M_k = M - k + 1$ and $n_k = n - k + 1$ for the null hypothesis H_k.

Under a null hypothesis H_k the quantity $\ell_k^* - \zeta_k$ is thus a random draw from an ordinary (two parameter) gamma distribution with shape parameter α [Eq. (7a)] and scale parameter β_k [Eq. (7b)]. The estimated dimension of the signal subspace $\hat{\kappa}$ is then specified by Eq. (4), where the quantity p_k corresponds to probability in this distribution (i.e., the right-tail area) above $\ell_k^* - \zeta_k$.

The Pearson-III approximation to the Tracy–Widom distribution is quite good, especially for the right-tail quantiles that are of primary interest in the present setting (Chiani 2014). The correspondence in the left tail is less good, due in part to the fact that a Tracy–Widom variable can in principle be any real number, whereas Eq. (6) requires $\ell_k^* > \zeta_k$. As a consequence, occasionally a sample value of $\ell_k^* - \zeta_k$ may be negative, in which case the interpretation should be that the corresponding $p_k \approx 1$.

Since the applicability of the Tracy–Widom distribution assumes that both n and M are large, it is of interest to examine its accuracy for the small to moderate sample size and/or data dimensions that are typically encountered. Table 1 shows relative specification error (%) for 95th percentiles computed using the Pearson-III

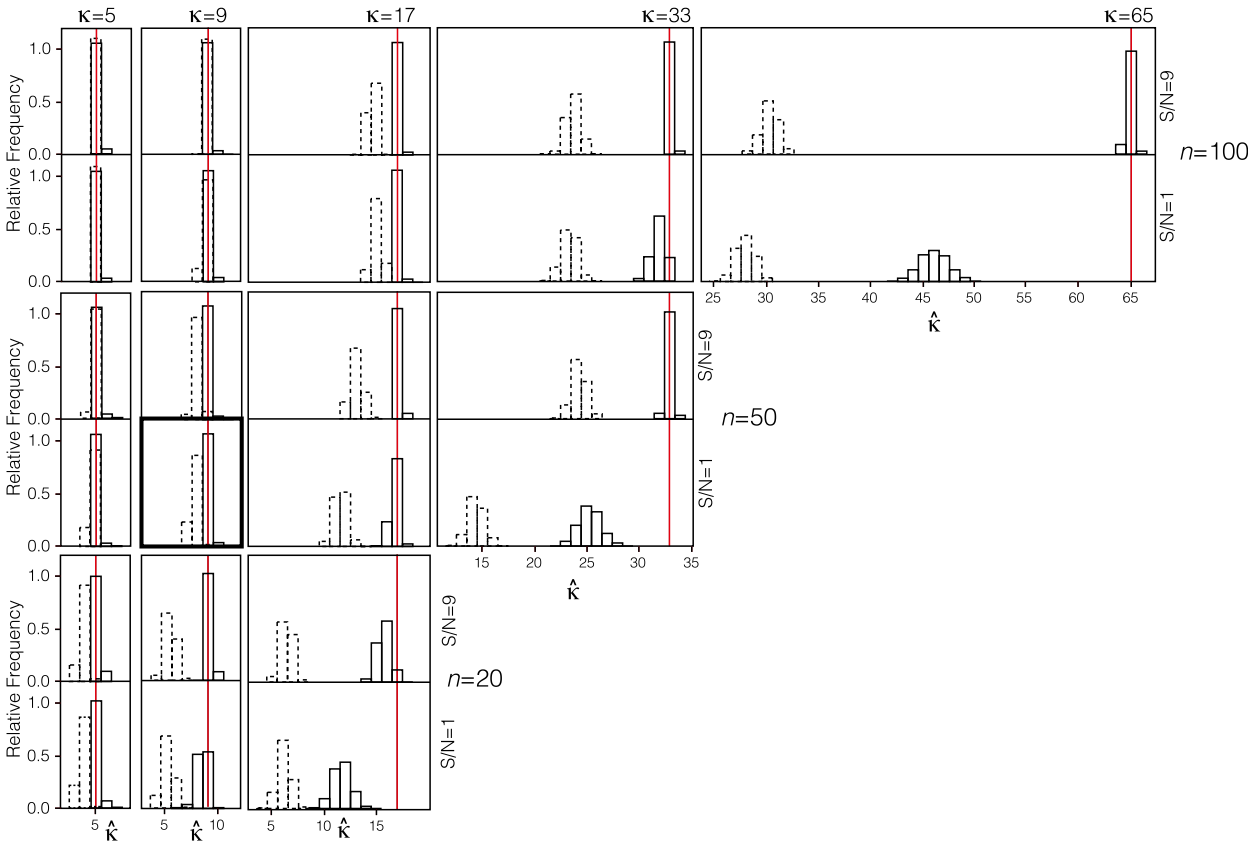

FIG. 1. Relative frequencies of estimated signal dimension, by Rule N (dashed) and the Tracy–Widom modification (solid). Correct signal dimensions are indicated by the red lines. Parameter combination indicated by the heavy box outline pertains to results in Fig. 2.

