Lognormal and Mixed Gaussian–Lognormal Kalman Filters

STEVEN J. FLETCHER,[a] MILIJA ZUPANSKI,[a] MICHAEL R. GOODLIFF,[a] ANTON J. KLIEWER,[a] ANDREW S. JONES,[a] JOHN M. FORSYTHE,[a] TING-CHI WU,[a] MD. JAKIR HOSSEN,[a] AND SENNE VAN LOON[a]

[a] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado*

(Manuscript received 14 March 2022, in final form 1 December 2022)

ABSTRACT: In this paper we present the derivation of two new forms of the Kalman filter equations; the first is for a pure lognormally distributed random variable, while the second set of Kalman filter equations will be for a combination of Gaussian and lognormally distributed random variables. We show that the appearance is similar to that of the Gaussian-based equations, but that the analysis state is a multivariate median and not the mean. We also show results of the mixed distribution Kalman filter with the Lorenz 1963 model with lognormal errors for the background and observations of the z component, and compare them to analysis results from a traditional Gaussian-based extended Kalman filter and show that under certain circumstances the new approach produces more accurate results.

KEYWORDS: Kalman filters; Data assimilation; Numerical weather prediction/forecasting

1. Introduction

Over the last decade or so, there have been advances in the variational (VAR) forms of data assimilation to allow for non-Gaussian behavior of the errors; specifically, there has been much development that allows for lognormally distributed errors, as well as for Gaussian errors combined with lognormal errors such that these errors can be minimized simultaneously (Fletcher and Zupanski 2006a,b, 2007; Fletcher 2010; Fletcher and Jones 2014). A full summary of the development of the mixed Gaussian–lognormal variational systems from full field 3DVAR to incremental 4DVAR can be found in Fletcher (2017).

The aforementioned theory has been applied in a microwave retrieval system for temperature and mixing ratio from the AMSU-A brightness temperatures (Kliewer et al. 2016). There, three theories are compared: 1) the mixed distribution approach, which seeks the mode, or most likely state, of the analysis distribution, 2) a logarithmic transform applied to the mixing ratio, which seeks the median of the analysis distribution (Fletcher and Zupanski 2007), and 3) a Gaussian model for the errors of temperature and mixing ratio. It is shown that the mixed distribution approach performs better than the other two mentioned approaches in fitting to observations.

However, the ability to extend the lognormal theory to the Kalman filters, and hence to the ensemble-based data assimilation systems, has not been forthcoming. The basis of the lognormal variational work was provided in the seminal paper Cohn (1997), which contains the first appearance of a definition for a lognormally distributed observational error associated with a direct observation, which is in the form of a ratio, not a difference. This definition, together with a change of variables for the background state, allowed for a version of the Kalman equations that accounts for direct observations with lognormal errors. The reader is referred to Cohn (1997) for the full details of this derivation. The main difference between the work in Cohn (1997) and the work presented here is twofold: 1) we allow for nonlinear observation operators in the Kalman filter equations, and 2) we allow for a combination of Gaussian and lognormal distributed background and observational errors.

The equations for the Kalman filter (Kalman 1960; Kalman and Bucy 1961) can be derived from either a control theory approach or a least squares formulation (Fletcher 2017), where the latter is referring to minimizing the trace of the analysis error covariance matrix with respect to the Kalman gain matrix. The Kalman filter is therefore seeking the mean of the analysis distribution to minimize the errors. For a Gaussian distribution, the mean, mode, and median are equivalent. However, for skewed distributions it is not the case that the three descriptive statistics are equal. For left skewed distributions, the mode is less than the median, which is less than the mean, whereas for right skewed distributions the opposite is true. In section 2, we show that if one tries to follow the initial steps of the least squares approach, it is impossible to derive a lognormal-based Kalman filter because one cannot separate out the analysis and forecast error covariance matrices from the logarithmic operator in order to evolve them by the linear model. Another stumbling point is that it is not possible to define a weighted sum of predicted states and new observations to determine an estimate of the state at the current time and still be consistent with a lognormal estimate.

Michael R. Goodliff's current affiliation: Cooperative Institute for Research in the Environmental Studies, University of Colorado Boulder and NOAA/Physical Sciences Laboratory, Boulder, Colorado.

Anton J. Kliewer's current affiliation: Cooperative Institute for Research in the Atmosphere, NOAA/OAR/ESRL/Global Systems Laboratory, Boulder, Colorado.

Ting-Chi Wu's current affiliation: Ministry of Science and Technology, Taiwan.

Md. Jakir Hossen's current affiliation: I.M. System's Group, Inc., College Park, Maryland.

Corresponding author: Steven J. Fletcher, steven.fletcher@colostate.edu

DOI: 10.1175/MWR-D-22-0072.1

We shall briefly summarize an alternative approach, which is referred to as the lognormal Kalman filter (Kondrashov et al. 2011), where a logarithmic transform is introduced to a model variable, and the analytical differential equation is rederived for the new variables. This approach is utilizing the property that the logarithm of a lognormally distributed random variable is a Gaussian distributed random variable and so the Kalman filter can be applied to the new variable. An inverse operation is utilized to obtain a value of the original model variable. This approach would not be practical for a numerical weather or ocean prediction model because it requires re-derivation of the associated prognostic differential equations for the state in the new variable.

Given these setbacks we recall that the Kalman filter is equivalent to 3DVAR when the static background error covariance matrix is replaced with a flow dependent background error covariance matrix, and examine a cost function–based approach to obtain the lognormal analytical state in terms of a lognormal-based Kalman gain matrix, and take the expectation of this state to obtain an expression that is similar to the Gaussian-based analysis error covariance matrix. We say "similar" in that the cost function will be defined to obtain the median, or unbiased state, of the analysis distribution. We should note here that the lognormal distribution is defined in terms of the vector of means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ for $\ln \mathbf{x}$, not the vector of random variables $\mathbf{x}$. We shall show that the lognormal-based analysis error covariance matrix is equivalent to the inverse Hessian matrix of the associated cost function scaled by the inverse of the derivatives of $\ln \mathbf{x}$. We shall also show that the estimate for the lognormal Kalman gain matrix minimizes the trace of the lognormal analysis covariance matrix.

The remainder of this paper is as follows. In section 3 we shall present the derivations of the lognormal and the mixed Gaussian–lognormal versions of the Kalman filter equations. In section 4 we shall examine the performance of the mixed Gaussian–lognormal Kalman filter equations against the Gaussian extended Kalman filter with the Lorenz 1963 model (Lorenz 1963) for different observational error variances and for different frequencies of observations. Also in this section we test the robustness of the new scheme against the extended Kalman filter over 5000 assimilation runs for the different observational error variance experiments and for 5000 perturbed true and background states. It is assumed in these experiments that the z component of the model is better modeled with a lognormal distribution than a Gaussian. The paper is finished with conclusions and ideas for future work.

2. Difficulties with lognormal-based Kalman filters

a. *Statistical derivation of the forecast error covariance matrix*

As indicated in the introduction, a first approach to derive a lognormal-based Kalman filter data assimilation system would be to follow the statistical derivation that is summarized in Fletcher (2017), but with the lognormal equivalent in parallel. The starting point is to define the Gaussian distributed analysis errors as

$$\boldsymbol{\varepsilon}_a \equiv \mathbf{x}_a - \mathbf{x}_t, \tag{1}$$

where $\mathbf{x}_t$ is the true state that is being sought, and $\mathbf{x}_a$ is the analysis state at the current time.

Introducing the time component, along with the background or forecast state $\mathbf{x}_b = \mathbf{M}_{n,n-1}\mathbf{x}_a^{n-1}$, in terms of the numerical model $\mathbf{M}$ operating on the analysis state at the previous time step, yields the background/forecast error as

$$\boldsymbol{\varepsilon}_b^n \equiv \mathbf{M}\mathbf{x}_a^{n-1} - \mathbf{x}_t^n, \tag{2}$$

where $\mathbf{M} \equiv \mathbf{M}_{n,n-1}$ is a linear or linearized numerical model matrix that operates from time step t^{n-1} to time step t^n.

Introducing the lognormal equivalent of (1) and (2), we have

$$\boldsymbol{\varepsilon}_{al}^n \equiv \mathbf{x}_{al}^n \oslash \mathbf{x}_t^n, \tag{3}$$

$$\boldsymbol{\varepsilon}_{bl}^n \equiv (\mathbf{M}\mathbf{x}_{al}^{n-1}) \oslash \mathbf{x}_t^n, \tag{4}$$

where $\oslash$ is the Hadamard division operator, which is a componentwise division operator. As we are assuming that the components of the analysis and true state are lognormally distributed, then we have the property that all of these entries are greater than zero. The subscript l above, and throughout, is referring to the components associated with the lognormal formulations. The reason that the analysis error in (3) is defined as a ratio is as a result of the lognormal distribution being geometric, which implies that the ratio or product of two independent lognormal random variables is itself a lognormal random variable. For the Gaussian distribution the equivalent property is that for two independent Gaussian random variables, their difference or sum is also a Gaussian random variable.

Using the definition of the analysis state at time $t = t^{n-1}$ for the Gaussian variable:

$$\mathbf{x}_a^{n-1} = \mathbf{x}_t^{n-1} + \boldsymbol{\varepsilon}_a^{n-1}, \tag{5}$$

the background errors can be written in terms of the previous time's true state and analysis error as

$$\boldsymbol{\varepsilon}_b^n = \mathbf{M}\mathbf{x}_t^{n-1} + \mathbf{M}\boldsymbol{\varepsilon}_a^{n-1} - \mathbf{x}_t^n. \tag{6}$$

This makes it possible to factorize the background error as

$$\boldsymbol{\varepsilon}_b^n = \mathbf{M}\boldsymbol{\varepsilon}_a^{n-1} + \boldsymbol{\varepsilon}_m^n, \tag{7}$$

where

$$\boldsymbol{\varepsilon}_m^n \equiv \mathbf{M}\mathbf{x}_t^{n-1} - \mathbf{x}_t^n, \tag{8}$$

is the model error.

To derive the analysis error covariance matrix in the lognormal Kalman filter, it is desirable to factorize the lognormal analysis error in a similar fashion as in (7). However, using

the definition of the background error (4) together with the analysis state at time $t = t^{n-1}$:

$$\mathbf{x}_{al}^{n-1} = \mathbf{x}_{t}^{n-1} \quad \boldsymbol{\varepsilon}_{al}^{n-1}, \tag{9}$$

it is clear that this is not possible. Indeed, in this way one finds

$$\ln\boldsymbol{\varepsilon}_{bl}^{n} = \ln\mathbf{M}(\mathbf{x}_{t}^{n-1} \quad \boldsymbol{\varepsilon}_{al}^{n-1}) - \ln\mathbf{x}_{t}^{n} \neq \ln\mathbf{M}\mathbf{x}_{t}^{n-1} + \ln\mathbf{M}\boldsymbol{\varepsilon}_{al}^{n-1} - \ln\mathbf{x}_{t}^{n}, \tag{10}$$

where the inequality comes from the fact that the linear model does not commute through the Hadamard product operator, and as such the model cannot act on both factors separately. Then also the logarithm cannot be expanded, making sure the background error cannot be factorized as in (6). Note that we consider the logarithm of the background error because the Kalman filter equations are defined through the expectations of the relevant Gaussian random variable, which for the lognormal distribution is $\ln\mathbf{x}$, and not the lognormal random variable $\mathbf{x}$.

