Ensemble forecasting and flow-dependent estimates of initial uncertainty

Martin Leutbecher, Roberto Buizza and Lars Isaksen

ECMWF, Shinfield Park, Reading
RG2 9AX, United Kingdom
M.Leutbecher@ecmwf.int

ABSTRACT

The initialisation of ensemble forecasts requires a representation of initial uncertainties. Whether data assimilation techniques can provide accurate estimates of initial uncertainty depends on the adequacy of the representation of all relevant sources of uncertainty in the assimilation process.

The first part of this study examines idealised assimilation and ensemble forecasting experiments with the Lorenz-95 system. In this highly idealised framework, it is possible to generate accurate, flow-dependent estimates of initial uncertainty. Sampling the distribution associated with the estimate of initial uncertainty leads then to a perfect ensemble prediction system. Then, the sensitivity of the ensemble forecast skill to the specification of the initial uncertainty is quantified. This is done through a series of ensemble forecast experiments using deliberately degraded estimates of initial uncertainty. The results show that the ensemble forecast skill benefits mostly in the early forecast ranges up to 48 h from an accurate flow-dependent estimate of initial uncertainty.

The second part of this study looks at the ECMWF Ensemble Prediction System (EPS). The reliability of the ensemble standard deviation as predictor for the expected ensemble mean RMS error is diagnosed as function of the forecast lead time for the operational EPS. Then, EPS experiments utilising initial perturbations from an ensemble of 4D-Var assimilations with perturbed observations are evaluated. Results based on 20 cases indicate that ensemble forecasts using only initial perturbations from the 4D-Var ensemble are significantly underdispersive. An EPS configuration which combines initial perturbations from an ensemble of analyses with perturbations from initial singular vectors appears to be best. It leads to a slight improvement of probabilistic skill in the extra-tropics and a significant improvement in the tropics compared to the operational configuration which uses initial and evolved singular vectors.

1 Introduction

Ensemble forecasting was foreseen about four decades ago as the only practical way to get an estimate of the probability density function (pdf) of the atmospheric state at a future time (Lewis 2005). The first operational ensemble prediction systems were established 15 years ago. There are several competing techniques for obtaining initial conditions from which the ensemble forecasts are started (Leutbecher and Palmer 2007). Ensemble forecasts consist of multiple integrations of numerical weather prediction models starting from slightly different initial conditions which are in the vicinity of a best estimate of the true atmospheric state.

If the pdf of the atmospheric state at initial time was known, one could simply draw a random sample of initial states from the initial pdf p_0 in order to obtain an ensemble of initial conditions. The initial pdf p_0 is expected to be strongly flow-dependent, i.e. it will vary from day-to-day. In the real world, it is quite difficult to obtain accurate estimates of p_0 because the sources of uncertainty cannot be quantified well enough. This encompasses for instance correlated observation errors or errors due to imperfections of the numerical weather

prediction model that affect p_0 through the use of a model generated prior, the background state. Therefore, operational ensemble prediction systems can only hope to approximately sample p_0.

This raises the question how important accurate estimates of p_0 are in order to generate skilful ensemble forecasts. In Section 2, we will investigate this question in the context of an idealised low dimensional system, the Lorenz-95 system. In this context, the sources of uncertainty can be accurately quantified. An extended Kalman filter will be used as data assimilation system. Thus, p_0 will be approximated by a Gaussian distribution

$$p_0^{\mathrm{KF}} = \mathrm{N}(\mathbf{x}^a, \mathbf{P}^a), \tag{1}$$

The mean $\mathbf{x}^a$ and covariance $\mathbf{P}^a$ are predicted by the Kalman filter. The benchmark ensemble prediction system is then defined as a 100-member ensemble with initial conditions sampled from p_0^{KF}. The benchmark ensemble prediction system is compared with several other systems which use a degraded estimate of p_0. This section will provide some guidance concerning the improvements in ensemble skill which could be expected from using accurate flow-dependent estimates of initial uncertainty in an ensemble prediction system.

