CrossMark

Quantification of Optimal Choices of Parameters in Lognormal Variational Data Assimilation and Their Chaotic Behavior

Steven J. Fletcher[1] · **Anton J. Kliewer**[1] ·
Andrew S. Jones[1]

Abstract An important property of variational-based data assimilation is the ability to define a functional formulation such that the minimum of that functional can be any state that is desired. Thus, it is possible to define cost functions such that the minimum of the background error component is the mean, median or the mode of a multivariate lognormal distribution, where, unlike the multivariate Gaussian distributions, these statistics are not equivalent. Therefore, for lognormal distributions it is shown here that there are regions where each one of these three statistics are optimal at minimizing the errors, given estimates of an a priori state. Also, as part of this work, a chaotic signal was detected with respect to the first guess to the Newton–Raphson solver that affect the accuracy of the solution to several decimal places.

1 Introduction

Data assimilation (DA) plays an important role in how operational Numerical Weather and Ocean Prediction (NWP) and (NOP) centers produce their forecasts (Lorenc 1986;

✉ Steven J. Fletcher
steven.fletcher@colostate.edu

Anton J. Kliewer
anton.kliewer@colostate.edu

Andrew S. Jones
andrew.s.jones@colostate.edu

1 Cooperative Institute for Research in the Atmosphere, Colorado State University, 1375 Campus Delivery, Fort Collins, CO 80523-1375, USA

Lorenc et al. 2000; Rabier et al. 2000; Rosmond and Xu 2006; Gauthier et al. 2007; Rawlins et al. 2007; Kleist et al. 2009). At the core of the data assimilation system is the minimization of the posterior errors. As the errors are treated as random variables, they must have a probability density function (PDF) or distribution assigned to them (Lorenc 1986; Fletcher 2017). Currently, in the NWP and NOP centers the prior errors are assumed to be multivariate Gaussian distributed.

Gaussian-based variational (VAR) DA systems can be derived from either a functional formulation or from a Bayesian-based approach; see Fletcher (2017) for a detailed derivations of both approaches. The Bayesian-based approach involves PDFs and is designed to find the maximum a posteriori (MAP). Note the likelihood is $p(y \mid x)$, while VAR finds the mode of $p(x \mid y)$. Also note that Bayesian inference tries to find an accurate approximation to the posterior. The mode is just one (and for high-dimensional systems an extremely unlikely) state, but it is possible to also find the mean, or a median, state of the posterior distribution using Bayes's theorem.

The functional approaches can be designed to have whatever form of statistic to minimize the analysis (or posterior) errors that is desired. These approaches are not equivalent to finding the descriptive statistics of the posterior distribution except for the modal approach. An important feature of the multivariate Gaussian distribution is that the three descriptive statistics are the same; however, this is not the case for non-symmetric distributions. Both the Bayesian and functional formulations for the minimization of the errors result in a cost function that has to be minimized. Two of the most commonly used schemes in operational NWP to find the minimums are conjugate-gradient, and limited memory Newton–Raphson, based algorithms.

During the last decade or so there has been work to develop VAR DA systems with probabilistic models for positive-valued random variables that are based upon lognormal PDFs. The first development was for lognormal observational errors, which is based upon Cohn (1997), with Gaussian background errors (Fletcher and Zupanski 2006a). The next stage of development was to combine a Gaussian PDF with a lognormal PDF to form a mixed PDF and hence a mixed distribution based DA system (Fletcher and Zupanski 2006b). In Fletcher and Zupanski (2006b) the employment of this mixed distribution was for observational errors. The extension to background errors was presented in Fletcher and Zupanski (2007) and demonstrated in a 3-dimensional VAR full field DA system with the Lorenz 63 model (Lorenz 1963), where full field is referring to solving for a value for the whole of a dynamical field, or parameter, instead of finding a small increment to the field, where the latter approach is referred to as incremental VAR.

There is also a transform approach, in which lognormally distributed random variables are transformed into Gaussian distributed random variables by taking the logarithm of the lognormal random variable. This is equivalent to finding the median of the posterior distribution as shown in Fletcher and Zupanski (2007).

The derivation for a mixed distribution-based 4DVAR DA scheme was presented in Fletcher (2010), where the formulation was also tested with the Lorenz 63 model. In both Fletcher and Zupanski (2007) and Fletcher (2010) the mixed distribution approach was compared against the transform technique. The reason for this is that the variational DA schemes in Fletcher and Zupanski (2006a, b, 2007) and Fletcher (2010) are based upon a MAP approach. Finally, the incremental versions of the mixed distribution

3-dimensional and 4-dimensional VARs were introduced in Fletcher and Jones (2014) and compared against a Gaussian configuration with the Lorenz 63 model. A summary of all of the papers just mentioned can be found in Chapter 21 of Fletcher (2017).

The reason for developing non-Gaussian based systems is in part due to what are referred to as "dropouts". A dropout is where an operational data assimilation produces a large negative increment for positive definite fields, such that when the increment is added back to the background state, the resulting new state is negative and thus unphysical, which then causes the numerical model to become unstable. An example of a dropout can be found in Song et al. (2012) with a biochemical ocean variational data assimilation system where the Gaussian approach from time to time resulted in an unphysical analysis state.

Recently, there has been work to implement the mixed Gaussian-lognormal approach, with synthetic and real brightness temperature data, into the Cooperative Institute for Research in the Atmosphere's (CIRA) 1-dimensional optimal estimator (C1DOE) (Kliewer et al. 2016), which is based upon a 1-dimensional VAR formulation and uses the Newton–Raphson algorithm. One-dimensional VAR requires an a priori state, and because the synthetic observations were created from a known state, the estimate for the a priori state was chosen to be a randomly perturbed value from the known state. However, when the transform approach was compared against the MAP approach, it was seen that the MAP approach was not optimal at minimizing the errors, even though the problem had been set up so that it was believed that it would be.

Given the constraint of not being able to outperform the median approach as expected by theory, an approximation to C1DOE for the univariate background and observational error case was developed for test purposes. This simple approximation led to the findings that are presented here. It appears that there are regions of values for the a priori state such that each of the three descriptive statistics, mean, mode or median will minimize lognormally distributed background errors. Thus, given the general definitions for the three descriptive statistics of a lognormal distribution, it is possible to rearrange them with respect to the three parameters: the a priori state, x_{ap}, the background error variance, σ_b^2, and the observational error variance, σ_o^2, such that the minimums of their associated cost function are at the true state and the regions referred to above can be derived for the three statistics for both the perfect and imperfect observation cases.

However, while undertaking the experiments described, it became apparent that there was a sensitivity to the first guess to the Newton–Raphson solver that affected the accuracy of the final solution. This sensitivity was investigated through creating what are referred to as error planes, where the error refers to the differences between the preprescribed true state and the converged solution. When the errors are plotted against the first guess, as well as against different magnitudes of measurement error, the resulting plots appear to indicate a chaotic behavior to the sensitivity to the first guess. Upon further investigation, it appears that the chaotic structure is a form of Newton fractal.

Given this brief description of the properties of the lognormal based VAR, the remainder of this paper is as follows: Sect. 2 contains a brief overview of the Bayesian formulation for the Gaussian errors along with introducing lognormal based variational DA followed by the extension to the lognormal background errors combined with

Gaussian observation errors case. The cost functions for minimizing the background errors with respect to the lognormal median and mean are also introduced. In Sect. 3 the cost function for the one-dimensional approximation to C1DOE that minimizes the background errors by the mean, median and mode are introduced, and the optimal regions are presented with respect to the different parameters. The initial sensitivity to the first guess to the Newton–Raphson solver is also presented here.

In Sect. 4 the chaotic structures that have been discovered for the simple univariate case are presented, and it is shown that a chaotic behavior is present when any of the three parameters are optimized given the other two parameters and the measurement error. Upon re-examining the perfect observation case it is shown that there are also Newton fractals underlying the sensitivity of the first guess to the Newton–Raphson solver here as well, but that the sensitivity can result in a loss of up to seven figures of accuracy at every decimal place!