To factorize the time evolution of the lognormal analysis error, another definition of the background error is necessary. A possible workaround is to instead define the background state at the next analysis time $t = t^n$ as the true state at that analysis time multiplied by the evolution of the analysis error from the previous analysis time $\mathbf{x}_{b}^{n} = \mathbf{x}_{t}^{n} \quad \mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1}$. In this case, one has to multiply by the lognormal model error to keep everything consistent. Then, the logarithm of the background error can be written as

$$\ln\boldsymbol{\varepsilon}_{bl}^{n} = \ln\mathbf{x}_{t}^{n} + \ln\mathbf{M}\boldsymbol{\varepsilon}_{al}^{n-1} - \ln\mathbf{x}_{t}^{n} + \ln\boldsymbol{\varepsilon}_{ml} = \ln\mathbf{M}\boldsymbol{\varepsilon}_{al}^{n-1} + \ln\boldsymbol{\varepsilon}_{ml}, \tag{11}$$

where the lognormal model error $\boldsymbol{\varepsilon}_{ml}$ is defined as

$$\ln\boldsymbol{\varepsilon}_{ml}^{n} \equiv \ln\mathbf{M}\mathbf{x}_{t}^{n-1} - \ln\mathbf{x}_{t}^{n}. \tag{12}$$

This now implies that it is not possible to move the numerical model out of the logarithm, which causes a problem as it is not the evolution of the logarithm of the analysis errors that is needed, but rather the logarithm of the evolution of the analysis error.

For the Gaussian filter, the forecast error covariance matrix can now be formed by multiplying (7) with its transpose and then applying the statistical expectation operator $E[\cdot]$, which yields

$$\begin{aligned}\mathbf{P}_{f}^{n} &\equiv E[\boldsymbol{\varepsilon}_{b}^{n}(\boldsymbol{\varepsilon}_{b}^{n})^{\mathrm{T}}] = E[(\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1} + \boldsymbol{\varepsilon}_{m}^{n})(\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1} + \boldsymbol{\varepsilon}_{m}^{n})^{\mathrm{T}}],\\ &= \mathbf{M}E[\boldsymbol{\varepsilon}_{a}^{n-1}(\boldsymbol{\varepsilon}_{a}^{n-1})^{\mathrm{T}}]\mathbf{M}^{\mathrm{T}} + E[\boldsymbol{\varepsilon}_{m}^{n}(\boldsymbol{\varepsilon}_{m}^{n})^{\mathrm{T}}],\\ &= \mathbf{M}\mathbf{P}_{a}^{n-1}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}^{n}\end{aligned} \tag{13}$$

where $\mathbf{P}_{a}^{n-1}$ is the analysis error covariance matrix at time $t = t^{n-1}$, and $\mathbf{Q}^{n}$ is the model error covariance matrix at time $t = t^{n}$, and it is assumed that the analysis error and the model error are uncorrelated.

Applying the same approach to (11) yields lognormal forecast and model error covariance matrices as

$$\begin{aligned}\mathbf{P}_{fl}^{n} &\equiv E[(\ln\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1} + \ln\boldsymbol{\varepsilon}_{m}^{n})(\ln\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1} + \ln\boldsymbol{\varepsilon}_{m}^{n})^{\mathrm{T}}],\\ &= E[\ln\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1}(\ln\mathbf{M}\boldsymbol{\varepsilon}_{a}^{n-1})^{\mathrm{T}}] + E[\ln\boldsymbol{\varepsilon}_{m}^{n}(\ln\boldsymbol{\varepsilon}_{m}^{n})^{\mathrm{T}}],\\ &= \mathbf{P}_{al}^{n} + \mathbf{Q}_{l}^{n}\end{aligned} \tag{14}$$

where it is assumed that the logarithm of the analysis and model errors are uncorrelated. It is clear from (14) that the definition for the forecast error covariance matrix does not explicitly contain the numerical model acting on the analysis error covariance matrix from the previous analysis time. It is, however, implicit in the state $\mathbf{x}^{n}$, as this is the result of evolving the true state and the analysis error from the previous time step. It should be noted here that given that it is not possible to interchange the model and the logarithm, for the remainder of the paper the new approach will be with the nonlinear model as there is no advantage of using the linear model to interchange operators. As future applications are more likely to be nonlinear, we remove the errors that are introduced through linearizations through this approach.

b. Kalman gain matrices

Another problem with trying to derive a lognormal version of the Kalman filter arises in the analysis step; as shown in Fletcher (2017), where if given a predicted state $\mathbf{x}_{b}^{n+1}$ that is associated with observations up to time step t^{n}, and assuming that an observation has been received at time $t = t^{n+1}$, then an estimate of the state at $t = t^{n+1}$, given the observation at time $t = t^{n+1}$ is required. In Kalman filter theory this step is started by assuming that the estimate is a weighted sum of the predicted state and the new observation, that is to say

$$\mathbf{x}_{a}^{n+1} = \mathbf{K}_{b}^{n+1}\mathbf{x}_{b}^{n+1} + \mathbf{K}_{o}^{n+1}\mathbf{y}^{n+1}, \tag{15}$$

where $\mathbf{y}^{n+1}$ is the observation at time $t = t^{n+1}$. We should note here that the observation could be either a direct or indirect observation of the predicted state. However, for a lognormal approach the equivalence of (15) would be in terms of a weighted sum of the logarithm of the predicted state and the observations such that

$$\ln\mathbf{x}_{al}^{n+1} = \ln\mathbf{K}_{bl}^{n+1}\mathbf{x}_{bl}^{n+1} + \ln\mathbf{K}_{ol}^{n+1}\mathbf{y}^{n+1}, \tag{16}$$

where the $\mathbf{K}$ matrices in (15) and (16) are referred to as gain matrices. Thus it is not possible to manipulate the equations to obtain the expressions for the gain matrices due to the logarithms being present and acting on the state, and it is not possible to interchange the operators to overcome this problem. Thus this approach cannot continue through the steps of the derivation of the Gaussian-based Kalman filter equations to seek an equivalent lognormal version of the Kalman gain matrix and filter equations.

c. Change of variable-based lognormal Kalman filter

As mentioned in the introduction, there is an alternative approach that has been proposed to obtain a form of a lognormal-based Kalman filter in Kondrashov et al. (2011). This approach, which recently has been used for a reanalysis of ring current

phase space densities in Aseev and Shprits (2019), should more accurately be referred to as the change of variable lognormal Kalman filter. In Kondrashov et al. (2011) the authors state that a striking feature of the radiation belts is that values of observed electron fluxes, and modeled particle space density (PSD), vary by several orders of magnitude. The corresponding error distributions are therefore not Gaussian, while standard data assimilation methods, such as the Kalman filter and its various adaptations to large-dimensional and nonlinear problems, are essentially based on least squares minimization of Gaussian errors, which was shown in the last section. Thus it is not possible to change these approaches directly to ensure consistency with a lognormal distribution.

As mentioned in Kondrashov et al. (2011), a possible technique that is quite often used for dealing with lognormal random variables is to use the property that the logarithm of a lognormal random variable is a Gaussian random variable. However, as shown in Fletcher and Zupanski (2007), when transforming from lognormal random variables to Gaussian random variables, minimizing the cost function in variational data assimilation, or finding the covariance matrix and the mean state with an ensemble Kalman filter system, then the descriptive statistic that is found in lognormal space once the inverse transform is applied, is the median state. A problem with this approach for the ensemble Kalman filter formulation is that the model is in terms of the original lognormal variable and not the transform logarithmic variable.

The motivation for considering a lognormal model for the PSD is justified through the statements from Kondrashov et al. (2011) that since this variable is always positive, and generally its variations, as measured by the standard deviation, increase as its mean value increases. However, Gaussian distributed variables can be negative and have a standard deviation that does not change as the mean changes. Lognormal errors arise when sources of variation accumulate multiplicatively, whereas Gaussian errors arise when these sources are additive.

To overcome the problem just mentioned, the authors of Kondrashov et al. (2011) propose rewriting the time evolution parabolic partial differential equation:

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L}\left[L^{-2} D_{\mathrm{LL}} \frac{\partial f}{\partial L}\right] - \frac{f}{\tau_L}, \tag{17}$$

where PSD is $f = f(L, t;\mu, J)$ in the Van Allen radiation belts, at fixed values of the adiabatic invariants μ and J. The radial variable L is the distance in the equatorial plane, measured in Earth radii R_E, from the center of Earth to the magnetic field line around which the electron moves at time t, D_{LL} is a radial diffusion term, and τ_L is the characteristic lifetime of the linear decay of J and μ, through considering the problem for $\log f$.

Thus, Kondrashov et al. develop a lognormal-based dynamical model for the change of variable $S = \log f$ through the chain rule for partial derivatives in time and space, applied to (17), which results in

$$\frac{\partial S}{\partial t} = L^2 \frac{\partial}{\partial L}\left[L^{-2} D_{\mathrm{LL}} \frac{\partial S}{\partial L}\right] - \frac{1}{\tau_L} + D_{\mathrm{LL}} \left(\frac{\partial S}{\partial L}\right)^2. \tag{18}$$

In Kondrashov et al. (2011) they show positive results from this approach compared to using the standard Gaussian-based Kalman filter equations. However, in numerical weather and ocean prediction it is not realistic to change all of the positive definite variables and then rederive the associated prognostic differential equations. Also as work in Goodliff et al. (2020) has shown, it is possible that the distribution of the positive definite variables can change with the dynamics, and may not always be lognormal. Therefore, also in numerical weather and ocean prediction the underlying distribution used should be able to change dynamically, which is unfeasible in a change of variable approach.

In the next section we shall present an approach that builds off of a cost function for a lognormal median of the posterior distribution for the situation where both the background, and observational, errors are lognormally distributed. We shall also show that this approach ensures that the derived covariances are consistent with a multivariate lognormal distribution.

