



# Verification against perturbed analyses and observations

N. E. Bowler, M. J. P. Cullen, and C. Piccolo

Met Office, Fitzroy Road, Exeter, EX1 3PB, UK

Correspondence to: N. E. Bowler (neill.bowler@metoffice.gov.uk)

Received: 6 December 2013 – Revised: 13 April 2015 – Accepted: 15 June 2015 – Published: 24 July 2015

**Abstract.** It has long been known that verification of a forecast against the sequence of analyses used to produce those forecasts can under-estimate the magnitude of forecast errors. Here we show that under certain conditions the verification of a short-range forecast against a perturbed analysis coming from an ensemble data assimilation scheme can give the same root-mean-square error as verification against the truth. This means that a perturbed analysis can be used as a reliable proxy for the truth. However, the conditions required for this result to hold are rather restrictive: the analysis must be optimal, the ensemble spread must be equal to the error in the mean, the ensemble size must be large and the forecast being verified must be the background forecast used in the data assimilation. Although these criteria are unlikely to be met exactly it becomes clear that for most cases verification against a perturbed analysis gives better results than verification against an unperturbed analysis.

We demonstrate the application of these results in a idealised model framework and a numerical weather prediction context. In deriving this result we recall that an optimal (Kalman) analysis is one for which the analysis increments are uncorrelated with the analysis errors.

## 1 Introduction

Verification of forecasts is an important aspect in the development of those forecasts. Any improvement in the forecasting system should be tested to demonstrate that the forecasts are genuinely improved. Each forecast is typically launched from an analysis state which is a combination of observations with a previous short-range forecast from the system. A common practice is to use the analysis from such a system as the truth against which to verify (for instance see Buizza et al., 2005). Since each analysis depends on the forecasts from previous cycles this is a dangerous practice, particu-

larly at short forecast lead times (Bowler, 2008). Nonetheless the convenience of performing verification against a state which is available on the model grid means that this remains a common practice with its attendant problems (as observed in Clayton et al., 2013).

One solution to the problem of verification against analyses is to verify forecasts against observations. The observations do not depend on the forecast, and so provide an independent measurement of the true state of the system<sup>1</sup>. However, observations themselves are contaminated by errors. Methods exist to account for the effect of these errors on verification statistics (Ciach and Krajewski, 1999; Saetra et al., 2004; Bowler, 2006; Candille and Talagrand, 2008). However, these errors are often poorly known, so accounting for their effect is difficult. Additionally, there are often few conventional observations over the oceans, which means that verification statistics can be blind to these areas.

As an alternative solution to these problems, we offer the idea of performing the verification against a perturbed analysis.

## 2 Verification against perturbed analysis

We are looking to verify a forecast  $x^f$  using the root-mean-square (RMS) error. This forecast is a single realisation, and so could either be a forecast from a deterministic system or an ensemble mean forecast. Ideally one would verify this forecast against the true state of the system  $x^t$ , but this state is generally unknown. Given that the truth is unknown we choose to verify instead some other state, in this case an analysis. We consider that rather than having a single analysis we have an ensemble of analyses and verify against a randomly chosen analysis ensemble member. We assume that the anal-

<sup>1</sup>Although any time correlation in observation errors can create a correlation between forecast and observation errors.

ysis ensemble represents its own errors correctly. Since we are considering mean-square errors, then we only need this last statement to hold to second order; that is, we require that

$$\langle |\bar{\mathbf{x}}^a - \mathbf{x}^t|^2 \rangle = \langle |\mathbf{x}_i^a - \bar{\mathbf{x}}^a|^2 \rangle, \quad (1)$$

where  $|\mathbf{x}|^2 = \mathbf{x}^T \mathbf{x}$  denotes the inner product where  $T$  indicates the matrix transpose, and the angle brackets  $\langle \cdot \rangle$  indicate the average over a large number of cases. The ensemble states are denoted by  $\mathbf{x}_i^a$  where  $i$  is the ensemble member number and the overbar ( $\bar{\mathbf{x}}$ ) indicates the ensemble mean.

Given the above definitions we consider the RMS error calculated against a perturbed analysis, that is a randomly chosen member of the analysis ensemble. The mean-square error of the forecast against this analysis is

$$\begin{aligned} \langle |\mathbf{x}^f - \mathbf{x}_i^a|^2 \rangle &= \langle |(\mathbf{x}^f - \bar{\mathbf{x}}^a) - (\mathbf{x}_i^a - \bar{\mathbf{x}}^a)|^2 \rangle \\ &= \langle |\mathbf{x}^f - \bar{\mathbf{x}}^a|^2 \rangle + \langle |\mathbf{x}_i^a - \bar{\mathbf{x}}^a|^2 \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\mathbf{x}_i^a - \bar{\mathbf{x}}^a) \rangle. \end{aligned} \quad (2)$$

In this case we are considering the verification against a given, chosen ensemble member  $i$ , not against each ensemble member in turn. However, since all ensemble members are typically exchangeable, this distinction is not important. We do not include a time index in this notation since all quantities are valid at the same time.

