A web-based tool to examine probability distributions of weather and climate data

Catherine A. Smith, Gilbert P. Compo, and Prashant D. Sardeshmukh

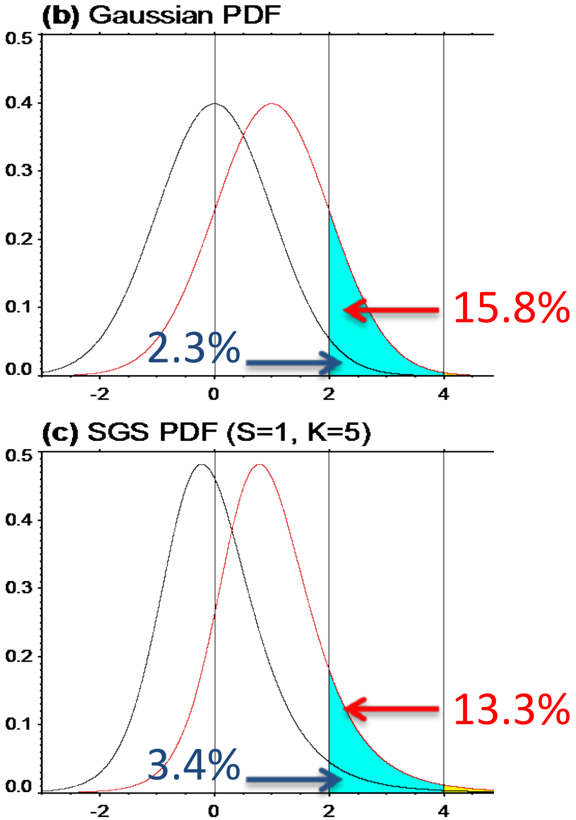
CIRES/U. of Colorado &

NOAA/ESRL PSD, Boulder CO

http://www.esrl.noaa.gov/psd/data/writ/distributions/

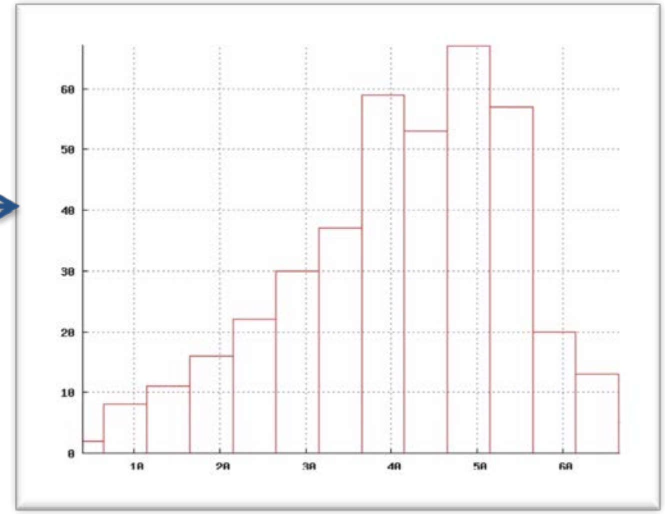
Importance of distributions

- Climate change
- Extreme Events
- El Niño/ La Niña…
- Model Analysis
- Cutoffs (HDD, Freezes, Speed)
- N/S or E/W switches.



Determining "True" Distributions

Create a histogram from the data

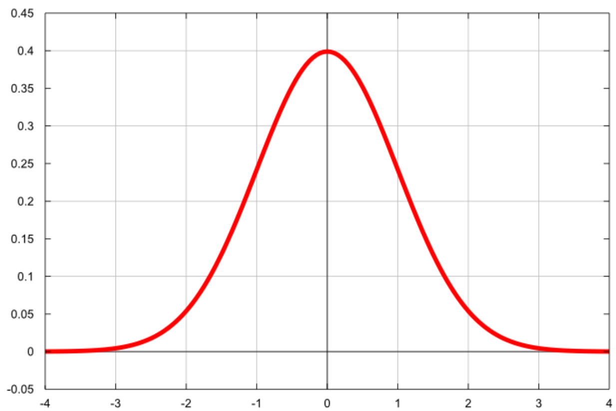


Problems: Sample Size and Sample Variability

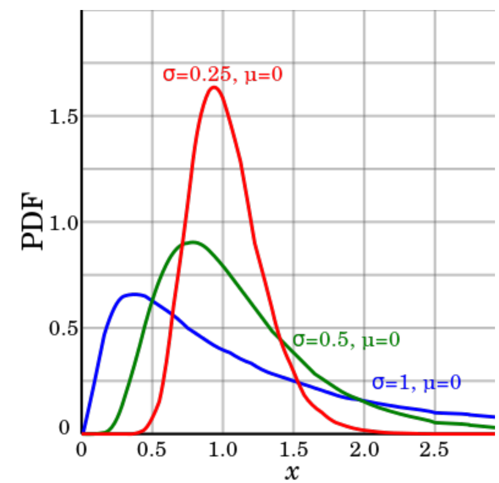
- Some possible extreme value never happened before!
- Error bars are large for probability calculations, especially at the tails.

