NON-GAUSSIAN 4D VAR

STEVEN J. FLETCHER and MANAJIT SENGUPTA

With thanks to Milija and Dusanka Zupanski, Andy Jones, Laura Fowler
and Thomas Vonder Haar

DoD Center for Geosciences/Atmospheric Research (CG/AR)

COOPERATIVE INSTITUTE FOR RESEARCH IN THE ATMOSPHERE

COLORADO STATE UNIVERSITY

PRESENTATION OUTLINE

- MOTIVATION FOR THE WORK
- REAL LIFE EXAMPLES OF NON-GAUSSIAN VARIABLES
- MATHEMATICAL ILLUSTRATIONS OF THE DRAWBACKS OF CURRENT METHODS
- PROBABILISTIC APPROACH USING CONDITIONAL INDEPENDENCE
- 4-D HYBRID LOGNORMAL – NORMAL DATA ASSIMILATION

MOTIVATION

An important assumption made in variational and ensemble data assimilation is that the state variables and observations are Gaussian distributed

Note: The difference between two Gaussian variables is also a Gaussian variable.

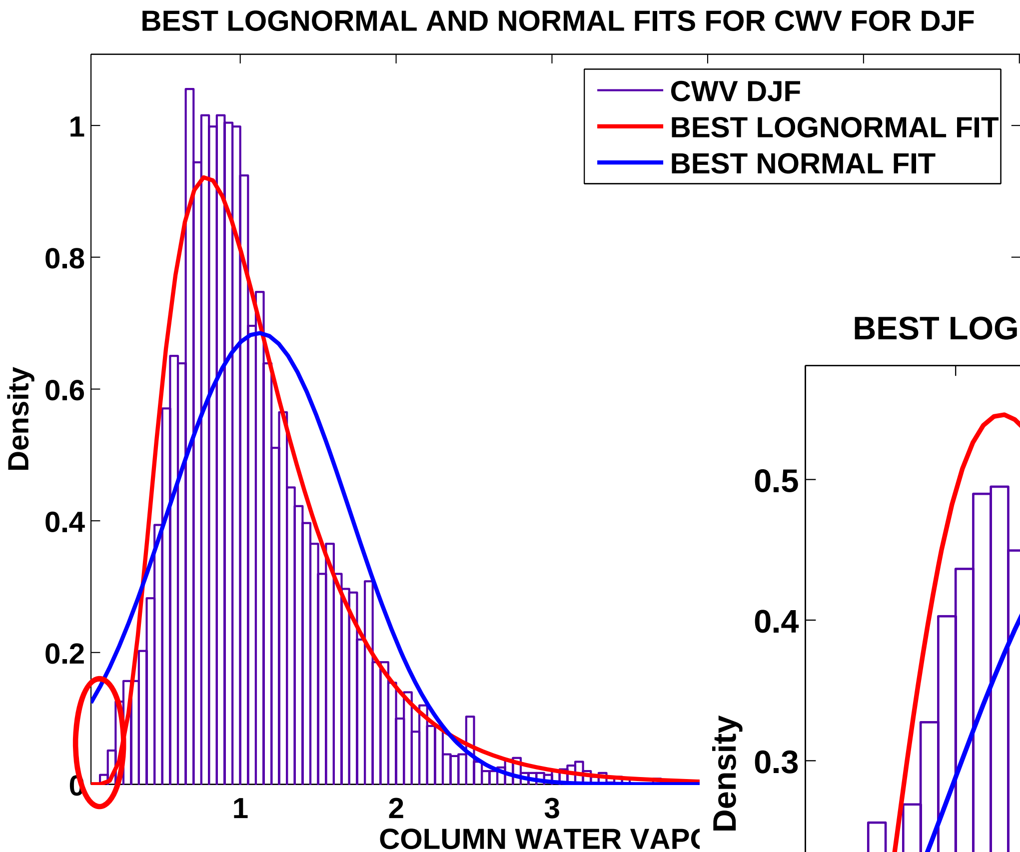
Is this true for all state variables?

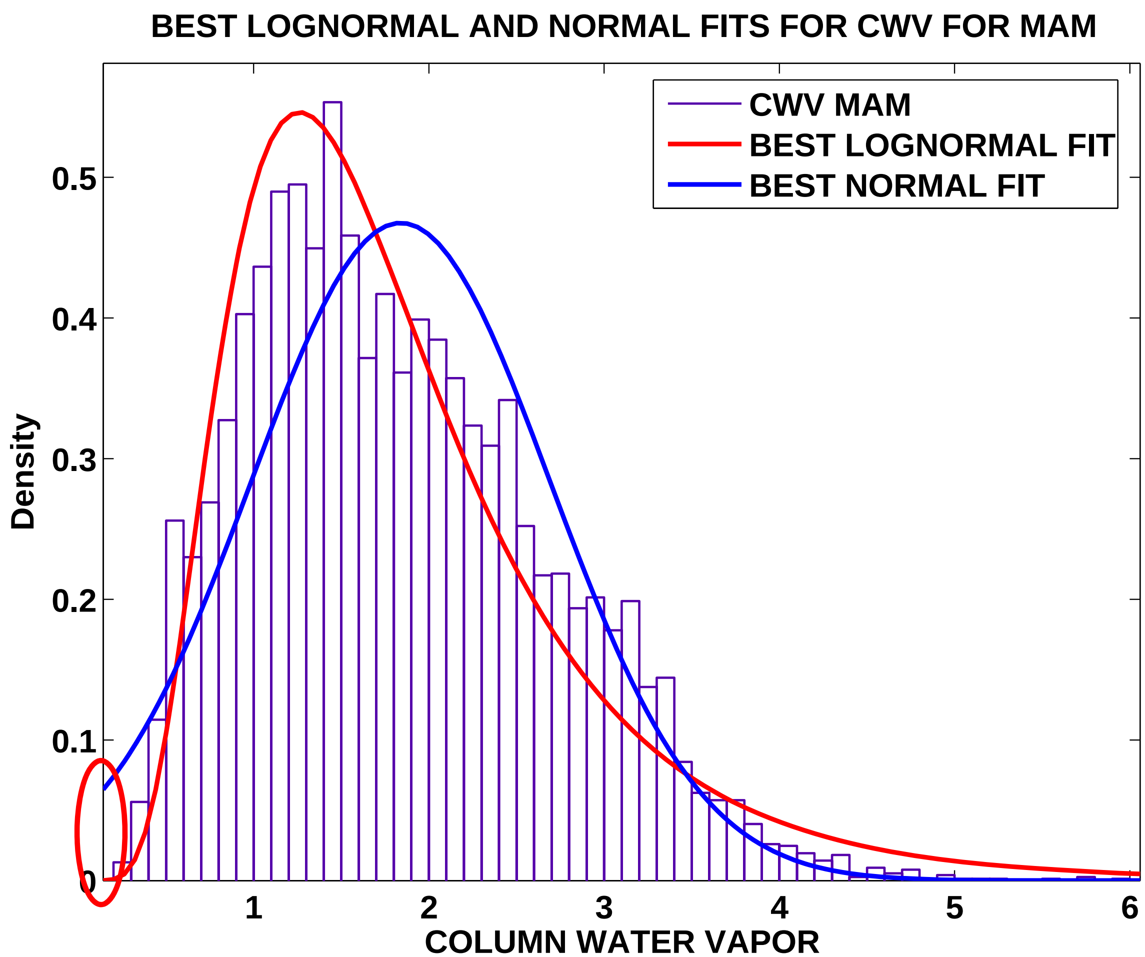
Is this true for all observations of the atmosphere?

