Observational Quantification of Non-Gaussian Errors Within a Humidity-Temperature 1DVAR Retrieval System over Japan

Anton J. Kliewer, Steven J. Fletcher, Andrew S. Jones, John M. Forsythe
Cooperative Institute for Research in the Atmosphere (CIRA)
Colorado State University (CSU), Fort Collins CO

Sixth WMO Symposium on Data Assimilation
College Park, MD
October 7-11, 2013

INTRODUCTION

The emphasis of the NSF-funded research activities are focused on non-Gaussian data assimilation extensions to the existing DPEAS CIRA one-dimensional variational (1DVAR) Optimal Estimator (C1DOE).

The Gaussian data assimilation assumption is made not just in the larger DA systems, but also within many 1D variational retrieval systems, such as C1DOE, which are Maximum Likelihood estimation (MLE) Bayesian approaches. Thus, these 1D systems offer unique observational test beds for data assimilation methodology research studies.

The baseline DPEAS data fusion methods include the use of 1DVAR data assimilation for satellite sounding data sets, and numerous real-time statistical analysis methods. Our new development activities to extend the current 1DVAR algorithm to use and test a new non-Gaussian data assimilation method are outlined below. After this initial observational testing stage, we plan more extensive 3DVAR WRF-based non-Gaussian extensions, leveraging what we learn in the 1DVAR non-Gaussian data assimilation DPEAS framework.

GOALS

1. Perform synthetic comparison tests.
2. Develop a mixed distribution 1DVAR approach using the C1DOE framework components to see the impact of the different assumptions for the humidity component of the retrieval system.
3. Compare this new mixed distribution approach against the assumed Gaussian and the transform approaches for different dynamical scales and for different seasons.
4. Use conclusions from goals 1 and 2 to guide the design of a mixed distribution approach to the larger 3D or 4D VAR weather data assimilation systems.

MOTIVATION OF RETRIEVAL AREA

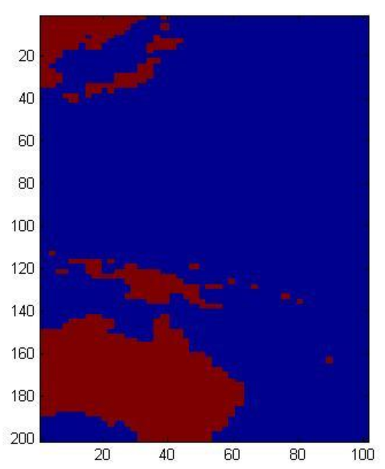
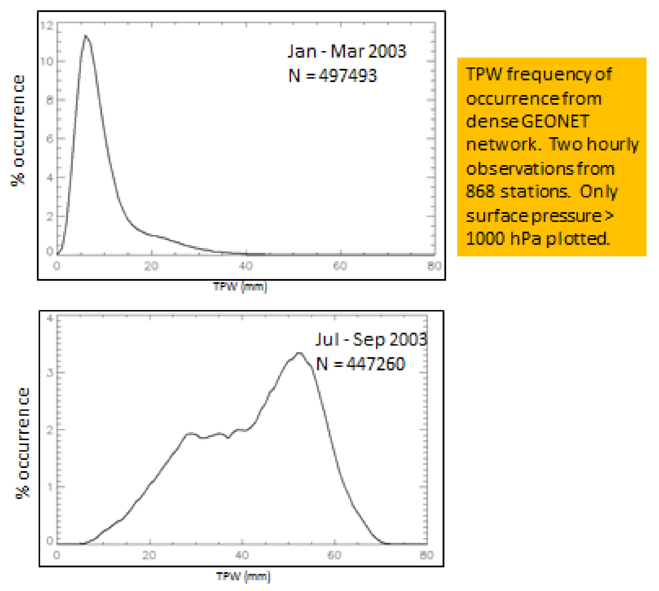


Figure 1: **T**he data has b
remapped to a 0.5 degree
grid (~ 55 km resolution at
equator) for this experime
the native resolution of
AMSU-B is 16 km at nadir
higher resolution would be
possible in the future

Figure 2: Distribution of TPW from the GEONET network around Japan, about 800 stations, from 2003 for winter and summer. Lognormal and mixed-distributions are clearly seen.



