

# Estimation of analysis and forecast error variances

By MALAQUIAS PEÑA<sup>1\*</sup> and ZOLTAN TOTH<sup>2</sup>, <sup>1</sup>*IMSG at Environmental Modeling Center, NCEP/NWS/NOAA, College Park, MD, USA*; <sup>2</sup>*Global Systems Division, ESRL/OAR/NOAA, Boulder, CO, USA*

(Manuscript received 17 June 2013; in final form 30 September 2014)

## ABSTRACT

Accurate estimates of error variances in numerical analyses and forecasts (i.e. difference between analysis or forecast fields and nature on the resolved scales) are critical for the evaluation of forecasting systems, the tuning of data assimilation (DA) systems and the proper initialisation of ensemble forecasts. Errors in observations and the difficulty in their estimation, the fact that estimates of analysis errors derived via DA schemes, are influenced by the same assumptions as those used to create the analysis fields themselves, and the presumed but unknown correlation between analysis and forecast errors make the problem difficult. In this paper, an approach is introduced for the unbiased estimation of analysis and forecast errors. The method is independent of any assumption or tuning parameter used in DA schemes. The method combines information from differences between forecast and analysis fields ('perceived forecast errors') with prior knowledge regarding the time evolution of (1) forecast error variance and (2) correlation between errors in analyses and forecasts. The quality of the error estimates, given the validity of the prior relationships, depends on the sample size of independent measurements of perceived errors. In a simulated forecast environment, the method is demonstrated to reproduce the true analysis and forecast error within predicted error bounds. The method is then applied to forecasts from four leading numerical weather prediction centres to assess the performance of their corresponding DA and modelling systems. Error variance estimates are qualitatively consistent with earlier studies regarding the performance of the forecast systems compared. The estimated correlation between forecast and analysis errors is found to be a useful diagnostic of the performance of observing and DA systems. In case of significant model-related errors, a methodology to decompose initial value and model-related forecast errors is also proposed and successfully demonstrated.

*Keywords:* uncertainty of analysis, forecast verification, estimation methods, data assimilation, ensemble forecasts

## 1. Introduction

Assessing the performance of analysis and forecast systems rely on accurate estimates of error variances in numerical analysis and forecast fields. Additionally, accurate estimates of error variances are necessary to optimise the performance of Data Assimilation (hereafter DA) and Numerical Weather Prediction (hereafter NWP) systems. Examples include setting the background error variance in DA schemes at levels reflecting the true short-term forecast error variances (Fisher, 1996), and generating the initial perturbation variance in ensemble forecast systems consistent with the true analysis error variances.

### 1.1. Analysis error estimation

Most studies assessing the quality of NWP analyses follow one of three approaches. First, observations can be used as

proxy for truth and compared to analysis fields. This approach is useful during experimental observing campaigns or during coordinated enhanced observation periods. It requires that observations not be assimilated by the DA system that produced the analysis to be assessed. As DA schemes try to use all available observations, this restriction may preclude its use in assessing the analysis in an operational context. The other restriction is that not all the model variables are observable variables.

The second group of studies assessing analysis quality use methods related to the DA schemes themselves. This approach allows assessments continuous in time and space and on the scales of motions resolved by the numerical analysis. A drawback in the approach is that estimates of analysis errors in operational NWP systems are computationally prohibitive. Several simplifying approximations have been proposed to enable the estimation of analysis error variance in the context of variational DA schemes (e.g. Fisher and Courtier, 1995; Riishøjgaard, 2000). A vast literature of sequential DA methods has also emerged in

\*Corresponding author.  
email: malaquias.pena.mendez@noaa.gov

the past years where error variance estimates are provided in terms of an ensemble of analyses and forecasts (Evensen, 1994, 2003; Houtekamer et al., 1995). These methods operate in a severely truncated phase space (relatively small number of ensemble members compared to the huge number of model variables), requiring the introduction of various techniques to avoid filter divergence (i.e. inflation of ensemble forecast variance; Li et al., 2009) that renders any analysis error variance estimates qualitative.

Ideally, estimation of analysis errors should be independent on the DA scheme used and should not use the same assumptions made to produce the analysis. These assumptions include critical estimates of error variances for the observations (including representativeness errors), the background errors and the correlation of the errors. These quantities are, in fact, used as tunable parameters in complex DA systems (Parrish and Derber, 1992; Houtekamer et al., 1995).

A third approach is to use short-range forecasts (e.g. 6-hour atmospheric forecasts) as an indicator of analysis quality as it is strongly affected by errors in the analysis. A limitation is that forecast errors are not only a function of analysis errors; they are also influenced by errors and approximations in numerical model formulation. Other, more significant limitations of this approach are discussed in the next subsection.

Analysis error estimations have also been diagnosed using Observation System Simulation Experiments (OSSE; Daley and Mayer, 1986), by extrapolating back to initial conditions evolving short-range forecasts (Leith, 1978) and by computing differences among ensembles of analysis (Lorenc and Swinbank, 1984), and ensembles of forecasts (Wei et al., 2010). More recently, Desroziers et al. (2005) utilise the linear estimation framework (Talagrand, 1997) to determine diagnostic relationships between analysis, background and observation error covariances based on the statistics of O–B, O–A and B–A, where A, B and O stands for analysis, background and observation error covariances, respectively. Li et al. (2009) obtained the observational error variances by simultaneously optimising the inflation factor in an EnKF DA assimilation scheme.

### 1.2. Forecast error estimation

To evaluate the quality of numerical forecasts, either observations or numerical analyses can be used as a proxy for truth. Verifying forecasts against observations has the same limitations as discussed in Section 1.1 with respect to assessing analysis quality except that all observations can be considered independent (i.e. not used in creating the initial condition for the forecast). Note also that the error in any particular observation is larger than that in

numerical analyses that optimally combine information from various observing systems and a short-term numerical forecast (i.e. the prior or background field). Analyses offer a comprehensive estimate of nature (on the scales resolved); on the other hand, they may be influenced by systematic errors associated with the use of imperfect numerical models for the generation of the background forecasts. Either proxy, however, contains errors, which must be considered for the proper assessment of the skill in numerical forecasts.

The need to consider the effect of observational errors in verification statistics was noted by Candille and Talagrand (2008). They generalise probabilistic forecast metrics by defining and using an observational probability distribution accounting for the errors in observations. Candille and Talagrand (2008) found that considering the observational uncertainty lowers both forecast reliability and forecast discrimination capacity. Bowler (2006) explicitly accounts for the observational uncertainty in categorical verification forecasts. In Bowler’s approach, forecast errors against observations are partitioned into forecast errors against truth and observational uncertainty. Bowler showed that the presence of uncertainty in the observations systematically lowers the performance scores. In both studies, the observational uncertainty is presumed.

In the vast majority of studies in which NWP forecasts have been evaluated using numerical analysis fields as proxy for truth since the advent of numerical DA, the difference between forecasts and their verifying analysis is used to assess forecast error variances. This is based on the often implicit assumption that the numerical analysis fields have no (or have only negligible) errors in them. This assumption is violated especially at short lead times where forecast and analysis errors are of comparable magnitude. Ignoring the presence of analysis errors (and their correlation with forecast errors) in assessing forecast errors leads to underestimation of forecast errors in regions where observations are either scarce or not given enough weight when assimilated as has been reported frequently (e.g. Simmons, 1999). Biased error estimates in turn can lead to incorrect tuning of analysis and forecast systems (e.g. the inappropriate setting of analysis background error variance in DA, or initial perturbation variance in ensemble systems).

An approach to take into account the correlation between analysis errors and forecast errors in the evaluation of the accuracy of analyses and forecasts was introduced by Simmons and Hollingsworth (2002). In their study, the analysis error variance is estimated as follows. The mean square difference,  $d^2$ , between two sets of analyses from distinct analysis and forecast systems is decomposed into three terms using the law of difference of variances:  $d^2 = a_1^2 + a_2^2 + 2r_a a_1 a_2$ , where  $a_i$  is the analysis error of each system. The last (covariance) term involves a correlation parameter,  $r_a$ , between the two sets of analysis errors,

which cannot be determined as the analysis errors of the individual systems are not known. In Simmons and Hollinsworth (2002),  $r_a$  is assumed to be equivalent to the correlation of short-range forecast error variances (where forecast errors are defined as the difference between forecast and verifying analysis fields).

The current study continues this line of inquiry, recognising that the difference between forecast and verifying analysis fields may be a systematically biased estimate of the true and unknown forecast error, defined here as the difference between a NWP forecast field and reality, represented on the scales resolved by the DA and forecast system. In contrast to the true forecast error, we refer to the difference between forecast and verifying analysis fields, which in most studies is assumed to be a good and unbiased estimate of the true error as ‘perceived error’.

While the true analysis error for any particular case is unknown due to errors and other shortcomings in observations and DA procedures, its *statistical expectation* may be estimated in an unbiased way using simple principles. In the approach proposed below, perceived errors are considered as ‘observations’, which are then related to true analysis and forecast errors using ‘prior knowledge’: our scientific understanding of the evolution of true forecast error as a function of lead time and some basic observation regarding the effect of DA on analysis and forecast errors.

This paper introduces concepts and a method for the unbiased estimation of true analysis and forecast error variances based on a set of forecast and verifying analysis data. The paper is organised as follows. In Section 2, formal definitions and background for the estimation of the true analysis and forecast error variances are presented. In Section 3, the accuracy of the methodology is tested in a simple model environment, while in Section 4 the methodology is applied to operational global NWP systems. Section 5 provides a summary and conclusions.