To continue the analysis, we consider that there exists the truth state,  $\mathbf{x}^t$ , against which we would ideally conduct the verification. Using this we expand one of the terms appearing on the right-hand side of Eq. (2):

$$\begin{aligned} \langle |\mathbf{x}^f - \bar{\mathbf{x}}^a|^2 \rangle &= \langle |(\mathbf{x}^f - \mathbf{x}^t) - (\bar{\mathbf{x}}^a - \mathbf{x}^t)|^2 \rangle \\ &= \langle |\mathbf{x}^f - \mathbf{x}^t|^2 \rangle + \langle |\bar{\mathbf{x}}^a - \mathbf{x}^t|^2 \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \mathbf{x}^t)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle. \end{aligned} \quad (3)$$

Combining Eqs. (2) and (3), we find that

$$\begin{aligned} \langle |\mathbf{x}^f - \mathbf{x}_i^a|^2 \rangle &= \langle |\mathbf{x}^f - \mathbf{x}^t|^2 \rangle + \langle |\bar{\mathbf{x}}^a - \mathbf{x}^t|^2 \rangle \\ &\quad + \langle |\mathbf{x}_i^a - \bar{\mathbf{x}}^a|^2 \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\mathbf{x}_i^a - \bar{\mathbf{x}}^a) \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \mathbf{x}^t)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle. \end{aligned} \quad (4)$$

The last term in this equation can be further re-arranged:

$$\begin{aligned} \langle (\mathbf{x}^f - \mathbf{x}^t)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle &= \langle ((\mathbf{x}^f - \bar{\mathbf{x}}^a) + (\bar{\mathbf{x}}^a - \mathbf{x}^t))^T \\ &\quad (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle \\ &= \langle |\bar{\mathbf{x}}^a - \mathbf{x}^t|^2 \rangle \\ &\quad + \langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle. \end{aligned} \quad (5)$$

We have previously assumed that the ensemble of analyses is ideal (Eq. 1). Using this assumption and substituting Eq. (5) into Eq. (4), various terms cancel and we find

$$\begin{aligned} \langle |\mathbf{x}^f - \mathbf{x}_i^a|^2 \rangle &= \langle |\mathbf{x}^f - \mathbf{x}^t|^2 \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\mathbf{x}_i^a - \bar{\mathbf{x}}^a) \rangle \\ &\quad - 2 \langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle. \end{aligned} \quad (6)$$

So, if the last two terms in this equation are zero (or cancel), then we would expect that verifying against a perturbed analysis would give the same result as verification against the truth.

In the second to last term, the second bracket is the difference between a random analysis ensemble member and the ensemble mean. If this term were averaged over all the choices of the random member, then it is easy to see that this term is zero, since the mean of the second bracket would be precisely zero. If all the ensemble members are equivalent to each other, then this term should disappear if the number of cases is large enough.

If the final term also vanishes, then we can consider that the data-assimilation system is in some sense optimal. If the final term were not zero, then it would be possible to make the ensemble mean analysis closer to the truth by post-processing it using the difference  $\mathbf{x}^f - \bar{\mathbf{x}}^a$ . A statistically optimal analysis will not benefit from post-processing in this way because it is by design as close to the truth as possible, and so the final term must also be zero. This is a somewhat different definition of an ‘‘optimal’’ data assimilation scheme from the usual. This difference is explored in more detail in Sect. 4.

Therefore, we conclude that verification against a perturbed analysis will give the same RMS error as verification against the truth if the analysis ensemble is ideal (the spread equals the error of the mean analysis) and the analysis is statistically optimal (could not be improved by simple post-processing). In a sense Eq. (6) is a simple result, since we have assumed that the analysis ensemble correctly represents the errors in the ensemble mean analysis. However, this re-arrangement allows us to see that all that is required for perturbed analysis to be a good proxy for the truth is for two cross-terms to be zero. The first of these is straightforwardly zero; the condition for the second to be zero is more challenging, as will be seen below.

### 3 Verification against perturbed observations

It might be thought that, since a true observation is statistically indistinguishable from a random member of a set of perturbed observations, then verification against perturbed observations would also be equivalent to verification against the truth. However, we show that this is not the case.

Consider the final term in Eq. (6). If we replace the references to the analysis with the observations, then this term becomes  $\langle (\mathbf{H}\mathbf{x}^f - \mathbf{y})^T (\mathbf{y} - \mathbf{H}\mathbf{x}^t) \rangle$  where  $\mathbf{y}$  are the observations and  $\mathbf{H}$  is the observation operator, which we will assume to be linear for simplicity. Now, we choose to define

the observation using  $\epsilon^o$ , its departure from the truth

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \epsilon^o. \quad (7)$$

Using this definition, we find

$$\langle (\mathbf{H}\mathbf{x}^f - \mathbf{y})^T (\mathbf{y} - \mathbf{H}\mathbf{x}^t) \rangle = \langle (\mathbf{H}(\mathbf{x}^f - \mathbf{x}^t) - \epsilon^o)^T (\epsilon^o) \rangle. \quad (8)$$

If we assume that forecast and observation errors are uncorrelated, then this reduces to

$$\langle (\mathbf{H}\mathbf{x}^f - \mathbf{y})^T (\mathbf{y} - \mathbf{H}\mathbf{x}^t) \rangle = - \langle (\epsilon^o)^T \epsilon^o \rangle = -Tr(\mathbf{R}), \quad (9)$$

which is the trace of the observation error covariance matrix. Therefore verification against perturbed observations will not give the same result as verification against the truth.

Although the use of perturbed observations is unhelpful, it is possible to subtract the estimated observation error from the RMSE calculated using unperturbed observations. This has been used successfully by some authors (for instance see Bowler et al., 2008), but retains the limitation that observations do not universally cover the globe.

#### 4 Definitions of an optimal analysis

Earlier we indicated that an analysis for which  $\langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle = 0$  should be considered an optimal analysis, since it would not be possible to improve this analysis by a simple post-processing. This is the same as saying that the analysis increments are orthogonal to the analysis errors. However, a more usual definition of an optimal analysis is one which uses the Kalman gain in calculating the analysis state. In the following we will demonstrate that these two definitions of an optimal analysis are equivalent. The orthogonality of analysis increments and errors for an optimal filter has been known for many years (see for instance Kailath, 1968). We include a derivation of this fact here as it highlights certain assumptions which need to be made.

