RMetS
Royal Meteorological Society

Linearity aspects of the ensemble of data assimilations technique

L. Megner,[a]* D. G. H. Tan,[b] H. Körnich,[c] L. Isaksen,[b] A. Horányi,[b] A. Stoffelen[d] and G.-J. Marseille[d]

[a]*Department of Meteorology, Stockholm University, Sweden*
[b]*European Centre for Medium-Range Weather Forecasts, Reading, UK*
[c]*Swedish Meteorological and Hydrological Institute, Norrköping, Sweden*
[d]*Royal Netherlands Meteorological Institute, De Bilt, the Netherlands*

*Correspondence to: L. Megner, Department of Meteorology, Stockholm University, Svante Arrhenius Väg 16C, Stockholm 10691, Sweden.
E-mail: linda@misu.su.se

We examine the linearity of the Ensemble of Data Assimilations (EDA) technique with respect to the amplitude of the applied observation perturbations. We provide explicit examples to assess the linear relationship between such modifications of the observing system and the resulting changes in the EDA ensemble spread. The results demonstrate that, for a state-of-the-art numerical weather prediction (NWP) system, such linearity between the applied observation perturbations and the EDA ensemble spread holds well for temporal and spatial regimes relevant to global medium-range weather prediction: specifically, for forecast lead-times of up to approximately 5 days, in the vertical throughout the troposphere up to the lower and middle stratosphere and for broad horizontal scales.

Key Words: data assimilation; impact assessment; ensemble; EDA technique

Received 7 February 2013; Revised 20 February 2014; Accepted 24 February 2014; Published online in Wiley Online Library

1. Introduction

Realizing further advances in global numerical weather prediction depends on making progress across a range of scientific and technical issues. Two prominent issues are (i) evolution of the Global Observing System and (ii) use of ensemble-based techniques, not only for providing uncertainty estimates of 'deterministic forecasts' but also as a component of hybrid data assimilation (Buehner *et al.*, 2010; Barker *et al.*, 2012; Bonavita *et al.*, 2012). Tan *et al.* (2007) estimated the expected impact of the European Space Agency's planned Atmospheric Dynamics Mission (ADM)–Aeolus Doppler Wind Lidar observing system using an ensemble-based approach, specifically an early version of ECMWF's Ensemble of Data Assimilations (hereafter EDA), but did not have the resources to test all of the underlying assumptions. The approach is increasingly being adopted for assessing other potential observing-system modifications (Harnisch *et al.*, 2013), so it is timely to complement the previous work by addressing some of the assumptions not examined to date. By reporting such work in this article, we seek to strengthen the scientific foundations underpinning the use of EDA-related techniques for assessing future observing systems.

This article arose from interest in the suitability of EDA-related approaches for assessing future observing-system modifications. Such assessments, whether EDA-related or not, invariably involve simulating the quality of the future observational data and most require the defining of reference atmospheric states for evaluating the expected impact of the future observations. In this context, the traditional approach has been to conduct Observing System Simulation Experiments (OSSEs), where a 'nature run' serves as both the reference atmospheric state (for verification purposes) and also input to the simulations of the entire observing system (baseline and modified: Stoffelen *et al.*, 2006; Masutani *et al.*, 2010). Use of the EDA approach as an alternative to OSSEs (Tan *et al.*, 2007) was in part motivated by the need for methods that do not incur the computational cost of generating the nature run and associated simulations of the entire baseline set of observations.

A key underlying assumption when applying the EDA technique to assess future observing systems concerns its linearity with respect to the magnitude of the observation perturbations. This assumption has not been tested extensively to date and is thus the focus of the current article. At first glance, it might seem rather straightforward to devise suitable experiments to examine such linearity properties. However, in developing the experimental protocols we encountered a number of subtleties, which we explain in some detail below for the benefit of readers who may wish to use EDA-related approaches in future. At the same time, we retain in our protocols some valuable aspects of the widely accepted approach for validating OSSEs, for example by examining the applicability of the assumption for a well-understood but reduced observing system (here chosen to be the Global Observing System with radiosondes and wind-profilers excluded). Use of a reduced observing system also facilitates examination of calibration issues in that it permits the generation of a quasi-independent reference dataset (here chosen to be the analyses from a separate data assimilation experiment for the full Global Observing System) for the purposes of defining forecast error.

In section 2, we present background aspects of EDA theory needed to understand our experimental set-up and results, which are presented in sections 3 and 4 respectively. Conclusions are given in section 5.

2. Background

2.1. Error, uncertainty characterization and EDA spread

We stress the importance of keeping separate the concepts of analysis/forecast error, uncertainty characterization and ensemble spread. We use the term *analysis/forecast error* to refer to a specific realization of the difference between an analysis/forecast and the (typically unknown) truth, expressed quantitatively for relevant geophysical parameters. From large samples of analyses/forecasts, together with limited information about the truth (i.e. from actual but imperfect observations), the concept of error statistics naturally arises (including quantification of correlated and uncorrelated components). *Uncertainty characterization* encompasses attempts to describe error statistics, in terms of quantitative statistical entities such as moments of statistical distributions (e.g. means, variances, skewness, kurtosis, covariances) but also in terms of qualitative elements such as recognition of the difficulty in faithfully representing geophysical forcings and processes (two examples being specification of sea-surface temperature and subgrid parametrization). Finally, we take the *unscaled ensemble spread* to have its usual statistical meaning, namely the standard deviation with respect to the ensemble mean.