In Sect. 5 the sensitivity of how close to the optimal value the parameters have to be before the chaos starts to dominate the accuracy of the solution is presented. The paper is finished with conclusions, along with thoughts for the implications of these findings for operational data assimilation systems.

2 Variational Data Assimilation

In this section an overview of variational data assimilation for the Gaussian, lognormal and mixed Gaussian-lognormal distributions is presented for the Bayesian approach, along with a brief explanation of the functional method.

2.1 Gaussian Based Variational Data Assimilation

As mentioned in the Introduction, there are two different approaches to derive the associated cost function for the variational data assimilation systems. The starting point for both the Bayesian and the functional approach is the definition of the errors. If the errors are assumed to be multivariate Gaussian distributed, then the background and observational errors are defined as

$$\boldsymbol{\varepsilon}^{\mathbf{b}} \equiv \boldsymbol{x} - \boldsymbol{x}_{\mathrm{ap}} \sim \mathrm{MG}\left(\mathbf{0}, \mathbf{B}\right), \qquad \boldsymbol{\varepsilon}^{\mathbf{o}} \equiv \boldsymbol{y} - \mathbf{h}\left(\boldsymbol{x}\right) \sim \mathrm{MG}\left(\mathbf{0}, \mathbf{R}\right), \qquad (1)$$

where $\boldsymbol{\varepsilon}^{\mathbf{b}}$ represents the background error, which is that associated with the numerical model's current estimate of the true state, $\boldsymbol{x}^{\mathrm{t}}$, given by $\boldsymbol{x}$ and where $\boldsymbol{x}_{\mathrm{ap}}$ is an a priori estimate of the true state. The observation error, $\boldsymbol{\varepsilon}^{\mathbf{o}}$, comprises two types of error: the measurement errors associated with the devices used to take the observation $\boldsymbol{y}$, and the representative error, which is associated with both the observation operator, $\mathbf{h}\left(\boldsymbol{x}\right)$, which is the approximation of the atmospheric or oceanic variable to the indirect observation of the atmospheric or oceanic variable, not an exact match for the observation, and the mismatch of dynamical scales between what the true observation is observing and those that the numerical model can resolve due to its grid resolution. The $\mathbf{B}$ and $\mathbf{R}$ matrices are the background and observational error covariance matrices respectively, and are defined by

$$\mathbf{B} \equiv \mathbb{E}\left[\left(\boldsymbol{\varepsilon}^{\mathbf{b}} - \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathbf{b}}\right]\right)^2\right], \text{ and } \mathbf{R} \equiv \mathbb{E}\left[\left(\boldsymbol{\varepsilon}^{\mathbf{o}} - \mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathbf{o}}\right]\right)^2\right], \qquad (2)$$

where $\mathbb{E}$ represents the expectation operator.

One approach to minimize the errors is to apply Bayes's theorem (Lorenc 1986; Fletcher 2017), which is given by

$$P(A \mid B) \propto P(B \mid A) P(A), \qquad (3)$$

where $P(\cdot)$ represents a probability density function, A and B are events that the probabilities of being true are sought, and $P(A \mid B)$ is the conditional probability density function that the event A is true given that the event B is true. In variational data assimilation the event A is $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{t}}$, while the event B is $\boldsymbol{y} = \boldsymbol{y}^{\mathrm{t}}$ as stated in Lorenc (1986). Therefore, given the expressions above (3) becomes

$$P\left(\boldsymbol{x} = \boldsymbol{x}^{\mathrm{t}} \mid \boldsymbol{y} = \boldsymbol{y}^{\mathrm{t}}\right) \propto P\left(\boldsymbol{y} = \boldsymbol{y}^{\mathrm{t}} \mid \boldsymbol{x} = \boldsymbol{x}^{\mathrm{t}}\right) P\left(\boldsymbol{x} = \boldsymbol{x}^{\mathrm{t}}\right). \qquad (4)$$

According to Lorenc (1986) (4) defines a multivariate PDF, which is denoted by $P_a(\boldsymbol{x})$, where this PDF is specifying all that is known about the *analysis*, and that VAR data assimilation is seeking the *best* estimate, $\boldsymbol{x}_a$, which is given by the mode of $P_a(\boldsymbol{x})$, and is the state such that $P_a(\boldsymbol{x})$ is maximum. The expression above is in terms of probabilities and as such the probability density functions that are to be used to calculate these probabilities have to be defined.

Given the assumption that the background and observational errors are multivariate Gaussian distributed, and that they are independent of each other, along with the definition of the multivariate Gaussian distribution,

$$\mathrm{MG}\,(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \frac{1}{|\boldsymbol{\Sigma}|^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\}, \qquad (5)$$

where n is the number of variables, $\boldsymbol{\mu}$ is the vector of means and $\boldsymbol{\Sigma}$ is the covariance matrix, and applying the log likelihood approach, which is to take the negative logarithm of (3), results in the problem of finding the state that minimizes a cost function. For the error definitions given above, the associated cost function is

$$J(\boldsymbol{x}) = \underbrace{\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{ap}})^{\mathrm{T}} \mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{ap}})}_{\text{BACKGROUND}} + \underbrace{\frac{1}{2}(\boldsymbol{y} - \mathbf{h}(\boldsymbol{x}))^{\mathrm{T}} \mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{h}(\boldsymbol{x}))}_{\text{OBSERVATIONAL}}. \qquad (6)$$

2.2 Lognormal Distribution Based Cost Functions

As mentioned in the Introduction, the basis for Gaussian PDF-based variational DA schemes is the definition of the errors. This is also true for lognormal PDF-based DA schemes. However, as a lognormal distribution is not defined for values less than or equal to zero, the definition for the errors must be such that the random variable

is always positive. The original definitions for lognormally distributed observational errors come from Cohn (1997) and Fletcher and Zupanski (2006a), while the definition for lognormally distributed background errors comes from Fletcher and Zupanski (2007), and are given by

$$
\boldsymbol{\varepsilon}_L^{\mathbf{b}} \equiv \frac{x}{x_{\mathrm{ap}}} \sim \mathrm{MLN}\left(\mathbf{0}, \mathbf{B}_L\right), \qquad \boldsymbol{\varepsilon}_L^{\mathbf{o}} \equiv \frac{y}{\mathbf{h}(x)} \sim \mathrm{MLN}\left(\mathbf{0}, \mathbf{R}_L\right), \tag{7}
$$

where

$$
\mathbf{B}_L \equiv \mathbb{E}\left[\left(\ln \boldsymbol{\varepsilon}_L^{\mathbf{b}} - \mathbb{E}\left[\ln \boldsymbol{\varepsilon}_L^{\mathbf{b}}\right]\right)^2\right], \text{ and } \mathbf{R}_L \equiv \mathbb{E}\left[\left(\ln \boldsymbol{\varepsilon}_L^{\mathbf{o}} - \mathbb{E}\left[\ln \boldsymbol{\varepsilon}_L^{\mathbf{o}}\right]\right)^2\right]. \tag{8}
$$

Note that the covariance matrices are with respect to $\ln x$, not x. Finally the associated multivariate lognormal distribution is defined by

$$
\mathrm{MLN}\left(x; \boldsymbol{\mu}_L, \boldsymbol{\Sigma}_L\right) = \prod_{i=1}^{N}\left(\frac{1}{x_i}\right) \frac{1}{|\boldsymbol{\Sigma}_L|^{\frac{N}{2}}(2\pi)^{\frac{N}{2}}}
$$
$$
\times \exp\left\{-\frac{1}{2}(\ln x - \boldsymbol{\mu}_L)^{\mathrm{T}} \boldsymbol{\Sigma}_L^{-1}(\ln x - \boldsymbol{\mu}_L)\right\}. \tag{9}
$$

Formulating the Bayesian problem as described in the introduction with the definitions in (7) and using the PDF defined in (9) to provide the probability, then next step is to seek the mode of the posterior distribution as set out in Lorenc (1986) which results in the following cost function that has to minimized

$$
J_L(x) = \frac{1}{2}\left(\ln x - \ln x_{\mathrm{ap}}\right)^{\mathrm{T}} \mathbf{B}_L^{-1}\left(\ln x - \ln x_{\mathrm{ap}}\right) + \left\langle\left(\ln x - \ln x_{\mathrm{ap}}\right), \mathbf{1}_N\right\rangle
$$
$$
+ \frac{1}{2}(\ln y - \ln \mathbf{h}(x))^{\mathrm{T}} \mathbf{R}_L^{-1}(\ln y - \ln \mathbf{h}(x)) + \left\langle(\ln y - \ln \mathbf{h}(x)), \mathbf{1}_{N_{\mathrm{o}}}\right\rangle, \tag{10}
$$

where $\mathbf{1}$ is a vector of 1s, N is the number of state variables and N_{o} is the number of observations.

