# Non-Gaussian Based Data Assimilation Lecture 1: Theory

Steven J. Fletcher
Cooperative Institute for Research in the Atmosphere
Colorado State University

Cooperative Institute for Research in the Atmosphere

# Overview of Lecture

- Motivation
- Evidence for non-Gaussian Behaviour
- Distributions and Descriptive Statistics
- Full Field Non-Gaussian Variational Data Assimilation.

# Motivation

Why should we consider going non-Gaussian?

In the atmosphere and ocean we have variables that are positive-definite. This means that they cannot obtain values less than or equal to zero.

Positive semi-definite variables can obtain the value of zero but not less than.

We therefore need an error structure and data assimilation systems that cannot allow our analysis variables to take an unphysical value.

# Motivation

Quite often it is said that just because the physical field may show non-Gaussian behaviour this does not mean that the errors will be non-Gaussian!!

It is well know that if you have two independent Gaussian distributed random variables (RV) then if we define a new RV that is the difference between the two Gaussian RVs then this is also a Gaussian RV.

However, we cannot say that if we assume that our error is Gaussian then our true state and background are also Gaussian distributed. – Well that is not quite true.

We shall now introduce moment generating functions which are important in "going back in the other direction"

# Motivation

**Moment Generating Functions (MGF):**

The definition for the moments of a PDF is given by

$$E[x] = \int_a^b x f(x) dx$$

The Moment Generating Function is defined by

$$m_x(t) = E[\exp\{tx\}] \equiv \int_a^b \exp\{tx\} f(x) dx$$

Where a and b are the lower and upper bounds of where the pdf is valid.

We now consider two very important properties of MGFs,

# Motivation

Let c and d be constants, and let $M_X(t)$ be the moment generating function for the random variable $X$, then the MGF for the random variable $Y = c + dX$ is

$$M_Y(t) = e^{ct} M_X(dt)$$

Let $X$ and $Y$ be independent random variables with MGFs $M_X(t)$ and $M_Y(t)$ then if we define a new random variable $Z = X + Y$ then the MGF of $Z$ is given by

$$M_Z(t) = E[e^{tZ}] = E\left[e^{t(X+Y)}\right] \equiv E[e^{tX}]E[e^{tY}] \equiv M_X(t)M_Y(t)$$

Therefore, the MGF of the sum (and subtraction through using the first property on the last page) of two random variables is the product of the MGF of the two random variables.

# Motivation

Given the properties of moment generating functions we now introduce the ***Uniqueness Theorem*** which is vital to understand why if we detect a non-Gaussian signal in one of our fields then it is quite difficult for our errors to be Gaussian.

Uniqueness Theorem: *Suppose that the two random variables $X$ and $Y$ have the moment generating functions given by $M_X(t)$ and $M_Y(t)$ respectively.* If $M_X(t) = M_Y(t)$ *for all values of* $t$, *then* $X$ *and* $Y$ *have the same distribution.*

# Motivation

So what does the uniqueness theorem mean to us in data assimilation? We start with defining the MGF for a Gaussian distribution:

$$M_G(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}.$$

If we assume that our errors are Gaussian distributed then they have a moment generating function as describe above.

If we have an error structure defined as $\varepsilon_b \equiv x^t - x_b$ then we require both sides to have Gaussian MGFs so that the properties of MGF earlier hold but also then the uniqueness theorem holds.

We assume that our errors are *unbiased* which for the Gaussian distribution means $\mu_{\varepsilon_b} = 0$.

# Motivation

And we are assuming that we have detected a Gaussian signal for the background field which is distributed $x_b \sim G\left(\mu_b, \sigma_b^2\right)$ and has an associated moment generating function, as does the errors. Therefore, by the uniqueness theorem we require both sides of the equation below to be equal

$$\exp\left\{0t + \frac{\sigma^2 t^2}{2}\right\} = M_{x_t}(t)\exp\left\{-\mu_b t + \frac{\sigma_b^2 t^2}{2}\right\}$$

Therefore, the question is can $M_{x_t}(t)$ be anything other than a Gaussian MGF?

Unfortunately yes it can. There are two possible answers for $M_{x_t}(t)$ so that the uniqueness theorem holds for the expression above:

$$M_{x_t}(t) = \exp\left\{\mu_b t + \frac{\sigma_t^2 t^2}{2}\right\} \quad or \quad M_{x_t}(t) = \exp\{\mu_b t\}$$

Where $\sigma^2 = \sigma_t^2 + \sigma_b^2$ or $\sigma^2 = \sigma_b^2$. What this is saying is that if a Gaussian distribution is detected for the background field and we have assumed Gaussian errors then the true state can only be Gaussian distributed or it can be a constant!!

# Motivation

Therefore, if we detect a lognormal signal for our background field and we assume Gaussian errors then the only way the uniqueness theorem can hold is if the true state is a Gaussian minus a lognormal!

This raises the bigger question: Why is your background distribution not at least the distribution type for the true field?

Another problem that arises when considering lognormal distributions is that there does not exist a MGF for this distribution. The reason being is that there exists another distribution which has the same moments as the lognormal distribution to infinity!

# Evidence for Non-Gaussian Behaviour

## Question 1

Synoptic scale: Are all the synoptic variables Gaussian? NO

Humidity: To assimilate this variable we have to use the logarithm of the variable (Polavarapu *et al* 2005). This indicates that this variable is LOGNORMAL, also evidence in Kliewer *et al.* (2016)

Wind component: Combined with the moisture flux then this variable is showing sign that the wind components may be LOGNORMAL (Raymond 1997). In Kliewer *et al* (2016) there is evidence that wind speed is non-Gaussian but not necessarily lognormal.

Comment: These variables are either positive definite or semi positive definite!!

# Evidence for Non-Gaussian Behaviour

## Question 1

Meso-scale: Are all meso-scale variables Gaussian? NO

In the paper by Miles *et al.* (2000) there is a large summary of cloud variables that are not Gaussian, specifically LOGNORMAL and GAMMA. As early as the 1970s rain and cloud variables had been identified as LOGNORMAL, Mielke *et al.* (1977)

# Evidence for Non-Gaussian Behaviour

## Question 2

Are all observations Gaussian? NO

Direct Observations: The variables which we have already mentioned are not Gaussian and therefore a direct observation of them is also not Gaussian.

Retrievals: By having to take the logarithm of the humidity to apply 1D Variational data assimilation (1D VAR), the retrieved field is not Gaussian. It is lognormal.

# Evidence for Non-Gaussian Behaviour

## Question 2

Optical Depth: In Stephens *et al.* (2002) the pdf of this variable is presented and is clearly showing a lognormal structure.

Infra-Red Flux Differences: From the same paper.

Cloud base height: In Sengupta *et al.* (2004) these observations shows signs of a lognormal structure.

Liquid water path: Same paper, shows a sharp positive skewness associated with a lognormal distribution with large variance.