## 2. Methodology

The estimation of analysis and forecast errors in this study is based on relationships between observed quantities (in this case, the perceived error measurements) and the variables of interest or target variables (true analysis and forecast errors). The relationships between the observed and unknown variables are considered part of prior knowledge about the nature of analysis errors and subsequent forecast errors, and how they are altered by the application of DA. As it will be seen, the prior knowledge will play a crucial role in facilitating the estimation process by lowering the number of unknowns and reducing uncertainty in the estimation process.

### 2.1. Decomposition of the perceived errors

To establish the relationship between the observed and unknown variables, we start with decomposing the perceived error variance,  $d_i^2$  as a function of true analysis and forecast errors (Ciach and Krajewski, 1999; Bowler, 2008):

$$d_i^2 = (F_i - A)^2 = ((F_i - T) - (A - T))^2 \equiv (x_i - x_0)^2 \quad (1)$$

where  $A$ ,  $F_i$  and  $T$  are, respectively, the verifying analysis, the forecast at lead  $i$  and the corresponding true state, all valid at the same verification time. The parentheses denote averages over space or time:  $(\cdot) = \frac{1}{K} \sum_{k=1}^{K=K} (\cdot)$ , where  $K$  is the number of observations in the sample. The forecast lead time is defined as  $t_i = i \cdot \Delta t$ , where  $i$  is the number of cycles of DA after initial time and  $\Delta t$  is the length of the analysis cycle (e.g.  $\Delta t = 6h$  for a typical DA cycle for global forecast systems). On the right hand side of eq. (1),  $x_0 = A - T$  and  $x_i = F_i - T$  are defined as the true analysis error and true forecast error, respectively. Because  $T$  is unknown both quantities can only be estimated statistically, which is the main topic of this study.

Taking a step beyond Ciach and Krajewski (1999), we use the law of sum of variances to rewrite the right hand side of (1):

$$d_i^2 = x_0^2 + x_i^2 - 2\rho_i x_0 x_i \quad (2)$$

Equation (2) denotes an inverse problem, where the measured perceived errors,  $d_i$ , are used to estimate the unknown functions  $x_0$  and  $x_i$  and their correlation,  $\rho_i$ . Consider  $i = 1$  ( $t_i = 6h$ ) in eq. (2). To facilitate the estimation of the three unknowns, one can introduce more equations like eq. (2) valid for various other lead times. By doing so, however, additional unknown variables ( $x_i$  and  $\rho_i$ ) are also introduced. Unless we can identify additional relationships between some of the variables the estimation problem cannot be solved.

### 2.2. Error evolution

To reduce the number of unknown variables in the set of equations like eq. (2), we seek to establish a temporal relationship between errors in forecasts with different lead times. The behaviour of growth in true forecast errors of atmospheric variables, particularly those arising from initial condition uncertainties, has been studied extensively both from theoretical (Lorenz, 1969a) and experimental points of view (e.g. using the spread of an ensemble; Lorenz, 1969b, 1982). The evolution of true errors from initial (analysis) to later (forecast) times with perfect models is governed by well-known perturbation or error growth dynamics (Lorenz, 1982; Trevisan et al., 1992), which can be described by a few parameters.

*2.2.1. Exponential error growth.* It is understood that linearly evolving (without nonlinear saturation of) errors in forecasts made with perfect models can be described by an exponential function (Lorenz, 1963):

$$x_1^2 = x_0^2 e^{\alpha t_1} \quad (3)$$

The expected evolution of errors in the short, mostly linear stage of error growth can therefore be described with just two parameters: the growth rate ( $\alpha$ ) and the initial analysis error ( $x_0$ ).

*2.2.2. Logistic error growth.* Beyond short lead times, exponentially growing forecast errors are affected by nonlinear processes leading to the saturation of errors. The effect of nonlinear saturation of errors is well understood and can be described by a logistic error growth model (Lorenz, 1982):

$$x_1^2 = \frac{S_\infty \cdot c}{e^{-\alpha t_1} + c} \quad (4)$$

where  $c = x_0^2 / (S_\infty - x_0^2)$ . With the logistic relationship, the mean evolution of errors can be described in a general fashion with just one additional parameter, the error saturation level ( $S_\infty$ ), beyond those used in the exponential error growth model ( $x_0$ ,  $\alpha$ ).

*2.2.3. Error growth in the presence of model errors.* NWP models offer only an approximate description of natural processes. NWP forecasts thus are affected not only by errors originating from imperfect initial conditions but also by errors arising from the use of imperfect NWP models. In contrast to initial-value-related errors, the emergence and temporal evolution of model-related errors are less understood. Assuming small errors associated with model imperfections are added at each time step of a numerical forecast, Leith (1978) proposed a constant growth rate for model-related errors. Dalcher and Kalnay (1987) and Reynolds et al. (1994) convolute this linear term with logistic error growth that describes the evolution of initial-value-related errors.

While acknowledging the relevance of linearly accumulating model-related errors, in the present study we focus our attention on ‘model drift’ that can be dominant especially in short lead time forecasts of tropical phenomena as convection is generally poorly parameterised in global scale models. Such drift is associated with a systematic difference between the attractors of nature and its model (Toth and Peña, 2007). We speculate that the drift originates when an analysed initial state is off the model’s attractor, the further away the state is from a model

trajectory the faster the fall-back motion is. To describe this behaviour, we propose a saturating exponential function:

$$x_1^2 = s_\infty - a \cdot e^{-\beta t_1}, \quad (5)$$

where  $s_\infty$  is the asymptotic difference between corresponding states on the attractors of nature and its model,  $s_\infty - a$ ;  $a < s_\infty$  is the magnitude of the drift-related error present in the initial condition, and  $1/\beta$  is the ‘e-folding’ time of errors from initial time to saturation.

*2.2.4. General error growth model.* While the leading singular vectors tend to dominate the dynamics of initial forecast errors (Szunyogh et al., 1997), the growth of drift-related errors might be more influenced by trailing vectors. It is convenient to assume that at least at short lead times the two types of errors do not interact. A general expression for forecast error evolution in such a case then takes the form of a sum of a logistic model [eq. (4)] representing the growth of initial errors ( $x_{i,in}$ ), and a saturating exponential model [eq. (5)] representing model drift-related error growth ( $x_{i,mo}$ ):

$$x_1^2 = x_{i,in}^2(t; x_0, \alpha, S_\infty) + x_{i,mo}^2(t; a, \beta, s_\infty) \quad (6)$$

Error growth in this general case will be controlled by six parameters, three for the logistic component of the error and three for the drift-related component.

### 2.3. Correlation between analysis and forecast errors

To further reduce the number of unknowns in eq. (2), we consider a prior knowledge regarding the impact of combining observations with a first guess (i.e. the DA process) on forecast errors. In particular, we consider the relationship between the correlation of analysis errors and errors in various lead time forecasts verifying at the time of the analysis. We first note the trivial fact that lacking any observations (i.e. no changes made to the first guess by DA), analyses are identical to the first guess and therefore the correlation of the error in various lead-time forecasts verifying at the time of a particular subsequent analysis is, by definition, one. The amount and quality of observational data entering into, and the quality of the DA scheme, determine the ‘effectiveness’ of a DA step, which is reflected in the correlation between the error in the first guess forecast and the error in the resulting analysis,  $\rho_1$ , eq. (2). The more effective the observing and DA systems are, the more error from the first guess forecast is eliminated, resulting in lower  $\rho_1$  values. Lower  $\rho_1$  is a necessary but not sufficient condition for improved DA efficiency as  $\rho_1$  can be decreased not only by removing existing first guess errors but also by the addition of new errors (e.g. noise) into the analysis not present in the first guess.

Let us assume that in a statistical sense successive DA steps use a similar amount but independent observational information and operate with the same effectiveness. Note also that in modern DA, observations are combined with a first guess forecast and therefore some of the errors originating from previous analyses are dynamically carried over or ‘cycled’ from one analysis time to the next (see Toth and Kalnay, 1993). Then, the cumulative effect of successive DA cycles is a succession of ‘rotations’ of forecast errors, as described next. To arrive to a quantitative relationship for the correlation we can view  $\rho_i$  as the cosine angle between an analysis error vector and a forecast error vector. The projections of the first guess,  $x_1$ , onto the analysis errors are  $A_1 = x_1 \rho_1$  and  $A_2 = x_1 \rho_2$ , respectively. The proportion of error reduced during the first cycle is the ratio between  $A_1$  and the first guess, whereas the proportion of error reduced during the second cycle is the ratio between  $A_2$  and  $A_1$ . Since a similar amount of independent information from observations is added on each analysis cycle, one can expect that the same proportion of the errors will be reduced in the two cycles. Thus,  $A_1/x_1 = A_2/A_1$ , or  $\rho_2 = \rho_1^2$ . A similar relationship can be obtained for the other lead times, resulting in the following equation:

$$\rho_i = \rho_1^i, \quad (7)$$

where  $\rho_i$  is the correlation between the error in a forecast of length  $i$  (in units of length of DA cycle) and the error in the analysis. We emphasise that this relationship is valid only for DA systems that use short-range forecasts initialised from the previous analysis as a background field. Notice that at sufficiently long lead times [eq. (7)] drops to zero and the perceived error variance in eq. (2) is simply the sum of the analysis and forecast error variances. Figure 1b indicates a good agreement with the true correlation for the Lorenz model when the analysis error is small.

