Identification and interpretation of nonnormality in atmospheric time series

Identification and interpretation of non-normality in atmospheric time series

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

Non-normal 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 non-normality in winter temperature from radiosonde time-series is examined through the use of linear filtering. Filtering either low or high frequencies generally suppresses what is otherwise statistically significant non-normal variability in temperature. Admitted frequencies that maximize skewness and kurtosis are, however, distinct from one another. This structure of non-normality is partly attributable to geometric relations between filtering and the appearance of skewness, kurtosis, and higher order moments in time series data. Other aspects of filtered skewness and kurtosis appear more specifically attributable to the presence of non-normal temperature variations at the highest resolved frequencies in the presence of atmospheric memory. A non-normal autoregressive model and a multiplicative noise model are both consistent with the observed frequency structure of non-normality.

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 non-normality can give insight into relevant physical processes. *Sardeshmukh and Sura* [2009] demonstrate that the appearance of non-normal characteristics of atmospheric temperature data can be explained as arising from the effects of multiplicative noise acting at the fastest resolved time scales of their model. [*Schneider et al.*, 2015], however, have demonstrated that temperatures filtered to the canonical synoptic time scales appears normal, whereas deviations from normality are apparent when including variability at time scales longer than 15 days.

These interpretations appear to conflict with one another, except that it becomes relevant to distinguish tendencies toward normality in filtered data that are indicative of physical processes from those introduced generically through the filtering of time series data. *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." Such a tendency for filtering to alter the appearance of normality has been noted for other signals [e.g., *Rozanov*, 1961; *Mallows*, 1967; *Papoulis*, 1972], to include *Donohoe and Battisti* [2009] describing how the asymmetry in the distribution of synoptic cyclones and anticyclones depends on the choice of temporal or spatial filtering techniques.

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 non-normal structure [*Kotulski and Sobczyk*, 1981; *Garth and Bresler*, 1997]. There remains the possibility, however, that distinct structure in the non-normality of filtered records can provide physical insight. In the following we examine radiosonde time series, first demonstrating statistically significant non-normality, then analytically describing how this non-normal structure can be expected to decay upon filtering, and finally describing more nuanced structures associated with non-normality using two simple numerical models.

2. Radiosonde Temperature Data

As an initial example, Fig. 1 shows the distribution of wintertime radiosonde temperature data. All radiosonde data were obtained from the Integrated Global Radiosonde Archive [*Durre et al.*, 2006] and are subset to only 12:00 GMT soundings at 850 millibars. Time periods are chosen for each station to provide a long record while minimizing the effect of missing values. Only 2% of samples are missing during the 1958–2009 interval used for the Barrow, Alaska station shown in Fig. 1), and these are infilled by linear interpolation. We remove the annual cycle and its first two harmonics with a Fourier notch filter, however seasonal results are similar when the annual cycle is retained (supplemental Fig. S1). The sample distribution of the Barrow data involves a positive skew and is found to be significantly non-normal ($p < 0.05$). 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. Controlling for autocorrelation by sampling only every five days still leads to rejection of the null of normality at $p < 0.05$, and similar results to these are obtained when using a Lilliefors test of normality.

This example is consistent with a more comprehensive study [*Perron and Sura*, 2013] that found statistically significant deviations from non-normality in daily data for nine atmospheric variables, including temperature, in reanalyses data. Surface temperature from the U.S. Global Historical Climatology Network (GHCN) has also been shown to exhibit significant non-normality [*Huybers et al.*, 2014].

The 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 gives results that are consistent with a normal distribution ($p = 0.14$). Filtering is performed using a simple top hat filter whereby Fourier components outside the pass-band 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

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

[2]Department of Atmospheric Sciences, University of Washington, Seattle , WA, USA

0094-8276/15/$5.00

appreciable differences are obtained when using more sophisticated filters, such as a forward-backward pass with a Butterworth filter.

An adjusted version of the KS test is used to assess non-normality in filtered records, accounting for the fact that band-pass filtering decreases the effective sample size of the data. In particular, the critical value for rejecting the null hypothesis of normality is computed using an effective sample size equal to the total number of retained Fourier coefficients. Although this adjustment is important for completeness, inability to reject normality is not merely because of the reduction in degrees of freedom—in this case, similar results are obtained when assuming that all samples are independent.