At present, ECMWF's primary use of the EDA technique is to provide flow-dependent initial-state uncertainty estimation for the Ensemble Prediction System (Buizza *et al.*, 2008) and to provide a scaled ensemble spread as a proxy for background-error statistics within the high-resolution deterministic data assimilation system (Isaksen *et al.*, 2010; Bonavita *et al.*, 2012). The scaled ensemble spread is a calibrated transformation of the unscaled ensemble spread as described by Isaksen *et al.* (2010). The unscaled ensemble spread is generated through three distinct processes: (i) perturbations to the boundary forcings and assimilated observations, (ii) propagation of initial-condition spread by the forecast model and (iii) a stochastic physics representation of 'model error'. The initial-condition spread arises from the ensemble analyses and includes the effects of all these processes from previous analysis cycles. In practice, the ensemble typically includes one additional 'control' member in which the perturbations from these processes are not applied. For a detailed mathematical description of EDA, see e.g. Žagar *et al.* (2005) and Tan *et al.* (2007).

Improving the correspondence between EDA unscaled ensemble spread and actual background-error statistics of a high-resolution 'deterministic' data assimilation system remains an active area of research (Bonavita *et al.*, 2012). Aspects currently under investigation include better representations of (i) stochastic physics, (ii) correlated observation error, (iii) correlated model error (iv) sampling error due to small ensemble sizes and (v) increased spatial resolution of the ensemble background fields. A measure of the need for improvements in these representations is provided by the calibration transformation from unscaled to scaled ensemble spread. The ECMWF implementation of this transformation consists of a specification of calibration factors (multiplicative scaling coefficients) based on obtaining a statistically consistent spread–skill relationship. For a perfect ensemble, the transformation would be the identity operation, i.e. all calibration factors would be unity. The calibration currently results in different factors for the Tropics, Northern Hemisphere and Southern Hemisphere Extratropics, respectively. It also incorporates altitude-dependence, reflecting spatial variations in observation coverage and formulation of the stochastic physics (as will be discussed in connection with Figure 1 in section 3.2) and dependence on the geophysical parameter under consideration. Factors of order 1.3 are currently applied; these are a substantial improvement over those applied in earlier implementations (Isaksen *et al.*, 2007) (which were of order 2), reflecting the research progress made in recent years (Bonavita *et al.*, 2011, 2012). Moreover, we anticipate that current developments in EDA will soon reduce the need for any statistical calibration. It is worth noting that a similar type of inflation that treats both sampling and model errors is required for current Ensemble Kalman Filter implementations (Whitaker and Hamill, 2012).

2.2. EDA for assessing future observing systems

Notwithstanding the need for further improvements in the EDA technique, it has been recognized that assessment of future observing systems is a valuable secondary application (Tan *et al.*, 2007; Harnisch *et al.*, 2013). In this context, the addition/withholding of a sizeable component of the observing system translates into a change in the unscaled ensemble spread and, through the associations described in section 2.1, provides some expectations of analysis/forecast error reduction.

The correspondence between changes in unscaled ensemble spread and traditional measures of observing-system impact is not always straightforward or intuitive. It is important to remember that an EDA ensemble is typically constructed to represent uncertainty in the unperturbed control member. In the standard practice for assessing observing-system components with the EDA technique, each observation is perturbed (degraded) by a random noise contribution commensurate with the size of the observation error (as specified by the underlying data assimilation system). Observations from a future observing system, which must be simulated (as is the case in OSSEs), are perturbed in the same way. The standard diagnostic in the EDA approach is the change in the ensemble spread arising from the introduction of the additional observing system (smaller spread being associated with reduced uncertainty and thus beneficial impact).

In such practice, each ensemble member is constructed to be 'sub-optimal' in two senses: (i) with respect to the control member, in the sense that the observations are degraded, and (ii) with respect to the 'optimal' analyses/forecasts that could be achieved with the degraded observations by re-adjusting the assumed observation errors. We ask the reader to keep in mind that these aspects are actually important ingredients for making the EDA approach suitable for observation impact assessment.

The situation is further complicated when changes in scaled ensemble spread are used as a proxy for observing-system impact, because of the additional role of the calibration transformations. Fortunately, the promising research developments mentioned in section 2.1 give us confidence that scaling of unscaled ensemble spread will soon become unnecessary. This places us in a position to examine the properties of unscaled ensemble spread for assessing the linearity properties of EDA with respect to observation perturbation magnitude. However, we remind the reader that, in an EDA system involving a substantial calibration transformation, the magnitude of unscaled ensemble spread will not be comparable to actual forecast errors.

It remains to define the experimental set-up and diagnostics that will permit us to examine the linearity properties of EDA with respect to observation perturbation magnitude.

3. Experimental set-up

3.1. EDA configuration

The EDA configuration used in this study was an instance of the ECMWF Ensemble of 4D-Var Data Assimilations (Isaksen *et al.*, 2010), i.e. using explicit perturbations of the assimilated observations, the sea-surface temperature field and the model physics. (In all other respects, the underlying data assimilation system was kept fixed.) All assimilated observations were randomly perturbed by sampling a Gaussian distribution with zero mean and standard deviation equal to the expected observation error multiplied by an observation perturbation scaling factor P. Observation-error correlations were taken into account for atmospheric motion vectors (Bormann *et al.*, 2003). Boundary-condition fields for sea-surface temperature were also