With the relationship given in eq. (7), the temporal correlation between true errors in a forecast and true errors in subsequently made verifying analysis fields can be described with a single correlation coefficient  $\rho_1$ , which as described above, is related to the effectiveness of the observing and DA schemes in removing chaotically growing (or other) errors from the background field.<sup>1</sup>

<sup>1</sup> $\rho_i$  could be estimated through Observing System Simulation Experiments (OSSE). OSSEs, however, are expensive, would need to be repeated as data assimilation and numerical modelling systems evolve, and are based on various assumptions that may introduce unknown errors into the estimations.

#### 2.4. Estimation of parameters

Sections 2.1 through 2.3 describe the parameters involved in eq. (2). We here describe how those parameters come together. We define an estimator  $\hat{d}_i^2$  of the observed perceived error variance [eq. (2)]:

$$\hat{d}_i^2 = x_0^2 + x_i^2 - 2\rho_1^i x_0 x_i \quad (8)$$

where  $x_i = x_1$  ( $x_0$ ,  $\alpha$ , etc.), contains prior knowledge of the error evolution and whose parameters depend on the choice of the error evolution model [eqs. (3–6)]. The other parameter is  $\rho_1^i$ , which is given by eq. (7). To solve for the unknowns we use a set of equations like eq. (8), valid for estimated perceived errors of forecasts of different lead times. For stable statistical estimates, the number of eq. (8) included will be chosen so that it is well above the number of unknowns. The estimated perceived error variance  $\hat{d}_i^2$  based on various hypothetical values for the unknowns in eq. (8) must compare with the sample-based observed values of perceived error  $d_i$  on the left-hand side of eq. (2) in such a way as to minimise the following cost function:

$$J = \max(|d_i^2 - \hat{d}_i^2| \cdot w_i^{-1}) \quad (9)$$

The cost function has two important aspects. First, the absolute values of the squared difference (so-called  $L_\infty$  norm) rather than the more common square difference (so-called  $L_2$  norm) will ensure a good fit for the entire range of  $i$ , including the two extremes (i.e. at the shortest and longest lead times) of the estimated perceived error data (R. Krzysztofowicz, personal communication). Second, we use a weight function,  $w$ , which varies with lead time and is proportional to the sampling error – the expression of  $w$  will be described in the following subsection. The minimisation of eq. (9) in this study is carried out using the Nelder-Mead Simplex method (Lagarias et al., 1998). This technique is sensitive to first guess parameters; it is thus important to ensure that the starting point of the minimisation of the cost function is located close to the absolute minimum.

For the sake of simplicity, whenever feasible, we will attempt to use the simplest, exponential error growth model [eq. (3)]. In experiments where the logistic [eq. (4)] or the general error growth models [eq. (6)] are used, we consider all available data points to sample well the saturation parameters. To prevent unrealistic solutions particularly when the six parameters model is used, the following common sense constraints are added to the cost function [eq. (6)]:  $0 < \rho_1 < 1$ ,  $x_0^2 > 0$  and  $s_\infty > x_0$ .

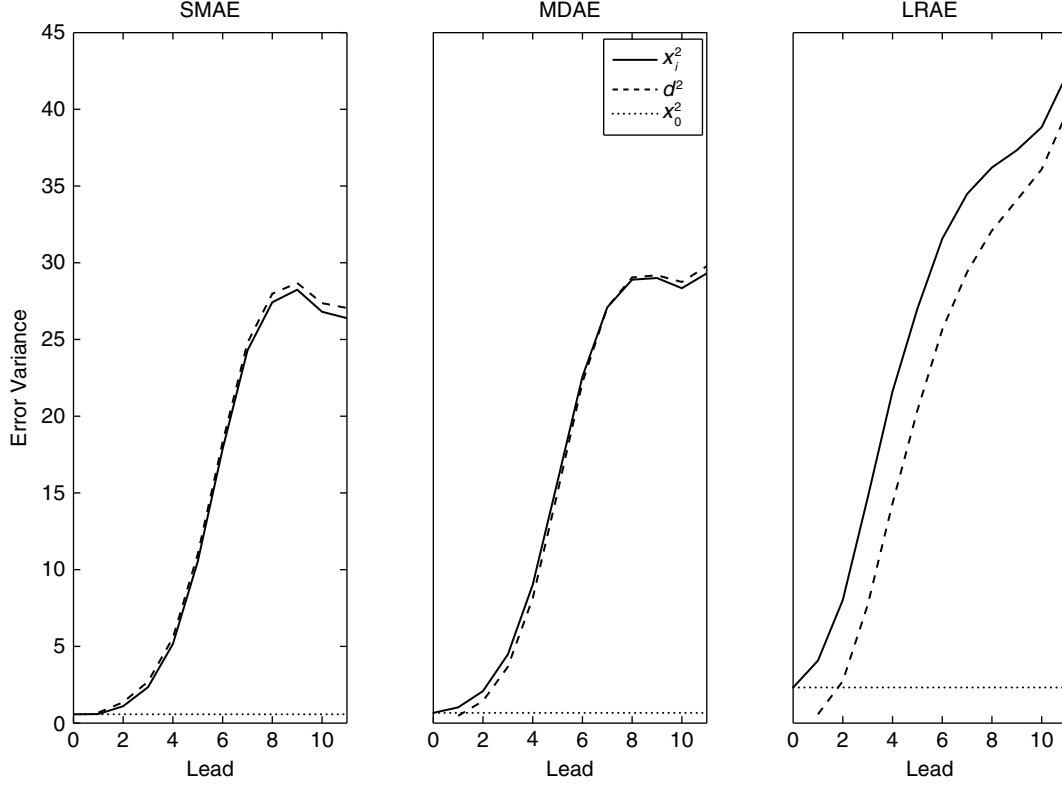


Fig. 1. Sample mean of true (solid line) and perceived (dashed line) forecast error variances, denoted respectively by  $x_i^2$  and by  $d_i^2$  as a function of lead time for the three experiments. The analysis error, denoted by  $x_0^2$ , is added (dotted line) as a reference. The sample includes  $10^4$  cases.

### 2.5. Error estimates associated with observed and estimated quantities

We here describe how we define  $w$  in eq. (9). As in any minimisation problem, sampling errors in observed values must be properly considered. Though differences between analysis and forecast fields can be computed exactly, their expected value, which is used as observed quantity in this study, is subject to sampling errors. The sampling uncertainty tends to grow with lead time due to larger variance and temporal correlation in those errors. Longer range time mean perceived errors, therefore, will have to be given smaller weight in the minimisation procedure.

To take this into consideration, the weights in eq. (9) are made proportional to the sample standard error of the mean (SEM), which reflects uncertainties in the time mean estimates of the expected perceived errors as a function of lead time. The  $SEM_i$  for lead time  $i$  is given as follows:

$$SEM_i = \frac{sd_i}{\sqrt{N}} \cdot f, \quad (10)$$

where,  $sd_i$  denotes the sample standard deviation at lead time  $i$ ,  $N$  is the sample size,  $f = \sqrt{(1+r_1)(1-r_1)^{-1}}$  is the

adjustment coefficient for serial correlation in the sample (Bence, 1995), and  $r_1$  is the autocorrelation in the sample. The weights  $w_i$  in eq. (9) are defined by considering the relative errors in time mean estimates of the expected perceived error for various lead times:

$$w_i = \frac{SEM_i}{\sum_i SEM_i} \quad (11)$$

As SEM values indicate the uncertainty in time mean estimates of perceived errors, they will also be used as expected deviations between the observed and estimated perceived error values (see Sections 3 and 4).

### 3. Application in a simulated forecast environment

In this section, the method described in Section 2 will be tested in a simulated environment where the true state is known. In such a setting, a direct evaluation of the estimation approach is possible.

### 3.1. Experimental design

The simulated environment is based on the Lorenz (1963) model. First, a long ( $5 \times 10^6$  time steps) integration is made to represent nature. Then, synthetic observations of all three model variables are created by adding noise to the nature run. The noise has a distribution  $N(0, \sigma_o^2)$ , where  $\sigma_o^2$  denotes the observation error variance, which will be set for the experiments to be analysed. The observations are assimilated every eight time steps using a 3-DVar DA scheme (e.g. Kalnay, 2002) with a forward operator  $\mathbf{H} = \mathbf{I}$ . The resulting analyses are used as initial conditions for forecasts out to 160 time steps (a length of 20 DA cycles.) The forecast model is also the Lorenz (1963) with identical parameter values as nature. In other words, these are perfect model experiments.

The background error covariance matrix used in the DA scheme is generated using lagged forecast differences (the NMC method; Parrish and Derber, 1992) from an independently generated simulation experiment like that described above except based on a different nature run. The resulting background error covariance matrix is multiplied by a tuning coefficient  $\kappa$  to control its effect in the DA procedure.

### 3.2. Experiments

With appropriate choices of tuning parameters  $\sigma_o$  and  $\zeta$ , various experiments characterised with analysis errors of different size can be carried out. To illustrate the concepts in this study three representative experiments are described for small, moderate and large analysis errors designated respectively as S, M and L. Table 1 shows the parameters specified in the DA scheme for these experiments and the resulting first guess ( $x_1^2$ ), and analysis ( $x_o^2$ ) sampling error variances. Larger prescribed observation error variances lead to larger first guess and analysis error variances. Resulting analysis error variances and corresponding forecast error variances in these experiments are shown in Fig. 1. The correlations between analysis errors and forecast errors

Table 1. Parameters prescribed to the data assimilation experiments and resulting sample error variances from a time series of  $10^4$  cycles.

