# Supporting Information for "Identification and interpretation of non-normality in atmospheric time series"

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

**Contents of this file**

1. Text S1 to S2
2. Table S1
3. Figures S1 to S8

**Text S1.**

**Analytical results for skewness**

For the bispectrum, we can write down analytical values for the relative skewness of the filtered data. Equation (C3) (main text) 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 passbands, each with a different expression for the total area of the admitted bispectrum. The regions are depicted in Fig.S1 as are analytical estimates of the filtered skewness. The lines a-e delimiting the different regions are

$$\begin{aligned} a: f_H &= 2f_L \\ b: f_H &= 2/3, f_L \leq 2/3 \\ c: f_L &= 2/3, f_H \geq 2/3 \\ d: f_H &= 2 - f_L \\ e: f_H &= 1 - f_L/2, \end{aligned} \tag{1}$$

while relative skewness in the six non-zero regions is

$$\begin{aligned} 1 &: \frac{(2-3f_H)^2}{4(f_H-f_L)^{3/2}} \\ 2 &: \frac{-2+3f_H-3f_H^2/4+3f_L-3f_Hf_L-3f_L^2/4}{(f_H-f_L)^{3/2}} \\ 3 &: \frac{1-3f_H+3f_H^2-3f_Hf_L+3f_L^2}{(f_H-f_L)^{3/2}} \\ 4 &: \frac{3(f_H-2f_L)^2}{4(f_H-f_L)^{3/2}} \\ 5 &: \frac{(2-3f_L)^2}{4(f_H-f_L)^{3/2}} \\ 6 &: \frac{-2+f_H(3-6f_L)+3f_L+9f_L^2/4}{(f_H-f_L)^{3/2}}, \end{aligned} \tag{2}$$

**Text S2.**

**Test for non-normality**

The KS test for non-normality employed here accounts for autocorrelation of the data, as well as the reduced number of degrees of freedom relative to the sample size in filtered data. The test is similar to the Lilliefors variation on the KS test [*Lilliefors*, 1967]. The test statistic is computed as the maximum deviation between the sample cumulative distribution function (CDF) and the CDF of a normal with identical mean and variance. A Monte-Carlo method is employed to compute the null distribution, by computing the maximum deviation between the sample CDF of phase randomized versions of the original data and the CDF of a normal distribution. The phase randomization ensures that the null model is normally distributed, while maintaining the same sample autocorrelation function and the same number of degrees of freedom as the original data. The test is consistent (5% type I errors at the nominal p=0.05 level) for the cases of autocorrelated data, subsampling for DJF, as well samples that have been normalized to unit variance and zero mean, thus avoiding the biases the standard KS test suffers from. A detailed analysis of the performance of this test is presented in figures S2 and S3.

## References

Lilliefors, H. W. (1967), On the kolmogorov-smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, *62*(318), 399–402.

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

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

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


0094-8276/16/$5.00

**Table S1.** 5th, 95th, 1st, and 99th percentiles of anomalies from the mean for a normal distribution with variance equal to the sample variance of winterthe data from those of a normal distribution with identical mean and variance

| | Medford | Barrow | Sapporo | Tateno |
|---|---|---|---|---|
| $T_5^N$ | -7.9 | -11.8 | -6.5 | -6.4 |
| $\Delta T_5$ | 0.4 | 0.8 | 0.6 | 0.8 |
| $T_{95}^N$ | 7.9 | 11.8 | 6.5 | 6.4 |
| $\Delta T_{95}$ | 0.3 | 0.3 | 0.6 | 0.8 |
| $T_1^N$ | -11.1 | -16.8 | -9.2 | -9.1 |
| $\Delta T_1$ | 1.1 | 2.3 | 1.0 | 1.9 |
| $T_{99}^N$ | 11.1 | 16.8 | 9.2 | 9.1 |
| $\Delta T_{99}$ | -0.5 | -0.5 | 0.7 | 1.5 |