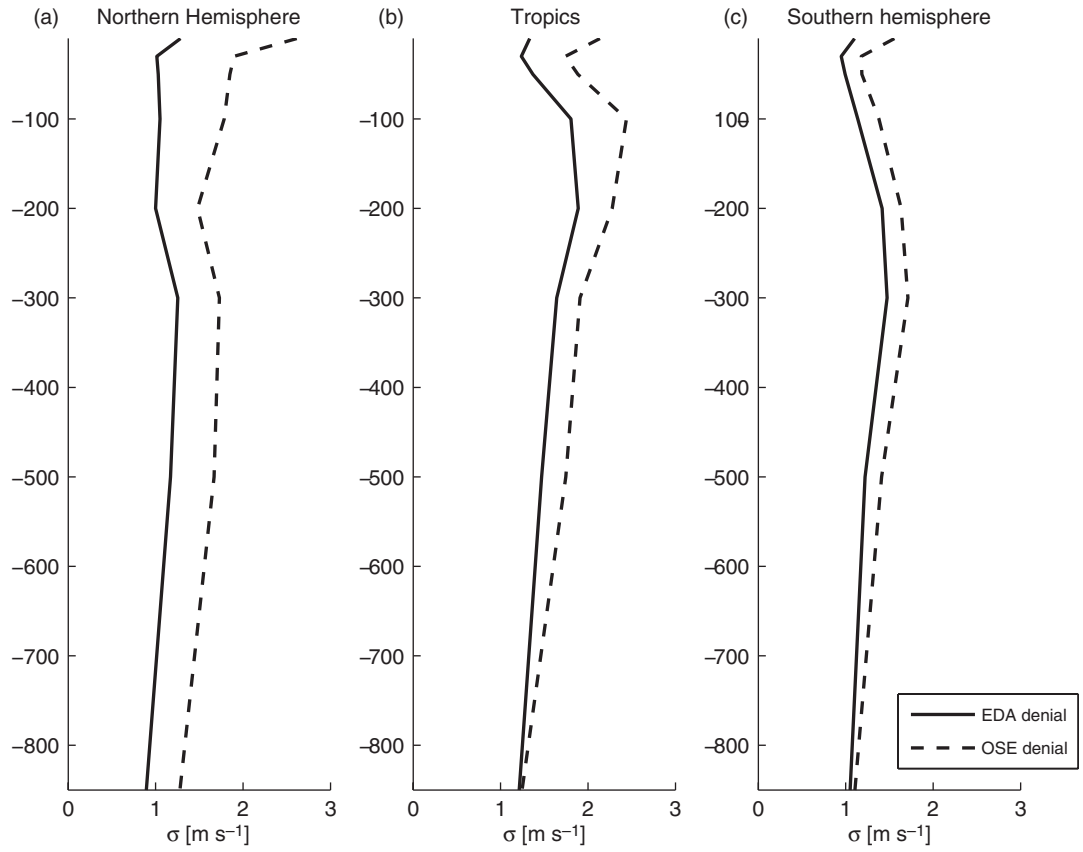

Figure 1. Unscaled ensemble spread $\sigma(P = 1.0)$ (solid curves) and estimated forecast error σ_{OSE} (dashed curves) for the 12 h zonal-wind forecasts. Panel (a) shows the Northern Hemisphere extratropics (90°N–30°N), (b) the Tropics (30°N–30°S) and (c) the Southern Hemisphere extratropics (30°S–90°S).

perturbed with correlated errors using the methodology currently applied in the ECMWF seasonal ensemble forecasting system (Vialard *et al.*, 2005). The forecast model uncertainties were represented by the Stochastic Perturbation of Physics Tendencies (SPPT) scheme (Palmer *et al.*, 2009; Shutts *et al.*, 2011), which perturbs parametrization tendencies at every time step of the model integration. The magnitude of the perturbation of the stochastic physics was reduced above 100 hPa and reached zero at 50 hPa, so that the only explicit perturbations above 50 hPa were those of the observations. The variational bias correction scheme (primarily used for satellite data) was only cycled for the control assimilation and was then used by all perturbed ensemble members. The data assimilation system component of the configuration was Cycle Cy35r2 of the ECMWF Integrated Forecast System (IFS), with analyses and innovation computations performed at *TL*399 spectral truncation (approximate model grid of 50 km spacing). The inner loops, which generate the analysis increments of the multi-incremental 4D-Var (Courtier *et al.*, 1994), had spectral truncations at *TL*95 and *TL*159, respectively. The forecast model component had 91 vertical levels that extended from the surface up to 0.1 hPa (around 65 km). Within the EDA framework, the perturbations to the observations, boundary conditions and forecast model in one data assimilation cycle give rise to an ensemble of analyses for that cycle and their influence propagates further to the ensemble of background fields for subsequent assimilation cycles (see Žagar *et al.*, 2005; Isaksen *et al.*, 2010). The spread of the EDA background-state fields typically reaches a steady-state level after a 3–7 day spin-up period, starting from identical background fields.

In our configuration, a single EDA ensemble consisted of a control (non-perturbed) assimilation and ten perturbed members and was conducted with a fixed value of the observation perturbation scaling factor *P*. Three distinct EDA ensembles were generated, corresponding to observation perturbation scaling factors of $P = 1.0$, 0.75 and 0.5, respectively. The assimilations were performed for 15 days (January 2007), of which the first 6 days are considered a spin-up period and thus excluded from the diagnostics. While the overall duration is sufficient for the current investigation of the broadest spatial scales, a longer duration would assist in examining smaller spatial scales.

3.2. *Linearity diagnostics and ensemble calibration*

The diagnostics examined are the (unscaled) ensemble spread from the three ensembles ($P = 1.0, 0.75, 0.5$), defined for a fixed forecast lead time as

$$\sigma(P) = \left[\frac{1}{K} \sum_{k=1}^{K} \frac{1}{N-1} \sum_{n=1}^{N} (x_n^{\mathrm{f}}(P) - \langle x^{\mathrm{f}}(P) \rangle)_k^2 \right]^{1/2}, \qquad (1)$$

where n ranges over all N ensemble members and k over all K analysis cycles and $\langle x^{\mathrm{f}} \rangle$ denotes the ensemble mean.