2.3 Lognormal Background Errors with Gaussian Observational Errors

It is also possible to define a cost function for lognormally distributed background errors combined with Gaussian distributed observational errors such that the minimum is the mode of the posterior distribution which is given by

$$
J_{\mathrm{LG}}(x) = \frac{1}{2}\left(\ln x - \ln x_{\mathrm{ap}}\right)^{\mathrm{T}} \mathbf{B}_L^{-1}\left(\ln x - \ln x_{\mathrm{ap}}\right) + \left\langle\left(\ln x - \ln x_{\mathrm{ap}}\right), \mathbf{1}_N\right\rangle
$$
$$
+ \frac{1}{2}(y - \mathbf{h}(x))^{\mathrm{T}} \mathbf{R}^{-1}(y - \mathbf{h}(x)), \tag{11}
$$

(Kliewer et al. 2016).

2.4 Functional Form of Variational Data Assimilation

The initial work into 4DVAR appears in Lewis and Derber (1985), where the observations are assimilated throughout a window rather than at a set time. The problem in Lewis and Derber (1985) is equivalent to the minimization of the functional

$$J\left(\boldsymbol{x}\left(t_0\right)\right) = \min_{\boldsymbol{x}(t_0)} \int \frac{1}{2} \sum_{i=0}^{t_a} \langle \mathbf{W}\left(t_i\right)\left(\boldsymbol{x}\left(t_i\right) \ -\boldsymbol{x}_{\mathrm{b}}\left(t_i\right)\right), \left(\boldsymbol{x}\left(t_i\right) - \boldsymbol{x}_{\mathrm{b}}\left(t_i\right)\right)\rangle \, \mathrm{d}\boldsymbol{x}, \quad (12)$$

where t_a is the analysis time, $\mathbf{W}$ is a weight matrix which can be changed depending on known accuracies, the expression $\langle \cdot, \cdot \rangle$ is the inner product operator, $\boldsymbol{x}_{\mathrm{b}}$ is the output from a numerical model, which has been started by some set of initial conditions, $\boldsymbol{x}_{\mathrm{b},0} \equiv \boldsymbol{x}_{\mathrm{b}}\left(t_0\right)$, and $\boldsymbol{x}$ is the analysis that has come from a simpler version of DA (Lewis and Derber 1985). The problem is then to seek the initial conditions that minimizes the weighted squared differences between the original analysis from a simpler DA scheme at several times and the coincident solutions to the numerical model.

Given the expression in (12) it is possible to define functionals such that the minimum of that functional is any state that is desired. Therefore, by applying calculus of variational techniques it is possible to show that the solution of these functionals are equivalent to different sets of cost functions (Fletcher 2010, 2017). Thus, it is possible to define cost functions such that their minimums are the median of a lognormal based posterior distribution. Why this is important is that one approach for lognormally distributed errors is to transform them into Gaussian random variables through the property that the logarithm of a lognormal random variable is a Gaussian random variable. It is shown in Fletcher and Zupanski (2007) that this optimal value in Gaussian space inverts to a median in lognormal space.

However, a property of medians of multivariate distributions is that they are non-unique. This can easily be seen from the definition of the multivariate median

$$\widehat{\boldsymbol{x}} \text{ such that } \int_{a_N}^{\hat{x}_N} \int_{a_{N-1}}^{\hat{x}_{N-1}} \ldots \int_{a_2}^{\hat{x}_2} \int_{a_1}^{\hat{x}_1} f\left(\boldsymbol{x}\right) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_{N-1} \mathrm{d}x_N = 0.5, \quad (13)$$

where $\widehat{\boldsymbol{x}} = \left(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\right)$ is the median state such that when the integral in (13) is evaluated at the entries in $\widehat{\boldsymbol{x}}$, where $f\left(\boldsymbol{x}\right)$ is a multivariate PDF, then the cumulative density function is equal to 0.5. The values of the a_i for $i = 1, 2, \ldots, N$ are the lower limits of the distribution under consideration. The non-uniqueness comes through considering a general entry of the random variable, say x_i, then given a chosen value, there exists values for the remaining $N - 1$, such that (13) will hold.

As shown in Fletcher and Zupanski (2007) and Fletcher (2010) it is possible to define a cost function that is based upon the definition of the lognormal distribution's median through a functional form, see Fletcher (2010) or Fletcher (2017) for a full derivation. For the current assumption for the background and observational errors, the associated cost function would be

$$J_{\text{md}}(x) = \frac{1}{2} \left(\ln x - \ln x_{\text{ap}}\right)^{\text{T}} \mathbf{B}_L^{-1} \left(\ln x - \ln x_{\text{ap}}\right) + \frac{1}{2} \left(y - \mathbf{h}(x)\right)^{\text{T}} \mathbf{R}^{-1}$$
$$\times \left(y - \mathbf{h}(x)\right). \tag{14}$$

It is possible to define an approximate cost function so that the minimum is the mean of the posterior distribution. It should be noted that the multivariate mean is actually a vector of mean states not a mean vector. Therefore, it is possible to define a cost function componentwise that uses the mean to be the minimizer of the lognormal background errors (Fletcher 2010) as

$$J_{\text{mn}} = \sum_{i=1}^{N} \frac{\left(\ln x_i - \ln x_{\text{ap},i}\right)^2}{\mathbf{B}_{L,i,i}} - \frac{1}{2} \left(\ln x_i - \ln x_{\text{ap},i}\right) + \frac{1}{2} \left(y - \mathbf{h}(x)\right)^{\text{T}}$$
$$\times \mathbf{R}^{-1} \left(y - \mathbf{h}(x)\right). \tag{15}$$

Given the three cost functions defined in (10), (14) and (15), it is possible to identify regions of values for the different parameters such that each of the three descriptive statistics mentioned above minimize the errors, and these regions are presented in the next section.

3 Identification of Areas of Optimization

During the synthetic brightness temperature experiments in Kliewer et al. (2016) there were problems with the modal-approach mixed distribution based version of C1DOE not being able to outperform the median based approach, even though the situation had been set up such that it should. To test why this was the case, a synthetic univariate version was coded in MATLAB to investigate what could be wrong minimizing the lognormal background errors with the mode, median and mean, which uses the following three cost functions

$$J_{\text{mo}}(x) = \frac{1}{2} \frac{\left(\ln x - \ln x_{\text{ap}}\right)^2}{\sigma_{\text{b}}^2} + \left(\ln x - \ln x_{\text{ap}}\right) + \frac{1}{2} \frac{\left(y - h(x)\right)^2}{\sigma_{\text{o}}^2}, \tag{16a}$$

$$J_{\text{md}}(x) = \frac{1}{2} \frac{\left(\ln x - \ln x_{\text{ap}}\right)^2}{\sigma_{\text{b}}^2} + \frac{1}{2} \frac{\left(y - h(x)\right)^2}{\sigma_{\text{o}}^2}, \tag{16b}$$

$$J_{\text{me}}(x) = \frac{1}{2} \frac{\left(\ln x - \ln x_{\text{ap}}\right)^2}{\sigma_{\text{b}}^2} - \frac{1}{2} \left(\ln x - \ln x_{\text{ap}}\right) + \frac{1}{2} \frac{\left(y - h(x)\right)^2}{\sigma_{\text{o}}^2}, \tag{16c}$$

respectively.