To calculate an analysis state we use the following formula:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f). \quad (10)$$

In this equation and the following paragraphs we refer to  $\mathbf{x}^a$  and  $\mathbf{x}^f$  without an overbar because this derivation can apply to any forecast and analysis and not simply one coming from an ensemble system.  $\mathbf{K}$  is the gain matrix applied to the innovations – this does not need to be the optimal (Kalman) gain. As in Eq. (7) the observation is defined by its departure from the truth,  $\epsilon^o$ . This allows us to re-arrange Eq. (10) as

$$\mathbf{KH}(\mathbf{x}^a - \mathbf{x}^t) = (\mathbf{I} - \mathbf{KH})(\mathbf{x}^f - \mathbf{x}^a) + \mathbf{K}\epsilon^o. \quad (11)$$

We post-multiply this equation by  $(\mathbf{x}^f - \mathbf{x}^a)^T$  and take the average over a large number of cases. This yields

$$\begin{aligned} \mathbf{KH} &< (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle \\ &= (\mathbf{I} - \mathbf{KH}) \langle (\mathbf{x}^f - \mathbf{x}^a)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle \\ &+ \mathbf{K} \langle \epsilon^o(\mathbf{x}^f - \mathbf{x}^a)^T \rangle, \end{aligned} \quad (12)$$

where we have assumed that  $\mathbf{K}$  and  $\mathbf{H}$  are constant in time. Note that in this equation the terms appear as  $\langle \mathbf{x} \mathbf{x}^T \rangle$ , which is the outer product where previously we have been dealing with terms like  $\langle \mathbf{x}^T \mathbf{x} \rangle$ , which is the inner product. Now, to deal with the terms on the right-hand side of this equation, we re-arrange the analysis Eq. (10) to be

$$\mathbf{x}^f - \mathbf{x}^a = \mathbf{KH}(\mathbf{x}^f - \mathbf{x}^t) - \mathbf{K}\epsilon^o. \quad (13)$$

We can square this equation, and take the average over a long time series to give

$$\begin{aligned} \langle (\mathbf{x}^f - \mathbf{x}^a)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle &= \mathbf{KH} \langle (\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T \rangle \\ &+ \mathbf{H}^T \mathbf{K}^T + \mathbf{K} \langle \epsilon^o(\epsilon^o)^T \rangle + \mathbf{K}^T, \end{aligned} \quad (14)$$

where we have assumed that the forecast and observation errors are uncorrelated. We re-write the forecast and observation covariance matrices using their usual terms  $\mathbf{B}$  and  $\mathbf{R}$  to give

$$\langle (\mathbf{x}^f - \mathbf{x}^a)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle = \mathbf{KHBH}^T \mathbf{K}^T + \mathbf{KRK}^T. \quad (15)$$

Returning to Eq. (13) we may multiply this by  $\epsilon^o$  to get the estimate of the second term as

$$\begin{aligned} \langle \epsilon^o(\mathbf{x}^f - \mathbf{x}^a)^T \rangle &= \langle \epsilon^o(\mathbf{x}^f - \mathbf{x}^t)^T \rangle \mathbf{H}^T \mathbf{K}^T \\ &- \langle \epsilon^o(\epsilon^o)^T \rangle \mathbf{K}^T. \end{aligned} \quad (16)$$

If we assume that forecast and observation errors are uncorrelated, then we find that

$$\langle \epsilon^o(\mathbf{x}^f - \mathbf{x}^a)^T \rangle = -\mathbf{RK}^T. \quad (17)$$

Substituting Eqs. (15) and (17) into Eq. (12) we find that

$$\begin{aligned} \mathbf{KH} &< (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle \\ &= (\mathbf{I} - \mathbf{KH})(\mathbf{KHBH}^T \mathbf{K}^T + \mathbf{KRK}^T) - \mathbf{KRK}^T. \end{aligned} \quad (18)$$

Expanding the right-hand side and cancelling terms, we get

$$\langle (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle = \mathbf{BH}^T \mathbf{K}^T - \mathbf{KHBH}^T \mathbf{K}^T - \mathbf{KRK}^T. \quad (19)$$

In Eq. (19) we have not made any assumption about the form of  $\mathbf{K}$ , and the terms labelled  $\mathbf{B}$  and  $\mathbf{R}$  are the true forecast- and observation-error covariance matrices. Previously we argued that Eq. (19) is zero if the gain matrix is equal to the Kalman

gain. So, we substitute the Kalman gain  $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$  for some of the terms in Eq. (19) to give

$$\begin{aligned} \langle (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^a)^T \rangle &= [\mathbf{B}\mathbf{H}^T \\ &\quad - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T \\ &\quad - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}] \mathbf{K}^T \\ &= 0. \end{aligned} \quad (20)$$

So, if we assume that the gain used in the data assimilation is optimal, then the key cross-term in Eq. (6) is zero. This is one of the conditions required for verification against a perturbed analysis to give the same RMS error as verification against the truth.

Now, Eq. (20) states that the outer product of the analysis errors with the analysis increment is zero. However, for the verification against a perturbed analysis to be a suitable substitute for verification against the truth we require the inner product of these two terms to be zero. If we have two vectors  $\mathbf{y}$  and  $\mathbf{x}$  then stating that the average of the outer product of these vectors is zero,  $\langle \mathbf{y}\mathbf{x}^T \rangle = 0$ , is the same as stating that

$$\langle y_i x_j \rangle = 0 \text{ for all } i, j. \quad (21)$$

If the inner product is to be zero, then we require that

$$\langle \mathbf{y}^T \mathbf{x} \rangle = \langle \sum_{i=1}^N y_i x_i \rangle = 0. \quad (22)$$

This demonstrates that Eq. (20) implies that  $\langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle = 0$ .