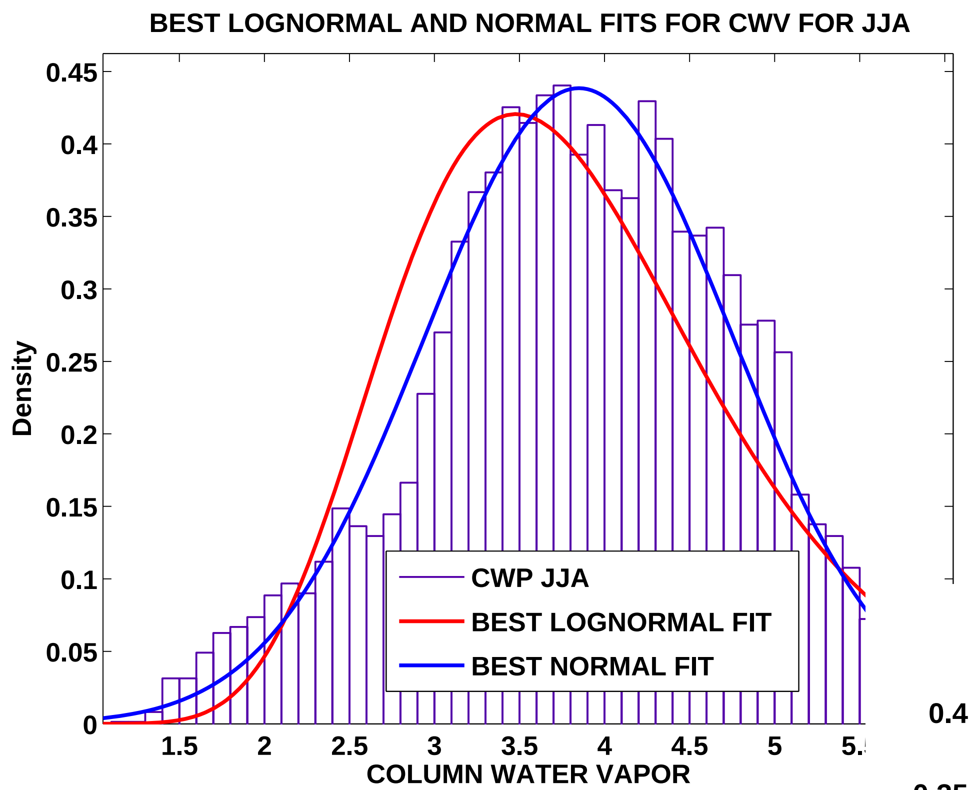
REAL LIFE EXAMPLES

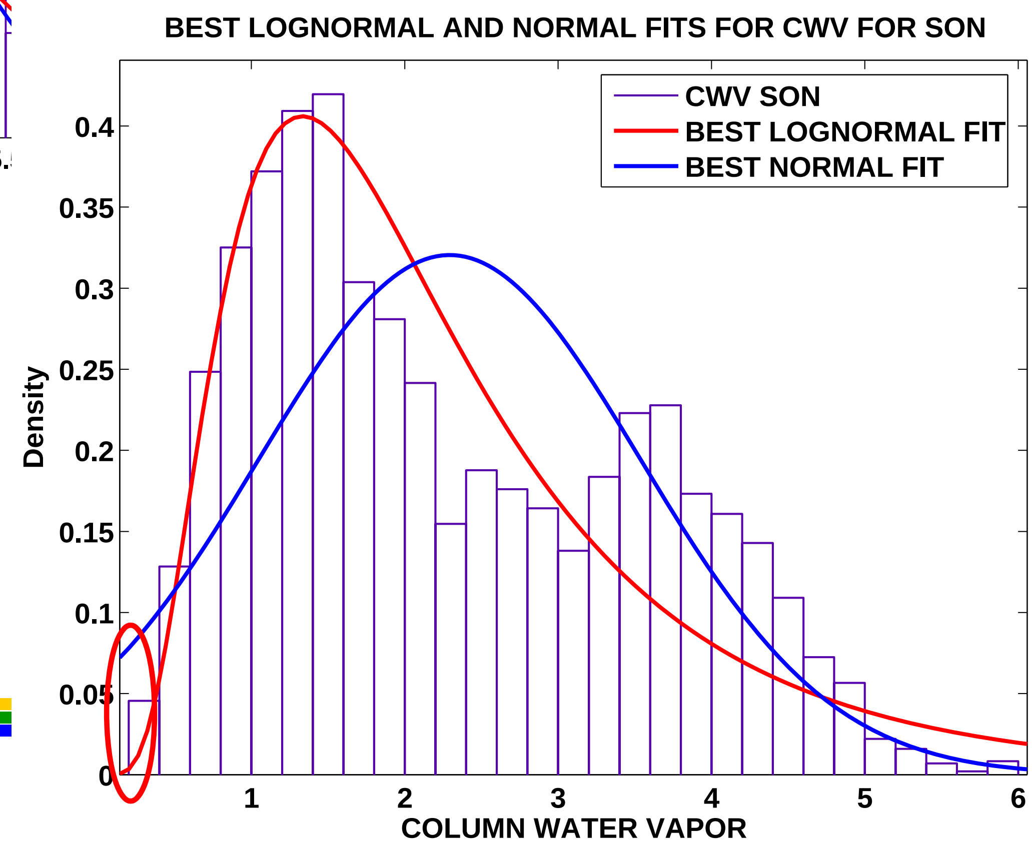
This data is column water vapour climatologies from the Oklahoma ARM-SGP site from 1997-2000 where the data are of when a boundary layer cloud was present

We have taken the observations and have sorted them by season as well as for the whole year. The data was collected using a microwave radiometer.


BEST LOGNORMAL AND NORMAL FITS FOR CWV FOR DJF
CWV DJF
BEST LOGNORMAL FIT
BEST NORMAL FIT
Density
0
0.2
0.4
0.6
0.8
1
1
2
3
COLUMN WATER VAPO