The minimums of (16a), (16b) and (16c) are found through setting their gradients to zero where each are given by

$$\frac{\mathrm{d}J_{\text{mo}}}{\mathrm{d}x} = \left(\frac{1}{x}\right) \left(\frac{\left(\ln x - \ln x_{\text{ap}}\right)}{\sigma_{\text{b}}^2} + 1\right) - H \frac{\left(y - h(x)\right)}{\sigma_{\text{o}}^2}, \tag{17a}$$

$$\frac{\mathrm{d}J_{\mathrm{md}}}{\mathrm{d}x} = \left(\frac{1}{x}\right)\frac{(\ln x - \ln x_{\mathrm{ap}})}{\sigma_{\mathrm{b}}^2} - H\frac{(y - h\,(x))}{\sigma_{\mathrm{o}}^2}, \tag{17b}$$

$$\frac{\mathrm{d}J_{\mathrm{me}}}{\mathrm{d}x} = \left(\frac{1}{x^{\mathrm{t}}}\right)\left(\frac{(\ln x - \ln x_{\mathrm{ap}})}{\sigma_{\mathrm{b}}^2} - \frac{1}{2}\right) - H\frac{(y - h\,(x))}{\sigma_{\mathrm{o}}^2}, \tag{17c}$$

respectively.

An important property of the lognormal distribution in relation to the three descriptive statistics, and in general for left skewed distribution, is that

$$\mathrm{MODE} \leq \mathrm{MEDIAN} \leq \mathrm{MEAN}.$$

For the lognormal distribution, these statistics are given by

$$\exp\left\{\mu - \sigma^2\right\} < \exp\left\{\mu\right\} < \exp\left\{\mu + \frac{\sigma^2}{2}\right\}. \tag{18}$$

It is the inequality in (18) that is important in determining the bounds on the three statistics and when they are optimal for minimizing the errors.

3.1 Perfect Observations Case

"Perfect observations" refers to the case where there are neither measurement nor representative errors and that at the true solution, x^{t}, the observational component is zero, i.e. $y = h\left(x^{\mathrm{t}}\right)$. Therefore, the two parameters, x_{ap} and σ_{b}, are free to be chosen so that the minima of (16a)–(16c) are at x^{t}. An important feature to note here is that while all three cost functions are equal to zero at $x_{\mathrm{ap}} = x^{\mathrm{t}}$, when $x = x^{\mathrm{t}}$, this does not imply that the gradients of their cost functions are also equal to zero.

Therefore, considering x_{ap} first such that the minima of (16a)–(16c) are at the true state, given the definition for the mean state, $\exp\left\{\ln x_{\mathrm{ap}} + \frac{\sigma_{\mathrm{b}}^2}{2}\right\} \equiv x_{\mathrm{ap}}\exp\left\{\frac{\sigma_{\mathrm{b}}^2}{2}\right\}$, implies that the zero of (17c) at x^{t} is $x_{\mathrm{ap}} = x^{\mathrm{t}}\exp\left\{-\frac{\sigma_{\mathrm{b}}^2}{2}\right\}$. Now, considering the median-based approach, for (17b) to have a solution at x^{t} it is required that $x_{\mathrm{ap}} = x^{\mathrm{t}}$. Finally, considering the modal based approach, for the minimum of (17a) to be at x^{t} then $x_{\mathrm{ap}} = x^{\mathrm{t}}\exp\left\{\sigma_{\mathrm{b}}^2\right\}$.

Thus, the mean-based approach has the smallest absolute errors for values of x_{ap} from just above zero to $x_{\mathrm{ap}} = x^{\mathrm{t}}\exp\left\{-\frac{\sigma_{\mathrm{b}}^2}{4}\right\}$ where the median-based approach now has the smallest absolute errors up to $x_{\mathrm{ap}} = x^{\mathrm{t}}\exp\left\{\frac{\sigma_{\mathrm{b}}^2}{2}\right\}$. Thus, for all $x_{\mathrm{ap}} \geq x^{\mathrm{t}}\exp\left\{\frac{\sigma_{\mathrm{b}}^2}{2}\right\}$ the modal based approach has the smallest absolute errors.

The ranges for which the three different statistics have the smallest errors associated with solving (17a)–(17c) using the Newton–Raphson method with the nonlinear
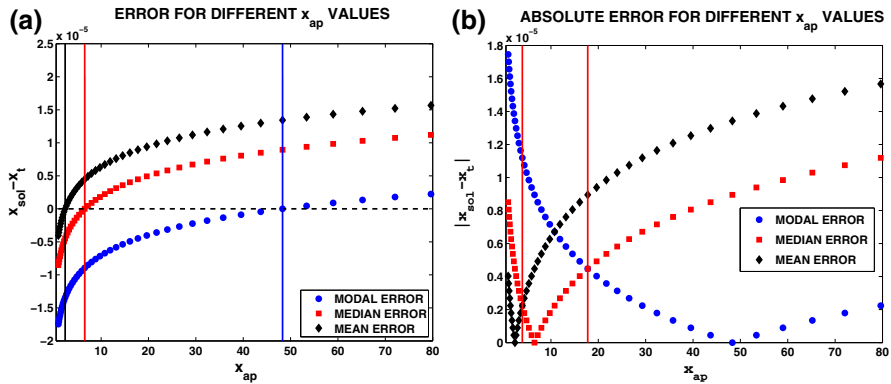
Fig. 1 Plots to illustrate the ranges where the mean (black), median (red) and the mode (blue) based approaches for different values of x_{ap} for $\sigma_b^2 = 2$ minimize the errors. Plot **a** is the difference plot, while plot **b** is the absolute error

observation operator, $h(x) = x^2$, are presented in Fig. 1. All first guesses are $x_{fg} = 1$, with $\sigma_b^2 = 2$ and $\sigma_o^2 = 0.01$. Note that no observational error (measurement error) has been introduced here, but there is representative error in that $h(x) \neq y$, except when $x = x^t$. The true state is generated from a lognormal distribution random number generator in MATLAB using $\mu = -0.75$ and $\sigma = 1.4454$. These values have been found from real data associated with C1DOE. The true solution here is $x^t = 6.5$. It should also be noted that there is no error added to the observations.

The errors and absolute errors have been plotted in Fig. 1a and b, respectively, for different values of x_{ap}, where $x_{ap} = x_t \exp\{(j - 10) * 0.05 * \sigma_b^2\}$ for $j = 1, 2, \ldots, 100$, where the lines in Fig. 1a represent the exact values for x_{ap} where the minimum of the associated cost function is at the true state, while in Fig. 1b the red lines represent the region where the median-based approach minimized the absolute errors the most. This is to illustrate where the different descriptive statistic-based approaches take over from each other, with respect to magnitude, in minimizing the errors. A clear feature that is present in both plots is that the area where the mean-based approach minimizes the errors is quite small, but that the regions just identified are also evident.

Another clear feature in Fig. 1b is that for this configuration the mode-based approach requires a large overestimation of the true state. However, this is illustrating that for the lognormal background error formulation the a priori state is not an approximation to the true state, but is an approximation to the true mean of the lognormal distribution. Therefore, in the lognormal situation it appears that the median-based approach is optimal at minimizing the errors if $x_{ap} \approx x^t$ and is consistent with the findings from Kliewer et al. (2016).