Under assumptions of linearity, the corresponding ensemble variances (squared spread) are expressible in the form (see Appendix, Eq. (A20) for details)

$$\sigma^2(1.0) = K_1^2 \sigma_\eta^2 + K_2^2 \sigma_\xi^2 + \text{higher order terms}, \qquad (2)$$

$$\sigma^2(0.75) = 0.75^2 K_1^2 \sigma_\eta^2 + K_2^2 \sigma_\xi^2 + \text{higher order terms} \qquad (3)$$

and

$$\sigma^2(0.5) = 0.5^2 K_1^2 \sigma_\eta^2 + K_2^2 \sigma_\xi^2 + \text{higher order terms}, \qquad (4)$$

where $P^2 K_1^2 \sigma_\eta^2$ corresponds to the variance induced by the observation perturbations and $K_2^2 \sigma_\xi^2$ to the variance induced by stochastic physics and boundary perturbations.

Motivated by Eqs (2) and (4), we define an estimate of $\sigma^2(0.75)$, here denoted by $\hat{\sigma}^2$, in terms of a linear combination of $\sigma^2(1.0)$ and $\sigma^2(0.5)$:

$$\hat{\sigma}^2 = \alpha \sigma^2(0.5) + (1 - \alpha) \sigma^2(1.0), \qquad (5)$$

where α is $(1 - 0.75^2)/(1 - 0.5^2)$. Hence, subject to negligibility of the higher order terms, the validity of the linearity assumption then hinges on whether $\hat{\sigma}$ is a good approximation to $\sigma(0.75)$. Note that the EDA configuration may also be investigated for linearity with respect to the stochastic physics and boundary perturbation inputs, but this is not within the scope of this manuscript, which focuses on observation inputs.

As mentioned in the Introduction, the observing system input examined in our linearity investigations consists of the Global Observing System with radiosondes and wind-profilers excluded. This choice was made to facilitate examination of calibration of ensemble spread for our experimental set-up. To do so, it is necessary to compare unscaled ensemble spread with a corresponding estimate of forecast error. Accordingly, we also conducted one further data assimilation experiment with the full Global Observing System, thereby generating a quasi-independent reference dataset of 'baseline analyses' x^{ba}. This allows us to estimate the forecast error of the control member of the $P = 1.0$ ensemble and compare it with the usual metric for an observing-system experiment involving denial of particular observations:

$$\sigma_{\mathrm{OSE}} = \left[\frac{1}{K} \sum_{k=1}^{K} [(x^{\mathrm{f}} - x^{\mathrm{ba}})_k - \overline{(x^{\mathrm{f}} - x^{\mathrm{ba}})}]^2 \right]^{1/2}. \qquad (6)$$

The use of a maximally independent baseline experiment as verification reference dataset was recently recommended by a WMO expert team in preference to the alternative technique of self-verification (using the analysis of each experiment: Andersson and Sato, 2012). We do our best to adopt this recommendation by verifying the forecasts x^{f} of the $P = 1.0$ 'radiosonde-denial' system against the analyses x^{ba} from the baseline 'full-observing' system.

Before proceeding to the next section, where we examine the validity regime of the EDA technique's linearity assumption with respect to observation perturbation magnitude, we therefore assess the significance of the calibration transformation that would make unscaled ensemble spread comparable to forecast error and compare these two quantities.

The solid and dashed curves in Figure 1 show, respectively, the unscaled ensemble spread $\sigma(1.0)$ and the control-member forecast-error estimate σ_{OSE} (both for a forecast lead time of 12 h and for the zonal-wind component). It is apparent that the calibration factors required to transform unscaled ensemble spread to forecast-error estimate are fairly constant for the Northern Hemisphere, apart from an increase above 100 hPa. There is a more pronounced dependence on altitude in the Tropics and Southern Hemisphere. Overall, the calibration factor is of order 1.5, a value that is in keeping with the trend from higher values in the past (of order 2 in 2006, see section 2.1) and those currently applied operationally (of order 1.3).

In Figure 2 we compare the geographical distribution of the unscaled ensemble spread $\sigma(1.0)$ (first column) and the control-member forecast-error estimate σ_{OSE} (second column). In order to simplify the visual comparison of the right- and left-hand panels, varying colour scales have been used. In general, the two approaches agree rather well on the spatial distribution of the forecast quality, at least in the troposphere and up to the lower stratosphere. However, at 10 hPa there are clearly spatial differences between $\sigma(1.0)$ and σ_{OSE}, with σ_{OSE} being much larger at northern latitudes. At the same time, Figure 1 showed that the calibration factor increases at these altitudes, meaning that the magnitudes of the two quantities differ. There could be many reasons for these differences, including the following.

(1) The stochastic physics based perturbations used in the EDA ensemble are reduced at higher altitudes, as described in section 3.
(2) The observation errors for radiosondes in the stratosphere might be underestimated.
(3) The spatially uncorrelated perturbations give insufficient increments at larger scales. This would mean that the large-scale Rossby waves that propagate to the stratosphere are less perturbed than they should be, which would lead to insufficient spread in the stratosphere. Since the easterlies in the summer hemisphere do not allow planetary waves to propagate into the stratosphere, this would explain the observed asymmetry between the Northern/winter and Southern/summer Hemispheres.
(4) The large-scale planetary wave perturbations may need a longer forecast time to spin up, as they are not baroclinically unstable and not as quickly triggered by the observation perturbation as the shorter baroclinic waves.

These differences show that, for a current state-of-the-art NWP system, calibration factors that vary with both altitude and geographical location are needed to interpret the EDA spread as a forecast-error estimate. However, as discussed above, ensemble spread is intimately connected to the stochastic representation of forecast uncertainty, a research area that is fast progressing. Thus we can expect these differences to decrease as spatial resolution and general skills in perturbation modelling progress.

We have already noted that, in an EDA system involving a substantial calibration transformation, the magnitude of unscaled ensemble spread will not be comparable to actual forecast errors. Nevertheless, the linearity properties of interest in this article are independent of the precise values of the calibration factors, and so we present the linearity diagnostics in terms of unscaled ensemble spread.

