AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL068880

Key Points:
- Atmospheric temperature variability is nonnormal on daily time scales
- Apparent normal variability on synoptic scales is an artifact of filtering
- Generating mechanism for nonnormality does not necessarily act at frequency where greatest nonnormality is observed

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5

Correspondence to:
C. Proistosescu,
cproist@fas.harvard.edu

Citation:
Proistosescu, C., A. Rhines, and P. Huybers (2016), Identification and interpretation of nonnormality in atmospheric time series, *Geophys. Res. Lett.*, *43*, 5425–5434, doi:10.1002/2016GL068880.

Received 8 DEC 2015
Accepted 20 APR 2016
Accepted article online 23 APR 2016
Published online 20 MAY 2016

Identification and interpretation of nonnormality in atmospheric time series

Cristian Proistosescu[1], Andrew Rhines[2], and Peter Huybers[1]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA, [2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA

Abstract Nonnormal characteristics of geophysical time series are important determinants of extreme events and may provide insight into the underlying dynamics of a system. The structure of nonnormality in winter temperature is examined through the use of linear filtering of radiosonde temperature time series. Filtering either low or high frequencies generally suppresses what is otherwise statistically significant nonnormal variability in temperature. The structure of nonnormality is partly attributable to geometric relations between filtering and the appearance of skewness, kurtosis, and higher order moments in time series data, and partly attributable to the presence of nonnormal temperature variations at the highest resolved frequencies in the presence of atmospheric memory. A nonnormal autoregressive model and a multiplicative noise model are both consistent with the observed frequency structure of nonnormality. These results suggest that the generating mechanism for nonnormal variations does not necessarily act at the frequencies at which greatest nonnormality is observed.

1. Introduction

Departures from normality in temperature have important implications for the frequency of extreme events [*Ruff and Neelin*, 2012], and discerning the statistical characteristics of nonnormality can give insight into relevant physical processes [*Sura and Perron*, 2010]. A wide variety of mechanisms have been proposed to explain the observed nonnormality [*Sura and Hannachi*, 2015]. One class of mechanisms posits that nonnormality arises on synoptic time scales, either from the effects of state-dependent or multiplicative noise [e.g., *Sura et al.*, 2005; *Sardeshmukh and Sura*, 2009; *Sura and Perron*, 2010] or from additive noise with nonnormal perturbations [e.g., *Luxford and Woollings*, 2012]. The second class of mechanisms ascribes nonnormality to low-frequency variability, such as nonnormal distributions arising as mixtures of Gaussian processes, introduced through nonlinear regime shifts or nonstationarity [e.g., *Hannachi*, 2010].

Sardeshmukh and Sura [2009] note that the relationship between skewness and kurtosis in observations of midlatitude variability is consistent with multiplicative noise acting on synoptic time scales. *Schneider et al.* [2015], however, have shown that midlatitude temperature records filtered to the canonical synoptic time scales appear normal within 3 standard deviations, whereas deviations from normality are apparent when including variability at time scales longer than 15 days. Similarly, *Rennert and Wallace* [2009] use geopotential height data filtered to three different frequency bands to show that skewness in atmospheric variability is related to cross-frequency coupling between intermediate and low-frequency time scales.

Considering these seemingly conflicting interpretations it is useful to distinguish structure indicative of physical processes from structure related to the generic effects of filtering time series. *Rosenblatt* [1961] noted that "[i]t appears to be part of the engineering folklore that a narrow band-pass filter applied to a stationary random input yields an output that is approximately normally distributed." The tendency of filtering to alter the perceived normality of the data has been noted for other signals [e.g., *Rozanov*, 1961; *Mallows*, 1967; *Papoulis*, 1972]. *Donohoe and Battisti* [2009] describe how the asymmetry in the distribution of synoptic cyclones and anticyclones depends on the choice of temporal or spatial filtering techniques, and a similar tendency toward normality has been noted for spatial filtering in the form of gridding [*Director and Bornn*, 2015; *Cavanaugh and Shen*, 2015].

An overall tendency for filtering to make time series data appear more normal can be explained from a Fourier perspective, whereby filtering nullifies interactions between frequencies that are necessary for representing