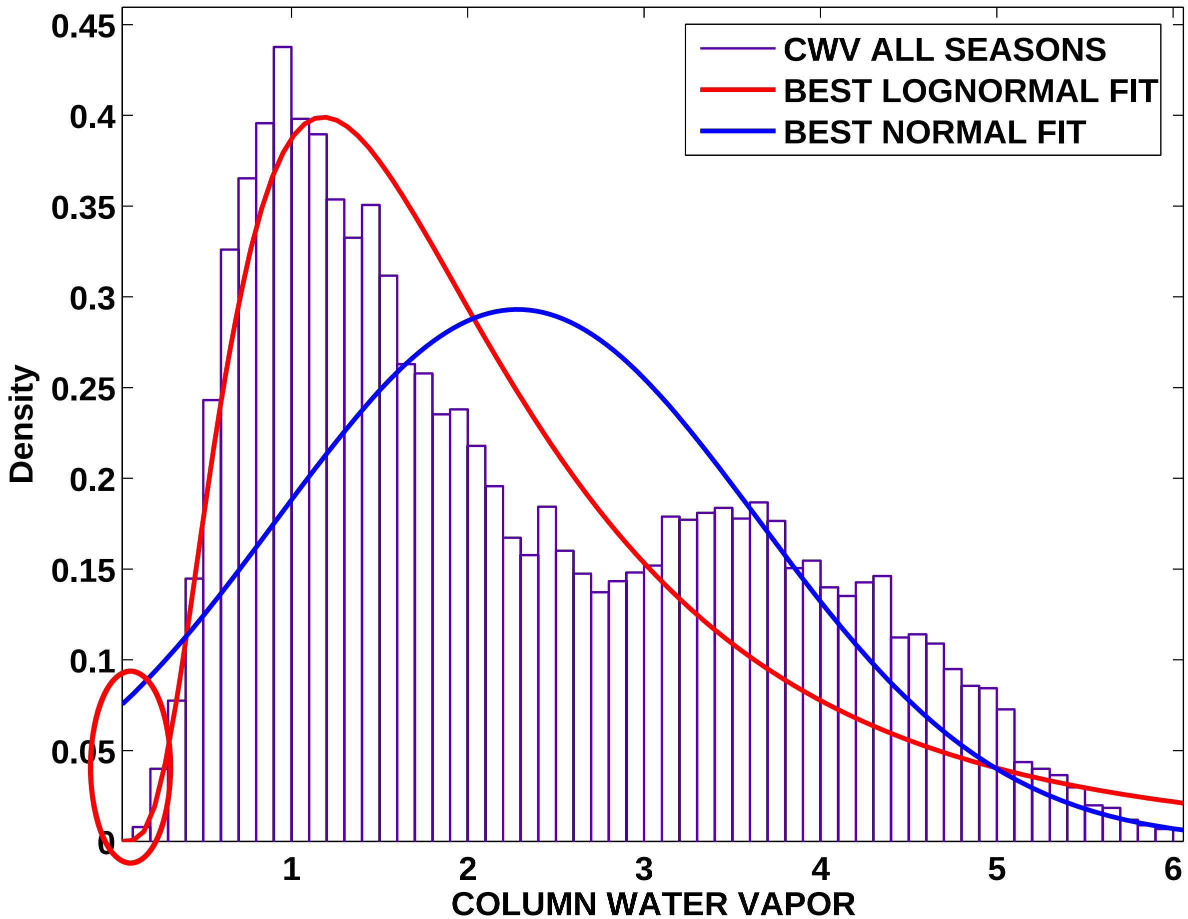

BEST LOGNORMAL AND NORMAL FITS FOR CWV FOR MAM
CWV MAM
BEST LOGNORMAL FIT
BEST NORMAL FIT
Density
0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
COLUMN WATER VAPOR





