

RESEARCH ARTICLE

Partition of analysis and forecast error variance into growing and decaying components

Jie Feng^{1,2}  | Zoltan Toth¹ | Malaquias Peña³ | Jing Zhang¹¹Global Systems Division,
ESRL/OAR/NOAA, Boulder, Colorado²School of Meteorology, University of
Oklahoma, Norman, Oklahoma³Department of Civil and Environmental
Engineering, University of Connecticut,
Storrs, Connecticut**Correspondence**J. Feng, School of Meteorology, University
of Oklahoma, 120 David Boren Blvd.
Norman, OK 73072, USA.
Email: jie.feng@ou.edu**Funding information**Global Systems Division,
ESRL/OAR/NOAA, Boulder, CO, USA**Abstract**

Due to the scarcity of and errors in observations, direct measurements of errors in numerical weather prediction (NWP) analyses and forecasts with respect to nature (i.e. “true” error) are lacking. Peña and Toth (2014) introduced an inverse method called SAFE-I where true errors are (a) theoretically assumed to follow exponential error growth, and (b) estimated from the perceived errors (i.e. forecast minus verifying analysis) that they affect. While decaying or neutral errors, by definition will not have a significant impact on longer-range forecast errors, they can still accumulate in, and negatively influence NWP data assimilation–forecast cycles. In a new, generalized version of the inverse method (SAFE-II), analysis and forecast error variance is decomposed into exponentially growing and decaying components, assuming they are independent as they comprise vectors from the leading and trailing ends of the Lyapunov spectrum, respectively. SAFE-II uses the initial variance and decay rate associated with non-growing perturbations to describe and estimate their behaviour. The assumptions behind SAFE-II are first validated in a simulated environment. SAFE-II is then applied to estimate the error variance in both simulated and operational analyses/forecast environments. Perceived error measurements are found to be statistically consistent (at the 95% significance level) with the SAFE-II error behaviour model, which offers a more accurate description of error variance than SAFE-I that neglects decaying errors. At various levels and for different variables, decaying errors are found to constitute up to 60% of the total analysis error variance, much of which decays during the first 12–18 hr of forecast integrations.

KEYWORDS

data assimilation, ensemble forecasts, error estimation, forecast verification, uncertainty of analysis

1 | INTRODUCTION

Due to the intermittency of, and errors in, available observations, the true state of the atmosphere, however alluring

it is, remains unknown. The state of the atmosphere is estimated using data assimilation (DA, current state, or analysis) and numerical weather prediction (NWP, future states or forecasts) tools. Both the assessment and

improvement¹ of the quality of DA and NWP tools and products depend on reliable estimates of analysis and forecast error variance. In most studies, such errors are estimated with the variance between NWP analysis or forecast states that are being evaluated and verifying observations or NWP analysis fields (in the case of forecast verification). Since errors in some verifying observations or analysis fields are of comparable magnitude to those in analysis or short-range forecast fields that are being evaluated, such an approach is convoluted and yields questionable results.

Peña and Toth (2014, PT14) introduced a method hereafter called Statistical Analysis and Forecast Error (SAFE-I) estimation that relates the measured *perceived forecast error variance* (forecast minus verifying analysis) to *true error variance* (forecast minus reality interpreted on the model grid). SAFE-I is independent of any assumptions used in analysis or forecast systems. The measured perceived forecast error variance is modelled by several unknowns. To reduce the number of unknowns in the statistical estimation process, it uses prior knowledge about the evolution of errors in analysis–forecast systems. The unknown parameters are estimated via the minimization of the difference between the sample mean (e.g. over a season) of measured and modelled (via the unknown parameters) perceived error variance. Feng *et al.* (2017) extended the application of SAFE-I from area mean to pointwise error estimation and quantified the spatial distribution of analysis and short-range forecast error variance at a 95% confidence level.

For simplicity, SAFE-I assumes that in short-range (i.e. out to 2 or 3 days) synoptic-scale forecasts all analysis errors grow at a close to exponential rate. Analysis errors are therefore assessed in a “growing equivalent” sense. The effect of non-growing analysis errors, if any, will implicitly manifest in modified estimates of the growing error component. In the presence of a significant level of decaying analysis errors, this may lead to an overestimation of initial growing error variance, and an underestimation of the growth rate.

Analysis fields are a weighted sum of observations and NWP first-guess forecast fields. It is generally accepted that NWP analyses contain both random or decaying, and dynamically conditioned, growing errors (Toth and Kalnay, 1993; 1997; Buizza *et al.*, 2005; Houtekamer *et al.*, 2005; Wei *et al.*, 2008; Peña *et al.*, 2010). The former generally signifies a lack of dynamical balance in analysis

fields. These errors are believed to originate from errors in observations (e.g. Hunt *et al.*, 2007; Stewart *et al.*, 2013), or statistical DA approximations,² hence can be considered random from a model dynamics point of view. Therefore, these errors project onto the stable (or decaying) manifold of the system (Toth and Kalnay, 1997; Kalnay, 2003). Growing errors originate from amplifying errors in first-guess forecasts, projecting onto the unstable (or growing) subspace (Pires *et al.*, 1996; Toth and Kalnay, 1997; Kalnay, 2003; Trevisan and Uboldi, 2004; Feng *et al.*, 2018). As Pires *et al.* (1996) showed, improved DA techniques lead to a reduction of the proportion of errors that decay in the overall analysis error.

When decaying errors are present in the analysis, over a transient period the overall error may either decay or exhibit slower than exponential growth due to the rapid collapse of random errors (e.g. Vannitsem and Nicolis, 1994; Trevisan and Legnani, 1995; Houtekamer *et al.*, 2005; Palatella *et al.*, 2013). Such a transient period is followed by exponential error growth,³ characteristic of the system's dynamics associated with the leading local Lyapunov vectors (Toth and Kalnay, 1997; Kalnay, 2003; Snyder and Hamill, 2003; Ding and Li, 2007; Li and Ding, 2011; Feng *et al.*, 2014).

Forecast errors also display a transitional decaying phase in Observing System Simulation Experiments (OSSEs) where true error is directly measurable (see, e.g., Privé and Errico, 2013). When initial perturbations are dynamically less conditioned (i.e. have significant projection on the stable manifold due to, e.g., the addition of simulated observational noise), the ensemble spread may also exhibit transitional behaviour (e.g. Houtekamer *et al.*, 2005; Hamill and Whitaker, 2011).

Decaying components of analysis error or perturbation variance rapidly disappear during the initial phase of forecast integrations (typically in less than a day). But their accurate estimation can (a) improve the accuracy of the analysis and short-range forecast error variance estimation, (b) diagnose the effectiveness of DA schemes (in the spirit of Pires *et al.*, 1996), and (c) provide guidance as to the appropriate level of growing, dynamically conditioned perturbations (as opposed to quickly disappearing noise) in initial ensemble perturbation generation methods. In

¹For example, the reliable specification of analysis error variances offers a reference for the rescaling of initial ensemble perturbations (Molteni *et al.*, 1996; Toth and Kalnay, 1997; Wei *et al.*, 2008). Also, the accurate quantification of short-range forecast error variances can orient the tuning of background forecast error covariance in DA (Fisher, 1996; Whitaker *et al.*, 2008).

²Examples include the use of “covariance localization” in ensemble Kalman filters (EnKF) for the reduction of spurious long-distance covariances (Houtekamer and Mitchell, 2001). Such schemes may introduce an imbalance among different variables (Mitchell *et al.*, 2002; Lorenc, 2003). The use of incomplete balance constraints may also leave gravity waves in the analysis that appear as noise to hydrostatic models (Huang and Lynch, 1993; Kleist *et al.*, 2009).

³In nonlinear systems, as the level of error becomes comparable to the size of the attractor, nonlinear interactions moderate exponential growth (Lorenz, 1982; Dalcher *et al.*, 1988).

particular, the diagnosis of decaying errors is a prerequisite for their reduction and for making analysis fields dynamically more balanced.

This study is based on the recognition that analysis errors generally project onto the full spectrum of local Lyapunov vectors (LLVs: Wolf *et al.*, 1985; Legras and Vautard, 1996), from the fastest growing to the fastest decaying directions (Toth and Kalnay, 1997; Vannitsem and Nicolis, 1997; Hamill *et al.*, 2002; Kalnay, 2003; Ding *et al.*, 2017; Feng *et al.*, 2018). This is because each analysis step introduces some noise into the analysis field, randomly projecting onto the full spectrum of directions in the phase space. The forecast step amplifies dynamically growing error patterns while dissipating errors in other directions, thus rotating the overall error toward the growing subspace. Such potentially complex error behaviour is approximated here by assuming that the total error variance is the sum of two orthogonal error components (SAFE-II). The first component is exponentially growing, characterized by the leading Lyapunov vector (LV: Lorenz, 1996; Toth and Kalnay, 1997; Ziehmann *et al.*, 2000; Kalnay, 2003; Feng *et al.*, 2014), estimated by SAFE-I, while the other component introduced here is exponentially decaying, considered as a composite of errors across all the neutral and trailing LVs.

The modelling of the decaying errors in SAFE-II is introduced in Section 2. Section 3 describes the Global Forecast System (GFS) that is used operationally at the National Centers for Environmental Prediction (NCEP), in which SAFE-II will be tested. The SAFE-II assumptions are validated in a GFS-based OSSE environment (Cucurull *et al.*, 2017) where “ground truth” is known exactly (Section 4). Experimental SAFE-II results from both simulated and operational systems, including a comparison with SAFE-I output, are presented and analysed in Section 5, followed by preliminary conclusions in Section 6 and discussion in Section 7.

2 | METHODOLOGY

2.1 | Statistical Analysis and Forecast Error estimation algorithm (SAFE-I)

Let \mathbf{F} , \mathbf{A} and \mathbf{T} denote the forecast, analysis, and true state of reality, all valid at the same time and interpolated onto a common model grid. The true (\mathbf{x}_i) and perceived (\mathbf{f}_i) errors in an $i \cdot \Delta t$ lead time forecast (where Δt is the length of the DA cycle) are then defined as:

$$\mathbf{x}_i = \mathbf{F}_i - \mathbf{T}_i, \quad (1)$$

$$\mathbf{f}_i = \mathbf{F}_i - \mathbf{A}_i. \quad (2)$$

Since the true state of reality is not known exactly, the true error is not measurable. For each lead time, PT14 introduces the following relationship between the true analysis and forecast error variances and the perceived forecast error variance measurements:

$$f_i^2 = x_0^2 + x_i^2 - 2\rho_i \cdot x_0 \cdot x_i, \quad (3)$$

where f_i^2 , x_0^2 and x_i^2 are the spatial and temporal mean of error variance corresponding with \mathbf{f}_i , \mathbf{x}_0 and \mathbf{x}_i , and ρ_i is the sample mean correlation between \mathbf{x}_0 and \mathbf{x}_i . The unknown parameters are estimated by minimizing the difference between the measured (f_i^2) and the modelled (\hat{f}_i^2) perceived error variance in Equation like (3).

Note that the number of unknowns in a series of Equation like (3) exceeds the number of measured quantities. Here we follow PT14 and use a simplifying set-up as well as prior knowledge about error growth and DA (in the form of several assumptions, see Table 1) to dramatically reduce the number of unknowns in a series of Equation like (3).

Simplifying set-up. As in SAFE-I, forecasts are verified against analysis fields from the same DA-forecast system that is used for the initialization of the forecasts:

$$\mathbf{F}_0 = \mathbf{A}_0. \quad (4)$$

We opt to use analysis fields instead of observations as a proxy for reality as, by design, they have a lower error. Choosing verifying analysis fields from the same system that initializes the forecasts reduces the number of unknowns, potentially reducing errors in their statistical estimation.

