Comparisons of Mixed Gaussian-Lognormal, Logarithmic Transform, and Gaussian fits all based temperature-mixing ratio Microwave retrieval Systems.

Steven J. Fletcher, Michael R. Goodliff, Anton J. Kliewer, Andrew S. Jones and John M. Forsythe

Cooperative Institute for Research in the Atmosphere, Colorado State University

Plan of talk

- Introduction to the mixed Gaussian-lognormal Distribution
- Properties of the mixed distribution
- Plots of the mixed distributions
- Applying the mixed distribution to a variational formulation
- Application of the mixed distribution to microwave brightness temperature based temperature-humidity retrievals
- Comparison with Gaussian and the logarithmic transform approach.
- CIRA Data Assimilation Testbed - CDAT

Mixed Gaussian-Lognormal distribution

The mixed distribution in its bivariate formulation is defined by

$$MX(\mu_G, \mu_L, \sigma_G, \sigma_L, \rho_{mx}) \equiv \frac{1}{\sqrt{|\mathbf{\Sigma}_{mx}| 2\pi x_2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} x_1 - \mu_G \\ \ln x_2 - \mu_L \end{pmatrix}^T \mathbf{\Sigma}_{\mathrm{mx}}^{-1} \begin{pmatrix} x_1 - \mu_G \\ \ln x_2 - \mu_L \end{pmatrix}\right\}$$

Where

$$\mathbf{\Sigma}_{mx} = \begin{pmatrix} VAR(X_1) & COV(X_1, \ln X_2) \\ COV(X_1, \ln X_2) & VAR(\ln X_2) \end{pmatrix}$$

Note that the variance of the lognormal component is with respect to $\ln X_2$, and that the covariance between the Gaussian and the lognormal random variables is between X_1 and $\ln X_2$.

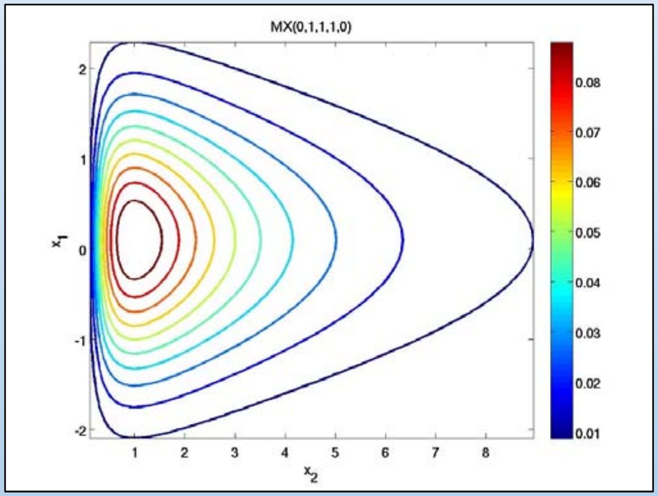
Properties of the Mixed Distribution

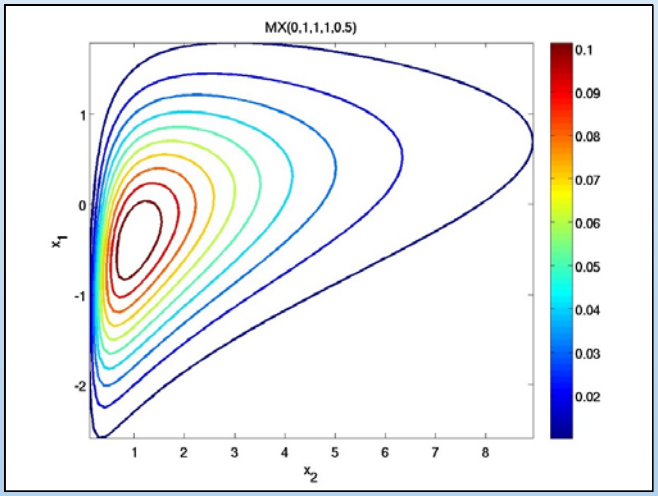
An important property of the mixed distribution is the definitions of the three descriptive statistics. The mean for each component can be found through forming the marginal and joint pdfs which can be shown to be Gaussian and lognormal, or vice-versa. Therefore the mean, mode and median are given by

