Bishop Craig (Orcid ID: 0000-0002-7898-6493)

Data assimilation strategies for state dependent observation error variances

Craig H. Bishop[1]

School of Earth Sciences and ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Melbourne, Victoria, Australia.

[1]*Correspondence to:* Craig H. Bishop, School of Earth Sciences and ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Parkville 3010, Victoria, Australia; craighuntlybishop@gmail.com.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3424

Keywords: data assimilation; bounded variables; observation error; observation error of bounded variables; state dependent observation error.

Abstract:

The Ensemble Kalman Filter (EnKF) and 4D-Var Data Assimilation (DA) approaches require that a fixed observation error variance be specified for each observation. To highlight the need to consider the state dependence of observation error variances, we prove that the error variance of unbiased observations of bounded variables tends to zero as the unknown true value of the variable approaches the bound. How then, should state dependent observation error variances be specified for the EnKF and 4DVar? In an idealized system, three distinct strategies for choosing the observation error variance R are considered: (i) choose R to be the ensemble mean of the observation error variances associated with each member of an ensemble forecast, (ii) choose R to be the observation error variance that would occur if the truth was equal to the average of the ensemble mean and the observed value, or (iii) (impractically) choose R to be the true observation error variance associated with the (unknown) true state. It is shown that choice (iii) is the worst choice while (i) is the best choice. It is then shown that the Kalman gain of the EnKF is the best linear unbiased estimator of the state only when its R is the mean of all the observation error variances implied by the prior distribution of truth. This is a general result that supports the

idealized experiment's findings. Because EnKF and 4DVar Gaussian assumptions are grossly inaccurate for near zero semi-positive definite variables with state dependent R, the paper also compares the performance of two variations of the EnKF (the ln-EnKF and the GIGG-EnKF) that correct aspects of these inaccuracies including the tendency of R to diminish as the truth approaches zero. It was found that the GIGG-EnKF out-performs the ln-EnKF by a significant margin.

1.Introduction

Data Assimilation (DA) is the process of estimating a distribution of model states by combining a prior distribution of model states with information from error prone observations of the model variables. Accurate knowledge of the distribution of error prone observations given a true state is fundamental to DA accuracy as it is this distribution that determines the reliability of the observation. All DA schemes can be viewed as an approximation to Bayes' theorem which gives the probability density function (pdf) $\rho(\mathbf{x}\,|\,\mathbf{y})$ describing the probability that the vector list $\mathbf{x}^t$ of (unknown) true values of the model variables is equal to $\mathbf{x}$ given a vector $\mathbf{y}$ of imperfect observations of a corresponding set of functions $H(\mathbf{x}^t)$ of the true state $\mathbf{x}^t$. Bayes' theorem states that $\rho(\mathbf{x}\,|\,\mathbf{y})$ is related to the prior or forecast pdf of possible true states $\rho(\mathbf{x})$ through the equation

$$\rho(\mathbf{x}\mid\mathbf{y})=\frac{L(\mathbf{y}\mid\mathbf{x})\rho(\mathbf{x})}{\iiint\limits_{V} L(\mathbf{y}\mid\mathbf{x})\rho(\mathbf{x})dV} \quad (1)$$

where $L(\mathbf{y}\mid\mathbf{x})$ gives the pdf of imperfect observations **y** assuming that the true state was given by **x**. The observation likelihood pdf $L(\mathbf{y}\mid\mathbf{x})$ describes the inaccuracy of the observations for all possible values of **x**. The covariance of the errors in the observations **y** given a fixed true state **x** is the observation error covariance associated with this true state. Variations of observation error covariance with **x** can profoundly affect the posterior pdf $\rho(\mathbf{x}\mid\mathbf{y})$ of true states implied by the observations. Currently, the DA schemes most frequently used in atmospheric and oceanic science are 4D-Var (e.g. Rabier et al., 2000, Rawlins et al., 2007) and the EnKF (e.g. Whitaker et al., 2008, Schraff et al., 2016). These schemes have been primarily used under the assumption that the observational uncertainty can be described by a Gaussian distribution and that its variance is independent of the state.

The problem of deducing exactly how observation error covariance matrices vary with the state and/or precisely what form $L(\mathbf{y}\mid\mathbf{x})$ should take is non-trivial. Nevertheless, there is a growing body of work that suggests that for some observation types, variations of the observation uncertainty are significant and should not be ignored.

When the value of the observed variable is bounded, the range of errors is also bounded. For example, consider the range of possible observation errors associated with the observed value y of a non-negative variable whose true value is given by y^t. In this case, the smallest possible value of the observation error $\varepsilon^o = y - y^t$ is equal to $-y^t$ because the smallest possible value of y is zero. Thus, as y^t approaches zero the most negative possible error also approaches zero. In the appendix, we use this fact to prove that "*unbiased observations of any variable with a known physical bound must have observation error variances that tend to zero as the true value of the observed variable approaches this physical bound*". The proof is consistent with findings by Mears et al. (2001) for observations of wind speeds and Hyer et al. (2011) for observations of aerosol optical depth. Both authors find that the observation error standard deviations for such semi-positive definite quantities decrease as the variable approaches zero. In addition, the proof is consistent with Bouttier et al. (2012) (Table 1) who assume that the observation error standard deviation for observations of rain linearly increases with observed rain rate.

Other evidence for the state dependence of observation error variance comes from Geer and Bauer (2011), Okamoto et al. (2014) and Harnisch et al. (2016) who all presented strong evidence that the standard deviation of the observation error for a large range of satellite radiance observations had a strong dependence on the cloud state. Minamide and Zhang (2017) chose flow dependent observation error variances by comparing the variance of an ensemble forecast with the size of the difference between the observation and the mean of mean ensemble forecast.

Operational 4DVar and EnKF DA schemes were originally designed for observations whose distribution about the true value is Gaussian and whose error covariances are independent of the state. Indeed, Courtier et al's (1994) derivation of incremental 4DVar and the EnKF derivations given in Burgers et al. (1998), Houtekamer and Mitchell (2001), Bishop et al. (2001), Anderson

(2001), Whitaker and Hamill (2002) and Hunt et al. (2007) all overlook the possibility that the observation error covariance matrix might be a function of the unknown true state.

This paper addresses fundamental aspects of possible strategies for choosing observation error variances for EnKF and variational algorithms when the observation error variance varies with the true state. Mostly, it focusses on the type of observation error variance state dependence likely to occur when the true state lies less than a few observation error standard deviations from the bound. In such circumstances, the distribution of imperfect observations given a true state will need to be non-Gaussian to avoid observed values that exceed the known physical bound. Hence, we also consider strategies for dealing with state-dependent observation error variances based on simple modifications to the EnKF that replace the Gaussian pdf assumption with pdf assumptions that assume (a) that the probability of observing values beyond the physical bound is zero, and (b) that the observation error standard deviation increases linearly with the distance of the true state from the bound.

In Section 2, we describe a simple univariate DA experiment in which the variable considered has a prior ensemble forecast distribution that is semi-positive definite (bounded from below at zero) and an observation whose observation error standard deviation is proportional to the value of the unknown true state. Section 3 compares the DA performance of three different strategies for assigning observation error variances. Consistent with our finding that the observation error standard deviation of unbiased observations must tend to zero as the true state tends to zero, in the idealized experiments of Section 3, the absolute observation error variance increased with the value of the true state, but the relative observation error variance remains constant. Section 4 provides a theoretical explanation of these specific results by proving the general result that, in any system, the Kalman gain of the EnKF is the best linear unbiased estimator of the state only when its observation error variance R is the mean of all the observation error variances implied by the prior distribution of truth. Some readers may prefer to read Section 4 before Section 3. In section 5 we assess the relative performance of simple variations on current DA schemes that avoid assigning non-zero probabilities to unphysical values and that directly accommodate constant relative observation error variances. Concluding remarks follow in Section 6.