4. Linearity results

The goal of this section is to help the reader to assess the validity regime of the EDA technique's linearity assumption with respect to observation perturbation magnitude. Figures 2–4 show a range of characteristics of unscaled EDA spread. Taken together, the figures show in condensed form the linearity properties of EDA spread in the time dimension (as measured by forecast 'length' or 'lead time') and also the spatial dimensions (vertical profiles and horizontal maps at a range of pressure levels).

The solid curves in Figure 3 show the 500 hPa global average of the EDA spread as a function of forecast length (lead time) for the three different values of P. Linearity can be assessed by comparing the curve for $P = 0.75$ with the approximation $\hat{\sigma}$ (dashed curve). The linear relationship gradually degrades as the forecast length is increased, but holds reasonably well up to forecast lengths of approximately 5 days. For 10 day forecasts, the two ensembles with larger perturbations ($1 \times \eta$ and $0.75 \times \eta$) both show the same spread. This degradation of the linear relationship when reaching longer forecasts is to be expected, since the impact of observations on the spread of the system is approaching saturation. One could consider it surprising that linearity holds for forecast lengths as large as 5 days.

For insight into the vertical dependence, Figure 4 shows how the global average of the EDA spread varies with altitude (pressure) at a forecast length of 12 h. As expected, the EDA spread is reduced with decreasing perturbations at all altitudes. The dashed black line represents the approximation $\hat{\sigma}$. As seen, this is very close to $\sigma(P = 0.75)$ throughout the domain. In terms of global average, the linearity relationship thus appears to hold reasonably well in both the troposphere and the stratosphere.

For the horizontal dependence, Figure 2 shows the spatial distribution of EDA spread for the experiment with full perturbations ($P = 1.0$, second column). The third column shows the linear approximation $\hat{\sigma}$, which is visually indistinguishable from $\sigma(P = 0.75)$ and so the latter is not shown. Instead, the spatial correlations between $\hat{\sigma}$ and $\sigma(P = 0.75)$ are given as the value C in the figure and are consistently high (above 0.99) at all altitudes. Further, we show in the fourth column that the relative difference between $\hat{\sigma}$ and $\sigma(P = 0.75)$ shows agreement to within 10% (note the different colour scale for relative difference in the

