Need for Caution in Interpreting Extreme Weather Statistics

Prashant.D.Sardeshmukh@noaa.gov,

Gil Compo, and Cécile Penland

CIRES Climate Diagnostics Center, University of Colorado
and Physical Sciences Division/ESRL/NOAA

Main Point

The PDFs of daily atmospheric anomalies are not Gaussian.

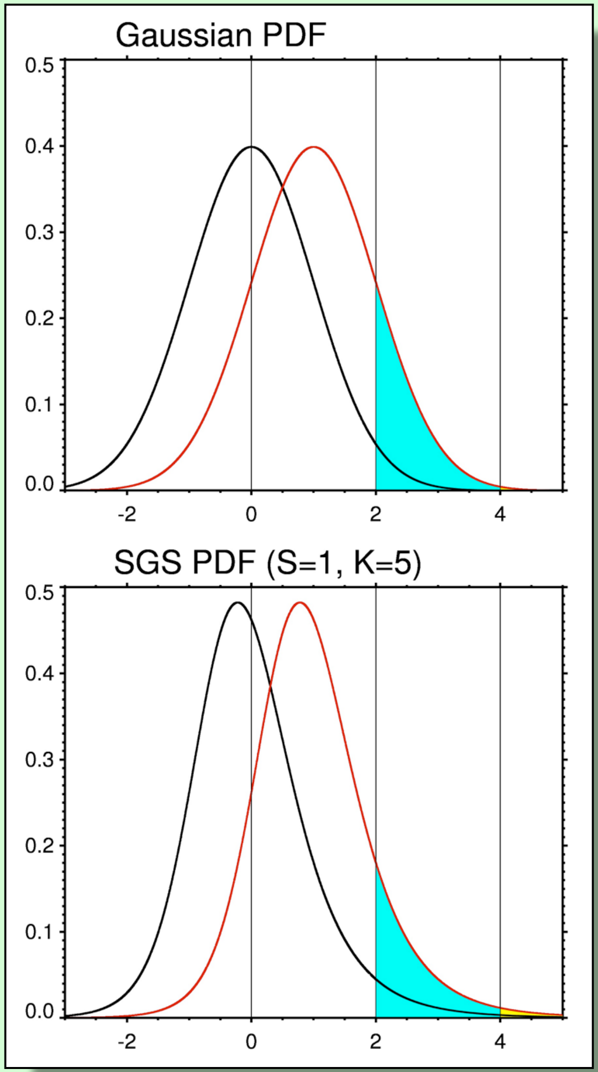
They are generally skewed and heavy tailed.

This has large implications for the statistics of extreme weather.

Non-Gaussianity has enormous implications for the probabilities of extreme values, and for our ability to estimate their changes using limited samples

Consider Gaussian vs non-Gaussian PDFs, ***both p(0,1)****, and shifted by 1 sigma*