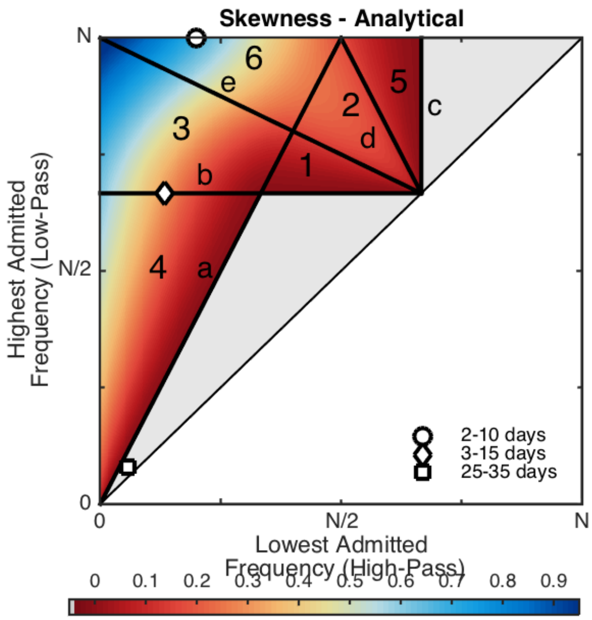


**Figure S1.** Skewness under band-pass filtering, relative to unfiltered values. Axes denote pass band limits, relative to the Nyquist frequency. Gray shaded area denotes values of skewness of exactly zero. Lines a-e delimiting different regions are given in Eqns. (1). The value of skewness in regions 1-6 are given in Eqns. (2)

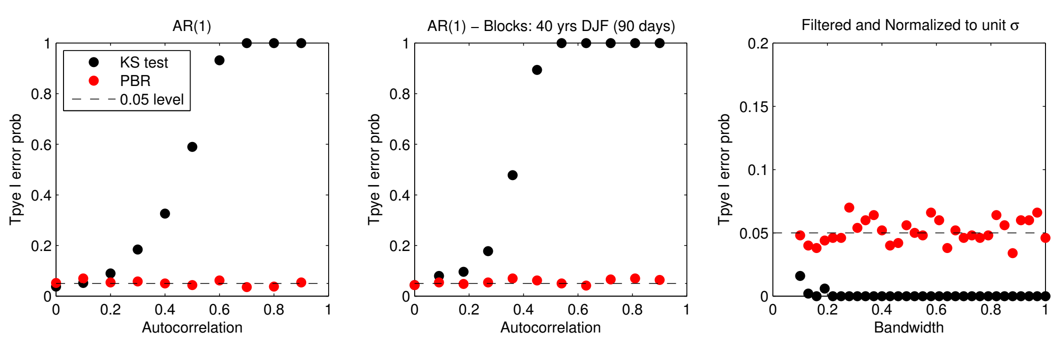


**Figure S2.** Probability of type-I errors (false positives) for the standard KS test and the KS test using phase randomization ($KS_{PR}$). At the nominal 95% critical value, a consistent test should obtain 0.05 false positives. **Left**: Normally distributed AR(1) process of length N=3500; **Middle**: Normally distributed AR(1) process of length 90*360=14,400, simulating 40 years of data. The data is then sub-sampled to 40 blocks of length 90 simulating an analysis of DJF temperature only. **Right**: I.I.D. Normally distributed data, filtered and normalized to unit variance. Probability of type-I error has been assessed using 500 iterations.