1. Simple experiment with state dependent observation error variance

(a) Experimental framework and data assimilation metrics

Consider a semi-positive definite variable whose prior pdf of truth is given by the black line in Figure 1a. This line is an example of a gamma pdf with both mean and variance equal to 1. In order to perform an ensemble DA experiment for such a variable we must be able to randomly sample such a pdf to create a K member ensemble. A subroutine for doing this is readily available in Matlab, for example. The blue bars in Figure 1a give the probability density histogram from one million random samples of the gamma pdf.

To perform a single DA trial in this simple system, we need a K member ensemble forecast of the observation, a true value y^t of the observed variable and an imperfect observation y.

The ensemble forecast, $y_i^f, i=1,2,...,K$ is created by randomly sampling the prior pdf of truth K times. To create an independent truth, we just take yet another independent random sample of the prior pdf of truth and label it y^t. This procedure ensures that our ensemble is perfect in the sense that is drawn from the same distribution as the truth even though none of its members will be equal to the truth. The pdf shown in Figure 1 is defined by two parameters: the prior mean $\langle y^f \rangle$ and the relative error variance $P_r = \frac{\mathrm{var}(y^f)}{\langle y^f \rangle^2}$. To ensure that the prior pdf changes from one DA trial to the next, for each DA experiment, we choose $\langle y^f \rangle$ to be a uniformly distributed random variable bounded between 0.25 and 2 while holding $P_r = 1$. It can be shown that P_r is equal to the inverse of the shape parameter of a gamma pdf. Hence, holding $P_r = 1$ ensures that the overall shape of the prior pdf shown in Figure 1 is maintained for all our experiments even though the prior mean is randomly changing from one experiment to the next. It also means that the absolute forecast error variance $P = P_r \langle y^f \rangle^2$ in our experiments varies between $(1/4)^2$ and 4 while the expected value of forecast error variance $E(P)$ over an infinite number of trials is given by

$$E(P) = \int_0^\infty \rho(\langle y^f \rangle) P_r \langle y^f \rangle^2 d\langle y^f \rangle = \frac{P_r}{1.75} \int_{1/4}^{2} \langle y^f \rangle^2 d\langle y^f \rangle = \frac{1}{5.25}\left[2^3 - \left(\frac{1}{4}\right)^3\right] \approx 1.52 \qquad (2).$$

We create an error prone observation y from the single true value y^t associated with each experiment by letting y be a random sample from an inverse gamma pdf that has a mean of y^t and a fixed relative error variance $R_r = \frac{\mathrm{var}\left(y - y^t\right)}{\left(y^t\right)^2} = \frac{R}{\left(y^t\right)^2}$ where R is the absolute observation error variance. Note that choosing the relative error variance to be a fixed number independent of the truth ensures that the absolute error variance $R = R_r\left(y^t\right)^2$ is proportional to the square of the unknown true value of the observed variable and hence goes to zero as the true value of the variable approaches its lower bound of zero. Also note that choosing the mean of the pdf of error prone observations y to be equal to the true value y^t of the observed variable ensures that we are modelling an *unbiased* observation. Thus, our observation error variance model is in accord with the assumptions of the Appendix. Figure 1b depicts the inverse-gamma pdf of possible error prone observations that would be selected from in the special case of $y^t = 1$ and $R_r = 1$.

The objective of ensemble DA in this case is to assimilate the observation y to obtain an ensemble of analyses, $y_i^a, i = 1, 2, ..., K$ of the observed variable. Arguably, the most fundamental DA performance metrics for an ensemble DA scheme is the mean square error (mse) of the mean of the ensemble of analyses $\overline{y^a} = \left(\frac{1}{K}\sum_{k=1}^{K} y_k^a\right)$ over a large number L of trials; specifically,

$$\mathrm{mse} = \frac{1}{L}\sum_{l=1}^{L}\left[\left(\overline{y^a} - y^t\right)_l\right]^2. \tag{3}$$

When mse is small the analysis ensemble mean is close to the unknown truth. For a perfect DA system, the variance of the analysis ensemble would be equal to the expected value of the square of the error of the ensemble mean. Hence, the average of the ensemble variances over a large number of trials ought to be equal to the corresponding mse. For that reason, a basic measure of the quality of the analysis ensemble is the normalized ensemble dispersion $\tilde{d}$ given by

$$\tilde{d} = \frac{(\text{ens variance})}{\text{mse}} = \frac{\frac{1}{L}\sum_{l=1}^{L}\left[\frac{1}{K-1}\sum_{k=1}^{K}\left(y_k^a - \overline{y^a}\right)^2\right]_l}{\text{mse}} \tag{4}$$

For a perfect ensemble DA scheme, this metric would tend to 1 as the number of trials L tended to infinity. Under these circumstances, if $\tilde{d} > 1$ the analysis ensemble is over-dispersive whereas if $\tilde{d} < 1$, it is under-dispersive.

(ii) Data assimilation approaches compared and contrasted

For each analysis, both EnKFs and 4DVar require a single specification of the observation error variance R and a single specification of the forecast error variance P. For all approaches considered here we shall approximate P using the sample variance of the ensemble forecast; in other words, we let

$$P = \frac{1}{K-1}\sum_{k=1}^{K}\left(y_k^f - y^b\right)^2 \tag{5}$$

where $y^b = \frac{1}{K}\sum_{k=1}^{K} y_k^f$ is the sample mean of the ensemble forecast. The standard perturbed observations EnKF produces the analysis ensemble using the equation

$$y_k^a = y_k^f + \frac{P}{R+P}\left[\left(y+\eta_k\right) - y_k^f\right], \quad k = 1,2,...,K \tag{6}$$

$\left(y+\eta_k\right)$ is the perturbed observation where the perturbation η_k is obtained by first drawing K random numbers $\tilde{\eta}_k, k = 1,2,...K$ from a normal distribution with mean zero and variance R and then setting

$$\eta_k = \left(\frac{K}{K-1}\right)\left(\tilde{\eta}_k - \overline{\tilde{\eta}_k}\right) \tag{7}$$

where $\overline{\tilde{\eta}_k}$ is the sample mean of $\tilde{\eta}_k, k = 1,2,...K$ and where the factor $\left(\frac{K}{K-1}\right)$ has been introduced to ensure that the expected value of the square of the random perturbations η_k is equal to R.

Note that in this very simple univariate experiment, there is no non-linear observation operator mapping from a control variable space to observation space. The ensemble is already in observation space. In these circumstances, there is no time dimension and no non-linearity and it can be shown that the analysis ensemble that would be produced by the perturbed observations form of 3DVar and 4DVar would be identical to that given by (6) provided the same values of R and P were chosen and that the cost function minimized was the usual 4DVar form that assumes Gaussian observation errors and Gaussian background errors (Courtier et al., 1994).

Here, we wish to compare the DA performance of (6) for three different methods of choosing *R*. Recalling that the true absolute observation error variance is related to the unknown truth by the equation $R = R_r\left(y^t\right)^2$, we consider three possibilities:

(i) $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$; i.e. let *R* be equal to its expected value over the prior ensemble forecast,

(ii) $R = R_r\left(\frac{y + y^b}{2}\right)^2$; i.e. the observation error variance that would occur if the unknown truth was equal to the average of the ensemble mean and the observed value, and

(iii) $R = R_r\left(y^t\right)^2$; i.e. we let *R* be equal to its true value (this choice would be unrealizable in practice but is nevertheless superficially appealing).

Note that method (ii) has similarity to work by Geer and Bauer (2011), Okamoto et al. (2014) and Harnisch et al. (2016) where the average of a cloud aspect of the background forecast and the observation was used as a predictor of observation error standard deviation.

The ensemble DA performance of (6) was examined for the above three approaches to specifying observation error variance. For a fixed value of the relative observation error variance R_r, ten sets of $L = 40,000$ trials were performed for each of the three approaches to defining observation error variance. In order to reduce random effects associated with small ensemble sizes, a relatively large ensemble of $K = 256$ members was used in all of the experiments. Three different fixed true values of R_r were considered: $R_r = 1, R_r = 0.1$ and $R_r = 0.01$.