Gaussian PDFs



P(x ≥ 2) = 2.3%
and increases by
a factor of 7

P(x ≥ 4) = 0.003%
and increases by
a factor of 43

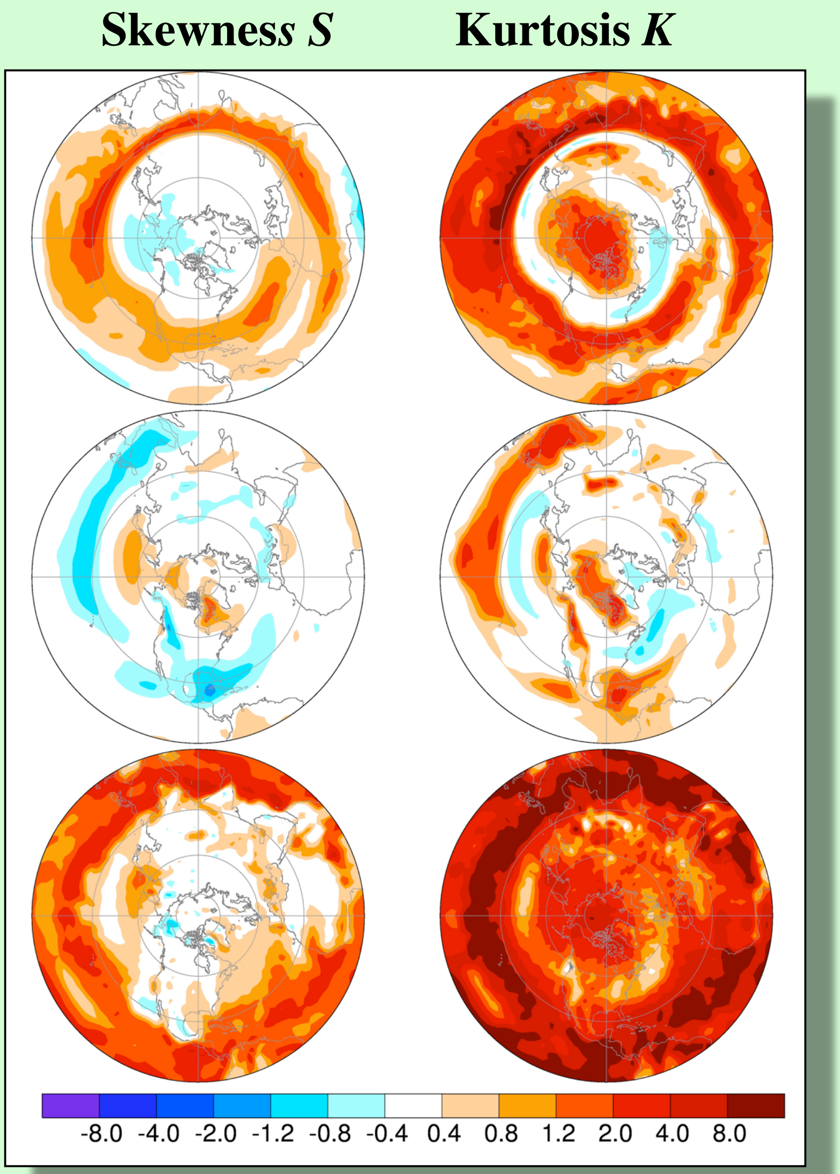
Non-Gaussian PDFs

skewed and heavy-tailed with
Skewness $S = 1$
Kurtosis $K = 5$

P(x ≥ 2) = 3.4%
and increases by
only a factor of 4

P(x ≥ 4) = 0.34 %
and increases by
only factor of 3

Skewness $S = \langle x^3 \rangle / \sigma^3$ and Kurtosis $K = \langle x^4 \rangle / \sigma^4 - 3$ of daily anomalies in winter computed over 137 winters (1871-2007) in the 20CR dataset (Compo et al 2011)