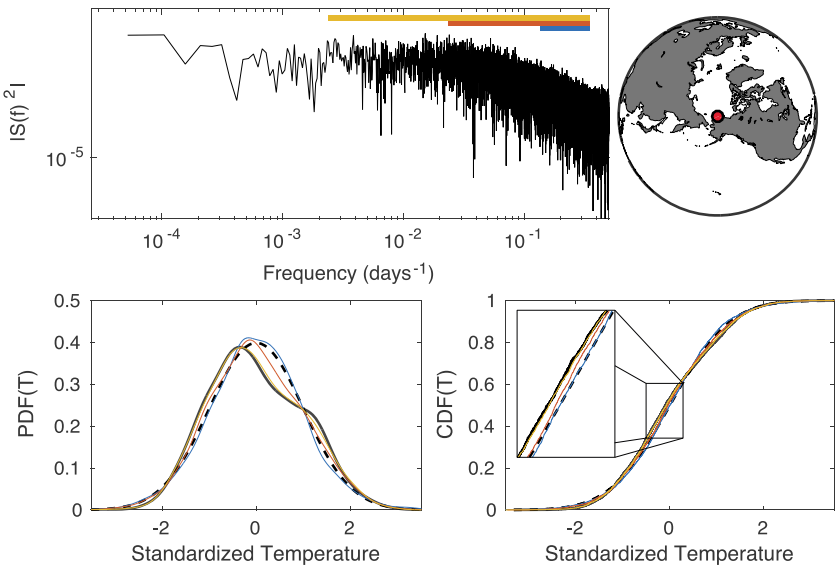

Figure 1. Filtering of radiosonde temperature time series and normality. (top) Power spectrum of a radiosonde time series of temperature at 850 mbar from Barrow, AK. The annual cycle is removed by notch filtering the annual frequency and its first two harmonics. The time series is then progressively filtered using a top-hat band-pass filter. The admitted frequency band for each filter is depicted by color bars indicating frequencies of 1/3–1/7 days $^{-1}$ (blue), 1/3–1/40 days^{-1} (orange), and 1/3–1/400 days^{-1} (yellow). (bottom) The probability density function and cumulative density function are depicted for winter (DJF) temperatures for the full data (solid black line), the filtered time series (solid colored lines), and a standardized normal distribution (dashed line).

nonnormal structure [*Kotulski and Sobczyk*, 1981; *Garth and Bresler*, 1997]. There remains the possibility, however, that distinct structure in the nonnormality of filtered records can provide physical insight. In the following we examine radiosonde time series, first demonstrating statistically significant nonnormality, then analytically describing how this nonnormal structure can be expected to decay upon filtering, and finally describing more nuanced structures associated with autocorrelation using two simple numerical models. We describe nonnormality using the third and forth moments of the distribution, i.e., skewness and kurtosis. Intuitively, skewness reflects the asymmetry of the distribution with positive values indicating greater weight in the upper tail, while kurtosis reflects overall weight placed in the tails relative to the center of the distribution.

2. Radiosonde Temperature Data

As an initial example, Figure 1 shows the distribution of wintertime radiosonde temperature from the Barrow, Alaska station. Data are obtained from the Integrated Global Radiosonde Archive [*Durre et al.*, 2006] and are subset to only 12:00 Greenwich Mean Time soundings at 850 mbar. The 1958–2009 interval is selected for analysis because only 2% of samples are missing, and these are infilled by linear interpolation. We remove the annual cycle and its first two harmonics by subtracting least squares fit sine waves, though results are similar when the annual cycle is retained (supporting information Figure S1). The sample distribution of the Barrow data involves a positive skew and is significantly nonnormal ($p < 0.01$). Significance is assessed using a Kolmogorov-Smirnov (KS) test [*Smirnov*, 1939], where the largest deviation between the sample cumulative distribution and that of a normal distribution is evaluated. In order to account for autocorrelation in the time series, the null distribution is built from normally distributed surrogate data generated by phase randomizing the original time series (see supporting information). Rejection of normality can appear either from features of subseasonal variability or interannual variability. For instance, normal variability that is subjected to an interannual trend in the mean would give the appearance of nonnormality [*Huybers et al.*, 2014]. Interseasonal variability, such as increased variance in winter can also lead to increased values of kurtosis [*Frisch and Sornette*, 1997]. Interannual contributions and interseasonal variability is suppressed in our analysis through only considering December through February winter months (DJF) and normalizing each seasonal realization to zero mean and unit variance.