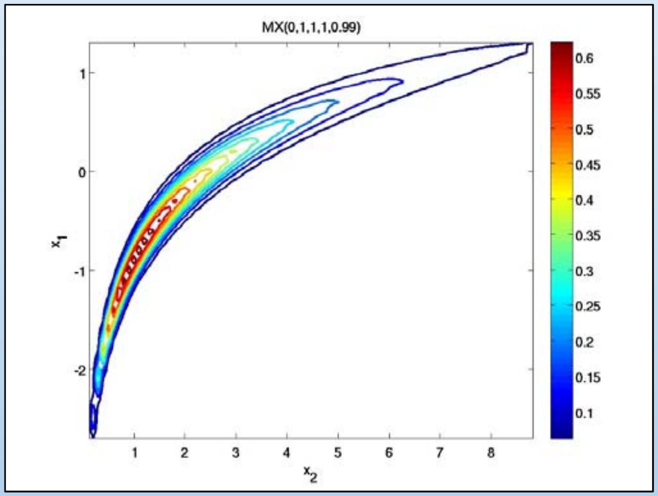
$$mean \equiv \begin{pmatrix} \mu_G \\ \exp\left\{\mu_L + \frac{\sigma_L^2}{2}\right\} \end{pmatrix}, \qquad median \equiv \begin{pmatrix} \mu_G \\ \exp\{\mu_L\} \end{pmatrix},$$

$$mode \equiv \begin{pmatrix} \mu_G - \rho\sigma_G\sigma_L \\ \exp \mu_L - \sigma_L^2 \end{pmatrix}$$