approximation to Tracy–Widom distributions, relative to the corresponding empirical distributions obtained through Monte Carlo simulation using the method described in the appendix. The positive tabulated values indicate that Tracy–Widom distributions overspecify this quantile, which leads to conservatism in the hypothesis tests (i.e., when $\kappa = k - 1$, the null hypothesis H_k is rejected less frequently that the nominal $\alpha_{\text{crit}} = 0.05$). However, this conservatism is slight unless either the sample size or the data dimension is fairly small, and the errors will be acceptable for many climatological applications. Similar results (not shown) are exhibited for other high quantiles. Fortunately, when either the sample size or data dimension is relatively small, the relevant reference distributions can be derived fairly quickly through Monte Carlo simulations, because (as outlined in the appendix) only the first sample eigenvalues $\hat{\lambda}_1$ need to be computed.

3. Performance in synthetic-data settings

In this section the performance of the Tracy–Widom-based modification to Rule N is investigated in synthetic-data settings, for which the correct κ can be known in advance. The data dimension is taken to be $M = 1500$, which is typical of the numbers of grid points in climate studies (e.g., Wallace and Gutzler 1981; Barnston 1994; Wilks 2008), and results will be shown for sample sizes n of 20, 50, and 100. Since $n << M$ these simulations are challenging settings for estimating PC truncations. Results will be shown for signal subspace dimensions, κ, of 5, 9, 17, 33, and 65 (provided $\kappa < n$); signal-to-noise (S/N) ratios

$$S/N = \sum_{k=1}^{\kappa} \lambda_k \Big/ \sum_{k=\kappa+1}^{M} \lambda_k \tag{9}$$

of 1 (half of total variance in the signal eigenvalues) and 9 (90% of total variance in the signal eigenvalues); and ratio of largest to smallest signal eigenvalues $\lambda_1/\lambda_\kappa = 8$, with intermediate signal eigenvalues varying linearly between the two extremes (results change little for values of this ratio in the range 2 through 32).

A large number of trials are computed for each combination of parameter values. In each trial, n synthetic M-dimensional data vectors $\boldsymbol{x}_i$ have been generated using

$$\mathbf{x}_i = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{E}^{\mathrm{T}}\mathbf{z}_i, \quad i = 1, \ldots, n, \tag{10}$$

where $\mathbf{\Lambda}^{1/2} = \mathrm{diag}\,(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots \sqrt{\lambda_\kappa}, \sigma_{\mathrm{noise}}, \sigma_{\mathrm{noise}}, \ldots \sigma_{\mathrm{noise}})$, from which synthetic sample covariance matrices and their $n - 1$ nonzero eigenvalues were computed. Each of the M columns of the $(M \times M)$ matrix $\mathbf{E}$ represents a generating-process eigenvector (i.e., the basis vectors estimated by the EOFs), the M elements of the diagonal matrix $\mathbf{\Lambda}^{1/2}$ are the square roots of the signal and noise eigenvalues in Eq. (1), and each $\mathbf{z}_i$ is an M-dimensional realization of independent standard Gaussian random variates. Although in a physically real setting the κ signal eigenvectors would usually exhibit larger spatial scales than the M–κ noise eigenvectors, here all M of these vectors have been generated through a Gram–Schmidt orthonormalization of random directions in M space, which does not affect the results for estimation of the signal dimensions.