Assumption 1. Model error. In this study, we focus on extratropical forecast variables verified against analysis fields that represent natural processes at the model's spatio-temporal resolution. For simplicity, under these conditions we assume that model error is negligible. In case total forecast error can be explained purely through the amplification of initial errors, the assumption will be considered validated. For other (e.g. tropical) variables or for processes not well resolved by the model (e.g. parameter or truncation errors), the model error can be explicitly represented as an additional term in Equation 5 below (see, e.g., PT14, and Vannitsem and Toth (2002), or Nicolis *et al.* (2009), respectively).

Assumption 2. Error evolution. Error variance in short-range forecasts of complex systems evolve exponentially and therefore can be described simply by two unknown parameters – the initial analysis error size x_0 , and the exponential growth rate α (Lorenz, 1963):

TABLE 1 Summary of Assumptions 1–6 behind the SAFE-II method related to the use of perceived error variance (PEV) and variance of lagged forecast difference (VLFD) measurements

| Subject | Estimation area | Assumption | Section introduced/ validated |
|---|-----------------|--|-------------------------------|
| 1. Model error | PEV | Negligible for studied variables | 2.1/4.1 |
| 2. Error evolution | PEV | Exponential growth/decay of initial error variance | 2.1,2.2/4.1 |
| 3. Data impact on analysis | PEV | Power law decorrelation of analysis error from increasing lead time forecast error | 2.1/4.1 |
| 4. Relationship between true and perceived error variance | VLFD | True and perceived error variances become similar with longer lead times | 2.3/4.2 |
| 5. Transient period | VLFD | Decaying errors diminish in first 24 hr of integration | 2.3/4.1 |
| 6. Divergence rate of model trajectories | VLFD | Divergence rate is similar between lagged forecasts vs. forecast and truth | 2.3/4.1,4.2 |

$$x_i^2 = x_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha}. \quad (5)$$

If necessary, Equation 5 can be augmented to represent the effect of nonlinear saturation (i.e. replace the exponential relationship with the logistic function), or model-related errors (PT14).

Assumption 3. Data impact on analysis. PT14 recognized that the repeated insertion of new observational information in successive DA-forecast cycles results in the progressive decorrelation of true error in a freely evolving forecast from the true error in verifying analyses valid at the same time. Assuming that a statistically similar amount of observational information is ingested in each DA cycle, the error decorrelation follows a power-law relationship:

$$\rho_i = \rho_1^i. \quad (6)$$

ρ_1 and ρ_i in Equation 6 indicate the angular extent to which the error in the latest analysis is rotated from the error in the first guess ($\Delta t = 6$ hr for a typical DA cycle in global forecast systems) or from earlier initialized longer-range forecasts all valid at the time of the analysis, respectively, due to one (or multiple) introduction(s) of observational information. The simplicity of the data impact relationship in Equation 6 is because the error that is assumed to be composed of leading LLVs, whether present in a DA-forecast cycle or in a “free” longer-range forecast, develops similarly in a quasi-linear fashion, over the same (or very similar) time-evolving flow.

With the relationships in Equations 4–6, the number of unknowns is significantly reduced, and the short-range

perceived error variance can be simulated with only three unknowns (x_0 , α , ρ_1):

$$\hat{f}_i^2 = x_0^2 + x_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha} - 2\rho_1^i \cdot x_0^2 \cdot \sqrt{e^{i \cdot \Delta t \cdot \alpha}}. \quad (7)$$

The unknown parameters are then estimated by minimizing the cost function:

$$J = \max(|f_i^2 - \hat{f}_i^2| \cdot w_i^{-1}), \quad (8)$$

where w_i^{-1} is the weight on the fitted perceived error variance at lead time $i \cdot \Delta t$, and $|\bullet|$ represents the absolute value. The minimization is carried out using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm (Byrd *et al.*, 1995). The choice of the L_∞ norm ($\max(\bullet)$; i.e. infinite norm) is motivated by a desire to get a good fit over the entire range of lead times (PT14). Simulated perceived error values on the right side of Equation 7 are expected to match their measurement counterparts only within the sampling uncertainty of the latter which is given by the standard error of the mean (SEM). For further details on SEM and w_i^{-1} , see Appendix A.

2.2 | Decomposition of analysis and forecast errors

As mentioned earlier, analysis errors are generally assumed to project with varying power on the full spectrum of LLVs, from the fastest growing to fastest decaying vectors (Toth and Kalnay, 1997; Vannitsem and Nicolis, 1997; Hamill *et al.*, 2002; Kalnay, 2003; Feng *et al.*, 2018).

Recognizing that true error variance at longer lead times is dominated by the fastest growing components of the total error (since decaying errors diminish early on), SAFE-I uses a most economical, 1-dimensional model to describe error evolution constrained in the subspace of the leading (i.e. fastest growing) LLVs.

Such a model, however, cannot describe the transitional behaviour arising early on in a forecast due to neutral or decaying analysis errors. To assess the behaviour of decaying errors and to enhance the accuracy of growing error variance estimates, here we propose a generalization of the SAFE-I algorithm. While SAFE-I assumes all errors are confined in the subspace of the leading LLVs and grow exponentially, the new method called SAFE-II introduces a second, exponentially *decaying* component orthogonal to the growing direction, accounting for all non-growing errors.

The total analysis error variance in SAFE-II is thus described as the sum of the growing and decaying components:

$$x_0^2 = g_0^2 + d_0^2, \tag{9}$$

where g_0^2 and d_0^2 are the initial growing and decaying error variances, respectively. In the forecast phase, the growing component expands exponentially, while the decaying component shrinks exponentially, yielding the following sum for the total true forecast error variance (*Assumption 2* behind SAFE-II):

$$x_t^2 = g_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha} + d_0^2 \cdot e^{i \cdot \Delta t \cdot \beta}, \tag{10}$$

where β is a negative value representing the exponential decay rate. The transitional behaviour of the total error (solid) due to the vanishing decaying errors (dotted) is illustrated in Figure 1. Following the initial transitional period during which most of the decaying errors disappear, the total error follows the evolution of the exponential component (dashed line in Figure 1). By substituting x_0^2 and x_t^2 in Equation 7 with Equations 9 and 10, the perceived error variance simulated with the two additional SAFE-II parameters (d_0 and β) can be written as:

$$\hat{f}_i^2 = g_0^2 + d_0^2 + g_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha} + d_0^2 \cdot e^{i \cdot \Delta t \cdot \beta} - 2\rho_1^i \cdot \sqrt{g_0^2 + d_0^2} \cdot \sqrt{g_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha} + d_0^2 \cdot e^{i \cdot \Delta t \cdot \beta}}. \tag{11}$$

2.3 | Use of additional measurements

Given the challenge of estimating two extra parameters compared to SAFE-I, we explored whether additional measurements beyond perceived errors could be used for the

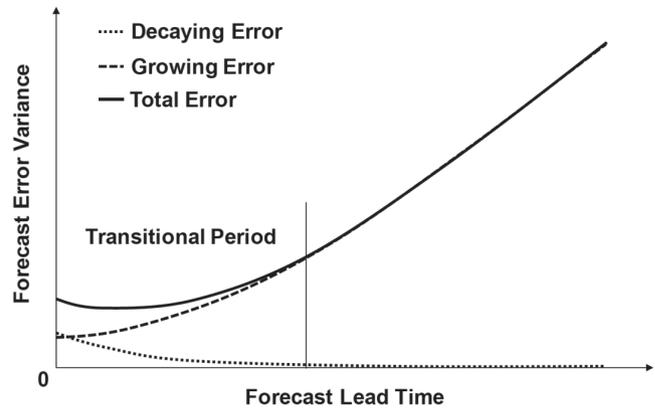


FIGURE 1 Schematic of the evolution of the true forecast error variance (solid) and its growing (dashed) and decaying (dotted) components

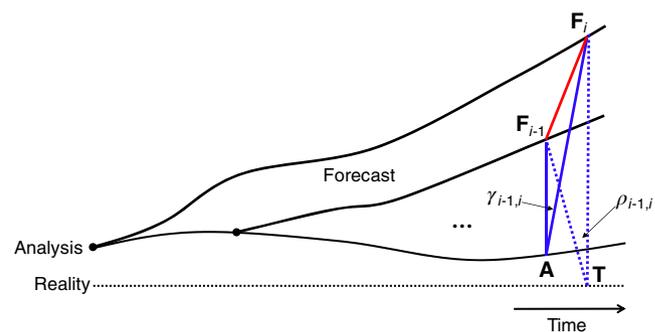


FIGURE 2 A 3D schematic of the relationship between the correlations of true (\mathbf{TF}_{i-1} and \mathbf{TF}_i , $\rho_{i-1,i}$) and perceived (\mathbf{AF}_{i-1} and \mathbf{AF}_i , $\gamma_{i-1,i}$) errors, all valid at the same time. \mathbf{F} , \mathbf{A} and \mathbf{T} represent forecast, analysed and true states, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

reduction of uncertainty in SAFE-II parameter estimation. Lagged Forecast Differences (i.e. the differences between two different lead time forecasts valid at the same time, hereafter LFD) is one such measurable quantity. In Figure 2, \mathbf{F}_{i-1} and \mathbf{F}_i are two such forecasts (with lead times of $(i-1) \cdot \Delta t$ and $i \cdot \Delta t$), while \mathbf{T} and \mathbf{A} denote the true and analysed states, respectively, all valid at the same time. In triangle $\mathbf{TF}_{i-1}\mathbf{F}_i$ (blue dotted lines), the LFD variance between \mathbf{F}_{i-1} and \mathbf{F}_i ($f_{i-1,i}^2$, red solid line) can be expressed as:

$$f_{i-1,i}^2 = x_{i-1}^2 + x_i^2 - 2\rho_{i-1,i} \cdot x_{i-1} \cdot x_i, \tag{12}$$

where $\rho_{i-1,i}$ is the correlation between \mathbf{TF}_{i-1} and \mathbf{TF}_i (blue dotted lines).

To enable the use of LFD measurements in SAFE-II without the introduction of $\rho_{i-1,i}$ as an additional unknown parameter, we introduced three additional assumptions.

Assumption 4. Relationship between true and perceived error variance. We first note that the correlation $\gamma_{i-1,i}$

between \mathbf{AF}_{i-1} and \mathbf{AF}_i (blue solid lines in Figure 2) can be readily calculated from perceived error measurements. We further note that while true forecast error variance grows exponentially as a function of lead time, analysis error variance x_0^2 remains the same. Therefore, at sufficiently long lead times, the true and perceived errors become similar in magnitude:

$$f_{i-1}^2 \sim x_{i-1}^2, \quad (13a)$$

$$f_i^2 \sim x_i^2, \quad (13b)$$

and therefore correlation $\rho_{i-1,i}$ can be well approximated by the measurable quantity $\gamma_{i-1,i}$. In this study we assume that the perceived and true forecast error variances become sufficiently similar at 2.25 ($i = 9$) and 2.5 days ($i = 10$) lead time, assuring that $\rho_{9,10} \approx \gamma_{9,10}$.

Assumption 5. Transient period. As further simplifications, we also assume that any transient error behaviour subsides within 24 hr.⁴

Assumption 6. Divergence rate of model trajectories. We also assume that the model, by reasonably capturing natural instabilities, reproduces the chaotic divergence of trajectories of the model and nature. Therefore, beyond 24 hr lead time, the three sides of triangle $\mathbf{TF}_{i-1}\mathbf{F}_i$ grow at the same pace, corresponding to the dynamically sustainable growth rate of the errors between model and reality (i.e. parameter α). It follows that correlations $\rho_{i-1,i}$ remain approximately unchanged beyond 24 hr ($i \geq 5$) and equal to $\rho_{9,10}$. Assumptions 1–6 are summarized in Table 1 and their validity will be investigated in section 4 in an OSSE environment.