Experiment	Prescribed		$x_1^2$	$x_o^2$
	$\sigma_o^2$	$\zeta$		
S	2	0.5	0.59	0.57
M	4	0.3	1.03	0.67
L	9	0.3	4.09	2.33

$\sigma_o^2$  is the observation error variance and  $\zeta$  is the tuning parameter of the background covariance error matrix.

are shown in Fig. 2, along with the theoretical correlation (dashed lines) given by eq. (7).

In experiment S (Fig. 1a), the perceived error is approximately the sum of the true analysis and forecast errors at all lead times. This is due to the low correlation value shown in Fig. 2a [cf. eq. (2)]. This case represents a NWP system with very low analysis error variance where the large information content of the observations results in low correlation between true analysis errors and true background (and longer range) forecast errors. In other words, in this situation information in the observations regarding the true state of the system can remove a large part of the error present in the background field.

In the M and even more so in the L experiment, perceived forecast errors underestimate the analysis and forecast errors. This situation occurs in NWP systems when its DA scheme can extract only a relatively small amount of information from the observations either because scarce or inaccurate observational data or inadequate DA methods. In such cases, the analysis relies more on the background forecast whose errors are largely left unchanged, leading to high analysis – forecast error correlation values (see Fig. 2b and c). For the L experiment the correlation is too large and the relationship [eq. (7)] is not valid.

### 3.3. Error estimation

To test the error estimation method introduced in Section 2, we pretend that truth and therefore the true errors and the correlation between analysis and forecast errors are unknown. We will statistically estimate the latter quantities based on observed perceived errors and prior knowledge described in Sections 2.2 and 2.3. In particular, as our focus is on estimating analysis and short-term forecast errors, we will assume that small initial errors grow exponentially [eq. (3)] and will limit the use of perceived error data to short lead times (up to five time units). We additionally assume that the correlation between analysis and forecast errors valid at the same time follows the relationship given in eq. (7).

Figure 3 shows the error estimations obtained for the S experiment as a function of lead time. First, we determine a good fit to the perceived errors by minimising eq. (9). The resulting fitting curve is the estimated perceived error variance  $\hat{\sigma}_t^2$  [eq. (8)], from which the other parameters are derived. As shown in Fig. 3a, the estimated perceived error variance (dashed) fits the measured perceived error variance indicating that the exponential curve can describe the forecast error growth with good accuracy. We added the SEM (10) interval for each lead time as a reference. In Fig. 3b and c, the intervals are determined by bootstrapping. We sample many possible values of the three ( $\alpha$ ,  $x_0$  and  $\rho_1$ ) independent parameters involved in eq. (8) such that the

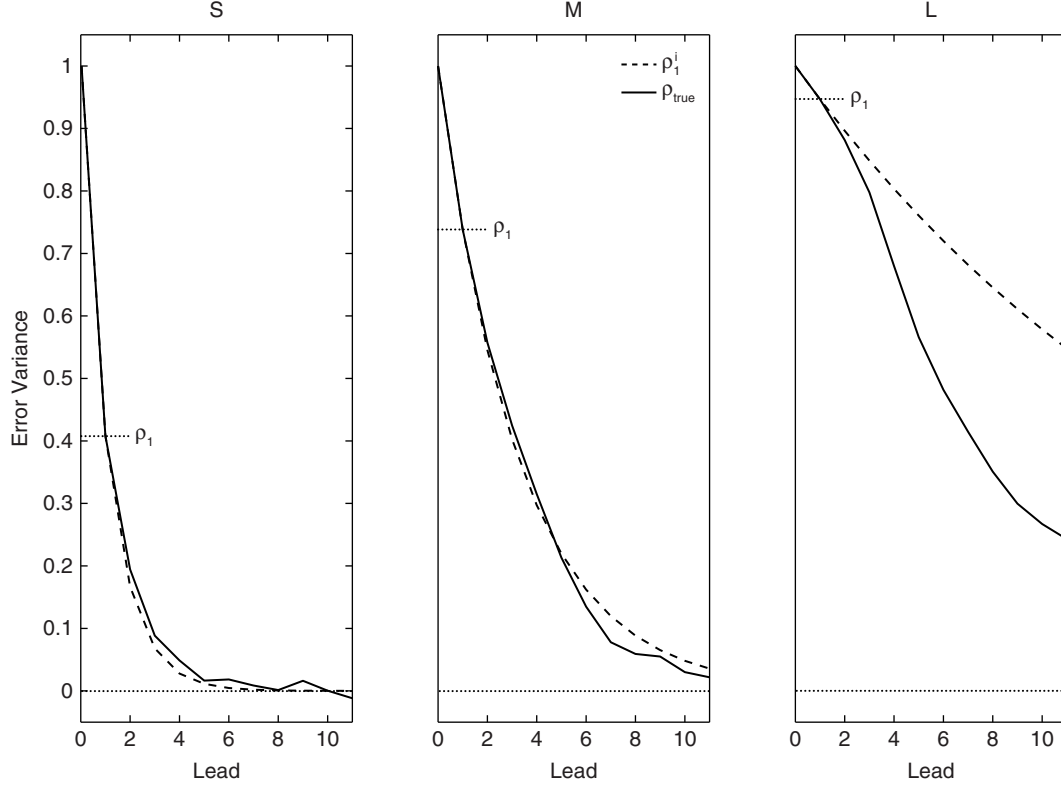


Fig. 2. Time correlation between analysis errors and forecast errors,  $\rho$ , as a function of lead time in the true system (continuous line) and in the empirical relationship [eq. (7); dashed line]. The correlation between the analysis errors and the first guess errors,  $\rho_1$ , is marked for each experiment.

modelled perceived error variances fall within the SEM interval, that is  $|d^2 - d_{\text{fit}}^2| < \text{SEM}$ . In Fig. 3b, the resulting forecast error (dashed) is close to the true forecast error at all the lead times considered. Likewise, Fig. 3c shows that the resulting estimated correlation of errors (dashed) accurately estimate the true correlation. Overall, the method applied to the observational perceived errors of experiment S produced estimated analysis error variances, forecast error variances and correlations that are close to the true values. In most of the lead times, the true forecast error variance falls within the bootstrapping interval.

The same method is applied to experiments M and L. Results from experiment M are not shown since the estimation results are as good as experiment S, except for the  $\rho_1$  parameter which is overestimated and is just outside the bootstrapping interval. Figure 4 shows the results for experiment L. In this case, the fit to the observed perceived errors is not acceptable. This is a clear indication that the exponential error growth curve is not appropriate and therefore the resulting analysis and forecast error estimates are not reliable. In this case, a more general error growth model should be used and thus a larger number of data points are required.

An important result from experiments L, M and S is that the quality of the fit to the observed values of the perceived error can be used to determine if the underlying forecast error growth (or other) assumptions are violated, in which case the true error estimates must be considered suspect. A statistically acceptable fit to the observed perceived error can therefore impart confidence in the true analysis and forecast errors estimated using the proposed algorithm.

#### 4. Application to real forecast systems

In this section, the method described in Section 2 and tested in the controlled environment in Section 3 will be used to estimate analysis and forecast errors from four NWP model. The model versions were in operation during the Fall of 2008 at the National Centers for Environmental Prediction (NCEP), the Canadian Meteorological Center (CMC), the European Center for Medium range Weather Forecast (ECMWF) and the Fleet Numerical Meteorology and Oceanography Center (FNMOC).

As the true error in this situation is unknown we will rely on the test of the fit to the observed perceived error to determine whether the true analysis and forecast error estimates are valid.