CSU/CIRA Steven J. Fletcher

BEST LOGNORMAL AND NORMAL FITS FOR CWV FOR ALL SEASONS

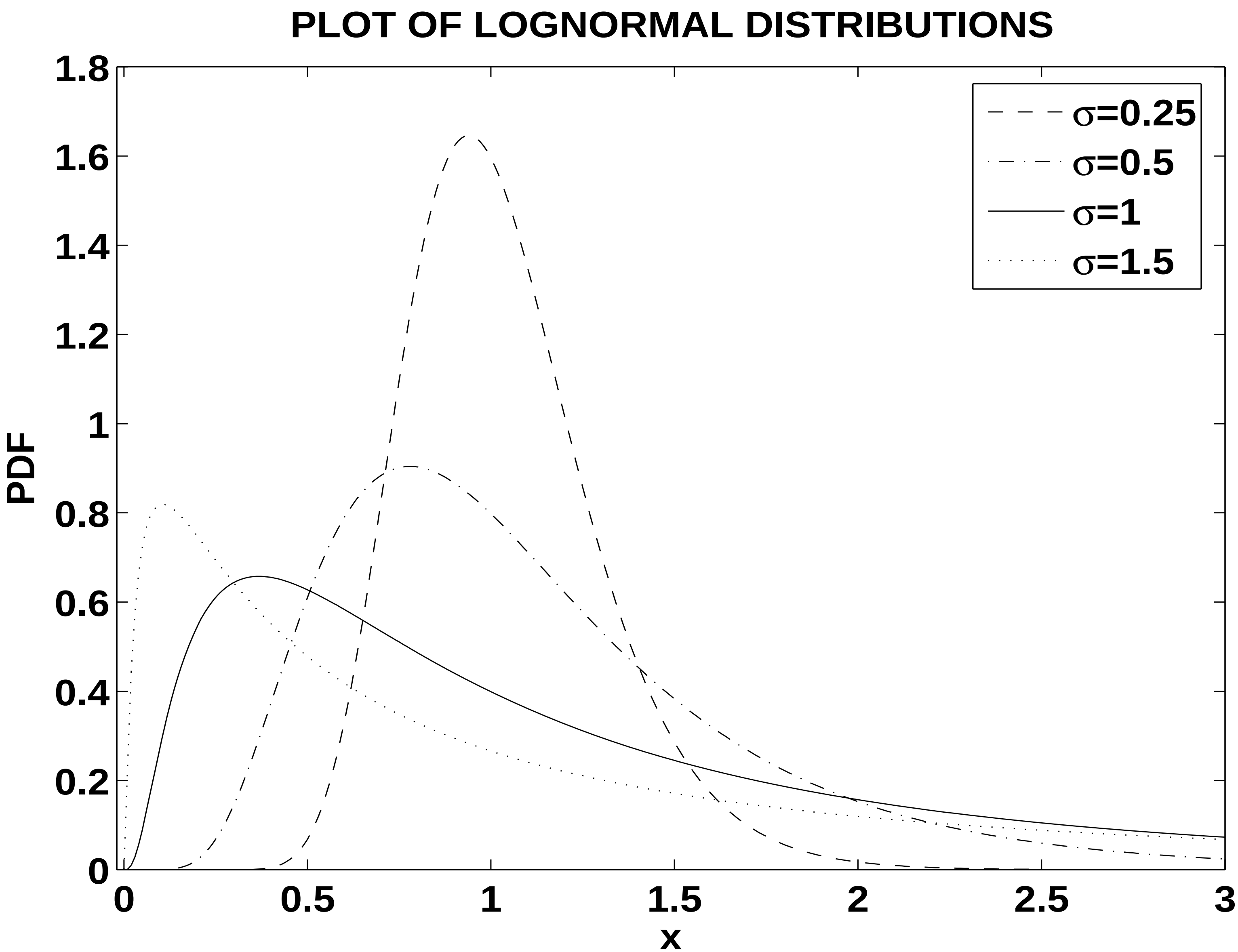


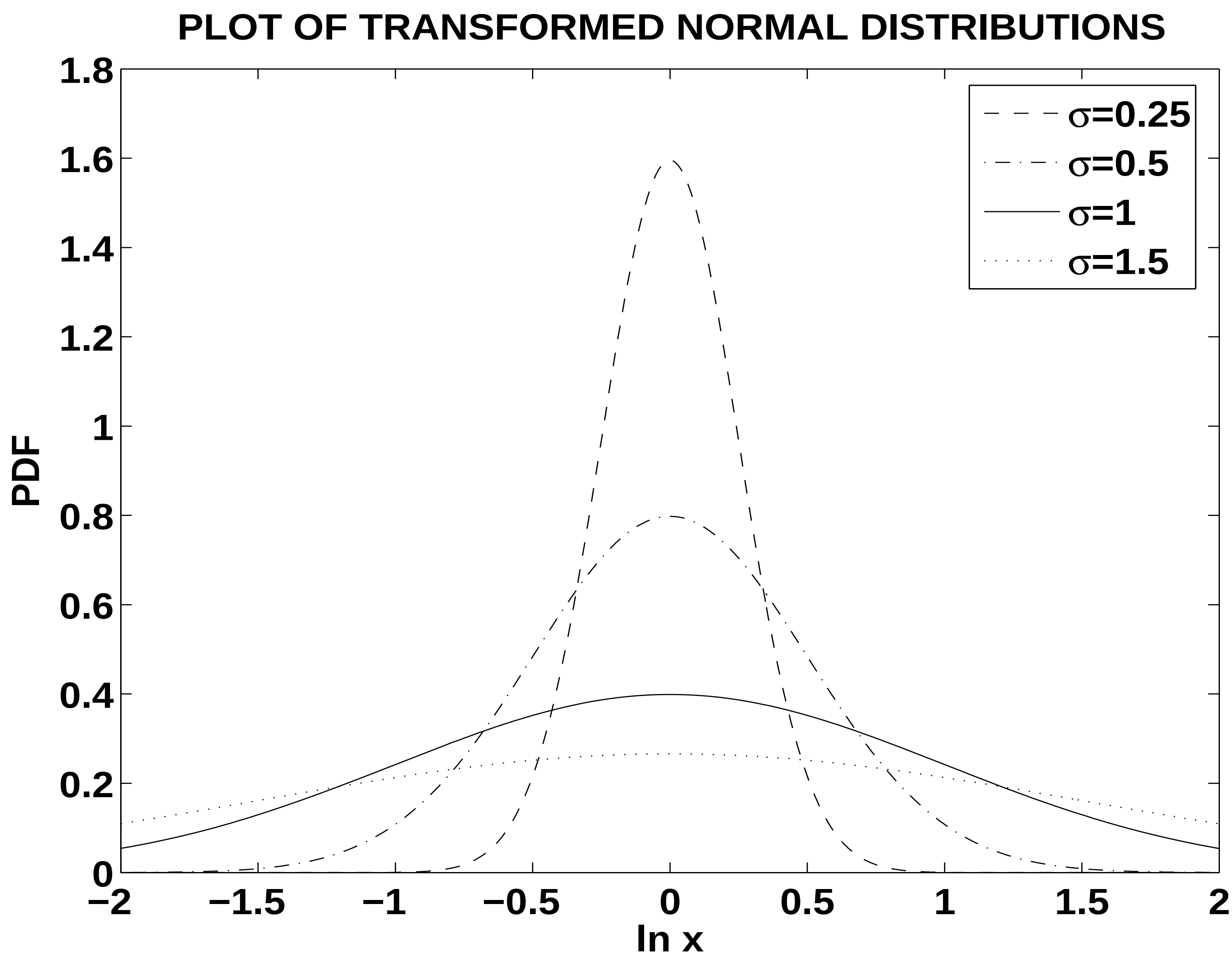
Current Techniques used with non-Gaussian Variables

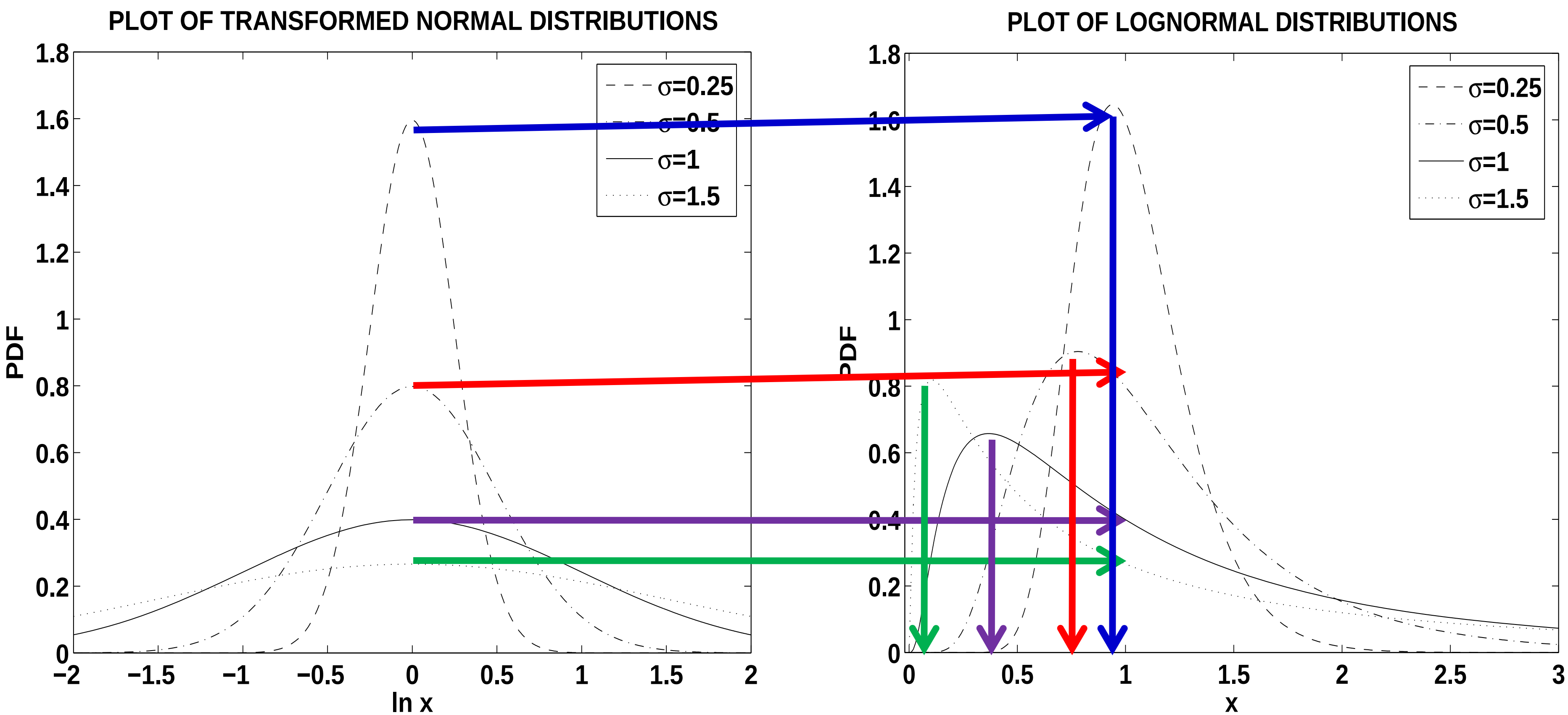
1) Transform by taking the LOGARITHM of the original state variable. This then makes the new variable ALMOST GAUSSIAN. Minimize the cost function with respect to this variable, TRANSFORM BACK and initialize with this state. STATE FOUND IS A NON-UNIQUE MEDIAN OF THE ORIGINAL VARIABLE, (Fletcher and Zupanski 2006a, 2007).

2) Assumed Gaussian assumption and BIAS CORRECT.

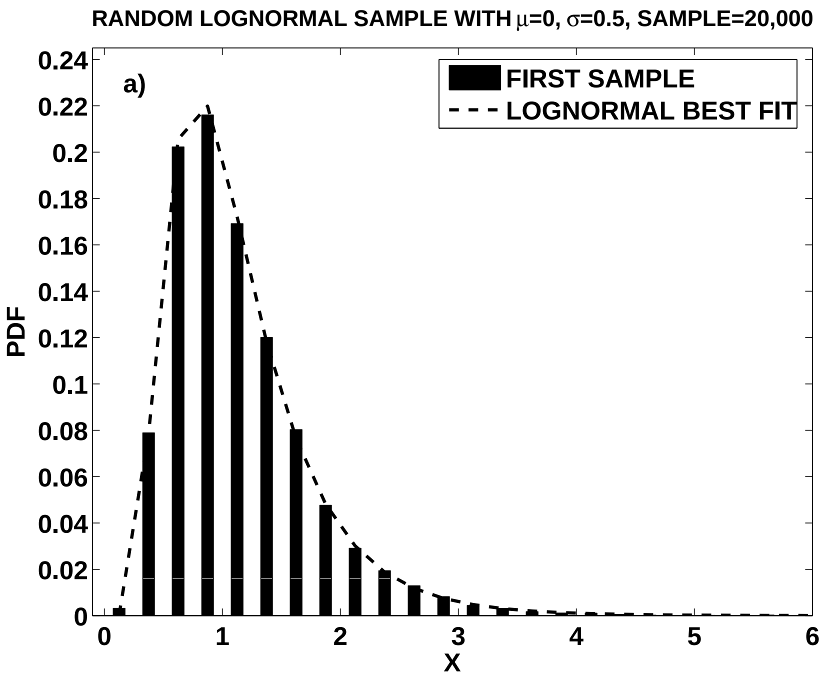
3) Using a Markov-Chain Monte-Carlo approaches (Posselt *et al.* 2008)