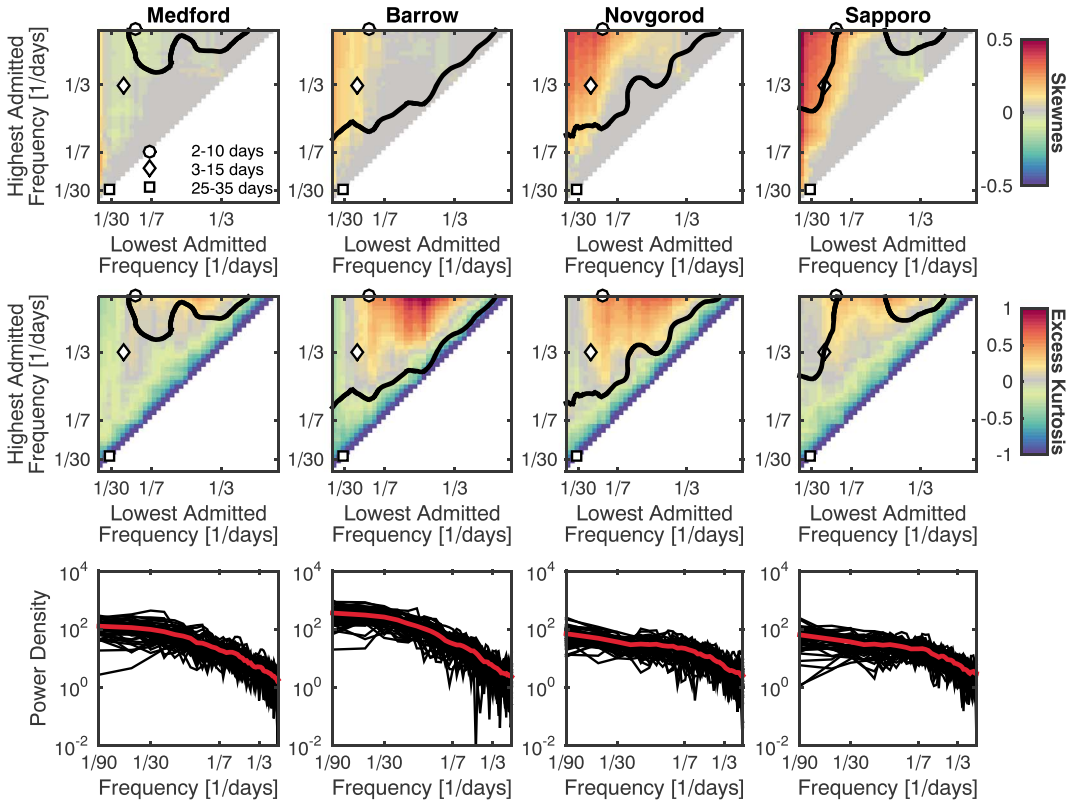

Figure 2. Effects of filtering on (top row) skewness and (middle row) excess kurtosis of 850 mbar DJF temperature from radiosonde stations for the full range of passbands. Axes denote passband limits. Each season is treated as an independent ensemble member, standardized to zero mean, and unit variance. Values of sample S and K averaged are over all seasons. Lines parallel to the diagonal have equal bandwidth for the passband, with the width of the passband increasing further away from the diagonal. Points below the solid black lines cannot be distinguished from samples drawn from a normal distribution under a Kolmogorov-Smirnov (KS) test, at 95% confidence level. Markers denote choices of filter passband discussed in the text. (bottom row) Spectral estimates of each individual season (black) and the mean spectrum (red). Note the logarithmic frequency axis for spectral estimates.

One indication of the relevance of understanding the origins of deviations from normality are that they reach 0.8 and 0.3°C for the 5th and 95th percentiles, respectively, and 2.3 and −0.5°C for the 1st and 99th percentiles (see also Table S1 in the supporting information). Such deviations can lead to considerable changes in the rates of exceedance of temperature thresholds, particularly in association with changes in mean climate [*Ruff and Neelin*, 2012]. Nonnormality in the Barrow data is also consistent with a more comprehensive study [*Perron and Sura*, 2013] that found statistically significant deviations from nonnormality in daily reanalyses data for nine atmospheric variables, including temperature. Surface temperature from the U.S. Global Historical Climatology Network (GHCN) has also been shown to exhibit significant nonnormality, as well as trends in higher-order moments and quantiles [*Huybers et al.*, 2014; *Cavanaugh and Shen*, 2014].

The Barrow radiosonde data become more consistent with a normal distribution upon filtering. For example, filtering to only retain synoptic scale anomalies at 1/3–1/15 day^{-1} frequencies alters the Barrow results to be substantially more consistent with a normal distribution ($p = 0.035$ using an adjusted KS test, $p = 0.15$ using a standard KS test). Filtering is performed using a simple top-hat filter whereby Fourier components outside the passband are set to zero. This approach is not optimal from the perspective of suppressing Gibbs phenomena but is amenable to later analytic calculations. Furthermore, no appreciable differences are obtained when using more sophisticated filters, such as a forward-backward pass with a Butterworth filter.

The example in Figure 1 can be generalized (Figure 2) to show how nonnormal components vary according to filtering. In addition to Barrow (USA, 1958–2009), we examine how skewness and excess kurtosis of three other long and nearly complete records from Medford (USA, 1958–2009), Novgorod (Russia, 1966–2009), and Sapporo (Japan, 1965–2009) change as a function of high and low cutoff frequencies specified for filtering.

Data are obtained and processed in a manner analogous to the Barrow example. A general pattern holds wherein skewness and kurtosis diminish with increasingly narrow filter bandwidths, but more nuanced structure is also present. Positive skewness appears when the lowest frequencies of 1/90 days^{-1} are admitted but which diminishes to near zero once low-frequency cutoffs of 1/7 days^{-1} or higher are specified. Relative to the low-frequency cutoff, changing the high-frequency cutoff has only a weak influence on skewness, except for the presence of skewness extending out toward a frequency band near 1/3–1/4 days^{-1}. In contrast to skewness, excess kurtosis is negative when admitting the lowest frequencies and is maximized when passing frequencies between 1/5 and 1/2 days^{-1}, roughly encompassing synoptic time scales. Excess kurtosis also becomes strongly negative when the passband becomes extremely narrow.

Significance of nonnormality in the radiosonde time series is evaluated for each filter combination using the previously described KS test. All records show statistically significant nonnormality in those regions having the greatest magnitudes of skewness and excess kurtosis. Higher-order moments are also present but decay more quickly under filtering and make only minor contributions to overall nonnormality.

3. Why Filtering Tends to Generate Normality

The tendency of a time series to become more normally distributed after filtering can be understood through the use of higher-order spectra [*Brillinger*, 1965; *Garth and Bresler*, 1997]. The power spectrum, bispectrum, and trispectrum of a process $x(t)$ can be defined in terms of the Fourier Transform $\hat{x}(f)$ as,

$$P(f_1) = \hat{x}(f_1) \cdot \hat{x}^*(f_1), \tag{1a}$$

$$B(f_1, f_2) = \hat{x}(f_1) \cdot \hat{x}(f_2) \cdot \hat{x}^*(f_1 + f_2), \tag{1b}$$

$$T(f_1, f_2, f_3) = \hat{x}(f_1) \cdot \hat{x}(f_2) \cdot \hat{x}(f_3) \cdot \hat{x}^*(f_1 + f_2 + f_3). \tag{1c}$$

Higher-order spectra follow a relation similar to Parseval's theorem relating the power spectrum and variance, wherein the surface integral over the bispectrum is related to skewness, S, and the volume integral over the trispectrum is related to excess kurtosis, K,

$$S = \frac{\int B(f_1, f_2) \cdot \mathrm{d}f_1 \cdot \mathrm{d}f_2}{\left(\int P(f_1) \cdot \mathrm{d}f_1\right)^{3/2}}, \tag{2a}$$

$$K = \frac{\int T(f_1, f_2, f_3) \cdot \mathrm{d}f_1 \cdot \mathrm{d}f_2 \cdot \mathrm{d}f_3}{\left(\int P(f_1) \cdot \mathrm{d}f_1\right)^{4/2}}. \tag{2b}$$