Note that rejection of normality can appear either from features of sub-seasonal variability or interannual variability. For instance, normal variability that is subject an interannual trend in the mean, amongst other variations, would give the appearance of non-normality (see [*Huybers et al.*, 2014] for a more complete discussion). Here we focus on only the seasonal temperature distribution, sub-selecting for the December through February winter months, and normalizing each seasonal realization to a mean of zero and variance of one.

The initial example (Fig. 1) can be generalized (Fig. 2) to show how non-normal components vary according to filtering. Specifically, we examine how skewness and excess kurtosis of four long and nearly complete records from Pittsburgh (USA), Medford (USA), Novgorod (Russia), and Sapporo (Japan) change as a function of high and low cut-off frequencies specified for filtering. A general pattern holds wherein skewness and kurtosis diminish with increasingly narrow filter bandwidths, but more nuanced structure is also present across the radiosonde time series.

All radiosonde data show skewness that is positive when the lowest frequencies of 1/90 days^{-1} are admitted but which diminishes to near zero once low-frequency cut-offs of 1/7 days^{-1} or higher are specified. Relative to the low-frequency cut-off, changing the high-frequency cut-off 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 to 1/2 days^{-1}, roughly encompassing synoptic timescales. Excess kurtosis also becomes strongly negative when the pass band becomes extremely narrow.

Significance of non-normality in the radiosonde time series is evaluated for each filter combination using the previously described adjusted KS test. With the exception of Pittsburgh, each record shows statistically significant non-normality ($p < 0.05$) 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 non-normality. In the following we explain the structure in skewness and kurtosis through the use of analytical and numerical models.

3. Analytical model of non-normality

The tendency of a time series to become more normally distributed after filtering can be understood through the use of its Fourier transform and 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 for the power spectrum, 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,

$$S = \frac{\int B(f_1, f_2) \cdot df_1 \cdot df_2}{\left(\int P(f_1) \cdot df_1\right)^{3/2}}, \tag{2a}$$

$$K = \frac{\int T(f_1, f_2, f_3) \cdot df_1 \cdot df_2 \cdot df_3}{\left(\int P(f_1) \cdot df_1\right)^{4/2}}. \tag{2b}$$

Deviations from a normal distribution are encoded in interactions between different frequency bands in the Fourier representation of time series and will be 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 Eqs. 1–2. Excluding a frequency f_i also 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 Fig. 3, and a similar depiction for the trispectrum is given in the supplementary material (Fig. S2). Although the filtering patterns are visually simple, integrals over higher order spectral volumes can become quite complicated [*Büeler et al.*, 2000].

Fig. 4 shows skewness and excess kurtosis as a function of filtered cut-off frequencies. There is excellent agreement between our analytically computed values and that obtained from a random time series containing 10^7 realizations from a Pearson distribution. The Pearson distribution is chosen because it permits for directly prescribing skewness and kurtosis, although the results are unchanged if the synthetic data is drawn from a range of other distributions such as beta, gamma or chi-squared. When the time series is 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 some of the main features of filtered skewness observed in the radiosonde data (Fig.4a). Filtering low frequencies more rapidly decreases skewness than filtering high frequencies because, as depicted in Fig. 3, the sum of low-frequency pairs occupies more area of the bispectrum not already excluded by filtering a single member of the pair than does the sum of high-frequency pairs. The analytical results also explain a lobe of high skewness that extend out towards 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 Eq. 1).

In contrast with skewness, correspondence between the kurtosis obtained from our analytical results and the radiosonde observations is poor. The analytical results show contours of kurtosis that align with contours of constant

filter bandwidth (Fig 4b), whereas the data show maximum kurtosis when filtering everything but frequencies corresponding to synoptic timescales. Negative values of excess kurtosis observed when filtering out all but a small number of frequencies (the diagonals in Fig. 2) reflects the fact that the excess kurtosis of a sine wave is -1.5. For the case of a normal process, wherein the distribution of Fourier coefficients is i.i.d., it can be shown that the expected sample excess kurtosis converges to the true process excess Kurtosis of zero as $K \propto -3/2n$, where n is the number of neighboring frequencies retained in the Fourier spectrum (see Appendix). Continuous frequency resolution in the analytical results means that they do not display this sinusoidal limit. The remainder of the mismatch between the analytical and observed kurtosis, however, points to substantive differences between the analytical model and actual temperature variability.

