4.7

AN ENSEMBLE DATA ASSIMILATION SCHEME THAT HAS MIXED GAUSSIAN AND NON-GAUSSIAN ERRORS

Steven J. Fletcher* and Milija Zupanski
Cooperative Institute for Research in the Atmosphere,
Colorado State University,
Fort Collins, CO

1. Introduction

It is well known that the foundations for three and four dimensional variational data assimilation, 3D and 4D VAR, arise from Bayesian probability theory. This problem is derived in (Lorenc, 1986), L86 hereafter, for general multivariate probability density functions, pdfs, for both the marginal and conditional distributions until we arrive at the definition of the errors. At this point the errors are then assumed to be additive and hence Normally distributed.

In (Cohn, 1997) it is suggested that it may be that we have data which is lognormally distributed and hence the errors are multiplicative. Increasingly there has been more data sets to back the claim about the lognormal distributed variables, (Sengupta et al., 2004) but as early as 1977 variables in the atmosphere have been identified as lognormally distributed, (Mielke et al., 1977).

Given these non Normal variables then we have to address how to correctly assimilate them. In (Fletcher and Zupanski, 2006a), FZ06a hereafter, we start to address this problem for multivariate lognormally distributed variables. We are able to derive a cost function similar to the one associated with the multivariate Normal mode. There is much discussion as to which statistic is the best to use to represent the distribution. With the multivariate lognormal there is only one statistic that can represent the distribution and that is its mode.

The reason for such a bold statement is due to a property that statisticians have known about the lognormal distribution since the 1960s. This distribution is not uniquely determined by its moments, (Heyde, 1963). This then causes problems if we are seeking the first moment. We would not know if we were approximating the lognormal distribution or the other distribution that satisfies the moment equations.

In FZ06a we derive the cost function associated with the mode of the multivariate lognormal distribution. We know that this state is the most likely and hence this is where the pdf is at a maximum. We show that the cost function associated with the lognormal distribution is similar to the Normal cost function and that the Jacobian and the Hessian also have a similar structure. The reason for seeking these matrices is for their use in minimisation routine such as quasi-Newton and conjugate gradients methods but also they are used in Hessian preconditioning, (Zupanski, 2005).

Deriving the cost function for the multivariate lognormal distribution is the first step. In atmospheric and oceanic modelling we have a large number of observations and model states which are not all from the same type of distribution and hence we can either assimilate the different types separately and ignore the correlations between the variables, try and find a transform so that all of the variables are distributed of the same type or as we have accomplished in (Fletcher and Zupanski, 2006b), FZ06b hereafter, we can define new multivariate distributions which are combination of other distributions.

In FZ06b we define and prove a multivariate distribution which has p Normal and q lognormal variates. The distribution retains many of the properties of the two subdistributions but also contains new properties that are not obvious at first glance but when the mathematics is applied the origin of the new properties become clear.

In this paper we show how to incorporate the new distribution into the Maximum Likelihood Ensemble Filter, MLEF, (Zupanski, 2005; Zupanski et al., 2005) which is under development at the Cooperative Institute for Research in the Atmosphere, CIRA, Colorado State University, CSU, and Florida State University, FSU. This is an ensemble filter which instead of using the ensemble mean as in the Ensemble Transform Kalman Filter, ETKF, (Bishop et al., 2001), to generate our statistics we use the state with the maximum likelihood. This state is currently found through solving the non-linear quadratic cost function associated with Normal errors.

In the next section we briefly summarise the Bayesian framework as set out L86 and show the derivation to the cost function associated with Normal errors. In Section 3 we present a summary of the derivation and key results from FZ06a to do with the lognormal errors. In Section 4 we present the hybrid distribution and derive the associated cost function as well as the Jacobian and Hessian to show that the structure of these matrices is similar to that of the Normal case. We finish with conclusions and plans on how to test this distribution in the MLEF.

2. Normal Framework

In this section we provide a summary of the derivation in L86 of the Bayesian framework that leads to the nonlinear quadratic cost function that is used when we have Normally distributed variables. Although it is thought that

**Corresponding author address:* Dr. Steven J. Fletcher, Cooperative Institute for Research in the Atmosphere, Colorado State University, 1375 Campus Delivery, Fort Collins, CO 80523-1375; Email: fletcher@cira.colostate.edu

it is the errors that are distributed Normally it is not. The important fact is that the state variables themselves are Normally distributed. We are using the property of the Normal distribution that the sum of two Normal variable is itself a Normal variable.

2.1 ***Bayesian probability***

The starting point in L86 is to consider the problem of finding the set of initial states such that the subsequent forecast is the 'best' possible. Due to the problem of the forecast being imperfect we have to compensate by introducing observations of the physical system. Therefore let the state vector be $\mathbf{x}$ where $\mathbf{x}^T = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N)$ given N is the total number of state variables, $\mathbf{y}$ is the vector of observations vector, $\mathbf{y}^T = (\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_{N_o})$ and N_o is the total number of observations such that $N_o << N$.

We now require a relationship between the model states, $\mathbf{x}$ and the observations, $\mathbf{y}$. This relationship is

$$\mathbf{y} = \mathbf{h}(\mathbf{x}), \tag{1}$$

where $\mathbf{h}(\mathbf{x})$ is a vector of non-linear interpolations or transformations from the model states to the observations i.e.

$$\mathbf{h}(\mathbf{x}) = \begin{pmatrix} h_1(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N) \\ h_2(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N) \\ \vdots \\ h_{N_o}(\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_N) \end{pmatrix}. \tag{2}$$

If the relationship between the observations and the model state variables is linear then $\mathbf{h}(\mathbf{x})$ is a matrix vector multiplication, $H\mathbf{x}$ where H is a rectangular matrix of dimensions $N_o \times N$. This then gives us the problem, according to L86, of finding the 'best' $\mathbf{x}$ which inverts (1) for a given $\mathbf{y^o}$ where $\mathbf{y^o}$ is the physical observation which contain errors.

The problem is then defined from Bayes theorem

$$P(\mathrm{A} \mid \mathrm{B}) \propto P(\mathrm{B} | \mathrm{A}) P(\mathrm{A}). \tag{3}$$

In the case that we are interested in we have the event A as $\mathbf{x} = \mathbf{x^t}$ and the event B as $\mathbf{y} = \mathbf{y^o}$. We can then define (3) as

$$P\left(\mathbf{x} = \mathbf{x^t} \middle| \mathbf{y} = \mathbf{y^o}\right) \propto P\left(\mathbf{y} = \mathbf{y^o} | \mathbf{x} = \mathbf{x^t}\right) P\left(\mathbf{x} = \mathbf{x^t}\right), \tag{4}$$

where the superscript t represents the 'true' solution and o represents observed value. With this (4) defines a N dimensional multivariate pdf, denoted as $P_a(\mathbf{x})$, where a represents the analysis.