In this calculation the forecast  $\mathbf{x}^f$  is the one used in calculating the new analysis. Given that the analysis referred to in the last term of Eq. (6) is an ensemble mean, then  $\mathbf{x}^f$  should be the ensemble mean background forecast to the data assimilation. That is, we must re-write Eq. (6) as

$$\begin{aligned} \langle |\bar{\mathbf{x}}^f - \mathbf{x}_i^a|^2 \rangle &= \langle |\bar{\mathbf{x}}^f - \mathbf{x}^t|^2 \rangle \\ &\quad - 2 \langle (\bar{\mathbf{x}}^f - \bar{\mathbf{x}}^a)^T (\mathbf{x}_i^a - \bar{\mathbf{x}}^a) \rangle \\ &\quad - 2 \langle (\bar{\mathbf{x}}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle \\ &\approx \langle |\bar{\mathbf{x}}^f - \mathbf{x}^t|^2 \rangle, \end{aligned} \quad (23)$$

where  $\bar{\mathbf{x}}^f$  is the ensemble-mean background for the ensemble data assimilation. Thus the above argument does not apply to deterministic forecasts or longer lead time forecasts. The issue of longer lead times is discussed further in Sect. 7.

This derivation also informs how the analysis ensemble is created. Following Eq. (10) the update of the ensemble mean will follow

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f), \quad (24)$$

where  $\mathbf{K}$  is the optimal (Kalman) gain matrix. In Sect. 2 we assumed that the analysis ensemble perturbations are drawn from the same distribution as the analysis errors. One way to ensure this (Berre et al., 2006) is to update each ensemble member according to

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y} + \mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f), \quad (25)$$

where  $\mathbf{y}_i$  is a perturbation to the observations created using the (true) observation error covariance matrix,  $\mathbf{R}$ . Note that in both the above equations  $\mathbf{K}$  is the Kalman gain calculated using the true (unknown) background and observation error covariance matrices. This matrix is approximated in the ensemble Kalman filter and ensemble-variational methods used with geophysical models (Houtekamer et al., 1996; Evensen, 1994). In the following tests we use  $\mathbf{K} = \mathbf{B}\mathbf{K}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$  with  $\mathbf{B}$  and  $\mathbf{R}$  fixed.

## 5 Testing using a simple model

A toy-model data assimilation system was created to test whether the above assumptions can hold in an idealised context. For this, the logistic map was used (see for instance Peitgen et al., 1992). The logistic map is a single-variable chaotic map, iterated according to

$$x_{n+1} = Cx_n(1 - x_n), \quad (26)$$

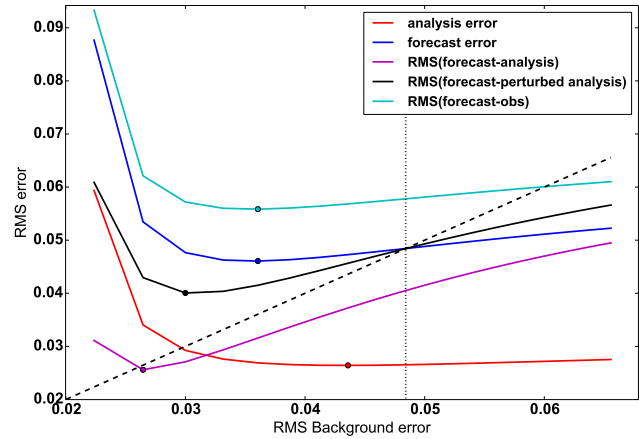
where  $C$  is a constant. The basin of attraction for this map is the range (0, 1), and states  $x > 1$  will diverge towards infinity. The map is chaotic when  $C > 3.57$  (approx.) and has a Hausdorff dimension of about 0.538 (Grassberger and Procaccia, 1983). In our experiments we choose  $C = 3.7$  as for this value the map exhibits chaotic behaviour.

We initialise an ensemble by randomly choosing states in the interval (0, 1). The logistic map is applied to each member to create a forecast ensemble. The forecast ensemble is transformed into an analysis ensemble by each member assimilating a perturbed observation. The observations are created by adding a perturbation to the run of the truth model. These perturbations are distributed according to  $\sim N(0, 0.001)$ . The perturbed observations are created from the observations by adding a perturbation sampled from the same distribution. The assimilation always uses a fixed background error variance,  $B$ , and we test the formulas derived above by varying the value of  $B$ . A fixed  $B$  is a poor approximation to the true background errors. This assimilation will not be optimal and we may find that  $\langle (\mathbf{x}^f - \bar{\mathbf{x}}^a)^T (\bar{\mathbf{x}}^a - \mathbf{x}^t) \rangle$  is non-zero. We examine this later. Observations are assimilated every time step and Eq. (26) is used to iterate both the ensemble members and the truth run. The first 2000 assimilation cycles are rejected as a spin-up period. Analysis states which fall outside the basin of attraction are reset to lie within

it. The assimilation is run for a further 200 000 assimilation cycles and 400 ensemble members are used. Confidence intervals were calculated using the bootstrap method assuming each of the assimilation cycles gives an independent sample of the analysis error. Since we use a long run the estimated confidence intervals are very narrow, and correspond approximately to the line width in the plots. Therefore these are not shown in order to aid clarity. In order to be consistent with the results of the previous section the only forecasts verified are the ensemble mean background forecasts. All results shown here have used the logistic map. Similar results have also been found with the models of Lorenz (1963, 1995).