To illustrate how strong the assumption that x_{ap} has to be quite close to the true state for the median-based approach to be optimal at minimizing the errors the absolute error plots for five different values of σ_b^2, 0.1, 0.5, 1, 1.5, 2 are shown in Fig. 2. It is quite clear that the region where the median-based approach is optimal can be quite small for small background error variances. However, for values of x_{ap} that are greater than
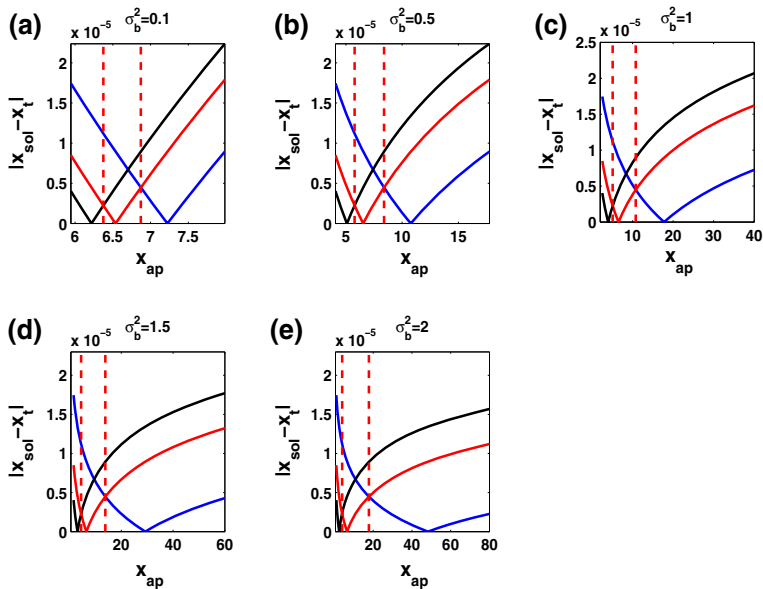
Fig. 2 Plots of the ranges where the three statistics based approaches minimize the absolute error. Mean (black), median (red) and mode (blue) where the dashed red lines represent the range of values for x_{ap} that the median based is optimal with $x_t = 6.5$ for **a** $\sigma_b^2 = 0.1$, **b** $\sigma_b^2 = 0.5$, **c** $\sigma_b^2 = 1$, **d** $\sigma_b^2 = 1.5$ and **e** $\sigma_b^2 = 2$

the upper bound for the median-based approach, the solution from the modal-based approach always has the smallest error.

As can be seen from Fig. 2a where $\sigma_b^2 = 0.1$, the range for which the median-based approach is optimal is smaller than for $\sigma_b^2 = 2$. The exact values for the upper and lower bounds for the five different choices for the background error variances are presented in Table 1. It is clear from both Fig. 2 and Table 1 that for $x_t = 6.5$, the range for $\sigma_b^2 = 0.1$ is only 0.4938, compared to 13.7264 for $\sigma_b^2 = 2$.

Another way of considering the bounds is to rearrange them to find bounds on the size of the errors $\frac{x_t}{x_{\text{ap}}}$, which for the median-based approach these are

$$\exp\left\{-\frac{\sigma_b^2}{2}\right\} \leq \frac{x_t}{x_{\text{ap}}} \leq \exp\left\{\frac{\sigma_b^2}{4}\right\}. \tag{19}$$

The error bounds for the variances shown in Table 1 are displayed in Table 2. It appears that for the median-based approach to have the smallest errors, the lognormal background error, for the perfect observation situation, can be as large as 65% bigger or 63% smaller.

However, it is also possible to bound the background error variance, σ_b^2, instead of x_{ap}. The values for σ_b^2 so that the three statistics based approaches are optimal are again obtained through evaluating the gradients in (17a)–(17c) at x_t. The first important feature to note here is that (17b) cannot be rearranged so that σ_b^2 is isolated.

Table 1 Summary of the upper and lower bounds for the optimization of the median based approach to minimize the errors

σ_b^2	Lower bound	Upper bound	Range
0.1	6.3395	6.8333	0.4938
0.5	5.7362	8.3462	2.6100
1	5.0622	10.7167	5.6545
1.5	4.4674	13.7605	9.2931
2	3.9424	17.6688	13.7264

Table 2 Summary of the upper and lower error bounds where the median-based approach minimizes the errors

σ_b^2	Lower bound	Upper bound
0.1	0.9512	1.0253
0.5	0.7788	1.1331
1	0.6065	1.2840
1.5	0.4724	1.4550
2	0.3679	1.6487

This then implies that for the median-based approach there is no optimal value for σ_b^2. Therefore, the median-based approach is only optimal if $x_{ap} = x_t$. For the mode-based approach, (17a), if $\sigma_b^2 = \ln\left(\frac{x_{ap}}{x_t}\right)$ then it is possible to converge for the lognormal approach $\forall x_{ap} > x_t$. Finally, for the mean-based approach, the relationship for σ_b^2 is

$$\sigma_b^2 = \ln\left[\left(\frac{x_t}{x_{ap}}\right)^2\right], \forall x_t > x_{ap}.$$

To show the important feature that if $x_{ap} > x_t$, the modal approach has a minimum at the true solution when σ_b^2 is as defined before, a Newton–Raphson solver is used for the univariate case for values of x_{ap} that are greater than the true solution. The true state, x_t, was chosen from the same lognormal distribution mentioned above where $x_t = 3.2601$, where the values for x_{ap} are 4, 8, 12, 16 and 20. The results are shown in Fig. 3.

While the magnitude of the errors in Fig. 3 (y-axis) are only 10^{-3} it should be noted that when $\sigma_b^2 = \ln x - \ln x_{ap}$, the solutions can be as accurate as machine precision if an optimal first guess is chosen. If the first guess is within what appears to be an optimal zone then the solution can be as accurate as 10^{-9}. Further away from these areas the solution may only have an accuracy of 10^{-3}. The sensitivity to the first guess, which was referred to in the introduction, shall be expanded upon in Sect. 3.3.

3.2 Observational Measurement Error

For the remainder of this paper, the focus is on the modal approach. This is first to do with the fact that it is possible to find expressions for the three parameters x_{ap}, σ_b^2 and σ_o^2, when there is a measurement error, denoted by ε_m, such that the minimum of the cost function
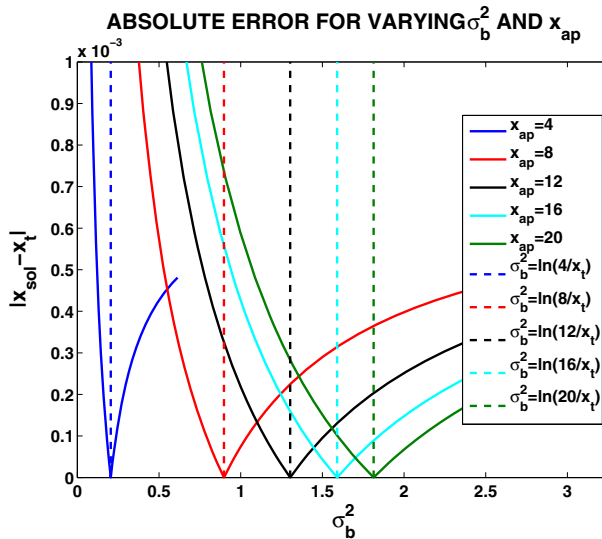
Fig. 3 Plot of the absolute errors for $x_{ap} = 4, 8, 12, 16, 20$ which have an affect on the σ_b^2 that are needed so that the solution of the cost function is the true solution

$$J_{\text{obsm}}(x) = \frac{1}{2}\left(\frac{(\ln x - \ln x_{ap})^2}{\sigma_b^2}\right) + (\ln x - \ln x_{ap}) + \frac{1}{2}\left(\frac{(y_t + \varepsilon_m - h(x))^2}{\sigma_o^2}\right),$$

(20)

is at the true solution, but secondly due to an unusual behavior that is detected with respect to the first guess to the Newton–Raphson solver applied to finding the minimum of (20).