We now seek the mode of the analysis distribution. That is we seek $\mathbf{x}_a = \mathbf{x}$ such that

$$\frac{dP_a}{d\mathbf{x}_a} = 0, \quad \frac{d^2 P_a}{d\mathbf{x}_a^2} < 0. \tag{5}$$

2.2 ***Normally Distributed Errors***

We start with the probability $P\left(\mathbf{x} = \mathbf{x}^t\right)$ which represents our knowledge about $\mathbf{x}$ before the observations are taken. This can be considered as the background error, ε_b,

$$\varepsilon_b \equiv \mathbf{x} - \mathbf{x_b}, \tag{6}$$

where $\mathbf{x_b}$ represents the background state. As such we are considering deviations away from this background state where we have an associated probability of $P_b(\mathbf{x} - \mathbf{x_b})$.

For the observational error we consider the total error which is the combination of the instrumental and representativeness errors

$$P\left(\mathbf{y} = \mathbf{y^o} | \mathbf{x} = \mathbf{x^t}\right) = P_o(\mathbf{y^o} - \mathbf{h}(\mathbf{x})), \tag{7}$$

which is ε^o. We have assumed that the observational and the background errors are independent which is an acceptable assumption L86. Combining (6) and (7) enables us to express (4) as

$$P_a(\mathbf{x}) = P_o(\mathbf{y^o} - \mathbf{h}(\mathbf{x})) P_b(\mathbf{x} - \mathbf{x_b}) \equiv P_o(\varepsilon^o) P_b(\varepsilon_b). \tag{8}$$

We now assume multivariate Normal, MN, distributions for the probabilities so that $\boldsymbol{\varepsilon^o} \sim MN(\mathbf{0}, \mathrm{R})$ where we have zero mean and covariance matrix R. The background errors are distributed $\boldsymbol{\varepsilon_b} \sim MN(\mathbf{0}, \mathrm{B})$ with mean zero and covariance matrix B. Therefore

$$\begin{aligned} P_b(\boldsymbol{\varepsilon_b}) &\propto \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon_b}^T \mathrm{B}^{-1} \boldsymbol{\varepsilon_b}\right\} && (9) \\ &\equiv \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x_b})^T \mathrm{B}^{-1} (\mathbf{x} - \mathbf{x_b})\right\}, \\ P_o(\varepsilon^o) &\propto \exp\left\{-\frac{1}{2}\varepsilon^{oT} \mathrm{R}^{-1} \varepsilon^o\right\} && (10) \\ &\equiv \exp\left\{-\frac{1}{2}(\mathbf{y^o} - \mathbf{h}(\mathbf{x}))^T \mathrm{R}^{-1} (\mathbf{y^o} - \mathbf{h}(\mathbf{x}))\right\}. \end{aligned}$$

Therefore substituting (10) and (11) into (4) yields

$$\begin{aligned} P_a(\mathbf{x}) &\propto \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x_b})^T \mathrm{B}^{-1} (\mathbf{x} - \mathbf{x_b})\right. \\ &- \left.\frac{1}{2}(\mathbf{y^o} - \mathbf{h}(\mathbf{x}))^T \mathrm{R}^{-1} (\mathbf{y^o} - \mathbf{h}(\mathbf{x}))\right\}. \end{aligned} \tag{11}$$

Maximising P_a is the equivalent of minimising $-\ln$ of (11) and so this then gives the non-linear cost function as

$$\begin{aligned} \mathrm{J}(\mathbf{x}) &= \frac{1}{2}(\mathbf{x} - \mathbf{x_b})^T \mathrm{B}^{-1} (\mathbf{x} - \mathbf{x_b}) \\ &+ \frac{1}{2}(\mathbf{y^o} - \mathbf{h}(\mathbf{x}))^T \mathrm{R}^{-1} (\mathbf{y^o} - \mathbf{h}(\mathbf{x})). \end{aligned} \tag{12}$$

If we consider an unconstrained minimisation method, such as the non-linear conjugate gradient or quasi-Newton methods, to find the minimum of (12) we require the Jacobian and the Hessian of (12). The Jacobian vector of (12) can easily be verified as

$$\frac{\partial \mathrm{J}}{\partial \mathbf{x}} \equiv \mathrm{B}^{-1}(\mathbf{x} - \mathbf{x_b}) - \mathrm{H}^T \mathrm{R}^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x})), \tag{13}$$

where

$$\mathrm{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}, \tag{14}$$

is the Jacobian Matrix of $\mathbf{h}$ with dimensions $N_o \times N$ and $\frac{\partial \mathrm{J}}{\partial \mathbf{x}}$ has dimensions $N \times 1$ where we drop the superscript o as we are now only dealing with the physical observations.

The Hessian of (12), componentwise, is defined as

$$\frac{\partial^2 \mathrm{J}}{\partial \mathrm{x}_i \partial \mathrm{x}_j} \equiv \left[\mathrm{B}^{-1} + \mathrm{H}^T \mathrm{R}^{-1} \mathrm{H}\right]_{ij} - \left[\mathrm{G}_i \mathrm{R}^{-1} \left(\mathbf{y} - \mathbf{h}\left(\mathbf{x}\right)\right)\right]_j, \tag{15}$$

where G is the Hessian of $\mathbf{h}$ such that

$$\mathrm{G}_i \equiv \frac{\partial}{\partial \mathrm{x}_i}\left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right). \tag{16}$$

Therefore the dimensions of the full Hessian matrix of J is $N_o \times N$, where there are N of the G_i matrices with $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, N$.

3. Lognormally Distributed Errors

An important difference between Normal and lognormal errors is that we only have the additive property with the Normal errors. If we consider the univariate lognormal distribution

$$f\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\}, \tag{17}$$

where $x \in (0, \infty)$ and μ and σ are the mean and standard deviation of $\ln x$ then it is impossible to find the expectation of $X + Y$ where X and Y are two independent lognormal random variables unlike the Normal distribution which is given by

$$f\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\}, \tag{18}$$

where $x \in (-\infty, \infty)$ we can find $E\left(X + Y\right)$. For lognormal variables we consider the expectation of the ratio

$$E\left(\frac{X}{Y}\right) = E\left(X\right) E\left(Y^{-1}\right).$$

Therefore it can be shown that the ratio has the lognormal distribution

$$\begin{aligned} f\left(\frac{x}{y}\right) &= \frac{y}{\sqrt{2\pi}\left(\sigma_x - \sigma_y\right) x} \\ &\quad \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \ln y - \left(\mu_x - \mu_y\right)}{\sigma_x\left(-\sigma_y\right)}\right)^2\right\}. \end{aligned} \tag{19}$$

Therefore we have to define our errors in terms of the ratio as indicated in (Cohn, 1997).