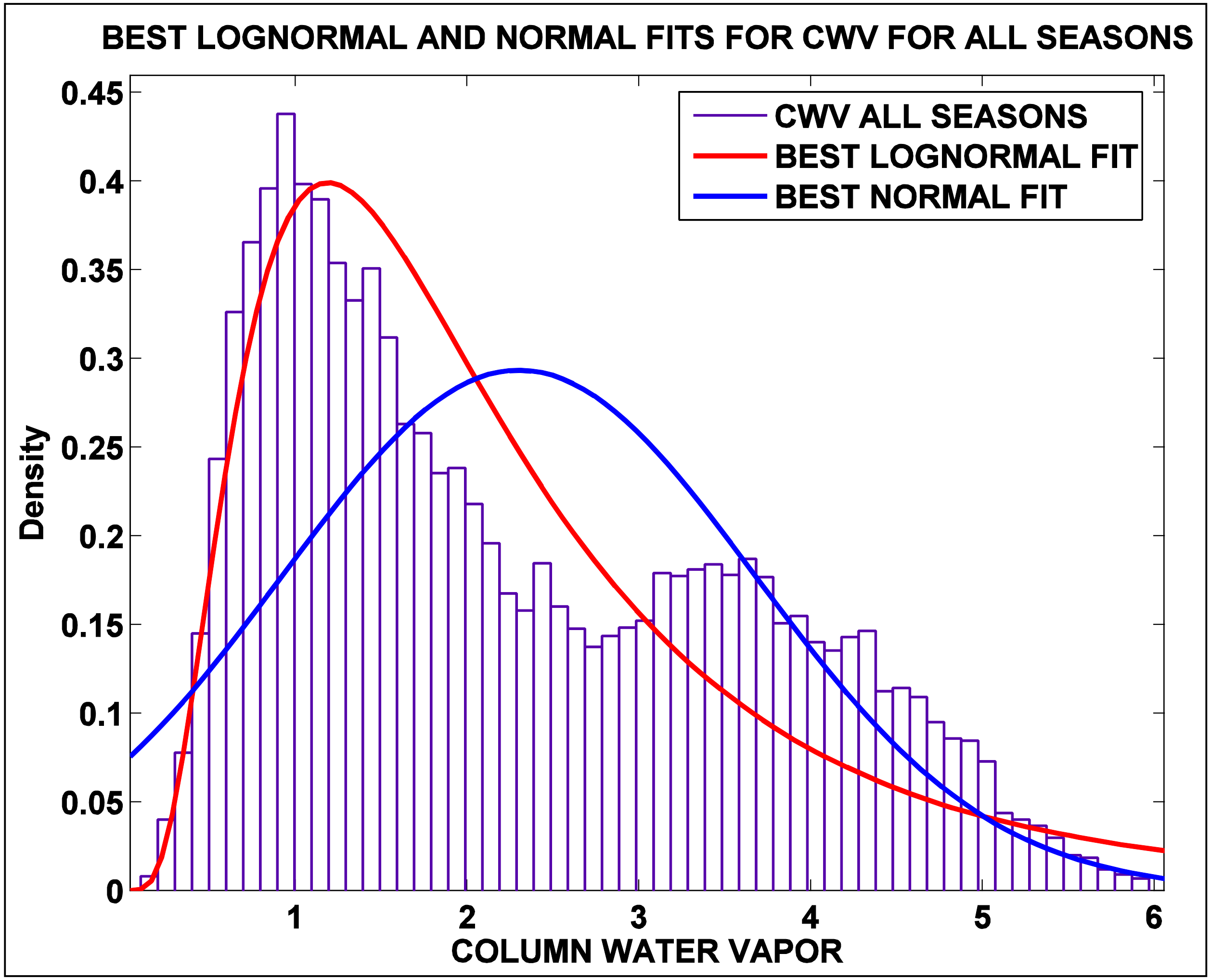
# Evidence for Non-Gaussian Behaviour

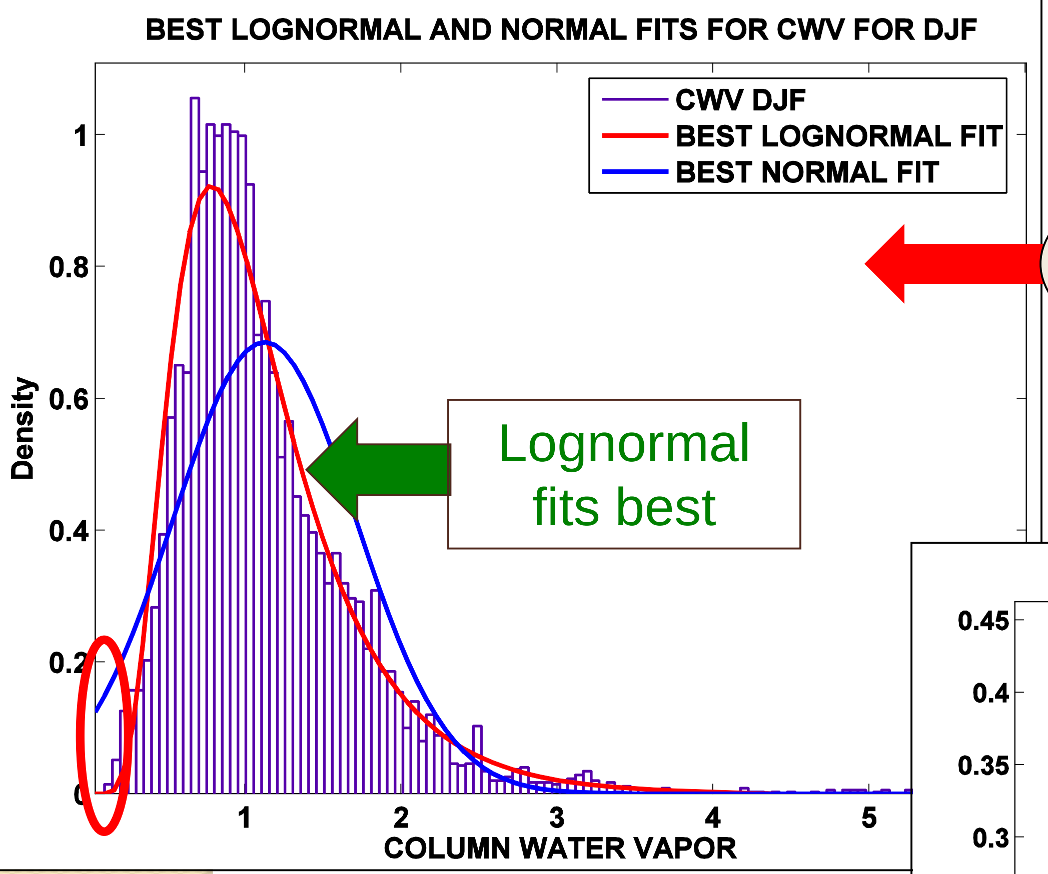
## Visual example

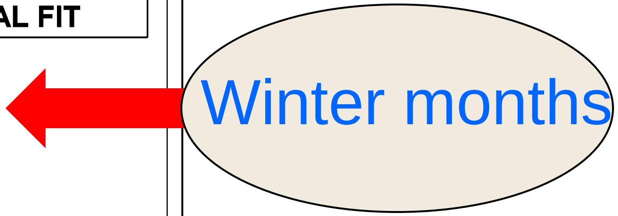
This data is column water vapour climatologies from the Oklahoma ARM-SGP site from 1997 – 2000 where the data are observed for days with boundary level clouds.

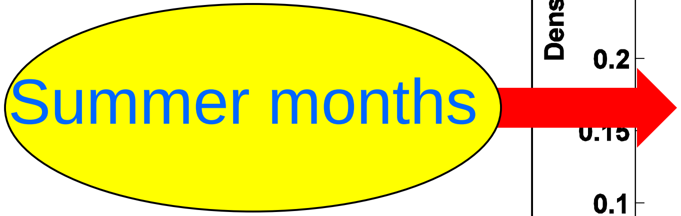
The data has been broken down by season as well as for the whole year. The data was collected from a microwave radiometer.

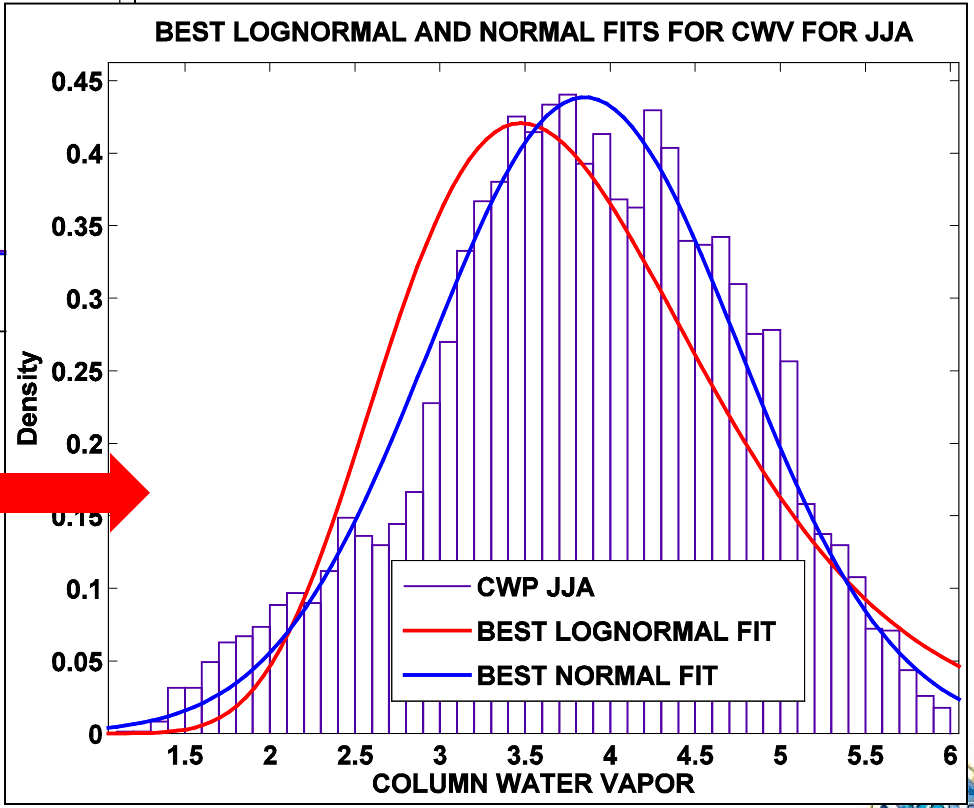
# Evidence for Non-Gaussian Behaviour




BEST LOGNORMAL AND NORMAL FITS FOR CWV FOR DJF
CWV DJF
BEST LOGNORMAL FIT
BEST NORMAL FIT
Density
COLUMN WATER VAPOR
Lognormal fits best