Figure 1 shows the RMS background-forecast and analysis errors as a function of  $B$ . When  $B$  is small the forecast and analysis errors (dark blue line and red line, respectively) are large and the system is sub-optimal for these values. Verification against a perturbed analysis gives a systematically lower RMS error (RMSE) than verification against the truth (dark blue line) for small values of  $B$ , since insufficient weight is given to the observations. The RMS error for verification against a perturbed analysis becomes equal to that when verifying against the truth for moderate values of  $B$  ( $\sim 0.049$ ). This point is also where the RMS error crosses the diagonal, indicating that the background errors used in the assimilation are equal to the actual background errors, and the assimilation is optimal. Verification against observations gives RMS errors which are systematically higher than all the other estimates. If observation errors are accounted for, then verification against observations becomes very similar to verification against the truth (not shown). Verification against unperturbed analyses gives smaller RMSEs than all the other methods.

The circles in Fig. 1 indicate the point at which the RMS errors are minimised for each curve. The minimum RMSE for verification against analysis (purple line) is a value of  $B$  around 0.026 which is much lower than the optimal value of  $B$  for verification against the truth. The black line shows verification against perturbed analyses and the minimum RMS error for this curve is when  $B$  is around 0.03. This is much larger than the value of  $B$  for the minimum RMS error for verification against (unperturbed) analyses. However, this value of  $B$ , around 0.03, is much lower than the optimal (Kalman) value of around 0.049. When verifying forecasts against the truth (dark blue line) the minimum value of the forecast error is found for  $B$  around 0.036, lower than the optimal (Kalman) value. This statement may seem counter-intuitive – the lowest forecast error is found when the value of  $B$  used in the analysis is not equal to the forecast error. However, recall that the logistic map is a non-linear map and that the Kalman filter is only optimal for linear models. We have found a similar result with other models (the models of Lorenz, 1963, 1995). For both these models the forecast error is minimised when the value of  $B$  used is larger than the actual forecast error (the value given for the Kalman filter). For the logistic map the value of  $B$  which minimises the analysis



**Figure 1.** RMS error of the forecast and analysis using the logistic model as a function of the background error standard deviation used in calculating the analysis. The red and blue lines show the RMSE for the analysis and forecast measured against the truth state. The other lines show the RMSE of the forecast, when verified against a different proxy for the truth. Verification is calculated over 200 000 analysis and forecast cycles.

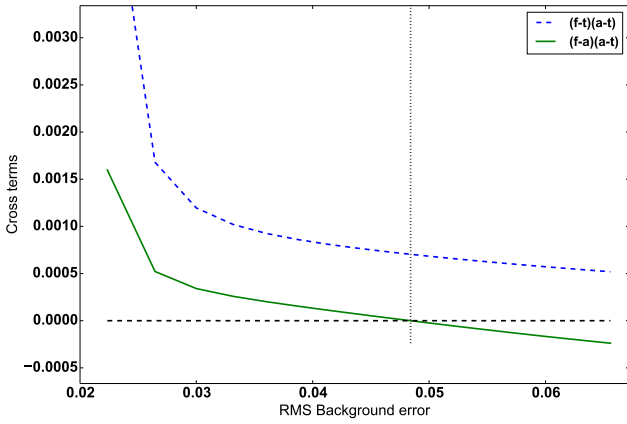
error is around 0.044, closer to the Kalman value than for the forecast error – this appears to be a result consistent across the different models.

The vertical line in Fig. 1 is the point at which the cross-term  $\langle (\bar{x}^f - \bar{x}^a)^T (\bar{x}^a - x^t) \rangle$  (last term of Eq. 23) is zero. We can see that this vertical line is at approximately the same value of  $B$  where the forecast and background errors are equal. This cross-term is plotted in Fig. 2, as the solid green line, as a function of  $B$ . Also plotted is the correlation between the forecast and analysis errors  $\langle (\bar{x}^f - x^t)^T (\bar{x}^a - x^t) \rangle$  (blue dashed line). This is non-zero for all the values of  $B$  run in these experiments. This demonstrates the problem in verifying against an unperturbed analysis that for all the values of  $B$  used here the errors in the forecast are correlated with the errors in the analysis.

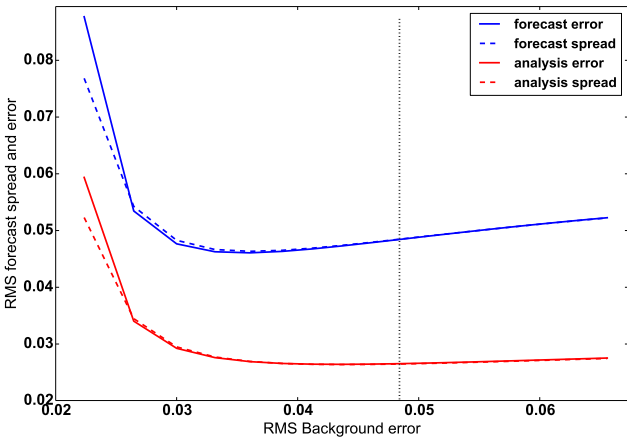
One of the conditions required for verification against perturbed analyses to give similar results to verification against the truth is for the analysis ensemble spread to equal the RMS analysis errors (Eq. 1). The analysis and forecast ensemble spread and error is plotted in Fig. 3. The ensembles appear to be well calibrated for most values of  $B$ . This may change if model error were introduced into the system.

## 6 Considering the effects of ensemble size

Next, we consider whether these results change substantially if fewer ensemble members are used. Results with a 10 member ensemble are shown in Fig. 4. This figure is rather similar to Fig. 1, with the most notable difference being that the vertical line no longer meets where the other lines cross.