In Section 3, we will turn to the operational ECWMF Ensemble Prediction System (EPS) and investigate to what extent the dispersion of the ensemble reliably predicts flow-dependent variations in uncertainty. Finally, some preliminary results concerning the benefit to the EPS of using initial conditions from an ensemble of data assimilations will be discussed in Section 4. Discussion and conclusions follow in Sections 5 and 6, respectively.

2 Some experiments with the Lorenz-95 system

The intention of this part, is to provide rough guidance on the importance of estimates of the initial pdf p_0 for ensemble prediction. We will focus on the perfect model scenario. Both the system and the forecast model are described by the following system of ordinary differential equations

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \tag{2}$$

where $i = 1, 2 \ldots N$. Here, we use a forcing amplitude of $F = 8$ and $N = 40$. Furthermore, cyclic boundary conditions are employed, $x_{41} = x_1, \quad x_{-2} = x_{38}, \quad x_{-1} = x_{39}$. A fourth-order Runge-Kutta scheme with a time step of 0.025 is used to solve the system. The system (2) was introduced by Lorenz (Lorenz 1995; Lorenz and Emanuel 1998; Lorenz 2005). The system is allegorical for the dynamics of "weather" at a fixed latitude. If we associate a unit time with 5 d, the system has an error doubling time of about 2 d. This is similar to the doubling times observed in current numerical weather prediction models (Simmons and Hollingsworth 2002). For any of the 40 variables in Eq. (2), the climatological mean is 2.3 and the standard deviation σ_{clim} is 3.6.

An extended Kalman filter is used here as data assimilation system for the Lorenz-95 system. The equations are standard and are given by Fisher et al. (2005), who used this system to illustrate the equivalence between Kalman smoother and weak constraint long-window variational assimilation. The experiments reported here use observations at every site and with a period of 6 h. The observation error standard deviation is set to 15% of the climatological standard deviation.

The system (2) is nonlinear and the covariance evolution in the extended Kalman filter will approximate the actual covariances. For stability reasons, it is necessary to add a source term $\mathbf{Q}$ in the covariance evolution equation. The source term is referred to as model error covariance and is represented here by a diagonal matrix $\mathbf{Q} = \sigma_q^2 \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. The standard deviation σ_q has been tuned in order to yield the best deterministic forecasts. A value of $\sigma_q = 0.001$ is used. The precise value of σ_q is not so important. However, with $\sigma_q = 0$, the Kalman filter will diverge after some time.

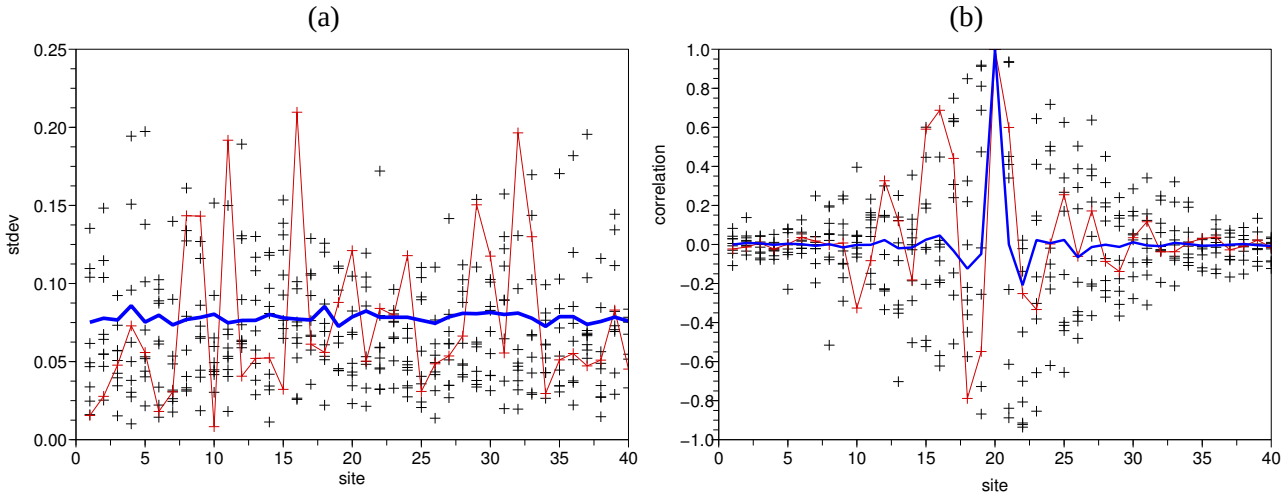

Figure 1: Analysis error standard deviation (a) and correlation with site 20 (b) predicted by the extended Kalman filter for the Lorenz-95 system. The average over 180 cases is shown as heavy blue lines. The values for ten subsequent cases (2 days apart) are shown by the symbols, one case of the ten is highlighted by a red line.