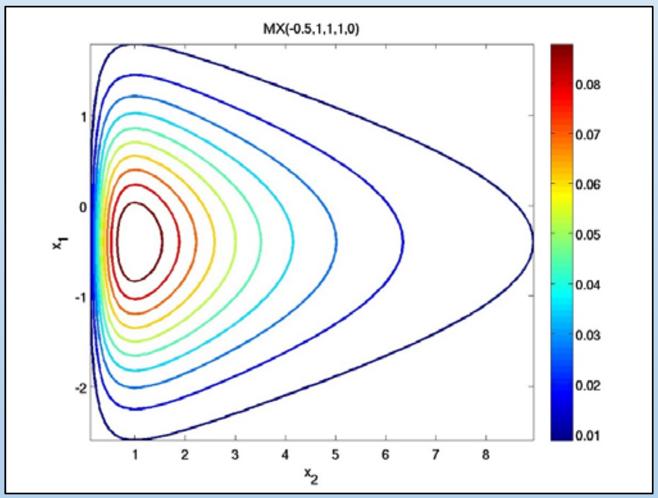
Plots of the Mixed Distribution

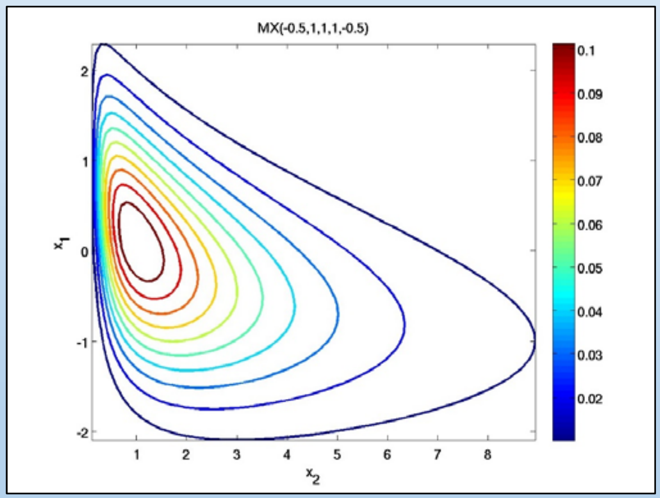


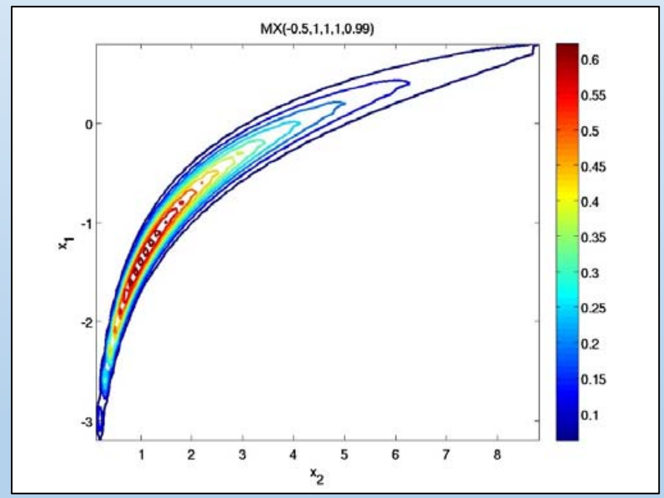




Plots of the Mixed Distribution







Applying the Mixed Distribution to VAR

To be able to apply the mixed distribution to a variational formulation we require the definitions for the errors along with the multivariate version of the mixed distribution. The background and observational errors are given by

$$\boldsymbol{\varepsilon}_b \equiv \begin{pmatrix} x^t_{p_1} - x^b_{p_1} \\ \dfrac{x^t_{q_1}}{x^b_{q_1}} \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_o \equiv \begin{pmatrix} y_{p_2} - h_{p_2}(x) \\ \dfrac{y_{q_2}}{h_{q_2}(x)} \end{pmatrix}$$

Where there are different number of Gaussian and observational background and observational errors, and that $N = p_1 + q_1$ and $N_o = p_2 + q_2$.

Applying the Mixed Distribution to VAR

The multivariate version of the mixed distribution is defined by

$$MX(\boldsymbol{\mu}_{mx}, \boldsymbol{\Sigma}_{\mathrm{mx}}) \equiv \frac{1}{\sqrt{|\boldsymbol{\Sigma}_{mx}|(2\pi)^{\frac{N}{2}}}} \prod_{i=p+1}^{N} \frac{1}{x_i} \exp\left\{-\frac{1}{2}\begin{pmatrix} \boldsymbol{x}_p - \boldsymbol{\mu}_p \\ \ln \boldsymbol{x}_2 - \boldsymbol{\mu}_q \end{pmatrix}^T \boldsymbol{\Sigma}_{\mathrm{mx}}^{-1} \begin{pmatrix} \boldsymbol{x}_p - \boldsymbol{\mu}_p \\ \ln \boldsymbol{x}_2 - \boldsymbol{\mu}_q \end{pmatrix}\right\}$$

Where

$$\boldsymbol{\mu}_{\mathrm{mx}} \equiv \begin{pmatrix} \boldsymbol{\mu}_p \\ \boldsymbol{\mu}_q \end{pmatrix} \quad \boldsymbol{\Sigma}_{\mathrm{mx}} \equiv \begin{pmatrix} \boldsymbol{\Sigma}_{pp} & \boldsymbol{\Sigma}_{pq} \\ \boldsymbol{\Sigma}_{qp} & \boldsymbol{\Sigma}_{\mathrm{qq}} \end{pmatrix}$$

which leads to the mode of the multivariate mixed distribution is given by

$$\boldsymbol{x_{mode}} = \begin{pmatrix} \boldsymbol{\mu}_p - \langle \boldsymbol{\Sigma}_{pq}, \boldsymbol{1}_q \rangle \\ \exp\{\boldsymbol{\mu}_q - \langle \boldsymbol{\Sigma}_{qq}, \boldsymbol{1}_q \rangle\} \end{pmatrix}$$