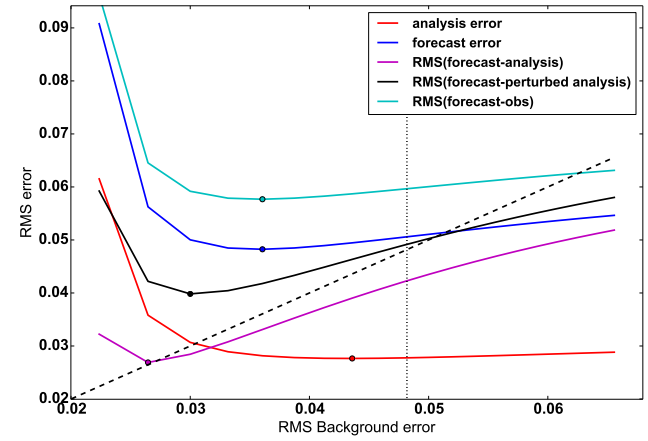


**Figure 2.** Important cross-terms calculated from a long analysis cycle using the logistic model, as a function of the background error standard deviation used in calculating the analysis. These are  $\langle (\bar{x}^f - x^t)(\bar{x}^a - x^t) \rangle$  (blue dashed) and  $\langle (\bar{x}^f - \bar{x}^a)(\bar{x}^a - x^t) \rangle$  (green solid).



**Figure 3.** RMS error and ensemble spread of the forecast and analysis using the logistic model, as a function of the background error standard deviation used in calculating the analysis. The ensembles were created by each ensemble member using the same assimilation method, assimilating perturbed observations.

To understand how ensemble size can affect the results, we need to return to estimates of the analysis error and spread. In Eq. (1) we relied on a cancellation of the analysis ensemble spread with the error of the ensemble mean. For a limited-size ensemble this cancellation does not hold precisely. As has been shown by Weigel (2011) the RMS error of an ensemble mean is slightly increased by effects related to the limited ensemble size. To show the limitations consider that the true state and each ensemble member are a random draw from the same distribution which has mean  $\mu$  and variance  $\sigma^2$ . We can thus write the truth as the mean of this distribution plus a deviation from the mean



**Figure 4.** RMS error of the forecast and analysis as plotted in Fig. 1, but using an ensemble with only 10 members.

$$x^t = \mu + s, \tag{27}$$

where  $\langle s \rangle = 0$  and  $\langle s^2 \rangle = \sigma^2$ . For an analysis ensemble member we would have

$$x_i^a = \mu + w_i, \tag{28}$$

where  $w_i$  is a random draw from the same distribution as  $s$ . Thus we may write the ensemble mean as

$$\begin{aligned} \bar{x}^a &= \mu + \bar{w} \\ &= \mu + \frac{1}{N} \sum_{i=1}^N w_i. \end{aligned} \tag{29}$$

We see that  $\bar{w}$  has mean zero and variance  $\sigma^2/N$  where  $N$  is the ensemble size. Using this Weigel (2011) showed that the mean-square error of the ensemble mean is

$$\begin{aligned} \langle (\bar{x}^a - x^t)^2 \rangle &= \langle \bar{w}^2 - 2\bar{w}s + s^2 \rangle \\ &= \sigma^2 \left(1 + \frac{1}{N}\right), \end{aligned} \tag{30}$$

since  $\langle \bar{w}s \rangle = 0$ . Due to the fact that the ensemble mean is not exactly equal to the mean of the distribution, the error of the ensemble mean is slightly larger than the variance of the distribution. This is a standard mathematical result (for instance see Hoel, 1984, p. 128). Using a similar argument lets us now consider the ensemble perturbations

$$\langle (x_i^a - \bar{x}^a)^2 \rangle = \langle w_i^2 - 2w_i\bar{w} + \bar{w}^2 \rangle. \tag{31}$$

From the definition of  $\bar{w}$  and recalling that the  $w_i$ 's are independent samples we see that

$$\langle w_i \bar{w} \rangle = \frac{1}{N} \langle w_i^2 \rangle = \frac{\sigma^2}{N} \quad (32)$$

and so

$$\langle (x_i^a - \bar{x}^a)^2 \rangle = \sigma^2 \left(1 - \frac{1}{N}\right). \quad (33)$$

So, the ensemble spread is slightly smaller than the variance of the distribution due to correlations between deviations of the ensemble mean from the distribution mean and the perturbations. This is often accounted for by using the unbiased estimator of the ensemble spread. Putting all this together, we find that for a well-calibrated ensemble

$$\frac{\langle (\bar{x}^a - x^t)^2 \rangle}{\langle (x_i^a - \bar{x}^a)^2 \rangle} = \frac{N + 1}{N - 1}. \quad (34)$$

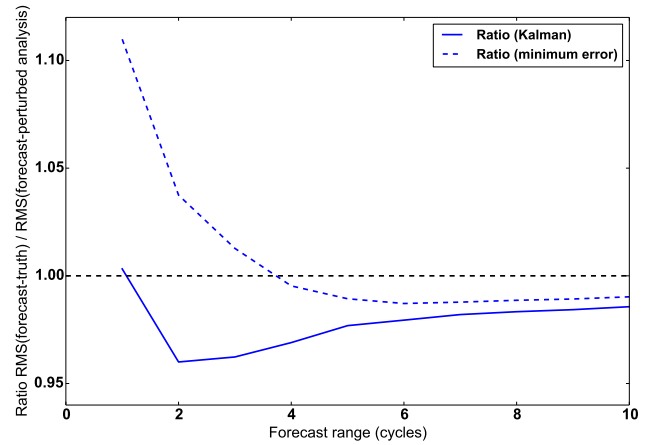
As the ensemble size goes to infinity this ratio tends to 1 and Eq. (1) holds. However, for a limited ensemble size these differences mean that verification against analysis is not the same as verification against the truth, even when the other conditions hold. This could be corrected for if the analysis spread is known.