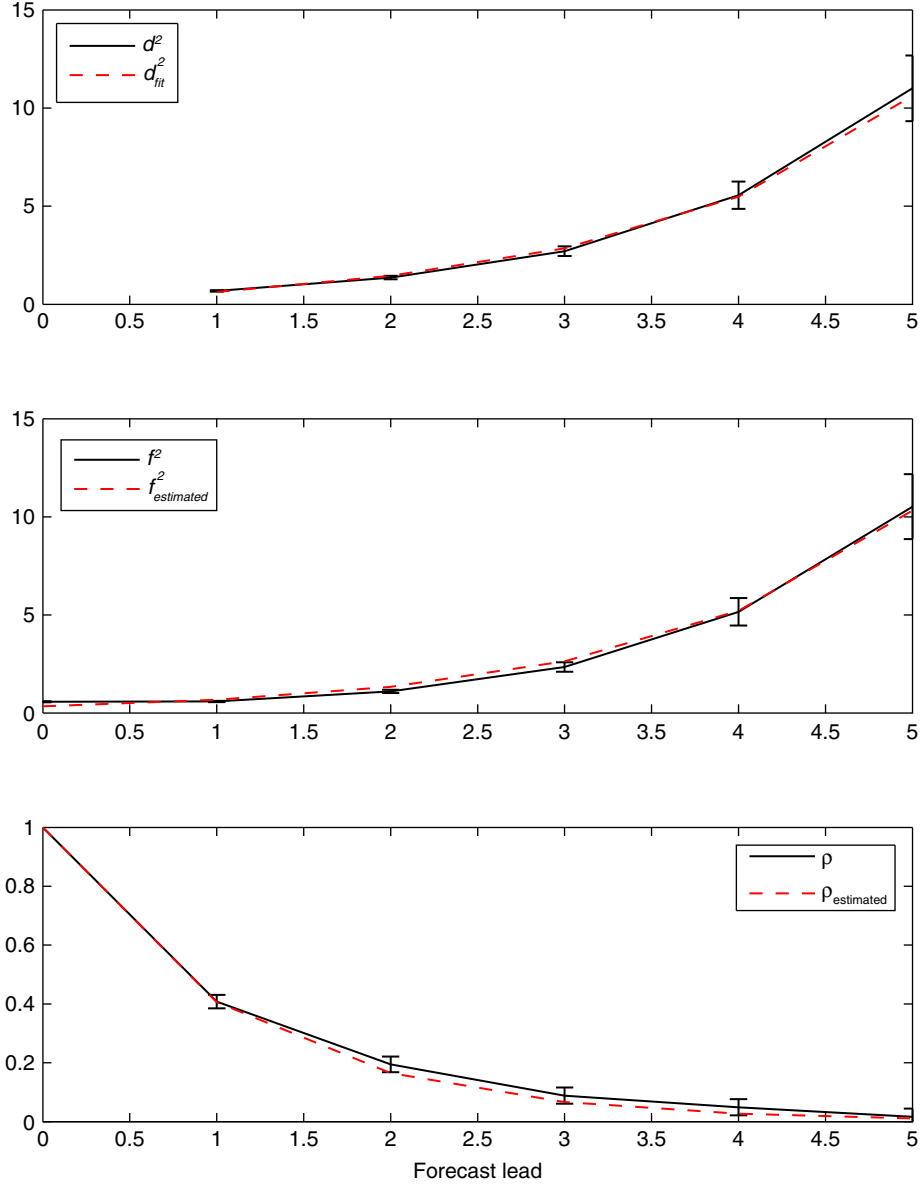


Fig. 3. Error variances and error correlation as a function of lead time for the S experiment using five lead points. Observed perceived error variance ( $d$ ) and modelled perceived error variance ( $d_{fit}$ ) (upper panel), forecast error variance ( $f^2$ ) and estimated forecast error variance ( $f_{estimated}^2$ ) (middle panel), true and estimated correlation (lower panel). Error bars in the upper panel are the SEM; for the other panels are the ranges of parameters such that  $|d^2 - d_{fit}^2| < \text{SEM}$ .

#### 4.1. Description of the data

The forecasts and the analysis verification data are regular grids of  $1 \times 1$  degrees of resolution in latitude and longitude. Two variables are analysed: 500 hPa geopotential height over the Northern Hemisphere (hereafter NH;  $30^\circ\text{N}$  to  $90^\circ\text{N}$ ) from the four available models, and the 10 m zonal component of the wind over the tropics (hereafter TR;  $30^\circ\text{S}$  to  $30^\circ\text{N}$ ) from the NCEP model only. The forecast perceived errors are computed as area-average error variances over the domains available. Forecast data consist

of 12 hours outputs of forecasts out to 16 d initialised at 0000 UTC daily from 1 Sep 2008 through 30 Nov 2008. Forecasts are verified against their own analysis.

#### 4.2. Negligible model error case: NH 500 hPa

In this section, the estimation method will be applied to the short-range forecast of NH 500 hPa height. Furthermore, the resulting error variances and parameters of each of the four models will be compared. Traditionally, comparative

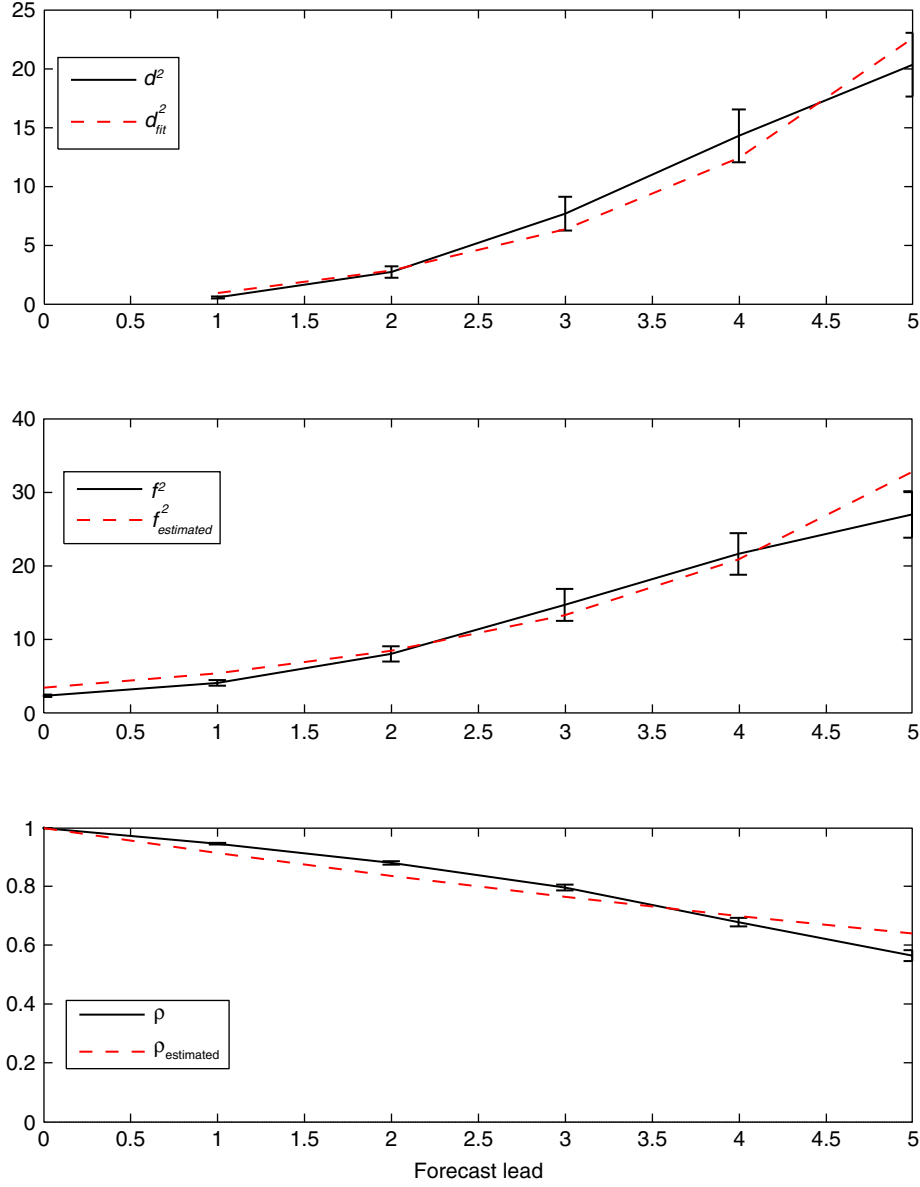


Fig. 4. Same as Fig. 3 but for the L experiment.

evaluations of different NWP systems are problematic as the choice of the analysis fields used influences the results and can unduly favour one or more of the prediction systems compared. The method introduced in this study mitigates this problem by using an estimated true state of the atmosphere as a common reference for all forecast systems verified and compared. Assuming that short-range forecasts of 500 hPa height have only negligible systematic errors and are not influenced significantly by non-linear processes here we will test the applicability of the exponential error growth model [eq. (3)].

We start the estimation process by computing both the observed perceived error variances and the SEM for

each model. Then, a modelled perceived error variance is obtained through optimisation of parameters to minimise eq. (9). Figure 5 displays the misfit (absolute value of the difference between the modelled perceived error variances and the observed perceived error variances) and the corresponding SEM as a function of lead time. Figure 5 indicates that in all cases and at all lead times the estimated perceived errors fits the observed perceived errors well within an acceptable range based on SEM. The small misfit confirms our hypothesis that exponential error growth [eq. (3)] is acceptable for this variable and the forecast lead times considered. The optimisation procedure determines the unknown parameters.