Based on the above assumptions and substituting x_i^2 in Equation 12 with Equation 10, the evolution of LFD between 1 to 2.5 days lead times can be modelled with only two parameters, g_0 and α , that are also used in the simulation of perceived forecast error variance (cf. Equation 11):

$$\widehat{f_{i-1,i}^2} = g_0^2 \cdot e^{(i-1) \cdot \Delta t \cdot \alpha} + g_0^2 \cdot e^{i \cdot \Delta t \cdot \alpha} - 2\gamma_{9,10} \cdot g_0^2 \sqrt{e^{(i-1) \cdot \Delta t \cdot \alpha} \cdot e^{i \cdot \Delta t \cdot \alpha}}, \quad (i = 5, 6, \dots, 10). \quad (14)$$

To distinguish between the lead times of perceived error and LFD measurements and associated weights, the index i is replaced with j in Equation 14 before an LFD term is incorporated into the SAFE-II cost function of

Equation 8:

$$J = \max(|f_i^2 - \widehat{f_i^2}| \cdot w_i^{-1}) + \max(|f_{j-1,j}^2 - \widehat{f_{j-1,j}^2}| \cdot w_{j-1,j}^{-1}) \quad (i = 1, 2, \dots, 10; j = 5, 6, \dots, 10). \quad (15)$$

Since the simulation of LFD (Equation 14) does not involve decaying errors, LFD variance measurements, if desired, can also be incorporated into the cost function of SAFE-I.

3 | DATASETS USED

SAFE-II will be applied to estimate true error variance in GFS analyses and forecasts, first in a simulated (OSSE), then in a realistic operational forecast environment at NCEP. Six, 12, ..., 60 hr perceived error measurements will be calculated over the extratropical Northern Hemisphere (NH; 30°–90°N) on a 1° × 1° regular latitude/longitude grid. A cosine weight of latitude is used when calculating the area mean error variance so as to avoid undue weights on data from higher latitudes. The choices of the spatial domain and lead time range minimize the effects of nonlinearities (Gilmour *et al.*, 2001) or model-related errors (Orrell *et al.*, 2001).

3.1 | OSSE data

SAFE-II assumptions will be validated (Section 4) and estimates evaluated (Section 5) using the OSSE set-up⁵ described in detail by Appendix B and Cucurull *et al.*, (2017). The three variables used are zonal wind (U), temperature (T), and geopotential height (GH). For the OSSE data used in this study, 6 hr analyses are used corresponding with lower boundary conditions between 3 July and 26 August 2005, with 7-day forecasts initialized only at every 0000 UTC. The SAFE-II cost function (Equation 15) is therefore modified to use 24-, instead of 6-hr lagged forecasts:

$$J = \max(|f_i^2 - \widehat{f_i^2}| \cdot w_i^{-1}) + \max(|f_{j-4,j}^2 - \widehat{f_{j-4,j}^2}| \cdot w_{j-4,j}^{-1}), \quad (i = 1, 2, \dots, 10; j = 8, 9, 10). \quad (16)$$

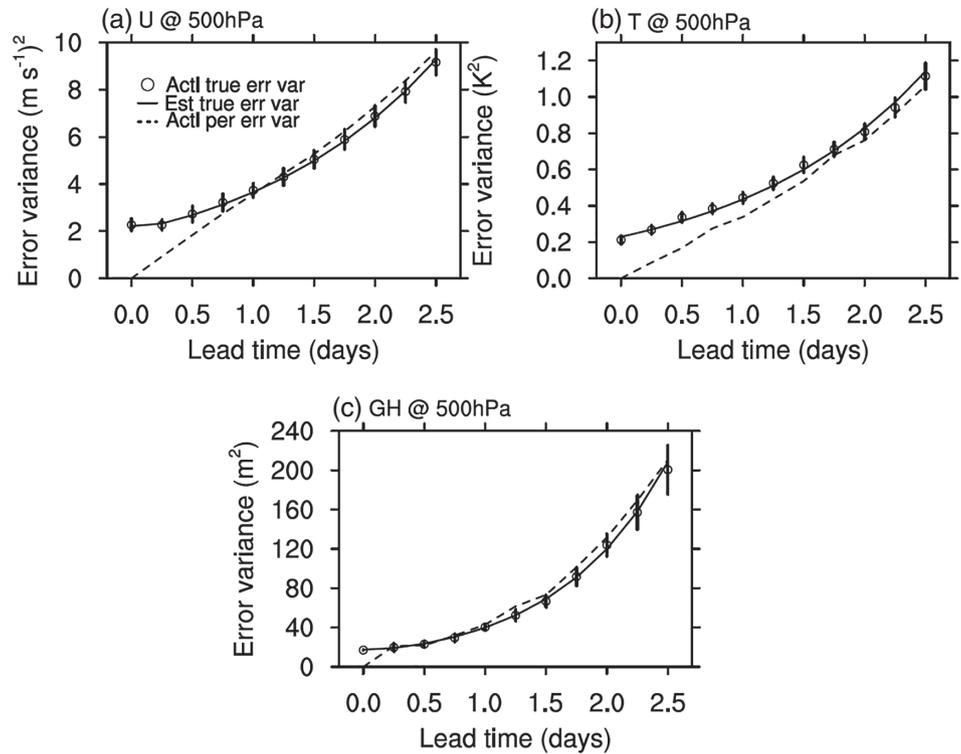
3.2 | Operational GFS data

In Section 5, SAFE-II will also be used to assess analysis and forecast error variance in the operational, T1534,

⁴While true forecast error is largely unaffected by decaying errors beyond 24 hours lead time, the perceived error remains affected by decaying errors present in the verifying analyses, allowing for the estimation of decaying parameters in SAFE-II.

⁵Short of having access to data from OSSE, analysis and forecast errors could also be simulated with perturbed fields from an ensemble of analyses and forecasts (Houtekamer *et al.*, 2005; Feng *et al.*, 2017).

FIGURE 3 Sample-mean based estimates of ground truth for true forecast error variance (open circles with 95% vertical confidence intervals as vertical bars) along with the corresponding fitted values (solid line) for variables (a) U, (b) T, and (c) GH, at 500 hPa in the OSSE environment. For comparison, perceived error variance measurements are also shown as dashed lines



64-level resolution GFS system (Yang, 2016). Analysis/forecast data are sampled every 6 hr and cover the 1 September–30 November 2015 period.

Note that as mentioned before, the OSSE experiments introduced in Section 3.1 use an earlier version of the NCEP NWP system (e.g. without a hybrid DA). Therefore, SAFE-II error estimates from the OSSE and operational environments cannot be directly compared.

4 | VALIDATION OF SAFE-II ASSUMPTIONS

An OSSE environment offers an ideal ground for the evaluation of the assumptions behind SAFE-II since not only the perceived error, but various characteristics of the true forecast error (e.g. error variance and the correlation between analysis and forecast errors) are also directly measurable.

4.1 | Basic assumptions

A key assumption (*Assumption 2* in Sections 2.1 and 2.2) states that the true forecast error variance can be considered as a sum of exponentially growing and decaying error components (Equation 10). In an OSSE environment, we can directly assess the validity of *Assumption 2* by fitting the error evolution relationship in Equation 10 to the time mean of *true error variance measurements*, through minimizing the following cost function:

$$J = \max(|x_i^2 - \hat{x}_i^2| \cdot \omega_i^{-1}), (i = 0, 1, \dots, 10), \quad (17)$$

where x_i^2 and \hat{x}_i^2 are the measured and modelled true forecast error variances, respectively, and ω_i^{-1} is the weight related to the SEM-based sampling error (refer to Appendix A) of the measured true error variances.

Figure 3 shows the sample (time) mean of directly measured true analysis and forecast error variance, along with 95% confidence intervals reflecting the effect of sampling errors, and a simulated error variance curve fitted to the sample mean of the measurements using the error decomposition of Equation 10, as a function of 0–60 hr lead time, for three selected variables. All simulated values fall within the 95% confidence intervals, indicating that *Assumption 2* about the decomposition of forecast errors (Section 2.2) is consistent with the experimental data.

The four estimated parameters with SAFE-II are listed in Table 2 (the first 4 values in “Fit SAFE-II” rows). These will be used as a reference for ground truth in the evaluation of SAFE-II estimates in Section 5. The performance of SAFE-I is also shown in Table 2. As expected, when decaying errors are absent (see variable T in Table 2), the two versions of SAFE identify the same exponential error growth (with identical fitted x_0^2 and α values). In the presence of decaying analysis errors (variables U and GH), the SAFE-I error growth model still offers a statistically acceptable fit except for the estimated total analysis error variance of GH. The good fitting of exponential growing forecast error also justifies that *Assumption 1* can be considered

TABLE 2 Comparison between fitted (fit) and reference (ref) values of error parameters for zonal wind (U), temperature (T), and geopotential height (GH) at 500 hPa in the observing system simulation experiments (OSSE) using SAFE-I and SAFE-II, respectively

| | g_0^2 | $e^{\Delta t \cdot \alpha}$ | d_0^2 | $e^{\Delta t \cdot \beta}$ | $g_0^2 + d_0^2$ | $\frac{d_0^2}{g_0^2 + d_0^2}$ | LFD var growth |
|---------------|---------|-----------------------------|---------|----------------------------|-----------------|-------------------------------|----------------|
| U: | | | | | | | |
| Fit: SAFE-I | 2.09 | 1.157 | 0.0 | - | 2.09 (8.0%) | 0.0 | 1.146 |
| Fit: SAFE-II | 1.96 | 1.168 | 0.248 | 0.221 | 2.21 (2.6%) | 11.2% | |
| Ref / 1.96SEM | - | - | - | - | 2.27 / 0.265 | - | |
| T: | | | | | | | |
| Fit: SAFE-I | 0.229 | 1.174 | 0.0 | - | 0.229 (9.0%) | 0.0 | 1.169 |
| Fit: SAFE-II | 0.229 | 1.174 | 0.0 | - | 0.229 (9.0%) | 0.0 | |
| Ref / 1.96SEM | - | - | - | - | 0.210 / 0.024 | - | |
| GH: | | | | | | | |
| Fit: SAFE-I | 14.9 | 1.288 | 0.0 | - | 14.9 (12.9%) | 0.0 | 1.278 |
| Fit: SAFE-II | 13.1 | 1.318 | 4.34 | 0.368 | 17.5 (2.2%) | 24.8% | |
| Ref / 1.96SEM | - | - | - | - | 17.1 / 2.10 | - | |

Note: g_0^2 and d_0^2 denote the growing and decaying components, with $\Delta t = 6$ hr growth and decay rates of $e^{\Delta t \cdot \alpha}$ and $e^{\Delta t \cdot \beta}$. The values in brackets indicate the percentage of estimation error compared to ground truth (reference). Entries with “-” indicate where parameter values are not available. The rightmost column lists the growth rate of lagged forecast difference (LFD) variance per 6 hr. The units of the error variances (g_0^2 and d_0^2) for U, T and GH are $(m \cdot s^{-1})^2$, K^2 and m^2 , respectively.

to be valid. However, with its additional two parameters, SAFE-II provides a considerably improved simulation of the analysis error compared to SAFE-I (2–3%, instead of 8–13% deviation from the reference measured true error, see corresponding numbers in parentheses in Table 2).

When decaying errors are present (variables U and GH) but not considered, SAFE-I arrives at higher initial growing error variance and lower growth rate estimates than SAFE-II (Table 2). SAFE-II finds the largest proportion of decaying errors in GH (about 24.8% of the total analysis error variance), followed by U (11.2%). As an example, Figure 4 illustrates the evolution of the estimated growing, decaying and total error variance for GH by SAFE-II from 0 to 1.5 days. It is analogous to the behaviour of growing and decaying components of true forecast error variance in the schematic figure (Figure 1). Variable U has qualitatively similar error evolution. It is consistent for the variables that decay is so fast that within 24 hr the percentage of decaying errors drops below 1% of the total forecast error variance. This validates *Assumption 5* in Section 2.3. The growth rate of LFD variance (the rightmost column in Table 2) has only up to 3% deviation from that of the true forecast error variance for all variables (Table 2) which indicates *Assumption 6* is reliable.