Applying the Mixed Distribution to VAR

If we now follow the standard log-likelihood approach for variational data assimilation through Bayes theorem then we obtain the 3DVAR cost function for the mixed distribution as

$$J_{mx}(\boldsymbol{x^t})$$
$$= \frac{1}{2}\begin{pmatrix} \boldsymbol{x}_{p_1}^t - \boldsymbol{x}_{p_1}^b \\ \ln \boldsymbol{x}_{q_1}^t - \ln \boldsymbol{x}_{q_1}^b \end{pmatrix}^{\mathrm{T}} \boldsymbol{B}_{\mathrm{mx}}^{-1} \begin{pmatrix} \boldsymbol{x}_{p_1}^t - \boldsymbol{x}_{p_1}^b \\ \ln \boldsymbol{x}_{q_1}^t - \ln \boldsymbol{x}_{q_1}^b \end{pmatrix} + \left\langle \begin{pmatrix} \boldsymbol{x}_{p_1}^t - \boldsymbol{x}_{p_1}^b \\ \ln \boldsymbol{x}_{q_1}^t - \ln \boldsymbol{x}_{q_1}^b \end{pmatrix}, \begin{pmatrix} \boldsymbol{0}_{p_1} \\ \boldsymbol{1}_{q_1} \end{pmatrix} \right\rangle$$
$$+ \frac{1}{2}\begin{pmatrix} \boldsymbol{y}_{p_2} - \boldsymbol{h}_{p_2}(\boldsymbol{x}^t) \\ \ln \boldsymbol{y}_{q_2} - \ln \boldsymbol{h}_{q_2}(\boldsymbol{x}^t) \end{pmatrix}^{T} \boldsymbol{R}_{mx}^{-1} \begin{pmatrix} \boldsymbol{y}_{p_2} - \boldsymbol{h}_{p_2}(\boldsymbol{x}^t) \\ \ln \boldsymbol{y}_{q_2} - \ln \boldsymbol{h}_{q_2}(\boldsymbol{x}^t) \end{pmatrix}$$
$$+ \left\langle \begin{pmatrix} \boldsymbol{y}_{p_2} - \boldsymbol{h}_{p_2}(\boldsymbol{x}^t) \\ \ln \boldsymbol{y}_{q_2} - \ln \boldsymbol{h}_{q_2}(\boldsymbol{x}^t) \end{pmatrix}, \begin{pmatrix} \boldsymbol{0}_{p_2} \\ \boldsymbol{1}_{q_2} \end{pmatrix} \right\rangle$$