Figure 1 shows frequencies of estimated signal dimensions for the various parameter combinations according to Eq. (4), for Rule N (dashed) and the Tracy–Widom modification (solid histograms), over 1000 trials for each parameter combination. The vertical red lines in each panel locate the correct specification, $\hat{\kappa} = \kappa$. For $\kappa = 5$ (leftmost column) both methods perform well, although for $n = 20$ the undue conservatism of Rule N begins to be evident. For the four larger values of κ, Rule N chooses $\hat{\kappa}$ values that are much too small, whereas the Tracy–Widom modification yields excellent results except when κ is a large fraction of n and the signal-to-noise ratio is low. In these cases ($n = 20$ and $\kappa = 17$, $n = 50$ and $\kappa = 33$, $n = 100$ and $\kappa = 65$) the performance of Rule N is even worse. Only a small portion of the conservatism of the Tracy–Widom $\hat{\kappa}$ estimates in these low n/κ trials is due to use of the parametric distribution: the average $\hat{\kappa}$ is larger by approximately 0.5 when the Monte Carlo empirical distributions for the $k = (\kappa + 1)$th (i.e., first noise) eigenvalue (used as the reference "truth" in Table 1) are used to construct each hypothesis test.

The Tracy–Widom modification of Rule N exhibits good performance even though no adjustments have been made to account for the effect of computing the sequence of multiple hypothesis tests specified by Eq. (4). Figure 2 illustrates that this result is a fortuitous consequence of the nonapplicability of the Tracy–Widom distribution for describing the sampling variation of the $k = (\kappa + 2)$th (second noise) eigenvalue. This figure pertains to the parameter combination $n = 50$, $\kappa = 9$, and $S/N = 1$, which is indicated by the heavy box outline in Fig. 1, but is representative of the other parameter combinations also. Figure 2a shows the relative frequencies of p_{10} [i.e., the test p value for the first noise eigenvalue in Eq. (4)] over 10 000 trials, when using the

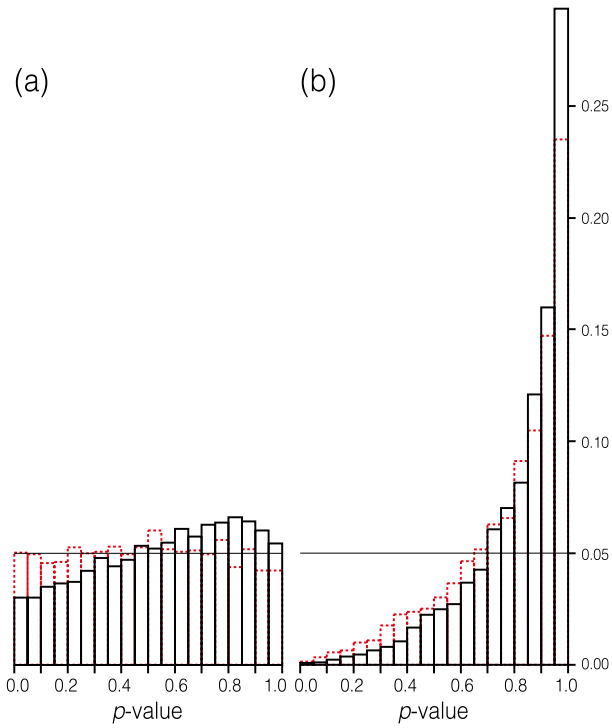

FIG. 2. Relative frequencies of p values computed using Monte Carlo (red dashed) and Tracy–Widom (black solid) reference distributions for (a) p_{10} for the H_{10} null hypothesis, and (b) p_{11} for the H_{11} null hypothesis, when $\kappa = 9$; with $n = 50$, and $S/N = 1$. Light horizontal lines locate ideal behavior.

empirical distribution of p_{10} (red dashed histograms) and the Tracy–Widom approximation to it (black solid histograms). The thin horizontal line located at $\alpha_{\mathrm{crit}} = 0.05$ locates ideal behavior for this histogram, and the close approximation of the dashed histogram to this ideal behavior supports the conjecture of Kritchman and Nadler (2008) noted in section 2. The height of the leftmost bar in Fig. 2a reflects the achieved test size when $\alpha_{\mathrm{crit}} = 0.05$, and shows that the actual test level is very close to $\alpha_{\mathrm{crit}} = 0.05$ when the empirical distribution of ℓ_{10}^* is used, and that only about 3% of these true null hypotheses are incorrectly rejected when the tests are computed using the Tracy–Widom distribution. This slight test conservatism is consistent with the results shown in Table 1.