3.1 *Lognormal observational errors*

We now consider the case where we have lognormal observational errors and Normal background errors. Therefore the conditional pdf is multivariate lognormal. Following the justification for the ratio as the variable which we consider, we define the lognormal errors as

$$y_i = \mathbf{h}\left(\mathbf{x}\right)_i \varepsilon_i^o \quad \Rightarrow \quad \boldsymbol{\varepsilon}^o = \frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)} \sim MLN\left(\mathbf{0}, \mathrm{R_L}\right), \tag{20}$$

where $\mathbf{0}$ is a vector of zeros and $\mathrm{R_L}$ is the multivariate lognormal covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \cdots & \rho_{N_o-1}\sigma_1\sigma_{N_o} \\ \rho_1\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2N_o-3}\sigma_2\sigma_{N_o} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N_o-1}\sigma_{N_o}\sigma_1 & \rho_{2N_o-3}\sigma_{N_o}\sigma_2 & \cdots & \sigma_{N_o}^2 \end{pmatrix} \tag{21}$$

where $\sigma_i,\ i = 1, 2, \ldots, N_o$ are the associated standard deviations of the components of $\ln \mathrm{x}_i$, $\rho_j,\ j = 1, 2, \ldots, \frac{1}{2}N_o\left(N_o - 1\right)$ are the correlations and the vector division in (20) is componentwise. The reason for the mean vector being zero is due to the problem that it is not possible for the statistic that we are approximating to have the three properties of minimum variance, maximum likelihood or be unbiased which we can with the Normal and most symmetric distributions. We can still find the most likely state $\mathbf{x}$ and have $\ln \mathbf{x}$ unbiased. In component form the observational error vector is

$$\varepsilon_i^o = \frac{\mathrm{y}_i}{\mathrm{h}_i\left(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N\right)}, \tag{22}$$

where $i = 1, 2, \ldots, N_o$.

We now require the multivariate version of the lognormal distribution. This is given by

$$\begin{aligned} f\left(\mathbf{x}\right) &= \frac{1}{\left(2\pi\right)^{\frac{N}{2}}\left|\mathrm{R_L}\right|}\left(\prod_{i=1}^{N_o}\frac{h_i\left(\mathbf{x}\right)}{y_i}\right) \\ &\quad \exp\left\{-\frac{1}{2}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)^T \mathrm{R_L}^{-1}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)\right\}. \end{aligned} \tag{23}$$

As we have mentioned and we go into more detail of the justification in FZ06a, we seek the mode of (23) rather than the median or the mean. Therefore we solve the dual problem of finding the minimum of $-\ln$ (23). This then gives us the problem

$$J\left(\mathbf{x}\right) = \frac{1}{2}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)^T \mathrm{R_L}^{-1}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right) + \left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)^T \mathbf{1}, \tag{24}$$

where $\mathbf{1}^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}$ and gives the full cost function with the Normal background as

$$\begin{aligned} J\left(\mathbf{x}\right) &= \frac{1}{2}\left(\mathbf{x} - \mathbf{x_b}\right)^T \mathrm{B}^{-1}\left(\mathbf{x} - \mathbf{x_b}\right) \\ &+ \frac{1}{2}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)^T \mathrm{R_L}^{-1}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right) \\ &+ \left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right)^T \mathbf{1}, \end{aligned} \tag{25}$$

where B is the background covariance matrix and $\mathbf{x_b}$ is some background state vector.

It is shown in FZ06a that the Jacobian of (25) is

$$\frac{\partial J}{\partial \mathbf{x}} = \mathrm{B}^{-1}\left(\mathbf{x} - \mathbf{x_b}\right) - \hat{\mathrm{H}}^T \mathrm{R_L}^{-1}\left(\ln\frac{\mathbf{y}}{\mathbf{h}\left(\mathbf{x}\right)}\right) - \hat{\mathrm{H}}^T \mathbf{1}, \tag{26}$$

where $\hat{\mathrm{H}}$ is similar to the Jacobian of the observation operator as for Normal errors and is defined as

$$\hat{\mathrm{H}} = \begin{pmatrix} \frac{1}{h_1}\frac{\partial h_1}{\partial \mathrm{x}_1} & \frac{1}{h_1}\frac{\partial h_1}{\partial \mathrm{x}_2} & \cdots & \frac{1}{h_1}\frac{\partial h_1}{\partial \mathrm{x}_N} \\ \frac{1}{h_2}\frac{\partial h_2}{\partial \mathrm{x}_1} & \frac{1}{h_2}\frac{\partial h_2}{\partial \mathrm{x}_2} & \cdots & \frac{1}{h_2}\frac{\partial h_2}{\partial \mathrm{x}_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h_{N_o}}\frac{\partial h_{N_o}}{\partial \mathrm{x}_1} & \frac{1}{h_{N_o}}\frac{\partial h_{N_o}}{\partial \mathrm{x}_2} & \cdots & \frac{1}{h_{N_o}}\frac{\partial h_{N_o}}{\partial \mathrm{x}_N} \end{pmatrix}, \quad (27)$$

and is of dimensions $N \times N_o$.