Winter months


Summer months


BEST LOGNORMAL AND NORMAL FITS FOR CWV FOR JJA
CWP JJA
BEST LOGNORMAL FIT
BEST NORMAL FIT
Density
COLUMN WATER VAPOR

## Distributions and Descriptive Statistics

We shall only be considering the continuous probability distributions for this lecture. The three descriptive statistics for continuous distributions are defined as

The mean, which is also the first moment of the distribution, is also referred to as the minimum variance estimator and is given by

$$\mu = E(x) = \int_a^b x f(x) dx, \qquad x \in [a, b]$$

# Distributions and Descriptive Statistics

The next statistic is the median. This is the unbiased estimator. For continuous distributions the median is the value of $x$ such that

$$\int_{a}^{x_{med}} f(z)dz = \frac{1}{2}.$$

The third and final descriptive statistic is the mode. This is the maximum likelihood estimator and is defined by

$$\left.\frac{df}{dx}\right|_{x=x_{mod}} = 0.$$

# Distributions and Descriptive Statistics

Question: How do the three statistic operators transfer to multivariate theory? Not very well!!!

Mean:

$$\boldsymbol{\mu} = E(\boldsymbol{x}) = \int_{\boldsymbol{a}}^{\boldsymbol{b}} \int_{\boldsymbol{c}}^{\boldsymbol{d}} \cdots \int_{\boldsymbol{e}}^{\boldsymbol{f}} \boldsymbol{x} f(\boldsymbol{x})\, d\boldsymbol{x},$$

However, this expression is not defined in mathematics, so for the multivariate situation we actually have the vector of means

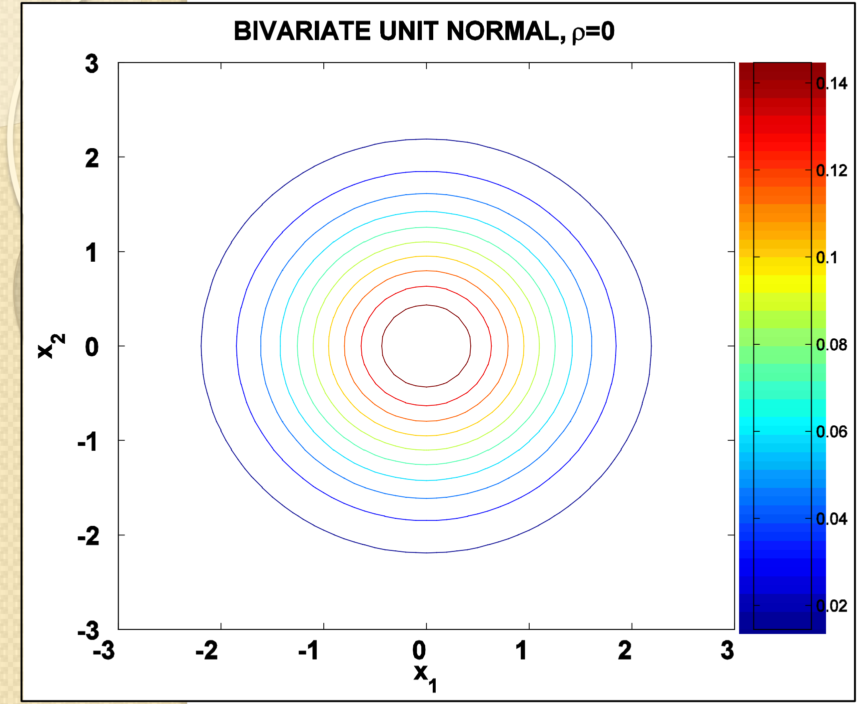
$$\mu_i = \int_{a_i}^{b_i} x_i f(x) dx_i, \qquad i = 1,2, \dots, N$$

## Multivariate Medians

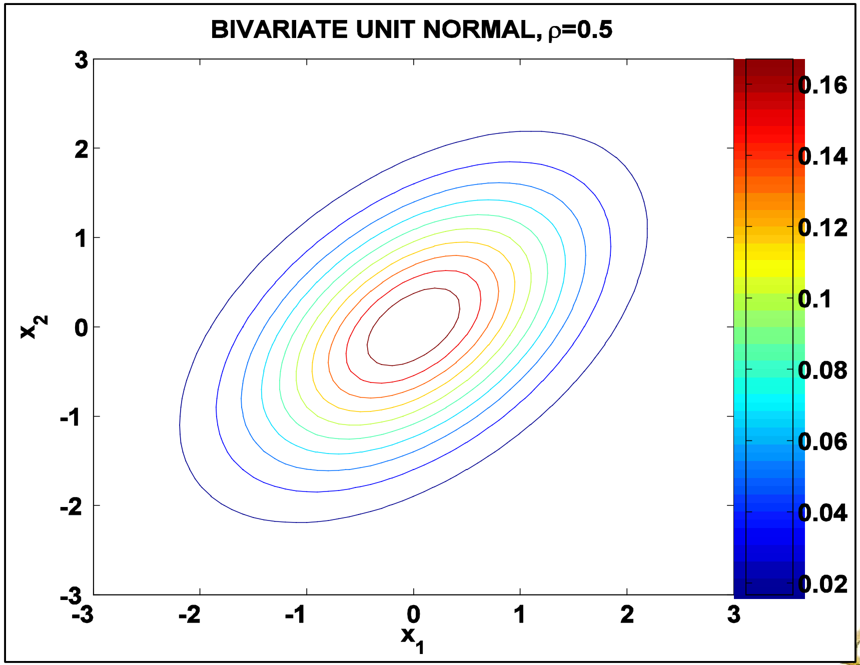
This is the worst of the three statistics. Although it is *unbiased* it is also non-unique, even for the multivariate Gaussian distribution. The multivariate definition is

$$\boldsymbol{x_{med}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$such\ that\ \int_{a_1}^{x_1}\int_{a_2}^{x_2}\cdots\int_{a_n}^{x_n} f(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{2}.$$



BIVARIATE UNIT NORMAL, ρ=0
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
x1
x2
0.14
0.12
0.1
0.08
0.06
0.04
0.02

BIVARIATE UNIT NORMAL, ρ=0.5
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
x1
x2
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02

## Distributions and Descriptive Statistics

Is all hope lost? **No**

Our saviour is the mode!

There is a simple multivariate extension of the definition of the mode from the univariate case to the multivariate.

$$\left.\frac{df}{d\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_{\boldsymbol{mod}}} = \mathbf{0}$$

Another advantage of the mode for a uni-modal distribution is that it is unique!

# Distributions and Descriptive Statistics

We now start to consider a non-Gaussian distribution. The first is the lognormal on its own. The univariate lognormal distribution is defined as

$$LN(\mu_L, \sigma_L^2) \equiv \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}\exp\left\{-\frac{1}{2}\left(\frac{(\ln x - \mu_L)\hat{}2}{\sigma_L^2}\right)\right\}$$

where $x \in (0, \infty)$, and $\mu_L \equiv E[\ln x]$and $\sigma_L^2 \equiv E[(\ln x - E[\ln x])^2]$.