Fit a Distribution

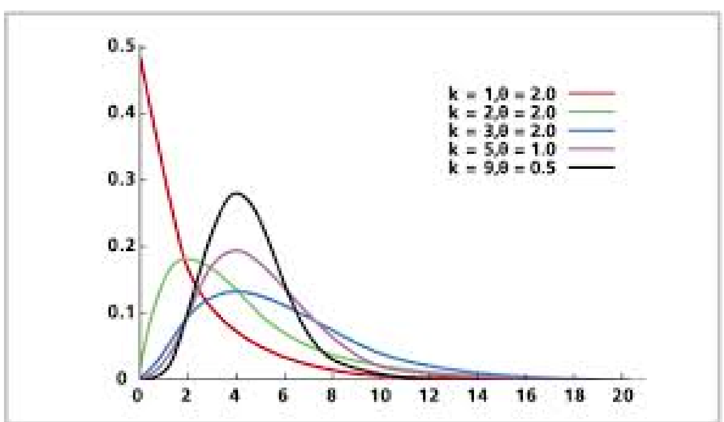
Gaussian Distribution



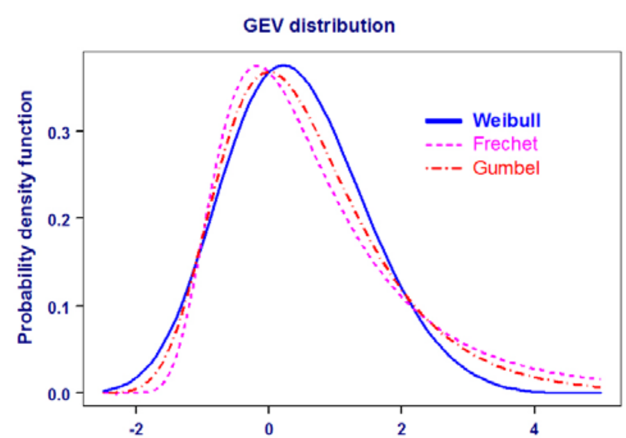
Log-Normal Distribution



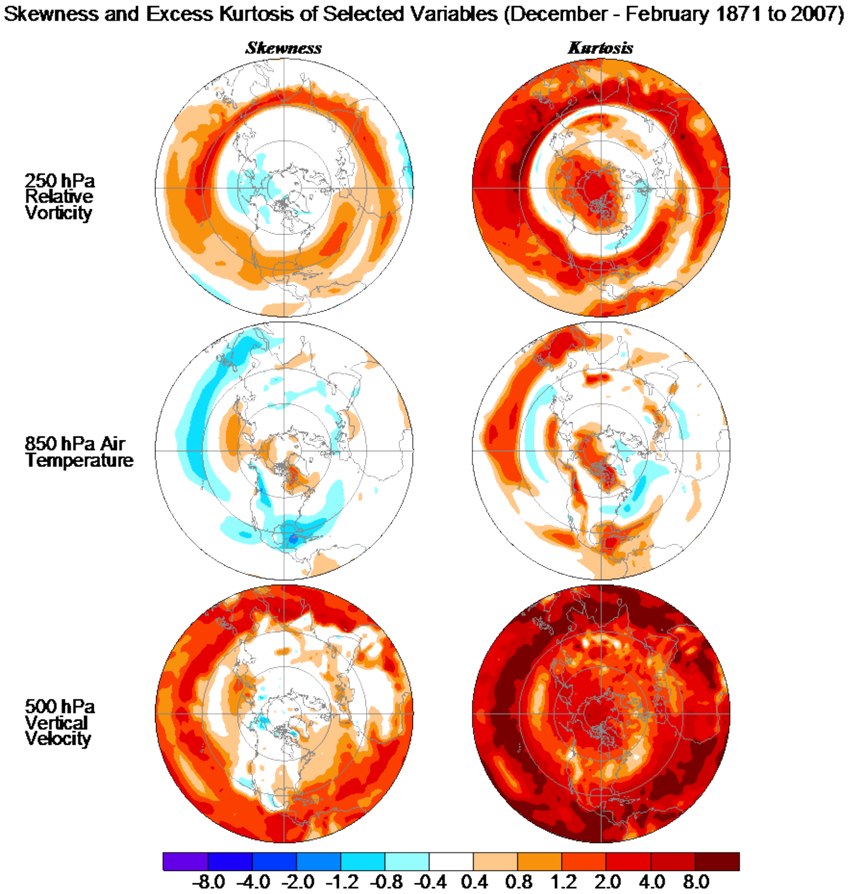
Gamma Distribution



Generalized extreme value



Are Daily Atmospheric Variables Gaussian?