4. Numerical models of non-normality

An obvious deficiency in the foregoing analytical results is the presumption of independent realizations. Spectral estimates of radiosonde temperature (Figs. 1, 2) have red power spectra, with increasing energy toward lower frequencies, that can be understood as arising from integrating synoptic weather anomalies [*Hasselmann*, 1976]. This memory in a system increases the contribution of low frequency variability relative to high frequencies and can be described as an autoregressive order one process,

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

where η represents independent samples from a Pearson distribution. Specifying $\rho = 0$ would give the same kurtosis structure as discussed with respect to Fig. 4, at least up to issues of random variations associated with finite samples; setting $\rho = 0.5$ does not appreciably change this result; but setting $\rho = 0.9$ leads to a kurtosis structure with a maximum at synoptic periods, in agreement with observations (Fig. 5). The reduction in kurtosis when retaining the lowest frequencies can be understood from the effects of the central limit theorem tending to make values normal when multiple independent realizations are summed together, and the fact that these sums will tend to have greater variance and, thereby, greater control of the distribution than higher frequency fluctuations.

Finally, we consider a a discretized version of the model presented by *Sardeshmukh and Sura* [2009] that involves correlated additive and multiplicative (CAM) noise,

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

Variables η_1 and η_2 represent independent realizations of a standard normal distribution. The first term represents the integrative tendency of the process, the second and third term respectively represent the normal and non-normal innovations, and the last term ensures the process is stationary. Non-normality 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 (Fig. 5).

5. Further discussion and conclusions

Both the non-normal autoregressive model and multiplicative noise models fit the observations through inheriting non-normality at the most highly resolved frequencies. This results can be reconciled with *Schneider et al.* [2015] finding normal variability at synoptic time scales but non-normal 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 (Fig. 2). Indeed, both our analytical (Fig. 4) and numerical (Fig. 5) results indicate that skewness and kurtosis are reduced to about half their peak values when filtering to these frequencies. In contrast, filtering to $1/2 - 1/10$ days^{-1} generally results in significant non-normality (Fig. 2) for the radiosonde data, emphasizing how otherwise seemingly arbitrary choices in cut-off frequencies can substantially alter statistical results. The findings of non-normal variability for a 1/25-1/35 day^{-1} pass-band results from the discrete frequency basis only retaining a small number of non-zero frequency terms, and, thus, tending towards a distribution determined by the sinusoidal Fourier basis.

For completeness, we also verify that applying the aforementioned filters to the five ERA-Interim 850hPa grid boxes used by *Schneider et al.* [2015] leads to results consistent with our radiosonde findings. These findings indicate that higher order moments should not be neglected when considering changes in daily extremes, though they do not alter the main conclusions regarding the physical mechanisms for change in mid-latitude variability that *Schneider et al.* [2015] present.

Despite the consistency between our models and observations, it remains unclear whether the highest resolved frequency of 1/2 day^{-1} is near that of the process giving rise to the non-normal distribution. Mechanisms generating non-normality will depend on multiple location- and scale-specific factors. It will stem in part from large-scale dynamics that plausibly operate at the resolved frequencies, and in part from small-scale non-normal perturbations, such as those associated with atmospheric turbulence that display non-normality at much high frequencies [*Chu et al.*, 1996]. In future work, it would be useful to examine higher resolution temperature records with respect to their non-normal contribution.

The frequency structure associated with skewness and kurtosis are, of course, not the only metrics by which to characterize the non-normality 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]. CAM noise can reproduce this parabolic clustering, although a number of other processes may also give rise to similar relations [*Sattin et al.*, 2009].

Design of further tests for better distinguishing between different classes of non-normal generating processes seems useful. One promising technique in this regard is prewhitening of a record in order to remove integrative tendencies toward normality. A similar approach is taken by *Director and Bornn* [2015] for correcting for the effects of spatial averaging on higher order moments. 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 non-stationarity associated with diurnal, seasonal, or longer timescale 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 pass-band 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,

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

The odd moments vanish on account of symmetry but excess Kurtosis is $-3/2$. The characteristic function of Eq. 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, $\prod_{j=1}^{n} J_0(a_j s)$.

Assuming all a_j equal unity and uniformly distributed phases, the k-th moment becomes,

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

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 ten frequencies are included. There will be phase dependence between different frequencies for non-normal distributions, and synthetic tests confirm that convergence is generally weaker 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,

$$H_2(f_1,f_2) = \hat{h}(f_1)\cdot\hat{h}(f_2)\cdot\hat{h}^*(f_1+f_2), \tag{B2a}$$
$$H_3(f_1,f_2,f_3) = \hat{h}(f_1)\cdot\hat{h}(f_2)\cdot\hat{h}(f_3)\cdot\hat{h}^*(f_1+f_2+f_3). \tag{B2b}$$

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

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

with f_L, f_H the low- and high-frequency cutoffs. Assuming $B(f_1,f_2)=1$ and $T(f_1,f_2,f_3)=1$, it is possible to derive estimates for skewness and kurtosis.