2. **DA performance of the three methods for choosing *R***

The blue, green and pink curves on Fig. 2a respectively give the mse of eq. (6)'s analysis ensemble mean using $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$, $R = R_r\left(\frac{y+y^b}{2}\right)^2$ and $R = R_r\left(y^t\right)^2$ to define the absolute observation error variance, *R,* in the case that the relative observation error variance $R_r = 1$. Fig. 2a shows that the largest mse results from using $R = R_r\left(y^t\right)^2$. This choice uses the true observation error variance associated with the true state. Hence, this impractical but superficially appealing choice of the true observation error variance is actually the worst choice. The Geer and Bauer (2011) inspired $R = R_r\left(\frac{y+y^b}{2}\right)^2$ choice gives a mse similar to the $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$ choice but Fig. 2b shows that while the $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$ choice delivers the desired normalized dispersion ratio of $\tilde{d} = 1.0$, the $R = R_r\left(\frac{y+y^b}{2}\right)^2$ choice delivers an under-dispersive ensemble with $\tilde{d} = 0.66$. Hence, with $R_r = 1$, the best choice is

$R = \frac{1}{K}\sum_{k=1}^{K} R_r \left(y_k^f\right)^2$, the second best choice is $R = R_r \left(\frac{y + y^b}{2}\right)^2$ and the worst choice is $R = R_r \left(y^t\right)^2$.

Fig. 3 is the same as Fig. 2 except it pertains to the experiment with $R_r = 0.1$. In this case the true observation error variance is just one tenth of $\left(y^t\right)^2$ and Fig 3 shows that this makes the mses significantly lower those obtained for the $R_r = 1$ case. The results pertaining to the method used to choose R in this $R_r = 0.1$ case are similar to those for the $R_r = 1$ case except now the mse of the $R = R_r \left(\frac{y + y^b}{2}\right)^2$ choice is higher than that of the $R = \frac{1}{K}\sum_{k=1}^{K} R_r \left(y_k^f\right)^2$ in all 10 independent cases whereas for the $R_r = 1$ case, it was higher in only 6 out of 10 cases.

The results for the $R_r = 0.01$ case are shown in Fig. 4. Although this increased observational precision reduces mses for all methods, the ranking of the methods used to choose R are identical to those obtained for the $R_r = 0.1$ case.

In summary, these experiments show that, by a significant margin, choosing $R = \frac{1}{K}\sum_{k=1}^{K} R_r \left(y_k^f\right)^2$ is superior to choosing $R = R_r \left(\frac{y + y^b}{2}\right)^2$. Furthermore, the unrealizable but superficially appealing choice of letting R be equal to the true absolute error variance associated with the unknown true state would be the worst choice of all.

3. Proof that ideal EnKF R is the average of observation error variance over the prior

As previously mentioned, in the absence of non-linearity and non-Gaussianity, equation (6) delivers the maximal likelihood estimate of the state associated with 3DVar. It is also the standard EnKF update equation. Eq (6) can be derived by either taking the derivative of the 3DVar cost-function and setting it to zero or by finding the linear mapping that minimizes analysis error variance. Daley (1991), for example, gives both these derivations. Here, we review the minimum error variance derivation to clarify why taking $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ was the best of the three choices considered in Section 3.

The EnKF equation (6) is consistent with the linear-regression-like hypothesis that the true state $y^t=\left\langle y^a\right\rangle-\varepsilon^a$ is a linear stochastic process described by

$$y^t=\left\langle y^f\right\rangle+Gv-\varepsilon^a, \tag{8}$$

where $v=\left(y-\left\langle y^f\right\rangle\right)$ is the unbiased innovation $\left(\left\langle v\right\rangle=0\right)$and $\varepsilon^a=\left\langle y^a\right\rangle-y^t$ is the error of the posterior mean $\left\langle y^a\right\rangle$ that has zero correlation with the innovation v $\left(\left\langle\varepsilon^a v\right\rangle=0\right)$ and a mean of zero $\left(\left\langle\varepsilon^a\right\rangle=0\right)$. G is the gain. The symbol $\left\langle y^a\right\rangle$ denotes the posterior or analysis mean while

$\langle y^f \rangle$ denotes the prior mean. Throughout this paper, the angle brackets denote the expectation operator. Note that since $y^t = \langle y^a \rangle - \varepsilon^a$, (8) can be rearranged into the equivalent forms

$$\langle y^a \rangle = \langle y^f \rangle + Gv \text{ and } \varepsilon^a = \varepsilon^f + Gv. \tag{9}$$

where $\varepsilon^f = \langle y^f \rangle - y^t$. The equivalence of (9) and (8) serves to highlight the close connection between linear regression analysis and the EnKF. From (9), the expected value of the analysis error variance over the prior pdf of truth is the quadratic function of G given by

$$\left\langle \left(\varepsilon^a\right)^2 \right\rangle = \left\langle \left(\varepsilon^f\right)^2 \right\rangle + G^2 \left\langle v^2 \right\rangle + 2G \left\langle v\varepsilon^f \right\rangle . \tag{10}$$

Thus, the analysis error variance is minimized when G satisfies

$$\frac{\partial \left\langle \left(\varepsilon^a\right)^2 \right\rangle}{\partial G} = 2G\left\langle v^2 \right\rangle + 2\left\langle v\varepsilon^f \right\rangle = 0. \tag{11}$$

Hence, (9) delivers the minimum error variance estimate of the truth when

$$G = -\left\langle \varepsilon^f v \right\rangle \left\langle v^2 \right\rangle^{-1}. \tag{12}$$

Now noting that

$$v = y - \left\langle y^f \right\rangle = \left(y - y^t\right) + \left(\left\langle y^f \right\rangle - y^t\right) = \varepsilon^o - \varepsilon^f \tag{13}$$

and assuming that $\left\langle \varepsilon^o \varepsilon^f \right\rangle = 0$, it follows that

$$\left\langle \varepsilon^f v \right\rangle = -\left\langle \left(\varepsilon^f\right)^2 \right\rangle = -P \text{ and } \left\langle v^2 \right\rangle = \left\langle \left(\varepsilon^o\right)^2 \right\rangle + \left\langle \left(\varepsilon^f\right)^2 \right\rangle = \left\langle \left(\varepsilon^o\right)^2 \right\rangle + P \tag{14}$$

Where P is the expected value of the square of the forecast error over the entire prior distribution of states. Hence, $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$ must be the expected value of the square of the observation error over the entire prior distribution of states. Consequently, if there is a different observation error variance associated with each possible prior state then $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$ *is given by the expected value of the observation error variance over the entire prior distribution of possible states*. Using (14), (12) and (10) the mse of the analysis is minimized when

$$G = P\left[P + \left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle\right]^{-1} \text{ and } \left\langle \left(\varepsilon^{a}\right)^{2} \right\rangle = P - P\left[P + \left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle\right]^{-1} P \ . \tag{15}$$

Hence, if we denote $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$ by the symbol R then equation (9) states that the minimum error variance state estimate is given by

$$\left\langle y^{a} \right\rangle = \left\langle y^{f} \right\rangle + P\left(P + R\right)^{-1} v \, . \tag{16}$$