Stochastically Generated Skewed Distribution ("SGS")

"Math Page"

Sardeshmukh, P.D., G.P.Compo, and C. Penland, 2015: **Need for caution in interpreting extreme weather statisti**cs. *J. Climate*, in press,

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

$$\mathcal{N}\ (E,g,b) = \frac{2\pi\ \nu^{1/2} (2b)^{-(2\nu+1)}\ \Gamma(2\nu+1)}{\Gamma(\nu+1-iq/2)\ \Gamma(\nu+1+iq/2)}$$

Derived from a stochastically perturbed damped linear Markov process x(t) whose change dx in a small time interval dt is given by :

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

Stochastically Generated Skewed Distribution ("SGS")

- **The noise is related to the base state so a skewed or kurtosed distribution is possible.**
- **It is a first order correction to a Gaussian.**
- **As E→0, the SGS distribution is a Gaussian.**

Use S(Skewness) and K(Kurtosis) to calculate E,g,b →

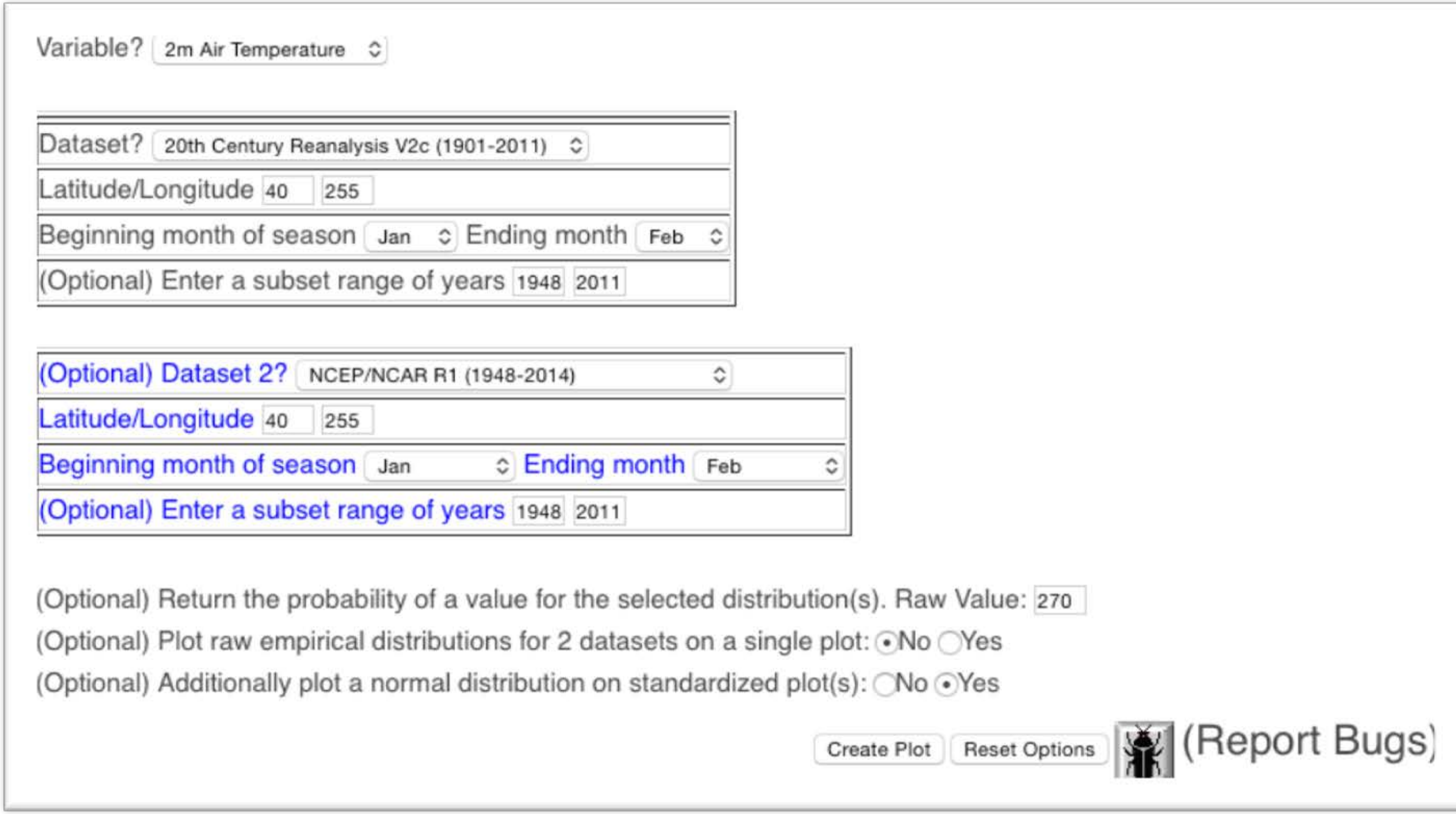
$$S = \frac{2E}{\left(1-E^2\right)}\frac{g}{\sigma} \quad \text{, and}$$