For the bispectrum, we can write down analytical values for the relative skewness of the filtered data. (B3) defines seven different regions in the $[f_L,f_H]\in\{[0,f_N]\times[0,f_N]|f_L<f_H\}$ space of possible pass-bands, each with a different expression for the total area of the admitted bispectrum. The regions are depicted in Fig.4a, as are analytical estimates of the filtered skewness. The lines a-e limiting the different regions are

$$a: f_H = 2f_L \tag{B4a}$$
$$b: f_H = 2/3, f_L \le 2/3 \tag{B4b}$$
$$c: f_L = 2/3, f_H \ge 2/3 \tag{B4c}$$
$$d: f_H = 2 - f_L \tag{B4d}$$
$$e: f_H = 1 - f_L/2, \tag{B4e}$$

while relative skewness in the six non-zero regions is

$$1: \frac{(2-3f_H)^2}{4(f_H-f_L)^{3/2}} \tag{B5a}$$
$$2: \frac{-2+3f_H-3f_H^2/4+3f_L-3f_Hf_L-3f_L^2/4}{(f_H-f_L)^{3/2}} \tag{B5b}$$
$$3: \frac{1-3f_H+3f_H^2-3f_Hf_L+3f_L^2}{(f_H-f_L)^{3/2}} \tag{B5c}$$
$$4: \frac{3(f_H-2f_L)^2}{4(f_H-f_L)^{3/2}} \tag{B5d}$$
$$5: \frac{(2-3f_L)^2}{4(f_H-f_L)^{3/2}}, \tag{B5e}$$
$$6: \frac{-2+f_H(3-6f_L)+3f_L+9f_L^2/4}{(f_H-f_L)^{3/2}} \tag{B5f}$$

Panels 4b,c show estimates of skewness and kurtosis obtained by numerical integration. For each $[f_L,f_H]$ pair spanning a set of 50×50 possible combinations, the skewness and kurtosis of the filtered data are computed by integrating equations (B1). H_2 and H_3 are computed via (B2,B3) on a discrete grid of 1000×1000 and $100\times100\times100$ frequency bins respectively. The numerically integrated and analytical results agree within the expected uncertainty associated with the integration.

Acknowledgments. We would like to thank Pedram Hassanzadeh, Tim Cronin, Mike Wallace, Jerry Mitrovica and the Harvard Extremes Reading Group for productive conversations and feedback on the project. The work was supported by NSF grant number AGS-1304309. AR 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].

References

Brillinger, D. R. (1965), An introduction to polyspectra, *The Annals of mathematical statistics*, pp. 1351–1374.

Büeler, B., A. Enge, and K. Fukuda (2000), Exact volume computation for polytopes: a practical study, in *Polytopescombinatorics and computation*, pp. 131–154, Springer.

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 resources research*, *32*(6), 1681–1688.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, et al. (2011), The era-interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Director, H., and L. Bornn (2015), Connecting point-level and gridded moments in the analysis of climate data*, *Journal of Climate*, *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, *Monthly Weather Review*, *137*(11), 3874–3887.

Durre, I., R. S. Vose, and D. B. Wuertz (2006), Overview of the integrated global radiosonde archive, *Journal of Climate*, *19*(1), 53–68.

Garth, L. M., and Y. Bresler (1997), The degradation of higher order spectral detection using narrowband processing, *Signal Processing, IEEE Transactions on*, *45*(7), 1770–1784.

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, *Journal of Climate*, *27*(19), 7368–7384.

Kotulski, Z., and K. Sobczyk (1981), Linear systems and normality, *Journal of Statistical Physics*, *24*(2), 359–373.

Mallows, C. (1967), Linear processes are nearly gaussian, *Journal of Applied Probability*, pp. 313–329.

Papoulis, A. (1972), Narrow-band systems and gaussianity, *Information Theory, IEEE Transactions on*, *18*(1), 20–27.

Pearson, K. (1916), Mathematical contributions to the theory of evolution. xix. second supplement to a memoir on skew variation, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pp. 429–457.

Perron, M., and P. Sura (2013), Climatology of non-gaussian atmospheric statistics, *Journal of Climate*, *26*(3), 1063–1083.