Another key assumption (*Assumption 3* in Section 2.1) states that the correlation between true analysis and

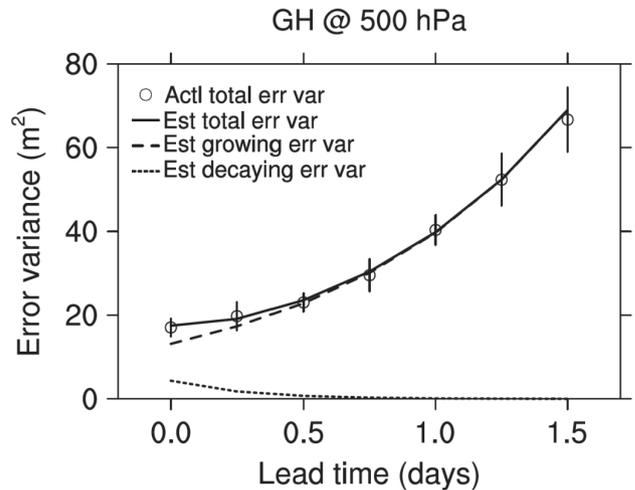
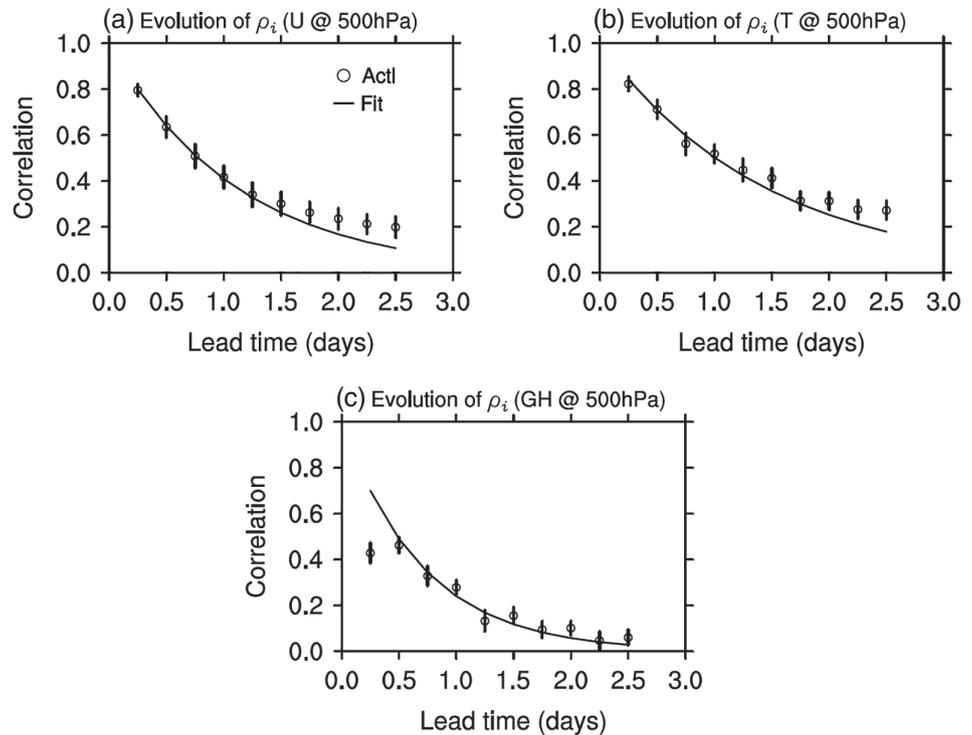


FIGURE 4 Estimates of growing (dashed line), decaying (dotted line) and total (solid line) error variance along with the corresponding fitted values (hollow circles with 95% confidence intervals as vertical bars) for GH at 500 hPa in the OSSE environment

forecast errors (ρ_i) exponentially decays with increasing lead time (Equation 6). In an OSSE environment, “ground truth” correlation values ρ_i can be diagnosed from true and perceived error variance measurements using a transformed version of Equation 3:

FIGURE 5 Sample-mean fitted (solid line) and measured (circle) ρ_i for variables (a) U, (b) T, and (c) GH, at 500 hPa in the OSSE. Vertical bars represent 95% confidence intervals



$$\rho_i = \frac{(x_0^2 + x_i^2 - f_i^2)}{(2 \cdot x_0 \cdot x_i)}. \quad (18)$$

To test *Assumption 3*, we simulate ρ_i with the exponential decorrelation relationship of Equation 6 and then fit the simulated curve to the time mean of the diagnosed quantities (i.e. ground truth from Equation 18) by minimizing a cost function analogous to Equation 17.

The results in Figure 5 reveal a reasonable correspondence between sample-based mean and simulated values of the correlation between analysis and forecast errors. For the two model prognostic variables (U and T), the fitted values of ρ_i are within the 95% sampling error interval of its sample-based mean values throughout the first 2 days. This indicates that the exponential degradation of ρ_1 (*Assumption 3*) is consistent with the experimental data. Returning to Figure 3 we observe that due to the relatively high correlation between analysis and short-range forecast errors (Figure 5), perceived errors for the model variables are significantly lower than the true errors (Figure 3). At 6 hr lead time, for example, the perceived error measurements for U and T provide a 2–3-fold underestimate of the true error variance. This will be further discussed in the context of the operational forecast system in Section 5.1.2.

Note that the third variable shown in Figure 5, GH, is not a directly analysed variable; rather, it is derived from analysis control variables. Simulated ρ values for GH are nevertheless consistent with the ground truth, albeit only at and beyond 12 hr lead time. The deviation of 6 hr ρ is possibly due to some random noise or bias introduced in

the calculation of GH (e.g. a particular discretization of the hydrostatic equation: Wee *et al.*, 2012) in the OSSE that makes the ρ assumption invalid.

4.2 | LFD-related assumptions

Recall from Section 2.3 that correlation $\rho_{9,10}$ between lagged true forecast errors \mathbf{x}_9 and \mathbf{x}_{10} in cost function (Equation 14) is specified by the correlation $\gamma_{9,10}$ between lagged perceived errors \mathbf{f}_9 and \mathbf{f}_{10} , valid at the same time. Since in an OSSE environment both angles are directly measurable, the accuracy of approximating correlation “ ρ ” with “ γ ” can be tested. Figure 6 displays ρ and γ as a function of lead time with forecasts lagged 24 hr apart (since the OSSE forecasts are available only once, instead of four times per day) for variables U, T and GH at 500 hPa height.

At 36/60 hr lead time, the correlation between lagged true and perceived errors differs less than 0.025. This can be explained by the small differences between true and perceived error variances at and beyond 36 hr lead time (Figure 3), validating *Assumption 4* (Section 2.3), and thus $\rho_{9,10} \approx \gamma_{9,10}$.

The correlation between lagged true errors in Figure 6 exhibits less than 0.01 variations beyond 24/48 hr lead time. This indicates that once transient decay subsides, triangles $\mathbf{T}\mathbf{F}_{i-1}\mathbf{F}_i$ are approximately similar. This effectively validates *Assumption 6* about the close similarity of exponential expansion or growth rates in the attractor of a model and that of a model trajectory diverged from reality.

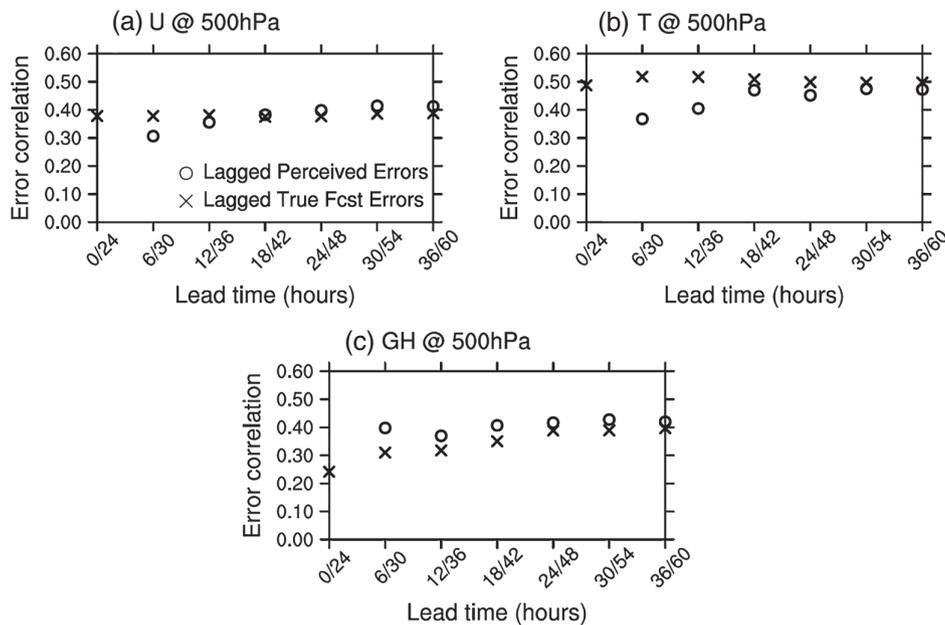


FIGURE 6 Comparison between the correlations of lagged perceived (circle) and true forecast (cross) errors with a lag of 24 hr as a function of lead time for variables (a) U, (b) T, and (c) GH, at 500 hPa in the OSSE

5 | ASSESSMENT OF ERROR VARIANCE IN OSSE AND OPERATIONAL ANALYSIS/FORECAST SYSTEMS

In this section, the SAFE-II algorithm described in Section 2 will be used to estimate Grid-point Statistical Interpolation (GSI)-GFS true analysis and forecast error variances. The estimates will be based on measurements of perceived error and LFD variances, first from an OSSE environment, then from the operational NCEP system.

5.1 | Error variance in selected variables

5.1.1 | OSSE environment

In Section 4.1, error decomposition Equation 10 was fitted directly to the time mean of true error variance measurements from an OSSE experiment. Here we will proceed as if we did not know reality and use perceived error variance measurements from the same OSSE analysis/forecast system to estimate the true error variance. In the error estimation experiments reported here, the truth will be used only in the evaluation of the results.

The quality of these practical estimates will be assessed by comparing them with a characterization of true growing (g_0^2) and decaying (d_0^2) error variances and their amplification and decay rates (α and β). As mentioned in Section 4.1, the fitted parameter values of SAFE-II in Table 2 are used as reference values for the estimates presented here. Along with these reference values, Table 3 shows the SAFE-I and SAFE-II error parameter estimates for 500 hPa analysis

variables U and T for the OSSE experiments. Estimates of GH are strongly influenced by the deviation of ρ at 6 hr (see discussion on Figure 5c) and thus are not shown. SAFE-II estimates of growing error variance and growing rate are closer to the reference values than the SAFE-I results, except for T where the results are identical since no decaying component is identified by SAFE-II. SAFE-II estimates are within 5–10% of the reference values for growing error variance and within 2% for their amplification rate.

Since decaying errors shrink fast and practically disappear within 24 hr (see Section 4.1), their estimation is especially challenging: only the first few perceived error variance measurement points provide meaningful information about their behaviour. LFD measurements used in the cost function (see Equation 14) are no help with the estimation of decaying parameters as they constrain only the estimation of the growing parameters. SAFE-II decaying error variance and decay rate estimates for model variable U contain relatively large, nearly 25% and 50% deviation from their reference values, respectively. The total error variance estimate for U is more accurate with SAFE-II than SAFE-I (1% vs. 9% error), though both are statistically reliable at the 95% confidence level. Their estimates of error correlation are similar.