$$K = \frac{3}{2}\left[\frac{1-E^2}{1-(3/2)E^2}\right]S^2 + \left[\frac{3E^2}{1-(3/2)E^2}\right] > \frac{3}{2}S^2 ,$$

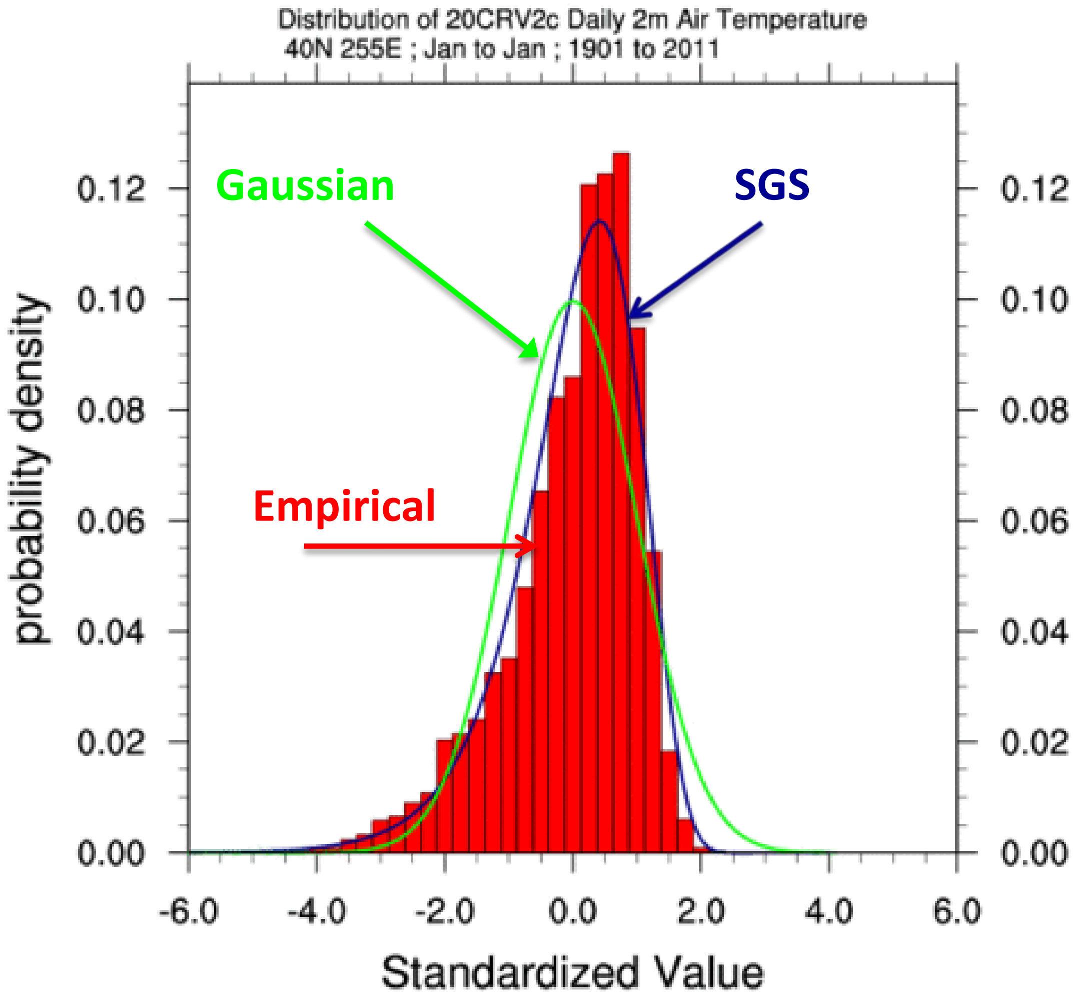
New data distribution web-tool for daily data