3. Lognormal and mixed Gaussian–lognormal Kalman filters

It appears from the attempted derivation in section 2 that it is not possible to obtain a lognormal version of the Kalman filter equations, that are based upon the first two moments of the multivariate Gaussian distribution through following a least squares approach, which would have involved showing that the derived Kalman gain matrix minimized the trace of the analysis error covariance matrix. However, what is important to recall here is that for the Gaussian distribution the three descriptive statistics are the same, that is to say the mode, median, and the mean are equal. This is not true for the lognormal distribution. It is shown in Fletcher and Zupanski (2006a) that these three statistics are quite different, and that they each have their own properties; the mode is the maximum likelihood state, the median is the unbiased state, and the mean is the minimum variance state in distribution theory. To make this important distinction more clear, we will review some of the characteristics of the lognormal distribution, before deriving the lognormal Kalman filter.

a. *Properties of the lognormal distribution*

It is quite often stated that an error is unbiased if its mean is equal to zero. This is not true for many non-Gaussian distributed random variables. As just stated, the median is the unbiased state, which implies that the cumulative density function at the median is equal to 0.5. For example, to say that a lognormal random variable, or error ε, is unbiased if $E[\varepsilon] = 0$, is equivalent to saying that

$$\exp\left(\mu + \frac{\sigma^2}{2}\right) = 0, \tag{19}$$

where $\mu = E[\ln \varepsilon]$ and σ^2 is the variance of $\ln \varepsilon$ and not ε, which cannot happen. It is not possible to have a zero mean for a lognormal random variable. However, it is possible that

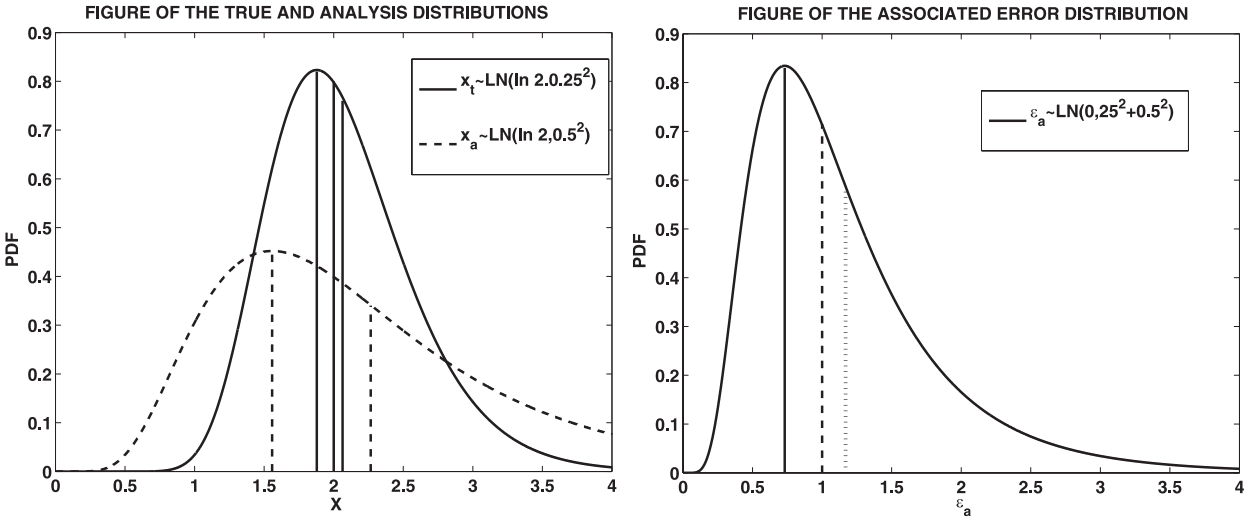

FIG. 1. (left) Plots to illustrate the differences in the modes, medians, and means for two different lognormal distributions that are representing the true state's distribution (solid curve) and the analysis state's distribution (dashed curve). (right) The distribution for the associated analysis error $\varepsilon_a = x_a/x_t$.

$E[\ln \varepsilon] = 0$, but applying this assumption to the background error would imply that the distributions of the true state and the background state are the same type and that they have the same median, not the same mean, in lognormal space, which is true in the Gaussian transformed space as well. This is because of the definition of the distribution of the ratio of two independent lognormal random variables, and its associated mean: if $x_b \sim \mathrm{LN}(\mu, \sigma_1^2)$ and $x_t \sim \mathrm{LN}(\mu, \sigma_2^2)$, then $\varepsilon_b = (x_b/x_t) \sim \mathrm{LN}(\mu - \mu, \sigma_1^2 + \sigma_2^2) = \mathrm{LN}(0, \sigma_1^2 + \sigma_2^2)$.

To help illustrate this point, we have plotted two lognormal distributions that are supposed to represent the probability density functions (PDFs) of the true state (solid curve) and the analysis state (dashed curve), where the two states have the same Gaussian mean, $\mu_t = \mu_a = \ln 2$, but with different Gaussian variances, $\sigma_t^2 = 0.25^2$ and $\sigma_a^2 = 0.5^2$, respectively, in Fig. 1. We have also plotted lines for the mode, median, and mean of the two distributions in this order in their respective line styles.

It is clear from the left panel of Fig. 1 that the two distributions have the same median, but that neither their modes nor their means are equal. In the right panel of Fig. 1 we have plotted the associated distribution for the equivalent analysis error $\varepsilon_a = x_a/x_t$. While the distribution has a median at 1, the most likely state is to the left of this value, indicating that the state with the highest probability of occurring for the analysis error is not equality, which implies a bias in the analysis.

In Fletcher and Jones (2014) it is shown that when following a modal approach for the incremental formulation of mixed Gaussian–lognormal 4DVAR, the analysis error distribution, or the posterior distribution as it is also referred to as, had a mode at 1. This is therefore indicating that the most likely answer from the data assimilation system was something close to the true state.

b. Lognormal Kalman filter—Median-based approach

Given this brief explanation and illustrations of the descriptive statistics of the lognormal distribution, it is clear that that the way to derive an equivalent set of Kalman filter–type equations for lognormally distributed errors, using the forecast error covariance matrix evolution described earlier, is to start with defining a cost function whose solution is the median of the analysis error distribution (Fletcher 2017), as this is equivalent to $\ln \boldsymbol{\varepsilon}_a$, and then find the associated covariance matrix for $\ln \boldsymbol{\varepsilon}_a$. In Fletcher (2017) it is shown that the median analysis state for lognormally distributed background and observation errors is the minimum of the following cost function:

$$J(\mathbf{x}) = \frac{1}{2}(\ln\mathbf{x} - \ln\mathbf{x}_b)^{\mathrm{T}}\mathbf{P}_{fl}^{-1}(\ln\mathbf{x} - \ln\mathbf{x}_b) + \frac{1}{2}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x})]^{\mathrm{T}}\mathbf{R}_l^{-1}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x})], \tag{20}$$

where $\mathbf{h}(\mathbf{x})$ is the nonlinear observation operator that maps the state to observation space, and $\mathbf{P}_{fl}$, now defined in terms of the nonlinear numerical model, is given by

$$\mathbf{P}_{fl}^n = [\ln M(\mathbf{x}_b^{n-1} \quad \boldsymbol{\varepsilon}_a^{n-1}) - \ln\mathbf{x}_t^n][\ln M(\mathbf{x}_b^{n-1} \quad \boldsymbol{\varepsilon}_a^{n-1}) - \ln\mathbf{x}_t^n]^{\mathrm{T}} + \mathbf{Q}_l^n. \tag{21}$$

Differentiating (20) with respect to $\mathbf{x}$ results in

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{W}_b^{-\mathrm{T}}\mathbf{P}_{fl}^{-1}(\ln\mathbf{x} - \ln\mathbf{x}_b) - \mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x})], \tag{22}$$

where $\mathbf{H}$ is the Jacobian of the observation operator, defined by $\mathbf{H}_{i,j} \equiv \partial\mathbf{h}_i(\mathbf{x})/\partial x_j$, and

$$\mathbf{W}_b^{-1} = \begin{pmatrix} x_1^{-1} & 0 & 0 & \cdots & 0 \\ 0 & x_2^{-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & x_n^{-1} \end{pmatrix},$$

$$\mathbf{W}_o^{-1} = \begin{Bmatrix} [\mathbf{h}_1(\mathbf{x})]^{-1} & 0 & 0 & \cdots & 0 \\ 0 & [\mathbf{h}_2(\mathbf{x})]^{-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & [\mathbf{h}_{N_o}(\mathbf{x})]^{-1} \end{Bmatrix}, \quad (23)$$

where n is the total number of state variables, and N_o is the total number of observed variables.

As it will be required later to finalize the derivation of the lognormal Kalman filter, it can be shown that the Hessian matrix of (20) is

$$\nabla_{\mathbf{x}}^2 J(\mathbf{x}) = \mathbf{W}_b^{-\mathrm{T}} \mathbf{P}_{fl}^{-1} \mathbf{W}_b^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{W}_o^{-\mathrm{T}} \mathbf{R}_l^{-1} \mathbf{W}_o^{-1} \mathbf{H}. \quad (24)$$

To compare later with the analysis covariance matrix, it is useful to write down a scaled Hessian matrix as

$$\mathbf{W}_b^{\mathrm{T}} \nabla_{\mathbf{x}}^2 J(\mathbf{x}) \mathbf{W}_b = \mathbf{P}_{fl}^{-1} + \mathbf{W}_b^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{W}_o^{-\mathrm{T}} \mathbf{R}_l^{-1} \mathbf{W}_o^{-1} \mathbf{H} \mathbf{W}_b. \quad (25)$$

An important feature to notice about (25) is that the Hessian matrix is positive definite because of the following short proof. Let $\mathbf{A}$ be a positive definite matrix, defined by $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$, $\forall \mathbf{x} \neq 0$, and it is known that $\mathbf{A}^{-1}$ is also positive definite. Now let $\mathbf{B} = \mathbf{W}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{W}$, and consider the scalar expression $\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{W}\mathbf{x} = (\mathbf{W}\mathbf{x})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{W}\mathbf{x}) = 0$. Because we know that $\mathbf{A}^{-1}$ is positive definite, then the only way this expression can be 0 is if $\mathbf{W}\mathbf{x}$ is equal to $\mathbf{0}$, but if we assume that $\mathbf{W}$ has full column rank, then this implies that $\mathbf{x} = \mathbf{0}$. This then implies that $\mathbf{B} = \mathbf{W}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{W}$ must be positive definite.

Returning to (24) we have that the diagonal matrices multiplying $\mathbf{P}_f^{-1}$ are of full column rank, which from the proof above, implies that the first term in (24) is positive definite. Turning to the second term, we know that $\mathbf{W}_0$ has full column rank as it is a diagonal matrix. With respect to $\mathbf{H}$, we know that there has to be a sensitivity to at least one state variable, and that the same observation is not assimilated twice; therefore, $\mathbf{H}$ must have full rank as well. Thus the second term in (24) is also positive definite. Given that the sum of two positive definite matrices is also a positive definite matrix, then the Hessian matrix must be positive definite. By these same arguments, the scaled Hessian matrix in (25) must also be positive definite.

As with the minimization of the Gaussian cost function in 3DVAR, setting the gradient of (20) equal to zero, and calling the state that achieves this the analysis state, denoted by $\mathbf{x}_a$, yields

$$\nabla_{\mathbf{x}} J(\mathbf{x}_a) = \mathbf{W}_b^{-\mathrm{T}} \mathbf{P}_{fl}^{-1} (\ln \mathbf{x}_a - \ln \mathbf{x}_b) - \mathbf{H}^{\mathrm{T}} \mathbf{W}_o^{-\mathrm{T}} \mathbf{R}_l^{-1} [\ln \mathbf{y} - \ln \mathbf{h}(\mathbf{x}_a)] = 0. \quad (26)$$

The next step in deriving the lognormal version of the Kalman filter equations is to state that we wish to associate the analysis state with the observations in terms of some form of lognormal-based Kalman gain matrix $\mathbf{K}_l$. Thus, we require

$$\ln \mathbf{x}_a - \ln \mathbf{x}_b = \mathbf{K}_l [\ln \mathbf{y} - \ln \mathbf{h}(\mathbf{x}_a)]. \quad (27)$$

The reason for the form in (27) is that applying a logarithm to the lognormal random variables results in Gaussian random variables, thus, one is free to consider a linear combination of these new variables, similar to the original Gaussian approach. This form also enables us to form the analysis covariance matrix that is consistent for a lognormal distribution.