It is readily shown that provided the ensemble randomly samples the prior distribution of truth, in the limit of an infinite ensemble size, the mean of (6) would be identical to (16). Hence, this derivation shows that the choice of R that minimizes the mse of the analysis mean given by (6) is $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$. If the observation error variance was independent of the value of the truth, then $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$ would be equal to the true flow-independent observation error variance. However, when the true observation error variance is a varying function of the truth then $\left\langle \left(\varepsilon^{o}\right)^{2} \right\rangle$ depends on the prior distribution of truth; in other words, *it depends on the degree of the data assimilator's prior ignorance of the state.* Thus, the above derivation of the well-known eq (6) serves to explicitly highlight where and how R depends on our prior ignorance. As far as the author is aware, the dependence of R on prior ignorance has received little or no attention in previous derivations of (6). In the special case considered in sections 2 and 3, the observation error variance is the quadratic function of the truth given by

$$R\left(y^t\right)=\left(y^t\right)^2 R_r \tag{17}$$

where the symbol $R\left(y^t\right)$ is used to indicate that the observation error variance depends on y^t. Hence, in this case, the expected value of observation error variance over the prior pdf of truth is

$$\left\langle\left(\varepsilon^o\right)^2\right\rangle=\left\langle R_r\left(y^t\right)^2\right\rangle=R_r\left\langle\left(y^t\right)^2\right\rangle=R_r\left[\left\langle y^t\right\rangle^2+\operatorname{var}\left(y^t\right)\right] \qquad (18)^2$$

where $\operatorname{var}\left(y^t\right)$ denotes the variance of the prior pdf of truth. Thus, in this case, the accuracy with which the flow dependent R can be estimated is directly related to the accuracy with which both the prior mean and the prior variance can be estimated. If a large ensemble is available that samples the prior distribution then the sample mean, and sample variance of the ensemble would provide a flow dependent estimate of both the prior mean and prior variance that could be used to estimate the right-hand side of (18). If one could only afford a single ensemble member (K=1)

[2] Note that while (18) specifies a direct correlation between the variances of the forecast and observational errors, it does not imply a correlation between forecast errors and observation errors. This is possible because (18) does not impose a correlation between the signs of forecast and observation errors.

then might use this single member as an approximation to $\langle y^t \rangle$ or even an average of the single member and the observed value as in Geer and Bauer (2011) and a static, average estimate B of $\mathrm{var}(y^t)$. As noted in Bishop and Satterfield (2013), when $\mathrm{var}(y^t)$ changes from one case to the next and the ensemble size is large enough to accurately estimate $\langle y^t \rangle$ but too small to accurately estimate $\mathrm{var}(y^t)$, then an estimate of $\mathrm{var}(y^t)$ based on a Hybrid linear combination of B and the ensemble variance can be obtained that is superior to one based on either alone. Further research will be needed to more precisely determine the best way to estimate the right-hand side of (18) for specific forecasting systems.

For the experiments of Section 3, the ensemble size (K=256) was large enough to make the assumption

$$\left\langle \left(\varepsilon^o\right)^2 \right\rangle \cong \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2 \tag{19}$$

very accurate. Hence, in the light of the theoretical analysis of this section, the superiority of the assumption $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$ over the other assumptions could have been predicted with a high degree of certainty.

4. Non-Gaussian variations on the EnKF pdfs for near-zero bounded variables

Here, we consider two variations on the EnKF that are more suitable for near-zero bounded variables in that they assume (a) that the probability of observing values beyond the physical bound is zero, and (b) that the observation error standard deviation increases linearly with the distance of the true state from the bound giving a constant relative observation error variance rather than a state varying absolute observation error variance. The primary motivation for this section is to indicate the performance gains to be realized from non-Gaussian methods that implicitly account for the dependence of observation error variance with the true state near physical bounds. One method to be considered is Bishop's (2016) Gamma-Inverse-Gamma and Gaussian (GIGG) variation on the EnKF and the other is an "*observation transform*" approach in which one transforms the observation to a variable whose error variance is approximately independent of the state. Since the variable transform approach was not considered in Bishop (2016), this section also provides a first comparison between the performance of the GIGG-EnKF and a variable transform approach.

Bishop's (2016) GIGG-EnKF perfectly matches the assumptions of the simple problem examined in sections 2 and 3 perfectly so we expect it to perform well. The GIGG-EnKF is specifically designed to accommodate relative observation error variances. We do not include details of this algorithm here but refer the reader to Bishop (2016) for the details of its implementation. We note that while the GIGG-EnKF can be incorporated within EnKFs that assimilate observations serially quite easily, an easy way of incorporating it within 3D-Var or 4D-Var has not yet been identified.

For the observation transform approach, consider the Taylor series for the logarithm of the observation about the logarithm of the truth given by

$$\ln(y) = \ln(y^t) + \frac{(y - y^t)}{y^t} - \frac{1}{2}\left[\frac{(y - y^t)}{y^t}\right]^2 + \dots \tag{20}$$

Hence, if the observation error variances are small enough to make the non-linear terms in the series negligible, $\ln(y) - \ln(y^t) = \frac{(y - y^t)}{y^t}$ and hence

$$\left\langle \left[\ln(y) - \ln(y^t)\right]^2 \right\rangle = \left\langle \left[\frac{(y - y^t)}{y^t}\right]^2 \right\rangle = R_r. \tag{21}$$

which is independent of the state for the simple problem considered in sections 2 and 3. Consequently, if we took the logarithm of all the ensemble members and the observation, we could apply equation (6) to those transformed variables using R_r as an invariant observation error variance in place of *R*. Having formed the analysis ensemble in log space using (6), the ensemble in real space is readily obtained by taking the inverse logarithm of each of the ensemble members. Although this approach could be used in both variational and EnKF DA schemes, for the sake of a concise name, we shall refer to this approach as the ln-EnKF approach for the remainder of this paper.

The cyan line on Fig. 2a shows that the mse for the ln-EnKF is substantially lower than that for the EnKF with $R = \frac{1}{K}\sum_{k=1}^{K} R_r \left(y_k^f\right)^2$ with $R_r = 1$. However, Fig. 2b shows that the normalized ensemble dispersion for the ln-EnKF is $\tilde{d} = 2.1$ so that the analysis ensemble variance is, on average, more than two times as large as the mse. . Closer examination of the problem showed that while in log space the analysis ensemble variance was approximately equal to the analysis error variance. However, *after the inverse-logarithm had been applied to each ensemble member the mean square error of the analysis mean was no longer equal to the ensemble variance*. The disparity between mse and ensemble variance produced by the non-linear transformation would likely degrade the performance of a cycling ensemble DA scheme, so it is not clear that the ln-EnKF would be able to maintain a smaller mse than the EnKF with $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ which has an essentially perfect normalized analysis ensemble dispersion of $\tilde{d}=1.0$.

Fig. 3 shows that with a moderately accurate observing instrument having $R_r=0.1$, the ln-EnKF mse is now no better than the EnKF with $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ but the ln-EnKF is still over-dispersive ($\tilde{d}=1.2$). Hence, in this case, the EnKF with $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ is clearly superior. Fig. 4 shows that with a very accurate observing instrument having $R_r=0.01$, there is no significant performance difference between the EnKF with $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ and the ln-EnKF.

In summary, for the two performance metrics (mse and $\tilde{d}$) over the three R_r values considered, the EnKF with $R=\frac{1}{K}\sum_{k=1}^{K}R_r\left(y_k^f\right)^2$ was superior to the ln-EnKF in 5 out of the 6 measures. This result shows that deleterious effects of non-linear transformations that render the observation error variance independent of the state may be harmful to ensemble DA performance. This failure to provide ensemble variances that match mean square error is, arguably, an even more serious defect of the non-linear transformation approach than the fact, described in Bishop (2016), that non-linear transformations of unbiased observations create *biased* observations. The primary problem of the transformation approach was that, unlike an EnKF with accurately specified observation and forecast error covariances, the ensemble variance is not guaranteed to be equal to the analysis error variance. Bishop (2016) showed that the GIGG-EnKF, which does not need to apply a non-linear transform to the observations, provides ensemble variances that match analysis error variance even when its assumption that the prior pdf is a gamma pdf is violated. Unsurprisingly, in this case where the assumptions of the GIGG-EnKF are perfectly satisfied, Fig's 2-4 show that correcting the assumptions of the DA scheme regarding observation and forecast uncertainty without making any variable transformations results in a scheme that is always superior to the other methods under the mse metric while maintaining acceptable $\tilde{d}$ values of normalized ensemble dispersion between 0.98 and 0.99.

5.Concluding remarks

For a variety of reasons, the observation error variance can depend on the unknown value of the true state. Here, the inescapability of such dependencies was made clear via a proof that the variance of the error of unbiased observations of bounded variables must tend to zero as the true value of the variable approaches the bound.