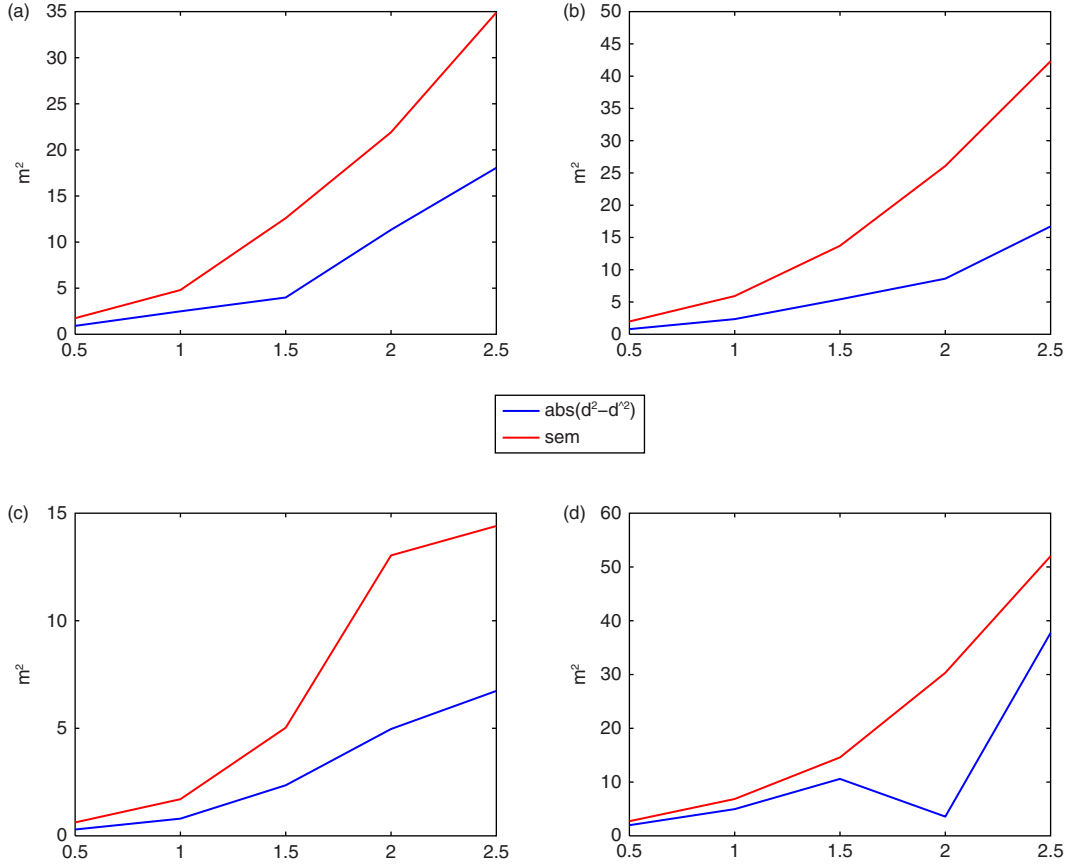


Fig. 5. Absolute value of difference between the observed perceived error variance and modelled perceived error variance, and the standard error of the mean (SEM) for (a) NCEP, (b) CMC, (c) ECMWF and (d) FNMOC. Abscissa axis corresponds to forecast lead time in units of days.

The resulting parameters are shown in Table 2. In the second column, the estimated analysis error variance indicates the ECMWF model has the smallest value, followed by CMC, NCEP and FNMOC. These results are qualitatively consistent with intercomparison results from earlier studies (e.g. Buizza et al., 2005). In the third column, the ECMWF forecasts exhibit the fastest error growth, followed by CMC, FNMOC and NCEP. Correspondingly, the error doubling times in the fourth column are shortest in the

Table 2. Parameters estimated for the NH 500 hPa inter-comparison of NWP systems: analysis error variance ( $x_0^2$ ), error growth ( $\alpha$ ), error doubling times (EDT), correlation between the analysis and first-guess errors ( $\rho_1$ ) and corresponding explained error variance (EEV).

MODEL	$x_0^2$ (m <sup>2</sup> )	$\alpha$ (day <sup>-1</sup> )	EDT (days)	$\rho_1$	EEV (%)
NCEP	38.0	0.25	2.7	0.56	31
CMC	29.5	0.27	2.6	0.47	22
ECMWF	11.5	0.30	2.3	0.22	5
FNMOC	49.2	0.26	2.8	0.60	36

ECMWF. We interpret the differences in the growth rate of short-range forecast errors in terms of the DA systems and the forecast models used at the various centres. Analyses produced by more advanced 4DVar schemes at ECMWF and CMC (in contrast to the 3DVar systems used at the time at FNMOC and NCEP) exhibit lower overall analysis error variance. The errors in these more advanced systems, however, may project more strongly onto the dynamically fast growing error space, leading to faster overall forecast error growth. As Pires et al. (1996) showed, the more advanced a DA system is, the lower its overall error variance is and the more those errors concentrate in the fastest growing subspace. This is because advanced DA techniques that exploit more information about the dynamics of the atmosphere (e.g. 4DVar) are very efficient in eliminating random (non-growing) errors, thus concentrating more of the smaller remaining error variance in the growing sub-space. Ensemble Kalman filter methods are designed to reduce the growing errors more than the non-growing errors.

The use of numerical models with stronger chaotic behaviour may also lead to faster error growth. Differences in model formulation may also contribute to differences in forecast error growth. Higher resolution models, for example, produce more small-scale variability that is not predictable beyond very short lead times, possibly contributing to faster error growth (Szunyogh and Toth, 2001). Additional experiments where, for example, DA and modelling systems are interchanged between the NWP centres may offer information as to the sources of differences in the rate of forecast error growth. The last two columns of Table 2 show the correlation between the analysis and first-guess errors ( $\rho_1$ ), and the corresponding explained error variance. The values of these parameters for the ECMWF model are considerably smaller than for the other models.

Figure 6 displays the error variances of the observed perceived, the estimated perceived and the estimated true forecast for each of the models. The forecast error variance extrapolated to lead 0 coincides with the analysis error variance. In the FNMOC, NCEP and CMC models, the perceived error variances are smaller than the estimate forecast errors at very short leads. This indicates that

verifications based only on the perceived error would underestimate both the analysis and the very short range forecast error variances. This is due to the effect of correlations between errors in short range forecasts and verifying analyses (cf. 2) that is overlooked by the traditional verification approaches. For the ECMWF, the estimated true error is consistent with the perceived error.

Figure 7 shows the correlation between short-range forecast and analysis errors as a function of lead time. As discussed previously, the correlation parameter is a good diagnostic for the quality of the DA and NWP systems. High correlation in some of the models coincides with underestimation of analysis and short-range forecasts.

#### 4.3. Large model errors: TR 10 mU

In this section, we apply the method to the 10 m zonal wind component (hereafter 10 mU) in the tropics. This variable, as most others in the tropics, may be strongly affected by model errors. Therefore, we will use the general error growth model [eq. (6)] that accounts for both initial value and model-related analysis and forecast errors. Due to the

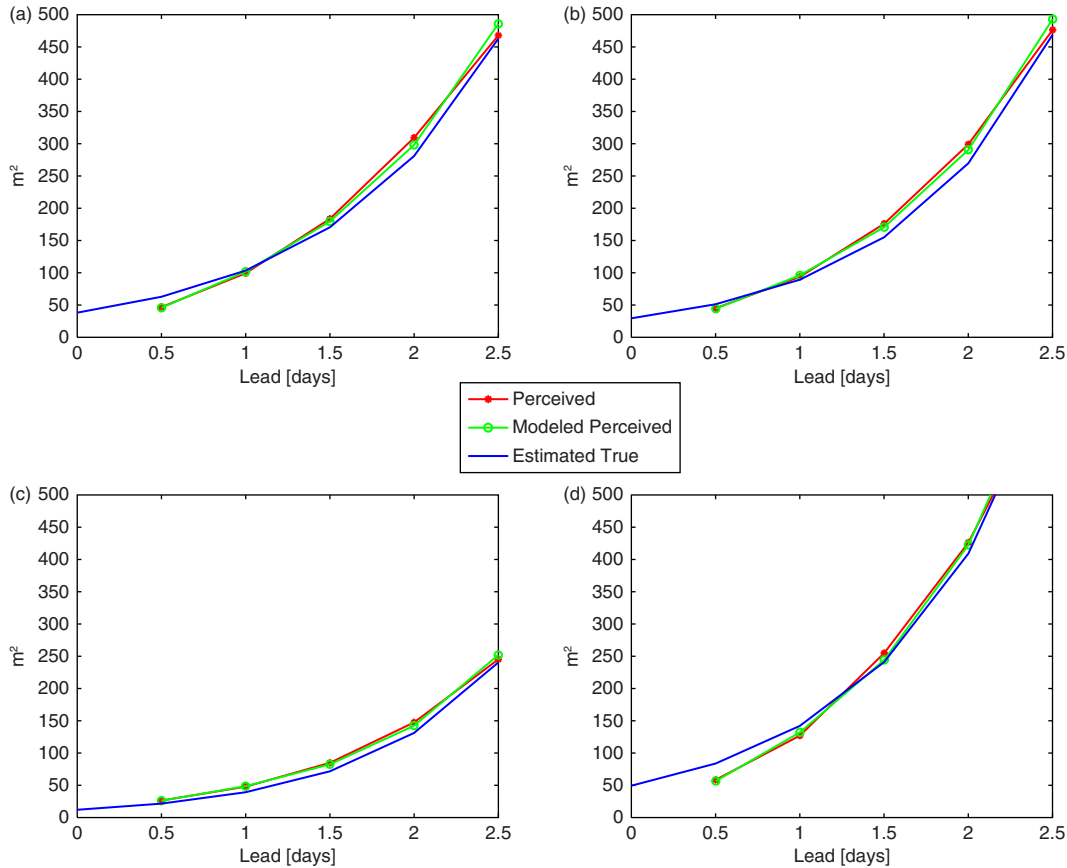


Fig. 6. Forecast (observed perceived, fitted perceives and true) error variances as a function of lead time for four global models: (a) NCEP, (b) CMC, (c) ECMWF and (d) FNMOC for models operational in 2008.