Deviations from a normal distribution are encoded in interactions between different frequency bands in the Fourier representation of time series and are modified by filtering. The specific effects of filtering $x(t)$ upon skewness and kurtosis can be obtained by replacing $\hat{x}(f)$ with a filtered version, $\hat{y}(f) = \hat{h}(f) \cdot \hat{x}(f)$, in equations (1) and (2). Excluding a frequency f_i eliminates all interacting frequencies pairs, such that the unfiltered area of the bispectrum generally diminishes more rapidly than that of the power spectrum. For example, a low-pass filter that retains only a fraction β of unfiltered frequencies will lead to a bispectrum with only $3\beta^2/4$ of its area being unfiltered for $\beta < 2/3$. A geometric depiction of the effects of band-pass filtering upon the bispectrum is given in Figure 3, and a similar depiction for the trispectrum is given in the supporting information (Figure S2). Although the patterns are visually simple, their analytical computation involves integrals over higher-order spectral volumes, which can become rather involved [*Büeler et al.*, 2000]. An analytical solution for bispectrum filtering is given in the supporting information.

Figure 4 shows the fraction of skewness and excess kurtosis admitted as a function of filtered cutoff frequencies. There is excellent agreement between these analytically computed values and those obtained from filtering a random time series containing 10^7 realizations from a Pearson distribution. The Pearson distribution is chosen because it is specifically derived as a class of models used to fit a nonnormal distribution based on skewness and kurtosis and because it provides a good fit to the radiosonde record. The results remain unchanged, however, if the simple models are forced with beta, gamma, or chi-square distributions. Similarly, the results hold for the case of correlated additive and multiplicative noise, where non-normality arises from interactions within the model, irregardless of whether prescribed in the forcing. When the time series is

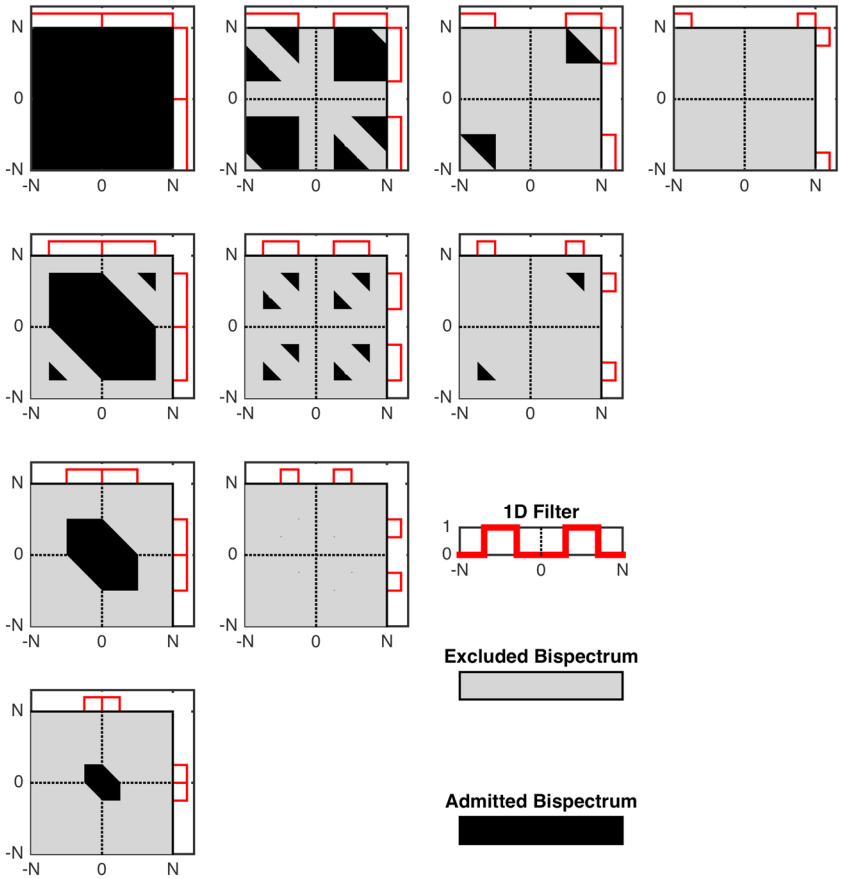

Figure 3. Effects of filtering on the bispectrum. Frequency axes are normalized relative to the Nyquist frequency (N), and the solid color depicts the part of the bispectrum respectively that is allowed to pass under a band-pass filter. (first to fourth rows) Greater filtering of high frequencies. (first to fourth column) Greater filtering of low frequencies. Note that both negative and positive frequencies contribute to the bispectrum.

reduced to 10^5 points, the match becomes more noisy, with the implication that the finite radiosonde time series we analyze can be expected to have variable skewness and kurtosis structure even if the underlying statistical distributions are identical.

The analytical results describe the main features of the skewness observed in the filtered radiosonde data (Figure 4, left). Filtering low frequencies more rapidly decreases skewness than filtering high frequencies because, as depicted in Figure 3, the sum of low-frequency pairs excludes more area of the bispectrum not already excluded by filtering a single member of the pair. The analytical results also explain a lobe of high skewness that extend out toward frequencies in the vicinity of half the Nyquist frequency, or in the case of the daily sampled radiosonde time series, 1/4 days^{-1}. This lobe is again a consequence of asymmetries in how filtering certain frequencies also excludes energy at interaction frequencies (see equation (1)).