Current operational DA methods are based on analyses that often assume that the observation error variance R is independent of the true state. The question of how best to define the observation error variance used in these schemes when state dependent observation error variance is present was addressed. Experiments were performed using a simple univariate system with a semi-positive definite variable and an observing instrument whose observation error standard deviation was proportional to the unknown true state. Three possible ways of defining R were tested: (i) the average of the true observation error variances over the prior ensemble forecast $\left(R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2 \right)$, (ii) the observation error variance that would occur if the unknown truth was equal to the average of the ensemble mean and the observed value $\left(R = R_r\left(\frac{y+y^b}{2}\right)^2 \right)$, and (iii) the true observation error variance $\left(R = R_r\left(y^t\right)^2 \right)$. The DA performance of these three approaches was assessed using mean square error and ensemble dispersion metrics. Under these metrics, it was found that assuming $R = \frac{1}{K}\sum_{k=1}^{K} R_r\left(y_k^f\right)^2$ gave better performance than assuming $\left(R = R_r\left(\frac{y+y^b}{2}\right)^2 \right)$ which, in turn, gave better performance than assuming $R = R_r\left(y^t\right)^2$. Note that method (ii), the second-best method, has similarity to work by Geer and Bauer (2011), Okamoto et al. (2014) and Harnisch et al. (2016) where the average of a cloud aspect of the background forecast and the observation was used as a predictor of observation error standard deviation.

To reveal the underlying reason for these results, a derivation of the Kalman gain of the EnKF was given that highlights the fact that a necessary condition for the EnKF to yield the best linear unbiased estimator of the state is that its observation error variance parameter R be equal to the the mean of the distribution of observation error variances implied by the prior distribution of truth. This general theoretical result provides a clear goal for observation error variance specification for EnKFs in the presence of state dependent observation error variances.

In linear systems, the states that minimizes the 4D-Var or 3D-Var cost functions are identical to those given by a Kalman smoother and filter, respectively. Hence, in such circumstances, the only way they can deliver the best linear unbiased estimator of the state is if they set the observation error variance parameter R to be equal to the mean of the distribution of observation error variances implied by the prior distribution of truth. However, if the system were non-linear, the 4D-Var and 3D-Var cost functions are intended to identify the most likely state (the mode) rather than the state that minimizes error variance (the mean). To obtain the most likely state one needs to penalize the departure of each candidate state from the observed state using the observation error variance associated with the candidate state. Hence, in 4D-Var or 3D-Var systems designed to find the mode one would need to inform the algorithm of the observation error variance associated with the current best guess of the state rather than the mean of the observation error variances over the prior. Thus, in the 2nd and higher outer loops, the observation error variances one would specify in a mode-seeking 4D-Var and 3D-Var would likely be different to those used in a minimum error variance seeking EnKF.

We also considered an alternative approach in which one applies a non-linear transform to the observation to render its error variance independent of the state. For the simple system considered, the logarithmic transform provided this property. The logarithm of the state and the observation could then be fed into the EnKF equations to provide an update of the logarithm of the state variable. However, while this approach gave competitive mean square analysis error, it also yielded systematically over-dispersive analysis ensembles. Notably, it was found that while the ensemble variance was equal to the mean square error of the analysis mean in log space, this important reliability constraint was strongly violated when the inverse logarithm of each of the ensemble members was taken. This failure to provide ensemble variances that match mean square error is, arguably, an even more serious defect of the non-linear transformation approach than the fact, described in Bishop (2016), that non-linear transformations of unbiased observations create *biased* observations.

Bishop's (2016) GIGG-EnKF was used to indicate the performance gains that could be made if DA algorithms could incorporate more accurate descriptions of forecast and observation uncertainty without using bias inducing and reliability destroying non-linear transforms. Specifically, the GIGG-EnKF gave the smallest mean square analysis errors for all the tested values of the relative observation error and analysis ensemble dispersion that matched the mean square error of the analysis mean.

Possible implications of this paper's findings for DA research and operations are as follows:

a. State dependent observation error variance should be anticipated and estimated whenever the observation is of a bounded variable. For both bounded and unbounded variables, more research is needed to improve methods for deducing how observation error covariances and/or the observation likelihood pdf $L(\mathbf{y}\,|\,\mathbf{x})$ change with changes in the true state.

b. The observation error variance that should be used in today's EnKF and variational DA schemes is the expected value of the observation error variance over the prior distribution of truth. This means that the part of the DA system that assigns observation error variances must have access to ensemble forecasts of the observed variable.

c. Basic research is needed to develop new DA algorithms that can better accommodate the typically non-Gaussian uncertainty distributions associated with state dependent observation error variances.

All of the Matlab code used in generating the figures and quantitative results for this paper can be obtained free of charge from the author via email.

Acknowledgments: The author gratefully acknowledges funding support from the Chief of Naval Research through the NRL Base Program, PE 0601153N and also the Australian Research Council's Centres of Excellence Scheme (CE170100023).

Appendix

Theorem: The observation error variance of unbiased semi-positive definite observations of semi-positive definite finite variables must tend to zero as the true value of the variable being observed approaches zero.

Proof: For each possible fixed value of the truth y^t, collect J independent error prone observations $y_j, j = 1,2,...,J$ and compute the corresponding set of independent observation errors $\varepsilon_j^o = y_j - y^t, j = 1,2,...,J$. The set C of all observation errors $\varepsilon_j^o = y_j - y^t, j = 1,2,...,J$ for any fixed truth y^t is composed of two disjoint sets. The set A of errors $\varepsilon_m^o = y_m - y^t, m = 1,2,...,M$ whose error values are all greater than or equal to zero and the set B of errors $\varepsilon_n^o = y_n - y^t, n = 1,2,...,N$ of N whose error values are all less than zero. In this proof, the error subscript index n is used to indicate that the error belongs to the set B of negative errors whereas the subscript m is used to indicate that the error belongs to the set A of semi-positive definite observation errors that are greater than or equal to zero. The intersection of sets A and B is the empty set while their union is the complete set C of observation errors. Note that when the sample size J is increased, M and or N will also increase. Also note that, in general, a different underlying distribution of observation errors will exist for each possible value of the truth. Whenever truth is changed, one needs to generate a new sample of J observation errors to investigate the characteristics of the pdf associated with the new value of truth. This proof pertains to how the variance of the distributions of observation errors change as the truth approaches zero. By the standard definition of variance, the observation error variance for a fixed value of truth y^t is given by

$$
\begin{aligned}
R &= \lim_{J\to\infty}\left(\frac{1}{J-1}\right)\left[\sum_{j=1}^{J}\left(\varepsilon_j^o\right)^2 - \left(\frac{1}{J}\sum_{j=1}^{J}\varepsilon_j^o\right)^2\right] \\
&= \lim_{J\to\infty}\left(\frac{1}{J}\right)\left\{\left[\sum_{m=1}^{M}\left(\varepsilon_m^o\right)^2 + \sum_{n=1}^{N}\left(\varepsilon_n^o\right)^2\right] - \left(\frac{1}{J}\left[\sum_{m=1}^{M}\varepsilon_m^o + \sum_{n=1}^{N}\varepsilon_n^o\right]\right)^2\right\} \\
&= \lim_{J\to\infty}\left\{\left[\left(\frac{M}{J}\right)\frac{\sum_{m=1}^{M}\left(\varepsilon_m^o\right)^2}{M} + \left(\frac{N}{J}\right)\frac{\sum_{n=1}^{N}\left(\varepsilon_n^o\right)^2}{N}\right] - \left[\left(\frac{M}{J}\right)\frac{\sum_{m=1}^{M}\varepsilon_m^o}{M} + \left(\frac{N}{J}\right)\frac{\sum_{n=1}^{N}\varepsilon_n^o}{N}\right]^2\right\} \\
&= \left[a\left\langle\left(\varepsilon_m^o\right)^2\right\rangle + b\left\langle\left(\varepsilon_n^o\right)^2\right\rangle\right] - \left[a\left\langle\varepsilon_m^o\right\rangle + b\left\langle\varepsilon_n^o\right\rangle\right]^2
\end{aligned}
\tag{A22}
$$

where $a = \lim_{J\to\infty}\{M/J\}$ and $b = \lim_{J\to\infty}\{N/J\}$ are the probabilities of semi-positive and negative errors, respectively, and the angle brackets indicate the expectation operator. For an unbiased observation

$$\left\langle\varepsilon_j^o\right\rangle = a\left\langle\varepsilon_m^o\right\rangle + b\left\langle\varepsilon_n^o\right\rangle = 0 . \tag{A2}$$