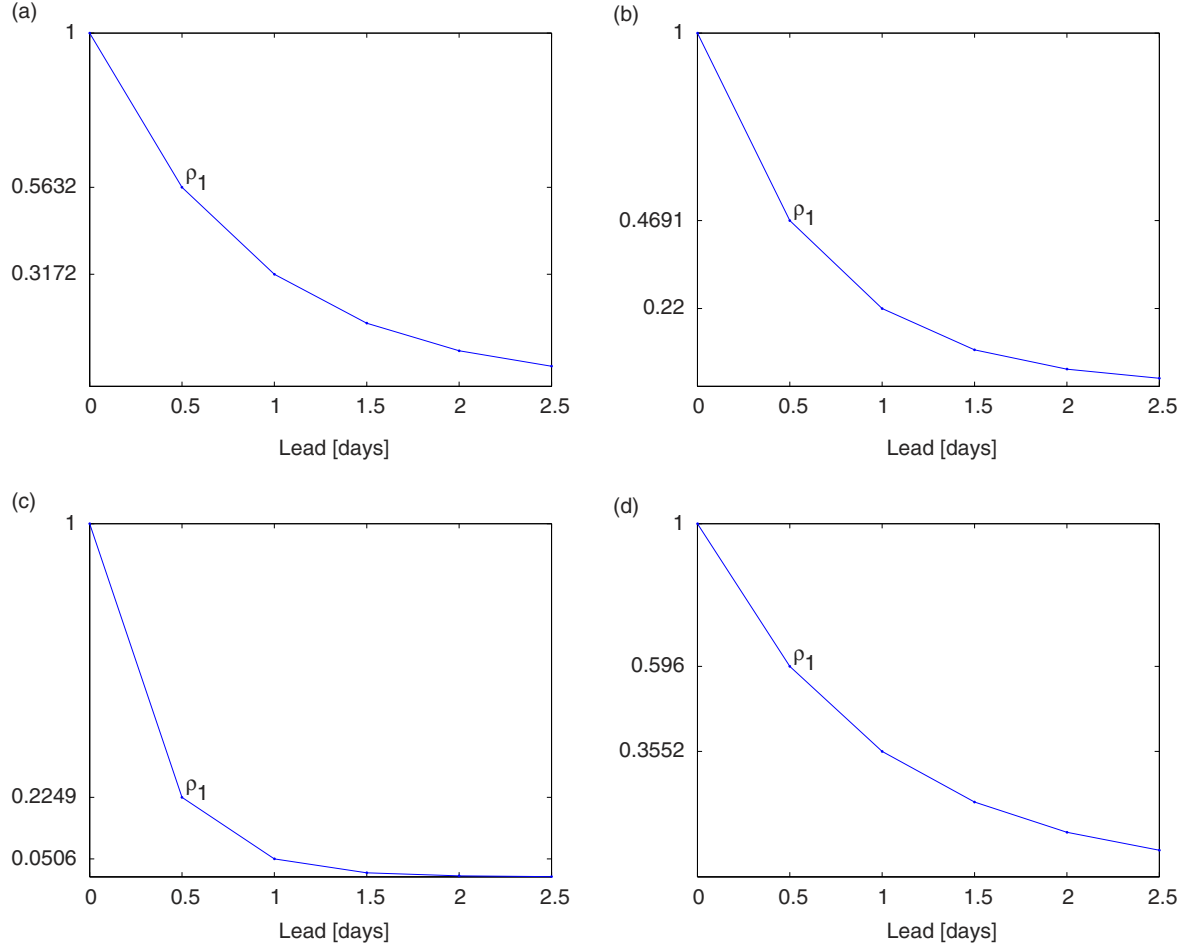


Fig. 7. Resulting correlation of forecast errors and verifying analysis errors for (a) NCEP, (b) CMC, (c) ECMWF, (d) FNMOC.

larger number of parameters involved in eq. (6) compared to the simpler exponential model in eq. (3), we will use the full 16 d forecast data points available. The Nelder-Mead Simplex minimisation algorithm in this case becomes more sensitive to the choice of the first guess value. To prevent the solution from getting trapped in a local minimum and to accelerate the optimisation process, we proceed in two steps. In the first step, we fit the general error growth model to perceived errors only in the 1 to 6 d lead time range. These initial parameter estimates are then used as first guesses in the second step of the minimisation process where perceived errors from all available (1 to 16 d) lead times are used. The purpose of the first minimisation step is to ensure reliable estimates for the model error component, which may dominate the short lead errors, whereas the second minimisation is aimed at refining especially the estimates of the saturation values for both components of the general error growth model. Figure 8c indicates that the resulting modelled perceived error variance is acceptably close to the observed perceived error variance at all lead times compared to the reference SEM.

In Fig. 8a, we note that both perceived and forecast error variances for 10 mU, especially at shorter lead times, have a concave increase with lead time. This suggests that at short leads errors are dominated by model drift. This is consistent with the study of Reynolds et al. (1994), which gives estimates of error variances much larger in the tropical than in the extratropical variables. Unlike initial-value-related errors that are dominated by leading Lyapunov vector type patterns (Szunyogh et al., 1997; Toth and Kalnay, 1997), model-related errors, as noted in Section 2.2.3, may be dominated by decaying, short-lived trailing singular vectors. We interpret the very low values of correlation between analysis and short range forecast errors in Fig. 8b as indications that model drift-related errors can be rapidly changing in the phase space and therefore depend strongly on the specific way the initial condition (analysis) is off the corresponding model trajectories. Overall, the results in Fig. 8 indicate that the general error growth model [eq. (6)] is able to describe complicated error behaviour and can provide insight into the initial value and model-related errors.

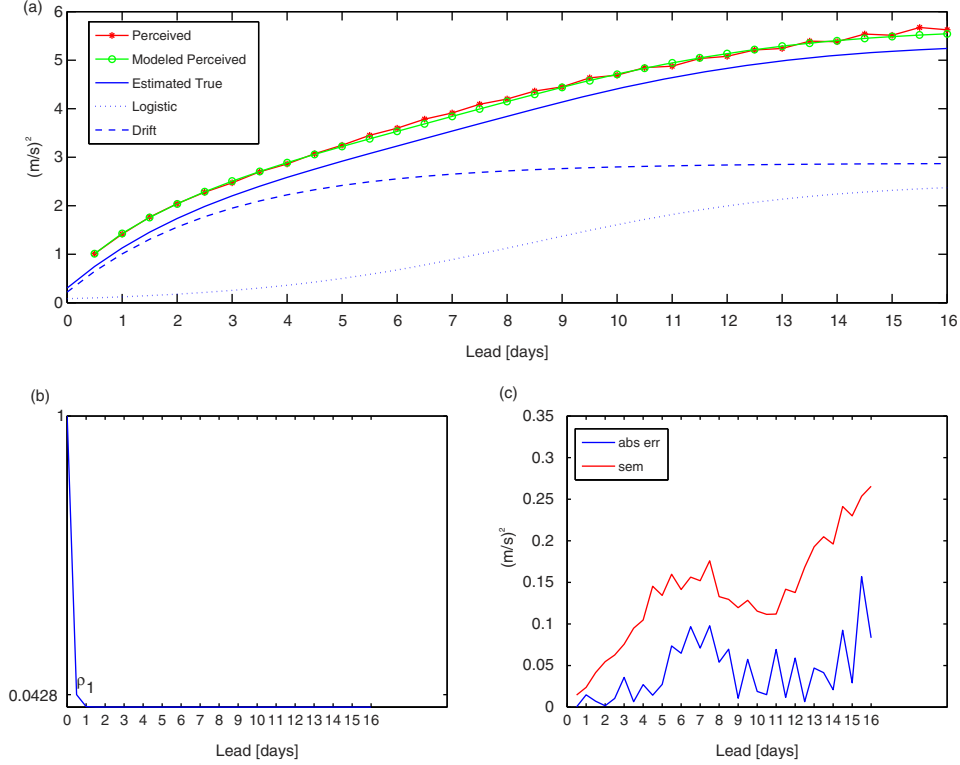


Fig. 8. (a) Seven-parameter model function applied to TR 10 mU wind forecasts from the NCEP model; (b) resulting analysis-forecast correlation; and (c) misfit compared to SEM (as in Fig. 5).

## 5. Conclusions and discussion

Accurate estimation of analysis and forecast errors are critical for a variety of applications, including the correct evaluation of forecasting systems, the tuning of DA systems and the proper initialisation of ensemble forecasts. The method introduced here relies on a small number of assumptions that are independent of assumptions used in DA schemes. The method recognises that perceived forecast error variances can be separated into true analysis errors (analysis minus truth on the model grid) and true forecast errors plus a term involving the correlation of the two errors. Since the true analysis and forecast errors are unknown, the method formulates the estimation of errors as an inverse problem in which the perceived errors are the ‘observations’ from where the rest of the unknowns are inferred. The method uses prior knowledge regarding the temporal evolution of forecast error variances to limit the number of possible function solutions. Those functions,  $x_i$  [eqs. (3–6)] are the error growth functions that have been established independently. To further reduce the number of unknowns, an assumption for the evolution of the correlation between analysis and forecast errors is introduced to permit solutions with a limited sample size.

With these two assumptions, a set of eq. (8) is written where the perceived error is modelled as a function of the unknown variables, which are then determined by a procedure that minimises the difference between the observed and modelled perceived error values. When the simplest error growth model [eq. (3)] is used, three basic parameters result from the minimisation procedure: the analysis error ( $x_0$ ), the growth rate ( $\alpha$ ) and the temporal correlation coefficient between the analysis error and the first guess error, valid at the same time ( $\rho_1$ ). The three parameters interrelate through eq. (3) and eq. (8). The method was evaluated in a simple model environment where the truth is known. Results from experiments simulating NWP systems with varying degrees of DA efficiency (i.e. with different levels of analysis errors) indicate that the method can (1) determine whether the error growth model applied is consistent with the observed perceived error data, and if so, (2) provide estimates of the true analysis and forecast errors within bounds expected from sampling errors.