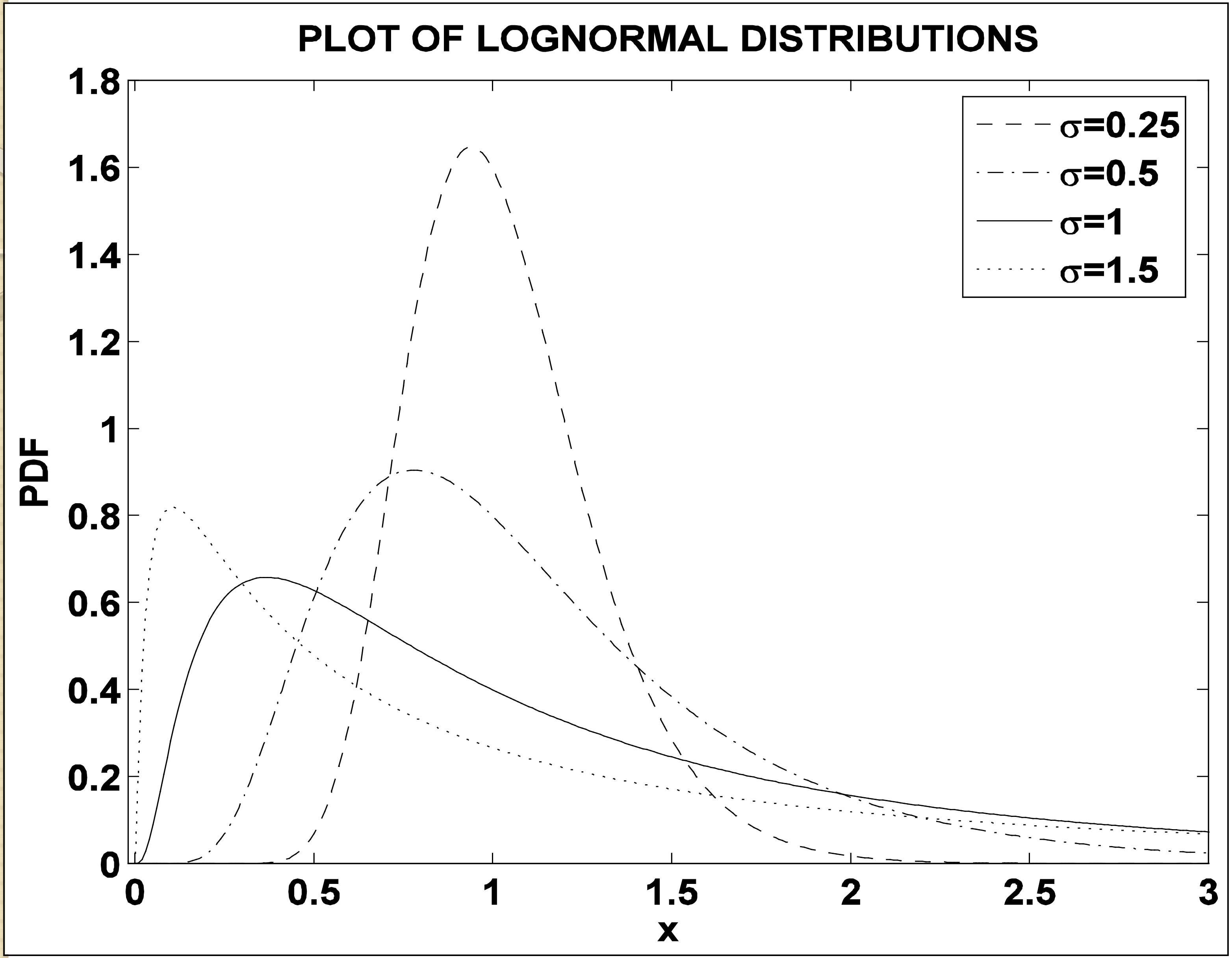
It can easily be shown that the mode, median and mean are

$$x_{mode} = \exp\{\mu_L - \sigma_L^2\}$$

$$x_{median} = \exp\{\mu_L\}$$

$$x_{mean} = \exp\left\{\mu_L + \frac{\sigma_L^2}{2}\right\}$$

As we can see the mode is the only statistic that is degenerate with respect to the variance. It is because of this property that the lognormal variational approach was adapted for non-Gaussian pdfs.


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

## Distributions and Descriptive Statistics

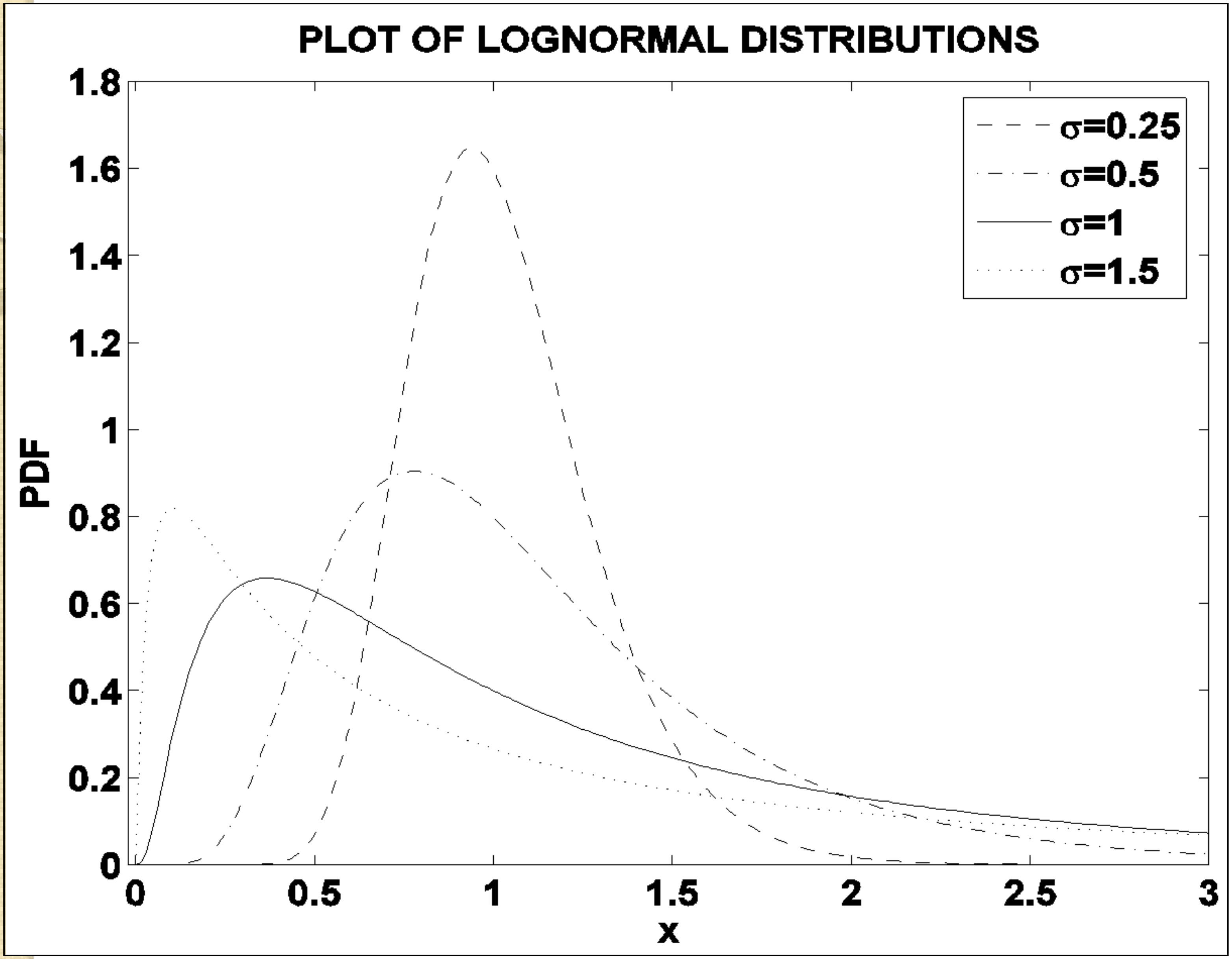
The multivariate lognormal distribution is defined as