5.1.2 | NCEP operational analysis/forecast system

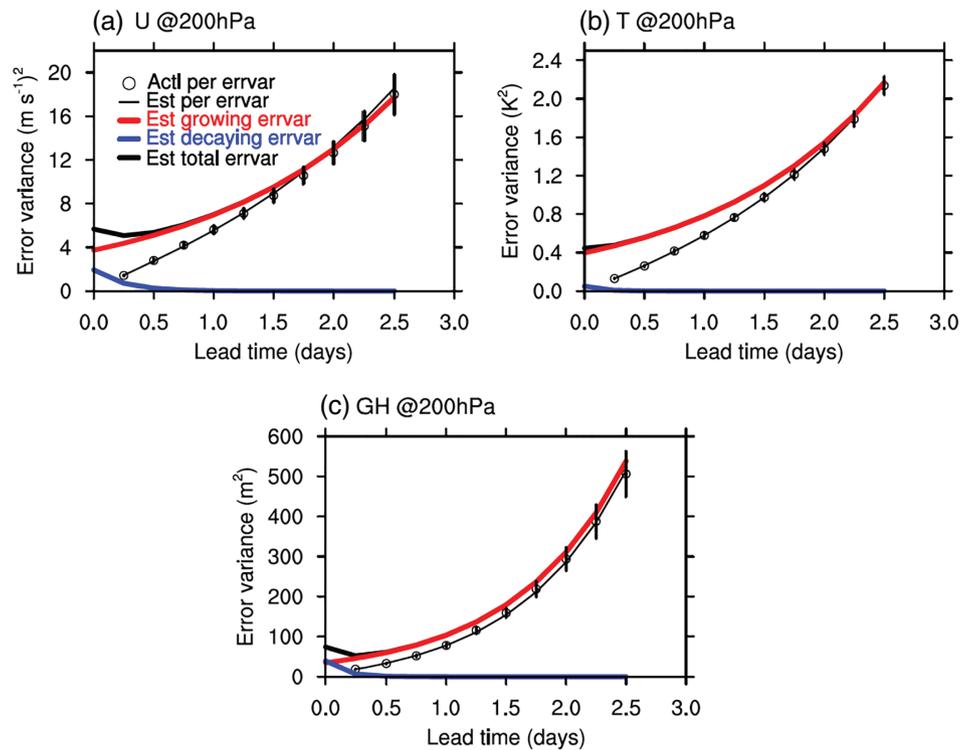
The main results of the study are visualized in Figure 7. The 200 hPa (a) U, (b) T, and (c) GH variables in the NCEP operational system are chosen for demonstration as decaying errors constitute a sizable portion of analysis error at

TABLE 3 SAFE-I and SAFE-II estimates of error evolution parameters for 500 hPa U and T in the OSSE experiments

| | g_0^2 | $e^{\Delta t \cdot \alpha}$ | d_0^2 | $e^{\Delta t \cdot \beta}$ | x_0^2 | $d_0^2 / (g_0^2 + d_0^2)$ | ρ_1 |
|-----------|---------|-----------------------------|---------|----------------------------|--------------|---------------------------|---------------|
| U: | | | | | | | |
| SAFE-I | 2.07 | 1.149 | 0.0 | - | 2.07 | 0.0 | 0.792 |
| SAFE-II | 2.06 | 1.153 | 0.19 | 0.34 | 2.25 | 8.4% | 0.804 |
| Ref | 1.96 | 1.168 | 0.25 | 0.22 | 2.27 (0.27) | 11.2% | 0.796 (0.023) |
| T: | | | | | | | |
| SAFE-I | 0.21 | 1.165 | 0.0 | - | 0.21 | 0.0 | 0.810 |
| SAFE-II | 0.21 | 1.165 | 0.0 | - | 0.21 | 0.0 | 0.810 |
| Ref | 0.23 | 1.174 | 0.0 | - | 0.21 (0.024) | 0.0 | 0.824 (0.031) |

Note: Reference values (Ref) are the SAFE-II fitted parameter values from Table 2. The values in brackets indicate the 95% sampling uncertainty confidence intervals of Ref.

FIGURE 7 Temporal variation of sample-mean actual (circle) and simulated (black thin line) perceived error variances and estimated total (black thick line), growing (red) and decaying (blue) true forecast error variances over Northern Hemisphere for variables (a) U, (b) T, and (c) GH, at 200 hPa for GFS-GSI operational forecast system



this level (see Section 5.2). Figure 7 shows the perceived error variance measurements (black open circles) with a 95% confidence level sampling uncertainty (black vertical bars) as a function of lead time. The corresponding simulated perceived error variance (thin black curve), and the estimated total (thick black curve), growing (red) and decaying (blue) error variance are also shown.

For all variables and at all lead times, the simulated perceived error variance falls within the 95% SEM uncertainty intervals of the perceived error measurements, indicating that the SAFE-II error behaviour model is consistent with the experimental measurement data. The results in

Figure 7 confirm the error behaviour indicated by the schematic Figure 1. After a relatively short, 12–18 hr transitional period during which the decaying error component vanishes, the total error assumes an exponential growth.

Figure 7 confirms a finding from the OSSE experiments (Section 4.1) that the conventional measure of forecast performance, perceived error variance, can be a rather poor estimate of the true short-range forecast error variance. At 6 hr lead time, for example, the perceived error measurements provide a 3–4-fold underestimate of the true error variance similarly as in the OSSE environment (Figure 3).

| | g_0^2 | $e^{\Delta t \cdot \alpha}$ | d_0^2 | $e^{\Delta t \cdot \beta}$ | $g_0^2 + d_0^2$ | $d_0^2 / (g_0^2 + d_0^2)$ | ρ_1 |
|------------|---------|-----------------------------|---------|----------------------------|-----------------|---------------------------|----------|
| U: | | | | | | | |
| 200 hPa | 3.74 | 1.17 | 1.93 | 0.37 | 5.67 | 34.0% | 0.87 |
| 500 hPa | 3.67 | 1.16 | 0.0 | - | 3.67 | 0.0 | 0.83 |
| T: | | | | | | | |
| 200 hPa | 0.39 | 1.19 | 0.049 | 0.35 | 0.439 | 11.2% | 0.86 |
| 500 hPa | 0.25 | 1.21 | 0.0 | - | 0.25 | 0.0 | 0.84 |
| GH: | | | | | | | |
| 200 hPa | 34.60 | 1.32 | 39.47 | 0.14 | 74.07 | 53.3% | 0.87 |
| 500 hPa | 24.72 | 1.32 | 34.88 | 0.14 | 59.60 | 58.5% | 0.87 |

TABLE 4 Estimated error parameters for U, T and GH at 200 and 500 hPa in GFS operational forecasts using SAFE-II

The discrepancy is due to the fact that perceived error measurements do not reflect the presence of error in the verifying analysis fields that is relatively highly correlated with the error in the forecasts. It is not until 2 days lead time that the deviation of simulated or measured perceived error variance from the true error variance drops below 5% of the true error variance. The use of perceived error as an estimate of true forecast error thus leads to an underestimation of error variance and an overestimation of error growth (PT14). The overestimation of error growth when perceived error variances are used may partially explain the apparent lack of sufficient spread and perceived deficiency in perturbation growth in most ensemble systems studied (see, e.g., Buizza *et al.*, 2005), as well as the difference between “external” (i.e. verified against analyses of the atmosphere) vs. “internal” (verified against another model forecast) predictability and error growth noted by Lorenz (1982) and a series of follow-on studies.

5.1.3 | Comparison with results from OSSE analysis/forecast system

Table 4 summarizes the NCEP operational forecast system results for 200 hPa height variables displayed in Figure 7. For an easy comparison with results from the error evolution of an OSSE experiment in Section 4.1, results for 500 hPa height variables are also shown in Table 4. Comparing SAFE-II estimates from Table 3 with 500 hPa estimates from Table 4, we first note that both the total and growing analysis errors appear to be severely underestimated by the OSSE system for U and GH. This may be the result of tuning OSSE error variances to match operational perceived error variances that, as noted above, are significantly lower than true error variances.

Interestingly, error growth rate estimates for the NCEP operational system verified against operational analyses

for September–November 2015 (Table 4), and against an ECMWF high-resolution simulation with July–August lower boundary forcing (OSSE nature run, Table 3) display less than 3% difference for the two model variables U and T at the 500 hPa height level. This may be an indication that when properly assessed, external (model vs. reality) and internal (model vs. model) error growth, after all, may be rather similar. These results are also consistent with *Assumption 6* in Section 2.3.

Analysis vs. forecast error correlations are found slightly (about 0.03) higher in the operational system for all variables (cf. ρ_1 in Tables 3 and 4). For U, the OSSE analysis also contains a larger decaying component. The use of an improved hybrid DA scheme and increased model resolution in the operational vs. the OSSE set-up, as well as the addition of too much noise in the simulation of observational error in the OSSE system, may both contribute to the correlation and decaying error results above.

Table 4 also lists error variance and other estimated parameters for GH. While the GH analysis vs. forecast error correlation in Table 4 appears to be similar to or only slightly higher than those for the other variables, both the growth rate and the percentage of decaying error are markedly higher for GH than for the other variables. The latter result is qualitatively consistent with OSSE results in Table 2. Note that GH is not a GFS model or GSI analysis variable but rather is derived from model prognostic variables including temperature, surface pressure, and humidity (Houtekamer *et al.*, 2005) through the hydrostatic equation (Grell *et al.*, 1995). When the hydrostatic equation is integrated from the model surface to the top of the model to calculate GH, independent random error present in the prognostic variables may lead to a higher level of noise (i.e. decaying error) in GH compared with model prognostic variables. As for the GH growth rate, it corresponds to an error doubling time of 1.26 days, below Simmons *et al.* (1995)'s 1.5 days estimate

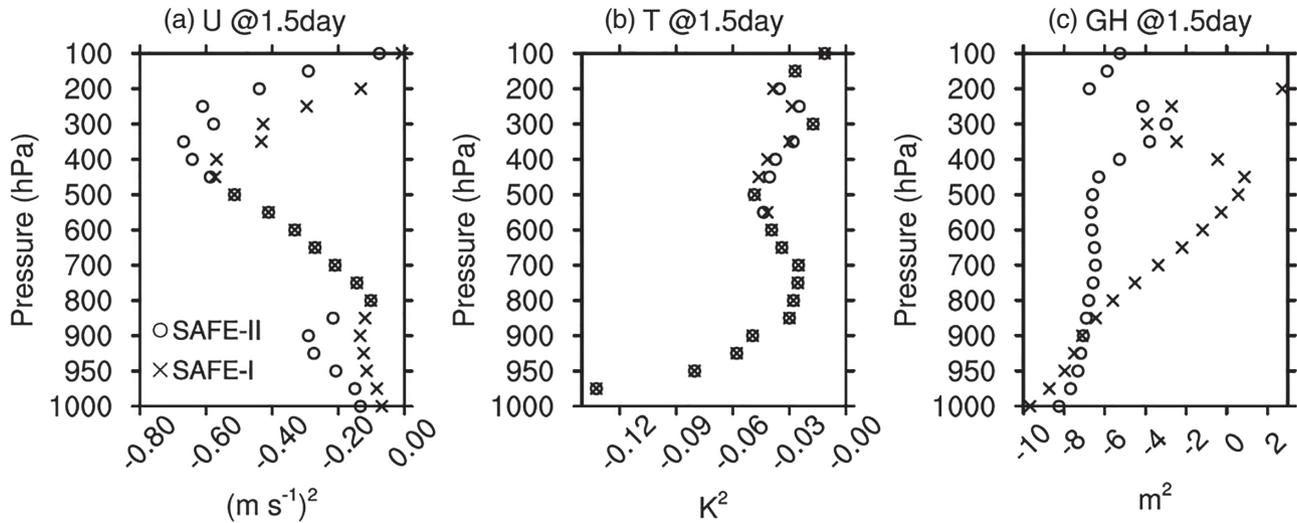


FIGURE 8 Profile of the differences between the absolute fitting errors and the 95% confidence interval of perceived error variances at 1.5 days by SAFE-II (circle) and SAFE-I (cross) for variables (a) U, (b) T, and (c) GH, for GFS-GSI operational forecast system

for 500 hPa height. It is not clear why the GH growth rate is significantly higher than that for the model variables.

5.2 | Vertical profile of analysis error variance

In this section, SAFE-II is used to estimate the vertical distribution of error variance from 1,000 to 100 hPa for the GFS-GSI operational forecast system. SAFE-I estimates are also given for a comparison.