Using this relation in (A1) gives

$$R = \left[a\left\langle\left(\varepsilon_m^o\right)^2\right\rangle + b\left\langle\left(\varepsilon_n^o\right)^2\right\rangle\right]. \tag{A3}$$

Rearranging (A2) gives

$$\left\langle\varepsilon_m^o\right\rangle = -\frac{b}{a}\left\langle\varepsilon_n^o\right\rangle \tag{A4}$$

Because the smallest possible value the observation y can take is zero, it follows that the range of $\varepsilon_n^o = y - y^t$ is given by $0 > \varepsilon_n^o \geq -y^t$. Consequently,

$$\lim_{y^t \to 0} \left\langle \varepsilon_n^o \right\rangle = 0 \text{ and } \lim_{y^t \to 0} \left[\left\langle \left(\varepsilon_n^o \right)^2 \right\rangle \right] = 0 \tag{A5}$$

Using the fact that $\lim_{y^t \to 0} \left[\left\langle \left(\varepsilon_n^o \right)^2 \right\rangle \right] = 0$ in (A3) gives

$$R = a \left\langle \left(\varepsilon_m^o \right)^2 \right\rangle. \tag{A6}$$

Even in the absence of a hard physical bound like non-negativity, all assimilated observations have a finite range. For example, all assimilated observations of surface temperatures on Earth lie well within the range of -150 C to 150 C. Such bounds are *climatological* bounds rather than physical bounds. Hence, observations of all semi-positive definite variables have some climatological upper bound $y_{\max}$. It follows that the maximum possible value of ε_m^o is $y_{\max}$ and that $\left\langle \varepsilon_m^o \right\rangle$ will always be less than or equal to $y_{\max}$.

Using the fact from (A5) that $\lim_{y^t \to 0} \left\langle \varepsilon_n^o \right\rangle = 0$ in the zero-bias condition (A4) yields three possible variable combinations that ensure that the observation errors remain unbiased as $y^t \to 0$:

Possibility 1: $\left\langle \varepsilon_m^o \right\rangle \to 0$.

Possibility 2: $a \to 0$ and $y_{\max} > \left\langle \varepsilon_m^o \right\rangle > 0$.

Possibility 3: Both $a \to 0$ and $\left\langle \varepsilon_m^o \right\rangle \to 0$ simultaneously.

If possibility 2 or 3 occur, then (A6) implies that $R \to 0$ as $y^t \to 0$. In preparation for considering possibility 1, let $w_0 = \int_0^{\varepsilon_{\text{infinitessimal}}} \rho\left(\varepsilon_m^o \mid y^t\right) d\left(\varepsilon_m^o\right)$ give the probability that ε_m^o is less than or equal to some positive infinitesimal value $\varepsilon_{\text{infinitessimal}}$ where $\rho\left(\varepsilon_m^o \mid y^t\right)$ gives the probability density of ε_m^o for a fixed value of y^t . Because probability densities integrate to unity, $1-w_0$ gives the probability that $\varepsilon_m^o > \varepsilon_{\text{infinitessimal}}$. Consequently, by definition,

$$\begin{aligned}
\left\langle \left(\varepsilon_m^o\right)^2 \right\rangle &= \int_0^{\infty} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right)^2 d\left(\varepsilon_m^o\right) \\
&= \int_0^{\varepsilon_{\text{infinitessimal}}} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right)^2 d\left(\varepsilon_m^o\right) + \int_{\varepsilon_{\text{infinitessimal}}}^{y_{\max}} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right)^2 d\left(\varepsilon_m^o\right) \\
&\le w_0 \varepsilon_{\text{infinitessimal}}^2 + \left(1-w_0\right) y_{\max}^2
\end{aligned} \tag{A7}$$

and

$$\begin{aligned}
\left\langle \varepsilon_m^o \right\rangle &= \int_0^{\infty} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right) d\left(\varepsilon_m^o\right) \\
&= \int_0^{\varepsilon_{\text{infinitessimal}}} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right) d\left(\varepsilon_m^o\right) + \int_{\varepsilon_{\text{infinitessimal}}}^{y_{\max}} \rho\left(\varepsilon_m^o\right)\left(\varepsilon_m^o\right) d\left(\varepsilon_m^o\right) \\
&\le w_0 \varepsilon_{\text{infinitessimal}} + \left(1-w_0\right) y_{\max}^2
\end{aligned} \tag{A8}$$

From (A8), it is clear that possibility 1's condition that $\left\langle \varepsilon_m^o \right\rangle \to 0$ as $y^t \to 0$ can only be met if as $y^t \to 0$, $\varepsilon_{\text{infinitessimal}} \to 0$ and $w_0 \to 1$. Hence, because $\left(1-w_0\right)$ tends to zero as w_0 tends to unity, (A7) gives

$$\left\langle \left(\varepsilon_m^o\right)^2 \right\rangle \le \lim\left(y^t \to 0\text{ , } \varepsilon_{\text{infinitessimal}} \to 0 \text{ and } w_0 \to 1\right)\left\{\left[w_0 \varepsilon_{\text{infinitessimal}}^2 + \left(1-w_0\right) y_{\max}^2 \right]\right\} = 0 \tag{A9}$$

Using (A9) in (A6) gives R=0 even when $a > 0$. Thus, regardless of whether it is possibility 1, 2 or 3 responsible for keeping the observation unbiased as $y^t \to 0$, the observation error variance $R \to 0$ as $y^t \to 0$ as was required.

Corollary: *Unbiased observations of any variable with a known physical bound must have observation error variances that tend to zero as the true value of the observed variable approaches this physical bound*

Proof: The case in which the true value of the observed variable z has a physical bound that prevents it obtaining a value *less than* the constant c is easily handled by mapping z to the new variable y using $y = z - c$. Since y has a lower bound of zero, the proof of the corollary then follows from the theorem. Similarly, the case in which the observed variable z has an upper physical bound equal to the value *d,* one can simply make the coordinate transformation $y = d - z$ to create a corresponding semi-positive definite variable to which the preceding proved theorem applies.

References

Anderson JL., 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Bishop, CH. 2016. The GIGG-EnKF: Ensemble Kalman Filtering for highly skewed non-negative uncertainty distributions. *Quart. J. Roy. Met. Soc.* **142**, 1395–1412. DOI: 10.1002/qj.2742

Bishop CH, Etherton, BJ, Majumdar SJ. 2001. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.* **129,** 420-436.

Bishop, C.H. and E. A. Satterfield, 2013: Hidden error variance theory 1: Exposition and analytic model. *Mon. Wea. Rev.*, **141,** 1454-1468.

Bouttier, F, Vié B, Nuissier O, Raynaud L. 2012. Impact of Stochastic Physics in a Convection-Permitting Ensemble. *Mon. Wea. Rev*., **140**, 3706–3721.

Burgers G., van Leeuwen PJ, Evensen G. 1998. Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Courtier, P., Thépaut, J.-N, Hollingsworth, A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J.R. Meteorol. Soc*., **120**: 1367–1387. doi: 10.1002/qj.49712051912

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press. 457 pp.

Geer, A. J. and Bauer, P. 2011. Observation errors in all-sky data assimilation. *Q.J.R. Meteorol. Soc.*, **137**: 2024–2037. doi: 10.1002/qj.830

Harnisch F, Weissmann M, Periáñez, Á, 2016. Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Q.J.R. Meteorol. Soc.,* **142**: 1797–1808. doi: 10.1002/qj.2776

Okamoto K, McNally AP, Bell W. 2014. Progress towards the assimilation of all-sky infrared radiances: an evaluation of cloud effects. *Q.J.R. Meteorol. Soc.,* **140**: 1603–1614. doi: 10.1002/qj.2242

Houtekamer PL, Mitchell HL. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter. *Physica D*, **230,** 112–126.