The Kalman filter predicts the analysis error covariance matrix $\mathbf{P}^a$. Figure 1 shows that the predicted $\mathbf{P}^a$ is indeed flow-dependent. For individual cases (red lines and black symbols), the standard deviation and the correlations depart significantly from the mean over a large sample of cases (heavy blue lines). Now, we consider the benchmark ensemble prediction system which consists of 100 forecasts which start from initial conditions sampled from the multivariate Gaussian distribution (1). Figure 2a shows the ensemble standard deviation and the ensemble mean RMS error as function of forecast lead time. Both curves agree very well throughout the entire forecast range. This indicates that the ensemble prediction system is statistically consist. The ensemble standard deviation starts from a value of about 3% of the climatological standard deviation σ_{clim} at initial time and increases throughout the 30 day forecast range to a value close to σ_{clim}. Conversely, the Brier skill score starts close to 1 and asymptotes towards 0 at 30 d (Fig. 2b).

In the following, the benchmark ensemble prediction system is compared with ensembles which sample initial conditions from a multinormal Gaussian distribution

$$p_0^{\mathbf{C}} = \mathrm{N}(\mathbf{x}^a, \mathbf{C}). \tag{3}$$

All considered distributions have the state $\mathbf{x}^a$ analysed by the Kalman filter as mean but the covariance $\mathbf{C}$ differs from the covariance predicted by the Kalman filter. The different settings for the covariance matrix are listed in Tab. 1. Experiment A refers to the benchmark experiment; the other experiments are referred to by letters B–H. In all experiments, an ensemble consists of 100 members. The Kalman filter was run for a period of 360 days. Ensemble forecasts are started every 2 days. Verification statistics is based on a sample of 180 cases for all experiments. The true state of the system is used for the verification. Therefore, the Brier skill score at initial time is not equal to 1 (Fig. 2).

A statistically consistent ensemble prediction system should be able to predict flow-dependent variations of uncertainty at all forecast ranges. For a perfectly reliable ensemble, i.e. members drawn from the same distribution from which the true state is drawn, it can be shown that the ensemble standard deviation matches the expected value of the ensemble mean RMS error (Leutbecher and Palmer 2007). This property can be exploited in order to construct a spread-reliability diagram. For a given forecast lead time, pairs of ensemble standard deviation and ensemble mean RMS error are considered for all 40 sites and all 180 cases. Then, the sample is stratified