### 7 Longer lead times

As was discussed in Sect. 4 the argument that the final term in Eq. (6) is zero requires the forecast being verified to be the background for the analysis. However, we might expect that this term is zero for longer lead times, since otherwise it should be possible to produce a superior analysis. To investigate this further we turn to the simple model tests used earlier.

Verification for longer lead times using the system described in Sect. 5 are given in Fig. 5. This shows the ratio of the RMSE measured against truth to the RMSE measured against perturbed analyses. This line is plotted for two choices of  $B$ . When the Kalman value of  $B$  is used the two verifications give the same RMS error at the first lead time (i.e. where the forecast is the background for the analysis). At longer lead times the RMS error when verifying against a perturbed analysis becomes larger than when verifying against the truth. This is caused by the final term in Eq. (6) giving a positive contribution to the verification against perturbed analysis. The interpretation is that  $x^t - \bar{x}^a$  and  $\bar{x}^f - \bar{x}^a$  are positively correlated – errors in the analysis are anti-correlated with differences between the forecast and the analysis. The correlation of analysis errors with forecast-analysis differences may be related to the use of a nonlinear model. The nonlinearity can lead to non-randomness of the errors which leads to the correlation.

Also shown in Fig. 5 is the ratio when  $B$  is chosen to be the value which gives the minimum forecast error – for the logistic map this value is lower than the Kalman value for



**Figure 5.** Ratio of the RMS errors of forecasts verified against truth and perturbed analyses using the logistic map for various lead times. For the solid line the background error was taken as the approximate Kalman value. For the dashed line  $B$  was taken for the value which minimises the short-period forecast error.

$B$ . In this case verification against perturbed analysis gives smaller RMSEs than verification against the truth at short lead times. At longer lead times the verifications cross over and the RMSE against perturbed analyses is greater than the RMSE against the truth.

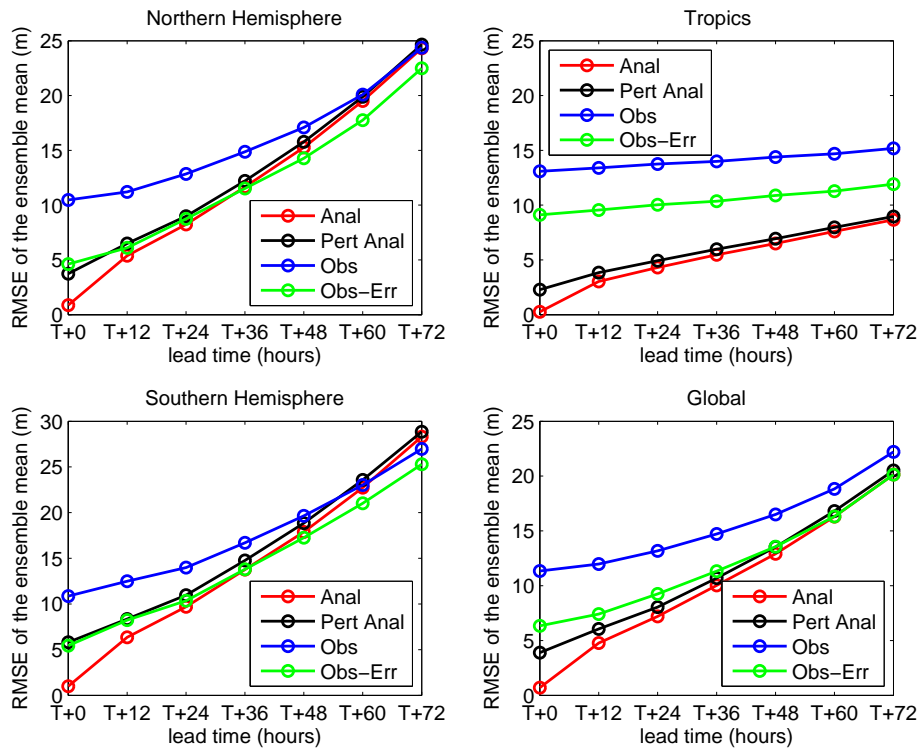
This behaviour at long lead times suggests that verification against a perturbed analysis is most useful at short lead times. Nonetheless it avoids the worst problems of verification against an unperturbed analysis. Therefore, we argue that it is still a useful replacement for that method of verification.

### 8 Verification of NWP forecasts

In order to understand whether this method can be applied to numerical weather prediction (NWP) systems we calculated the RMS error of a forecast ensemble mean against observations and perturbed analyses. The RMS error against analyses was calculated at observation locations so that the quantities are directly comparable.

Figure 6 shows the RMS error of the forecast ensemble mean as a function of lead time for 500 hPa geopotential height for the Met Office Global and Regional Ensemble Prediction System, MOGREPS (Bowler et al., 2008). At the time the forecasts were taken the MOGREPS ensemble consisted of a random sample of 11 members selected from 22 perturbed members used to cycle the ETKF every 6 h, plus the control member. The time average has been taken over 1 month of data. The different panels in Fig. 6 represent means over different geographical areas: Northern Hemisphere, tropics, Southern Hemisphere and the whole globe. Each panel shows the RMS error of the ensemble mean against the unperturbed analysis in red, the perturbed anal-





**Figure 6.** RMS errors of MOGREPS ensemble mean as a function of forecast lead time for forecasts of 500 hPa geopotential height. The forecast errors are reported for verification against observations and perturbed and unperturbed analyses.

yses in black, and the observations in blue, in green against the observations when the observation errors are accounted for. An observation error of 9.4 m (RMS) has been assumed.