- Plot empirical and fitted distributions of daily atmospheric data
- Fit data to different distributions
- Return requested probabilities of a value
- Return statistics
- Compare distributions
- Return all associated data and statistics in netCDF for users to experiment with

Web Interface



Distribution of Boulder, CO January 2m Air Temperature from 20CRV2c: 1901-2011



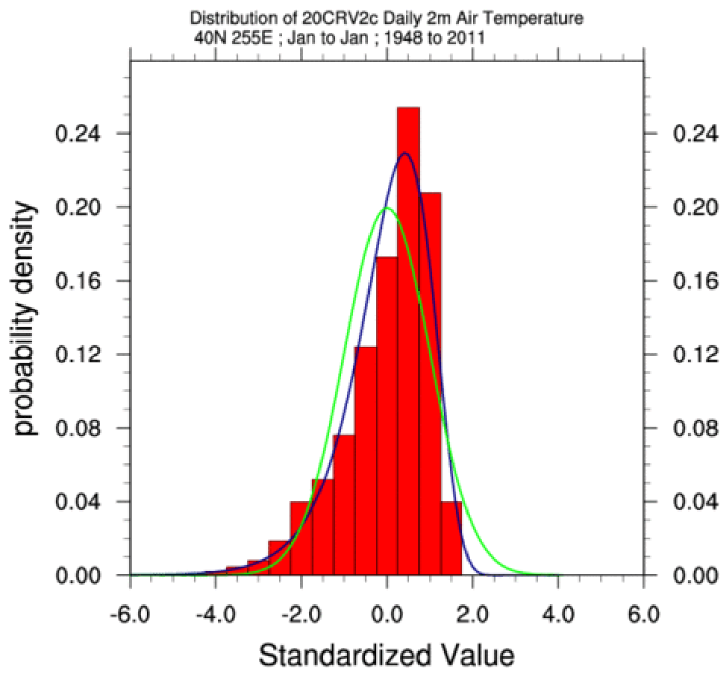
Statistics of Boulder, CO
January 2m Air Temperature: 1901-2011

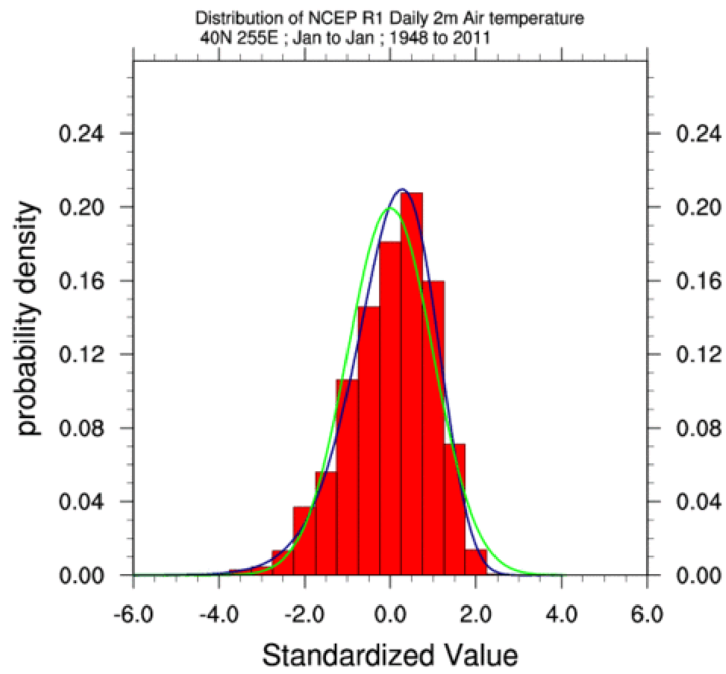
Time Series Statistics: 20th Century Reanalysis V2c 2m Air Temperature

Statistic	20CRV2c
N	3441
Mean	269.47
STD	4.44
Skewness	-1.053
Kurtosis	1.137
Minimum	247.48
Lower Decile	263.09
Median	270.47
Upper Decile	274.16
Maximum	279.64

SGS Statistic	20CRV2c
E	0.342
g	-1.361
b	0.175

Distributions Boulder, CO 20CRV2c vs. NCEP R1 January 2m Air Temperature: 1948-2011