250 mb
Vorticity

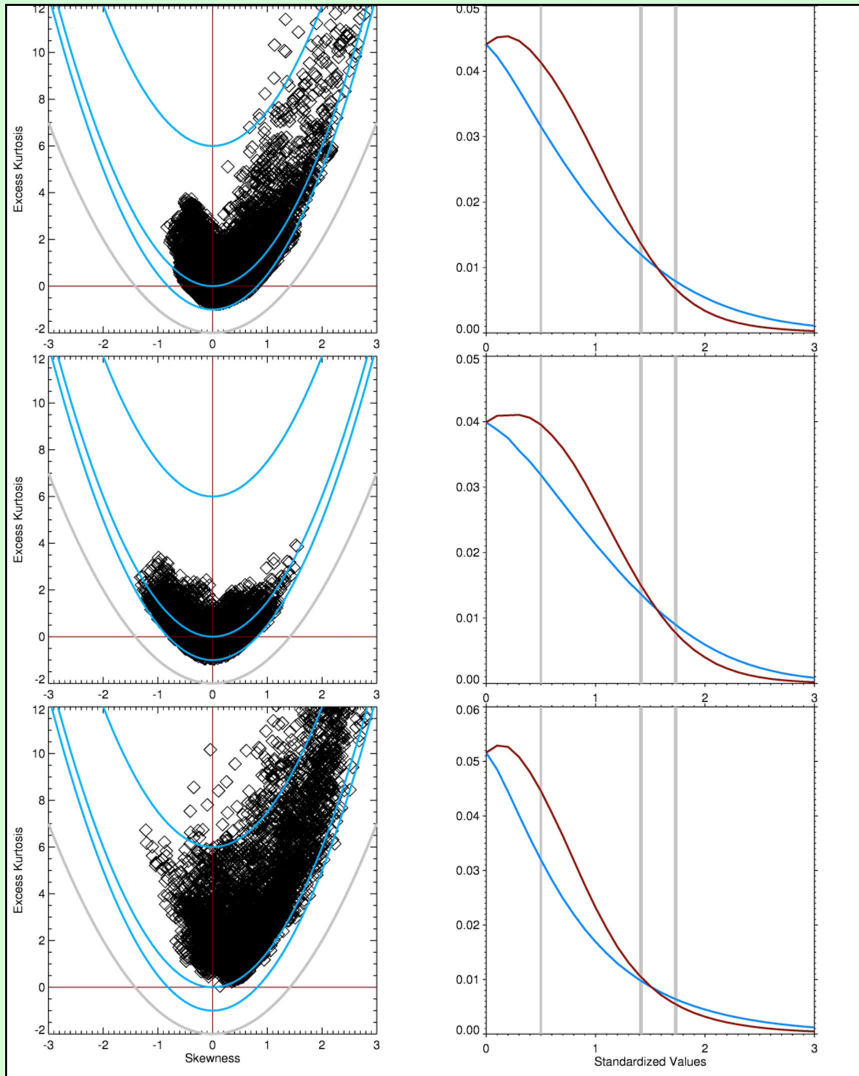
850 mb
Air temperature

500 mb
Vertical Velocity

Some distinctive features of the non-Gaussianity of standardized daily anomalies at all N.H. grid points

computed using 137 winters (1871-2007) of 20CR data

K* vs *S **Average Histograms**



250 mb Vorticity

850 mb Air temperature

500 mb Vertical Velocity

Note the parabolic inequality $K \geq 3/2\, S^2$

Note that the crossover point where p(x) = p(-x) lies between 1.4σ and 1.7σ

A generic "Stochastically Generated Skewed" (SGS) probability density function (PDF) suitable for describing non-Gaussian climate variability *(Sardeshmukh and Sura J. Clim 2009)*

$$p(x) \;=\; \frac{1}{\mathcal{N}}\left[(Ex+g)^2+b^2\right]^{-\left(1+\frac{\lambda}{E^2}\right)} \exp\left[-\,\frac{2\lambda g}{E^2 b}\;\arctan\left(\frac{Ex+g}{b}\right)\right]$$

If $E \to 0$, then $p(x) \;\to\;$ a Gaussian PDF

$\lambda > 0 \quad b \geq 0$

$g \geq 0$ or $g < 0$

$E \geq 0$

Such a PDF has power-law tails, its moments satisfy $K \geq (3/2)\, S^2$, and $p(x) = p(-x)$ at $\hat{x} \approx \sqrt{3}\,\sigma$

This PDF arises naturally as the PDF of the simplest 1-D damped linear Markov process that is perturbed by Correlated Additive and Multiplicative white noise ("CAM noise")

$$\frac{dx}{dt} \;=\; -\left(\lambda+\frac{1}{2}E^2\right)x \;+\; b\,\eta_1 \;+\; (Ex+g)\,\eta_2 \;-\; \frac{1}{2}Eg$$

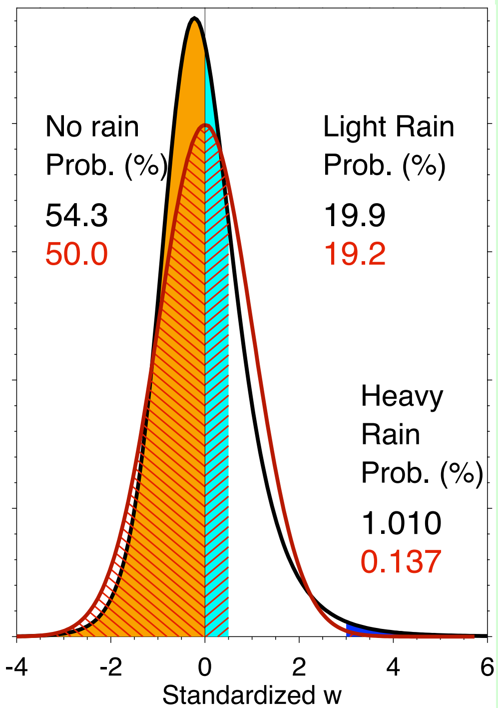
If $E \to 0$, this is just the evolution equation for Gaussian "red noise"