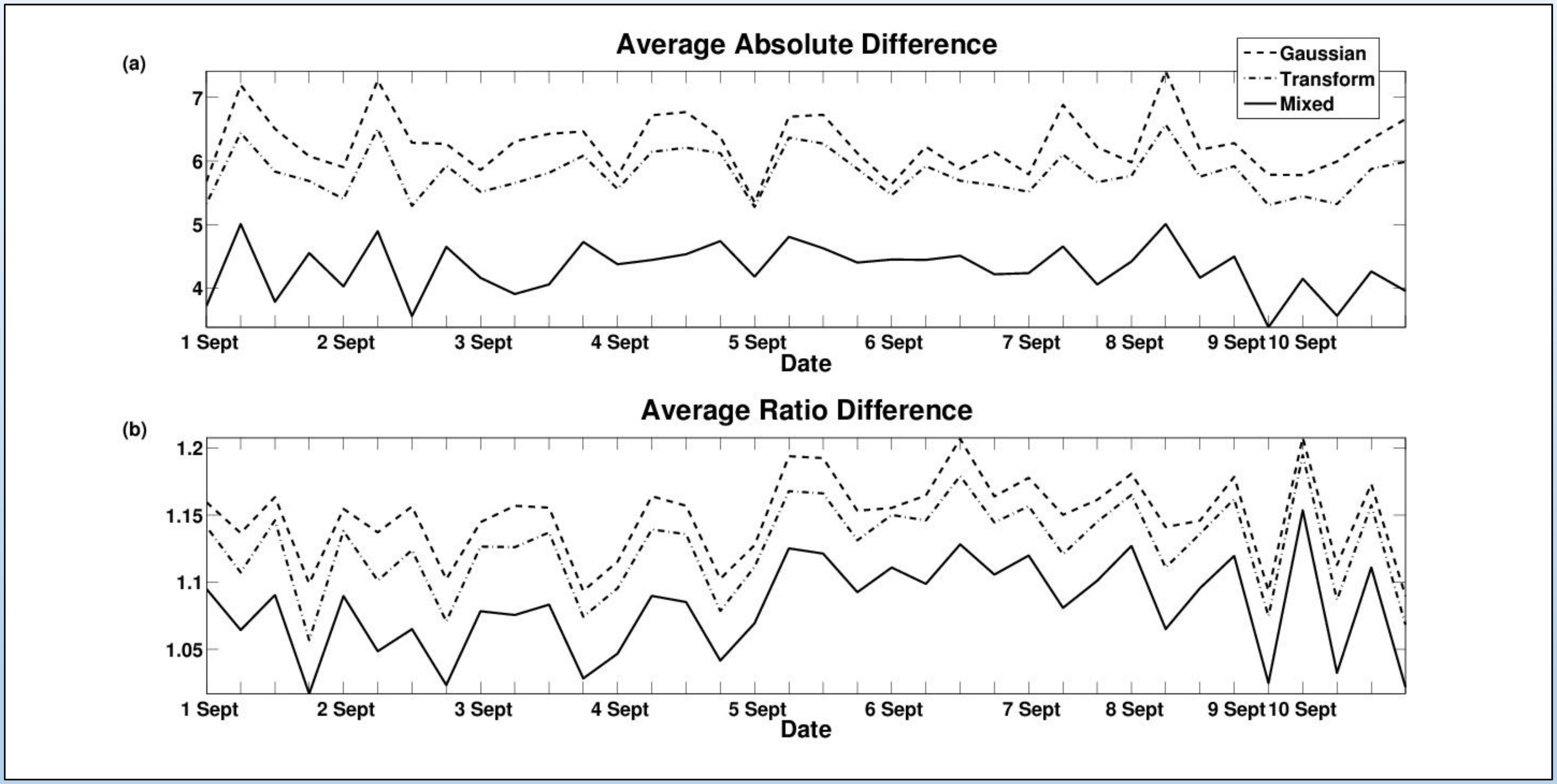
Application of the Mixed Distribution

The CIRA 1-Dimensional Optimal Estimator (C1DOE) is a 1DVAR retrieval system for mixing-ratio and temperature from microwave brightness temperatures. Its original version is a Gaussian fits all formulation.

We have now implemented the mixed distribution approach where we are assuming lognormal errors for the mixing-ratio, and Gaussian for the temperature.

Along with the mixed distribution and Gaussian fits all approaches we have also implemented the logarithmic transform approach for mixing-ratio (Kliewer et al 2016).