Figure 2b shows the corresponding relative frequencies for p_{11}, which pertain to the test statistic ℓ_{11}^*, when $\kappa = 9$. For both the empirical and Tracy–Widom distributions, the probability that the true null hypothesis H_{11} is rejected is vanishingly small. Similar plots for p_{12}, p_{13}, p_{14}, etc. (not shown) exhibit even more extreme negative skewness. Accordingly, even though Eq. (4) specifies a sequence of multiple tests, if the first true null hypothesis $H_{\kappa+1}$ is erroneously rejected the probabilities are

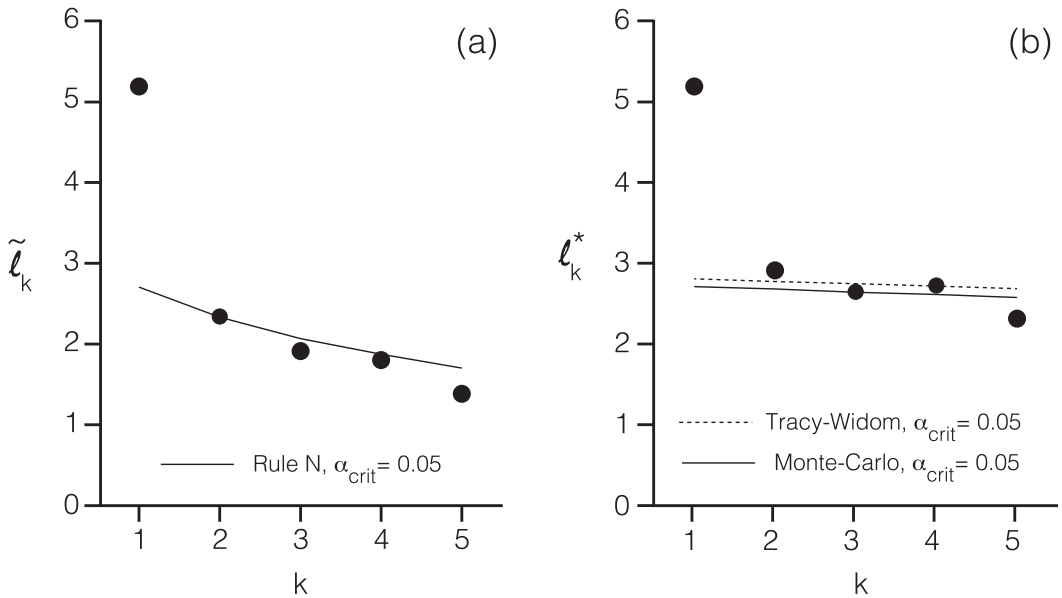

FIG. 3. Illustration of the use of (a) Rule N and (b) the modified Rule N, for the Bering Sea cyclone data of Overland and Preisendorfer (1982).

vanishingly small that $H_{\kappa+2}$, $H_{\kappa+3}$, etc., will also be rejected, obviating the multiple-testing problem in this setting.

4. Real-data example

Overland and Preisendorfer (1982) illustrated use of Rule N with a PCA relating to numbers of cyclones transiting an $M = 56$-member network of grid cells in the Bering Sea during October–February, during a period of $n = 23$ winters. Their results pertaining to the covariance matrix of cyclone counts are displayed graphically in Fig. 3a. Here the five leading sample eigenvalues, scaled according to Eq. (2) are plotted, together with the 95th percentile of the Monte Carlo distribution for these statistics. Accordingly $\tilde{\ell}_k$ points above the curve indicate $p_k \leq \alpha_{crit} = 0.05$, so that Eq. (4) yields $\hat{\kappa} = 2$. Overland and Preisendorfer (1982) concluded that only the two leading eigenvalues are statistically distinguishable from noise.

Figure 3b shows the corresponding result when Eq. (4) is evaluated using Tracy–Widom distributions (dashed curve), and 10 000-member Monte Carlo distributions (solid curve) for ℓ^*_k [Eq. (5)]. Table 1 indicates that for this small sample size and data dimension, the Tracy–Widom distribution overestimates the 95th percentile for the pure-noise ℓ^*_k statistics by approximately 4%, and so yields a test that is more conservative than the nominal $\alpha_{crit} = 0.05$. In Fig. 3b, Eq. (4) yields $\hat{\kappa} = 2$ when computing the p_k values from Tracy–Widom distributions, but $\hat{\kappa} = 4$ when using the direct Monte Carlo distributions. Because n and M are relatively small these Monte Carlo distributions can be computed very quickly, as outlined in the appendix.

5. Conclusions

This paper has proposed a modification of the popular Rule N for PCA truncation, based on the Tracy–Widom distribution for the largest noise eigenvalue of a sample covariance matrix. This new method improves upon Rule N because it accounts at each stage of the sequence of hypothesis tests for the results of previous tests in the sequence.