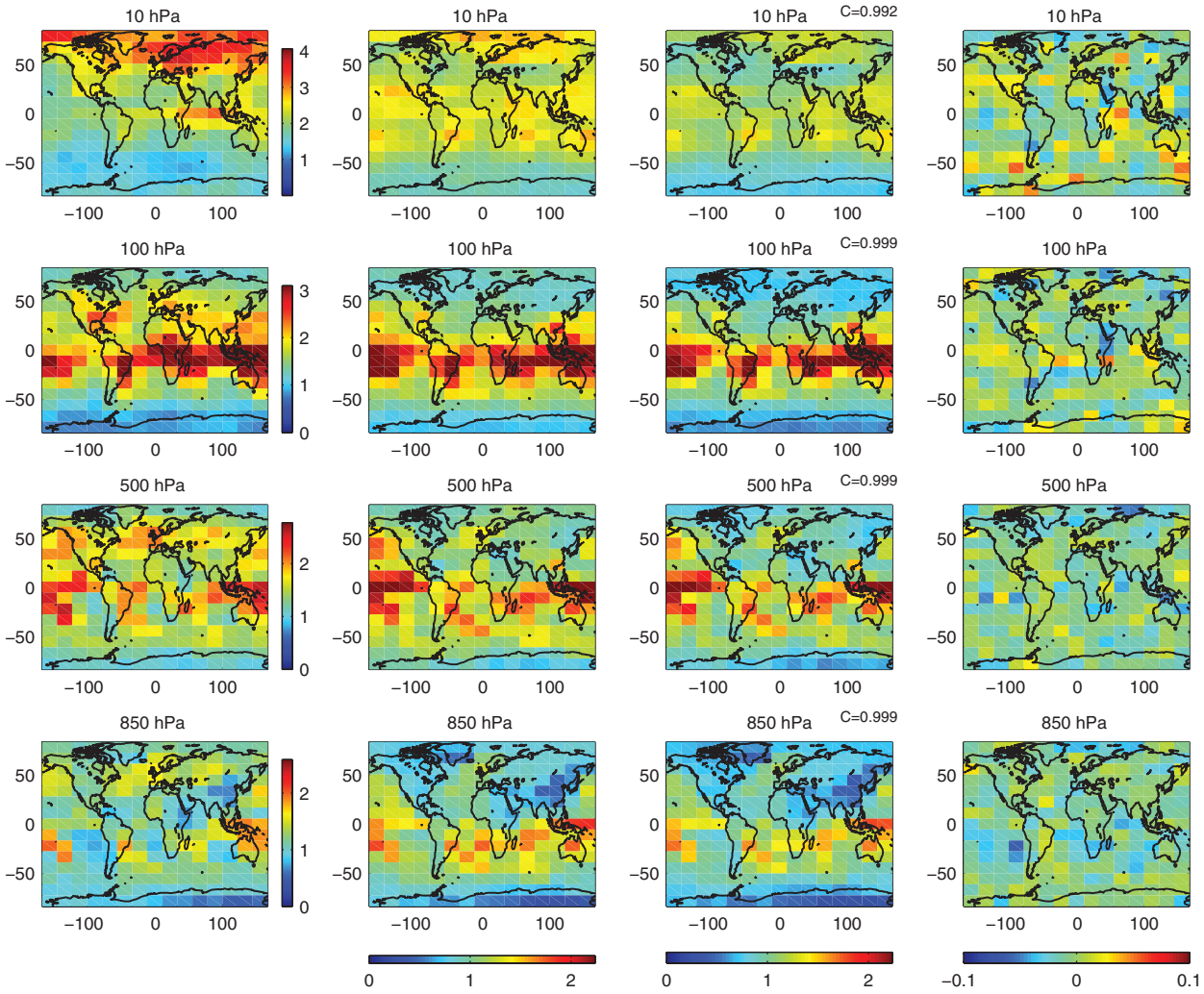

Figure 2. The 12 h EDA spread in zonal wind. From left to right, the columns show estimated control-member forecast error σ_{OSE}, unscaled ensemble spread $\sigma(P = 1.0)$, linearity approximation $\hat{\sigma}$ and the relative difference measure $[\hat{\sigma} - \sigma(P = 0.75)]/\sigma(P = 0.75)$. From top to bottom, the rows show 10, 100, 500 and 850 hPa. It is clear that this difference is small and mostly consists of random noise (note the different colour scale used for the rightmost column).

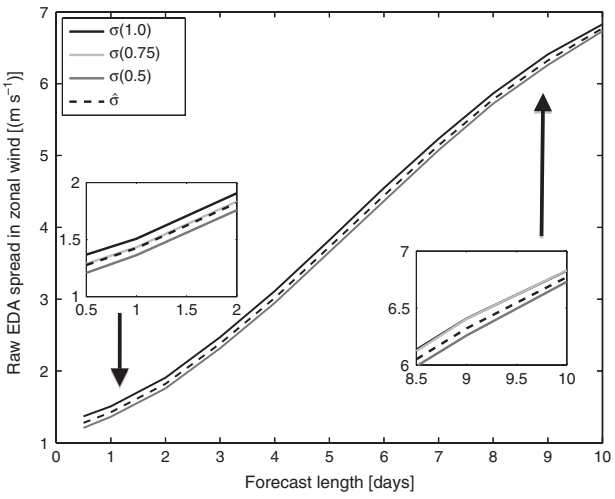

Figure 3. The colour-coded lines show the 500 hPa global-average EDA spread obtained from the experiments with $P = 1$, 0.75 and 0.5 times the original perturbation of the observations as functions of forecast length. The dashed black line shows the approximation $\hat{\sigma}$. The fact that this line almost completely overlaps with $\sigma(0.75)$ shows that there is indeed a linear relationship between the observation perturbation amplitude and the resulting EDA spread.

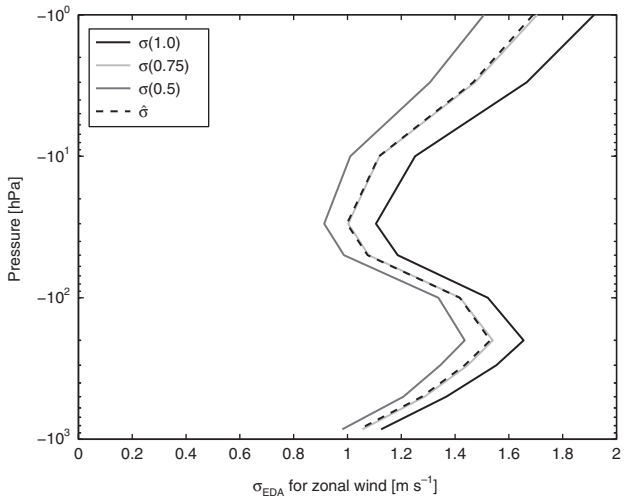

Figure 4. The colour-coded lines show the EDA spread in the 12 h forecast obtained from the experiments with $P = 1$, 0.75 and 0.5 times the original perturbation of the observations. The dashed black line shows the approximation $\hat{\sigma}$.

fourth column) and a regional dependence, comprising essentially random fluctuations. Thus the linear relationship between the observation perturbation magnitude and the EDA spread appears to hold not only in the global average, as shown in Figure 4, but also in every part of the globe.

5. Conclusions

We have studied the ability of the EDA approach to estimate forecast errors and examined the linear relationship between the magnitude of the observation perturbations and the resulting

ensemble spread. In order to check the assumed linearity between the observation perturbation magnitude and the resulting EDA spread, we compared three experiments: one following standard practice for determining the observation perturbations, selected to represent observation and model errors realistically, one where the perturbations were reduced to three-quarters of the standard perturbations and one where the perturbations were reduced to half of the standard perturbations. We find an almost perfect linear relationship between the perturbation magnitude and the resulting EDA spread. This relationship remains valid not only for the global average but also for broad spatial scales and in the time domain up to day 5. The linear relationship gradually degrades as the forecast length is increased beyond 5 days and is fully broken down in 10 day forecasts. A degradation of the linear relationship for longer forecasts is indeed to be expected, as the spread of the system approaches saturation.

It is emphasized that the underlying data assimilation system and forecast model used in this study behave as a linear system for input observation perturbations, without necessarily implying that the system is linear with respect to other changes, or optimal. For example, we consider it an open question whether EDA spread is linear for other system changes such as modifying assumed observation errors.

By demonstrating the fidelity of the linear approximation on 3–5 day time-scales, our results support the use of EDA for the assessment of observing-system changes in the context of medium-range weather forecasts. Such support will be strengthened further by improvements in EDA implementations, particularly those that reduce the need to invoke non-unity calibration factors in the transformation from unscaled to scaled ensemble spread.

Acknowledgements

This work was funded by ESA contract 20940 and 104080 as well as by the Swedish Space Board under contract Dnr 111/07. The authors thank Nedjeljka Žagar, Anne-Grete Straume, Massimo Bonavita and Harald Schyberg for illuminating discussions about ensembles of data assimilations and observing-system assessment. The comments from two anonymous referees helped us improve this article.

Appendix

This Appendix provides a short summary of the mathematical motivation behind EDA adapted from Žagar *et al.* (2005) and explains the theoretical reasoning behind Eqs (2)–(4).