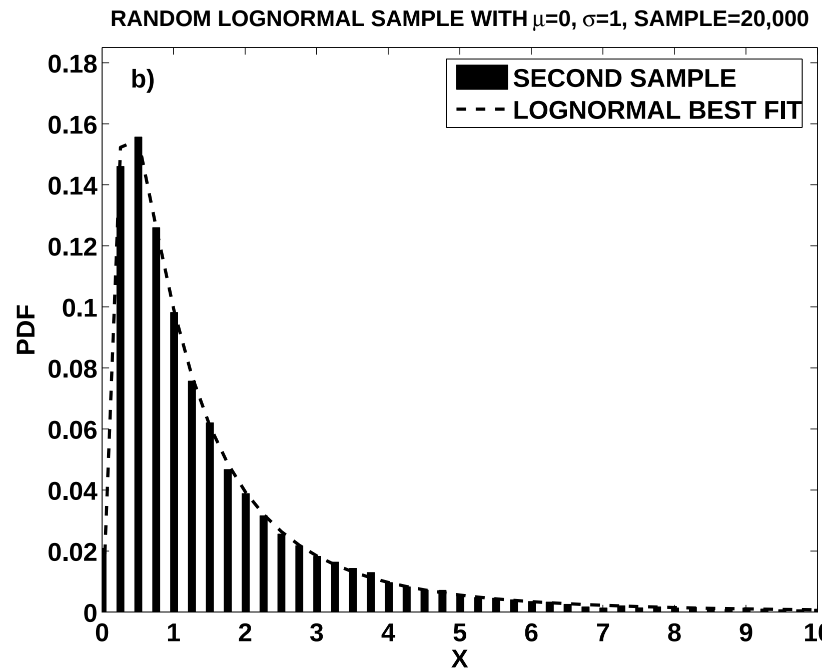

PLOT OF LOGNORMAL DISTRIBUTIONS
σ=0.25
σ=0.5
σ=1
σ=1.5
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
PDF
0
0.5
1
1.5
2
2.5
3
x


PLOT OF TRANSFORMED NORMAL DISTRIBUTIONS
PDF
ln x
σ=0.25
σ=0.5
σ=1
σ=1.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2