Rohatgi, V. K., and G. J. Székely (1989), Sharp inequalities between skewness and kurtosis, *Statistics & probability letters*, *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 of Probability & Its Applications*, *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, *Geophysical Research Letters*, *39*(4).

Sardeshmukh, P. D., and P. Sura (2009), Reconciling non-gaussian climate statistics with linear dynamics, *Journal of Climate*, *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, *Physica Scripta*, *79*(4), 045,006.

Schneider, T., T. Bischoff, and H. Potka (2015), Physics of changes in synoptic midlatitude temperature variability, *J. Climate*, *28*(6), 23122331, doi:10.1175/jcli-d-14-00632.1.

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

Sura, P., and P. D. Sardeshmukh (2008), A global view of non-gaussian sst variability, *Journal of Physical Oceanography*, *38*(3), 639–647.

Corresponding author: Cristian Proistosescu, Earth and Planetary Sciences Department, Harvard University, Cambridge, MA, 02318, USA (cproist@fas.harvard.edu)

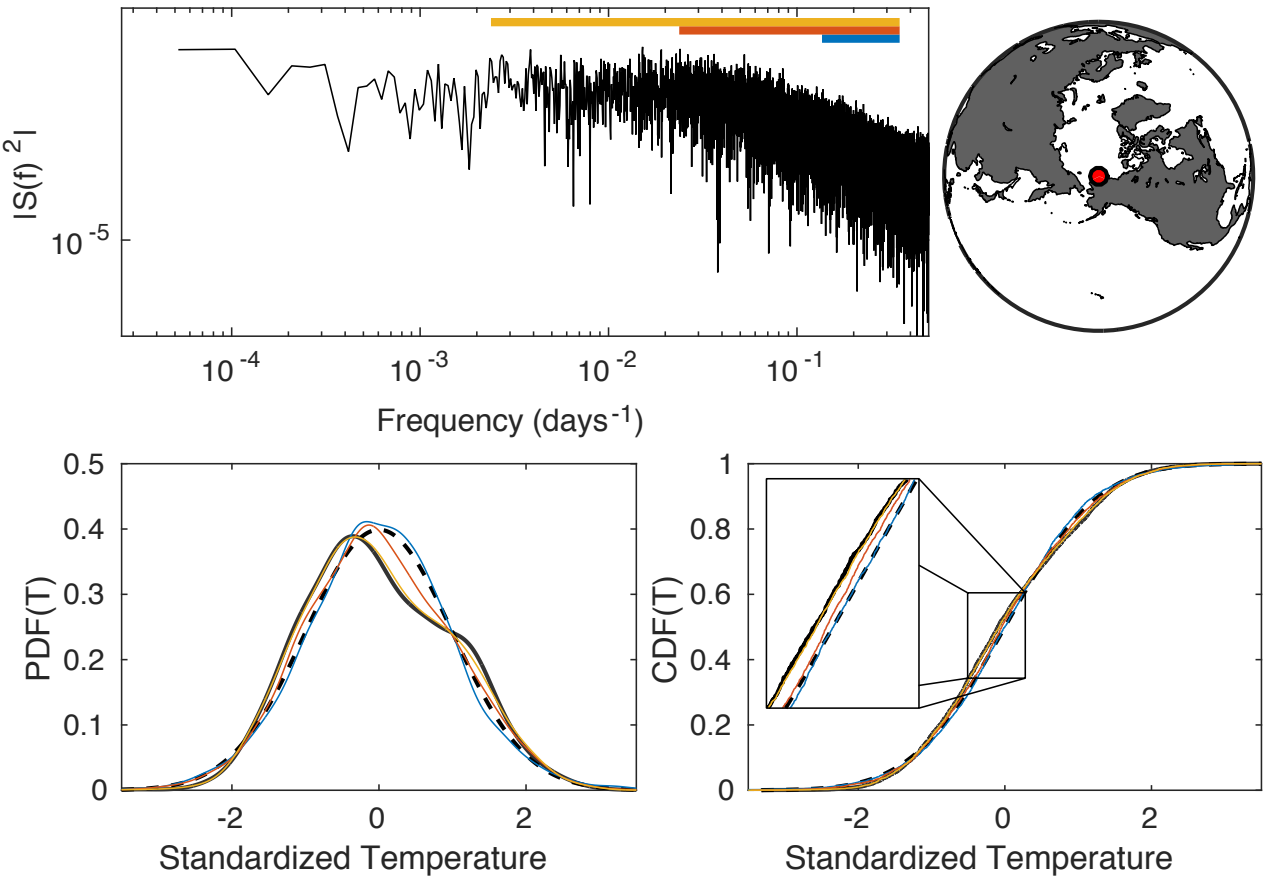

Figure 1. Filtering of radiosonde temperature time-series and normality. **Top:** Power spectrum of a radiosonde time series of temperature at 850 mbar from Barrow, AK. The annual cycle is removed by notch-filtering the annual frequency and its first two harmonics. The time-series is then progressively 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 (PDF) and cumulative density function (CDF) 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).