Statistics of 20CRV2c vs. NCEP R1
January 2m Air Temperature: 1948-2011

alue is 253: Standardized value is -3.530 for Distribution 1

.003047 : Empirical Probability of <= 253 for Distribution 1*

.005402 : SGS Probability of <= 253 for Distribution 1

.000192 : Gaussian Probability of <= 253 for Distribution 1

alue is 253: Standardized value is -2.681 for Distribution 2

.009577 : Empirical Probability of <= 253 for Distribution 2

.012273 : SGS Probability of <= 253 for Distribution 2

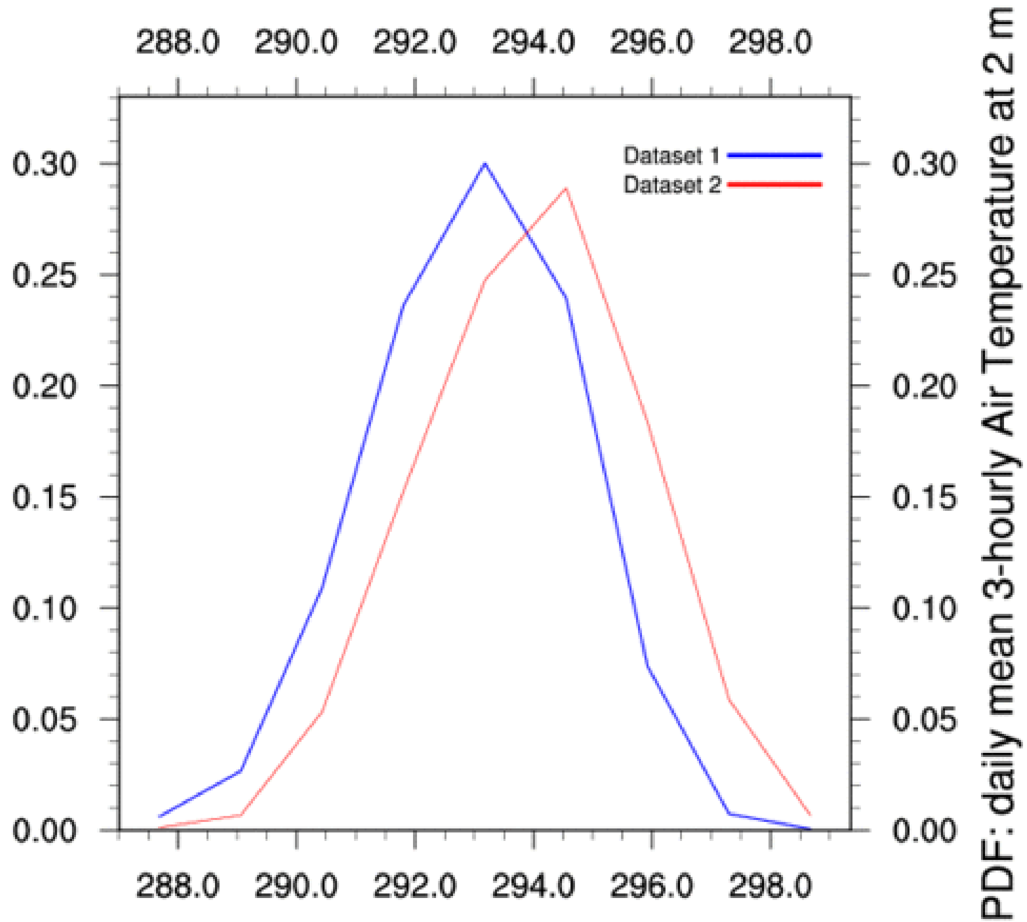
.004011 : Gaussian Probability of <= 253 for Distribution 2*

Empirical probability, is ranked value, scaled. If not an exact match

ime Series Statistics: 2m Air Temperature Jan to Jan 40;2

Statistic	20th Century Reanalysis V2c	NCEP/NCAR R1
N	1984	1984
Mean	269.31	266.48
STD	4.62	5.03
Skewness	-1.074	-0.599
Kurtosis	1.028	0.303
Minimum	247.48	244.40
Lower Decile	262.45	259.61
Median	270.48	267.13
Upper Decile	274.07	272.36
Maximum	277.27	278.85
E	0.347	0.210
g	-1.362	-1.364
b	0.160	0.308

Distributions of Early and Late 20th Century NYC, NY Daily Summer Air Temperatures from 20CRV2c



Statistics of Early vs. Late 20th Century NYC, NY Daily Summer Air Temperatures from 20CRV2c