In contrast with skewness, correspondence between the kurtosis obtained from our analytical results and the radiosonde observations is poor. The Fourier representation shows contours of kurtosis that align with contours of constant filter bandwidth (Figure 4, right), whereas the data show maximum kurtosis when filtering everything but frequencies corresponding to synoptic time scales. Negative values of excess kurtosis observed when filtering all but a small number of frequencies (the diagonals in Figure 2) reflect the fact that the excess kurtosis of a sine wave is −1.5. For the case of a normal process, wherein the Fourier coefficients are independently distributed, it can be shown that the expected sample kurtosis converges to the process excess Kurtosis of zero as $K \propto -3/2n$, where n is the number of consecutive frequencies retained in the Fourier spectrum (see Appendix A). Analytical results do not display this sinusoidal limit on account of continuous frequency resolution. The remainder of the mismatch between the analytical and observed kurtosis, however,

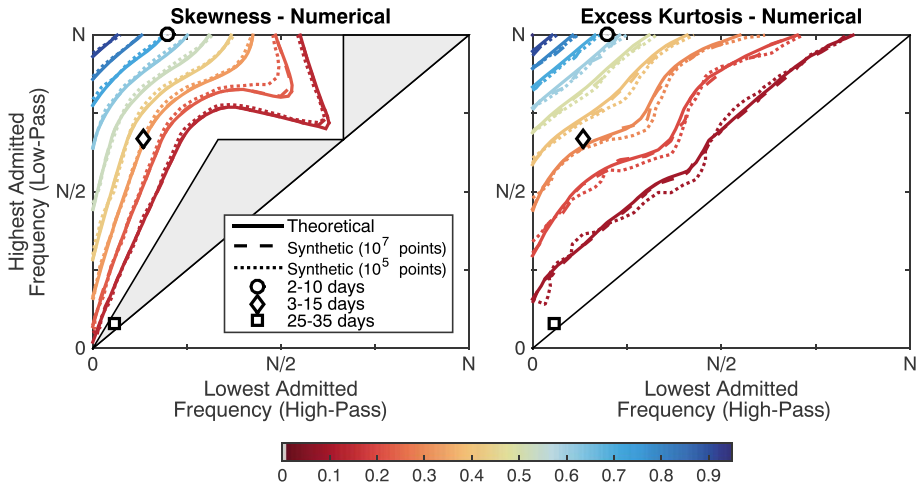

Figure 4. Skewness and excess kurtosis under band-pass filtering, relative to unfiltered values. Axes denote passband limits, relative to the Nyquist frequency. Gray shaded area denotes values of skewness of exactly zero. (left) Numerical computation of a filtered isotropic bispectrum and synthetic estimations based on independent nonnormal noise. (right) Same as middle but for excess kurtosis. Details of the numerical estimations are given in Appendix B.

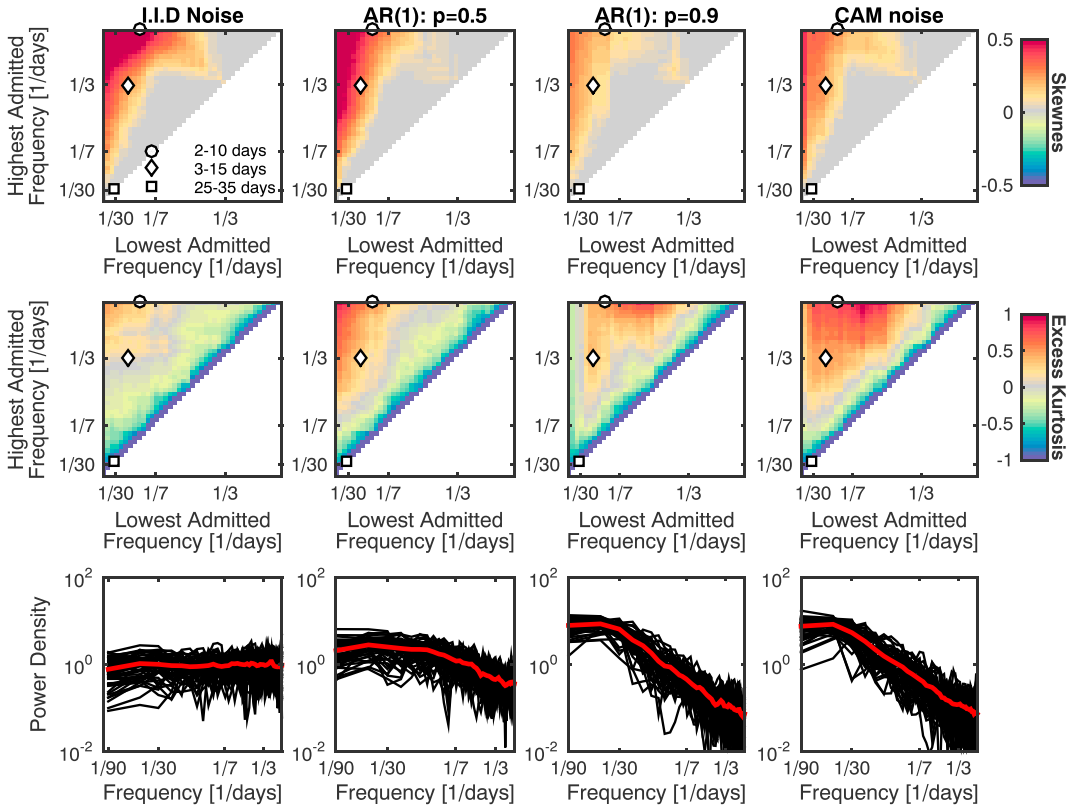

Figure 5. Same as Figure 2 but for synthetic data. (first to right columns) Results of models that are independent and identically distributed (i.i.d.), autoregressive order one with parameters of $\rho = 0.5$ and $\rho = 0.9$ (equation (3)), and correlated additive and cumulative noise (equation (4)). The i.i.d. and autoregressive models are driven by nonnormal realization from a Pearson distribution. For each model, 50 independent ensemble members of length 90 are generated, with each member representing a season. Plotted results are the average after analyzing each season, except for the bottom row which also shows individual realizations in black.

points to substantive differences between filtering independent realizations from a Pearson distribution and actual temperature variability.