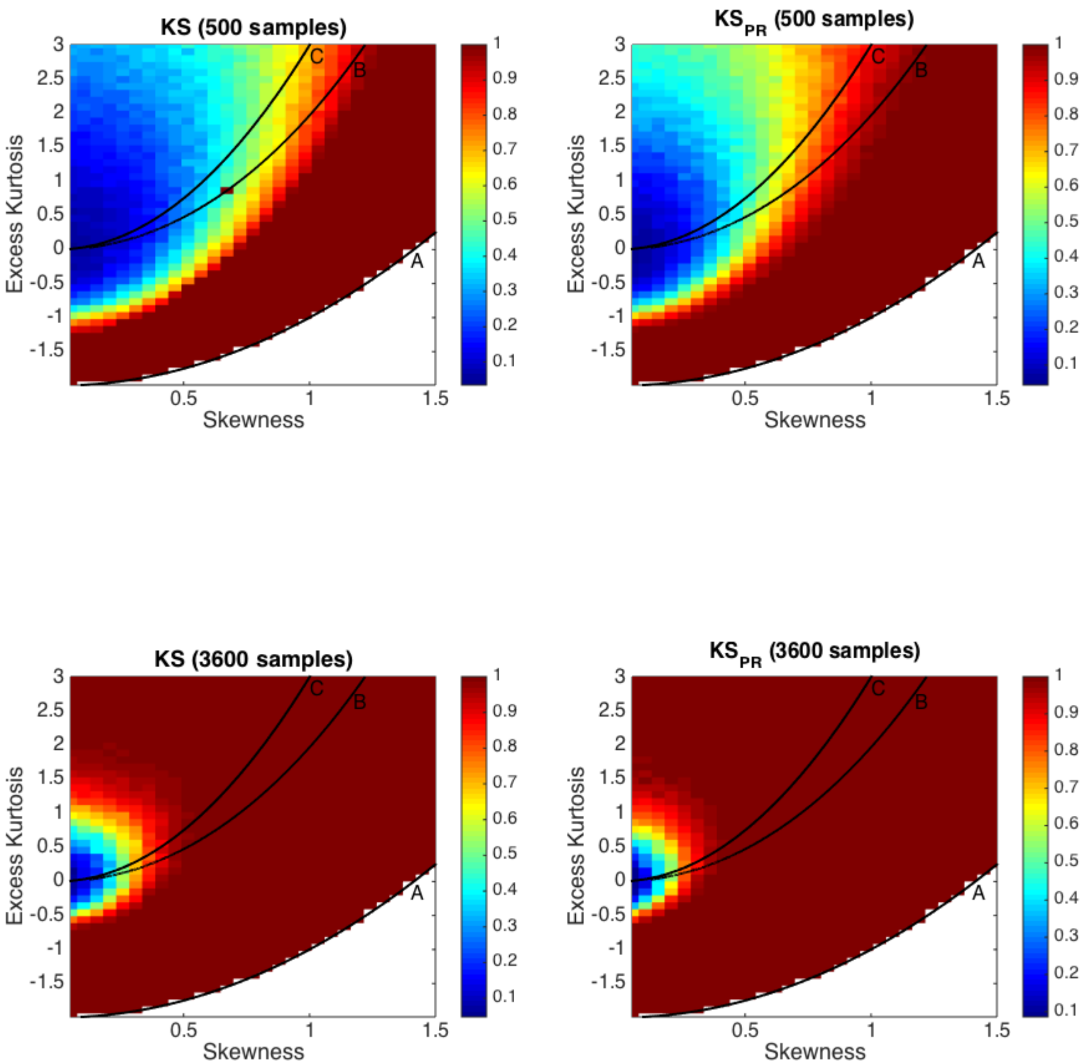


**Figure S3.** Estimates of the Statistical power of the test, defined as one minus the probability of false negatives, for a standard KS test **(left)** and the KS test using phase randomization (KS$_{PR}$) **(right)**. The data consists of 500 i.i.d. samples drawn from a Pearson'' distribution with zero mean, unit variance and varying skewness and kurtosis. The different regions depict the particular type of distributions. Curve **A** depicts the fundamental inquality of skewness and kurtosis. The region between **A** and **B** consists of Beta distributions; Curve B denotes the family of gamma distributions with unit variance (including chi-squared); Region between **B C** contains Fisher's F-distribution; Curve **C** contains inverse Chi-squared; The Region left of Curve **C** contains the Cauchy and Student's t-distribution.