Hyer EJ, Reid JS, Zhang J. 2011. An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, *Atmospheric Measurement Techniques*, **4**, 379-408, doi: 10.5194/amt-4-379-2011.

Mears CA, Smith DK, Wentz FJ. 2001. Comparison of Special Sensor Microwave Imager and buoy-measured wind speeds from 1987 to 1997, *J. Geophys. Res*., **106**(C6), 11719–11729, doi:10.1029/1999JC000097

Minamide M, Zhang F, 2017: Adaptive Observation Error Inflation for Assimilating All-Sky Satellite Radiance. **145**, 1063-1081

Rabier F, Järvinen H, Klinker E, Mahfouf J-F, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q.J.R. Meteorol. Soc*., **126**: 1143–1170. doi: 10.1002/qj.49712656415

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.R. Meteorol. Soc.*, **133**: 347–362. doi: 10.1002/qj.32

Schraff C, Reich H, Rhodin A, Schomburg A, Stephan K, Periáñez A, Potthast R. 2016. Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Q.J.R. Meteorol. Soc.*, **142**: 1453–1472. doi: 10.1002/qj.2748

Whitaker JS, Hamill TM. 2002. Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Whitaker JS, Hamill TM, Whitaker JS, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.,* **436**, 463-482. http://dx.doi.org/10.1175/2007MWR2018.1

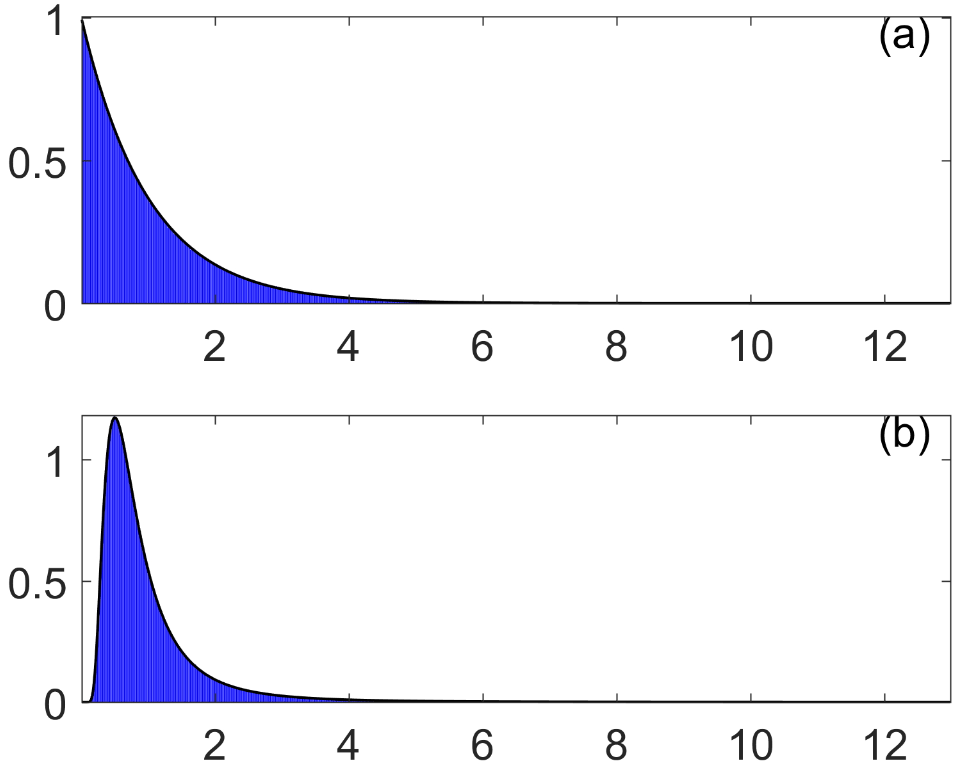
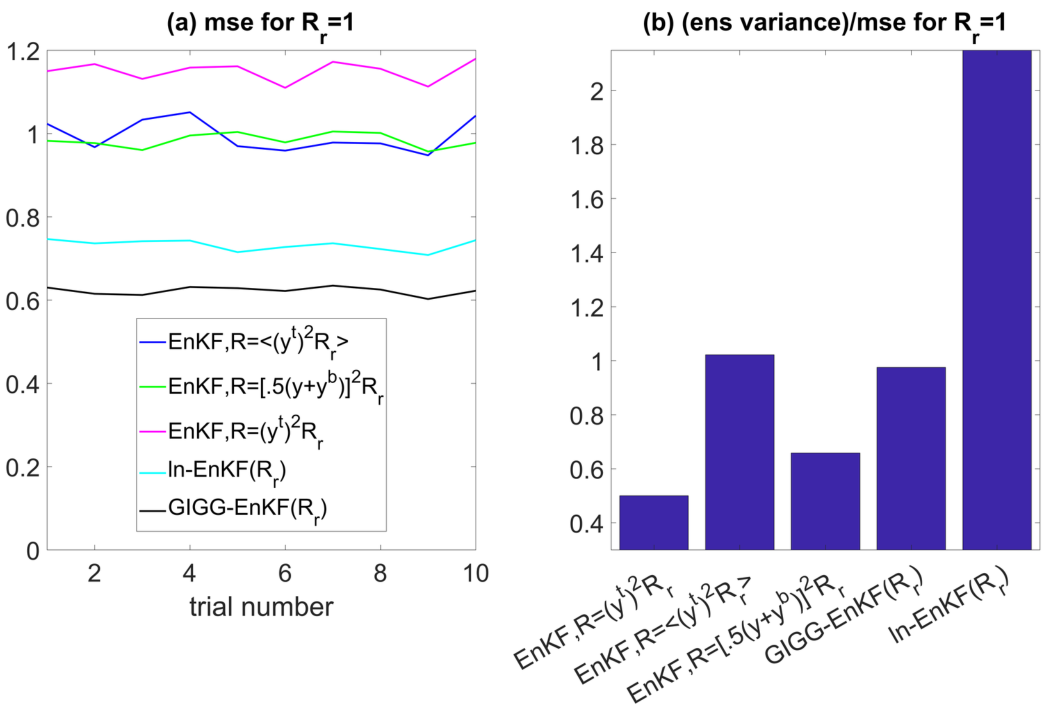


Fig1.tif



(a) mse for R_r=1
1.2
1
0.8
0.6
0.4
0.2
0
2
4
6
8
10
trial number
EnKF,R=<$(y^t)^2R_r$>
EnKF,R=$[.5(y+y^b)]^2R_r$
EnKF,R=$(y^t)^2R_r$
In-EnKF(R_r)
GIGG-EnKF(R_r)
(b) (ens variance)/mse for R_r=1
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
EnKF,R=$(y^t)^2R_r$
EnKF,R=<$(y^t)^2R_r$>
EnKF,R=$[.5(y+y^b)]^2R_r$
GIGG-EnKF(R_r)
In-EnKF(R_r)