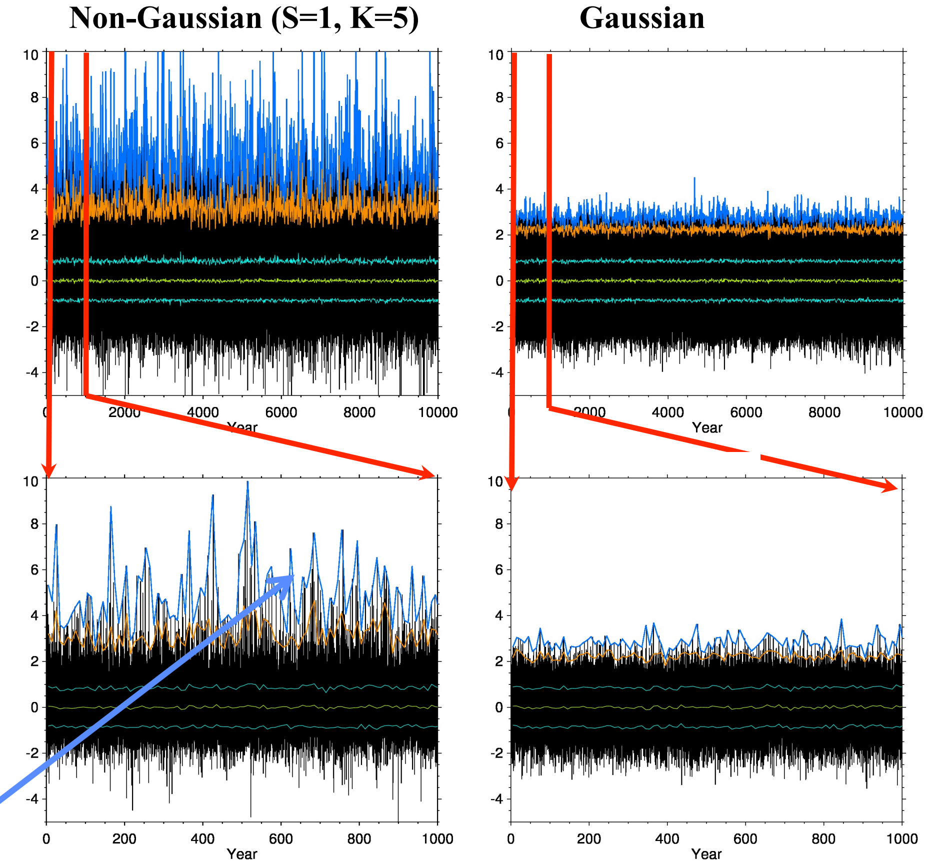
η_1 and η_2 are Gaussian white noises of unit amplitude.

The parameters of this model (and of the PDF) can be estimated using the first four moments of x and its correlation scale. **The model can then be run to generate Monte Carlo estimates of extreme statistics**

Sharply contrasting behavior of extreme *w* anomalies (and by implication, of extreme precipitation anomalies) obtained in 10^8-day runs (equivalent to 10^6 100-day winters) of the Gaussian and non-Gaussian models

Gaussian (red) and non-Gaussian (black, S=1, K=5) PDFs with same mean and variance