4. Numerical Models

An obvious deficiency in the foregoing analytical results is the presumption of independent realizations, whereas spectral estimates of radiosonde temperature have increasing energy toward lower frequencies (Figures 1 and 2), indicative of memory between subsequent values. We represent correlations in atmospheric data using an autoregressive process of order one, meant to represent the generally more energetic variability found at lower frequencies, either because of preferential damping of high-frequency variations or the presence of memory in the record, e.g., due to the long time scales associated with ocean thermal inertia [*Hasselmann*, 1976],

$$x(t+1) = \rho x(t) + \eta(t), \tag{3}$$

where η represents independent realizations from a Pearson distribution. Specifying $\rho = 0$ gives the same skewness and kurtosis structure discussed previously (Figure 4), at least up to variations associated with finite samples. Setting $\rho = 0.9$, however, gives results in agreement with observations (Figure 5), especially with respect to maximum kurtosis at synoptic periods.

We also consider a discretized version of the model presented by *Sardeshmukh and Sura* [2009] that involves correlated additive and multiplicative noise. This is equivalent to an AR(1) process forced with nonnormal noise, wherein the nonnormality arises from state dependency of perturbations. Such state dependency has been modeled as arising from stochastic damping of Rossby waves [*Sura et al.*, 2005] or stochastic advection of potential vorticity anomalies [*Sura and Perron*, 2010]. The discretized model can be written as

$$x(t+1) = \rho x(t) + b\eta_1(t) + (Ex(t) + g)\eta_2(t) - \frac{1}{2}Eg, \tag{4}$$

with η_1 and η_2 representing independent realizations of a standard normal distribution. The first term represents the autocorrelated nature of the process, the second and third term respectively represent the normal and nonnormal innovations, and the last term ensures the process is stationary. Nonnormality is introduced by a state-dependent and asymmetric amplification of the normal forcing, η_2. Specifying $\rho = 0.9$, $b = 0.1$, $E = 0.2$, and $g = 0.1$ gives output with a similar structure of skewness and kurtosis to the radiosonde data (Figure 5).

Both numerical models reproduce the lobe of increased kurtosis present in radiosonde data. Heuristically, this lobe seems to be a consequence of the fact that high-pass filtering autocorrelated data to time scales shorter than the decorrelation time introduces a strong variance heteroskedasticity.

5. Further Discussion and Conclusions

Both the additive and multiplicative noise models fit the observations through inheriting nonnormality from synoptic scale processes. This result can be reconciled with *Schneider et al.* [2015] finding normal variability at synoptic time scales but nonnormal variability at longer time scales through considering the effects of band-pass filtering. Filtering to 1/3–1/15 days^{-1} synoptic frequencies generally leads to insignificant deviations from normality in the radiosonde data (Figure 2), in agreement with analytical and numerical predictions (Figures 4 and 5). The findings of nonnormal variability for a 1/25–1/35 day^{-1} passband result from the discrete frequency basis only retaining a small number of nonzero frequency terms and, thus, tending toward a distribution determined by the sinusoidal Fourier basis. For completeness, we have also verified that applying the aforementioned filters to the five ERA-Interim 850 hPa grid boxes examined by *Schneider et al.* [2015] gives equivalent results.

Rennert and Wallace [2009] also examine the presence of nonnormality as a function of retained frequency. Using data filtered to retain frequencies below (30 day)$^{-1}$, between (6–30 day)$^{-1}$, and above (6 day)$^{-1}$, they attribute skewness to the coupling between the low and intermediate frequency bands. Nonnormality is then postulated to arise from processes with corresponding time scales, e.g., global teleconnections and Rossby waves. Our findings, however, indicate that nonnormality introduced at the highest resolved time scales in the presence of memory leads to the appearance of greatest skewness when low frequencies are retained (Figure 5). These results suggest that the generating mechanism for nonnormal variations does not necessarily act at the frequencies at which greatest nonnormality is observed.

Instead, our findings are consistent with processes introducing nonnormality on time scales faster than the local decorrelation time scale. *Schneider et al.* [2015] suggest that midlatitude temperature distributions are controlled by anomalies advected along a uniform gradient. Such a model may also account for nonnormality introduced at synoptic time scales as additive noise, for example, through asymmetries in advection length scale or deviations from a constant gradient introduced by the finite size of the planet or the presence of jets [*Luxford and Woollings*, 2012]. Multiplicative noise can also be introduced by covariance between the advection flow field and temperature gradient [*Sura and Perron*, 2010]. It remains unclear, however, whether the highest resolved frequency of 1/2 day^{-1} is near that of the process giving rise to the nonnormal distribution. Nonnormality could also be introduced by atmospheric turbulence at much higher frequencies [*Chu et al.*, 1996], and it will be useful to examine higher-resolution temperature records in future work.