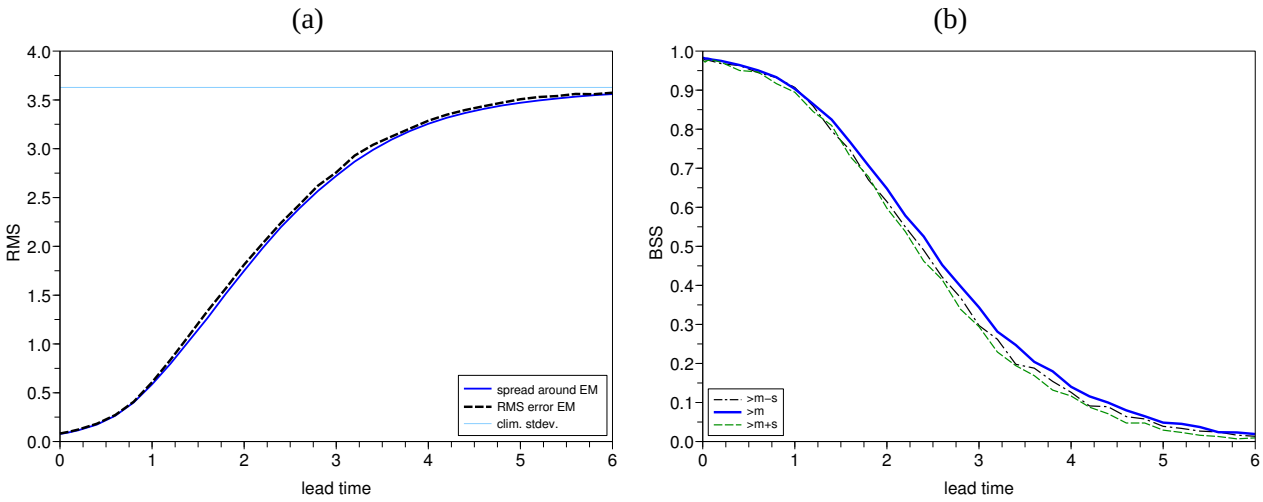

Figure 2: Ensemble standard deviation (spread) and ensemble mean RMS error (a) and Brier skill score (b) for a 100-member ensemble using the initial uncertainty estimated provided by the extended Kalman filter. The Brier Skill Score is shown for three events: anomalies exceeding $-1, 0, +1$ standard deviation of the climatological distribution.

Exp.	$\mathbf{C}(t_k)$	Description
A	$\mathbf{P}^a_k$	benchmark: covariance predicted by KF
B	$\overline{\mathbf{P}^a}$	time mean covariance
C	$\overline{\sigma^2_a}\mathbf{I}$	time and space average variance
D	$\mathbf{P}^a_{\Pi(k)}$	random permutation Π of KF-predicted cov.
E	$4\overline{\sigma^2_a}\mathbf{I}$	inflated time and space average variance
F	$\frac{1}{4}\overline{\sigma^2_a}\mathbf{I}$	deflated time and space average variance
G	$4\overline{\sigma^2_a}\mathbf{D}_G$	correct total variance but wrong spatial distribution
H	$4\overline{\sigma^2_a}\mathbf{D}_H$	correct total variance but wrong spatial distribution

Table 1: Specification of analysis error covariance in the ensemble forecast experiments A–H. $\mathbf{C}(t_k)$ denotes the covariance estimate used for the ensemble starting at time t_k.

by the ensemble standard deviation into 20 bins corresponding to an interval of the spread distribution. The first bin contains the pairs with spread not exceeding the 5% quantile of the spread distribution. The last bin contains the pairs where the spread exceeds the 95% quantile of the spread distribution. Then, the RMS of the ensemble standard deviation and the RMS of the ensemble mean error is computed for the pairs in each bin. For a perfectly reliable ensemble the resulting 20 pairs of ensemble standard deviation and ensemble mean RMS error should be on the diagonal. Note, that for small ensemble size one expects a deviation from the diagonal; this is discusses in detail in (Leutbecher and Palmer 2007).

Figure 3 shows spread-reliability diagrams for experiments A, B, C and D for different forecast lead times. Ensemble dispersion and ensemble mean RMS error (averaged over all 360 pairs in each bin) agree well for the benchmark ensemble, Exp. A. Experiments B and C, which use a time-independent covariance estimate in (3), show no variability of spread initially apart from sampling uncertainty, which is small for a 100 member ensemble. However, over the first 48 hours a significant variability of the ensemble dispersion develops. From 48 h onwards, spread and ensemble mean RMS error agree well for Exp. B and C. In Exp. D, the covariance $\mathbf{C}$ varies from day-to-day. However, these variations are completely independent of the actual variations of uncertainty because $\mathbf{C}$ is obtained from the Kalman filter predicted covariances $\mathbf{P}^a_k$ through a random permutation $\Pi(k)$ of the forecast dates t_k. Thus, the spread-reliability curve is horizontal at initial time for Exp. D. At a lead time of 48 h, the spread is still not perfectly reliable. However, at a lead time of 120 h, the spread is reliable in all four experiments. There may still be a marginal positive benefit at a forecast lead time around 5 days of using the Kalman filter predicted flow-dependent covariance estimate because the range of the spread distribution is slightly larger in Exp. A than in the other experiments.