Both Rule N and the proposed modification assume the particular separation of signal and noise subspaces expressed in Eq. (1). Experiments using synthetic data conforming to this assumption show excellent results unless the dimension of the signal subspace is a large fraction of the sample size and the signal-to-noise ratio is relatively small, and even under these circumstances the proposed modification strongly outperforms the original Rule N. However, both of these methods may perform poorly in settings where the signal and noise subspaces are defined differently.

Even though the proposed procedure is based on a sequence of hypothesis tests, it is not necessary to account for this test multiplicity because the bias of Tracy–Widom distributions for other than the largest noise eigenvalues prevents the procedure from

greatly overestimating the dimension of the signal subspace.

The Tracy–Widom distribution provides a good representation for the random variations of the leading noise eigenvalue when sample size and data dimension are both moderate to large. When either or both of these parameters are relatively small the Tracy–Widom distribution yields conservative tests and truncation points; that is, too few of the leading eigenvalues are regarded as representing signal, on average, which will lead to concluding that some signal modes are noise. In this circumstance the appropriate reference distributions can be easily and quickly generated by Monte Carlo methods, as described in the appendix.

The results presented here were based on underlying Gaussian-distributed data, but the sampling distributions of sample covariance-matrix eigenvalues are very similar for uniform, or random sign (± 1 with equal probability) data (Faber et al. 1994). However, if sensitivity to non-Gaussian data is suspected, the appropriate reference distributions can again be derived using Monte Carlo methods as described in the appendix.

Acknowledgments. I thank the anonymous reviewers, whose comments lead to an improved presentation. This research was supported by the National Science Foundation under Grant AGS-1112200.

APPENDIX

Monte Carlo Computation of the Reference Distributions

A different sampling distribution must be computed for each null hypothesis H_k [Eq. (3)], $k = 1, 2, 3, \ldots$, pertaining to the largest sample eigenvalue of a pure noise covariance matrix formed from random data of dimension M_k, with sample size n_k, where $M_k = M - k + 1$ and $n_k = n - k + 1$. If either the sample size n or the data dimension M is relatively small, the Tracy–Widom distribution may represent inadequately the sampling distribution of the largest noise eigenvalue, as indicated in Table 1. In this case discrete approximations to the appropriate distributions can be computed relatively cheaply using the power method (e.g., Golub and Van Loan 1983), because only the leading eigenvalue rather than all of the nonzero eigenvalues of each sample covariance matrix needs to be computed. This approach can also be used if the effects of non-Gaussian data are suspected to be significant, but with possibly large computational expense if the smaller of n_k and M_k is not relatively small.

Each Monte Carlo iteration (of perhaps 1000) begins with a $(n_k \times M_k)$ matrix $\mathbf{X}'$ of uncorrelated data anomalies (sample means of each of the M_k columns have been subtracted from the n_k values in that column). These underlying data may be random numbers from a Gaussian or other distribution, or bootstrapped samples from an actual dataset. A sample noise covariance matrix $\mathbf{S}$ is then formed as

$$\underset{(M_k \times M_k)}{\mathbf{S}} = \frac{1}{n_k - 1}\mathbf{X}'^{\mathrm{T}}\mathbf{X}' \quad \text{if} \quad n_k > M_k \tag{A1a}$$

or

$$\underset{(n_k \times n_k)}{\mathbf{S}} = \frac{1}{n_k - 1}\mathbf{X}'\mathbf{X}'^{\mathrm{T}} \quad \text{if} \quad n_k \le M_k. \tag{A1b}$$

Because the nonzero eigenvalues of Eqs. (A1a) and (A1b) are the same (e.g., von Storch and Zwiers 1999, p. 300; Wilks 2011, p. 554), the computations can be greatly reduced using Eq. (A1b) in the common circumstance that $n << M$. Because of the normalization in Eq. (5), the division by $n_k - 1$ is not actually necessary. Note that Eq. (A1b) does not describe a "T-mode" analysis (Compagnucci and Richman 2008), because the anomalies in $\mathbf{X}'$ are computed by subtracting the M column means of $\mathbf{X}$, and not the n row means.