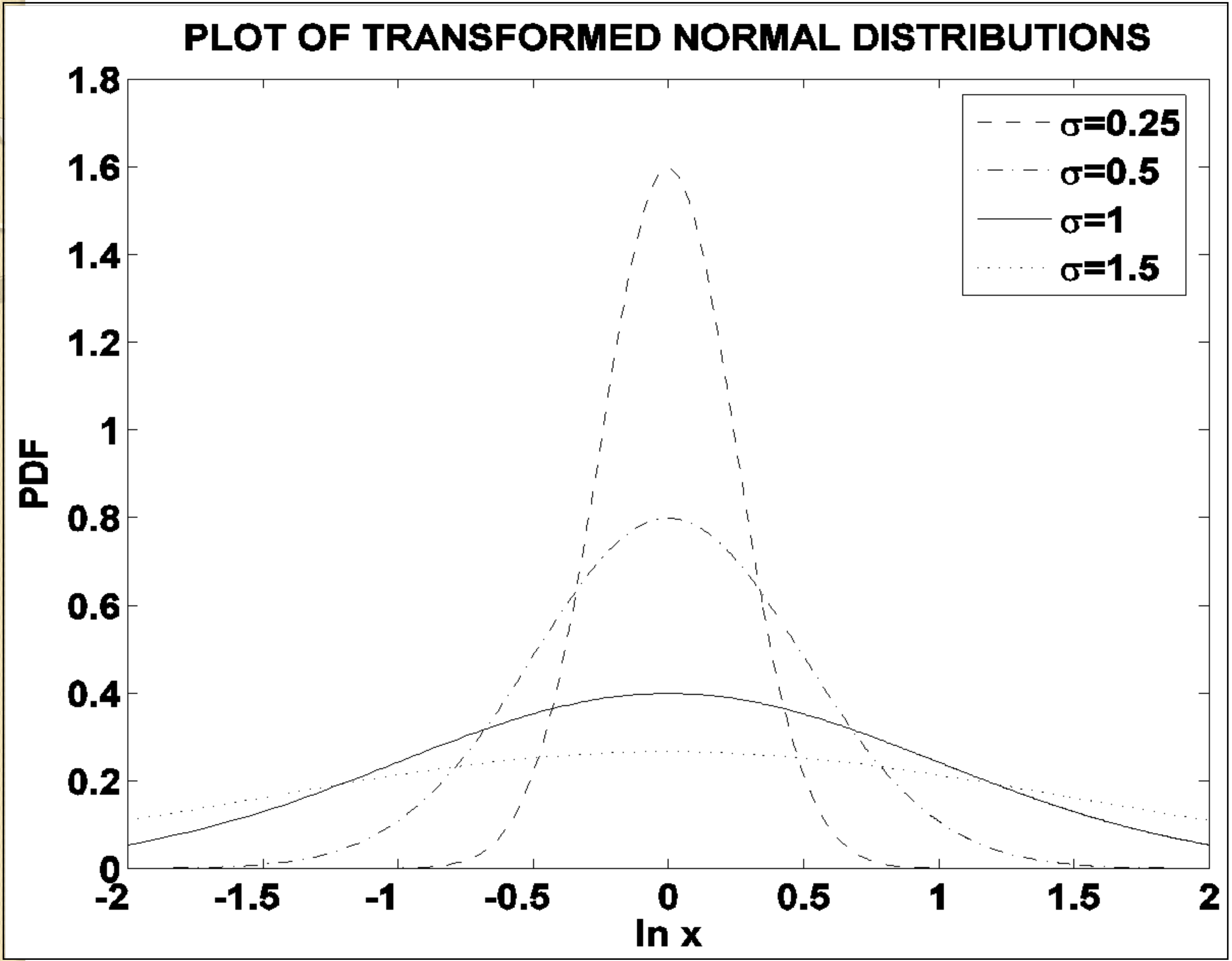
$$MLN(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \equiv \frac{1}{|\boldsymbol{\Sigma}_l|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}}\left(\prod_{i=1}^{n}\frac{1}{x_i}\right)\exp\left\{-\frac{1}{2}(\ln\boldsymbol{x}-\boldsymbol{\mu}_l)^T\boldsymbol{\Sigma}_l^{-1}(\ln\boldsymbol{x}-\boldsymbol{\mu}_l)\right\},$$

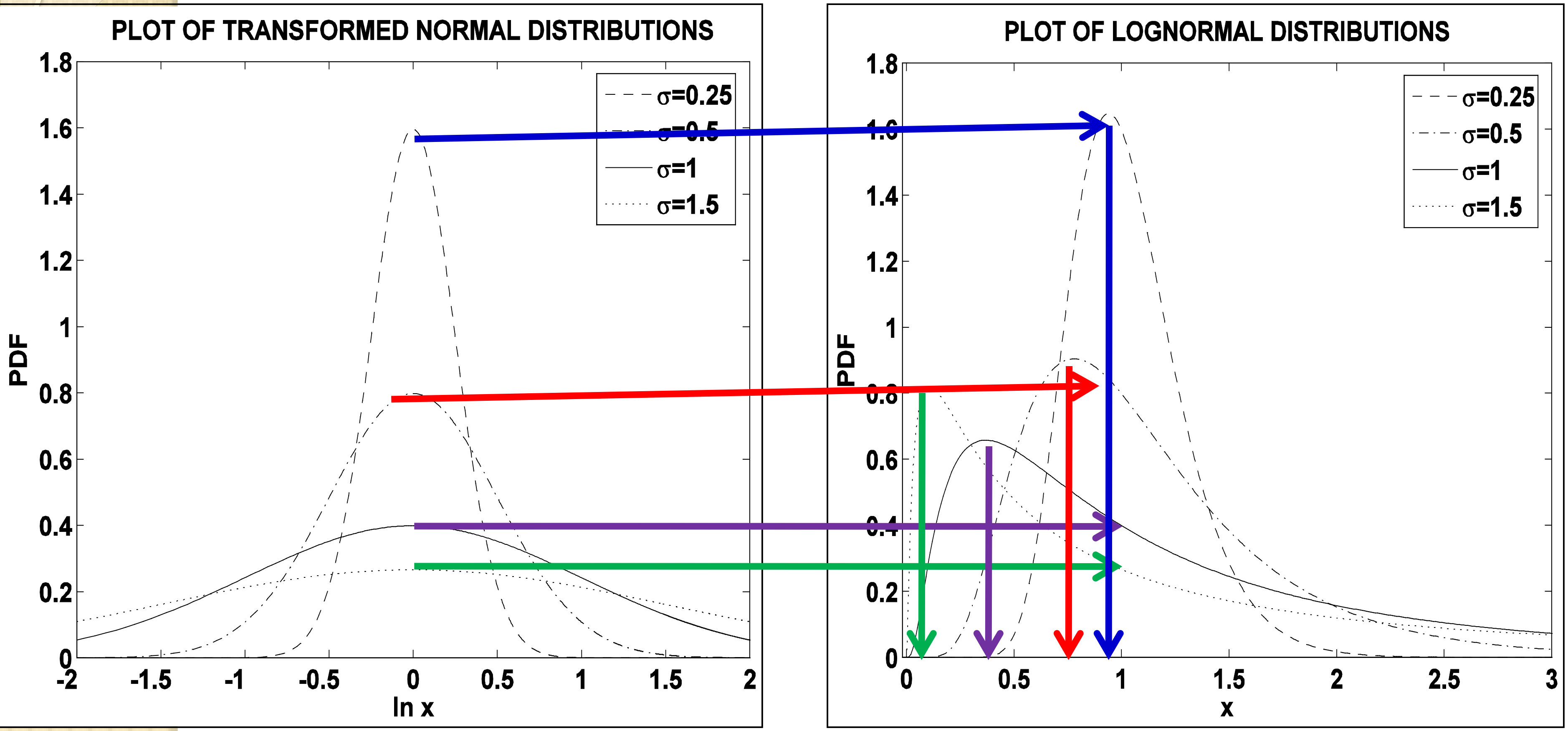
Where the multivariate lognormal mode and median are defined as

$$\boldsymbol{x}_{\boldsymbol{mode}} = \exp\{\boldsymbol{\mu}_{\boldsymbol{L}} - \langle\boldsymbol{\Sigma}_{\boldsymbol{L}}, \mathbf{1}\rangle\}$$
$$\boldsymbol{x}_{\boldsymbol{median}} = \exp\{\boldsymbol{\mu}_{\boldsymbol{L}}\}$$

An important property of the lognormal distribution is that if $\boldsymbol{x} \sim \boldsymbol{MLN} \Rightarrow \ln\boldsymbol{x} \sim \boldsymbol{MG}$. This transform has been used and is still currently used in some DA systems, however, it has a problem. The mode that is found in Gaussian transformed spaces can only invert back to the median in lognormal space.


PLOT OF LOGNORMAL DISTRIBUTIONS
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
σ=0.25
σ=0.5
σ=1
σ=1.5