The total variance of initial perturbations is close to the variance of the initial error in Exps. A–D. Even gross errors in the specification of initial uncertainty (e.g. Exp. D) turn into a reliable prediction of uncertainty after a time scale of 5 days. However, this is not the case for every choice of the covariance matrix $\mathbf{C}$. Experiments E and F demonstrate the persistent detrimental effect of over- and underdispersion, respectively. In the two experiments, the covariance matrix is static and diagonal; the total variance is too large/small by a factor of four. The unreliability of the spread is not lost with increasing forecast lead time in Exps. E and F in contrast to B, C and D (Fig. 4). Obviously, the ensemble standard deviation will eventually asymptote towards the climatological standard deviation in any experiment that uses initial perturbations of some kind.

Even if the total variance is correct, the wrong spatial distribution of the initial variance can cause a persistent degradation of the spread-reliability. Experiments G and H were designed to illustrate this effect. The variances is specified by $\mathbf{C} = 4\overline{\sigma_a^2}\mathbf{D}$, where $\mathbf{D} = \mathbf{D}_G$ and $\mathbf{D} = \mathbf{D}_H$ are diagonal matrices. In Exp. G, the variance is concentrated in sites 1–10, i.e. $\left[\mathbf{D}_G\right]_{jj} = 4$ for $j = 1, \ldots, 10$ and 0 otherwise. In Exp. H, the variance is non zero at every 4th site: $\left[\mathbf{D}_H\right]_{jj} = 4$ if $j \mod 4 = 1$ and 0 otherwise. In both experiments, the total variance is correct

$$\mathrm{tr}\left(\overline{\sigma_a^2}\mathbf{D}_G\right) = \mathrm{tr}\left(\overline{\sigma_a^2}\mathbf{D}_H\right) = \mathrm{tr}\left(\overline{\mathbf{P}^a}\right). \tag{4}$$

Figure 5 shows that the spread in Exp. H is more reliable than the spread in Exp. G. The latter experiment still exhibits significant unreliability at a lead time of 5 d.

Some of the ensemble forecasts, e.g. Exp. D and H, start with initial uncertainty estimates that are very inaccurate, yet they evolve into ensembles with a reliable spread distribution during the first 5 days. Other ensembles, like Exp. G, need significantly longer to evolve into an ensemble with a reliable spread distribution. Let us now try to understand the dynamical principles that determine the time-scale required to obtain a reliable spread distribution. The argumentation is purely qualitative and is not an attempt of a formal proof. Let us consider the ensemble forecast at a location k and forecast lead time t. Furthermore, let $\Omega_k(t)$ denote the region of points at initial time on which the forecast at location k and lead time t depends. In the Lorenz-95 system[1], the size