### 5.1. Fair intercomparison of various forecast systems

The comparative evaluation of forecasts from different NWP systems against analysis fields has been problematic (Simmons, 1999). First, the use of a common analysis

as a proxy for truth would be preferable; however, such a choice may offer an unfair advantage to one or more NWP systems whose initial conditions may be similar to the analysis of choice. And second, the other often used approach, verifying forecasts from each system against their own analysis is also questionable as it lacks a common reference as a proxy for truth.

These two issues are mitigated as the current method evaluates analyses and forecasts from each system independently of the others (no unfair advantage, first issue), and it implicitly uses the true state of the atmosphere on the resolved scales as a common reference or proxy for truth (proper common reference, second issue). This allows, for the first time, for the unbiased estimation of true analysis and forecast errors. The intercomparison of 500 hPa height forecasts from four leading NWP centres indicates that the scheme with the smallest analysis error (ECMWF) also has the lowest correlation between analysis and forecast errors. This relationship indicates that the most efficient DA schemes (lowest analysis error variance) can extract the most information from the observations that is independent of the first guess (lowest correlation between analysis and forecast errors).

Verifying against independent observations can be a good choice as well. However, as described in the introduction it is limited in space (spotty), in time and only for observable variables.

### 5.2. *Disentangling initial value and model-related errors*

While the temporal evolution of initial-value-related error variances in chaotic systems has been studied extensively, the behaviour of model-related forecast error variances and their interaction with initial-value-related error variances is less understood. The partitioning of initial value and model-related forecast error variances is especially challenging as manifested by the lack of applicable methodologies.

The proposed relationships regarding the temporal evolution of model drift-related forecast error variances [eq. (5)], and the lack of interaction between initial value and model-related error variances [eq. (6)] were tested using a variable thought to be strongly affected by model-related errors, the 10 m tropical wind. The results show that the assumed relationships are consistent with the observed perceived error values. The proposed method, for the first time offers a decomposition of initial value and model drift-related forecast error variances, suggesting that total forecast error variance in the tropics is dominated by model-related errors during the first week of the forecast.

### 5.3. *Potential extensions and applications*

Several important simplifications were made in this work that will require further analysis. For example, we applied the approach to two variables, NH 500 hPa and TR10 mU, whose behaviour is well known. It remains to be seen whether the method can be applied to more complex variables such as precipitation. The model drift error was assumed to behave as a saturating exponential function but it may behave differently. As noted earlier, the Simplex (Lagarias et al., 1998) algorithm used in this study can identify solutions close to the absolute minimum if presented with a reasonable first guess. We found this algorithm to be too sensitive to the first guess and required many iterations to arrive to what we considered the absolute minimum. To improve efficiency and accuracy of the estimates, in future studies we will test alternative minimisation algorithms.

All NWP error estimation experiments in this study used area average perceived errors (for the NH extratropics or the tropics). Sampling errors in area mean perceived error variances is much reduced compared to error variances for individual gridpoints. In future studies we will explore if the sample size typically available permits the estimation of error variances for smaller areas or individual gridpoints. Point-wise estimates of the true analysis and short range forecast error variances may be used to specify initial spread in ensemble forecasting and background error variances in DA schemes, respectively.

The scope of this work was also limited by the use of determinist forecasts. Ensemble forecasts offer additional information particularly the ensemble spread that could be used to assess the error growth more generally. While the ensemble spread is a useful piece of information to estimate the error growth rate parameter, we must make a distinction between error growth rate in the subspace generated by the ensemble members and the error growth rate between the truth and the model. The results presented in this paper are ‘static’ and not flow dependent. Despite this limitation, it might be possible to apply the error estimates to generate what is called ‘masks’ to constrain the amplitude of the perturbations in certain regions.

## 6. Acknowledgements

Yuejian Zhu provided the hemispheric analysis and forecast data used in this study. Discussions with Yuejian Zhu, Mozheng Wei, Dingchen Hou and Roman Krzysztofowicz are gratefully acknowledged. We acknowledge Eugenia Kalnay and two other anonymous reviewers who provided valuable comments that improved the original manuscript.

## References

- Bence, J. R. 1995. Analysis of short time series: correcting for autocorrelation. *Ecology*. **76**(2), 628–639.
- Bowler, N. E. 2006. Explicitly accounting for observation error in categorical verification of forecasts. *Mon. Weather Rev.* **134**, 1600–1606.
- Bowler, N. E. 2008. Accounting for the effect of observation errors on verification of MOGREPS. *Meteorol. Appl.* **15**, 199–205.
- Buizza, R., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Weather Rev.* **133**, 1076–1097.
- Candille, G. and Talagrand, O. 2008. Impact of observational error on the validation of ensemble prediction systems. *Q. J. Roy. Meteorol. Soc.* **134**, 959–971.
- Ciach, G. J. and Krajewski, W. F. 1999. On the estimation of radar rainfall error variance. *Adv. Water Resour.* **22**, 585–595.
- Dalcher, A. and Kalnay, E. 1987. Error growth and predictability in operational ECWF forecasts. *Tellus A*. **39**, 474–491.
- Daley, R. and Mayer, T. 1986. Estimates of global analysis error from the global weather experiment observational network. *Mon. Weather Rev.* **114**, 1642–1653.
- Desroziers, G., Berre, L., Chapnik, B. and Poli, P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. Roy. Meteorol. Soc.* **131**, 3385–3396.
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlos methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143–10162.
- Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**, 343–367.
- Fisher, M. 1996. The specification of background error variances in the ECMWF variational analysis system. In: *Proceedings ECMWF Seminar on Data Assimilation*. Reading, UK, Available from ECMWF, pp. 645–652.
- Fisher, M. and Courtier, P. 1995. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo 220. Available from ECMWF.
- Houtekamer, P. L., Lefaiivre, L. and Derome, J. 1995. The RPN ensemble prediction system. In: *Seminar Proceedings on Predictability*, Vol. II, Reading, UK, European Centre for Medium-Range Weather Forecasts, pp. 121–146.
- Kalnay, E. 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, New York, 340 pp.
- Lagarias, J. L., Reeds, J. A., Wrights, M. H. and Wright, P. E. 1998. Convergence properties of the Nelder-Mead Simplex Method in Low dimensions. *SIAM J. Optim.* **9**, 112–147.
- Leith, C. E. 1978. Objective methods for weather prediction. *Ann. Rev. Fluid. Mech.* **10**, 107–128.
- Li, H., Kalnay, E. and Miyoshi, T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. Roy. Meteorol. Soc.* **135**, 523–533.
- Lorenc, A. C. and Swinbank, R. 1984. On the accuracy of general circulation statistics calculated from FGGE data – a comparison of results from two sets of analyses. *Q. J. Roy. Meteorol. Soc.* **110**, 915–942.
- Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.
- Lorenz, E. N. 1969a. The predictability of flow which possesses many scales of motion. *Tellus*. **21**, 289–307.
- Lorenz, E. N. 1969b. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636–646.
- Lorenz, E. N. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus*. **34**, 505–513.
- Parrish, D. F. and Derber, J. 1992. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.* **120**, 1747–1763.
- Pires, C., Vautard, R. and Talagrand, O. 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A*. **48**, 96–121.
- Reynolds, C., Webster, P. and Kalnay, E. 1994. Random error growth in NMC's global forecasts. *Mon. Weather Rev.* **122**, 1281–1305.
- Riishojgaard, L. P. 2000. A method for estimating the analysis error variance in a physical space statistical data assimilation system. *Q. J. Roy. Meteorol. Soc.* **126**, 1367–1386.
- Simmons, A. J. 1999. Objective verification of deterministic forecasts. In: *Proceedings of Diagnosis of models and data assimilation systems*. ECMWF, Reading, UK, pp. 385–404.
- Simmons, A. J. and Hollingsworth, A. 2002. Some aspects of the improvement in skill of numerical weather prediction. *Q. J. Roy. Meteorol. Soc.* **128**, 647–677.
- Szunyogh, I. and Toth, Z. 2001. The effect of increased horizontal resolution on the NCEP global ensemble mean forecasts. *Mon. Weather Rev.* **130**, 1125–1143.
- Szunyogh, I., Kalnay, E. and Toth, Z. 1997. A comparison of Lyapunov and optimal vectors in a low-resolution GCM. *Tellus A*. **49**, 200–227.
- Talagrand, O. 1997. Assimilation of observations; an introduction. *J. Meteorol. Soc. Japan*. **75**, 191–209.
- Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330.
- Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **125**, 3297–3319.
- Toth, Z. and Peña, M. 2007. Data assimilation and numerical forecasting with imperfect models: the mapping paradigm. *Physica D*. **230**, 146–158.
- Trevisan, A., Malguzzi, P. and Fantini, M. 1992. On Lorenz's law for the growth of large and small errors in the atmosphere. *J. Atmos. Sci.* **49**, 713–719.
- Wei, M., Toth, Z. and Zhu, Y. 2010. Analysis differences and error variance estimates from multi-centre analysis data. *Aust. Meteorol. Ocean J.* **59**, 25–34.