DPEAS/C1DOE ESSENTIALS

C1DOE in its current form is a traditional 1DVAR retrieval system with enhanced global data inputs, that make numerous studies and intercomparisons possible due to the C1DOE algorithm residing within the DPEAS framework (see Fig. 1). The diagnostic capabilities within C1DOE are robust, with over 500 data and diagnostic fields being generated for each satellite field-of-view. It is also possible to combine GPS and Infrared satellite data sets within the same data flow, making cloud analysis and advanced surface effect studies much easier.

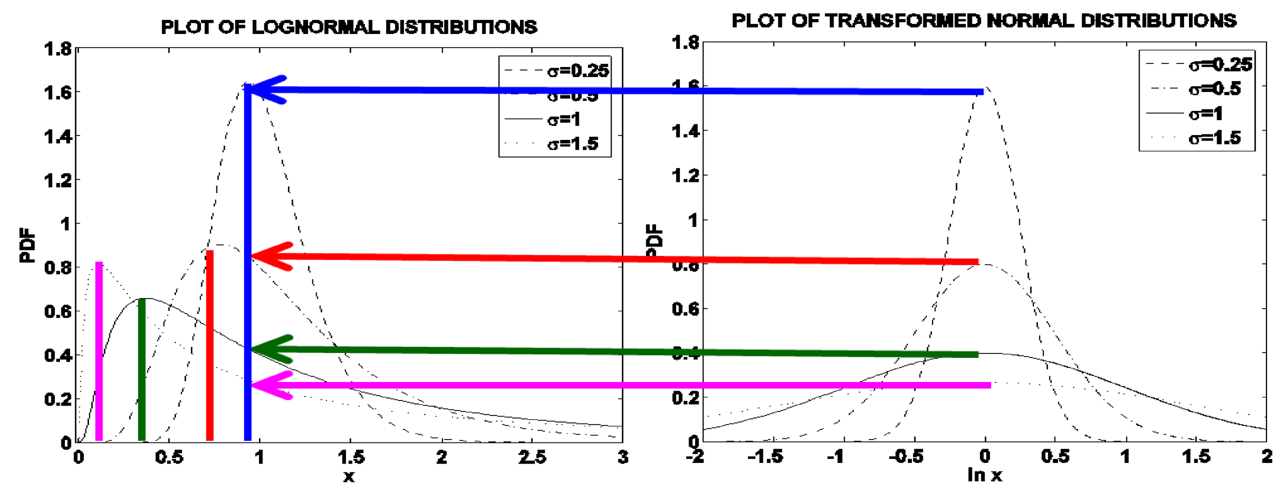


Figure 3: Illustration of the transform and implications between a) lognormal and b) Gaussian spaces. The horizontal blue, red, green, and magenta lines indicate the inverse transform from the transformed normal distribution back to the lognormal distribution space for lognormal distributions of σ = 0.25, 0.5, 1.0, and 1.5, respectively. When inverted back from the Gaussian transform analysis space, the transform approach finds the median in the lognormal space, and thus loses all the skewness information contained in the original lognormal distribution, where the vertical blue, red, green, and magenta lines indicate the respective original lognormal modes.

COST FUNCTIONS

Using a Bayesian approach we wish to understand the errors associated with the retrieved state vector. Bayes Theorem, $P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}$, gives the probability density function of the state when the measurement is given. *A priori* assumptions about the distribution of the background state leads to different posterior distributions. Maximizing the probability of the state vector *x* corresponds to minimizing the following cost functions for each approach. While the transform and logarithmic models have not been fully implemented yet compelling evidence is given for their necessity.

Gaussian:

$$J_N(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x_b})^T \boldsymbol{B}_N^{-1}(\boldsymbol{x}-\boldsymbol{x_b}) + \frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x}))$$

Lognormal Transformation: $(\boldsymbol{X} = \ln \boldsymbol{x})$

$$J_T(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{X}-\boldsymbol{X}_b)^T \boldsymbol{B}_T^{-1}(\boldsymbol{X}-\boldsymbol{X}_b) + \frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x}))$$

Lognormal:

$$J_L(\boldsymbol{x}) = \frac{1}{2}(\ln \boldsymbol{x} - \ln \boldsymbol{x_b})^T \boldsymbol{B}_L^{-1}(\ln \boldsymbol{x} - \ln \boldsymbol{x_b}) + \frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1}\left(\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})\right) + (\ln \boldsymbol{x} - \ln \boldsymbol{x_b})^T \boldsymbol{1}_N$$