To find expressions for the three parameters just mentioned so that the minimum of (20) is the true state, the gradient of (20) is evaluated at $x = x_t$, which gives

$$\left.\frac{\partial J_{\text{obsm}}}{\partial x}\right|_{x=x_t} = \frac{1}{x_t}\left(\frac{(\ln x_t - \ln x_{ap})}{\sigma_b^2} + 1\right) - H|_{x=x_t}\frac{\varepsilon_m}{\sigma_o^2} = 0,$$

(21)

where $y_t - h(x) = 0$. Through some standard rearrangement techniques it is easy to show that the optimal values for x_{ap}, σ_b^2 and σ_o^2 are

$$x_{\text{ap,opt}} \equiv x_t \exp\left\{\sigma_b^2 - \frac{x_t\, H|_{x=x_t}\, \varepsilon_m \sigma_b^2}{\sigma_o^2}\right\},$$

(22a)

$$\sigma_{\text{b,opt}}^2 \equiv \left(1 - \frac{x_t\, H|_{x=x_t}\, \varepsilon_m}{\sigma_o^2}\right)^{-1}(\ln x_{ap} - \ln x_t),$$

(22b)

$$\sigma_{\text{o,opt}}^2 \equiv \frac{x_t\, H|_{x=x_t}\, \varepsilon_m \sigma_b^2}{\ln x_t - \ln x_{ap} + \sigma_b^2}, \quad \text{for} \quad x_t \neq x_{ap} \exp\left\{-\sigma_b^2\right\}.$$

(22c)

ANALYSIS ERROR FOR DIFFERENT FIRST GUESSES
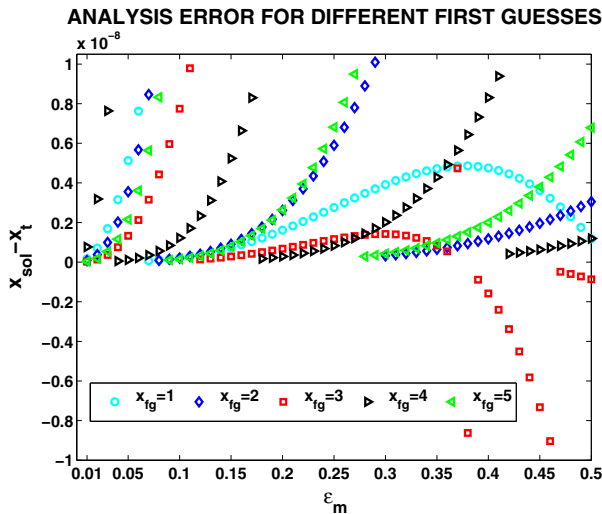
Fig. 4 Plot of the errors for increasing measurement errors for the x_{fg} values of 1, 2, 3, 4 and 5

An important feature to note about the these expressions is that when (22a) and (22b) are evaluated at $\varepsilon_m = 0$, the expressions are those identified for the perfect observation case.

To test the effectiveness of the formulation in (22a), values for $\varepsilon_m = 0.01, 0.02, \ldots,$ 0.05 were used. The main feature that was discovered is that as ε_m becomes larger, the associated optimal values for x_{ap} become very small. However, even with small values for x_{ap} it was still possible to converge, but not to machine precision, which was the case when $\varepsilon_m = 0$. It should be noted that to obtain the most accurate answer, there again appeared to be a sensitivity to the choice for the first guess for the Newton–Raphson solver.

3.3 Initial Detection of a Sensitivity to the First Guess for the Newton–Raphson Solver

It became apparent during the experiments for the optimized x_{ap} case that the choice for the first guess for the Newton–Raphson solver was more sensitive than previously indicated. It was found that the sensitivity was present at every decimal place. Through trial and error by changing each decimal place for the first guess for the Newton–Raphson solver, there were values such that the solver was started at values greater than these then the accuracy would be between 10^2 to 10^3 times worse than if the scheme was started at or less than these "cliff values". The magnitude of the error has been increased to approximately 10^{-11} when the measurement error is introduced. When starting from the wrong point at any decimal place the magnitude of the error increases to 10^{-8}.

To illustrate the impact of the first guess for the Newton–Raphson solver the residuals, $x_{\text{sol}} - x_t$, for the first guesses, $x_{\text{fg}} = 1, 2, 3, 4$ and 5, given the measurement errors, $\varepsilon_m = 0.01, 0.02, \ldots, 0.5$ are shown in Fig. 4, where the true solution

has again been randomly sampled from a lognormal distribution and has the value $x_t = 3.260104025221147$, where $\sigma_b^2 = 0.5$ and $\sigma_o^2 = 0.1$, with x_{ap} set to the expression in (22a).

A disturbing feature in Fig. 4 is the apparent lack of randomness in the error, but also the breaking points that appear to be present except for $x_{fg} = 3$, which has a very unusual behavior compared to the other four values. However, during the previous perfect observation experiments it had been possible to find certain values for x_{fg} such that the solution from the Newton–Raphson solver was equivalent to the true state to machine precision, but that it does not appear to be so for this configuration.

4 Chaotic Structures

As mentioned in the previous section, there appears to be an unusual structure underlying the accuracy of the solutions for the minimum of (16a). To investigate this behavior a series of plots were created for each of the three parameters at their optimal values for the x_t that was selected in the previous section. The plots are of error values for increasing first guess values, x_{fg}, for 0.001 to 20, stepping by 0.001, against increasing measurement error values, 0.01 to 2, stepping by 0.01. This configuration results in four million points to sample the error plane. The first parameter that is considered is the optimized x_{ap}.

4.1 Chaotic Structures for Optimized x_{ap}

Given the evidence in the previous section of an unusual sensitivity of the first guesses for the Newton–Raphson solver, the plots in Fig. 5 are of the whole error-plane on the left, while the right-hand-side plot is a zoomed in area around the values that were found from the trial and error experiments.

It should be noted that the reason that the x-axis starts at 1.411 is that this is the first x_{fg} value that enables all of the measurement errors to have a real-valued answer. A feature to be concerned about is the chaotic structure for first guess values that are less than the true state. There is still a non-uniform structure to the errors for $x_{fg} > x^t$. In Fig. 5 the green areas are those that result in the highest accuracy. This is true of all the contour plots presented in this section that have the parameters at the optimal values.

By following a line at $x_{fg} = 2$, 3, 4 and 5 in the left plot in Fig. 5 then it is quite clear that the values for the errors changed quite substantially for the different measurement errors, which then explains the behavior presented in Fig. 4. The unusual behavior for $x_{fg} = 3$ can be explained from the plots in Fig. 5. At $x_{fg} = 3$ it can be seen from Fig. 5 that the errors changes sign as the measurement error increases, while for the other choices in Fig. 4 then, for the range of the measurement error in that figure, the errors stay positive.