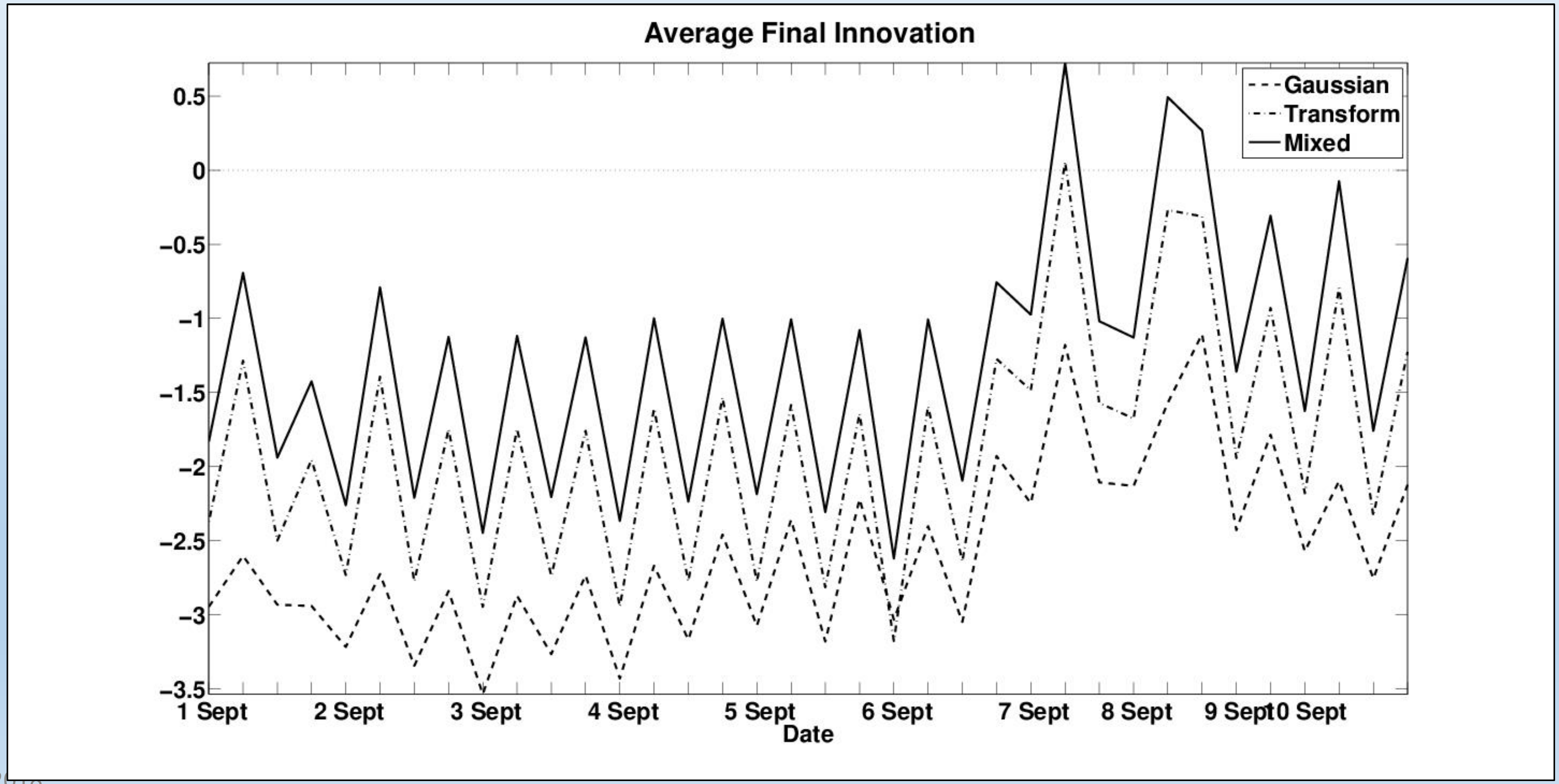
Application of the Mixed Distribution



Comparisons of the three retrieval methods against the Microwave Surface and Precipitation Products Systems (MSPPS) TPW product. Solid is the mixed approach, dot-dashed is the transform and the dashed is the Gaussian.

Application of the Mixed Distribution

AMSU-A Channel 6 (54.4GHz) (Temperature Channel in the troposphere) Final Innovations