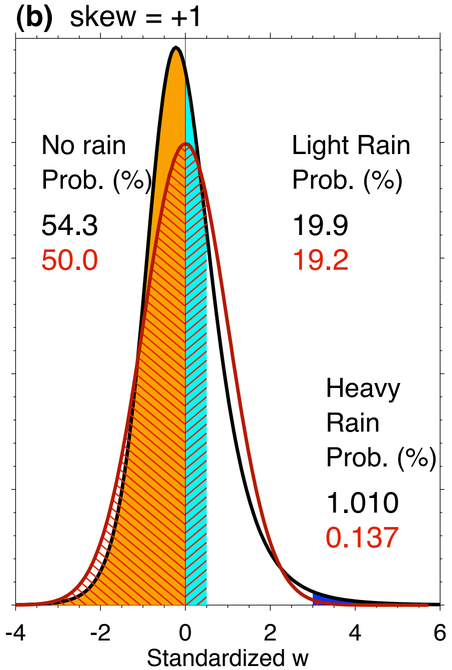
Note how one can obtain spurious 100-yr trends of decadal extremes in the non-Gaussian case *even in this statistically stationary world*.

Blue curves: Time series of decadal maxima
(i.e the largest daily anomaly in each decade = 1000 days
= 10 100-day winters)

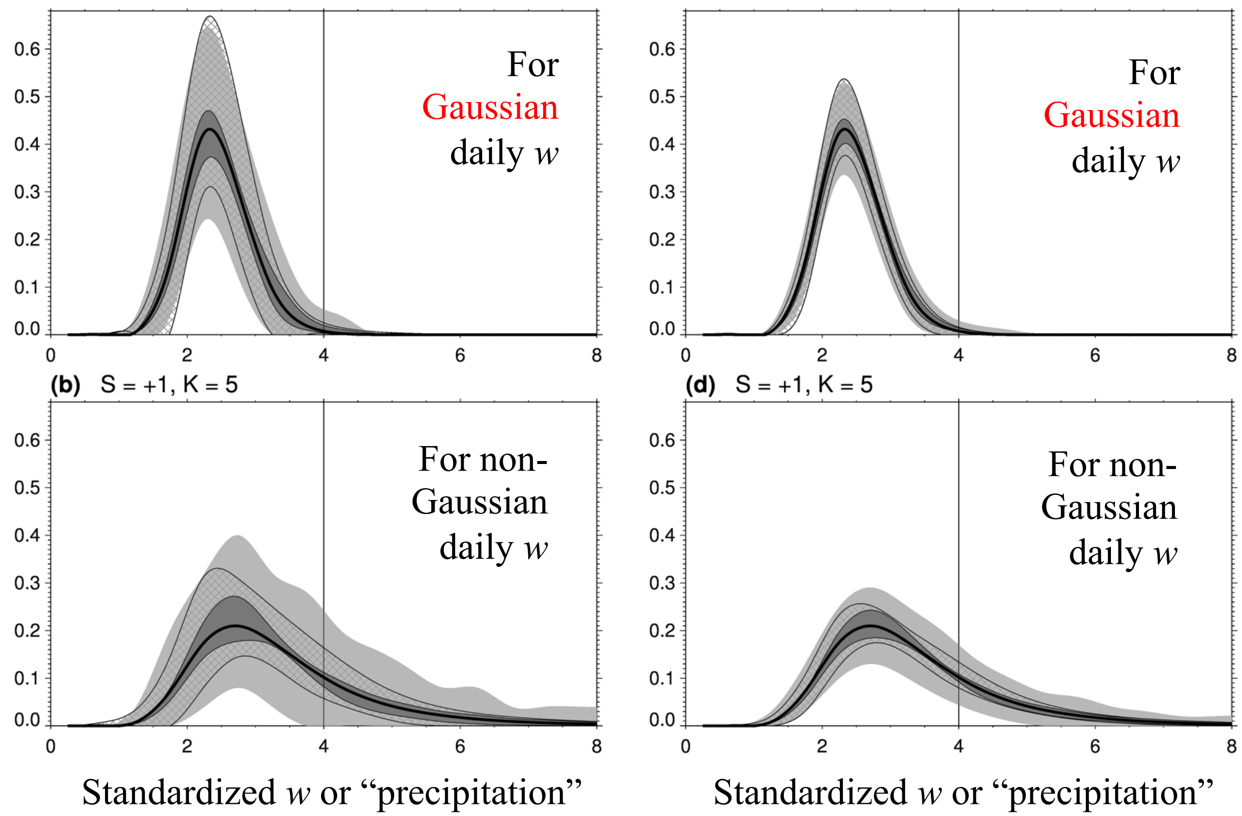
Orange curves: Time series of 99.5th decadal percentile
(i.e. the 5th largest daily anomaly in each decade)