The left plot of Fig. 6 is the zoomed-in area where the trial and error experiment was performed and where the values that were found were assumed to be the values with the smallest error are also plotted on top of the error plane. It is clear from the zoomed-in plot in Fig. 6 that there were regions near the trial and error points that
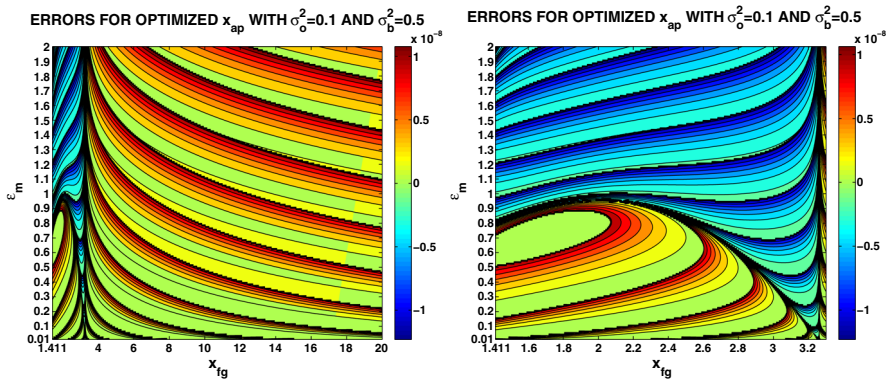
Fig. 5 Left: Contour plot of the Newton–Raphson errors for optimized x_{ap}. Right: Contour plot of the Newton–Raphson errors for $1.411 \leq x_{fg} \leq 3.3$
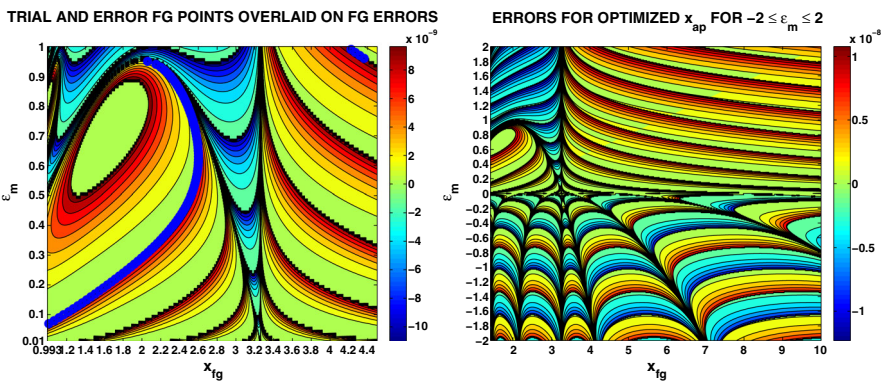

Fig. 6 Left Plot: Plot of the first guesses (blue circles) from the trial and error experiments with the error structure. Right Plot: Plot of the errors for the negative and positive measurement errors at the optimal value for x_{ap}

would have resulted in more accurate answers if the first guesses had been in these regions. The zoomed-in plot in Fig. 6 also addresses the reason why the x_{fg} values started to become smaller (the guesses were following the orbit of the lobe in the error plane); why there was no answer for $\varepsilon_m = 0.96$ and higher in the same area (the values turned negative, but also that the orbit stops at $\varepsilon_m = 0.95$), as well as showing the region of $x_{fg} > x^t$ that had the same error magnitudes.

The right hand side plot in Fig. 6 is of the error plane now with $-2 \leq \varepsilon_m \leq 2$ and for $1.411 \leq x_{fg} \leq 10$. This to illustrate that there is also a chaotic signal for negative measurement errors when x_{ap} is optimized. It appears that there is more sensitivity to the first guess for $\varepsilon_m < 0$, implying that finding the most accurate answer could be more difficult for $\varepsilon_m < 0$ than for $\varepsilon_m > 0$. It is also clear in the right-hand-side plot in Fig. 6 that at $\varepsilon_m = 0$ there is also some chaotic signal with respect to the first guess and the accuracy of the solution from the Newton–Raphson solver.

ERROR FOR OPTIMIZED σ_b^2 WITH σ_o^2=0.1, x_{ap}=1.25*x_t
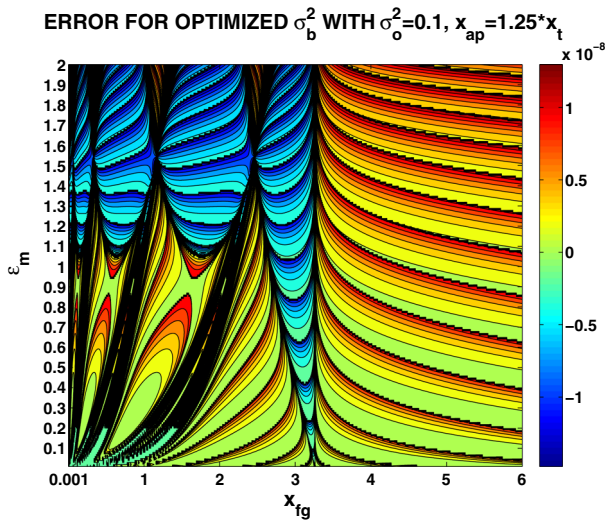
Fig. 7 Plot of the error plane for optimized σ_b^2 given $\sigma_o^2 = 0.1$ and $x_{ap} = 1.25 * x_t$ for $0.001 \leq x_{fg} \leq 6$

4.2 Chaotic Structures for Optimized σ_b^2

In this subsection the chaotic structures in the error plane when σ_b^2 is optimized when measurement error is present is presented. The first feature to notice in (22b) is that the relationship between σ_b^2 and ε_m is not exponential, which is the case between x_{ap} and ε_m. During the natural extension of the optimized x_{ap} experiments to the optimized σ_b^2 situation it became quite apparent that there was again a sensitivity to the first guesses used for the Newton–Raphson solver. Figure 7 is the error-plane where $\sigma_o^2 = 0.1$ and $x_{ap} = 1.25 * x_t$, where the starting value for x_{fg} is 0.001 as there are real solutions for all values of ε_m here.

The noticeable difference in Fig. 7, compared to the plot in Fig. 5, is that there appears to be more sensitive areas in the error plane for the optimized σ_b^2 case. These sensitive areas are those that appear to be black in color. These areas represent rapid changes in the error. This means that a small change in the first guess will produce quite different answers. Another noticeable difference is the structure of the chaos for $x_{fg} < x^t$ in both figures. However, it appears that the structures $x_{fg} > x^t$ are quite similar to those for the optimized x_{ap} case. The structure for the errors associated with $x_{fg} < x^t$ are quite different; in addition to the rapid chaotic signal, the areas where the most accurate solutions can be found are in different locations, which then implies values that could result in the more accurate solution for one optimized parameter are not the same for the other optimized parameter.

4.3 Chaotic Structures for Optimized σ_o^2

In this subsection the chaotic structures for the case when σ_o^2 is optimized to compensate for the presence of measurement error are presented (Fig. 8). The first noticeable
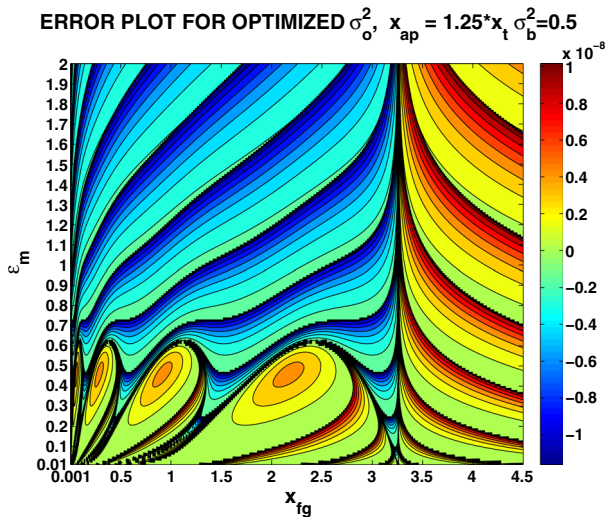
Fig. 8 Plot of the error plane for optimized σ_o^2 given $\sigma_b^2 = 0.5$ and $x_{\mathrm{ap}} = 1.25 * x_t$ for $0.001 \leq x_{\mathrm{fg}} \leq 6$

difference between Figs. 7 and 8 is the lack of the rapid changing of the errors with respect to the first guess in the latter figure. The lobe structures that are present in Fig. 8 appear to be similar to those present in the equivalent plots for x_{ap} but there is a significant difference. The difference between the two cases is that for the optimized σ_o^2 the error values at the center of the lobes are not as accurate as those at the center of the lobe in Fig. 5. Therefore, there are different values for x_{fg} for each optimized parameter that result in the most accurate solutions from the Newton–Raphson solver.