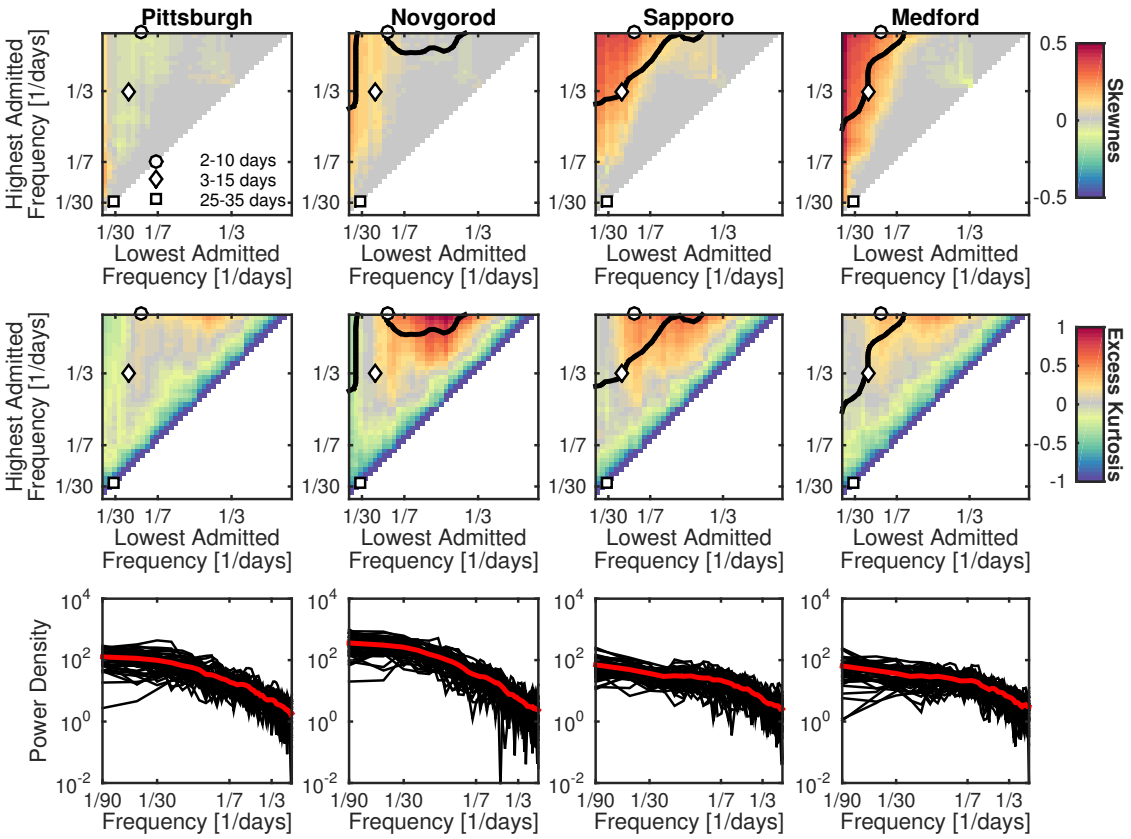

Figure 2. Effects of filtering on skewness (**Top**) and excess kurtosis (**Middle**) of 850 mbar DJF temperature from radiosonde stations for the full range of pass-bands. Axes denote pass band limits. Each season is treated as an independent ensemble member, standardized to zero mean and unit variance. Values of sample S and K averaged are over all seasons. Lines parallel to the diagonal have equal bandwidth for the pass-band, with the width of the pass-band 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, with degrees of freedom adjusted to match the number of Fourier coefficients in the filtered data. Markers denote choices of filter pass-band discussed in the text. **Bottom**: spectral estimates of each individual season (black) and the mean spectrum (red). Note the logarithmic frequency axis for spectral estimates.