Next we introduce the logarithmic geometric tangent linear approximation (Fletcher and Jones 2014), which enables the logarithm of the model fields to be operated on by the Jacobian of the observation operator $\mathbf{h}(\mathbf{x})$. The starting point is to consider the numerical geometric derivative of $\ln \mathbf{h}_j(\mathbf{x})$, for a component x_i, with a multiplicative increment Δx_i, such that

$$\frac{\ln \mathbf{h}_j(x_i \Delta x_i) - \ln \mathbf{h}_j(x_i)}{x_i(\Delta x_i - 1)} \approx \frac{1}{\mathbf{h}_j(x_i)} \frac{d\mathbf{h}_j(x_i)}{dx_i}. \quad (28)$$

Multiplying and dividing (28) by $[\ln(x_i \Delta x_i) - \ln x_i]$ and interchanging the denominators on the left-hand side, yields

$$\frac{\ln \mathbf{h}_j(x_i \Delta x_i) - \ln \mathbf{h}_j(x_i)}{\ln(x_i \Delta x_i) - \ln x_i} \frac{\ln(x_i \Delta x_i) - \ln x_i}{x_i(\Delta x_i - 1)} \approx \frac{1}{\mathbf{h}_j(x_i)} \frac{d\mathbf{h}_j(x_i)}{dx_i}. \quad (29)$$

Taking the limit of the second factor on the left-hand side of (29) with respect to $\Delta x_i \to 1$, and then multiplying by its inverse yields

$$\frac{\ln \mathbf{h}_j(x_i \Delta x_i) - \ln \mathbf{h}_j(x_i)}{\ln(x_i \Delta x_i) - \ln x_i} \approx \frac{1}{\mathbf{h}_j(x_i)} \frac{d\mathbf{h}_j(x_i)}{dx_i} x_i.$$

Rearranging the expression above yields

$$\ln \mathbf{h}_j(x_i \Delta x_i) - \ln \mathbf{h}_j(x_i) \approx \frac{1}{\mathbf{h}_j(x_i)} \frac{d\mathbf{h}_j(x_i)}{dx_i} x_i [\ln(x_i \Delta x_i) - \ln x_i], \quad (30)$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, N_o$. Collecting all of the different components for i and j results in the following general matrix-vector geometric tangent linear approximation to $\mathbf{h}(\mathbf{x} \quad \Delta\mathbf{x})$ as

$$\ln \mathbf{h}(\mathbf{x} \quad \Delta\mathbf{x}) - \ln \mathbf{h}(\mathbf{x}) \approx \mathbf{W}_o^{-1} \mathbf{H} \mathbf{W}_b [\ln(\mathbf{x} \quad \Delta\mathbf{x}) - \ln \mathbf{x}]. \quad (31)$$

Returning to the derivation of the lognormal Kalman gain matrix, we substitute $\mathbf{x}_a$ for $\mathbf{x} \quad \Delta\mathbf{x}$ and $\mathbf{x}_b$ for $\mathbf{x}$, along with the assumption that the analysis error, or $\Delta\mathbf{x}$, is close to one, which is consistent with the assumption made for the tangent linear approximation used in incremental variational data assimilation (Fletcher and Jones 2014; Fletcher 2017). Thus, we have that

$$\ln \mathbf{h}(\mathbf{x}_a) \approx \ln \mathbf{h}(\mathbf{x}_b) + \mathbf{W}_o^{-1} \mathbf{H} \mathbf{W}_b (\ln \mathbf{x}_a - \ln \mathbf{x}_b). \quad (32)$$

Now substituting (32) into (26) results in

$$\mathbf{W}_b^{-\mathrm{T}}\mathbf{P}_{fl}^{-1}(\ln\mathbf{x}_a - \ln\mathbf{x}_b) - \mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x}_b) - \mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b(\ln\mathbf{x}_a - \ln\mathbf{x}_b)] = 0. \quad (33)$$

Pre-multiplying (33) by $\mathbf{W}_b^{\mathrm{T}}$ and factorizing yields

$$(\mathbf{P}_{fl}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)(\ln\mathbf{x}_a - \ln\mathbf{x}_b) = \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x}_b)]. \quad (34)$$

Pre-multiplying (34) by the inverse of the first matrix factor results in a format as desired in (27), given by

$$(\ln\mathbf{x}_a - \ln\mathbf{x}_b) = (\mathbf{P}_{fl}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)^{-1} \times \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x}_b)], \quad (35)$$

and as such the lognormal Kalman gain matrix is of a similar form to the Gaussian version, but now containing the derivatives of the logarithms, given by

$$\mathbf{K}_l \equiv (\mathbf{P}_{fl}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)^{-1}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}, \equiv \mathbf{P}_{fl}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}(\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\mathbf{P}_{fl}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}} + \mathbf{R}_l)^{-1}. \quad (36)$$

To obtain the expression in (36) involves a double application of the Sherman–Morrison–Woodby formula. A proof for the Gaussian equivalent from the Physical Space Assimilation System (PSAS) can be found on page 755 in Fletcher (2017), but to obtain the expression above involves scaling by $\mathbf{W}_b$ and $\mathbf{W}_o^{-1}$. An important feature to note here is that it is required that $\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\mathbf{P}_{fl}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}} + \mathbf{R}_l$ is positive definite. From the argument presented earlier we know that the observational Jacobian matrix multiplied by the $\mathbf{W}$ matrices is still of full rank, and as such this implies that the expression above is positive definite.

It is the latter expression of the lognormal Kalman gain matrix $\mathbf{K}_l$ that will be used in the next step, which is to derive the lognormal equivalent of the analysis error covariance matrix. The starting point here is to take the lognormal analysis error definition from (3) and use (27) to substitute $\ln\mathbf{x}_a$. Thus, we have

$$\ln\boldsymbol{\varepsilon}_a = \ln\mathbf{x}_b - \ln\mathbf{x}_t + \mathbf{K}_l[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x}_b)]. \quad (37)$$

To obtain a similar form to that derived for the Gaussian approach, $\ln\mathbf{h}(\mathbf{x}_b)$ must be rewritten in terms of the true state and the background error. This is achieved through using the geometric tangent linear approximation (31) with $\mathbf{x}_b = \mathbf{x}_t \ \ \boldsymbol{\varepsilon}_{bl}$, which results in

$$\ln\mathbf{h}(\mathbf{x}_b) \approx \ln\mathbf{h}(\mathbf{x}_t) + \mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\ln\boldsymbol{\varepsilon}_{bl}. \quad (38)$$

Substituting (38) into (37) yields

$$\ln\boldsymbol{\varepsilon}_a = \ln\boldsymbol{\varepsilon}_{bl} + \mathbf{K}_l[\ln\mathbf{y} - \ln\mathbf{h}(\mathbf{x}_t) - \mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\ln\boldsymbol{\varepsilon}_{bl}], = (\mathbf{I} - \mathbf{K}_l\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)\ln\boldsymbol{\varepsilon}_{bl} + \mathbf{K}_l\ln\boldsymbol{\varepsilon}_{ol}. \quad (39)$$

To simplify the appearance of the derivation we now define $\mathbf{H} \equiv \mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b$. To form the lognormal analysis error covariance matrix, we have to take the expectation of $\ln\boldsymbol{\varepsilon}_a(\ln\boldsymbol{\varepsilon}_a)^{\mathrm{T}}$, which is

$$\mathbf{P}_{al} \equiv E[\ln\boldsymbol{\varepsilon}_a(\ln\boldsymbol{\varepsilon}_a)^{\mathrm{T}}], = (\mathbf{I} - \mathbf{K}_l\mathbf{H})E[\ln\boldsymbol{\varepsilon}_{bl}(\ln\boldsymbol{\varepsilon}_{bl})^{\mathrm{T}}](\mathbf{I} - \mathbf{K}_l\mathbf{H})^{\mathrm{T}} + \mathbf{K}_lE[\ln\boldsymbol{\varepsilon}_{ol}(\ln\boldsymbol{\varepsilon}_{ol})^{\mathrm{T}}]\mathbf{K}_l^{\mathrm{T}}, = \mathbf{I}(\mathbf{I} - \mathbf{K}_l\mathbf{H})\mathbf{P}_{fl}(\mathbf{I} - \mathbf{K}_l\hat{\mathbf{H}})^{\mathrm{T}} + \mathbf{K}_l\mathbf{R}_l\mathbf{K}_l^{\mathrm{T}}. \quad (40)$$

Following the same expansion of the products in (40) as for the Gaussian case, and noticing that the lognormal Kalman gain equation can be written as

$$\mathbf{K}_l = \mathbf{P}_{fl}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}_{fl}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_l]^{-1},$$

results in (40) becoming

$$\mathbf{P}_{al} = (\mathbf{I} - \mathbf{K}_l\mathbf{H})\mathbf{P}_{fl}. \quad (41)$$

We now show that the analysis error covariance matrix is equivalent to the inverse of the scaled Hessian matrix in (25). The first step is to expand $\mathbf{K}_l$ in (41) in terms of $\mathbf{H}$, and using the rule for the inverse of the product of two matrices, yielding

$$\mathbf{P}_{al} = \mathbf{P}_{fl} - \mathbf{P}_{fl}\mathbf{H}^{\mathrm{T}}\mathbf{R}_l^{-1}[\mathbf{H}\mathbf{P}_{fl}\mathbf{H}^{\mathrm{T}}\mathbf{R}_l^{-1} + \mathbf{I}]^{-1}\mathbf{H}\mathbf{P}_{fl}. \quad (42)$$

Next recalling the Sherman–Morrison–Woodbury formula:

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^{\mathrm{T}})^{-1} \equiv \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}[\mathbf{I} + \mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{U}]^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1},$$

where for (42) this implies that $\mathbf{A} = \mathbf{P}_{fl}^{-1}$, $\mathbf{U} = \mathbf{H}^{\mathrm{T}}\mathbf{R}_l^{-1}$, and $\mathbf{V} = \mathbf{H}^{\mathrm{T}}$, the lognormal analysis error covariance matrix can be rewritten as

$$\mathbf{P}_{al} = (\mathbf{P}_{fl}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}_l^{-1}\mathbf{H})^{-1}, = (\mathbf{P}_{fl}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_l^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)^{-1}, \quad (43)$$

where the expression inside the brackets on the right-hand side of (43) is the same as that of the scaled Hessian matrix of the cost function in (25).