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

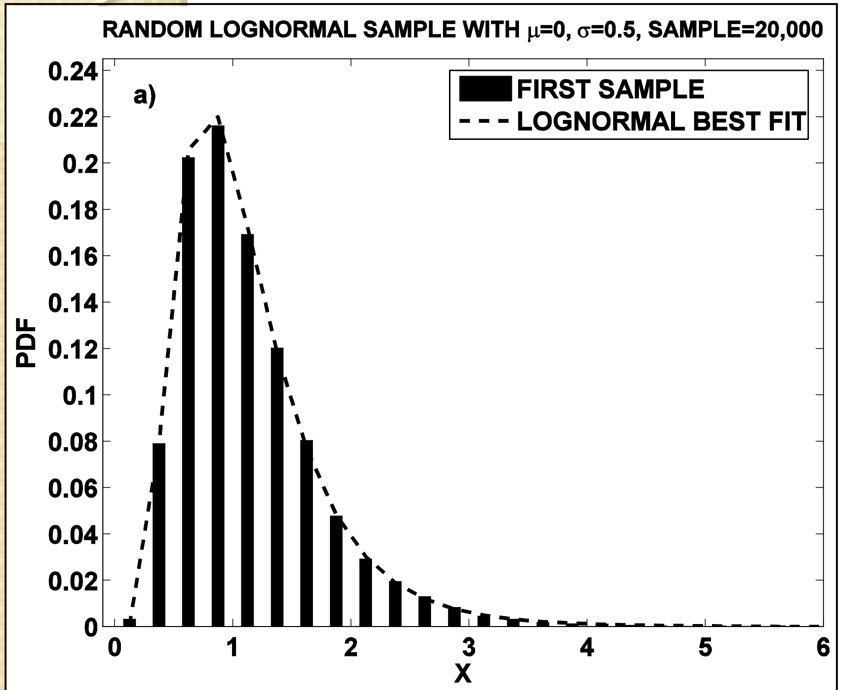


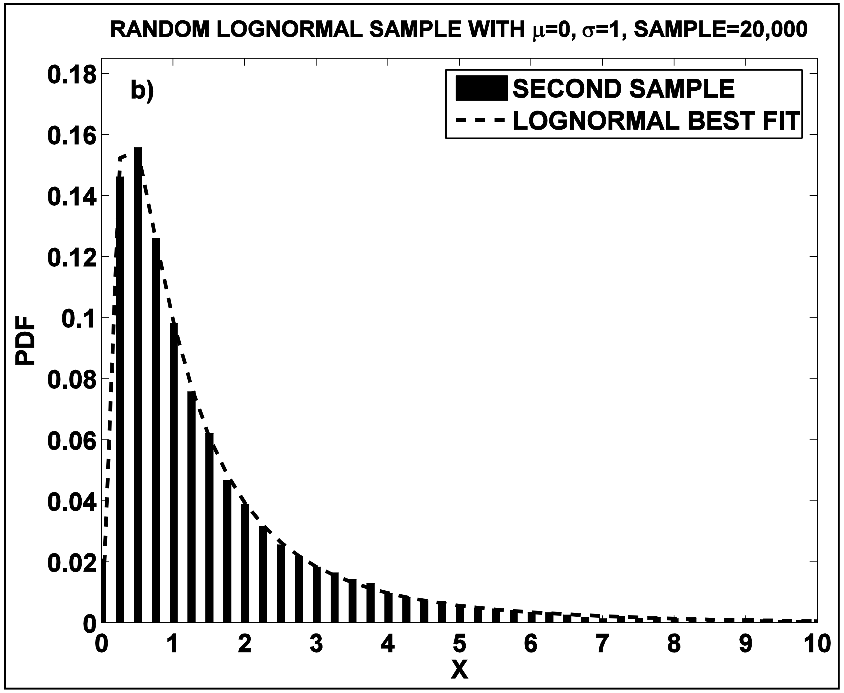

**All skewness information is lost**

# Full Field Non-Gaussian Variational Data Assimilation

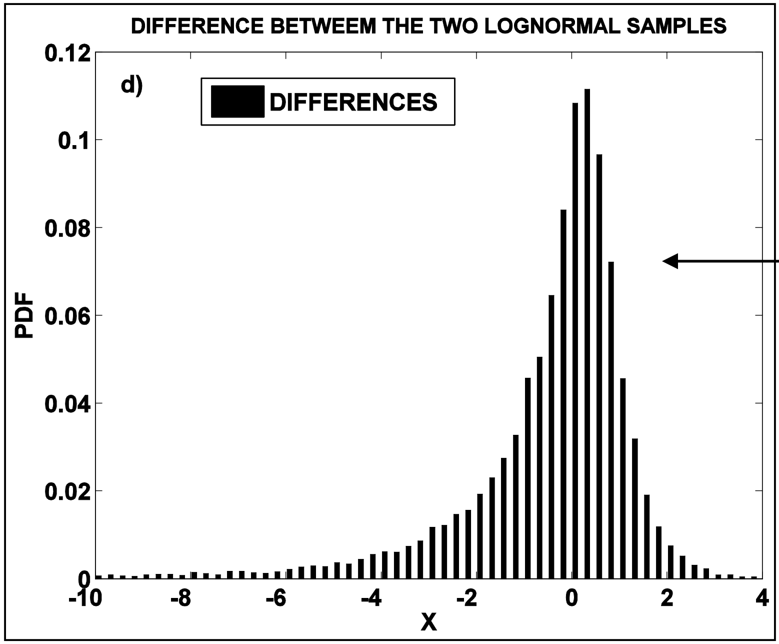
The starting point for lognormal based data assimilation is the definition of the errors. This was first proposed in Cohn (1997) where, due to the geometric behaviour of the lognormal distribution the errors could not be defined as the difference between two variables.

NOTE: there is no known distribution of the difference between two lognormal random variables. It is known NOT to be a Gaussian distribution or a lognormal distribution.




**ASSUMED GAUSSIAN APPROACH**




NOTE: DIFFERNCE IS NOT A GAUSSIAN

# Full Field Non-Gaussian Variational Data Assimilation

The basis for data assimilation is the definition of the errors. As the lognormal distribution is geometric, the additive property from Gaussian distribution theory does not hold. However, there is a geometric property that the ratio, and product, of two independent lognormally distributed random variables is also a lognormally distributed random variable.

Therefore we can define the background and observational errors geometrically as

$$\varepsilon_{b_i} \equiv \frac{x_i^t}{x_i^b}, i = 1,2,\ldots,N,$$

$$\varepsilon_j^o \equiv \frac{y_j}{h_j(\boldsymbol{x})}, j = 1,2,\ldots,N_o$$

Fletcher, S. J. and M. Zupanski, 2006a: A data assimilation method for lognormally distributed observational errors. *Q. J. Roy. Meteorol. Soc.*, **132,** 2505—2519.

# Full Field Non-Gaussian Variational Data Assimilation

The basis of 3D variational data assimilation (VAR DA) is Bayes theorem which is given by

$$P(A|B) \propto P(B|A)P(A)$$

Where the events $A$ is that $x_b = \boldsymbol{x_t}$ and the event is that $\boldsymbol{y^o} = \boldsymbol{y^t}$ and that combined into the conditional pdf above results in

$$P(\boldsymbol{x_b} = \boldsymbol{x_t} | \boldsymbol{y^o} = \boldsymbol{y^t}) \propto P(\boldsymbol{y^o} = \boldsymbol{y^t} | \boldsymbol{x_b} = \boldsymbol{x_t}) P(\boldsymbol{x_b} = \boldsymbol{x_t})$$

The next step is to form the maximum likelihood estimate by taking the negative logarithm of the expression above.

To find the cost function to minimize we have to substitute the definition of a multivariate lognormal probability density function along with the error definitions on the slide before which results in

$$J_L(\boldsymbol{x}) = \frac{1}{2}(\ln \boldsymbol{x_t} - \ln \boldsymbol{x_b})^T \boldsymbol{B_L^{-1}} (\ln \boldsymbol{x_t} - \ln \boldsymbol{x_b}) + \langle (\ln \boldsymbol{x_t} - \ln \boldsymbol{x_b}), \boldsymbol{1_N} \rangle + \frac{1}{2}(\ln \boldsymbol{y} - \ln \boldsymbol{h}(\boldsymbol{x}))^T \boldsymbol{R_L^{-1}} (\ln \boldsymbol{y} - \ln \boldsymbol{h}(\boldsymbol{x})) + \left\langle (\ln \boldsymbol{y} - \ln \boldsymbol{h}(\boldsymbol{x})), \boldsymbol{1_{N_o}} \right\rangle.$$

Fletcher, S. J. and M. Zupanski, 2007: Implications and impacts of transforming lognormal variables into normal variables in VAR. *Meteor. Zeit.*, **16**, 755—765.

## Full Field Non-Gaussian Variational Data Assimilation

However, we do not live in a one distribution fits all, and it is well known that there are many synoptic fields that are Gaussian distributed. In Fletcher and Zupanski (2006b) a new mixed Gaussian-lognormal distribution is proposed and proved.

We first consider the bivariate mixed distribution where we have one Gaussian distributed random variable and one lognormally distributed random variable. Therefore, the bivariate mixed distribution is given by

$$MX(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\mathbf{mx}}) \equiv \frac{1}{|\boldsymbol{\Sigma}_{\boldsymbol{mx}}|^{\frac{1}{2}} 2\pi} \frac{1}{x_2} \exp\left\{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \rho \frac{(x_1-\mu_1)}{\sigma_1} \frac{(\ln x_2 - \mu_2)}{\sigma_2}\right.\right.$$

# Mixed Gaussian-Lognormal distribution

The mixed distribution in its bivariate formulation is defined by

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

Where

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

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 $\ln X_2$ .