The power method (e.g., Golub and Van Loan 1983) is then used to extract the leading eigenvalue λ_1 of $\mathbf{S}$ for the current iteration. For a given value of k, these leading noise eigenvalues collectively compose the discrete Monte Carlo approximation to the desired distribution. Beginning with an arbitrary initial guess for the leading eigenvector $\mathbf{e}$, with $||\mathbf{e}|| = 1$, the algorithm proceeds by iterating until convergence:

$$\mathbf{v} = \mathbf{S}\mathbf{e}, \tag{A2a}$$

$$\lambda_1 = ||\mathbf{v}||, \tag{A2b}$$

$$\mathbf{e} = \mathbf{v}/\lambda_1. \tag{A2c}$$

Here $\mathbf{v}$ is an intermediate storage vector, and $||\mathbf{v}||$ indicates its Euclidean length. After the first Monte Carlo evaluation it is convenient to begin by taking $\mathbf{e}$ as the final value from the previous cycle.

REFERENCES

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564, doi:10.1175/1520-0442(1994)007<1513:LSSTCP>2.0.CO;2.

Chiani, M., 2014: Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution. *J. Multivar. Anal.*, **129**, 69–81, doi:10.1016/j.jmva.2014.04.002.

Compagnucci, R. H., and M. B. Richman, 2008: Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int. J. Climatol.*, **28**, 703–726, doi:10.1002/joc.1574.

Davis, R. E., 1976: Predictability of sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, **6**, 249–266, doi:10.1175/1520-0485(1976)006<0249:POSSTA>2.0.CO;2.

Faber, N. M., L. M. C. Buydens, and G. Kateman, 1994: Aspects of pseudorank estimation methods based on the eigenvalues of principal component analysis of random matrices. *Chemom. Intell. Lab. Syst.*, **25**, 203–226, doi:10.1016/0169-7439(94)85043-7.

Feng, J., Q. Wang, S. Hu, and D. Hu, 2016: Intraseasonal variability of the tropical Pacific subsurface temperature in the two flavours of El Niño. *Int. J. Climatol.*, **36**, 867–884, doi:10.1002/joc.4389.

Golub, G. H., and C. F. Van Loan, 1983: *Matrix Computations.* John Hopkins Press, 476 pp.

Johnstone, I. M., 2001: On the distribution of the largest eigenvalue in principal component analysis. *Ann. Stat.*, **29**, 295–327, doi:10.1214/aos/1009210544.

Jolliffe, I. T., 2002: *Principal Component Analysis.* Springer, 487 pp.

Kritchman, S., and B. Nadler, 2008: Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.*, **94**, 19–32, doi:10.1016/j.chemolab.2008.06.002.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Science Rep. 1, Statistical Forecasting Project, Department of Meteorology, MIT (NTIS AD 110268), 49 pp.

Obukhov, A. M., 1947: Statistically homogeneous fields on a sphere. *Uspethi Math. Nauk*, **2**, 196–198.

Ortega, P., F. Lehner, D. Swingedouw, V. Masson-Delmotte, C. C. Raible, M. Casado, and P. Yiou, 2015: A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature*, **523**, 71–77, doi:10.1038/nature14518.

Overland, J. E., and R. W. Preisendorfer, 1982: A significance test for principal components applied to a cyclone climatology. *Mon. Wea. Rev.*, **110**, 1–4, doi:10.1175/1520-0493(1982)110<0001:ASTFPC>2.0.CO;2.

Pérez-Hernández, M., and T. M. Joyce, 2014: Two modes of Gulf Stream variability revealed in the last two decades of satellite altimeter data. *J. Phys. Oceanogr.*, **44**, 149–163, doi:10.1175/JPO-D-13-0136.1.

Preisendorfer, R. W., and C. D. Mobley, 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, 425 pp.

——, F. W. Zwiers, and T. P. Barnett, 1981: *Foundations of Principal Component Selection Rules.* SIO Reference Series 81-4, Scripps Institution of Oceanography, 192 pp.

Tracy, C. A., and H. Widom, 1996: On orthogonal and symplectic matrix ensembles. *Commun. Math. Phys.*, **177**, 727–754, doi:10.1007/BF02099545.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, Y., R. M. Castelao, and Y. Yuan, 2015: Seasonal variability of alongshore winds and sea surface temperature fronts in eastern boundary current systems. *J. Geophys. Res. Oceans*, **120**, 2385–2400, doi:10.1002/2014JC010379.

Wilks, D. S., 2008: Improved statistical seasonal forecasts using extended training data. *Int. J. Climatol.*, **28**, 1589–1598, doi:10.1002/joc.1661.

——, 2011: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 676 pp.