5.2.1 | Fitting error

As described in Section 2.1, a critical part of SAFE-II is the evaluation of the fit of simulated perceived error curves to the sample-based (time) mean of perceived error measurements at all lead times considered. A fitting error say 95% of the time smaller than SEM at the 95% significance level indicates that experimental measurement data are consistent with the SAFE assumptions and error model. Figure 8 displays the difference between the absolute value of the fitting error of perceived error variance and the 1.96SEM confidence interval at the 1.5-day lead time for variables U, T and GH, for both SAFE-I and SAFE-II as follows:

$$|f_i^2 - \hat{f}_i^2| - 1.96\text{SEM}_i. \quad (19)$$

The results at other lead times are qualitatively similar. The fitting error is smaller than 1.96SEM for all vertical levels for both SAFE-I and SAFE-II, which indicates that both error models are consistent with the measurements

at the 95% confidence level. The lower negative values for SAFE-II indicate that the two additional parameters (the variance and decay rate of decaying errors) introduced in the present study offer a more complete error evolution model, attested by a closer fit to the perceived error measurements.

5.2.2 | Total error

Figure 9 displays the growing (red circles), decaying (blue circles), and total (black circles) analysis error variance for the three variables investigated: U, T and GH. SAFE-I estimates of the total error variance (that is all assumed to be growing) are shown as red crosses for a comparison. The 6 hr lead time perceived error variance measured as the difference between first guess and analysis fields is also provided (green plus signs) as a possible indicator of analysis quality.

Looking first at the total error of the two model variables, U has an absolute maximum around the upper-level jet (300 hPa), gradually/quickly dropping to lower/much lower values near the bottom/top of the domain. In contrast, T has two peaks, one presumably associated with the low-level jet (around 925 hPa), and a second one above the jet level (200 hPa). Interestingly, the ratio between the maximum and minimum total error variance in the vertical is in the 4–5 range for the two variables U and T. The vertical profile of GH is less pronounced, with an absolute and secondary maximum at 300 hPa and the surface, respectively.

As found earlier for selected variables in an OSSE setting (Section 5.1.1), when no decaying errors are

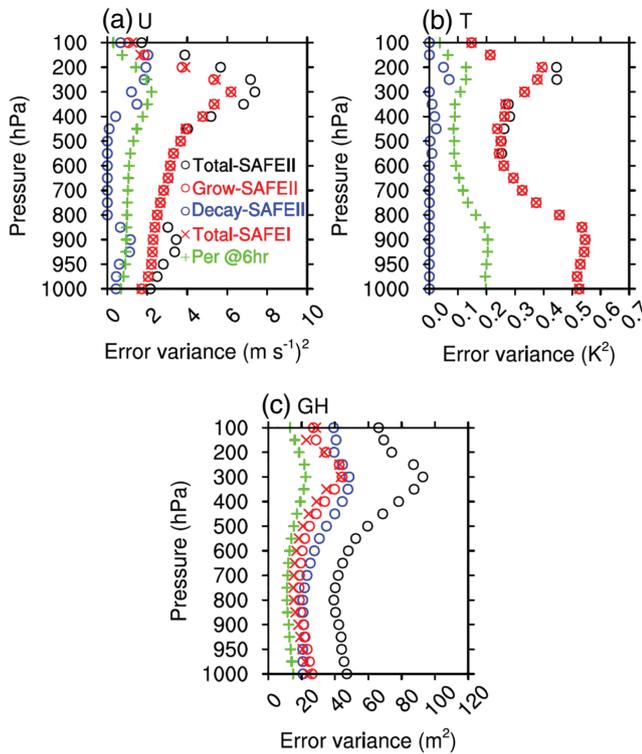


FIGURE 9 Same as Figure 8, but for estimated total (black circle), growing (red circle) and decaying (blue circle) analysis error variance by SAFE-II. Green and red crosses represent the 6 hr perceived error variance and the estimated analysis error variance by SAFE-I, respectively

detected, SAFE-I error estimates (red crosses in Figure 9) match the total error variance estimates of SAFE-II (black open circles). In the presence of decaying errors, SAFE-I still provides growing error estimates similar to SAFE-II; however, these estimates are lower than the total error since the decaying analysis errors are not directly accounted for.

Six-hour perceived error variance directly relates to the quality of background forecasts (or first guesses) in DA, and indirectly reflects error variance in the analysis. Profiles of 6 hr perceived error variance for the three variables correlate well (at 0.88 or higher values) with SAFE-II estimates of total analysis error variance profiles (Figure 9). As found in Section 5.1.2 for variables at 200 hPa height, perceived error measurements, however, are by a factor of 3–4 lower than estimates of true error through the entire profile of all variables. Such an underestimation can have profound impacts in the areas of data assimilation (underestimation of first-guess error variance), ensemble initialization (specification of too low initial ensemble spread), and OSSE system calibration (setting simulated analysis error variance at too low levels).

We mention that the estimated U and T total error variances in Figure 9 display similar vertical profiles to those

measured directly by Privé *et al.* (2013), see the thick solid lines in their Fig. 5a,d, and Privé and Errico (2013), see the heavy dashed lines in their Fig. 1a,d, in their OSSE studies. The actual error variance values from their studies, however, differ from the operational forecast system error estimates in Figure 9, just as was the case with the NCEP OSSE results (see related discussion in Section 5.1.3). Note that error levels may also differ due to distinctly different circulation regimes over the evaluation period of the operational vs. OSSE DA-forecast system.

5.2.3 | Growth rate

Beyond the variance and correlation of errors, SAFE-I and SAFE-II also provide estimates for the time evolution of error variance as a function of lead time. The 6 hr amplification factors for (a) U, (b) T, and (c) GH are displayed as a function of height in Figure 10. At all levels, GH has consistently faster error growth rate than the other two variables. For all variables, error growth peaks near the level of the midlatitude jet characterized by strong baroclinic instabilities at 300 hPa for U and GH, while around 450 hPa for T. Variations in growth rate across levels and variables reflect the instability properties of different dynamical processes, operating on various spatial scales. The slow error growth near the model top for the variables relative to other levels may be explained by the strongly diffusive model dynamics (Houtekamer *et al.*, 2005).

The model variables U and T also have a weaker maximum, near the low-level jet and surface, respectively. It is interesting to point out that for the two model variables, total analysis error variance has a corresponding maximum (typically 50 hPa for U and 150 hPa for T) above the double maxima observed in error growth. With vertically uniform observational coverage, analysis error maxima are expected to exactly collocate that of error growth. Given the density of *in situ* observations gradually decreases with altitude, the upward shift of analysis maxima from growth rate maxima is expected.

5.2.4 | Decaying errors

Just as shown for 200 hPa variables (Figure 7 and Table 4), the decaying errors are most prominent in GH fields throughout the entire atmosphere (Figure 9). This is even more visible in Figure 11 that depicts the vertical profile of the percentage of the decaying component in total analysis error variance (open circles) estimated by SAFE-II. As discussed in Section 5.1.3, decaying errors in GH may be accentuated by the formula used in their derivation from analysed model prognostic variables.

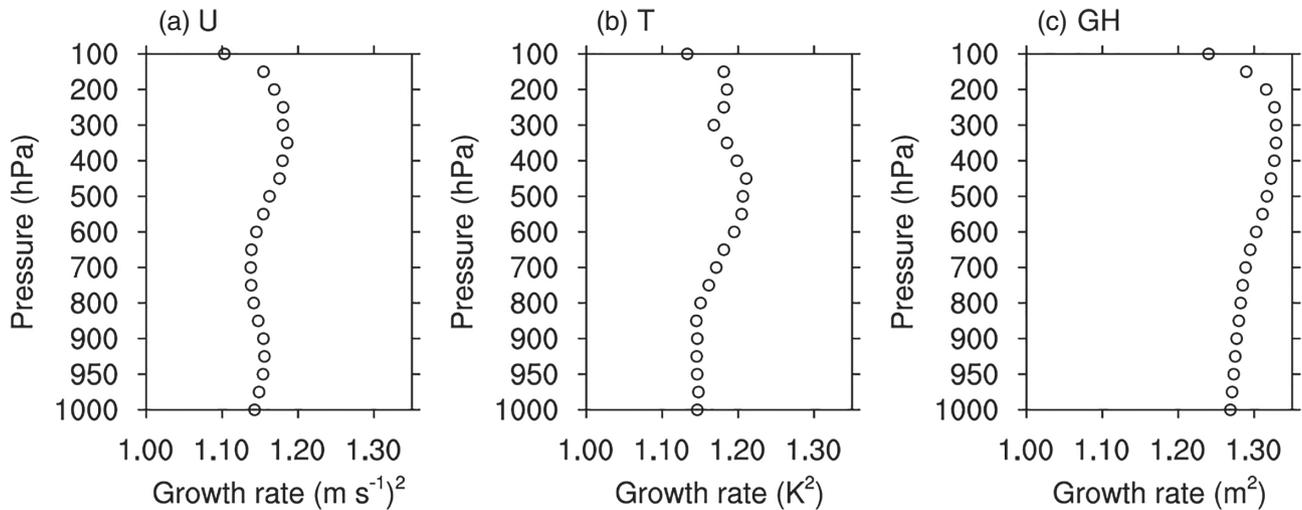


FIGURE 10 Same as Figure 8 but for the estimated growth rate of error variance per 6 hr

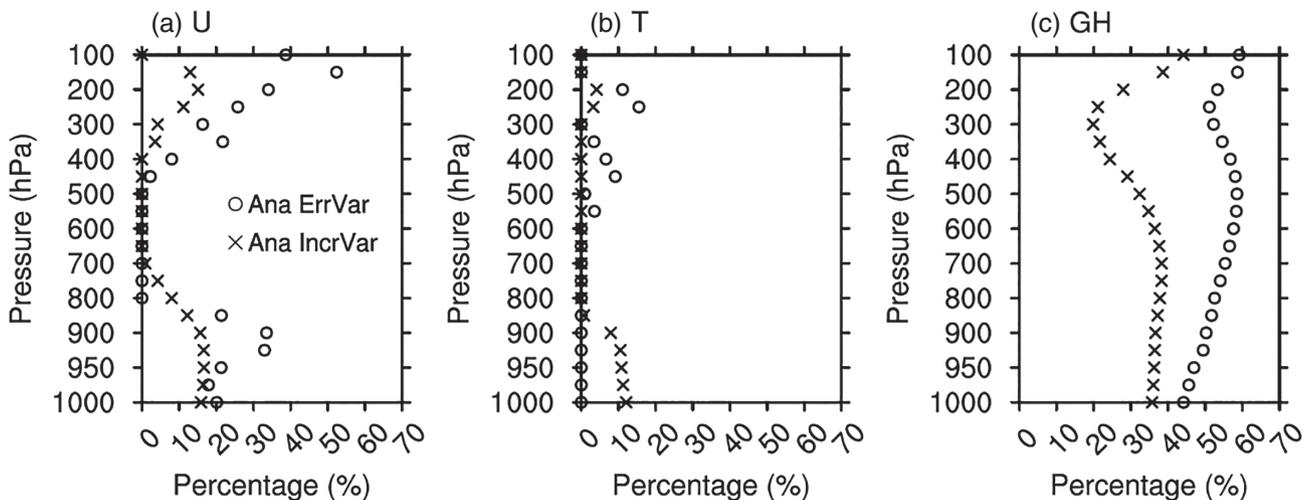


FIGURE 11 Same as Figure 8 but for estimated percentage of decaying components in total analysis error variance by SAFE-II (circle) and variance of analysis increment (cross)

In contrast, variable T is least affected by decaying errors, where they constitute less than 15% of the total analysis error, and only in the upper half of the atmosphere (Figures 9 and 11). U is in between, with two maxima situated around the upper- and lower-level jets, with a spike near the top of the model. The source of decaying errors in the analysed variables includes representativeness errors (especially near diverse topography: Quintana-Seguí *et al.*, 2008; Jiménez and Dudhia, 2012), approximations in balance constraints, observational noise, interpolations, localization, and other statistical and numerical procedures in DA. Interestingly, no decaying errors are found in the non-divergent mid-troposphere where commonly used balance constraints in the DA schemes may be most applicable. As noted in Section 5.1.1, due to their nature, decaying errors affect only analyses and short-range forecasts,

therefore their estimates are subject to a higher level of uncertainty. Further studies into the estimation of decaying errors are therefore warranted.