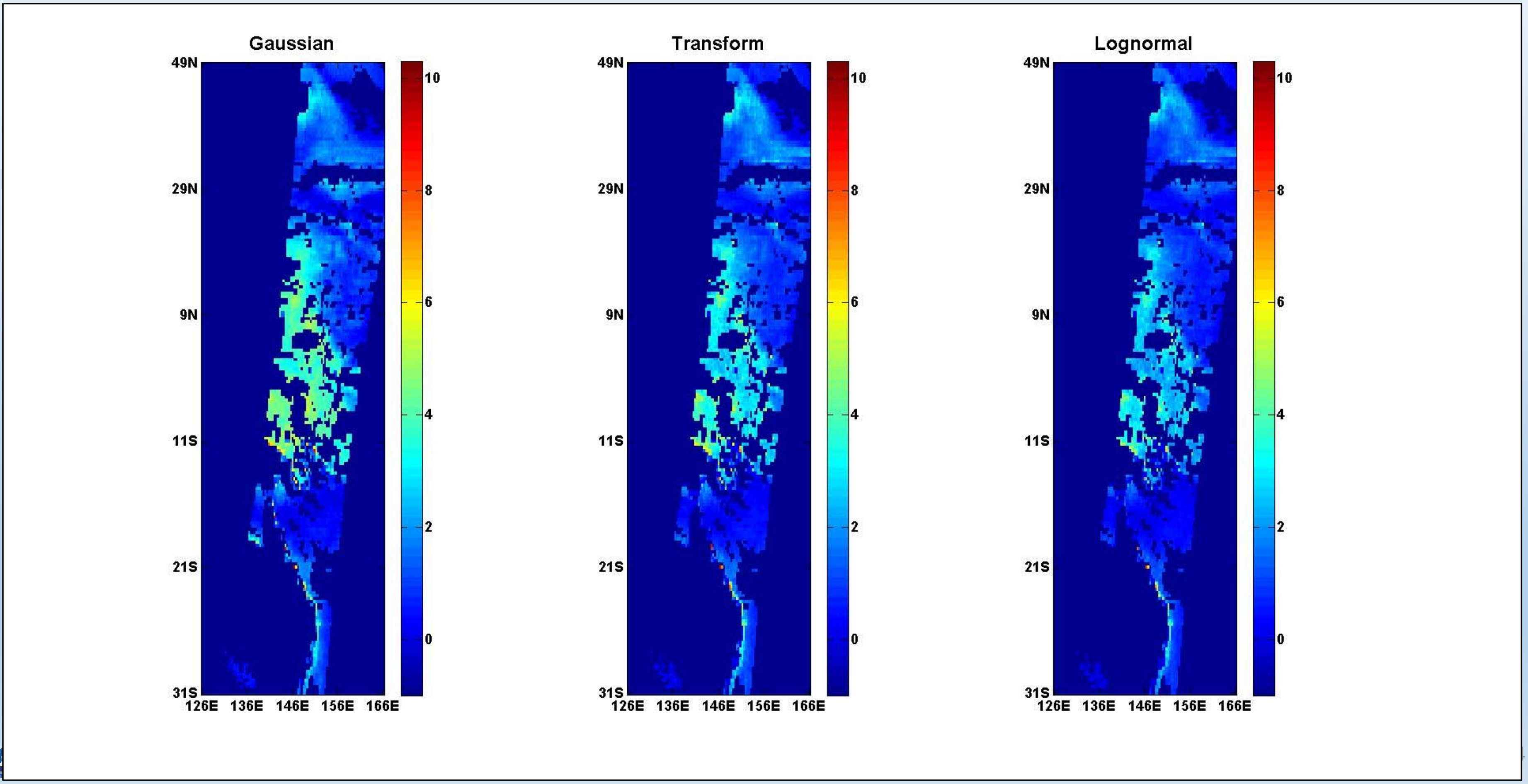
CIRA Data Assimilation Testbed (CDAT)

As part of a new NSF award to CIRA we are developing a website that will be running the three different versions of the retrieval system in near real time in different regions over the Earth.

Along with the output from the three different retrievals systems, there will also be the output from the detection algorithm presented in Mike Goodliff's poster. The purpose of also showing the detection algorithm output the website is for users to look at where a non-Gaussian signal has been detected and to look at the impact on the values from the retrieval systems to make decision of where the Gaussian fits all could be sub-optimal.

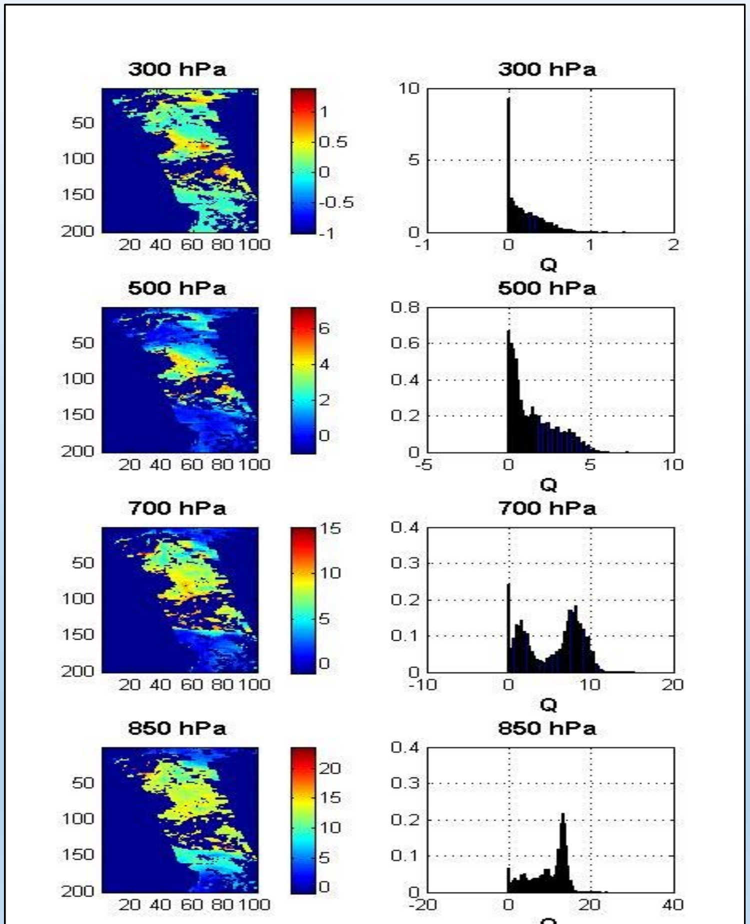
The website address will be www.cdat.cira.colostate.edu