[1] and also in the real atmosphere

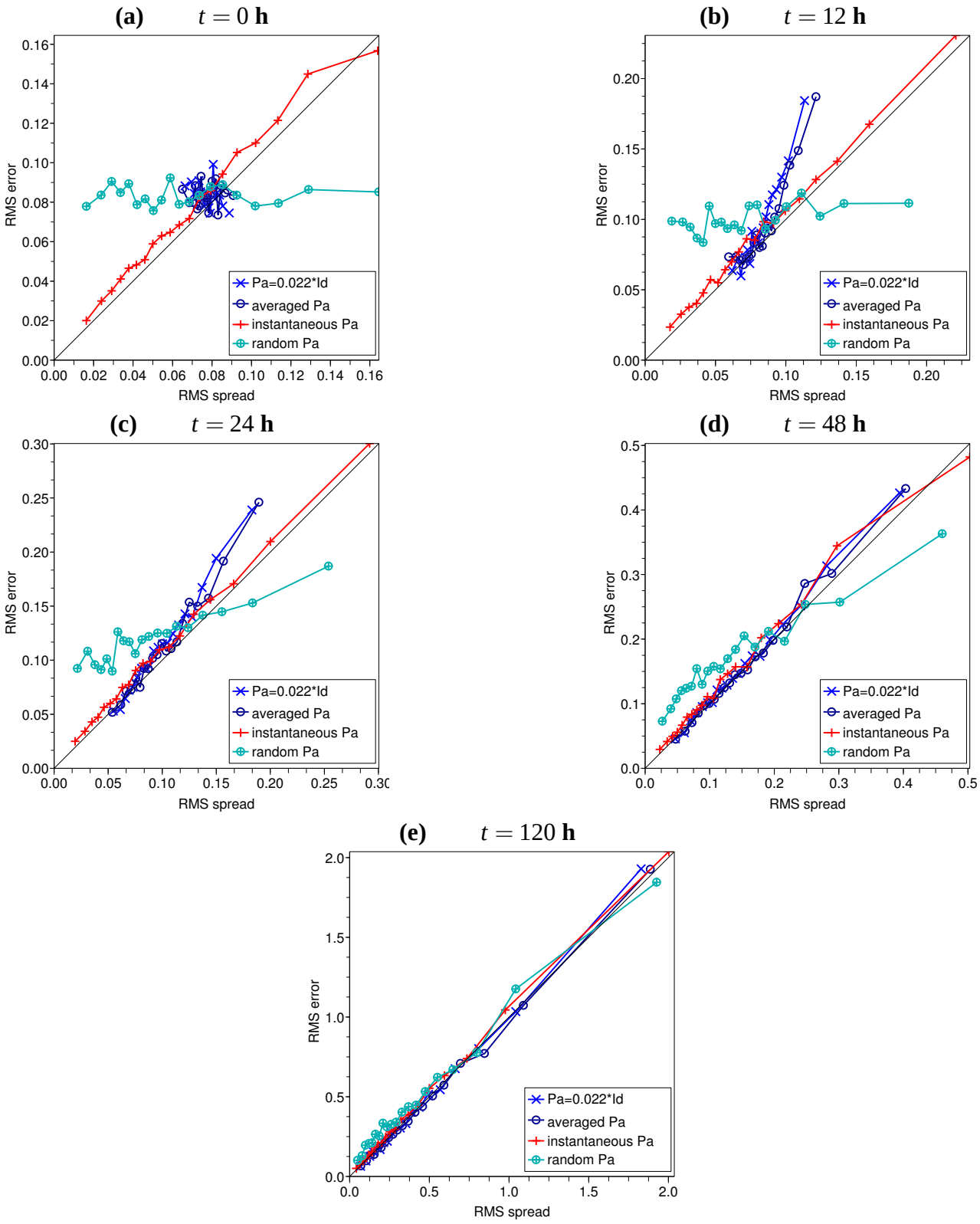

Figure 3: Spread-reliability diagrams for the experiments A (red), B (blue circles), C (blue, crosses) and D (light blue, circled plus) with the Lorenz-95 system. The forecast lead times are 0 h (a), 12 h (b), 24 h (c), 48 h (d) and 120 h (e).