It is possible to rearrange (26) to able to find the mode of the analysis distribution given by

$$\mathbf{x} = \mathbf{x_b} + \mathrm{B}\hat{\mathrm{H}}^T \left(\mathrm{R_L}^{-1} \ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x})} + \mathbf{1} \right). \quad (28)$$

However, (28) has the observation operator evaluated at the state that we seek. We now introduce a linearisation to $\ln \mathbf{h}(\mathbf{x})$ such that

$$-\ln \mathbf{h}(\mathbf{x}) \approx -\ln \mathbf{h}(\mathbf{x_b}) - \hat{\mathrm{H}}(\mathbf{x} - \mathbf{x_b}). \quad (29)$$

We also have the term $\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x})}^T \mathbf{1}$ to linearise. This term is also represented by

$$\left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x})} \right)^T \mathbf{1} \equiv \sum_{i=1}^{N_o} \ln \frac{y_i}{h_i(\mathbf{x})}, \quad (30)$$

therefore we need to linearise the sum in (30). We accomplish this by using the following first order in terms of partial derivatives of $\mathbf{x}$ approximation

$$\begin{aligned} \sum_{i=1}^{N_o} \ln \frac{y_i}{h_i(\mathbf{x})} &\approx \quad (31) \\ &\sum_{i=1}^{N_o} \left(\ln \frac{y_i}{h_i} - \frac{1}{h_i}\frac{\partial h_i}{\partial \mathbf{x_b}} (\mathbf{x} - \mathbf{x_b}) \right. \\ &+ \left. \frac{1}{2} (\mathbf{x} - \mathbf{x_b})^T \frac{1}{h_i^2} \left(\frac{\partial h_i}{\partial \mathbf{x_b}} \right) \left(\frac{\partial h_i}{\partial \mathbf{x_b}} \right)^T (\mathbf{x} - \mathbf{x_b}) \right). \end{aligned}$$

It is possible to write the summation in (31) in terms of matrix multiplications given by

$$\begin{aligned} \sum_{i=1}^{N_o} \ln \frac{y_i}{h_i(\mathbf{x})} &\approx \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} \right)^T \mathbf{1} - (\mathbf{x} - \mathbf{x_b})^T \hat{\mathrm{H}}^T \mathbf{1} \\ &+ \frac{1}{2} (\mathbf{x} - \mathbf{x_b})^T \hat{\mathrm{H}}^T \hat{\mathrm{H}} (\mathbf{x} - \mathbf{x_b}). \quad (32) \end{aligned}$$

Substituting (29) and (32) into (26) allows the linearised cost function to be written as

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2} (\mathbf{x} - \mathbf{x_b})^T \mathrm{B}^{-1} (\mathbf{x} - \mathbf{x_b}) \quad (33) \\ &+ \frac{1}{2} \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} \right)^T \mathrm{R_L}^{-1} \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} \right) \\ &- (\mathbf{x} - \mathbf{x_b})^T \hat{\mathrm{H}}^T \mathrm{R_L}^{-1} \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} \right) \\ &+ \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} - \hat{\mathrm{H}}(\mathbf{x} - \mathbf{x_b}) \right)^T \mathbf{1} \\ &+ \frac{1}{2} (\mathbf{x} - \mathbf{x_b})^T \hat{\mathrm{H}}^T \hat{\mathrm{H}} (\mathbf{x} - \mathbf{x_b}). \end{aligned}$$

The Jacobian of (33) is

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{x}} &= \mathrm{B}^{-1} (\mathbf{x} - \mathbf{x_b}) - \hat{\mathrm{H}}\mathrm{R_L}^{-1} \left(\ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} \right) \quad (34) \\ &+ \hat{\mathrm{H}}^T \mathrm{R_L}^{-1} \hat{\mathrm{H}} (\mathbf{x} - \mathbf{x_b}) - \hat{\mathrm{H}}^T \mathbf{1} + \hat{\mathrm{H}}^T \hat{\mathrm{H}} (\mathbf{x} - \mathbf{x_b}). \end{aligned}$$

Therefore rearranging to find the minimum gives

$$\begin{aligned} \mathbf{x} = \mathbf{x_b} &+ \left(\mathrm{I} + \mathrm{B}\hat{\mathrm{H}}^T \mathrm{R_L}^{-1} \hat{\mathrm{H}} + \mathrm{B}\hat{\mathrm{H}}^T \hat{\mathrm{H}} \right)^{-1} \times \\ &\left(\mathrm{B}\hat{\mathrm{H}}^T \mathrm{R_L}^{-1} \ln \frac{\mathbf{y}}{\mathbf{h}(\mathbf{x_b})} + \mathrm{B}\hat{\mathrm{H}}^T \mathbf{1} \right). \quad (35) \end{aligned}$$

Therefore we have been able to show that most of the familiar expressions from the Normal framework cross over to the lognormal framework. In FZ06a we present the Hessian matrix for (25) where the first order derivative terms are those that are present in the inversion in (35). The advantage of this means that we are able to generate a form of Hessian preconditioning, (Zupanski, 2005), to aid in the minimisation. We go into more detail about this preconditioner in Section 5.

4. Hybrid Normal-lognormal Errors

In (Fletcher and Zupanski, 2006b) we present and prove that we can have a multivariate hybrid distribution of p Normal variates and q lognormal variates. The multivariate version is given by

$$\begin{aligned} f_{p,q}(\mathbf{x}) \equiv &\frac{1}{(2\pi)^{\frac{N}{2}} |\mathrm{R_{hy}}|^{\frac{1}{2}}} \left(\prod_{i=1}^{q} \frac{1}{\mathrm{x}_i} \right) \\ &\exp \left\{ (\hat{\mathbf{x}} - \boldsymbol{\mu})^T \mathrm{R_{hy}}^{-1} (\hat{\mathbf{x}} - \boldsymbol{\mu}) \right\}, \quad (36) \end{aligned}$$

where R is of the same structure as (21) and

$$\hat{\mathbf{x}} \equiv \begin{pmatrix} \mathbf{x}_p \\ \ln \mathbf{x}_q \end{pmatrix}$$

and $\mathbf{x}_p \in \mathbb{R}^p$ and $\mathbf{x}_q \in \mathbb{R}^{q+}$. We show in FZ06b that the expectation of the individual Normal or lognormal distribution is that of their individual sub-distribution. The interesting feature is the mode of (36) which keeps the lognormal components having the same mode as they would for that type of distribution but the Normal components are scaled by the their covariances with the lognormal components,

$$\hat{\mathbf{x}}_{mo} = \boldsymbol{\mu} - \mathrm{R_{hy}} \begin{pmatrix} \mathbf{0}_p \\ \mathbf{1}_q \end{pmatrix}, \quad (37)$$

where $\mathbf{0}_p^T = \begin{pmatrix} 0 & 0 & \ldots & 0 \end{pmatrix}$ has dimensions $p \times 1$.