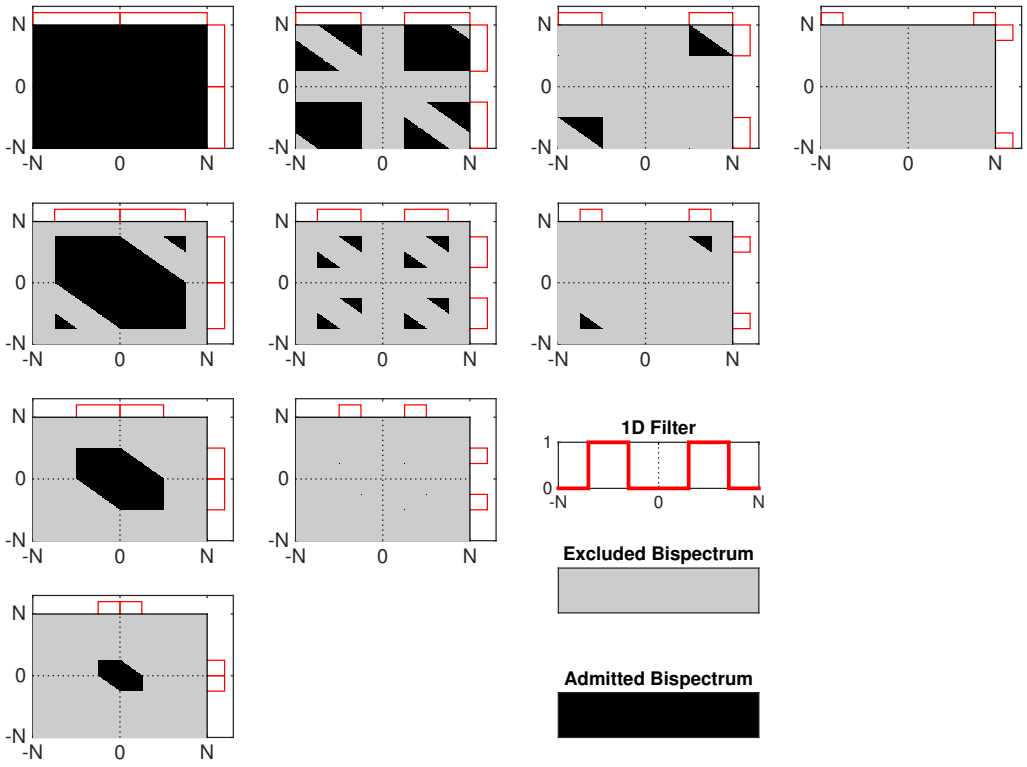

Figure 3. Effects of filtering on the bispectrum. Frequency axes are normalized relative to the Nyqvist frequency (N), and the solid color depicts the part of the bispectrum respectively that is allowed to pass under a band-pass filter. Rows from top to bottom indicate greater filtering of high frequencies, and columns from left to right are greater filtering of low frequencies. Note that both negative and positive frequencies contribute to the bispectrum.

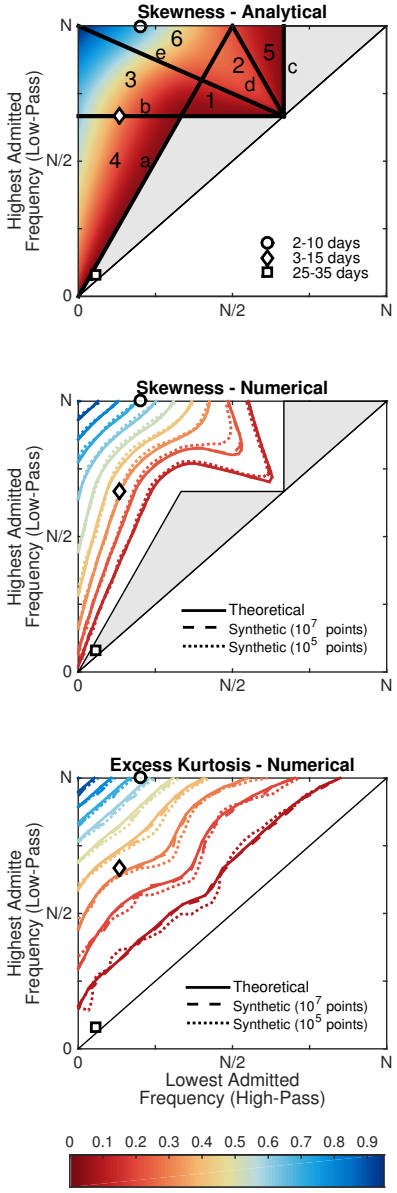

Figure 4. Skewness and excess kurtosis under band-pass filtering, relative to unfiltered values. Axes denote pass band limits, relative to the Nyqvist frequency. Gray shaded area denotes values of skewness of exactly zero. **Top**: Analytical computation of skewness from a filtered isotropic bispectrum. **Middle**: Numerical computation of a filtered isotropic bispectrum, and synthetic estimations based on realizations of an i.i.d. process. **Bottom**: Same as middle, but for excess kurtosis. Details of the analytical and numerical estimations are given in the appendix.