The final part of this derivation is to show that the expression for the lognormal Kalman gain matrix minimizes the trace of the analysis error covariance matrix. Thus differentiating (40) with respect to $\mathbf{K}_l$ and setting to zero yields

$$\frac{\partial\mathbf{P}_{al}}{\partial\mathbf{K}_l} = -2(\mathbf{I} - \mathbf{K}_l\mathbf{H})\mathbf{P}_{fl}\mathbf{H}^{\mathrm{T}} + 2\mathbf{K}_l\mathbf{R}_l = \mathbf{0}, \Rightarrow \mathbf{P}_f\mathbf{H}^{\mathrm{T}} = \mathbf{K}_l(\mathbf{H}\mathbf{P}_f\mathbf{H}^{\mathrm{T}} + \mathbf{R}_l), \Rightarrow \mathbf{K}_l = \mathbf{P}_f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}_f\mathbf{H}^{\mathrm{T}} + \mathbf{R}_l)^{-1}. \quad (44)$$

It is clear that the expression for the lognormal-based Kalman gain matrix in (36) is the same as that in (44), and thus the lognormal Kalman gain matrix minimizes in a least squares sense the analysis error covariances.

Thus in summary the equations for the analysis step of the lognormal Kalman filter are given by

$$\mathbf{P}_{fl}^{n} = [\ln M(\mathbf{x}_{a}^{n-1} \quad \boldsymbol{\varepsilon}_{a}^{n-1}) - \ln \mathbf{x}_{t}^{n}] \times [\ln M(\mathbf{x}_{a}^{n-1} \quad \boldsymbol{\varepsilon}_{a}^{n-1}) - \ln \mathbf{x}_{t}^{n}]^{\mathrm{T}} + \mathbf{Q}_{l}^{n}, \tag{45}$$

$$\ln \mathbf{x}_{a}^{n} = \ln \mathbf{x}_{b}^{n} + \mathbf{K}_{l}^{n}[\ln \mathbf{y} - \ln \mathbf{h}(\mathbf{x}_{a})], \tag{46}$$

$$\mathbf{K}_{l}^{n} = \mathbf{P}_{fl}^{n}\mathbf{W}_{b}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_{o}^{-\mathrm{T}}[\mathbf{W}_{o}^{-1}\mathbf{H}\mathbf{W}_{b}\mathbf{P}_{fl}^{n}\mathbf{W}_{b}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_{o}^{-\mathrm{T}} + \mathbf{R}_{l}]^{-1}, \tag{47}$$

$$\mathbf{P}_{al}^{n} = (\mathbf{I} - \mathbf{K}_{l}^{n}\mathbf{W}_{o}^{-1}\mathbf{H}\mathbf{W}_{b})\mathbf{P}_{fl}^{n}, \tag{48}$$

$$\mathbf{x}_{b}^{n} = M(\mathbf{x}_{a}^{n-1}). \tag{49}$$

As shown with the development of the lognormal forms of variational data assimilation, we do not live in an one type of distribution only world, and as such the next step is to combine the lognormal Kalman filter just derived with the Gaussian Kalman filter to be able to use the mixed Gaussian–lognormal distribution from Fletcher and Zupanski (2006b) to form a second non-Gaussian-based Kalman filter system.

c. Mixed Gaussian–lognormal Kalman filter

In this section we shall refer to the mixed Gaussian–lognormal Kalman filter as MXKF. The starting point for the derivation of the MXKF is the definition of the associated background, observational, model, and analysis errors. As we are assuming that the error that is to be minimized is from a mixed distribution, it implies that there are a set of Gaussian distributed errors and a set of lognormal distributed errors that need to be minimized simultaneously. This then implies that the true state, background, and analysis states are given by

$$\mathbf{x}_{tmx} \equiv \begin{matrix} \mathbf{x}_{tG} \\ \mathbf{x}_{tl} \end{matrix}, \quad \mathbf{x}_{bmx} \equiv \begin{matrix} \mathbf{x}_{bG} \\ \mathbf{x}_{bl} \end{matrix}, \quad \mathbf{x}_{amx} \equiv \begin{matrix} \mathbf{x}_{aG} \\ \mathbf{x}_{al} \end{matrix}, \tag{50}$$

where G represents the Gaussian distributed random variables, and l represents the lognormally distributed random variables, which implies that the associated mixed distributed errors are given by

$$\boldsymbol{\varepsilon}_{bmx}^{n} \equiv \begin{matrix} \mathbf{x}_{bG}^{n} - \mathbf{x}_{tG}^{n} \\ \ln \mathbf{x}_{bl}^{n} - \ln \mathbf{x}_{tl}^{n} \end{matrix}, \quad \boldsymbol{\varepsilon}_{amx} \equiv \begin{matrix} \mathbf{x}_{aG}^{n} - \mathbf{x}_{tG}^{n} \\ \ln \mathbf{x}_{al}^{n} - \ln \mathbf{x}_{tl}^{n} \end{matrix},$$

$$\boldsymbol{\varepsilon}_{omx} \equiv \begin{matrix} \mathbf{y}_{G} - \mathbf{h}_{G}(\mathbf{x}_{tmx}^{n}) \\ \ln \mathbf{y}_{l} - \ln \mathbf{h}_{l}(\mathbf{x}_{tmx}^{n}) \end{matrix}, \quad \boldsymbol{\varepsilon}_{mmx} \equiv \begin{Bmatrix} [M(\mathbf{x}_{tmx}^{n-1})]_{G} - \mathbf{x}_{tG}^{n} \\ \ln[M(\mathbf{x}_{tmx}^{n-1})]_{l} - \ln \mathbf{x}_{tl}^{n} \end{Bmatrix}. \tag{51}$$

It should be noted here that the number of Gaussian observation errors will, in most circumstances, be less than the dimension of the Gaussian component of the true state. This is also true for the lognormal distributed errors. Finally there may not be an equal number of Gaussian and lognormal background, or observational, errors

The mixed distribution-based forecast error covariance matrix can be shown to be

$$\mathbf{P}_{fmx}^{n} \equiv \begin{matrix} M(\boldsymbol{\varepsilon}_{aG}^{n-1}) & M(\boldsymbol{\varepsilon}_{aG}^{n-1}) \\ \ln M(\boldsymbol{\varepsilon}_{al}^{n-1}) & \ln M(\boldsymbol{\varepsilon}_{al}^{n-1}) \end{matrix}^{\mathrm{T}}, \tag{52}$$

where it can clearly be seen that there are covariances between the Gaussian and the lognormal forecast errors.

The next step is to define the equivalent cost function from the lognormal median approach to find the median of the mixed distribution, which is given by

$$J_{mx}(\mathbf{x}) = \frac{1}{2} \begin{matrix} \mathbf{x}_{tG} - \mathbf{x}_{bG} \\ \ln \mathbf{x}_{tl} - \ln \mathbf{x}_{bl} \end{matrix}^{\mathrm{T}} \mathbf{P}_{fmx}^{-1} \begin{matrix} \mathbf{x}_{tG} - \mathbf{x}_{bG} \\ \ln \mathbf{x}_{tl} - \ln \mathbf{x}_{bl} \end{matrix} + \frac{1}{2} \begin{matrix} \mathbf{y}_{G} - \mathbf{h}_{G}(\mathbf{x}_{t}) \\ \ln \mathbf{y}_{l} - \ln \mathbf{h}_{l}(\mathbf{x}_{t}) \end{matrix}^{\mathrm{T}} \mathbf{R}_{mx}^{-1} \begin{matrix} \mathbf{y}_{G} - \mathbf{h}_{G}(\mathbf{x}_{t}) \\ \ln \mathbf{y}_{l} - \ln \mathbf{h}_{l}(\mathbf{x}_{t}) \end{matrix}, \tag{53}$$

where the Jacobian of (53) can be shown to be

$$\nabla_{\mathbf{x}_{t}} J(\mathbf{x}) = \mathbf{W}_{b}^{-\mathrm{T}}\mathbf{P}_{fmx}^{-1} \begin{matrix} \mathbf{x}_{tG} - \mathbf{x}_{bG} \\ \ln \mathbf{x}_{tl} - \ln \mathbf{x}_{bl} \end{matrix} - \mathbf{H}^{\mathrm{T}}\mathbf{W}_{o}^{-\mathrm{T}}\mathbf{R}_{mx}^{-1} \begin{matrix} \mathbf{y}_{G} - \mathbf{h}_{G}(\mathbf{x}_{t}) \\ \ln \mathbf{y}_{l} - \ln \mathbf{h}_{l}(\mathbf{x}_{t}) \end{matrix}, \tag{54}$$

and the associated Hessian and scaled Hessian matrices are given by

$$\nabla_{\mathbf{x}_{t}}^{2} J(\mathbf{x}) = \mathbf{W}_{b}^{-\mathrm{T}}\mathbf{P}_{fmx}^{-1}\mathbf{W}_{b}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{W}_{o}^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_{0}^{-1}\mathbf{H},$$

$$\Rightarrow \mathbf{W}_{b}^{\mathrm{T}}\nabla_{\mathbf{x}_{t}}^{2} J(\mathbf{x})\mathbf{W}_{b} = \mathbf{P}_{fmx}^{-1} + \mathbf{W}_{b}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_{o}^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_{0}^{-1}\mathbf{H}\mathbf{W}_{b}, \tag{55}$$

where

$$\mathbf{W}_{b}^{-1} \equiv \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & x_{tl_1}^{-1} & & \\ & & & & \ddots & \\ & & & & & x_{tl_N}^{-1} \end{pmatrix},$$

$$\mathbf{W}_{o}^{-1} \equiv \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \mathbf{h}_{l_1}(\mathbf{x}_{t}) & & \\ & & & & \ddots & \\ & & & & & \mathbf{h}_{l_{N_o}}(\mathbf{x}_{t}) \end{bmatrix}.$$

The next stage is to form the analysis errors in terms of the background state combined with a weighting of the observations. Thus, for the mixed distribution approach this should be of the following form:

$$\begin{matrix} \mathbf{x}_{aG} - \mathbf{x}_{bG} \\ \ln \mathbf{x}_{al} - \ln \mathbf{x}_{bl} \end{matrix} = \mathbf{K}_{mx} \begin{matrix} \mathbf{y}_{G} - \mathbf{h}_{G}(\mathbf{x}_{a}) \\ \ln \mathbf{y}_{l} - \ln \mathbf{h}_{l}(\mathbf{x}_{a}) \end{matrix}. \tag{56}$$

As with the lognormal derivation, we introduce the geometric tangent linear approximations to the lognormal observation

operators and the additive tangent linear model to the Gaussian observation operators into the gradient of the mixed distribution cost function and set to zero. This results in

$$0 = \mathbf{W}_b^{\mathrm{T}}\mathbf{P}_{fmx}^{-1}\begin{matrix}\mathbf{x}_{aG}-\mathbf{x}_{bG}\\ \ln\mathbf{x}_{al}-\ln\mathbf{x}_{bl}\end{matrix} - \mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\begin{matrix}\mathbf{y}_G-\mathbf{h}_G(\mathbf{x}_b)\\ \ln\mathbf{y}_l-\ln\mathbf{h}_l(\mathbf{x}_b)\end{matrix} - \mathbf{H}\mathbf{W}_o^{-1}\mathbf{W}_b\begin{matrix}\mathbf{x}_{aG}-\mathbf{x}_{bG}\\ \ln\mathbf{x}_{al}-\ln\mathbf{x}_{bl}\end{matrix}. \tag{57}$$