Following the same justification as for the multivariate lognormal distribution we seek the mode of (36) for the observational conditional pdf following this hybrid distribution and as such we seek the minimum of the dual problem given by

$$\begin{aligned} J_{hy}\left(\mathbf{x}\right) &= \frac{1}{2}\left(\mathbf{x}-\mathbf{x_b}\right)^T \mathrm{B}^{-1}\left(\mathbf{x}-\mathbf{x_b}\right) \\ &+ \frac{1}{2}\tilde{\varepsilon}^{oT}\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} + \tilde{\varepsilon}^{o}\left(\begin{array}{c} \mathbf{0}_p \\ \mathbf{1}_q \end{array}\right), \end{aligned} \quad (38)$$

where

$$\tilde{\varepsilon}^{o} = \left(\begin{array}{c} \mathbf{y}_p - \mathbf{h}_p\left(\mathbf{x}\right) \\ \ln \mathbf{y}_q - \ln \mathbf{h}_q\left(\mathbf{x}\right) \end{array}\right). \quad (39)$$

We can easily find the Jacobian of (38) as

$$\frac{\partial J_{hy}}{\partial \mathbf{x}} = \mathrm{B}^{-1}\left(\mathbf{x}-\mathbf{x_b}\right) - \tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} - \hat{\mathrm{H}}_q^T\mathbf{1}_q, \quad (40)$$

where

$$\tilde{\mathrm{H}} = \left(\begin{array}{c} \mathrm{H}_p \\ \hat{\mathrm{H}}_q \end{array}\right), \quad \mathrm{H} = p \times N \ \text{ and } \ \hat{\mathrm{H}} = q \times N,$$

where H_p is as defined by (14) and $\hat{\mathrm{H}}_q$ is defined by (27).

The Hessian of (38) is given by

$$\begin{aligned} \left(\frac{\partial^2 J_{hy}}{\partial \mathbf{x}^2}\right)_{ij} &= \left[\mathrm{B}^{-1} + \tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\mathrm{H}} + \hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q\right]_{i,j} \\ &- \left[\tilde{\mathrm{G}}\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o}\right]_i - \left[\bar{\mathrm{G}}\right]_{i,j}, \end{aligned} \quad (41)$$

where

$$\begin{aligned} \tilde{\mathrm{G}} &= \left(\begin{array}{c} \mathrm{G}_p \\ \hat{\mathrm{G}}_q \end{array}\right), \\ \mathrm{G}_{i,j} = \frac{\partial^2 h_i}{\partial \mathrm{x}_j \partial \mathrm{x}_i}, \ i &= 1,2,\ldots,p \\ j &= 1,2,\ldots,N, \\ \hat{\mathrm{G}}_{i,j} = \frac{1}{h_i}\frac{\partial^2 h_i}{\partial \mathrm{x}_i \partial \mathrm{x}_j}, \ i &= p+1,p+2,\ldots,N_o, \\ j &= 1,2,\ldots,N, \\ \bar{\mathrm{G}}_{i,j} = \sum_{k=p+1}^{N_o}\frac{1}{h_k}\frac{\partial^2 h_k}{\partial \mathrm{x}_i \partial \mathrm{x}_j}, \ i &= 1,2,\ldots,N \\ j &= 1,2,\ldots,N. \end{aligned}$$

To find the solution we introduce the following three linearisations to the relative components

$$\begin{aligned} \mathbf{h}_p\left(\mathbf{x}\right) &\approx \mathbf{h}_p\left(\mathbf{x_b}\right) + \mathrm{H}_p\left(\mathbf{x}-\mathbf{x_b}\right), & (42) \\ \ln \mathbf{h}_q\left(\mathbf{x}\right) &\approx \ln \mathbf{h}_q\left(\mathbf{x_b}\right) + \hat{\mathrm{H}}_q\left(\mathbf{x}-\mathbf{x_b}\right), & (43) \\ \left(\ln \mathbf{h}_q\right)^T\mathbf{1}_q &\approx \left(\ln \mathbf{h}_q\left(\mathbf{x_b}\right)\right)^T\mathbf{1}_q + \left(\mathbf{x}-\mathbf{x_b}\right)^T\hat{\mathrm{H}}_q^T\mathbf{1}_q \\ &- \frac{1}{2}\left(\mathbf{x}-\mathbf{x_b}\right)^T\hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q\left(\mathbf{x}-\mathbf{x_b}\right). & (44) \end{aligned}$$

If we now substitute (42), (43) and (44) into (38) and (39) we have

$$\begin{aligned} J_{hy}\left(\mathbf{x}\right) &= \frac{1}{2}\left(\mathbf{x}-\mathbf{x_b}\right)^T\mathrm{B}^{-1}\left(\mathbf{x}-\mathbf{x_b}\right) \\ &+ \frac{1}{2}\tilde{\varepsilon}^{oT}\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} - \left(\mathbf{x}-\mathbf{x_b}\right)^T\tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} \\ &+ \frac{1}{2}\left(\mathbf{x}-\mathbf{x_b}\right)^T\tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\mathrm{H}}\left(\mathbf{x}-\mathbf{x_b}\right) \\ &+ \left(\ln\frac{\mathbf{y}_q}{\mathbf{h}_q}\left(\mathbf{x_b}\right)\right)^T\mathbf{1}_q - \left(\mathbf{x}-\mathbf{x_b}\right)^T\hat{\mathrm{H}}_q^T\mathbf{1}_q \\ &+ \frac{1}{2}\left(\mathbf{x}-\mathbf{x_b}\right)^T\hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q\left(\mathbf{x}-\mathbf{x_b}\right), \end{aligned} \quad (45)$$

where $\hat{\varepsilon}_o$ is evaluated at $\mathbf{x_b}$. Taking the first derivative with respect to $\mathbf{x}$ yields

$$\begin{aligned} \mathbf{0} &= \left(\mathrm{B}^{-1} + \tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\mathrm{H}} + \hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q\right)\left(\mathbf{x}-\mathbf{x_b}\right) \\ &- \left(\tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} + \hat{\mathrm{H}}_q^T\mathbf{1}_q\right). \end{aligned} \quad (46)$$

We can easily rearrange (46) to find the solution

$$\begin{aligned} \mathbf{x} &= \mathbf{x_b} + \left(I + \mathrm{B}\tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\mathrm{H}} + \mathrm{B}\hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q\right)^{-1} \\ &\times \left(\mathrm{B}\tilde{\mathrm{H}}^T\mathrm{R}_{\mathrm{hy}}^{-1}\tilde{\varepsilon}^{o} + \mathrm{B}\hat{\mathrm{H}}_q^T\mathbf{1}_q\right). \end{aligned} \quad (47)$$

Therefore most of the minimisation techniques that we use for the Normal minimisation techniques can be applied to both the multivariate lognormal and the multivariate hybrid distribution. As we have mentioned, the advantage of this hybrid distribution is that it allows us two assimilate to different variable types at the same time rather than separating them to their individual distribution type of transform the variable into a Gaussian variable.