The frequency structure associated with skewness and kurtosis are, of course, not the only metrics by which to characterize nonnormality associated with a time series. The relation between skewness and kurtosis in samples of both atmospheric and oceanic variability [*Sura and Sardeshmukh*, 2008; *Sardeshmukh and Sura*, 2009] has been shown to have a tighter clustering along a parabola than would be expected from purely mathematical considerations [*Pearson*, 1916; *Rohatgi and Székely*, 1989]. Correlated additive and multiplicative noise can reproduce this parabolic clustering, although a number of other processes may also give rise to similar relations [*Sattin et al.*, 2009].

Tests for better distinguishing between different classes of nonnormal generating processes have seen little progress [*Sura and Hannachi*, 2015]. Autocorrelation in the data seems to be a primary obstacle. One promising technique in this regard is appropriately prewhitening. Further possibilities include directly testing for multiplicative relationships, examining the extent to which spectral or higher-order spectral features suggest the presence of aliasing, and developing approaches for handling nonstationarity associated with diurnal, seasonal, or longer time scale changes in distributional properties.

Appendix A: Influence of Discrete Frequency Space

Filters can cause the resulting signal to tend toward that of a sinusoid when the passband only admits a small number of frequencies. We derive a scaling relationship to quantify what is meant by a small number. First, consider the distribution of a sinusoid with amplitude a,

$$\frac{1}{\pi a\sqrt{1-\left(\frac{y}{a}\right)^2}}. \tag{A1}$$

The odd moments vanish on account of symmetry but excess Kurtosis is $-3/2$. The characteristic function of equation (A1) equals $J_0\,(as)$, i.e., the zero-th Bessel function of the first kind scaled by amplitude a, and the characteristic function of the distribution of a sum of n such components is the product of their individual characteristic functions, $\prod_{j=1}^{n} J_0\left(a_j s\right)$.

Assuming all a_j equal unity and uniformly distributed phases, the kth moment becomes

$$\mu_k = \mathrm{E}[Y^k] = (-i)^k \frac{\mathrm{d}^k}{\mathrm{d}s^k}\left(\prod_{j=1}^{n} J_0\,(s)\right)\Bigg|_{s=0}. \tag{A2}$$

The fourth moment is solved for by setting $k = 4$, expanding using the chain rule while making use of identities relating derivatives of Bessel functions of the first kind of different orders, and evaluating at $s = 0$. Variance is similarly computed by setting $k = 2$. Excess kurtosis is then

$$K = \frac{\mu_4}{\sigma^4} - 3 = -\frac{3}{2n}, \tag{A3}$$

which is within 5% of a normal distribution once 10 frequencies are included. There will be phase dependence between different frequencies for nonnormal distributions, and synthetic tests confirm that convergence is slower under these conditions.

Appendix B: Analytical and Numerical Filtering Results

The bispectrum, B_F, and trispectrum, T_F, of the filtered time series are

$$B_F(f_1, f_2) = H_2(f_1, f_2) \cdot B(f_1, f_2), \tag{B1a}$$

$$T_F(f_1, f_2, f_3) = H_3(f_1, f_2, f_3) \cdot T(f_1, f_2), \tag{B1b}$$

where H_2 and H_3 are defined in terms of the filter $\hat{h}(f)$ as,

$$H_2(f_1, f_2) = \hat{h}\left(f_1\right) \cdot \hat{h}\left(f_2\right) \cdot \hat{h}^*\left(f_1 + f_2\right), \tag{B2a}$$

$$H_3(f_1, f_2, f_3) = \hat{h}\left(f_1\right) \cdot \hat{h}\left(f_2\right) \cdot \hat{h}\left(f_3\right) \cdot \hat{h}^*\left(f_1 + f_2 + f_3\right). \tag{B2b}$$

Since $\hat{x}(f)$ is periodic with period $2f_N$, the top-hat filter $\hat{h}(f)$ is defined accordingly:

$$\hat{h}\left(f\right) = \begin{cases} 1 \ , f_L < (f + f_N \mathrm{mod} \ \ 2f_N) - f_N < f_H \\ 0 \ , \text{otherwise,} \end{cases} \tag{B3}$$

with f_L, f_H the low- and high-frequency cutoffs.

Figure 4 shows estimates of skewness and kurtosis obtained by numerical integration. For each $[f_L, f_H]$ pair spanning a set of 50×50 possible combinations, theoretical estimates of skewness and kurtosis of the filtered data are computed by integrating according to equation (2). B_F and T_F are computed via (B1)–(B3) on a discrete grid of 1000×1000 and $100 \times 100 \times 100$ frequency bins, respectively, and assuming $B(f_1, f_2) = 1$ and $T(f_1, f_2, f_3) = 1$. These values are compared with estimates obtained by filtering synthetic data to the same set of $[f_L, f_H]$ pairs.

Acknowledgments
We would like to thank Pedram Hassanzadeh, Tim Cronin, Mike Wallace, David Battisti, Carl Wunsch, Jerry Mitrovica, and the Harvard Extremes Reading Group for productive conversations and feedback on the project. We would also like to thank Noah Diffenbaugh, Tapio Schneider, and an anonymous reviewer. The work was supported by NSF grant AGS-1304309. A.R. acknowledges support from the James S. McDonnell Foundation. Radiosonde data were obtained from the public Integrated Global Radiosonde Archive [*Durre et al.*, 2006], and ERA-Interim data were obtained from the public database of the European Centre for Medium-Range Weather Forecasts [*Dee et al.*, 2011]. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