5.2.5 | The decaying component of analysis increments

An analysis field (Kalnay, 2003) is the sum of a first-guess forecast that as we saw itself contains decaying errors, and the analysis increment (AI) which is identical to the 6 hr perceived forecast error. It is well understood that the introduction of excessive noise into the analysis via the AI in a cycled DA system can negatively affect the quality of the analysis (Houtekamer and Mitchell, 2001; Dee, 2005). Hence the reduction of noise in AI has been a prominent

but hard-to-achieve goal in DA. As the noise introduced by the data assimilation step via the AI into the analysis field contributes to the overall level of decaying errors in the analysis, the vertical profile of the proportion of decaying errors in the analysis is expected to be qualitatively similar to that in the AI.

Since both measurements and simulated values of 6 hr perceived forecast errors are an integral part of SAFE-II, a convenient methodology offers itself for the estimation of noise in the form of decaying errors in AI. The method is based on the simulation of the variance in 6 hr LFDs with Equation 10, and then fitting the simulated curves to the sample mean of different lead time LFD variance measurements. A cost function similar to Equation 17 is used where x_i^2 and ω_i^{-1} are replaced by the sample mean of 6 hr LFD variance and its SEM-based weight, respectively.

Figure 11 shows the proportion of the decaying error component in the variance in AI and analysis fields as crosses and open circles, respectively. As expected, the vertical profile of the proportion of the decaying component of the analysis error is similar to, though 10–30% higher than the decaying error component in AI for all variables investigated. Just like in the analysis fields, decaying errors in AI are more pronounced in the upper and lower parts of the model domain, roughly as those in analysis errors. Note, however, that no decaying analysis errors are diagnosed for low-level temperature, despite their presence in AI.

6 | CONCLUSIONS

The evaluation of and improvements to data assimilation, ensemble forecasting, and observing system simulation techniques require knowledge of error variance in NWP analysis and forecast fields. Since reality is unknown, such error variance (i.e. “true” error variance) is directly not measurable. As observations are sporadic, most systematic studies resort to estimating error variance by comparing forecast fields with verifying analysis fields (i.e. “perceived” error). Such an approach (a) cannot assess errors in the analysis, and (b) ignores the effect of analysis error on forecast error estimates.

Peña and Toth (2014) proposed an inverse procedure called Statistical Analysis and Forecast Error (SAFE-I) algorithm for the bias-free estimation of true analysis and forecast error variance. SAFE-I uses perceived error measurements (defined with respect to the verifying analysis), and models them with a few parameters describing the evolution of the true error in time: the initial error variance (g_0^2), the dynamical growth rate (α), and the correlation between analysis and background forecast errors (ρ_1). The unknown parameters are estimated by minimizing the difference between the measured and modelled (via

the unknown parameters) perceived error at various lead times. SAFE-I is independent of assumptions and methods used in observing, DA, or prediction systems.

An important assumption in SAFE-I is that at short lead times the true forecast error variance (variances between forecasts and reality at the same time) grows exponentially. This assumption, however, neglects any noise that observations or the analysis procedure may inject into the analysis. Such errors typically project onto the stable manifold of the system and thus rapidly decay, manifesting as a transitional behaviour in the evolution of the total error variance. In this article, we relax the error evolution assumption in SAFE-I by the introduction of decaying, in addition to the growing analysis errors. Specifically, the modified SAFE method (SAFE-II) models true forecast error variance as the sum of an exponentially growing, and an orthogonal decaying component, the latter of which is described by its variance (d_0^2) and decay rate (β). The estimation of the expanded set of parameters in SAFE-II is facilitated by the inclusion of additional measurements in the form of variances between lagged forecasts valid at the same time, linked with parameters g_0^2 and α .

When decaying errors are present, the true forecast error variance may display an initial transitional behaviour, during which total error may decay or exhibit slower than exponential growth while decaying errors diminish. Only after most decaying errors vanish, does the total error assume an exponential pattern of growth.

The performance of SAFE-II was evaluated using the NCEP GFS/GSI system. First, the assumptions behind SAFE-II were validated in an OSSE environment where reality is exactly known. Area mean (Northern Hemisphere extratropics) true analysis and forecast error variance were simulated by the error growth equation used in SAFE-II, and fitted to sample-based measurements of these quantities from an OSSE system. Error variance simulated by SAFE-II was found to be within sampling uncertainty of the sample-based measurements. This, along with other related results indicate that the assumptions behind SAFE-II are consistent with the experimental data.

Next, in the same OSSE environment, we pretended that truth is unknown and used only perceived error measurements and SAFE-I or SAFE-II to produce and validate against true error variance estimates. In the presence of decaying errors (variable U), all SAFE-II error parameter estimates were found to be more accurate than those with SAFE-I. For the two model variables tested (500 hPa U and T), SAFE-II estimates of total analysis error variance were within 1% of the actual measured values, while growth rate and error correlation values were within 2% of their reference values. Growing analysis error variance estimates deviated less than 5% from their reference values.

Decaying errors were found to diminish rapidly. Hence perceived error measurements are affected only at a few early lead times, leading to larger (up to 50%) uncertainty in decaying error variance and decay rate estimates.

In Section 5, SAFE-II was used to estimate the error variance in operational NCEP analyses and forecasts. U, T and GH perceived error variances simulated by SAFE-II were found to be within the sampling uncertainty of their measurement-based counterparts, indicating that the assumptions behind SAFE-II are consistent with the NCEP operational data. The key findings of this part of the study are as follows:

- The growth rate for the model variables U and T peaks around the upper-level jet in the areas of strongest baroclinic instabilities, with an error variance doubling time of around 23 hr. A weaker maximum appears around the lower-level jet. Error variance doubling time near the surface is around 32 hr. Forecasts for GH exhibit error growth faster than those for U and T at all levels.
- The maximum of total analysis error variance for U and GH are near the upper-level jet (250–300 hPa), consistent with the level of their fastest error growth rate. The maximum of total analysis error variance for T is near the low-level jet (~925 hPa). Interestingly, the total analysis error for the model variables U and T peaks just above the maxima in growth rate. This may be explained by a general decrease in the density of *in situ* observations with increasing altitude.
- Decaying errors constitute up to 40 and 15% of the total analysis error variance for wind and temperature variables in the upper (and for U, also in the lower) atmosphere, respectively. Decaying errors originate from observational noise, and approximations in DA procedures (e.g. improper balance constraints caused by model-related errors near the model top, and lack of proper specification of representativeness error in areas of complex topography). No decaying errors are observed in the non-divergent mid-tropospheric region. This may be related to the quasi-nondivergent nature of dynamics at these layers where balancing the analysis variables is simpler and more straightforward.
- GH has a higher (50–60%) proportion of decaying errors than the model variables. This may be due to the accumulation of independent noise from the model variables as GH is derived from them.

7 | DISCUSSION

Due to the limited number of short-lead-time perceived error measurements influenced by decaying errors, the

uncertainties in decaying parameter estimates are much higher than those in the estimates of the other parameters. Efficient approaches to constrain the estimates of decaying parameters and assess the uncertainty in such estimates need to be pursued further. The power-law relationship of the error decorrelation (i.e. *Assumption 3*) may also need to be refined as decaying errors may not exhibit the same exponential-like decorrelation behaviour as the growing errors.

Possible future applications of SAFE-II may also include grid-pointwise estimation of analysis error variance. Geographical localization of excessive noise in analysis fields (e.g. due to a lack of physical or dynamical balance) may aid in the diagnosis and correction of weaknesses in DA techniques. Solid estimates of analysis uncertainty may also benefit ensemble initialization techniques.

A recurring observation in this study is that the commonly used perceived error variance gives a rather poor estimate of the true error variance (e.g. 3–4-fold underestimation at 6 hr lead time) and a related overestimation of the error growth rate within the first 2 days due to the neglect of (a) analysis errors, and (b) the correlation between error fields in the analyses and forecasts. The use of perceived error as an estimator of true error can have significant consequences in a number of areas:

Data Assimilation. In DA, background error variances will be underestimated. As DA performance depends only on the ratio (but not the absolute value) of errors in the background field vs. the observations, the tuning of DA schemes may lead to an underestimation of observational (including representativeness) errors as well.

Observing System Simulation Experiments. If true error variance is assumed to be as low as perceived error variance measured in operational forecast systems, OSSE systems may be tuned to exhibit too low true error variance. This problem may be evidenced in NOAA's OSSE system (cf. column 7 in Tables 2 and 4).

Predictability. When the growth of perturbations such as lagged forecast differences (LFD) is compared with the growth of perceived error, the latter, since at short lead times perceived errors have unrealistically low values, appears to be significantly faster than the former. This situation has been widely interpreted in the literature as a sign that external predictability is shorter than internal predictability (i.e. the divergence of trajectories in nature is faster than in its numerical models, e.g. Simmons *et al.*, 1995). True error, however, amplifies much slower than perceived error (see, e.g., Figures 3 and 7), possibly eliminating the need for such hypothetical explanations. Implications may include a longer than currently thought limit on predictability.

Ensemble Forecasting. If the size of initial perturbations is set so that 6 hr ensemble variance matches 6 hr

perceived error variance, the ensemble, though it may appear reasonable when its spread is checked against perceived error, actually will start out under-dispersive. Irrespective of initial perturbation generation methods, the under-dispersiveness readily manifests itself at later lead times (e.g. Buizza *et al.*, 2005), however, when analysis error variance becomes negligible compared to forecast error variance. Conventionally, the situation is explained as insufficient perturbation growth due to model imperfection presumed related to the numerical models being more predictable than the atmosphere (i.e. too high internal predictability). The notion and an array of stochastic model perturbation methods (Buizza *et al.*, 1999; Shutts, 2005) have been proposed to hasten perturbation growth with the aim of remedying a problem that may not exist. Future studies can further explore the validity of the Statistical Analysis and Forecast Error (SAFE) estimation-based interpretations advanced above.

ACKNOWLEDGEMENTS

Lidia Cucurull, Ruifang Li and Tanya Peevey kindly provided the Observational System Simulation Experiment data and the corresponding references. Discussions with Krishna Kumar (NCEP), Jordan Alpert (NCEP), Fanglin Yang (NCEP), Si Shen (NCAR), and Roman Krzysztofowicz (University of Virginia) are gratefully acknowledged. We acknowledge the encouragement and support of Kevin Kelleher, former Director of GSD.