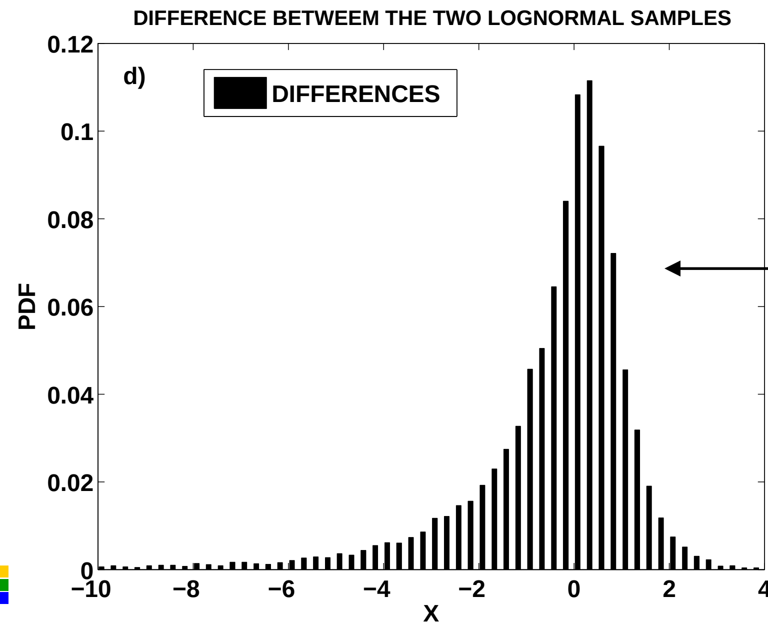


All skewness information is lost

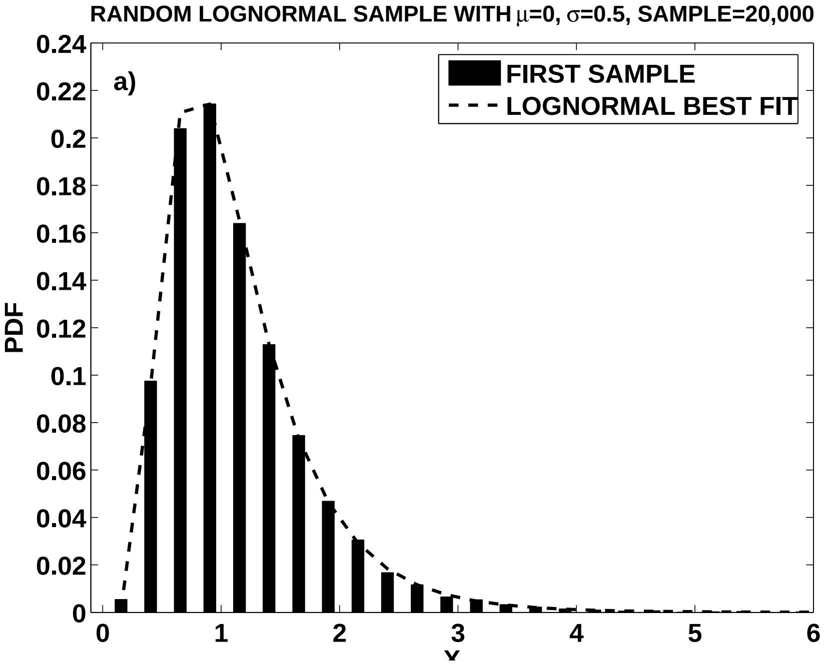


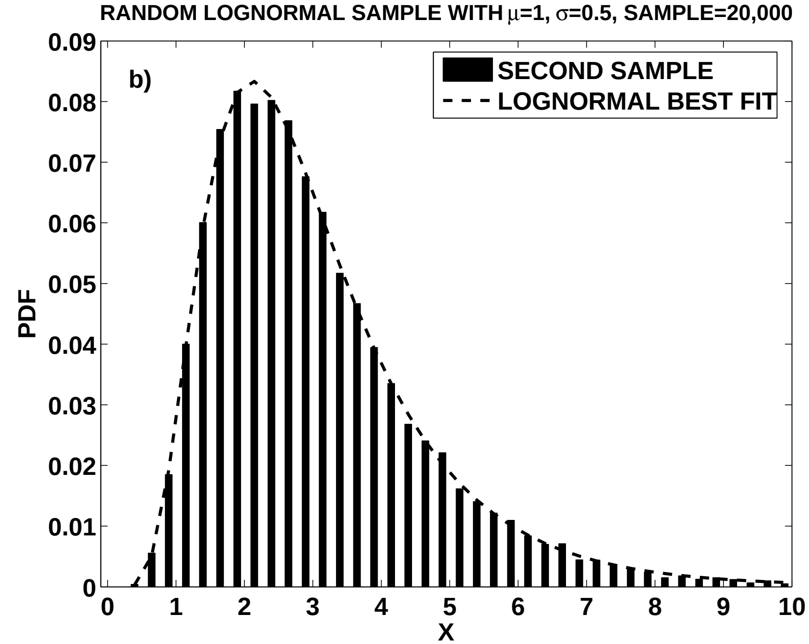


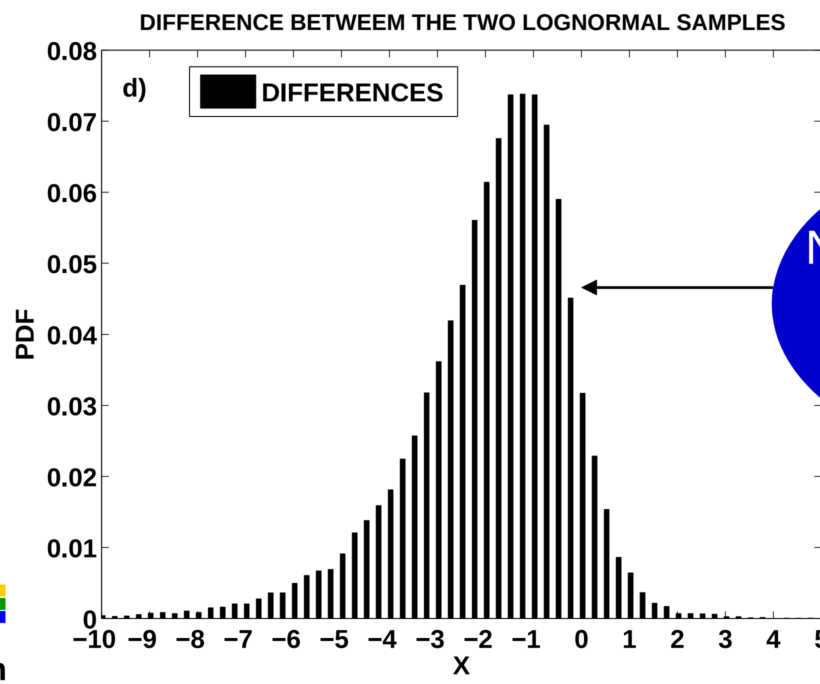
ASSUMED
GAUSSIAN
APPROACH



NOTE: DIFFERNCE
IS NOT
A GAUSSIAN







NOTE: DIFFERNCE IS NOT A GAUSSIAN

PROBLEMS ASSOCIATED WITH CURRENT TECHNIQUES

ASSUMED GAUSSIAN:

IMPACT 1: Wrong probabilities assigned to outliers.

IMPACT 2: Probabilities assigned to unphysical values.

IMPACT 3: Wrong statistic used to approximate the true distribution of the random variable.

MISCONCEPTIONS ABOUT LOGNORMAL DATA ASSIMILATION

1) The theory holds as the background solution is independent of the true solution, it is only an approximation and statistically has no information about the true solution.

2) The theory holds for the observational component as the observations are independent of the observations operator and vice-versa.

3) If two solutions have a relative error of 50% then we are still out by a factor of two in both cases no matter what order of magnitude.

4D LOGNORMAL DATA ASSIMILATION

Unlike with the three dimensional version of variational data assimilation, the four dimensional version is defined as a weighted least squares problem.

The Gaussian weighted least squares approach to 4D VAR is defined through a calculus of variation problem with initial conditions found through the adjoint.

This weighted least squares approach can be defined for a lognormal framework, which is defined by the following inner product

$$g_1(\boldsymbol{x_0}) = \iiint_A \sum_{i=1}^{N_o} \frac{1}{2} \left\langle \ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i\left(\boldsymbol{M}_i(\boldsymbol{x}_0)\right), \boldsymbol{R}_i^{-1}\left(\ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i\left(\boldsymbol{M}_i(\boldsymbol{x}_0)\right)\right) \right\rangle$$

As with the Gaussian case we know that the first variation of the functional defined on the previous slide is equivalent to

$$\delta g_1(\boldsymbol{x}_0) = \sum_{i=1}^{N_0} \left\langle \ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i(\boldsymbol{M}_i(\boldsymbol{x}_0)), \boldsymbol{W}_{o,i} \boldsymbol{H}_i \mathbf{M}_i \boldsymbol{R}_i^{-1} \delta \boldsymbol{x}_0 \right\rangle$$
$$= \left\langle \nabla g_1(\boldsymbol{x}_0), \delta \boldsymbol{x}_0 \right\rangle$$

Through using the properties of inner products we get that the gradient is

$$\nabla g_1(\boldsymbol{x}_0) = \sum_{i=1}^{N_o} \left(\boldsymbol{W}_{o,i} \boldsymbol{H}_i \mathbf{M}_i \boldsymbol{R}_i^{-1} \right)^T \left(\ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i(\boldsymbol{M}_i(\boldsymbol{x}_0)) \right)$$

The solution is a median and not the mode and hence is independent of the variance.

We need to define the functional as

$$g_2(\boldsymbol{x_0}) = \iiint_A \sum_{i=1}^{N_o} \frac{1}{2} \left\langle \ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i(\boldsymbol{M}_i(\boldsymbol{x}_0)) + \boldsymbol{R}^T \boldsymbol{1}, \boldsymbol{R}_i^{-1} (\ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i(\boldsymbol{M}_i(\boldsymbol{x}_0))) \right\rangle$$

Which then has a gradient of

$$\nabla g_2(\boldsymbol{x_0}) = \sum_{i=1}^{N_o} \left(\boldsymbol{W}_{o,i} \boldsymbol{H}_i \mathbf{M}_i \boldsymbol{R}_i^{-1} \right)^T \left(\ln \boldsymbol{y}_i - \ln \boldsymbol{h}_i(\boldsymbol{M}_i(\boldsymbol{x}_0)) + \boldsymbol{R}_i^T \boldsymbol{1} \right)$$

Current Gaussian approach

$$J_G\left(\mathbf{x}_0\right)=\sum_{i=1}^{N_0}\left\langle \mathbf{R}_{G,i}^{-1}\left(\mathbf{y}_i-\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right),\left(\mathbf{y}_i-\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right)\right\rangle$$

Improved Transform technique

$$J_{TR}\left(\mathbf{x}_0\right)=\sum_{i=1}^{N_0}\left\langle \mathbf{R}_{L,i}^{-1}\left(\ln\mathbf{y}_i-\ln\left(\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right)\right),\left(\ln\mathbf{y}_i-\ln\left(\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right)\right)\right\rangle$$

New Lognormal 4D VAR approach:

$$J_{LN}\left(\mathbf{x}_0\right)=\sum_{i=1}^{N_0}\left\langle \mathbf{R}_{L,i}^{-1}\left(\ln\mathbf{y}_i-\ln\left(\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right)+\mathbf{R}_{L,i}\mathbf{1}_{N_{0,1}}\right),\left(\ln\mathbf{y}_i-\ln\left(\mathbf{h}_i\left(M_{0,i}\left(\mathbf{x}_0\right)\right)\right)\right)\right\rangle$$

Term that gives the mode

PROBABILITY APPROACH

$$P\left(x_0, x_1, x_2, \ldots, x_{N_o} \middle| y_1, y_2, y_3, \ldots, y_{N_0}\right) =$$
$$P(x_0) P(x_1 | x_0) P(y_1 | x_1, x_0) P(x_2 | y_1, x_1, x_0)$$
$$\ldots P\left(y_{N_O} \middle| x_N, y_{N_o - 1}, x_{N_o}, \ldots, y_1, x_1, x_0\right)$$

The expression above is Bayes theorem for a multi-event probability situation. It can be simplified through using conditional independence.

$$P\left(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_o} \middle| \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{N_0}\right) = P(\mathbf{x}_0) \prod_{i=1}^{N_o} P(\mathbf{y}_i | \mathbf{x}_0)$$