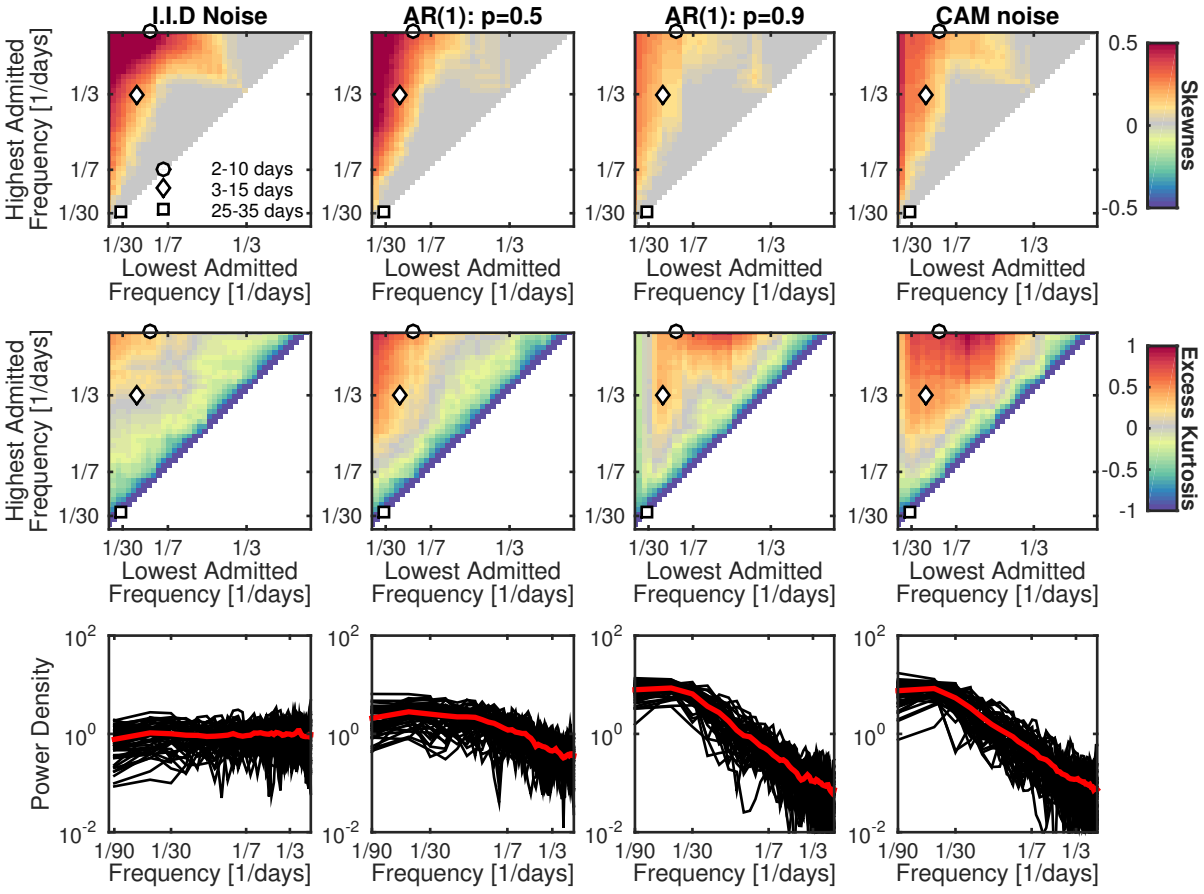

Figure 5. Same as Fig. 2, but for synthetic data. From left to right, columns present results models that are i.i.d., autoregressive order one with parameters of $\rho = 0.5$ and $\rho = 0.9$ (Eq. 3), and correlated additive and cumulative noise (Eq. 4). The i.i.d. and autoregressive models are driving by non-normal 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 panels which also shows individual realizations in black.