References

Brillinger, D. R. (1965), An introduction to polyspectra, *Ann. Math. Stat.*, *36*, 1351–1374.
Büeler, B., A. Enge, and K. Fukuda (2000), Exact volume computation for polytopes: A practical study, in *Polytopes — Combinatorics and Computation*, pp. 131–154, Springer, Basel, Switz.
Cavanaugh, N. R., and S. S. Shen (2014), Northern Hemisphere climatology and trends of statistical moments documented from GHCN-daily surface air temperature station data from 1950 to 2010, *J. Clim.*, *27*(14), 5396–5410.
Cavanaugh, N. R., and S. S. Shen (2015), The effects of gridding algorithms on the statistical moments and their trends of daily surface air temperature*, *J. Clim.*, *28*(23), 9188–9205.
Chu, C. R., M. B. Parlange, G. G. Katul, and J. D. Albertson (1996), Probability density functions of turbulent velocity and temperature in the atmospheric surface layer, *Water Resour. Res.*, *32*(6), 1681–1688.
Dee, D., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597.
Director, H., and L. Bornn (2015), Connecting point-level and gridded moments in the analysis of climate data*, *J. Clim.*, *28*(9), 3496–3510.
Donohoe, A., and D. S. Battisti (2009), The amplitude asymmetry between synoptic cyclones and anticyclones: Implications for filtering methods in feature tracking, *Mon. Weather Rev.*, *137*(11), 3874–3887.
Durre, I., R. S. Vose, and D. B. Wuertz (2006), Overview of the integrated global radiosonde archive, *J. Clim.*, *19*(1), 53–68.
Frisch, U., and D. Sornette (1997), Extreme deviations and applications, *J. Phys. I*, *7*(9), 1155–1171.
Garth, L. M., and Y. Bresler (1997), The degradation of higher order spectral detection using narrowband processing, *IEEE Trans. Signal Process.*, *45*(7), 1770–1784.
Hannachi, A. (2010), On the origin of planetary-scale extratropical winter circulation regimes, *J. Atmos. Sci.*, *67*(5), 1382–1401.
Hasselmann, K. (1976), Stochastic climate models Part I. Theory, *Tellus*, *28*(6), 473–485.
Huybers, P., K. A. McKinnon, A. Rhines, and M. Tingley (2014), U.S. daily temperatures: The meaning of extremes in the context of nonnormality, *J. Clim.*, *27*(19), 7368–7384.
Kotulski, Z., and K. Sobczyk (1981), Linear systems and normality, *J. Stat. Phys.*, *24*(2), 359–373.
Luxford, F., and T. Woollings (2012), A simple kinematic source of skewness in atmospheric flow fields, *J. Atmos. Sci.*, *69*(2), 578–590.
Mallows, C. (1967), Linear processes are nearly Gaussian, *J. Appl. Probab.*, *4*, 313–329.
Papoulis, A. (1972), Narrow-band systems and Gaussianity, *IEEE Trans. Inf. Theory*, *18*(1), 20–27.
Pearson, K. (1916), Mathematical contributions to the theory of evolution. XIX. Second supplement to a memoir on skew variation, *Philos. Trans. R. Soc. A*, *216*, 429–457.
Perron, M., and P. Sura (2013), Climatology of non-Gaussian atmospheric statistics, *J. Clim.*, *26*(3), 1063–1083.
Rennert, K. J., and J. M. Wallace (2009), Cross-frequency coupling, skewness, and blocking in the Northern Hemisphere winter circulation, *J. Clim.*, *22*(21), 5650–5666.
Rohatgi, V. K., and G. J. Székely (1989), Sharp inequalities between skewness and kurtosis, *Stat. Probab. Lett.*, *8*(4), 297–299.
Rosenblatt, M. (1961), Some comments on narrow band-pass filters, *Quart. Appl. Math*, *18*, 387–393.
Rozanov, Y. A. (1961), On the applicability of the central limit theorem to stationary processes which have passed through a linear filter, *Theory Probab. Appl.*, *6*(3), 321–322.

Ruff, T. W., and J. D. Neelin (2012), Long tails in regional surface temperature probability distributions with implications for extremes under global warming, *Geophys. Res. Lett.*, *39*, L04704, doi:10.1029/2011GL050610.

Sardeshmukh, P. D., and P. Sura (2009), Reconciling non-Gaussian climate statistics with linear dynamics, *J. Clim.*, *22*(5), 1193–1207.

Sattin, F., M. Agostini, R. Cavazzana, G. Serianni, P. Scarin, and N. Vianello (2009), About the parabolic relation existing between the skewness and the kurtosis in time series of experimental data, *Phys. Scr.*, *79*(4), 45006.

Schneider, T., T. Bischoff, and H. Płotka (2015), Physics of changes in synoptic midlatitude temperature variability, *J. Clim.*, *28*(6), 2312–2331, doi:10.1175/jcli-d-14-00632.1.

Smirnov, N. V. (1939), Estimate of deviation between empirical distribution functions in two independent samples, *Bull. Moscow Univ.*, *2*(2), 3–16.

Sura, P., and A. Hannachi (2015), Perspectives of non-Gaussianity in atmospheric synoptic and low-frequency variability, *J. Clim.*, *28*(13), 5091–5114.

Sura, P., and M. Perron (2010), Extreme events and the general circulation: Observations and stochastic model dynamics, *J. Atmos. Sci.*, *67*(9), 2785–2804.

Sura, P., and P. D. Sardeshmukh (2008), A global view of non-Gaussian SST variability, *J. Phys. Oceanogr.*, *38*(3), 639–647.

Sura, P., M. Newman, C. Penland, and P. Sardeshmukh (2005), Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes?, *J. Atmos. Sci.*, *62*(5), 1391–1409.