The assimilation system, denoted by f, uses the background state x^{b} and observations y to produce an analysis

$$x^{\mathrm{a}} = f(x^{\mathrm{b}}, y), \tag{A1}$$

which is used as initial state for a forecast model M that produces the forecast x^{f} for time T:

$$x^{\mathrm{f}}(T) = M_T(x^{\mathrm{a}}). \tag{A2}$$

Denoting the true state x^{t} and the true values of the observed quantities y^{t}, we can define the observation, background and analysis errors as

$$\epsilon^{\mathrm{o}} = y - y^{\mathrm{t}}, \tag{A3}$$
$$\epsilon^{\mathrm{b}} = x^{\mathrm{b}} - x^{\mathrm{t}}, \tag{A4}$$
$$\epsilon^{\mathrm{a}} = x^{\mathrm{a}} - x^{\mathrm{t}}. \tag{A5}$$

We now Taylor-expand Eq. (A1) to obtain the error of the analysis:

$$\epsilon^{\mathrm{a}} = \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial x^{\mathrm{b}}}\epsilon^{\mathrm{b}} + \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial y}\epsilon^{\mathrm{o}} + \mathcal{O}(\epsilon^2). \tag{A6}$$

Similarly we write the error of the forecast as

$$\epsilon^{\mathrm{f}}(T) = \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\epsilon^{\mathrm{a}} + \epsilon^{\mathrm{m}}(T) + \mathcal{O}(\epsilon^2), \tag{A7}$$

where $\epsilon^{\mathrm{m}}(T)$ is the model error. Let us now consider an analysis with random perturbations ζ and η to the background and observations, respectively:

$$\hat{x}^{\mathrm{a}} = f(x^{\mathrm{b}} + \zeta, y + \eta). \tag{A8}$$

The error of this analysis is

$$\hat{\epsilon^{\mathrm{a}}} = \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial x^{\mathrm{b}}}(\epsilon^{\mathrm{b}} + \zeta) + \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial y}(\epsilon^{\mathrm{o}} + \eta) + \mathcal{O}(\epsilon^2). \tag{A9}$$

The difference $\delta\hat{x}^{\mathrm{a}}$ between two analyses made with perturbations (ζ_1, η_1) and (ζ_2, η_2) is

$$\delta\hat{x}^{\mathrm{a}} = \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial x^{\mathrm{b}}}\delta\zeta + \frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial y}\delta\eta + \mathcal{O}(\epsilon^2), \tag{A10}$$

where $\delta\zeta = \zeta_2 - \zeta_1$ and $\delta\eta = \eta_2 - \eta_1$.

The forecast obtained with initial conditions given by the analysis $\hat{x^{\mathrm{a}}}$ and with an additional perturbation ξ to the physics in the model is

$$\hat{x^{\mathrm{f}}} = M_T(\hat{x^{\mathrm{a}}}) + \xi. \tag{A11}$$

The error of this forecast is

$$\hat{\epsilon^{\mathrm{f}}}(T) = \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\hat{\epsilon^{\mathrm{a}}} + \xi + \epsilon^{\mathrm{m}}(T) + \mathcal{O}(\epsilon^2). \tag{A12}$$

The difference between two forecasts, from independently perturbed analyses, is thus

$$\delta\hat{x^{\mathrm{f}}}(T) = \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\delta\hat{x^{\mathrm{a}}} + \delta\xi + \mathcal{O}(\epsilon^2). \tag{A13}$$

By inserting Eq. (A10) into Eq. (A13), we obtain

$$\begin{aligned}\delta\hat{x^{\mathrm{f}}}(T) =& \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial x^{\mathrm{b}}}\delta\zeta \\ &+ \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial y}\delta\eta + \delta\xi + \mathcal{O}(\epsilon^2).\end{aligned} \tag{A14}$$

The background field is not perturbed explicitly (see section 3), but becomes perturbed as perturbations in the model and observations propagate to the background. Thus, the perturbation of the background, $\delta\zeta$, is a function (g) of $\delta\xi$ and $\delta\eta$:

$$\begin{aligned}\delta\hat{x^{\mathrm{f}}}(T) =& \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial x^{\mathrm{b}}}g(\delta\eta, \delta\xi) \\ &+ \frac{\partial M_T(x^{\mathrm{t}})}{\partial x}\frac{\partial f(x^{\mathrm{t}}, y^{\mathrm{t}})}{\partial y}\delta\eta + \delta\xi + \mathcal{O}(\epsilon^2).\end{aligned} \tag{A15}$$

If there is a linear relationship between the perturbations $\delta\eta$ and $\delta\xi$ and the EDA spread, the functions f, M_T and g must behave sufficiently linearly. If this assumption is true, we can simplify Eq. (A15) to

$$\delta\hat{x^{\mathrm{f}}}(T) = \kappa_1\delta\eta + \kappa_2\delta\xi, \tag{A16}$$

where κ_1 and κ_2 are linear operators and terms $\mathcal{O}(\epsilon^2)$ have been omitted.