5 Sensitivity of the Chaotic Regions

In this section the sensitivity of the chaotic behavior is presented. The sensitivity is measured by how close to the optimal value for the three different parameters for the x_t mentioned in the last section does the chaotic behavior appear to be present. The results for the x_{ap} case are presented in Fig. 9. For the four plots in Fig. 9 it appears from the titles that the chaotic behavior starts to occur when x_{ap} is at 99.99% of the optimal value, but that some form of shock occurs at 99.9% of optimal. When the estimate of x_{ap} is at 99.999% then the structure of the chaotic behavior identified earlier is present, but not at the magnitude detected in Sect. 4.1. It is when the estimate for x_{ap} is at 99.9999% that the magnitudes of the errors are the same as Sect. 4.1.

Results for the other two parameters are not shown here, but what was noticeably different to the optimized a priori state case is that for both σ_b^2 and σ_o^2 cases is that the optimized value must be at 99.9999% before an indicator of the sensitivity to the first guess starts to appear. Thus this is indicating that exponential relationship between the optimized x_{ap} and ε_m, implies that the chaotic behavior occurs at less accurate estimates of the optimal state than for the approximations to the optimal values for σ_b^2.
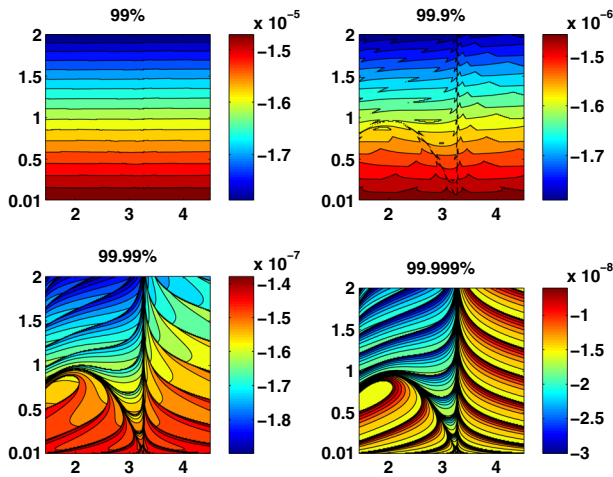
Fig. 9 Plot of the error plane as x_{ap} approaches the optimal value

6 Conclusions

In the first part of this paper it has been shown that there are bounds on the three descriptive statistics, mean, mode and median, for a lognormal background error with a Gaussian observational error formulation, where that specific statistic minimizes the errors. However, it was discovered that it is possible to find optimal expressions for the three parameters: a priori state, x_{ap}, background error variance, σ_{b}^2, and the observation error variance, σ_{o}^2, so that the minimum of the modal-based cost function was at the true state. This has been shown for both the perfect observation case, as well as when measurement error is present. Note that for the perfect observation case there was no expression possible for σ_{o}^2.

Given the optimal expressions for the three parameters, when solutions to the cost function were sought using a Newton–Raphson solver, it became apparent that there was a sensitivity to the first guess for the solver that affected the accuracy of its solution. Upon further investigation, it became clear that there was an underlying chaotic behavior associated with this minimization approach. It has also been shown that the chaotic behavior starts to appear in the error plane as the estimates for the optimal values for the three parameters are above 99% of the true value. Below 99%, and above 101%, of the optimal value then the performance of the Newton–Raphson solver appears to be uniform, and hence independent of x_{fg}, but results in drastically less accurate solutions.

While the results presented are for a specific true state, tests with larger and smaller true values, with different values for x_{ap}, σ_{b}^2 and σ_{o}^2 have been performed and some form of chaotic behavior is still present, but different in appearance.

The behavior with respect to the Newton–Raphson solver has been detected before (Gupta 2013), but not for the situation that has been presented here. This behavior is referred to as a Newton Fractal. Newton Fractals occur when the quadratic convergence of the Newton–Raphson scheme breaks down. However, this is a worrying feature, as

it implies that as the estimates of the parameters involved in data assimilation approach the optimal value such that the minimum of the cost function is at the true state, the Newton–Raphson method's property of quadratic convergence breaks down. Given the breakdown of the quadratic convergence, the sensitivity to the first guess begins, and effectively the fractal starts to dominate the solvability of the method, even though the problem is approaching the true state, which is the ultimate goal for a data assimilation system.

This chaotic behavior needs further investigation to see the impact on different data assimilation and retrieval systems that use a form of Newton–Raphson for their solver. Given the findings here, techniques to detect the chaotic signal to give guidance on the conclusions that are made about the solutions of the Newton–Raphson solver need to be developed. The chaotic signal could result in wrong conclusions about the impact of new changes to systems, which could be new observations or numerical model development, when what really is occurring is that the system is close to the true state, and that chaos is now dominating the solvability of the problem.

The work that has been presented here demonstrates the need for the ability to run multiple Bayesian probability models to illustrate the potential important impacts that occur to retrieval and data assimilation systems under these conditions to allow the user check if their conclusions about the affect their changes to the system have had are correct.

References

Cohn SE (1997) An introduction to estimation error theory. J Meteorol Soc Jpn 75:257–288

Fletcher SJ (2010) Mixed lognormal-Gaussian four-dimensional data assimilation. Tellus 62:266–287

Fletcher SJ (2017) Data assimilation for the geosciences: from theory to applications, 1st edn. Elsevier, Boston

Fletcher SJ, Jones AS (2014) Multiplicative and additive incremental variational data assimilation for mixed lognormal and Gaussian errors. Mon Weather Rev 142:2521–2544

Fletcher SJ, Zupanski M (2006a) A data assimilation method for log-normally distributed observational errors. Q J Roy Meteorol Soc 132:2505–2519

Fletcher SJ, Zupanski M (2006b) A hybrid normal and lognormal distribution for data assimilation. Atmos Sci Lett 7:43–46

Fletcher SJ, Zupanski M (2007) Implications and impacts of transforming lognormal variables into normal variables in VAR. Meteorologische Zeitschrift 16:755–765

Gauthier P, Tanguay M, Laroche S, Pellering S, Morneau J (2007) Entension of a 3D-var to 4D-var: implementation of 4D-var at the meteorological service of Canada. Mon Weather Rev 135:2339–2354

Gupta B (2013) Newton Raphson fractals: a review. Int J Emerg Trends Technol Comput Sci IJETTCS 2:119–123

Kleist DT, Parrish DF, Derber JC, Treadon R, Wu W-S, Lord S (2009) Introduction of the GSI into the NCEP global data assimilation system. Weather Forecast 24:1691–1705

Kliewer AJ, Fletcher SJ, Forsythe JM, Jones AS (2016) Application of mixed lognormal-Gaussian data assimilation techniques to synethic data with the CIRA 1-dimensional optimal estimation retrieval system. Q J R Meteorol Soc 142:274–286

Lewis JM, Derber JC (1985) The use of adjoints equations to solve a variational adjustment problem with advective conttstraints. Tellus 73A:309–322

Lorenc AC (1986) Analysis methods for numerical weather prediction. Q J R Meteorol Soc 126:1177–1194

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW (2000) The Met. Office global three dimensional variational data assimilation scheme. Q J R Meteorol Soc 126:2991–3012

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

Rabier F, Jarvinen H, Klinker E, Mahfouf J-F, Simmons A (2000) The ECMWF implementation of four dimensional variational assimilation. Part I: experimental results with simplified physics. Q J R Meteorol Soc 126A:1143–1170

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ (2007) The Met Office global four-dimensional variational data assimilation scheme. Q J Roy Meteorol Soc 133:347–362

Rosmond T, Xu L (2006) Development of NAVDAS-AR: non-linear formulation and outer loop test. Tellus 58A:45–58

Song H, Edwards CA, Moore AM, Fiechter J (2012) Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation. Ocean Modell 54:1–17