Factorizing (57) results in

$$[\mathbf{P}_{fmx}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b]\begin{matrix}\mathbf{x}_{aG}-\mathbf{x}_{bG}\\ \ln\mathbf{x}_{al}-\ln\mathbf{x}_{bl}\end{matrix} = \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{T}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\begin{matrix}\mathbf{y}_G-\mathbf{h}_G(\mathbf{x}_b)\\ \ln\mathbf{y}_l-\ln\mathbf{h}_l(\mathbf{x}_b)\end{matrix}. \tag{58}$$

Therefore, for the mixed distribution approach one finds

$$\begin{matrix}\mathbf{x}_{aG}-\mathbf{x}_{bG}\\ \ln\mathbf{x}_{al}-\ln\mathbf{x}_{bl}\end{matrix} = [\mathbf{P}_{fmx}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{T}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b]^{-1} \times \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\begin{matrix}\mathbf{y}_G-\mathbf{h}_G(\mathbf{x}_b)\\ \ln\mathbf{y}_l-\ln\mathbf{h}_l(\mathbf{x}_b)\end{matrix}. \tag{59}$$

Thus, the Kalman gain matrix for the mixed Gaussian–lognormal approach is

$$\mathbf{K}_{mx} \equiv [\mathbf{P}_{fmx}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b]^{-1}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}. \tag{60}$$

Through applying the Sherman–Morrison–Woodbury formula twice it is possible to write (60) in the more usable form,

$$\mathbf{K}_{mx} \equiv \mathbf{P}_{fmx}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}[\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\mathbf{P}_{fmx}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}} + \mathbf{R}_{mx}]^{-1}, \tag{61}$$

where the proof of the expression above can be found through using the derivation on page 755 from Fletcher (2017), and the proof of positive definiteness follows from the arguments from the lognormal Kalman filter analysis error derivation.

As with the lognormal approach, we shall introduce some notation to simplify the appearance of the derivation for the analysis error covariance matrix. We shall denote $\mathbf{H} \equiv \mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b$. Using the definition of the analysis errors for the mixed distribution from (51), the different forms of tangent linear approximations presented earlier, as well as the standard version from the Gaussian formulation, results in the mixed distribution analysis errors of the following form:

$$\begin{matrix}\boldsymbol{\varepsilon}_{aG}\\ \ln\boldsymbol{\varepsilon}_{al}\end{matrix} = [\mathbf{I}-\mathbf{K}_{mx}\mathbf{H}^{\mathrm{T}}]\begin{matrix}\boldsymbol{\varepsilon}_{bG}\\ \ln\boldsymbol{\varepsilon}_{bl}\end{matrix} + \mathbf{K}_{mx}\begin{matrix}\boldsymbol{\varepsilon}_{oG}\\ \ln\boldsymbol{\varepsilon}_{ol}\end{matrix}. \tag{62}$$

Thus, forming the product of the analysis error vector with its transpose and taking the expectation results in

$$\mathbf{P}_{amx} = [\mathbf{I}-\mathbf{K}_{mx}\mathbf{H}^{\mathrm{T}}]\mathbf{P}_{fmx}[\mathbf{I}-\mathbf{K}_{mx}\mathbf{H}^{\mathrm{T}}]^{\mathrm{T}} + \mathbf{K}_{mx}\mathbf{R}_{mx}\mathbf{K}_{mx}^{\mathrm{T}}, \tag{63}$$

where following the same arguments for the Gaussian and lognormal cases results in the analysis error covariance matrix of the following form:

$$\mathbf{P}_{amx} = [\mathbf{I}-\mathbf{K}_{mx}\mathbf{H}^{\mathrm{T}}]\mathbf{P}_{fmx}. \tag{64}$$

The final step is to confirm that the inverse of the analysis error covariance matrix is equivalent to the scaled Hessian of (53) in (55) which can easily be shown as the expression above is the same in appearance as the standard Gaussian and the lognormal version from the last section. Therefore, the analysis error covariance matrix for the mixed distribution approach is given by

$$\mathbf{P}_{amx} = (\mathbf{P}_{fmx}^{-1} + \mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}}\mathbf{R}_{mx}^{-1}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b)^{-1}. \tag{65}$$

Thus, in summary, the mixed Gaussian–lognormal-based Kalman filter equations are given by

$$\mathbf{P}_{fmx}^{n} = \begin{matrix}M(\boldsymbol{\varepsilon}_{aG}^{n-1}) & M(\boldsymbol{\varepsilon}_{aG}^{n-1})\\ \ln M(\boldsymbol{\varepsilon}_{al}^{n-1}) & \ln M(\boldsymbol{\varepsilon}_{al}^{n-1})\end{matrix}^{\mathrm{T}} + \mathbf{Q}_{mx}^{n}, \tag{66}$$

$$\begin{matrix}\mathbf{x}_{aG}^{n}\\ \ln\mathbf{x}_{al}^{n}\end{matrix} = \begin{matrix}\mathbf{x}_{bG}^{n}\\ \ln\mathbf{x}_{bl}^{n}\end{matrix} + \mathbf{K}_{mx}^{n}\begin{matrix}\mathbf{y}_G-\mathbf{h}_G(\mathbf{x}_b^{n})\\ \ln\mathbf{y}_l-\ln\mathbf{h}_l(\mathbf{x}_b^{n})\end{matrix}, \tag{67}$$

$$\mathbf{K}_{mx}^{n} = \mathbf{P}_{fmx}^{n}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{T}\mathbf{W}_o^{-\mathrm{T}}[\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b\mathbf{P}_{fmx}^{n}\mathbf{W}_b^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}_o^{-\mathrm{T}} + \mathbf{R}_{mx}]^{-1}, \tag{68}$$

$$\mathbf{P}_{amx}^{n} = [\mathbf{I}-\mathbf{K}_{mx}^{n}\mathbf{W}_o^{-1}\mathbf{H}\mathbf{W}_b]\mathbf{P}_{fmx}^{n}, \tag{69}$$

$$\begin{matrix}\mathbf{x}_{bG}^{n}\\ \mathbf{x}_{bl}^{n}\end{matrix} = M\begin{matrix}\mathbf{x}_{aG}^{n-1}\\ \mathbf{x}_{al}^{n-1}\end{matrix}. \tag{70}$$

In this section it has been shown that it is possible to derive a nonlinear version of the Kalman filter equations to be used with lognormal random variables, as well as with a combination of lognormal and Gaussian random variables. The appearance of the set of equations are similar to the Gaussian form, but the evolution of the analysis error covariance matrix is exact and not through the application of the linearized model.

In the next section the mixed Gaussian–lognormal Kalman filter equations will be tested against the original linearized Gaussian Kalman filter equations, referred to as the extended Kalman filter (EKF) with the Lorenz 1963 model. It has been shown in the development and testing of the mixed Gaussian–lognormal variational data assimilation systems that the z component of this model is highly non-Gaussian, and as such is a good test to assess the performance of these new equations.

4. Experiments with the Lorenz 1963 model

As just mentioned, the Lorenz 1963 model has been used extensively to test the development of the mixed Gaussian–

lognormal-based variational data assimilation systems. An important feature of the Lorenz 1963 model is that there are regions of the Lorenz attractor where the z component does not follow a Gaussian distribution, as has been shown in Fletcher (2010) and Goodliff et al. (2020). This component is always positive, while the x and y components have positive and negative values. In Fletcher (2010) there are four climatologies of the z component that are created from the Lorenz 1963 model with the initial conditions that we shall define soon, where it was clear that after 100 000 time steps this component appeared to have a global mode with a skewness to the left and then a secondary mode with a smaller occurrence rate than the global mode. When fitting a lognormal distribution to this data the global mode is very well captured with a slight underestimation of the secondary mode. When a Gaussian distribution was fitted to this data, both modes were underestimated and the Gaussian mode was in between the two modes and was assigning higher probabilities to states that did not occur that often. See Fig. 21.3 in Fletcher (2017) for this example.

This model is also a good choice due to its simplicity for a dynamical model that exhibits chaotic behavior. Another important property of this model is that it is very sensitive to the initial conditions from which it starts, and as such can give very different answers even by being out by a few decimal places from the *true* state. For an example of this sensitivity see Fletcher (2017). The continuous model equations are as given by

$$\frac{dx}{dt} = -\sigma(x - y), \tag{71}$$

$$\frac{dy}{dt} = \rho x - y - zx, \tag{72}$$

$$\frac{dz}{dt} = xy - \beta z, \tag{73}$$

where $x = x(t)$, $y = y(t)$, and $z = z(t)$ are the state variables, $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ are parameters.

The experiments will compare the analysis errors from the EKF, which uses the linearized model, against those from the MXKF, that uses the full nonlinear model. The two filters will be tested with different observational errors, where it is assumed that x and y components have Gaussian errors and the z component has lognormal errors. The observations are generated with different observational error variances, and with different times between analysis updates to determine the robustness of the new approach.

In this section we shall look at the sensitivity of the EKF and the MXKF to both observational error variance size, as well as time between observations. The numerical scheme that is used for the discretization of the nonlinear model is the second-order explicit Runge–Kutta scheme. The MXKF utilizes the nonlinear numerical model for all three components, while the EKF utilizes a linearized version of the model, which was calculated analytically and then discretized with the same scheme as the nonlinear model. The adjoint model was also calculated by hand. This is all coded in MATLAB. See Fletcher (2017) for more details on this calculation.

We shall consider four different configurations in this section; 1) $\sigma_o = 0.5$, 50 time steps between observations, 2) $\sigma_o = 2$, 25 time steps between observations, 3) $\sigma_o = 0.25$, 200 time steps between observations, and 4) $\sigma_o = 1$, 100 time steps between observations, where σ_o is the observational error standard deviation.

The true solution is started from the initial conditions: $x_{0t} = -5.4458$, $y_{0t} = -5.4841$, and $z_{0t} = 22.5606$, while the background solution starts from $x_{0b} = -5.9$, $y_{0b} = -5.0$, and $z_{0b} = 24.0$. These are the same set of initial conditions that have been used with the mixed Gaussian–lognormal variational data assimilation schemes. However, an extra feature that is used in these experiments, and is the same for both the EKF and the MXKF is an approximation for the model error term. We use the values that are suggested in Evensen and Fabio (1997):

$$\mathbf{Q} \equiv \begin{pmatrix} 0.1491 & 0.1505 & 0.0007 \\ 0.1505 & 0.9048 & 0.0014 \\ 0.0007 & 0.0014 & 0.9180 \end{pmatrix}. \tag{74}$$

a. Experiment 1: $\sigma_0 = 0.5$, 50 time steps between observations

In Fig. 2 we have two sets of plots, the first is of the z and x trajectories for the true states (red lines), the solution from the MXKF (blue lines), and the solution from the EKF (black lines), along with the observations (green circles). The second set of plots is of the z and x errors, where for the z component we consider the ratio to define the error, while for the x component the error is defined as the difference.