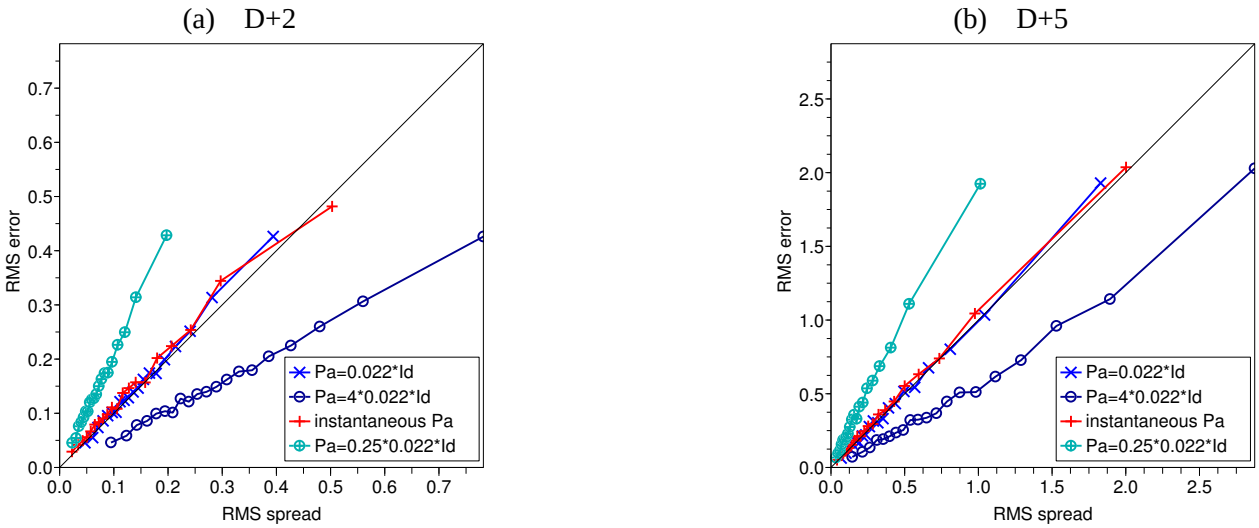

Figure 4: Spread-reliability diagrams for over- and underdispersive ensembles, Exps. E (blue, circles) and F (light blue, circled pluses). The forecast lead times are 48 h (a) and 120 h (b).

of $\Omega_k(t)$ is expected to increase with lead time t until it will eventually fill the entire domain. Figure 1a shows that deviations of the accurate Kalman filter estimate of the analysis error variance from its mean value $\overline{\sigma_a^2}$ are large on small scales, say, $\ll 10$ sites and small on the larger scales (> 10 sites). The leading order error of the ensemble spread is expected to be given by the error of the analysis error variance averaged over the domain of influence $\Omega_k(t)$ if we neglect that the dynamics can amplify uncertainty originating from certain parts of $\Omega_k(t)$ and attenuate uncertainty originating from other parts of $\Omega_k(t)$. In other words, errors in the initial covariance $\mathbf{C}$ potentially cancel if the spatial scale of the dominant variance errors is significantlyt smaller than the size of $\Omega_k(t)$. In Exps. B, C, D and H, the error in the initial variance is dominated by small scales which implies that after relatively short lead times of the order of 2 d, $\Omega_k(t)$ is large enough so that the variance averaged over $\Omega_k(t)$ is fairly accurate. In contrast, in Exp. G and also in Exps. E and F the variance averaged over $\Omega_k(t)$ for t of the order of 2 d will still have significant errors. This, simple theoretical explanation suggests that flow-dependent variations of initial variance are important up to a lead time at which the domains of influence for individual locations become significantly larger than the dominant spatial scale of flow-dependent anomalies of initial variance.

Now, we will examine whether the evolution from an somewhat unreliable spread distribution to a more reliable spread distribution is also present in an operational ensemble prediction system for the atmosphere.

3 The operational EPS

This section describes aspects of the spread-reliability of the operational Ensemble Prediction System (EPS) at ECMWF. The purpose of this brief interlude is to set the scene for the research experiments presented in the next section. Operational daily runs of the EPS started in May 1994. Since then, initial uncertainties were represented in the EPS with singular vectors of the propagator. The propagator of the tangent-linear system is defined over a 2-day interval starting at the forecast initial time. The dynamics is linearised with respect to a nonlinear forecast. The singular value decomposition of the propagator uses a non-dimensionalisation based on the total energy norm (Buizza and Palmer 1995; Palmer et al. 1998). In the currently operational EPS, the extra-tropical perturbations in each hemisphere are sampled from an isotropic Gaussian distribution in the