Before we move onto the penultimate section we have a word of waring about this last practice. Although it is true that if X is lognormal then $\ln X$ is Normal but if we seek the so call mode for $\ln X$ we are actually finding the median not the mode. Therefore when we transform back to the lognormal or model variable then we have performed our analysis about the median not the mode and therefore we are making different conclusions to what we think we are actually performing.

5. Maximum Likelihood Ensemble Filter

The Maximum Likelihood Ensemble Filter, MLEF, (Zupanski, 2005), has a similar structure to the Ensemble Kalman Tranform Filter, ETKF, (Bishop et al., 2001) and could be consider to be in a family of these types of filter but has a major difference in the fact that it uses the most likely state rather than the ensemble mean.

This most likely state or mode of the analysis probability distribution currently if we have all Gaussian errors is found by solving a version of a cost function similar to (12). The cost function is similar in that it appears to be the same but the background covariance matrix is evaluated in ensemble space and not the full model space.

To verify this statement we shall provide a brief summary of the derivation of the MLEF.

5.1 *Derivation of the MLEF*

The starting point for the derivation of the filter is from an approximation to the forecast error covariance evolution of the discrete Kalman Filter with Gaussian errors given by

$$P_f\left(k\right) = M_{k-1,k}P_a\left(k-1\right)M_{k-1,k}^T + Q\left(k-1\right), \quad (48)$$

where P_f is the forecast error covariance matrix, M is the non-linear model matrix, P_a is the analysis covariance matrix and Q is the Gaussian model error covariance matrix. We make the assumption at this point of the development of the filter to set the model error to zero. We now factorise P_f into its square root components, $P_f^{\frac{1}{2}}$,

$$P_f = MP_aM^T = \left(MP_a^{\frac{1}{2}}\right)\left(MP_a^{\frac{1}{2}}\right)^T = P_f^{\frac{1}{2}}P_f^{\frac{T}{2}}, \quad (49)$$

where $P_a^{\frac{1}{2}}$ is the square root analysis covariance matrix. We assume that this matrix can be written as

$$P_a^{\frac{1}{2}} = \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_S \end{pmatrix}, \quad \mathbf{p}_i = \begin{pmatrix} p_{1,i} \\ p_{2,i} \\ \vdots \\ p_{N,i} \end{pmatrix}, \quad (50)$$

where S is the total number of ensembles and such that $S << N$. Substituting (50) into (49) gives

$$\begin{aligned} P_f^{\frac{1}{2}} &= \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_S \end{pmatrix}, \\ \mathbf{b}_i &= M\left(\mathbf{x}_{k-1} + \mathbf{p}_i\right) - M\left(\mathbf{x}_{k-1}\right) \approx M\mathbf{p}_i, \end{aligned} \quad (51)$$

where $\mathbf{x}_{k-1}$ is the analysis from the previous cycle at time t_{k-1}. An important thing to note here is the fact that $P_f^{\frac{1}{2}}$ is obtained from S non-linear ensemble forecast runs and one control run. We also use these columns to initiate the ensembles for the next analysis cycle and hence the forecast error covariance matrix brings a form of flow dependency into the error analysis.

We now address how we find $\mathbf{x}_{k-1}$. As we mention the filter seeks the most likely dynamical state and hence we require the mode of the analysis pdf. We therefore use (12) to find $\mathbf{x}_{k-1}$ but now defined as

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\left(\mathbf{x} - \mathbf{x_b}\right)^T P_f^{-1}\left(\mathbf{x} - \mathbf{x_b}\right) \\ &+ \frac{1}{2}\left(\mathbf{y} - \mathbf{h}(\mathbf{x})\right)^T \mathrm{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x})\right), \end{aligned} \quad (52)$$

However, we do have P_f^{-1} but we do have $P_f^{\frac{1}{2}}$. We therefore introduce a Hessian preconditioner

$$\mathbf{x} - \mathbf{x_b} = P_f^{\frac{1}{2}}\left(I + C\right)^{-\frac{T}{2}}\boldsymbol{\zeta}, \quad (53)$$

where $\boldsymbol{\zeta}$ is the control variable defined in ensemble subspace and

$$C = P_f^{\frac{1}{2}}\mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}P_f^{\frac{1}{2}} = \left(\mathrm{R}^{-\frac{1}{2}}\mathrm{H}P_f^{\frac{1}{2}}\right)^T\left(\mathrm{R}^{-\frac{1}{2}}\mathrm{H}P_f^{\frac{1}{2}}\right). \quad (54)$$

We have the practical question of how do we define the matrix multiplication. In the previous analysis cycle $P_f^{\frac{1}{2}}$ is calculated as part of the algorithm. We therefore know the columns of $P_f^{\frac{1}{2}}$ and this helps us towards the calculation of $(I + C)^{-\frac{T}{2}}$. We there introduce the vector $\boldsymbol{z_i}$ which is defined as

$$\begin{aligned} \boldsymbol{z_i} &= \left(\mathrm{R}^{-\frac{1}{2}}\mathrm{H}P_f^{\frac{1}{2}}\right)_i = \mathrm{R}^{-\frac{1}{2}}\mathrm{H}\mathbf{b}_i \\ &\approx \mathrm{R}^{-\frac{1}{2}}\left(\mathbf{h}\left(\mathbf{x} + \mathbf{b}_i\right) - \mathbf{h}(\mathbf{x})\right). \end{aligned} \quad (55)$$

With this approximation in (55) we define the C matrix as

$$C = \begin{pmatrix} \boldsymbol{z}_1^T\boldsymbol{z}_1 & \boldsymbol{z}_1^T\boldsymbol{z}_2 & \dots & \boldsymbol{z}_1^T\boldsymbol{z}_S \\ \boldsymbol{z}_2^T\boldsymbol{z}_1 & \boldsymbol{z}_2^T\boldsymbol{z}_2 & \dots & \boldsymbol{z}_2^T\boldsymbol{z}_S \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{z}_S^T\boldsymbol{z}_1 & \boldsymbol{z}_S^T\boldsymbol{z}_2 & \dots & \boldsymbol{z}_S^T\boldsymbol{z}_S \end{pmatrix}. \quad (56)$$

The advantage of this way of writing the matrix this way is that it is a symmetric matrix and hence has an orthogonal eigenvalue decomposition. This then enables us to write C as $C = V\Lambda V^T$ where V is the matrix of the orthogonal eigenvectors of C and Λ is a diagonal matrix containing the eigenvalues of C. This then enables us to calculate

$$(I + C)^{-\frac{T}{2}} = V\left(I + \Lambda\right)^{-\frac{1}{2}}V^T. \quad (57)$$

The final details of the algorithm can be found in (Zupanski, 2005) and as such we will not go into the technical parts but just to say when the minimum value of the transformed version of the cost function is found by substituting (53) into (12) we update the analysis covariance square root matrix as

$$P_a^{\frac{1}{2}} = P_f^{\frac{1}{2}}\left(I + C\left(\mathbf{x}_{opt}\right)\right)^{-\frac{T}{2}}. \quad (58)$$

We now consider how to form the Hessian preconditioner for the multivariate lognormal situation as well as the hybrid error distribution.