The PDFs of winter maxima are VERY DIFFERENT if the PDFs of the daily values are Gaussian or “SGS”. They are also more accurately estimated by fitting SGS distributions to *all* daily values than by fitting GEV distributions to just maximum values

PDFs of daily *w* :
Gaussian
and non-Gaussian
(S=1, K=5)



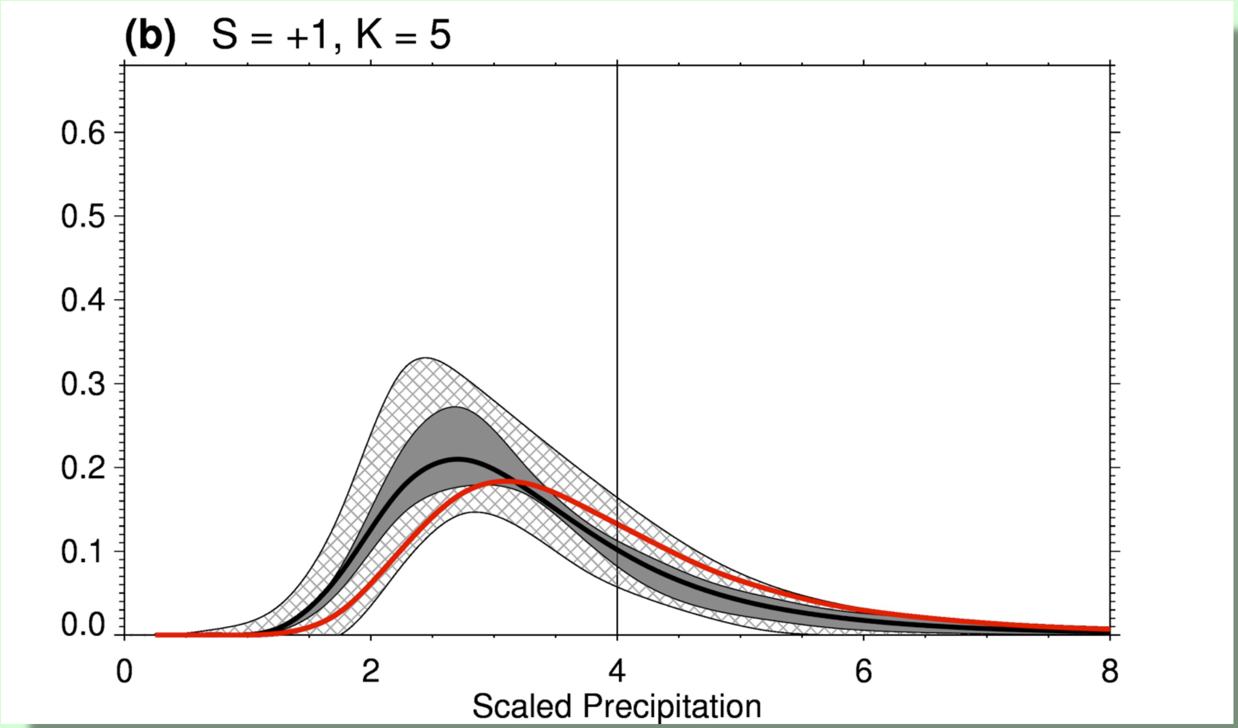
PDFs of daily winter *maxima* (Extreme Value PDFs) estimated using
25-winter records **100-winter records**



Black Curves: Extreme Value PDFs of winter *w* (or “precipitation”) maxima estimated from 10^6 model winters, when the PDF of daily *w* is Gaussian or non-Gaussian

Outer grey bands: 95% intervals of raw histogram-based estimates using the 25 or 100 winter maxima
Inner grey bands: 95% intervals of GEV PDFs fitted to the 25 or 100 winter maxima
Darkest grey bands: 95% intervals of Extreme Value PDFs derived from SGS distributions fitted to all daily values in the 25 or 100 winters

The sampling uncertainties of GEV-based Extreme Value PDFs estimated from 25-yr records are larger than the modest changes expected in the true PDF from , say, a 15% increase in precipitation (red curve). The corresponding uncertainties from an SGS-theory based estimation are much smaller.