Verification against observations gives RMS errors which are systematically higher than all other estimates, while verification against unperturbed analyses provides smaller RMS error than verification against observations and perturbed analyses. This is in agreement with Fig. 1. The exception is for the Southern Hemisphere, where the error against observations becomes smaller than the estimates against analyses after  $T + 60$  h. When observation errors are accounted for, the verification against the observations is very similar to the verification against perturbed analyses from  $T + 0$  h to  $T + 36$  h for the Northern and Southern hemispheres, while for longer lead times it gives lower RMS errors. This does not happen in the tropics since it is likely that verification includes the contribution of systematic errors which are not accounted for in the analysis perturbations. This is expected since 500 hPa geopotential height does not provide a good representation of what happens in the tropics.

The consistency of the RMS errors for short lead times in the northern and southern extra-tropics when calculated against perturbed analyses and observations (when subtracting observation error) suggests that this ensemble meets many of the required criteria. At longer lead times the RMS error against perturbed and unperturbed analyses gives larger

errors than for verification against observations, when subtracting observation error. This is consistent with the results in Fig. 5 – when analysis and forecast errors are no longer correlated the effect of analysis error is to over-estimate the RMSE.

## 9 Conclusions

We have shown that verification against a perturbed analysis gives the same RMS errors as verification against the truth, under certain conditions. These conditions require that the analysis ensemble is ideal (its RMS spread matches the RMS error in the mean analysis), that the analysis is optimal and that the ensemble size is large. Although NWP data assimilation systems are typically well tuned (to maximise forecast performance), none of these conditions is likely to hold exactly in practice. Additionally, the above results only apply to a forecast which is the background for the analysis against which it is verified.

In spite of these limitations we believe that this may be a useful approach to verification. Firstly it will give more realistic results than verification against an unperturbed analysis in most situations. Secondly the alternative is to verify against observations and explicitly account for the effect of observation error. Given the difficulty in estimating observation error and the fact that many parts of the world are



sparingly observed, this has its own limitations. The verification results for NWP forecasts indicate it gives very similar results to verification against observations, when observation error is accounted for, for short lead times in the extratropics. Given that the problems of verification against unperturbed analyses are most pronounced at short lead times, our method is potentially valuable for verification of short-term NWP forecasts.

It would be interesting to further explore some of the aspects of this method. For instance, what is the effect of using an analysis ensemble which is over-spread in some areas and under-spread in others? This study also demonstrated that for a non-linear model the Kalman filter solution may not minimise the system's forecast error. We feel that a better understanding of this result would be beneficial.

*Acknowledgement.* The analysis of limited ensemble size came about through discussion with Jonathan Flowerdew. Rob Darvell gave extensive assistance in the verification of the NWP forecasts.

Edited by: O. Talagrand

Reviewed by: five anonymous referees

## References

- Berre, L., Stefanescu, S., and Pereira, M.: The representation of the analysis effect in three error simulation techniques, *Tellus A*, 58, 196–209, 2006.
- Bowler, N. E.: Explicitly accounting for observation error in categorical verification of forecasts, *Mon. Weather Rev.*, 134, 1600–1606, 2006.
- Bowler, N. E.: Accounting for the effect of observation errors on verification of MOGREPS, *Meteorol. Appl.*, 15, 199–205, 2008.
- Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E.: The MOGREPS short-range ensemble prediction system, *Q. J. Roy. Meteorol. Soc.*, 134, 703–722, 2008.
- Buizza, R., Houtekamer, P., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y.: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, *Mon. Weather Rev.*, 133, 1076–1097, 2005.
- Candille, G. and Talagrand, O.: Impact of observational error on the validation of ensemble prediction systems, *Q. J. Roy. Meteorol. Soc.*, 134, 957–971, 2008.
- Ciach, G. J. and Krajewski, W. F.: On the estimation of radar rainfall error variance, *Adv. Water Resour.*, 22, 585–595, 1999.
- Clayton, A. M., Lorenc, A. C., and Barker, D. M.: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, *Q. J. Roy. Meteorol. Soc.*, 139, 1445–1461, 2013.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics, *J. Geophys. Res.-Oceans*, 99, 10143–10162, 1994.
- Grassberger, P. and Procaccia, I.: Measuring the strangeness of strange attractors, *Physica D*, 9, 189–208, 1983.
- Hoel, P. G.: *Introduction to mathematical statistics*, 5th Edn., Wiley, 1984.
- Houtekamer, P., Lefavre, L., Derome, J., Ritchie, H., and Mitchell, H.: A system simulation approach to ensemble prediction, *Mon. Weather Rev.*, 124, 1225–1242, 1996.
- Kailath, T.: An innovations approach to least-squares estimation, Part I: linear filtering in additive white noise, *IEEE T. Autom. Control*, 13, 646–655, 1968.
- Lorenz, E. N.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130–148, 1963.
- Lorenz, E. N.: Predictability: a problem partly solved, in: *Proceedings of the seminar on predictability*, vol. I, ECMWF, Reading, Berkshire, UK, 1–18, 1995.
- Peitgen, H. O., Jurgens, H., and Saupe, D.: *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, New York, 1992.
- Saetra, O., Hersbach, H., Bidlot, J.-R., and Richardson, D. S.: Effects of observation errors on the statistics for ensemble spread and reliability, *Mon. Weather Rev.*, 132, 1487–1501, 2004.
- Weigel, A. P.: Ensemble forecasts, in: *Forecast verification: A practitioner's guide in atmospheric science*, edited by: Jolliffe, I. T. and Stephenson, D. B., Wiley-Blackwell, Chichester, England, p. 144, 2011.