5.2 *Lognormal and Hybrid errors Hessian preconditioning*

We start by considering the two definitions we have for the Hessians associated with the two error types. For multivariate lognormal error we have

$$\mathrm{G_L} = P_f^{-1} + \hat{\mathrm{H}}^T\mathrm{R_L}^{-1}\hat{\mathrm{H}} + \hat{\mathrm{H}}^T\hat{\mathrm{H}}, \quad (59)$$

and for the hybrid the Hessian is

$$\mathrm{G_{hy}} = P_f^{-1} + \tilde{\mathrm{H}}^T\mathrm{R_{hy}^{-1}}\tilde{\mathrm{H}} + \hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_q. \quad (60)$$

If we now pre-multiply (59) and (60) by $P_f^{\frac{T}{2}}$ and post multiply by $P_f^{\frac{1}{2}}$ then we obtain

$$\begin{aligned} \mathrm{G_L} &= I + P_f^{\frac{T}{2}}\hat{\mathrm{H}}^T\mathrm{R_L}^{-1}\hat{\mathrm{H}}P_f^{\frac{1}{2}} + P_f^{\frac{T}{2}}\hat{\mathrm{H}}^T\hat{\mathrm{H}}P_f^{\frac{1}{2}}, \quad (61) \\ \mathrm{G_{hy}} &= I + P_f^{\frac{T}{2}}\tilde{\mathrm{H}}^T\mathrm{R_{hy}^{-1}}\tilde{\mathrm{H}}P_f^{\frac{1}{2}} + P_f^{\frac{T}{2}}\hat{\mathrm{H}}_q^T\hat{\mathrm{H}}_qP_f^{\frac{1}{2}}. \quad (62) \end{aligned}$$

This then means we can write (61) and (62) as

$$\begin{aligned} \mathrm{G_L} &= I + C_L, \quad (63) \\ \mathrm{G_{hy}} &= I + C_{hy}. \quad (64) \end{aligned}$$

From (59) and (62) we can now define the two new Hessian preconditioners as

$$\begin{aligned} \mathbf{x} - \mathbf{x_b} &= P_f^{\frac{1}{2}}\left(I + C_L\right)^{-\frac{T}{2}}\boldsymbol{\zeta}_L, \quad (65) \\ \mathbf{x} - \mathbf{x_b} &= P_f^{\frac{1}{2}}\left(I + C_{hy}\right)^{-\frac{T}{2}}\boldsymbol{\zeta}_{hy}. \quad (66) \end{aligned}$$

We also can apply the same approximation that we use for the Gaussian case for the other two distributions

$$\begin{aligned} C_L &= \left(\mathrm{R}_{\mathrm{L}}^{-\frac{1}{2}}\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right)^T\left(\mathrm{R}_{\mathrm{L}}^{-\frac{1}{2}}\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right) \\ &+ \left(\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right)^T\left(\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right), \qquad (67) \\ C_{hy} &= \left(\mathrm{R}_{\mathrm{L}}^{-\frac{1}{2}}\tilde{\mathrm{H}}P_f^{\frac{1}{2}}\right)^T\left(\mathrm{R}_{\mathrm{L}}^{-\frac{1}{2}}\tilde{\mathrm{H}}P_f^{\frac{1}{2}}\right) \\ &+ \left(\hat{\mathrm{H}}_qP_f^{\frac{1}{2}}\right)^T\left(\hat{\mathrm{H}}_qP_f^{\frac{1}{2}}\right). \qquad (68) \end{aligned}$$

However, this is not the most effective way to approximate the C_L matrix. We can also write (67) as

$$\begin{aligned} C_L &= P_f^{\frac{T}{2}}\hat{\mathrm{H}}^T\left(\mathrm{R_L}^{-1}+I\right)\hat{\mathrm{H}}P_f^{\frac{1}{2}} \qquad (69) \\ &= \left(\left(\mathrm{R}_{\mathrm{L}}^{-1}+I\right)^{\frac{1}{2}}\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right)^T\left(\left(\mathrm{R}_{\mathrm{L}}^{-1}+I\right)^{\frac{1}{2}}\hat{\mathrm{H}}P_f^{\frac{1}{2}}\right). \end{aligned}$$

From this we introduce the vector $\boldsymbol{z}$ this time as

$$\begin{aligned} \boldsymbol{z_i} &= \left(\left(\mathrm{R}_{\mathrm{L}}^{-1}+I\right)^{\frac{1}{2}}\hat{\mathrm{H}}^TP_f^{\frac{1}{2}}\right)_i = \left(\mathrm{R}_{\mathrm{L}}^{-1}+I\right)^{\frac{1}{2}}\hat{\mathrm{H}}\mathbf{b}_i \\ &\approx \left(\mathrm{R}_{\mathrm{L}}^{-1}+I\right)^{\frac{1}{2}}\left(\frac{\mathbf{h}\left(\mathbf{x}+\mathbf{b}_i\right)-\mathbf{h}\left(\mathbf{x}\right)}{\mathbf{h}\left(\mathbf{x}\right)}\right). \qquad (70) \end{aligned}$$

The manipulation of (69) for the hybrid distribution is

$$\begin{aligned} C_{hy} &= P_f^{\frac{T}{2}}\tilde{\mathrm{H}}^T\left(\mathrm{R}_{\mathrm{hy}}^{-1}+\tilde{I}\right)\tilde{\mathrm{H}}P_f^{\frac{1}{2}} \qquad (71) \\ &= \left(\left(\mathrm{R}_{\mathrm{hy}}^{-1}+\tilde{I}\right)^{\frac{1}{2}}\tilde{\mathrm{H}}P_f^{\frac{1}{2}}\right)^T\left(\left(\mathrm{R}_{\mathrm{hy}}^{-1}+\tilde{I}\right)^{\frac{1}{2}}\tilde{\mathrm{H}}P_f^{\frac{1}{2}}\right), \end{aligned}$$

where

$$\tilde{I} = \begin{pmatrix} 0_{p\times p} & 0_{p\times q} \\ 0_{q\times p} & I_{q\times q} \end{pmatrix}.$$