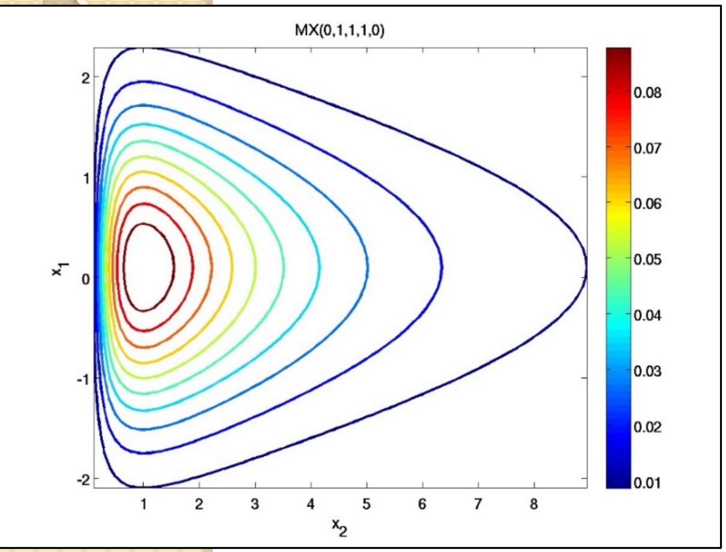
# Properties of the Mixed Distribution

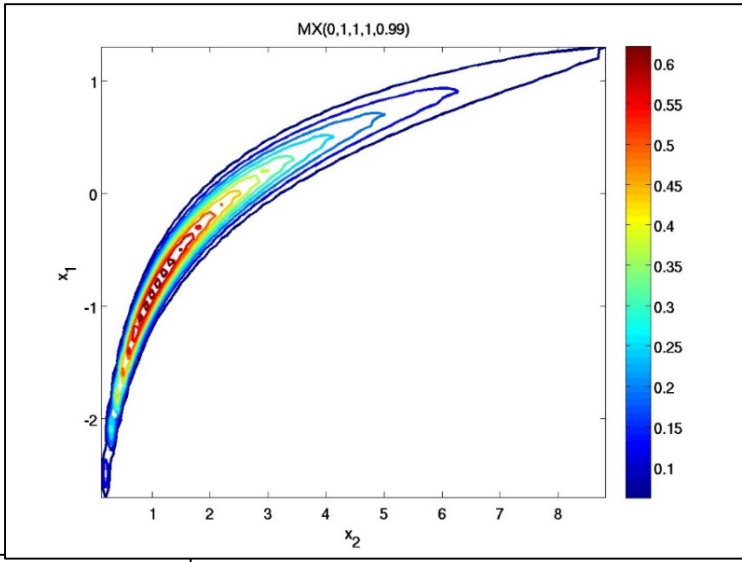
An important property of the mixed distribution is the definitions of the three descriptive statistics. The mean for each component can be found through forming the marginal and joint pdfs which can be shown to be Gaussian and lognormal, or vice-versa (Fletcher, 2017). Therefore the mean, mode and median are given by

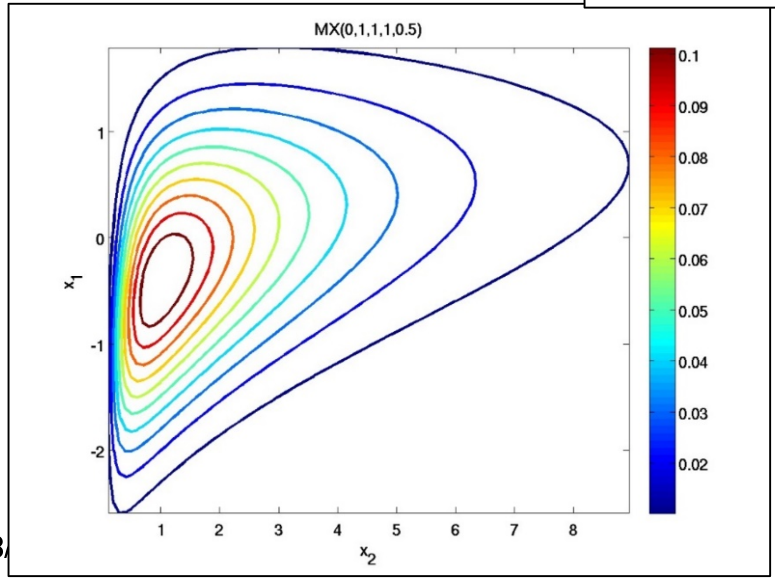
$$mean \equiv \begin{pmatrix} \mu_G \\ \exp\left\{\mu_L + \frac{\sigma_L^2}{2}\right\} \end{pmatrix}, \qquad median \equiv \begin{pmatrix} \mu_G \\ \exp\{\mu_L\} \end{pmatrix},$$

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

# Plots of the Mixed Distribution

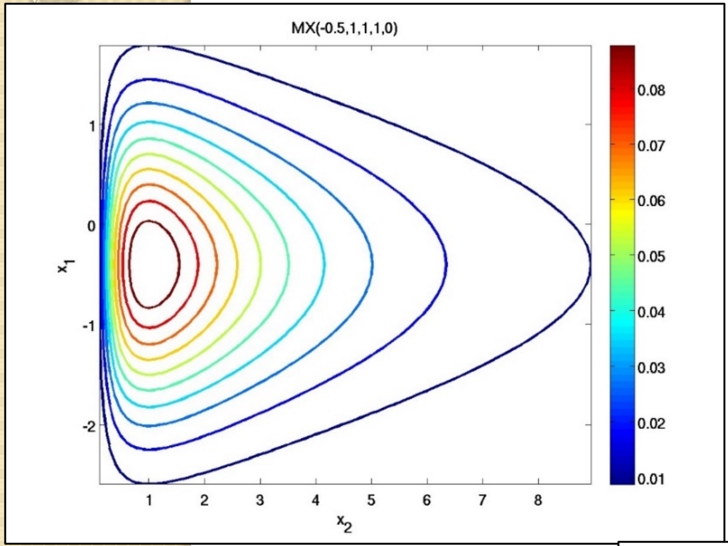


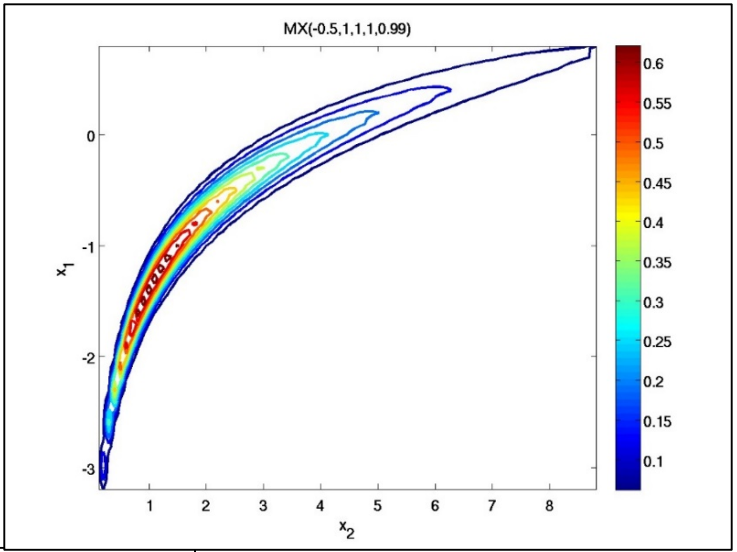



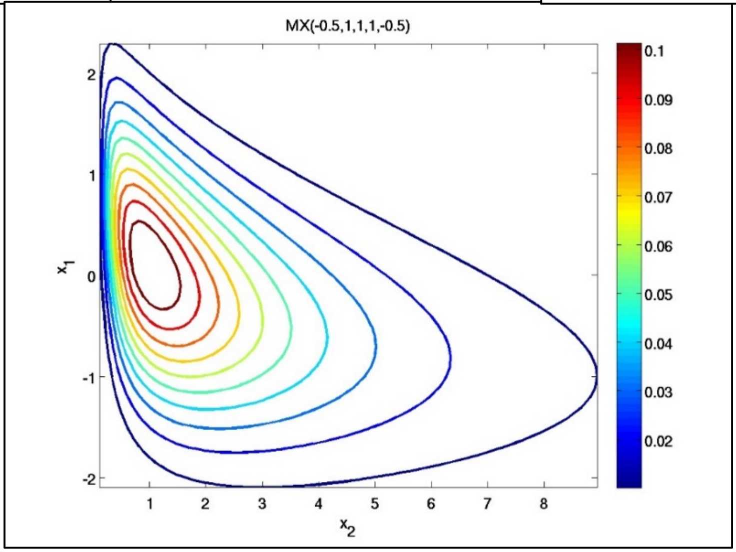




Cooperative Institute for Research in the Atmosphere

# 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

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

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

# Full Field Non-Gaussian Variational Data Assimilation

We have just seen from the figures on the previous slides that the mixed distribution appears to keep properties from both distributions, the symmetry about the mode but yet also the ability to have outliers assign probabilities that a Gaussian would miss.

The multivariate version of the mixed distribution starts with assuming that there are $p$ Gaussian random variables and $q$ lognormal random variables such that $N = p + q$. Therefore, the definition of the mixed distribution is

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

# Full Field Non-Gaussian Variational Data Assimilation

Given the definition for the multivariate mixed Gaussian-lognormal distribution we now consider the extension to the error definitions that would be needed to apply this distribution towards a 3D VAR format.