It is clear from the trajectory plots in Fig. 2 that the observations are not that accurate, but are frequent. Both solutions appear drawn toward the observations in that the error increases when the less accurate observations are assimilated compared to the more accurate ones. However, when we consider the error plots we can see that both approaches are impacted by the less than perfect observations, but that the MXKF solution is able to be more consistent for both the x and z components, as measured by the z error being close to one, and close to zero for the x component, with the MXKF scheme able to recover quicker.

b. Experiment 2: $\sigma_o = 2$, 25 time steps between observations

In this section we consider the case where we have more observations than in experiment one, but these observations are less accurate. These results are presented in Fig. 3 in the same configuration as in experiment one. We can see that while there are some quite inaccurate observations, the solutions from either approach are not able to go out of phase from the true solution. Again it is clear that the MXKF approach is able to stay more consistent than the EKF solutions.

c. Experiment 3: $\sigma_o = 0.25$, with 200 time steps between observations

In this experiment we are considering the case where we have fewer observations, but they are quite accurate. These

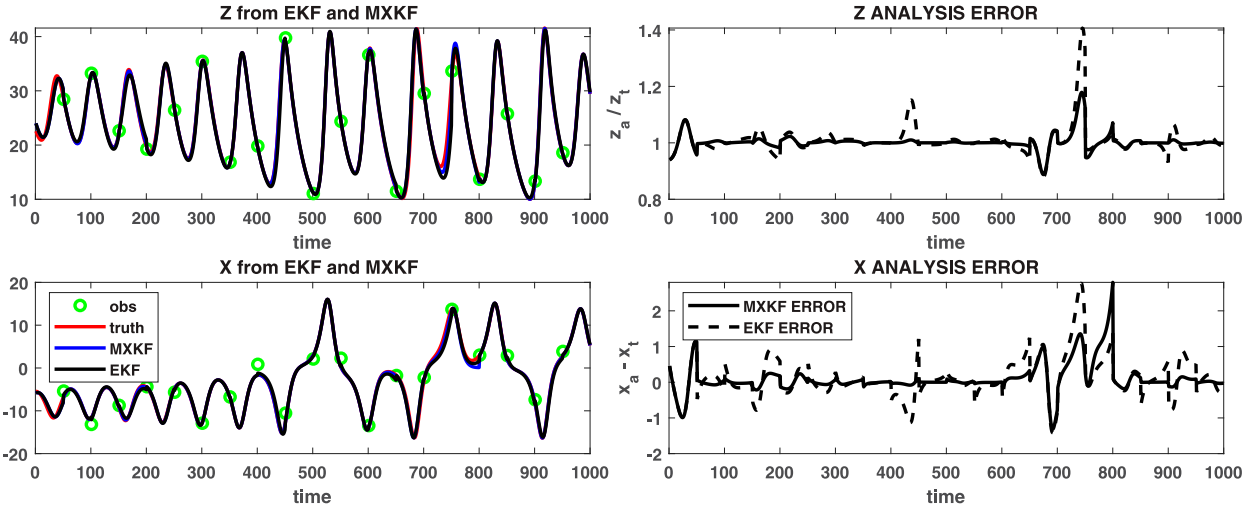

FIG. 2. (left) z and x true (red), mixed Gaussian–lognormal Kalman filter (MXKF) (blue), and extended Kalman filter (EKF) (black), observations (green circles). (right) z_a/z_t error and $x_a - x_t$ error plots for the analysis from the MXKF (solid black) and the EKF (black dashed) for $\sigma_o = 0.5$ with 50 time steps between observations.

results are presented in Fig. 4. As expected there is a decrease in the accuracy of both approaches, but we also have that neither of the solutions go out of phase. We again see that the MXKF is able to produce a more consistent solution than the EKF for both the x and z components.

d. *Experiment 4: $\sigma_o = 1$, 100 time steps between observations*

The results from this experiment are presented in Fig. 5 where we can see that we have the situation where the EKF does go out of phase with the true solution, even going on to the wrong attractor for a short while, before assimilating additional observations to bring it back toward the true solution. However, for this configuration we see that the MXKF approach does not go out of phase in neither the z nor the x component to the extreme that the EKF solution does and appears to be able to better assimilate the observations each time.

e. *Robustness testing 1: Observational error standard deviations and observational frequency*

In this subsection we present results from running experiments 1 to 4 with 5000 different random draws from the observational error distribution to test the robustness of the MXKF. To determine the robustness we calculate the analysis error as a lognormally distributed random variable, i.e., the

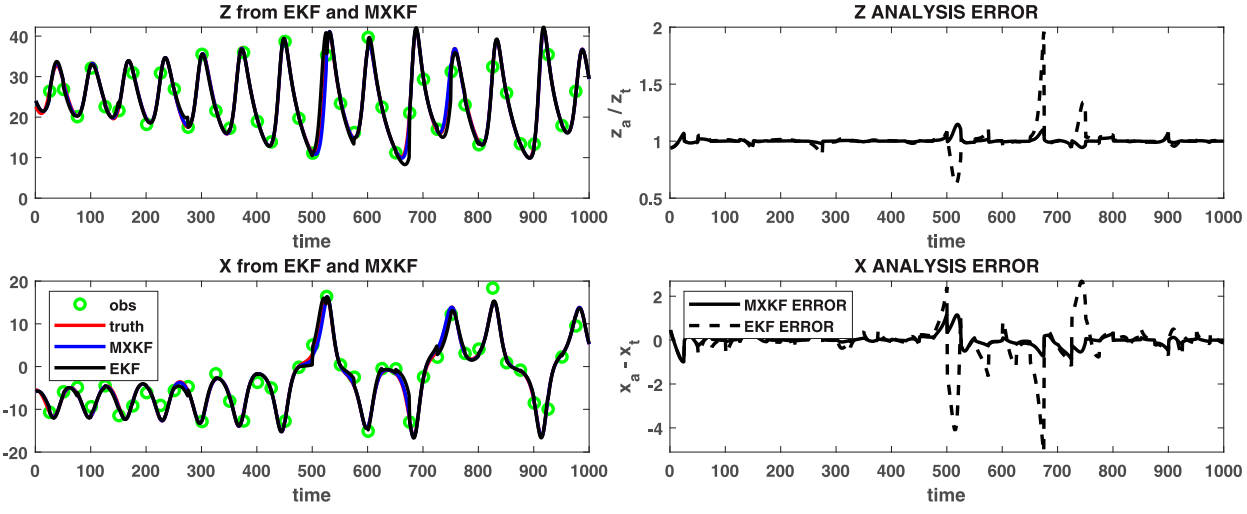

FIG. 3. (left) z and x true (red), mixed Gaussian–lognormal Kalman filter (MXKF) (blue), and extended Kalman filter (EKF) (black), observations (green circles). (right) z_a/z_t error and $x_a - x_t$ error plots for the analysis from the MXKF (solid black) and the EKF (black dashed) for $\sigma_o = 2$ with 25 time steps between observations.

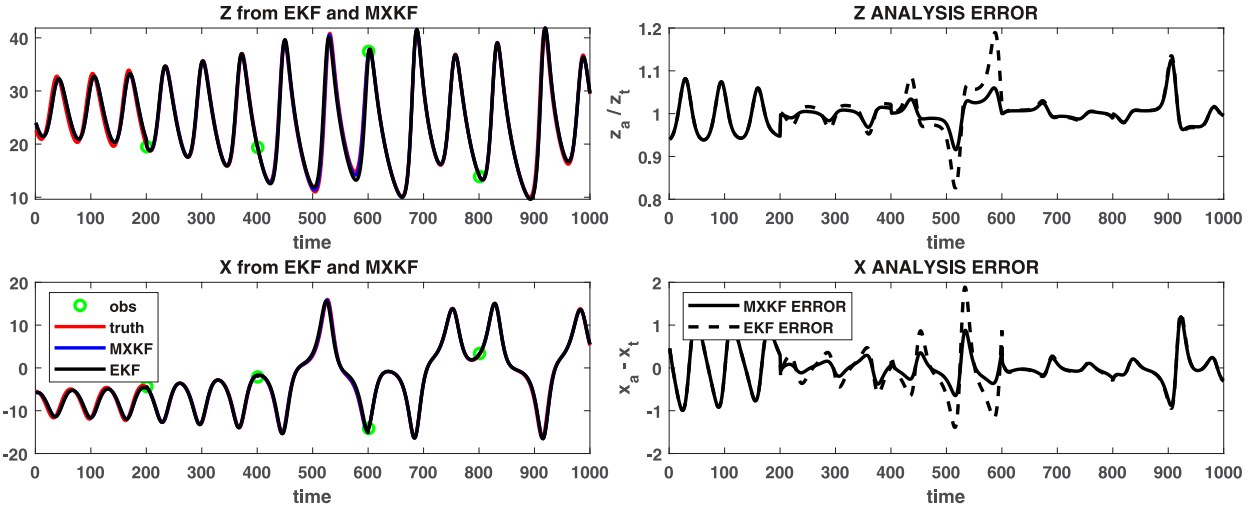

FIG. 4. (left) z and x true (red), mixed Gaussian–lognormal Kalman filter (MXKF) (blue), and extended Kalman filter (EKF) (black), observations (green circles). (right) z_a/z_t error and $x_a - x_t$ error plots for the analysis from the MXKF (solid black) and the EKF (black dashed) for $\sigma_o = 0.25$ with 200 time steps between observations.

ratio of the analysis to the true state for all 5000 solutions from the MXKF and the EKF, and from these errors we calculate the average minimum error and the average maximum error for the MXKF and the EKF. These results are summarized in Table 1 to highlight the spread from the average minimum analysis error to the maximum analysis error for the MXKF and the EKF.

We note here that during the 5000 evaluations for using experiment 3 and 4 configurations there was one instance for each where the MXKF did not converge. This aside, given that the analysis error is a ratio and if the scheme is performing well then the analysis error should be approximately equal to 1 as seen in the results in Fletcher and Jones (2014).