Fig2.tif

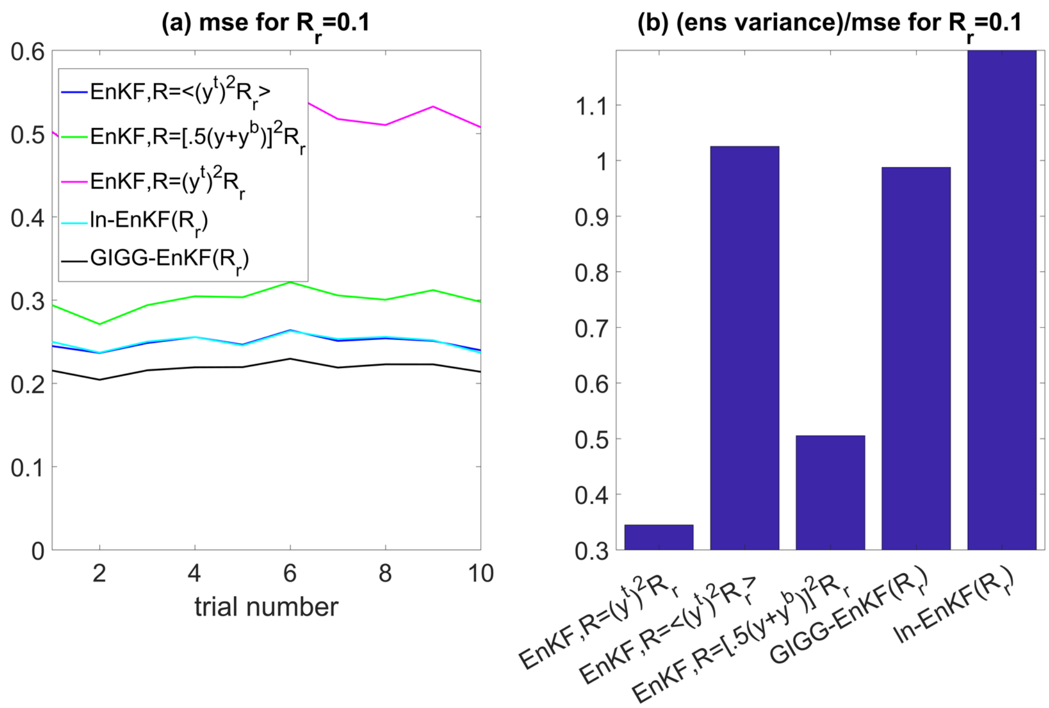
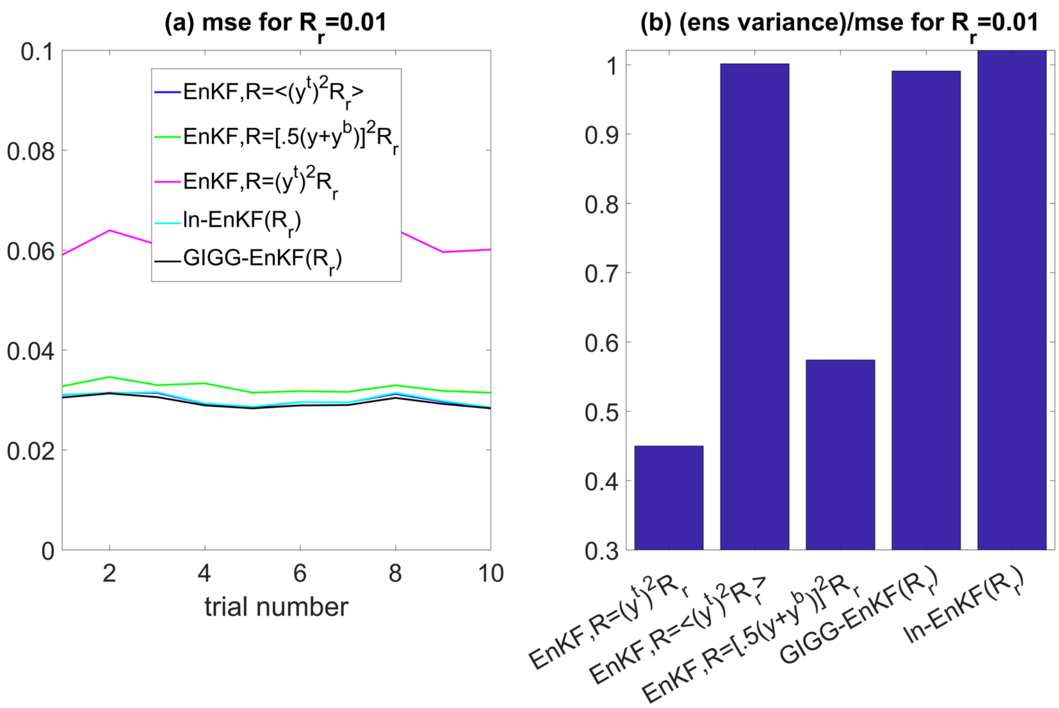

(a) mse for R_r=0.1
0.6
0.5
0.4
0.3
0.2
0.1
0
2
4
6
8
10
trial number
EnKF,R=<$(y^t)^2R_r$>
EnKF,R=[.5(y+y^b)]^2R_r
EnKF,R=$(y^t)^2R_r$
In-EnKF(R_r)
GIGG-EnKF(R_r)
(b) (ens variance)/mse for R_r=0.1
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
EnKF,R=$(y^t)^2R_r$
EnKF,R=<$(y^t)^2R_r$>
EnKF,R=[.5(y+y^b)]^2R_r
GIGG-EnKF(R_r)
In-EnKF(R_r)

Fig3.tif

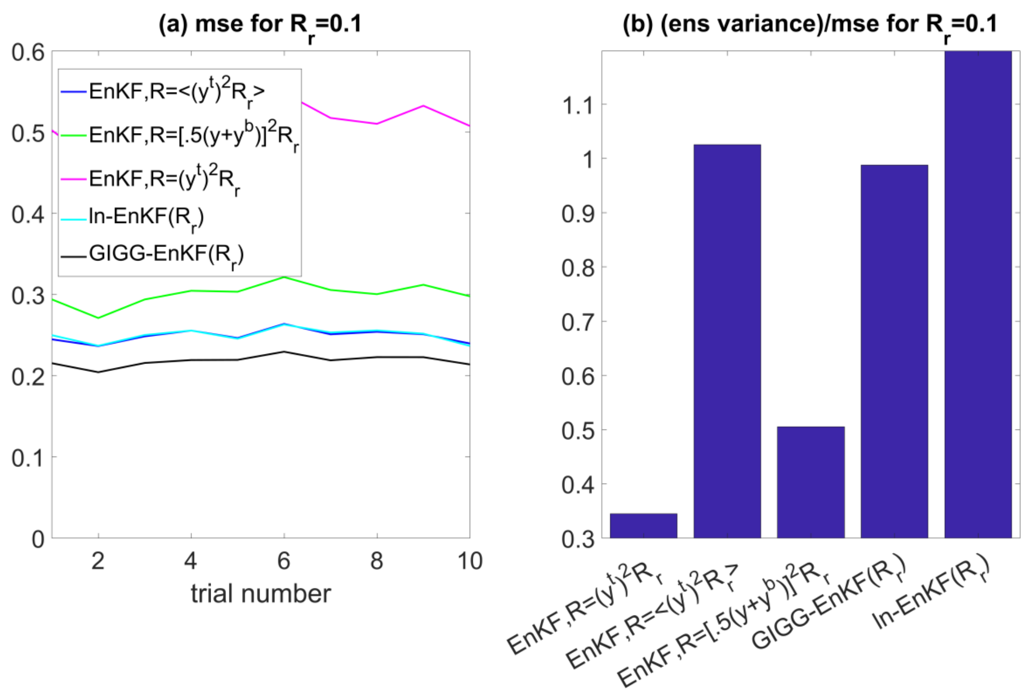


(a) mse for R_r=0.01
EnKF,R=<(y^t)^2R_r>
EnKF,R=[.5(y+y^b)]^2R_r
EnKF,R=(y^t)^2R_r
ln-EnKF(R_r)
GIGG-EnKF(R_r)
trial number
(b) (ens variance)/mse for R_r=0.01
EnKF,R=(y^t)^2R_r
EnKF,R=<(y^t)^2R_r>
EnKF,R=[.5(y+y^b)]^2R_r
GIGG-EnKF(R_r)
ln-EnKF(R_r)

Fig4.tif

Data assimilation strategies for state dependent observation error variances

Craig H. Bishop*



When the observation error variance *R* is a function of the unknown true state, *R is unknown.* It is shown that the *R* of unbiased observations of bounded variables must tend to zero as the true value approaches the bound. Three distinct strategies for choosing the *R* needed by EnKF and variational data assimilation schemes are considered. It is shown that letting *R* be the mean of the prior distribution of *R* values is the best of these choices.