The definitions for mixed background and mixed observation errors where there are $p_1$ Gaussian background errors, $q_1$ lognormal background errors, $p_2$ Gaussian observational errors, and $q_2$ lognormal observational errors, are

$$\boldsymbol{\varepsilon_b} \equiv \begin{pmatrix} \boldsymbol{x_{p_1}^t - x_{p_1}^b} \\ \dfrac{\boldsymbol{x_{q_1}^t}}{\boldsymbol{x_{q_1}^b}} \end{pmatrix}, \qquad \boldsymbol{\varepsilon_o} \equiv \begin{pmatrix} \boldsymbol{y_{p_2} - h_{p_2}(x)} \\ \dfrac{\boldsymbol{y_{q_2}}}{\boldsymbol{h_{q_2}(x)}} \end{pmatrix}$$

Therefore, following the maximum likelihood route that the Gaussian and the lognormal based costs functions were derived from results in the mixed distribution based 3D VAR cost function being

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

# Full Field Non-Gaussian Variational Data Assimilation

Originally 4D VAR was derived from a variational prospective where a predetermined weighted least squares functional was defined to be minimised that had a Gaussian formulation to it.

In Fletcher (2010) this approach was extended to define a functional who's solution was the mode of a multivariate lognormal and then a mixed Gaussian-lognormal distribution. However, a more general probability model was sought so that in the future we could substitute any PDF into this Bayesian expression to obtain the cost function for the maximum likelihood estimator.

We start with what is referred to as the multi-event version of Bayes Theorem

$$P\left(\boldsymbol{x_0} \middle| \boldsymbol{x_N}, \boldsymbol{x_{N-1}}, \dots, \boldsymbol{x_1}, \boldsymbol{y_{n_o}}, \boldsymbol{y_{n_o-1}}, \dots, \boldsymbol{y_1}\right)$$
$$\propto P(\boldsymbol{x_0})P\left(\boldsymbol{x_1} \middle| \boldsymbol{x_0}\right)P\left(\boldsymbol{y_1} \middle| \boldsymbol{x_2}, \boldsymbol{x_1}, \boldsymbol{x_0}\right)P(\boldsymbol{x_2}|\boldsymbol{x_1}, \boldsymbol{x_0}) \cdots$$
$$P\left(\boldsymbol{x_N} \middle| \boldsymbol{x_{N-1}}, \boldsymbol{x_{N-2}}, \dots, \boldsymbol{y_p}, \boldsymbol{x_k}, \dots, \boldsymbol{y_1}, \boldsymbol{x_2}, \boldsymbol{x_1}, \boldsymbol{x_0}\right)$$

Whilst this expression looks terrifying to evaluate we are able to simplify through some assumptions that are made in DA.

## Full Field Non-Gaussian Variational Data Assimilation

Assumptions:

1. That the observation error is only dependent on the model state at that time.
2. That the model error is only dependent on the previous model evaluation – a one step Markov chain.
3. Model states are not dependent on observations.
4. Perfect model implies that if the initial conditions are true then all future states are also true.

# Full Field Non-Gaussian Variational Data Assimilation

Therefore, we now have two probability models – one for the perfect model (strong constraint) and one for the imperfect model (weak constraint) that are not dependent on any distribution choice.

**Strong Constraint:**

$$P\left(\boldsymbol{X_0}\middle|\widehat{\boldsymbol{X_N}}\right) \propto P(\boldsymbol{X_0})\prod_{\boldsymbol{i=1}}^{\boldsymbol{N_o}} P(\boldsymbol{Y_i}|\boldsymbol{X_k})$$

Where $\widehat{\boldsymbol{X_N}} \equiv \boldsymbol{X_{N-1}}, \dots, \boldsymbol{Y_p}\boldsymbol{X_k}, \dots, \boldsymbol{Y_1}, \boldsymbol{X_1}, \boldsymbol{X_0}$ and $k$ is the equivalent time index of the ith observation.

**Weak Constraint:**

$$P\left(\boldsymbol{X_0}\middle|\widehat{\boldsymbol{X_N}}\right) \propto P(\boldsymbol{X_0})\prod_{\boldsymbol{i=1}}^{\boldsymbol{N_o}} P(\boldsymbol{Y_i}|\boldsymbol{X_k})\prod_{\boldsymbol{j=1}}^{\boldsymbol{N}} P\left(\boldsymbol{X_j}\middle|\boldsymbol{X_{j-1}}\right)$$

To obtain the cost function we must take the negative logarithm of these two expressions. Therefore, the generalised cost functions for strong and weak constraint 4D VAR are:

# Full Field Non-Gaussian Variational Data Assimilation

Strong Constraint:

$$J(\boldsymbol{X_0}) = -\ln P(\boldsymbol{X_0}) - \sum_{i=1}^{N_O} \ln P(\boldsymbol{Y_i}|\boldsymbol{X_k})$$

Weak Constraint:

$$J(\boldsymbol{X_0}) = -\ln P(\boldsymbol{X_0}) - \sum_{i=1}^{N_O} \ln P(\boldsymbol{Y_i}|\boldsymbol{X_k}) - \sum_{j-1}^{N} \ln P\left(\boldsymbol{X_j}\middle|\boldsymbol{X_{j-1}}\right)$$

# Full Field Non-Gaussian Variational Data Assimilation

**The resulting 3DVAR and 4DVAR cost functions are**

$$J(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{\varepsilon}_b^T\boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_b + \boldsymbol{\varepsilon}_b^T\begin{pmatrix}\boldsymbol{0}_{bp}\\ \boldsymbol{1}_{bq}\end{pmatrix} + \frac{1}{2}\boldsymbol{\varepsilon}_o^T\boldsymbol{R}^{-1}\boldsymbol{\varepsilon}_o + \boldsymbol{\varepsilon}_o^T\begin{pmatrix}\boldsymbol{0}_{op}\\ \boldsymbol{1}_{oq}\end{pmatrix}$$

**(Fletcher and Zupanski, 2007), where**

$$\boldsymbol{\varepsilon}_b = \begin{pmatrix}\boldsymbol{x}_p^t - \boldsymbol{x}_{bp}\\ \ln \boldsymbol{x}_q^t - \ln \boldsymbol{x}_{bq}\end{pmatrix} \; \boldsymbol{and} \;\; \boldsymbol{\varepsilon}_o = \begin{pmatrix}\boldsymbol{y}_{op} - \boldsymbol{h}_{op}(\boldsymbol{x})\\ \ln \boldsymbol{y}_{oq} - \ln \boldsymbol{h}_{oq}(\boldsymbol{x})\end{pmatrix}$$

**and**

$$\boldsymbol{J}(\boldsymbol{x}_0) = \frac{1}{2}\boldsymbol{\varepsilon}_{0b}^T\boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_{0b} + \boldsymbol{\varepsilon}_{0b}^T\begin{pmatrix}\boldsymbol{0}_{bp}\\ \boldsymbol{1}_{bp}\end{pmatrix} + \frac{1}{2}\sum_{i=1}^{N_o}\boldsymbol{\varepsilon}_{0i}^T\boldsymbol{R}_i^{-1}\boldsymbol{\varepsilon}_{oi} + \sum_{i=1}^{N_o}\boldsymbol{\varepsilon}_{oi}^T\begin{pmatrix}\boldsymbol{0}_{opi}\\ \boldsymbol{1}_{oqi}\end{pmatrix}$$

**(Fletcher, 2010) again where**

$$\boldsymbol{\varepsilon}_{0b} = \begin{pmatrix}\boldsymbol{x}_p^t(\boldsymbol{t}_0) - \boldsymbol{x}_{bp}(\boldsymbol{t}_0)\\ \ln \boldsymbol{x}_q^t(\boldsymbol{t}_0) - \ln \boldsymbol{x}_{bq}(\boldsymbol{t}_0)\end{pmatrix} \; \boldsymbol{and} \; \boldsymbol{\varepsilon}_{oi} = \begin{pmatrix}\boldsymbol{y}_{pi} - \boldsymbol{h}_{pi}\left(\boldsymbol{M}_i(\boldsymbol{x}(\boldsymbol{t}_0))\right)\\ \ln \boldsymbol{y}_{qi} - \ln \boldsymbol{h}_{qi}\left(\boldsymbol{M}_i(\boldsymbol{x}(\boldsymbol{t}_0))\right)\end{pmatrix}$$


Fletcher, S. J. and M. Zupanski, 2007: Implications and impacts of transforming lognormal variables into normal variables in VAR. *Meteor. Zeit.*, **16,** 755—765.

Fletcher, S.J., 2010: Mixed lognormal-Gaussian four-dimensional data assimilation. *Tellus,* **62A,** 266—287.