Taking the negative logarithm of the circled pdf in the previous slide results in

$$\boxed{\min\left\{J\left(\mathbf{x}_0\right)=-\ln P\left(\mathbf{x}_0\right)-\sum_{i=1}^{N_0}\ln P\left(\mathbf{y}_i\middle|\mathbf{x}_0\right)\right\}}$$

This can now be used to derive a 4D VAR system for any distributed random variable

$$P\left(\boldsymbol{x_0},\boldsymbol{x_1},\boldsymbol{x_2},\ldots,\boldsymbol{x}_{N_o}\middle|\boldsymbol{y_1},\boldsymbol{y_2},\boldsymbol{y_3},\ldots,\boldsymbol{y}_{N_0}\right)=$$

$$P\left(\boldsymbol{x_0}\right)\prod_{i=1}^{N_o}P\left(\boldsymbol{y}_i\middle|\boldsymbol{x_0}\right)$$

For the multivariate Gaussian case we have

$$P\left(\boldsymbol{x_0}\right)\propto\exp\left\{-\frac{1}{2}\left(\boldsymbol{x_0}-\boldsymbol{x_{b,0}}\right)^T\boldsymbol{B}_0^{-1}\left(\boldsymbol{x_0}-\boldsymbol{x_{b,0}}\right)\right\}$$

$$P\left(\boldsymbol{y}_i\middle|\boldsymbol{x_0}\right)\propto\exp\left\{-\frac{1}{2}\left(\boldsymbol{y_i}-\boldsymbol{h}_i\left(\boldsymbol{M}_i\left(\boldsymbol{x_0}\right)\right)\right)^T\boldsymbol{R}_i^{-1}\left(\boldsymbol{y_i}-\boldsymbol{h}_i\left(\boldsymbol{M}_i\left(\boldsymbol{x_0}\right)\right)\right)\right\}$$

For the multivariate lognormal case we have

$$P(\boldsymbol{x_0}) \propto \left(\prod_{j=1}^{N} \frac{\boldsymbol{x_{0,i}}}{\boldsymbol{x_{b,0,j}}} \right) \times$$

$$\exp\left\{ -\frac{1}{2} \left(\ln \boldsymbol{x_0} - \ln \boldsymbol{x_{b,0}} \right)^T \boldsymbol{B}_0^{-1} \left(\ln \boldsymbol{x_0} - \ln \boldsymbol{x_{b,0}} \right) \right\}$$

$$P(\boldsymbol{y}_i | \boldsymbol{x_0}) \propto \left(\prod_{k=1}^{N_{o,i}} \frac{h_{i,k}\left(\boldsymbol{M}\left(\boldsymbol{x}_0\right)\right)}{y_{i,k}} \right) \times$$

$$\exp\left\{ -\frac{1}{2} \left(\boldsymbol{y_i} - \boldsymbol{h}_i\left(\boldsymbol{M}_i\left(\boldsymbol{x_0}\right)\right) \right)^T \boldsymbol{R}_i^{-1} \left(\boldsymbol{y_i} - \boldsymbol{h}_i\left(\boldsymbol{M}_i\left(\boldsymbol{x_0}\right)\right) \right) \right\}$$

Results with the Lorenz 1963 model

We are going to be using the hybrid Gaussian-lognormal distribution and comparing it with the transform approach. The associated cost function is

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{\varepsilon}_b^T - \boldsymbol{\varepsilon}_b\right)^T \boldsymbol{B}_0^{-1}\left(\boldsymbol{\varepsilon}_b^T - \boldsymbol{\varepsilon}_b\right) + \boldsymbol{\varepsilon}_b^T \begin{pmatrix} \boldsymbol{0}_{b,p} \\ \boldsymbol{1}_{b,q} \end{pmatrix} + \sum_{i=1}^{N_o}\left(\boldsymbol{\varepsilon}_{o,i}^T - \boldsymbol{\varepsilon}_{o,i}\right)^T \boldsymbol{R}_i^{-1}\left(\boldsymbol{\varepsilon}_{o,i}^T - \boldsymbol{\varepsilon}_{o,i}\right) + \sum_{i=1}^{N_o} \boldsymbol{\varepsilon}_{o,i}^T \begin{pmatrix} \boldsymbol{0}_{o,p,i} \\ \boldsymbol{1}_{o,q,i} \end{pmatrix}$$

$$\boldsymbol{\varepsilon}_b = \begin{pmatrix} \boldsymbol{x}_{p_1} - \boldsymbol{x}_{b,p} \\ \ln \boldsymbol{x}_{q_1} - \ln \boldsymbol{x}_{b,q} \end{pmatrix} \quad \boldsymbol{\varepsilon}_{\boldsymbol{o},i} = \begin{pmatrix} \boldsymbol{y}_{p,i} - \boldsymbol{h}_{p,i}(\boldsymbol{x}) \\ \ln \boldsymbol{y}_{q,i} - \ln \boldsymbol{h}_{q,i}(\boldsymbol{x}) \end{pmatrix}$$

Example with the Lorenz 1963 model

The three non-linear differential equations are given by (Lorenz 1963)

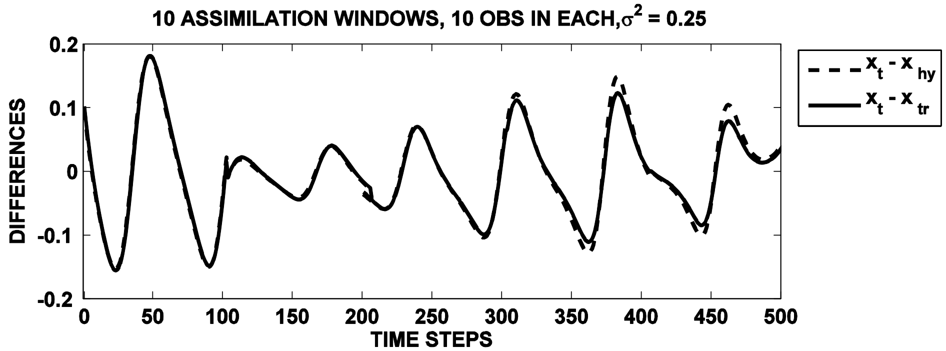
$$\dot{x} = -\sigma x + \sigma y$$
$$\dot{y} = -xz + \rho x - y$$
$$\dot{z} = xy - \beta z$$

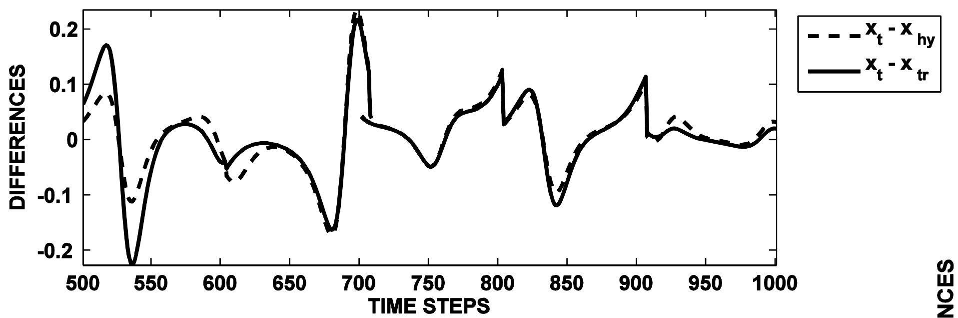
$$\beta = \frac{8}{3}, \quad \sigma = 10 \ \text{ AND } \ \rho = 28$$