We can again find a column approximation to the C_{hy} matrix as follows

$$\boldsymbol{z_i}\left(l\right) \approx \begin{cases} \mathrm{R}_{\mathrm{hy}}^{-\frac{1}{2}}\left(\mathbf{h}_l\left(\mathbf{x}+\mathbf{b}_i\right)-\mathbf{h}_l\left(\mathbf{x}\right)\right), \\ 1 \le l \le p, \\ \left(\mathrm{R}_{\mathrm{hy}}^{-1}+\tilde{I}\right)^{\frac{1}{2}}\left(\frac{\mathbf{h}_l(\mathbf{x}+\mathbf{b}_i)-\mathbf{h}_l(\mathbf{x})}{\mathbf{h}_l(\mathbf{x})}\right), \\ p < l \le N_o. \end{cases} \qquad (72)$$

Therefore, from (70) and (72) we can generate the associated C matrix as given by (56). We can therefore carrying on using the eigenvalue decomposition to find the inversion associated with the Hessian preconditioning and hence solve the associated cost functions (25) and (38).

6. Conclusions and Further Work

In this extended abstract we have expanded on the ideas set out in FZ06a and FZ06b to do with the problem of how to assimilate data that is non-Gaussian distributed. In FZ06a we tackle the problem of how to assimilate variables which are lognormal. It is a common misconception that we only have to worry about the errors being from the type of distribution that we are concerned with. This is not true. At the heart of this misconception is the assumption that the difference between to Gaussian variables is a Gaussian itself.

If we consider a lognormal variable then we have the problem that the difference between two lognormal variables is itself not lognormal. We have to consider with the product or the ratio to have this new variable as being lognormal. In Section 3 we summarise the results from FZ06a dealing with which is the best statistic to use to monitor a non-Gaussian distribution. We make the asscertion that the mode is the correct statistic as it is bounded, is a function of the covariances and the variances but also it is uniquely determined.

Another worrying chain of thought that has been applied to non-Gaussian data is to transform it into Gaussian, for the lognormal variable X simply $\ln X$ and then find the mode of $\ln X$. The problem with this is that in the Gaussian case then

$$mode = median = mean$$

and although we set up the problem to find the mode we are also finding the median as well. When we change back into lognormal space where our state or observation variables are we actually transform back to one of the non-unique medians of the multivariate lognormal distribution where we forget that for non-symmetric distributions we have

$$mode \le median \le mean.$$

A key feature of multivariate lognormal distribution is that the Jacobian and the Hessian appear to have similar structures to that of their Gaussian counterparts. This is also true for the hybrid distribution we introduce in Section 4, FZ06b. This hybrid enables us to assimilate the Gaussian and lognormal variables simultaneously and allows for covariances between the two different types.

Although the later sections deals with the implementation of the lognormal and hybrid distribution into the MLEF all of this theory is applicable to 3D and 4D variational methods as well as ensemble methods based on distribution theory.

The plans for this work is to introduce lognormal observational height errors in to CSU's 2D shallow water equations model on the sphere, (Heikes and Randall, 1995a; Heikes and Randall, 1995b) with the improved dynamics, (Ringler and Randall, 2002) combined with different Rossby-Haurwitz waves, (Williams et al., 1992). These waves are known to support different dynamics that are present in a full 3D primitive equations model and are well understood to be a first method of testing new ideas before applying to the larger more complicated models.

There are three experiments that we wish to compare between:

1. Assimilate lognormal errors in a Gaussian framework
2. Assimilate lognormal and then Gaussian variables separately
3. Assimilate using the hybrid distribution framework

By addressing these three problems we can identify any extra errors that are being introduced by incorrectly assimilating the errors by the wrong type.

Further research is possible in developing more hybrid distributions between the Gaussian distribution and other more complicated distributions say the Gamma or the Rayleigh. More work is needed on identifying what distribution the data is from so we can then firstly define the correct type of error i.e. additive or multiplicative and then develop the cost function for that distribution correctly and then finally develop the hybrid distribution between that distribution and the Gaussian.

REFERENCES

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the Ensembles Transform Kalman Filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Cohn, S. E., 1997: An introduction to estimation error theory. *J. Met. Soc. Japan*, **75**, 257–288.

Fletcher, S. J. and M. Zupanski, 2006a: A data assimilation method for lognormal variables and errors. *Submitted to Q. J. R. Meteorol. Soc.*

Fletcher, S. J. and M. Zupanski, 2006b: A hybrid multivariate Normal-lognormal distribution for data assimilation. *Submitted to Annals Appl. Prob.*

Heikes, R. and D. A. Randall, 1995a: Numerical Integration of the Shallow Water Equations on a Twisted Icosahedral Grid. Part I:Basic Design and Results of Tests. *Mon. Wea. Rev.*, **123**, 1862–1880.

Heikes, R. and D. A. Randall, 1995b: Numerical Integration of the Shallow Water Equations on a Twisted Icosahedral Grid. Part II:A Detailed Description of the Grid and an Analysis of Numerical Accuracy. *Monthly Weather Review*, **123**, 1881–1887.

Heyde, C. C., 1963: On a property of the lognormal distribution. *J. Roy. Statist. Soc. B.*, **25**, 392–393.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc*, **112**, 1177–1194.

Mielke, P. W., J. S. Williams, and S.-C. Wu, 1977: Covariance analysis technique based on bivariate lognormal distribution with weather modification applications. *J. App. Met.*, **16**, 83–187.

Ringler, T. D. and D. A. Randall, 2002: A Potential Enstrophy and Energy Conserving Numerical Scheme for Solutions of the Shallow Water Equations on a Goedesic Grid. *Mon. Wea. Rev.*, **130**, 1397–1410.

Sengupta, M., E. E. Clothiaux, and T. P. Ackerman, 2004: Climatology of warm boundary layer clouds at the ARM SGP site and their comparision to models. *J. Clim.*, **17**, 4760–4782.

Williams, D. L., J. B. Drake, J. J. Hack, R. Jakob, and P. N. Swartrauber, 1992: A standard test set for numerical approximations to the shallow water equations in spherical geometry. *J. Comp. Phy.*, **102**, 211–224.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.

Zupanski, M., S. J. Fletcher, I. M. Navon, B. Uzunoglu, R. P. Heikes, D. A. Randall, T. D. Ringler, and D. Daescu, 2005: Initiation of ensemble data assimilation. *In Print Tellus A*.