CDAT

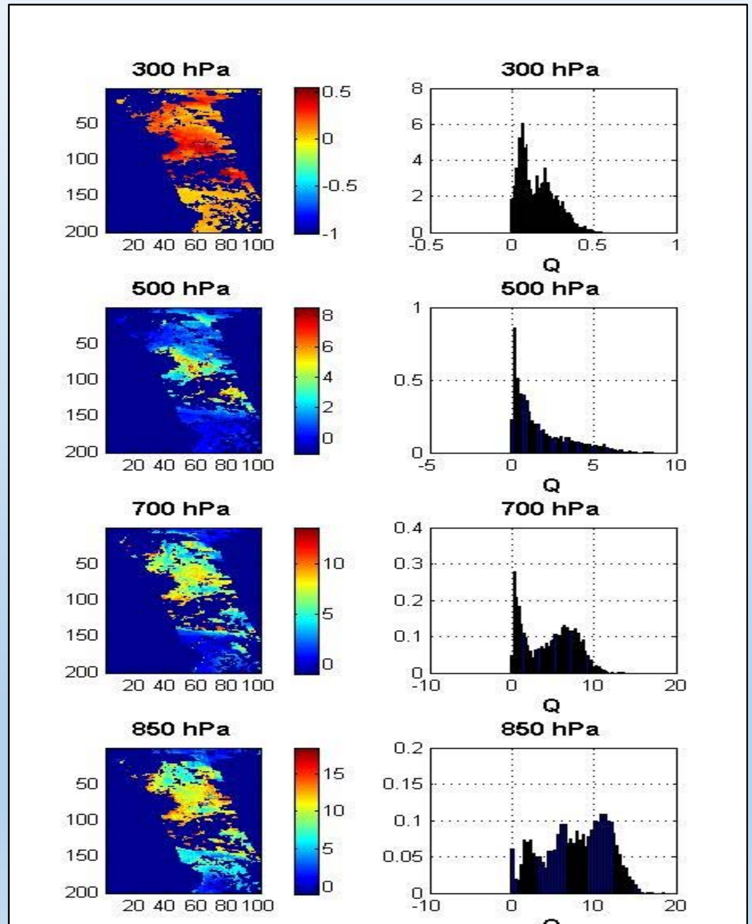


Gaussian
Transform
Lognormal
49N
29N
9N
11S
21S
31S
126E 136E 146E 156E 166E
10
8
6
4
2
0

www.cdat.cira.colostate.edu



Gaussian Fits all



Mixed Gaussian-Lognormal

Conclusions and Further Work

- Have presented a multivariate PDF that is a models the behavior of Gaussian and lognormal random variables simultaneously.
- Have shown that the Gaussian component of the mode of the distribution is a function of the covariances between the Gaussian and lognormal random variables.
- Presented results of applying the mixed distribution in a 1DVAR retrieval system, and compared its performance against Gaussian fits all and the logarithmic approach.
- Introduced the CIRA Data Assimilation Testbed (CDAT)

Conclusions and Further Work

- To develop the hybrid version of the mixed distribution based variational data assimilation
- Develop the buddy check and variational quality control measures
- Develop non-Gaussian detection algorithm to make the mixed distribution dynamically based.
- Implement the mixed distribution into the WRF-GSI for both the static and hybrid formulations.