0.970844 : Empirical Probability of <= 296 for Distribution 1
0.968458 : SGS Probability of <= 296 for Distribution 1
0.960004 : Gaussian Probability of <= 296 for Distribution 1
0.863524 : Empirical Probability of <= 296 for Distribution 2
0.877500 : SGS Probability of <= 296 for Distribution 2
0.875027 : Gaussian Probability of <= 296 for Distribution 2

Time Series Statistics: 2m Air Temperature Jul to Aug 40.7;286.1

Statistic	1901-1926	1986-2011
N	1612	1612
Mean	292.97	293.98
STD	1.71	1.80
Skewness	-0.263	-0.165
Kurtosis	-0.094	-0.149
Minimum	287.00	287.17
Lower Decile	290.67	291.55
Median	293.12	294.04
Upper Decile	295.04	296.29
Maximum	298.16	299.36
E	0.102	0.081
g	-1.282	-1.008
b	0.589	0.989

Coming Attractions

Soon:

- Subsets of Years (e.g El Niño/La Niña)
- KS measure of similarity of distributions
- More datasets: ERA-Interim, US station data
- Different output plot types
- More Variables

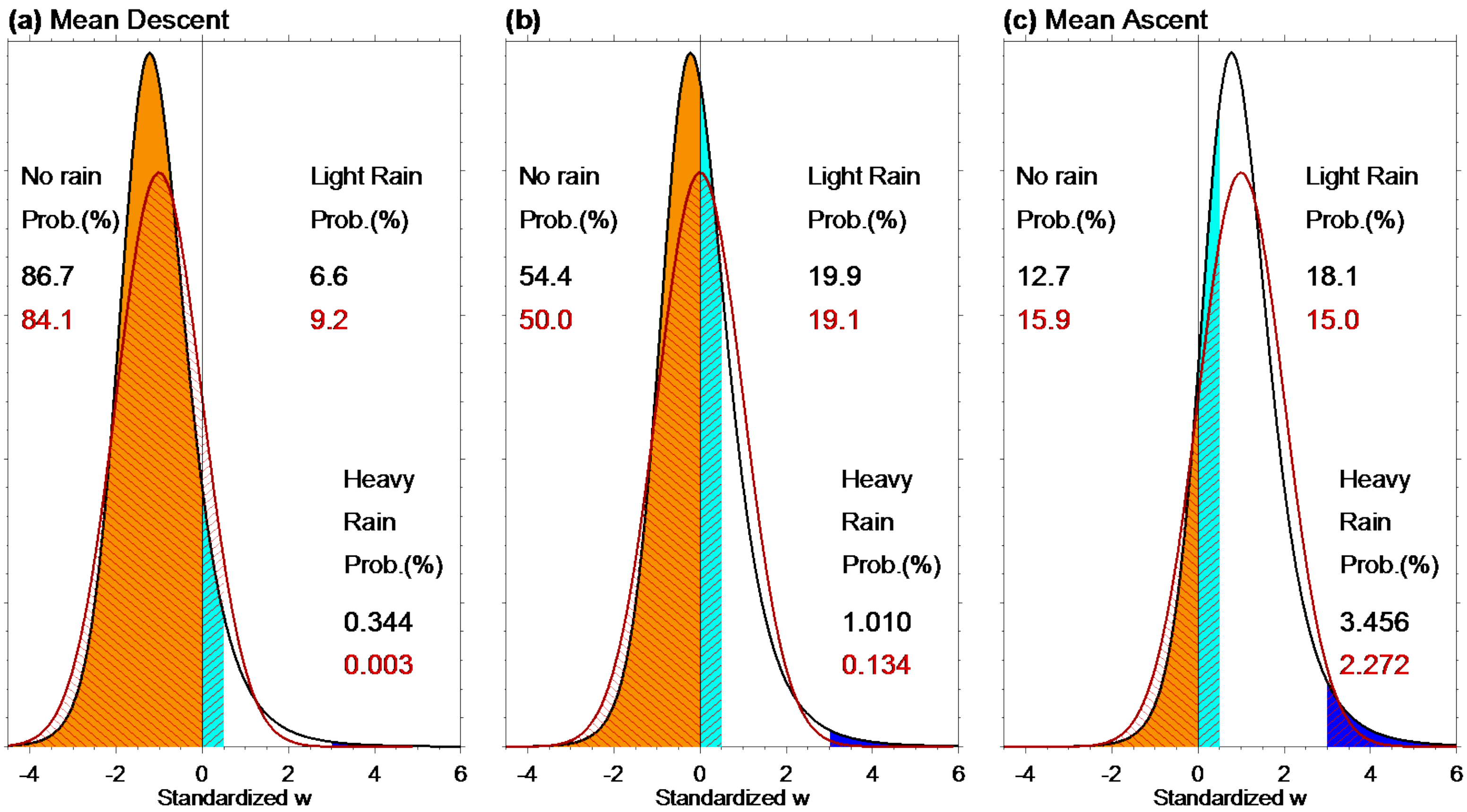
Planned:

- Improved Interface
- Error bars on SGS and Gaussian distributions
- Error bars on probabilities
- Area averages (e.g. global)
- Land/Ocean areas
- More statistics: Autocorrelation, trends…

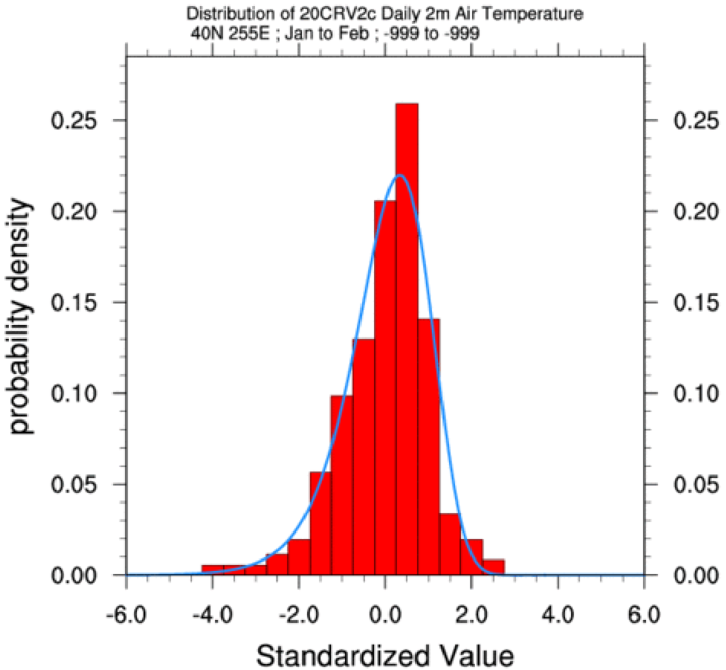
More Coming Attractions…

- Precipitation
  - Gamma and other distributions are used “because they work” for precipitation
  - But, omega can be fitted by an SGS distribution. And, the precipitation distribution corresponds to the omega distribution as in the next plot

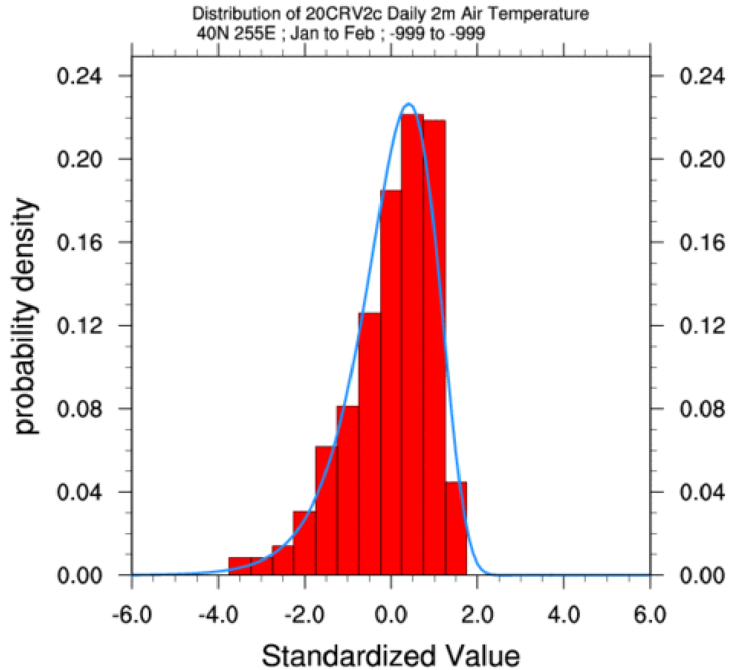
Omega and Precipitation Distributions



More Coming Attractions



Boulder, CO JF El Niño 2m T



Boulder, CO JF La Niña 2m T

Mean:	271.98	vs.	269.52
STD:	3.03	vs.	4.87
Skew:	-.80	vs.	-1.02
Kurtosis:	1.50	vs.	.83
Minimum:	260.0	vs.	251.6
Maximum:	279.6	vs.	277.1

- And that concludes my talk and the 40th CDW