$$x_0 = -5.4458, \quad y_0 = -5.4841 \ \text{ AND } \ z_0 = 22.5606$$

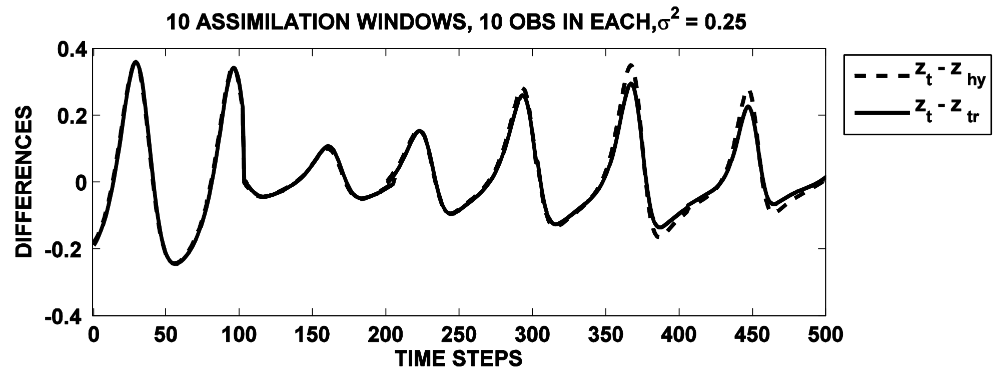
Going to assume x and y components and the associated obs are Gaussian, z is lognormal (Fletcher and Zupanski 2007)

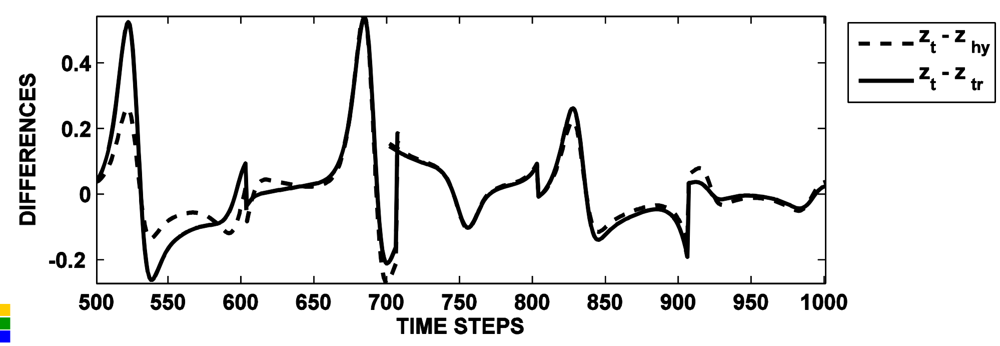
CG/AR





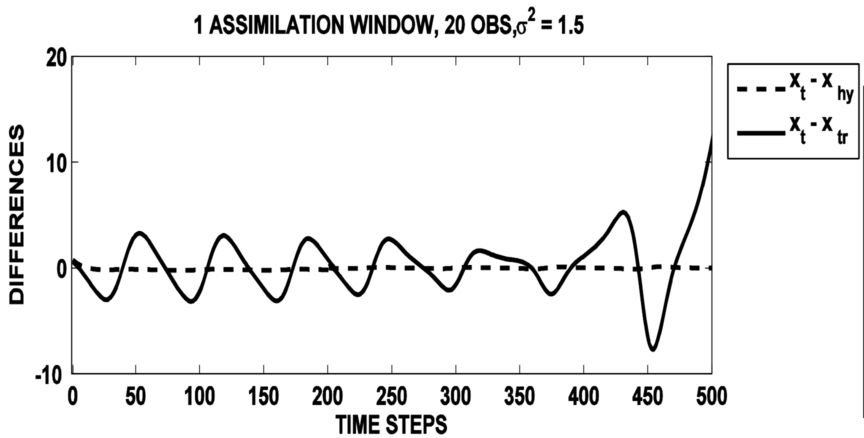
Plots of the differences in the trajectories with many accurate obs with short assimilation windows



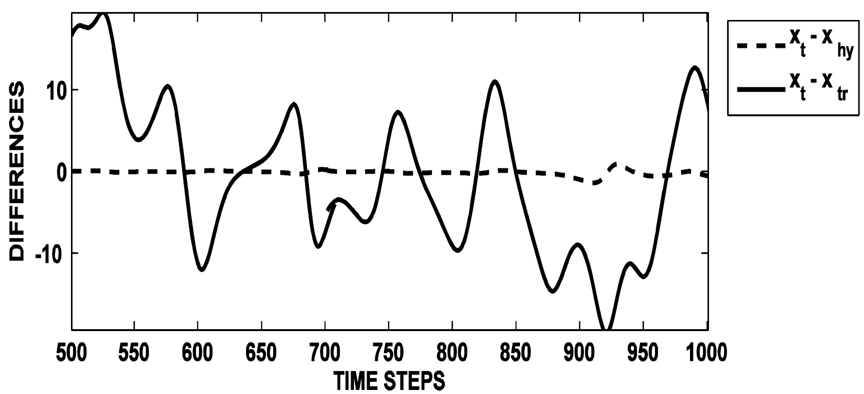


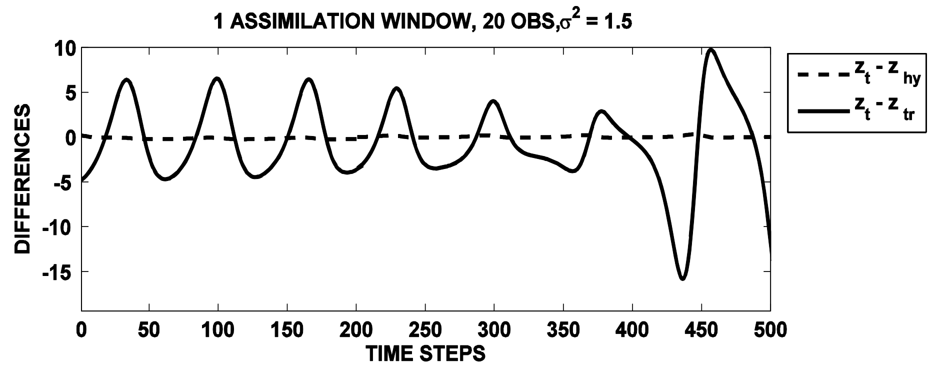
Note: transform and hybrid approach quite similar

BEST CASE SCENERIO

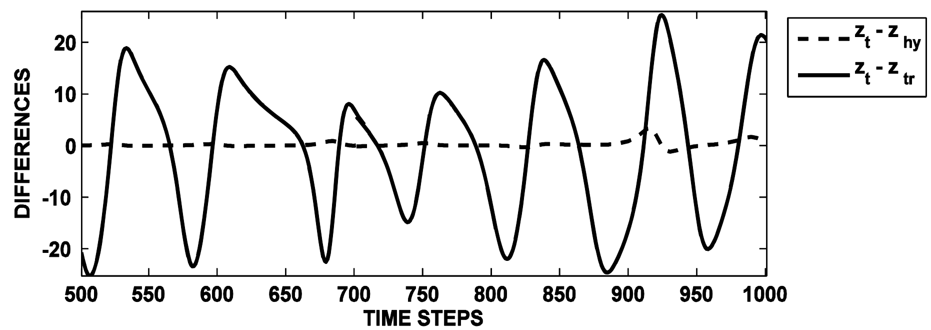


When assimilation window is large (1000ts) with fewer and less accurate obs, hybrid approach is more accurate





Transform approach converges quickly to the wrong solution



Conclusions and Further Work

- Careful which statistic to use to analyses
- Mode is closer to the true trajectory in the Lorenz 63 model
- Possible to assimilate variables of mixed types simultaneously
- Combine other distributions?? i.e. Gamma, Normal, Lognormal
- More consistent methods for finding positive definite variables
- No need to change background error covariance matrix (if already using the transform approach)