RETRIEVED VALUES

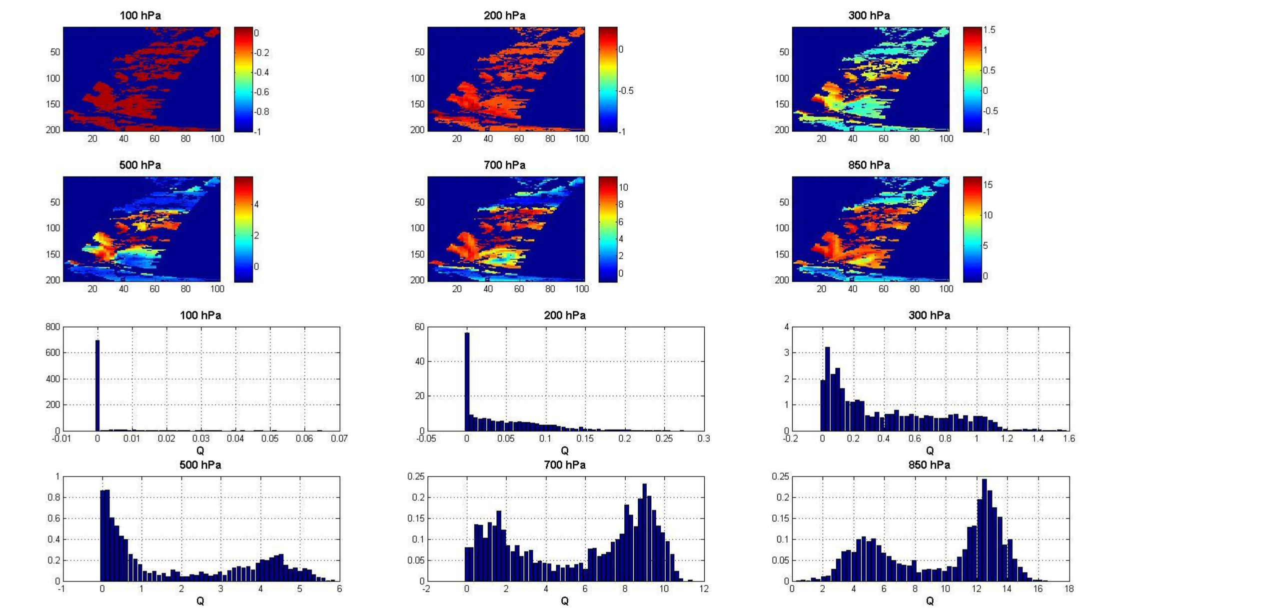


Figure 4: The top panel depicts mixing ratio (g/kg) from the Gaussian C1DOE retrieval system. The bottom panel contains relative frequencies for the corresponding levels in the atmosphere. Note the non-Gaussian behavior, specifically the mixed normal-lognormal distributions. This retrieval is from September 5th, 2005.

CONCLUSIONS

1. Spatial assumptions about the GDAS background distribution and the observational errors following a Gaussian background are clearly insufficient.
2. A lognormal-normal model would better capture the bi-modal distribution corresponding to areas before and after a front.
3. A lognormal-normal model may be required in different regions and may be seasonal-dependent.

References:

1. Donofrio, K., 2007: *A 1DVAR Optimal Estimation Retrieval of Non-Precipitating Water Vapor Profiles Over the Global Oceans Using AMSU,* Colorado State University, M.S. Thesis.
2. Fletcher, S. J., and M. Zupanski, 2006: A data assimilation method for log-normally distributed observational errors. *Quart. J. Roy. Meteor. Soc.,* **132**, 2505-2519.
3. Fletcher, S.J. and M. Zupanski, 2006: A hybrid normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett.* **7**,43-46.
4. Fletcher, S.J. and M. Zupanksi, 2007: Implications and impacts of transforming lognormal variables into normal variables in VAR. *Met. Zeit.* **16,** 755-765.
5. Fletcher, S. J. 2010: Mixed lognormal-Gaussian four-dimensional data assimilation. Tellus, 62A, 266-287.

This work is primarily supported by the National Science Foundation via grant AGS-1038790 at Colorado State University. This research was also supported by multiple grants from the NOAA/NESDIS Product System Development and Implementation (PSDI) program, a NASA ROSES grant, and a grant by the Air Force Weather Agency (AFWA) to the DoD Center for Geosciences / Atmospheric Research (CG/AR) at Colorado State University, as well as a subcontract from the National Center for Atmospheric Research (NCAR) to CSU.

Email Contact: Anton.Kliewer@ColoState.edu