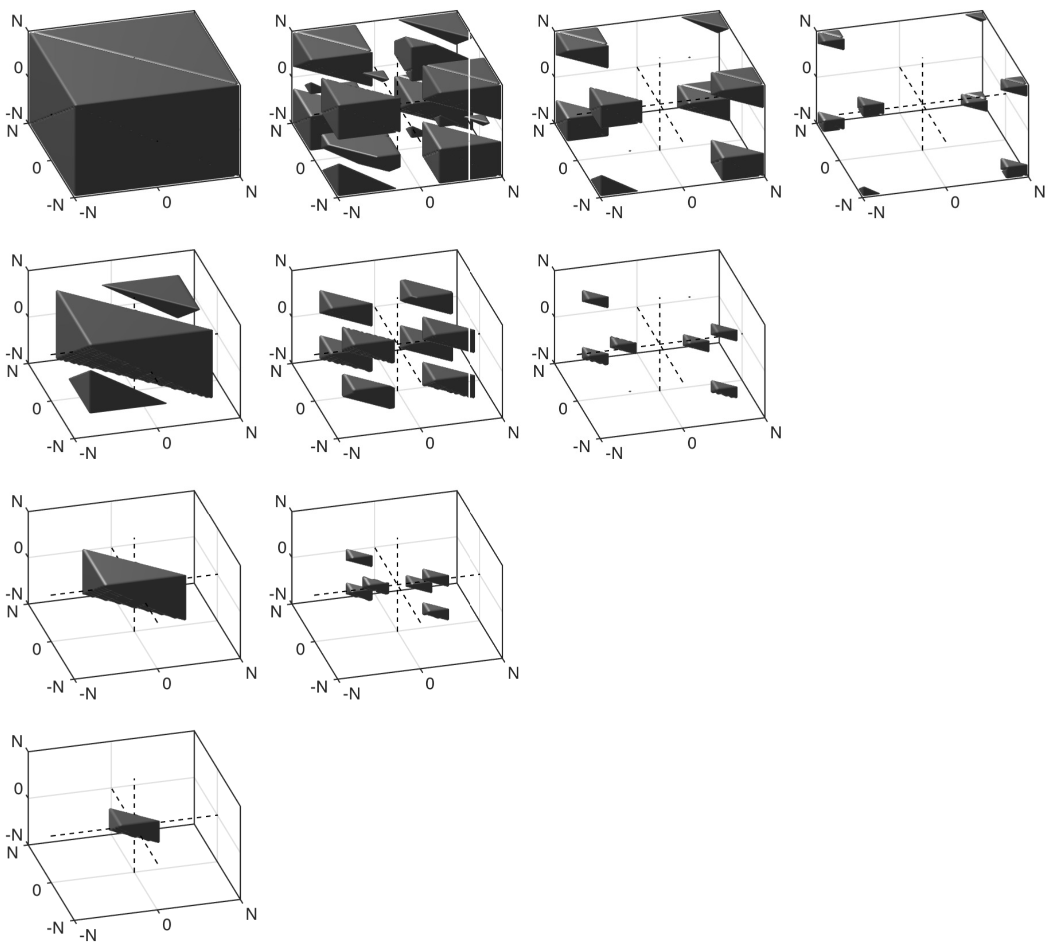

**Figure S4.** Same as Fig. 3, but depicting the effects of a top-hat band-pass filter on the trispectrum.

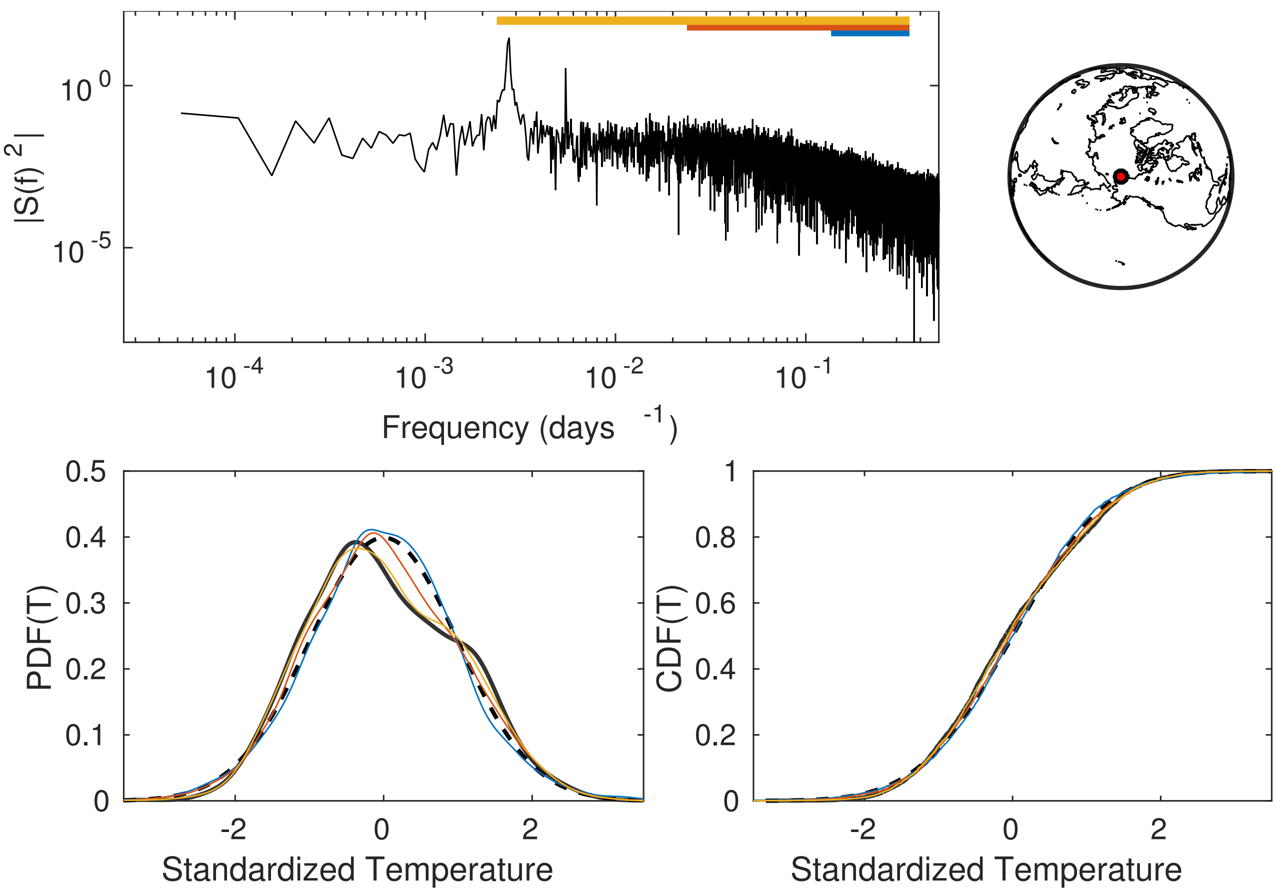


**Figure S5.** Same as Fig. 1, but without the annual cycle removed.