ORCID

Jie Feng  <https://orcid.org/0000-0002-2480-2003>

REFERENCES

- Andersson, E. and Matsutani, M. (2010) *Collaboration on observing system simulation experiments (joint OSSE)*. *ECMWF Newsletter*, 123, 14–16.
- Atlas, R. (1997) Atmospheric observations and experiments to assess their usefulness in data assimilation. *Journal of the Meteorological Society of Japan*, 75, 1–20.
- Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Zhu, Y.J. and Wei, M.Z. (2005) A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133, 1067–1097.
- Buizza, R., Miller, M. and Palmer, T.N. (1999) Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 125, 2887–2908.
- Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16(5), 1190–1208.
- Cucurull, L., Li, R. and Peevey, T.R. (2017) Assessment of radio occultation observations from the COSMIC-2 mission with a simplified observing system simulation experiment configuration. *Monthly Weather Review*, 145, 3581–3597.
- Dalcher, A., Kalnay, E. and Hoffman, R.N. (1988) Medium range lagged average forecasts. *Monthly Weather Review*, 116, 402–416.
- Dee, D. (2005) Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3323–3343.
- Ding, R.Q., Li, J. and Li, B.S. (2017) Determining the spectrum of the nonlinear local Lyapunov exponents in a multidimensional chaotic system. *Advances in Atmospheric Sciences*, 34(9), 1027–1034.
- Ding, R.Q. and Li, J.P. (2007) Nonlinear finite-time Lyapunov exponent and predictability. *Physics Letters*, 364A, 396–400.
- Feng, J., Ding, R.Q., Liu, D.Q. and Li, J.P. (2014) The application of nonlinear local Lyapunov vectors to ensemble predictions in the Lorenz systems. *Journal of the Atmospheric Sciences*, 71(9), 3554–3567.
- Feng, J., Li, J., Ding, R. and Toth, Z. (2018) Comparison of nonlinear local Lyapunov vectors and bred vectors in estimating the spatial distribution of error growth. *Journal of the Atmospheric Sciences*, 75(4), 1073–1087.
- Feng, J., Toth, Z. and Peña, M. (2017) Spatially extended estimates of analysis and short-range forecast error variances. *Tellus*, 69A, 1325301. <https://doi.org/10.1080/16000870.2017.1325301>.
- Fisher, M. (1996) The specification of background error variances in the ECMWF variational analysis system. In: *Proceedings ECMWF Seminar on Data Assimilation, 2–6 September 1996*. Reading: ECMWF, pp. 645–652.
- Gilmour, I., Smith, L.A. and Buizza, R. (2001) Linear regime duration: is 24 hours a long time in synoptic weather forecasting? *Journal of the Atmospheric Sciences*, 58, 3525–3539.
- Grell, G.A., Dudhia, J. and Stauffer, D. (1995) *A description of the fifth-generation PENN State/NCAR Mesoscale Model (MM5)*. NCAR Technical Note NCAR/TN-398+STR, 10 pp.
- Hamill, T., Snyder, M.C. and Morss, R.E. (2002) Analysis-error statistics of a quasigeostrophic model using three-dimensional variational assimilation. *Monthly Weather Review*, 130, 2777–2791.
- Hamill, T.M. and Whitaker, J.S. (2011) What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Monthly Weather Review*, 139, 117–131.
- Houtekamer, P.L. and Mitchell, H.L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129, 123–137.
- Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charon, M., Spacek, L. and Hansen, B. (2005) Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Monthly Weather Review*, 133, 604–620.
- Huang, X.-Y. and Lynch, P. (1993) Diabatic digital-filtering initialization: application to the HIRLAM model. *Monthly Weather Review*, 121, 589–603.
- Hunt, B., Kostelich, E. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230, 112–126.
- Jiménez, P.A. and Dudhia, J. (2012) Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *Journal of Applied Meteorology and Climatology*, 51, 300–316.
- Kalnay, E. (2003) *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press.

- Kleist, D.T., Parrish, D.F., Derber, J.C., Treadon, R., Errico, R.M. and Yang, R. (2009) Improving incremental balance in the GSI 3DVAR analysis system. *Monthly Weather Review*, 137, 1046–1060.
- Legras, B. and Vautard, R. (1996) A guide to Lyapunov vectors. In: Palmer, T (Ed.) *Proceedings of ECMWF Seminar on Predictability*, Vol. 1. Reading: ECMWF, pp. 143–156. [Available from ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom.].
- Li, J.P. and Ding, R.Q. (2011) Temporal–spatial distribution of atmospheric predictability limit by local dynamical analogues. *Monthly Weather Review*, 139, 3265–3283.
- Lorenc, A.C. (2003) The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129, 3183–3203.
- Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.
- Lorenz, E.N. (1982) Atmospheric predictability experiments with a large numerical model. *Tellus*, 34, 505–513.
- Lorenz, E.N. (1996) Predictability: a problem partly solved. In: Palmer, T. and Hagedorn, R. (Eds.) *Proceedings of ECMWF Seminar on Predictability*, Vol. I. Reading: ECMWF, pp. 1–18.
- Masutani, M., Woollen, J.S., Lord, S.J., Kleespies, T.J., Emmitt, G.D., Sun, H.B., Wood, S.A., Greco, S., Terry, J., Treadon, R. and Campana, K.A. (2006) *Observing System Simulation Experiments at NCEP*. College Park, MA: National Centers for Environmental Prediction. Office Note 451.
- Mitchell, H.L., Houtekamer, P.L. and Pellerin, G. (2002) Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly Weather Review*, 130, 2791–2808.
- Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T. (1996) The new ECMWF Ensemble Prediction System: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122, 73–119.
- Nicolis, C., Perdigo, R. and Vannitsem, S. (2009) Dynamics of prediction errors under the combined effect of initial condition and model errors. *Journal of the Atmospheric Sciences*, 66, 766–778.
- Orrell, D., Smith, L., Barkmeijer, J. and Palmer, T.N. (2001) Model error in weather forecasting. *Nonlinear Processes in Geophysics*, 8, 357–371.
- Palatella, L., Carrassi, A. and Trevisan, A. (2013) Lyapunov vectors and assimilation in the unstable subspace: theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46, 254020. <https://doi.org/10.1088/1751-8113/46/25/254020>.
- Peña, M. and Toth, Z. (2014) Estimation of analysis and forecast error variances. *Tellus*, 66A, 21767. <https://doi.org/10.3402/tellusa.v66.21767>.
- Peña, M., Toth, Z. and Wei, M. (2010) Controlling noise in ensemble data assimilation schemes. *Monthly Weather Review*, 138, 1502–1512.
- Pires, C., Vautard, R. and Talagrand, O. (1996) On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, 48A, 96–121.
- Privé, N. and Errico, R.M. (2013) The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A*, 65, 21740. <https://doi.org/10.3402/tellusa.v65i0.21740>.
- Privé, N.C., Errico, R.M. and Tai, K.-S. (2013) The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment. *Journal of Geophysical Research – Atmospheres*, 118, 5332–5346.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. and Morel, S. (2008) Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47, 92–107.
- Shutts, G.J. (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131, 3079–3102.
- Simmons, A.J., Mureau, R. and Petroliagis, T. (1995) Error growth estimates of predictability from the ECMWF forecasting system. *Quarterly Journal of the Royal Meteorological Society*, 121, 1739–1771.
- Snyder, C. and Hamill, T.M. (2003) Leading Lyapunov vectors of a turbulent baroclinic jet in a quasigeostrophic model. *Journal of the Atmospheric Sciences*, 60, 683–688.
- Stewart, L.M., Dance, S.L. and Nichols, N.K. (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus*, 65A, 19546. <https://doi.org/10.3402/tellusa.v65i0.19546>.
- Toth, Z. and Kalnay, E. (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological Society*, 74, 2317–2330.
- Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP: the breeding method. *Monthly Weather Review*, 125, 3297–3318.
- Trevisan, A. and Legnani, R. (1995) Transient error growth and local predictability: a study in the Lorenz system. *Tellus*, 47A, 103–117.
- Trevisan, A. and Uboldi, F. (2004) Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system. *Journal of the Atmospheric Sciences*, 61, 103–113.
- Vannitsem, S. and Nicolis, C. (1994) Predictability experiments on a simplified thermal convection model: the role of spatial scales. *Journal of Geophysical Research*, 99(D5), 10377–10385.
- Vannitsem, S. and Nicolis, C. (1997) Lyapunov vectors and error growth patterns in a T21L3 quasigeostrophic model. *Journal of the Atmospheric Sciences*, 54, 347–361.
- Vannitsem, S. and Toth, Z. (2002) Short-term dynamics of model errors. *Journal of the Atmospheric Sciences*, 59, 2594–2604.
- Wee, T.-K., Kuo, Y.-H., Lee, D.-K., Liu, Z., Wang, W. and Chen, S.-Y. (2012) Two overlooked biases of the Advanced Research WRF (ARW) model in geopotential height and temperature. *Monthly Weather Review*, 140, 3907–3918.
- Wei, M., Toth, Z., Wobus, R. and Zhu, Y. (2008) Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, 60A, 62–79.
- Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y. and Toth, Z. (2008) Ensemble data assimilation with the NCEP Global Forecast System. *Monthly Weather Review*, 136, 463–482.

- Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. (1985) Determining Lyapunov exponents from a time series. *Physica D*, 16, 285–317.
- Yang, F. (2016) Evaluation of hurricane forecast skills of NCEP GFS retrospective experiments for the FY2016 implementation. In: *32nd Conference on Hurricanes and Tropical Meteorology, 2016, San Juan, Puerto Rico*. Available at: <https://ams.confex.com/ams/32Hurr/webprogram/Paper293991.html>.
- Ziehmann, C., Smith, L.A. and Kurths, J. (2000) Localized Lyapunov exponents and the prediction of predictability. *Physics Letters*, 271A, 237–251.

How to cite this article: Feng J, Toth Z, Peña M, Zhang J. Partition of analysis and forecast error variance into growing and decaying components. *QJR Meteorol Soc.* 2020;146:1302–1321. <https://doi.org/10.1002/qj.3738>

APPENDIX A

A | Sampling uncertainty

Just as SAFE-I (Peña and Toth, 2014), SAFE-II estimates the unknown parameters of true analysis and forecast error variance by fitting perceived error variance modelled with the unknown parameters to sample-based measurements of perceived error variance. The expected error in finite sample-based estimates of the expected value of normally distributed variables is given by the Standard Error of the Mean (or Measurement, SEM):

$$SEM_i = sd_i \cdot f / \sqrt{N}, \quad (A1)$$

where sd_i represents the sample standard deviation in the sample at lead time i , N is the sample size, and $f = \sqrt{(1 + r_1)(1 - r_1)^{-1}}$ is an adjustment coefficient accounting for serial correlation (r_1) in the sample.

As the standard deviation of a finite sample-based mean tends to grow with lead time, observed quantities at longer time ranges will need to be given smaller weight in the minimization procedure. The standardized weights w_i in Equation 8 are defined as:

$$w_i = SEM_i / \sum_i SEM_i. \quad (A2)$$

Note that the definition of SEM_i and w_i can be generalized to other finite sample-based estimates of expected

value, like the lagged forecast difference and true forecast error variance *et al.* used in this study.

Since SEM values quantify the uncertainty in sample mean values, they can also be considered as confidence intervals when SAFE estimates are compared with the mean of measurements. Assuming that the finite-sample mean of perceived error variance follows a Gaussian distribution, the 95% confidence interval can be defined by adding and subtracting 1.96 times the SEM_i value to/from the perceived error variance measurements.

B | OSSE set-up

In OSSEs, analyses and forecasts are generated the same way as in an operational NWP system, except the role of real observations is taken by simulated observations. A long integration with a fine-resolution model other than that used in the NWP DA-forecast system is usually considered as truth (or nature), from which simulated observations are generated with the addition of noise meant to represent different sources of observational and representativeness errors (e.g. Atlas, 1997). Since truth is exactly known, when carefully designed, OSSEs offer a unique and fully controlled environment in which to evaluate the quality of NWP techniques.

Nature used in this OSSE system was created by the European Centre for Medium-range Weather Forecasts (ECMWF) operational model version c31r1 at T511 (about 40 km) horizontal and 91-level vertical resolution, with boundary forcing data from 1 May 2005 to 31 May 2006 (Masutani *et al.*, 2006; Andersson and Matsutani, 2010). The NWP modelling (GFS) and DA system (Gridpoint Statistical Interpolation analysis, GSI) are based on an earlier and reduced resolution (T382, about 52 km, and 64-level) version of NCEP's operational suite with a non-hybrid DA scheme. The observations assimilated include conventional, satellite, and COSMIC-2 (“Constellation for Observing System for Meteorology, Ionosphere, and Climate”) data generated from the nature run. Representativeness errors are inherent in the simulated observations due to a difference in resolution between the nature run and the NWP system. No systematic or random errors were otherwise added to nature for the simulated observations, except for satellite radiances. All observations are assimilated using a ± 1 hr window centred at nominal analysis times.