Models differ in their representation of observed non-Gaussian behavior, and do not "converge" simply by increasing resolution.

Below are some results from ECMWF (Wedi et al, 2010)

Higher resolution influence (250hPa vorticity)

Wedi et al 2010

Variance:

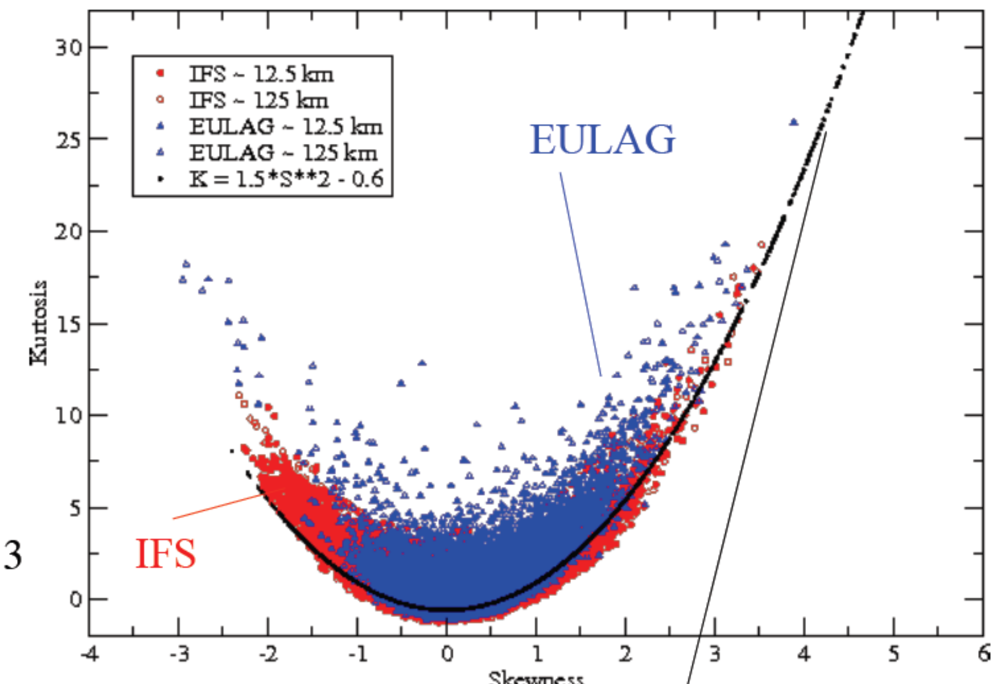
$$\sigma^2 = \overline{(x-\overline{x})^2}$$

Skewness:

$$S = \frac{\overline{(x-\overline{x})^3}}{\sigma^3}$$

Kurtosis:

$$K = \frac{\overline{(x-\overline{x})^4}}{\sigma^4} - 3$$



Predicted from linear stochastic models forced with Correlated Additive and Multiplicative ("CAM") noise (Sardeshmukh and Sura, 2009)

Summary

1. **The PDFs of daily anomalies are significantly skewed and heavy-tailed. This fact has enormous implications both for the probabilities of extremes and for estimating changes in those probabilities.**

2. We have demonstrated the relevance of "stochastically generated skewed" (SGS) distributions for describing daily atmospheric variability, that arise from simple extensions of a "red noise" process.

3. The parameters of these SGS distributions, and of the associated linear Markov model, can be estimated from the first four moments of the data (mean, variance, skewness, and kurtosis). The model can then be run to generate not only the appropriate SGS distribution, but also to estimate sampling uncertainties through extensive Monte Carlo integrations.

4. We have shown that extreme-value distributions can be estimated more accurately from limited-length records using such a Markov model than through direct GEV approaches.

5. To accurately represent extreme weather statistics and their changes, it is necessary for climate models to accurately represent the first four moments of daily variability. The good news is that for many purposes this may also be sufficient. The bad news is that currently they do not adequately capture the changes of even the first moment (the mean) on regional scales.