If the perturbations represent the observation and model errors fully, the vector $(\delta\hat{x^{\mathrm{a}}}, \delta\xi)^{\mathrm{T}}$ has a covariance matrix equal to twice that of $(\epsilon^{\mathrm{a}}, \epsilon^{\mathrm{m}}(T))^{\mathrm{T}}$ (Žagar *et al.*, 2005). Comparison between

Eqs (A7) and (A13) shows that the covariance matrix of $\delta\hat{x^{\mathrm{f}}}(T)$ is twice that of $\epsilon^{\mathrm{f}}(T)$. In other words, the variance of the differences between differently perturbed forecasts initialized with differently perturbed analyses can be written as

$$\sigma^2_{\delta\hat{x^{\mathrm{f}}}} = 2 \times \sigma^2_{\epsilon^{\mathrm{f}}}, \tag{A17}$$

where $\sigma^2_{\epsilon^{\mathrm{f}}}$ is the variance of the unperturbed forecast initialized with an unperturbed analysis. Given that the variance of the difference between random variables is twice the variance of the random variables in question, the variance of the EDA ensemble (for clarity reasons we use the notation σ^2_{EDA} here in the Appendix, although the EDA ensemble variance is denoted simply by σ^2 in the main part of the work) becomes

$$\sigma^2_{\mathrm{EDA}} = \sigma^2_{\epsilon^{\mathrm{f}}}. \tag{A18}$$

This means that the ensemble spread of the perturbed forecasts gives an estimate of the forecast error of the unperturbed assimilation system. From Eqs (A17) and (A18), the variance of $\delta\hat{x^{\mathrm{f}}}(T)$ is $2 \times \sigma^2_{\mathrm{EDA}}$. Thus,

$$2\sigma^2_{\mathrm{EDA}} = \sigma^2_{\delta\hat{x^{\mathrm{f}}}} = \kappa_1^2\sigma_\eta^2 + \kappa_2^2\sigma_\xi^2. \tag{A19}$$

In order to simplify the calculations, we set $\kappa_1^2 = 2K_1^2$ and $\kappa_2^2 = 2K_2^2$, where K_1 and K_2 are also linear operators, yielding

$$\sigma^2_{\mathrm{EDA}} = K_1^2\sigma_\eta^2 + K_2^2\sigma_\xi^2, \tag{A20}$$

which can be compared with Eqs (2)–(4).

References

Andersson E, Sato Y. 2012. 'Outcome and recommendations of the WMO workshop on the impact of various observing systems on NWP'. In *Proceedings of the Fifth WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO Technical Report 1/2012, Sedona, AZ. http://www.wmo.int/pages/prog/www/OSY/Reports/NWP-5_Sedona2012.html (accessed 1 October 2012) .

Barker D, Huang XY, Liu Z, Auligné T, Zhang X, Rugg S, Ajjaji R, Bourgeois A, Bray J, Chen Y, Demirtas M, Guo YR, Henderson T, Huang W, Lin HC, Michalakes J, Rizvi S, Zhang X. 2012. The weather research and forecasting (WRF) models community variational/ensemble data assimilation system: WRFDA. *Bull. Am. Meteorol. Soc.* **93**: 831–843.

Bonavita M, Raynaud L, Isaksen L. 2011. Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system: Some effects of ensemble size and day-to-day variability. *Q. J. R. Meteorol. Soc.* **37**: 423–434.

Bonavita M, Isaksen L, Hólm E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.* **138**: 1540–1559.

Bormann N, Saarinen S, Kelly G, Thépaut JN. 2003. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Weather Rev.* **31**: 706–718.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP Part II. *Mon. Weather Rev.* **138**: 1567–1586.

Buizza R, Leutbecher M, Isaksen L. 2008. Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 2051–2066, doi: 10.1002/qj.346.

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

Harnisch F, Healy SB, Bauer P, English SJ. 2013. Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations. *Mon. Weather Rev.* **141**: 4395–4413.

Isaksen L, Fisher M, Berner J. 2007. Use of analysis ensembles in estimating flow-dependent background error variance. In *Proceedings of the ECMWF Workshop on Flow-dependent Aspects of Data Assimilation*, Technical Report, list/14092007. ECMWF: Reading, UK. http://www.ecmwf.int/en/research/publications (accessed 14 September 2007).

Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. *Ensemble of data assimilations at ECMWF*, *Technical Memorandum* 636. ECMWF: Reading, UK. http://www.ecmwf.int/publications/ (accessed 22 February 2011) .

Masutani M, Woollen JS, Lord SJ, Emmitt GD, Kleespies TJ, Wood SA, Greco S, Sun H, Terry J, Kapoor V, Treadon R, Campana KA. 2010. Observing system simulation experiments at the National Centers for Environmental Prediction. *J. Geophys. Res.* **115**: D07101, doi: 10.1029/2009JD012528.

Palmer T, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts G, Steinheimer M, Weisheimer A. 2009. *Stochastic parametrization and model uncertainty*, *Technical Memorandum* 598. ECMWF: Reading, UK. http://www.ecmwf.int/publications/ (accessed 15 October 2009).

Shutts G, Leutbecher M, Weisheimer A, Stockdale T, Isaksen L, Bonavita M. 2011. Representing model uncertainty: Stochastic parametrizations at ECMWF. *ECMWF Newsl.* **129**: 19–24.

Stoffelen A, Marseille G, Bouttier F, Vasiljevic D, de Haan S, Cardinali C. 2006. ADM–Aeolus Doppler wind lidar observing system simulation experiment. *Q. J. R. Meteorol. Soc.* **132**: 1927–1947, doi: 10.1256/qj.05.83.

Tan DGH, Andersson E, Fisher M, Isaksen L. 2007. Observing-system impact assessment using a data assimilation ensemble technique: Application to the ADM–Aeolus wind profiling mission. *Q. J. R. Meteorol. Soc.* **133**: 381–390, doi: 10.1002/qj.43.

Vialard J, Vitart F, Balmaseda MA, Stockdale TN, Anderson DLT. 2005. An ensemble generation method for seasonal forecasting with an ocean–atmosphere coupled model. *Mon. Weather Rev.* **133**: 441–453.

Whitaker JS, Hamill TM. 2012. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev.* **140**: 3078–3089.

Žagar N, Andersson E, Fisher M. 2005. Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *Q. J. R. Meteorol. Soc.* **131**: 987–1011, doi: 10.1256/qj.04.54.