From the values in Table 1 it is clear that the MXKF on average for the situations considered here has a smaller spread between the average maximum and average minimum analysis error for all four experiments, bearing in mind the caveat above. It appears that the scheme performs best on average for the situation where there are inaccurate observations but more of them (experiment 2). The largest spread for the MXKF appears to be for experiment 3, accurate observations but less frequent.

f. *Robustness testing 2: Perturbing the true and background states' initial conditions*

In the results present here experiment 1 was used for the observational error standard deviation and frequency but the

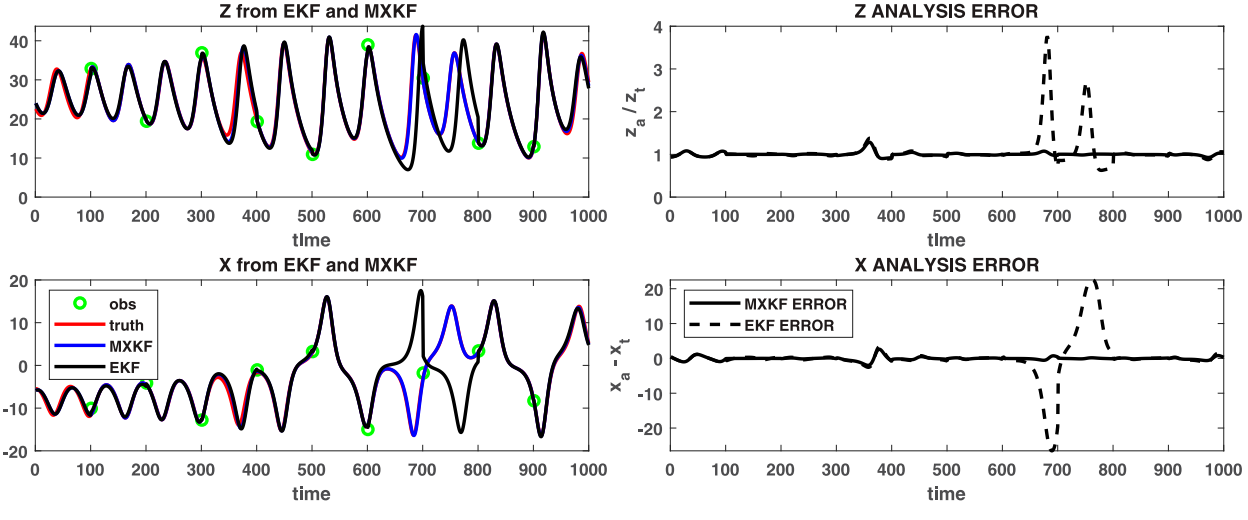

FIG. 5. (left) z and x true (red), mixed Gaussian–lognormal Kalman filter (MXKF) (blue), and extended Kalman filter (EKF) (black), observations (green circles). (right) z_a/z_t error and $x_a - x_t$ error plots for the analysis from the MXKF (solid black) and the EKF (black dashed) for $\sigma_o = 0.5$ with 50 time steps between observations.

TABLE 1. Summary of the average minimum and maximum analysis errors for experiments 1–4 over 5000 assimilation runs.

	MXKF		EKF	
Expt	Avg min	Avg max	Avg min	Avg max
1	0.836	1.25	0.723	1.432
2	0.904	1.112	0.786	1.268
3	0.662	1.403	0.527	1.556
4	0.766	1.278	0.605	1.505

initial conditions for the true state and the background state were randomly perturbed using the MATLAB function NORMRND with mean zero and three different standard deviations, $\sigma_p = 0.1, 0.5, 1$. Different perturbations were applied to the true initial conditions and the background initial conditions from those presented at the beginning of this section but were drawn from the same distribution. The performance metrics as for robustness testing 1 were applied here and are summarized in Table 2.

From Table 2 it is clear that there is a sensitivity in the MXKF to the initial conditions for the true and background state. It should be noted that the MXKF had an approximate 1% failure rate for all three configurations, while the EKF did not. As with robustness testing 1, the MXKF has a smaller spread between its average maximum and minimum analysis error compared to the EKF when it converged.

5. Conclusions and further work

In this paper we have been able to show that it is not possible to follow the linear least squares approach that is used to derive the Kalman filter, and the extended Kalman filter (EKF) equations, to derive a similar expression for lognormally distributed errors. However, we have been able to show that if we keep the nonlinear model and follow a cost function–based approach associated with the median from Fletcher (2010), then it is possible to derive a set of nonlinear equations for the update of the median of the lognormal analysis state together with its uncertainty. We were able to extend this to the mixed Gaussian–lognormal probability density function, where the associated Kalman filter equations are referred to as the MXKF.

We coded the new MXKF, along with the EKF, for the Lorenz 1963 model in MATLAB, and showed that for different configurations of observational error variances and time steps between observations, where the observational errors for the x and y components were Gaussian distributed, while for the z component these were lognormally distributed, the MXKF appeared to be more consistent with the true solutions for longer periods than the EKF. We should note that the EKF was using the linearized numerical model, while the MXKF was using the nonlinear numerical model. It appears that this has an effect on the performance of the EKF compared to the MXKF in that it appears to fit more to the observations, while the MXKF does not always pull straight to the observations.

To evaluate the general performance of the MXKF against the EKF a set of 5000 assimilation experiments was run for each of the four experimental configurations from section 4. It was shown that the MXKF had a smaller spread between the average minimum and average maximum analysis error for all four experimental configurations, but we note that there was one realization for both experiment 3 and 4 where the MXKF did not converge. This robustness test was followed up with a sensitivity study of the MXKF and the EKF to perturbed true state and background state initial conditions.

TABLE 2. Summary of the average minimum and maximum analysis errors for perturbing the true and background initial condition over 5000 assimilation runs using experiment 1's observational configuration.

	MXKF		EKF	
Expt	Avg min	Avg max	Avg min	Avg max
$\sigma_p = 0.1$	0.820	1.408	0.680	1.710
$\sigma_p = 0.5$	0.804	1.521	0.679	1.806
$\sigma_p = 1$	0.788	1.631	0.668	1.791

It has been shown in Fletcher and Jones (2014) that lognormal-based data assimilation systems can be quite sensitive to the accuracy of the observations of the different components of the Lorenz 1963 model near the transition zones between the two attractors, and it is possible that the lognormal Kalman filter could also be suffering from this here, and is left for further work to determine if this is the case.

The next step in this work is to build the theory for an ensemble-based approach to the MXKF equations and rigorously test them with different toy problems. Given the nonlinear nature of the equations, and the dependence on the equations being derived from a cost function, the most likely candidate would be the maximum likelihood ensemble filter (MLEF) from Zupanski (2005). The MLEF is comprised of two steps: the forecast step uses the standard definition of the update for the forecast error covariance matrix from the Kalman filter but uses the nonlinear model to evolve this matrix between analysis times, instead of a linear model, through an ensemble where each ensemble member's perturbation is a column from the analysis error covariance matrix from that assimilation cycle. The analysis step is to solve a flow dependent 3D VAR cost function projected into ensemble space through a Hessian preconditioner. The square root analysis error covariance is updated through the inversion of the Hessian preconditioner. The steps here are easily adaptable to the new MXKF equations for the updates of the analysis and forecast error covariance matrices.

The reason behind the non-Gaussian work over the last 15 years has been to develop more consistent data assimilation systems for positive definite variables. In the atmosphere it is well know that relative humidity is positive definite, that is to say it is always larger than zero, and as such we do not wish for a data assimilation system to produce an answer that is negative or equal to zero for this field. It has been shown in Kliewer et al. (2016) that through using a mixed Gaussian–lognormal 1DVAR for a temperature–mixing ratio it is possible to obtain better fits to both the temperature and the moisture channels through the covariances between the Gaussian and

lognormal random variables. A full description of the links between the two distributions can be found in Fletcher (2017).

As most of the operational numerical weather prediction centers use a form of hybrid ensemble/variational data assimilation algorithm, it became important for the mixed Gaussian–lognormal theory to move toward that approach. However, the major stumbling block has been the Kalman filter component to create the ensemble covariance. The work in this paper is the first step toward a Gaussian–lognormal hybrid 4DVAR.

Acknowledgments. The National Science Foundation Grant AGS-1738206 at CIRA/CSU supported authors 1, 3, 4, 5, and 6, while authors 2 and 7 were supported by NOAA's Hurricane Forecast Improvement Program Award NA18NWS4680059. Funding for authors 1, 8, and 9 came from National Science Foundation Grant AGS-2033405 at CIRA/CSU. We are grateful for the extremely helpful reviews from the three anonymous reviewers and are grateful to reviewer two for the proof to show the positive definiteness of the different analysis error covariance matrices.

Data availability statement. The code that support the findings of this study is openly available at the following URL: https://mountainscholar.org/handle/10217/234474.

REFERENCES

Aseev, N. A., and Y. Y. Shprits, 2019: Reanalysis of ring current electron phase space densities using Van Allen probe observations, convection model, and log-normal Kalman filter. *Space Wea.*, **17**, 619–638, https://doi.org/10.1029/2018SW002110.

Cohn, S. E., 1997: An introduction to estimation error theory. *J. Meteor. Soc. Japan*, **75**, 257–288, https://doi.org/10.2151/jmsj1965.75.1B_257.

Evensen, G., and N. Fabio, 1997: Solving for the generalized inverse of the Lorenz model. *J. Meteor. Soc. Japan*, **75**, 229–243, https://doi.org/10.2151/jmsj1965.75.1B_229.

Fletcher, S. J., 2010: Mixed lognormal-Gaussian four-dimensional data assimilation. *Tellus*, **62A**, 266–287, https://doi.org/10.1111/j.1600-0870.2010.00439.x.

——, 2017: *Data Assimilation for the Geosciences: From Theory to Applications*. Elsevier, 976 pp.

——, and M. Zupanski, 2006a: A data assimilation method for log-normally distributed observational errors. *Quart. J. Roy. Meteor. Soc.*, **132**, 2505–2519, https://doi.org/10.1256/qj.05.222.

——, and ——, 2006b: A hybrid multivariate normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett.*, **7**, 43–46, https://doi.org/10.1002/asl.128.

——, and ——, 2007: Implications and impacts of transforming lognormal variables into normal variables in VAR. *Meteor. Z.*, **16**, 755–765, https://doi.org/10.1127/0941-2948/2007/0243.

——, and A. S. Jones, 2014: Multiplicative and additive incremental variational data assimilation for mixed lognormal-Gaussian errors. *Mon. Wea. Rev.*, **142**, 2521–2544, https://doi.org/10.1175/MWR-D-13-00136.1.

Goodliff, M., S. Fletcher, A. Kliewer, J. Forsythe, and A. Jones, 2020: Detection of non-Gaussian behavior using machine learning techniques: A case study on the Lorenz 63 model. *J. Geophys. Res. Atmos.*, **125**, e2019JD031551, https://doi.org/10.1029/2019JD031551.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, https://doi.org/10.1115/1.3662552.

——, and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95–108, https://doi.org/10.1115/1.3658902.

Kliewer, A. J., S. J. Fletcher, A. S. Jones, and J. M. Forsthye, 2016: Comparison of Gaussian, logarithmic transform and mixed Gaussian-log-normal distribution based 1DVAR microwave temperature-water vapour mixing ratio retrievals. *Quart. J. Roy. Meteor. Soc.*, **142**, 274–286, https://doi.org/10.1002/qj.2651.

Kondrashov, D., M. Ghil, and Y. Shprits, 2011: Lognormal Kalman filter for assimilating phase space density data in the radiation belts. *Space Wea.*, **9**, S11006, https://doi.org/10.1029/2011SW000726.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Zupanski, M., 2005: Maximum likelihood ensemble filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726, https://doi.org/10.1175/MWR2946.1.