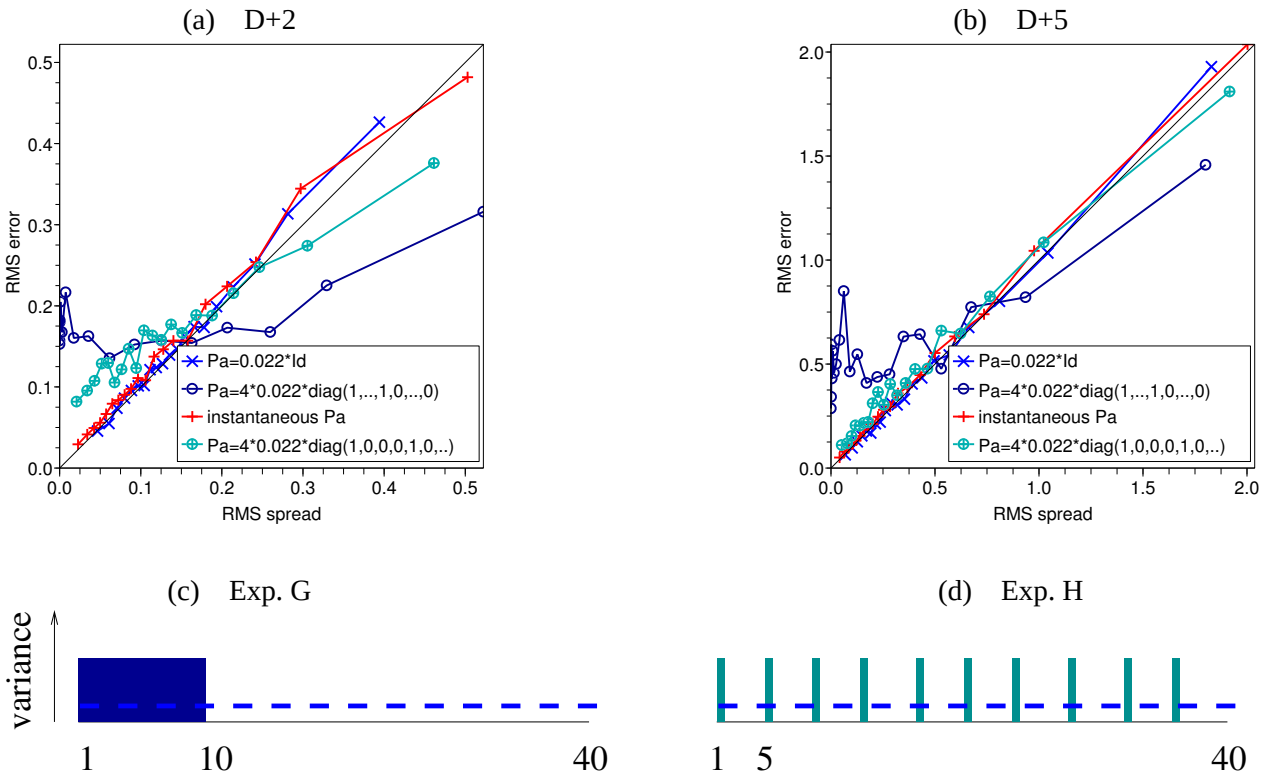

Figure 5: Spread-reliability diagrams for ensemble forecast experiments G (blue, circles) and H (light blue, circled pluses) with the Lorenz-95 system at lead time 48 h (a) and 120 h (b). The incorrect spatial distribution of the variance in Exp. G and H is illustrated in panels (c) and (d), respectively.

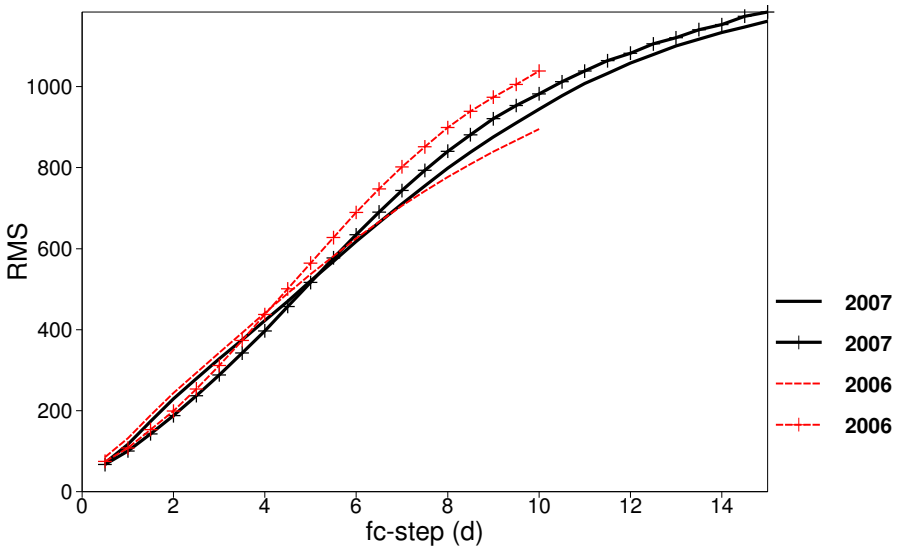

Figure 6: Ensemble standard deviation (plain lines) and ensemble mean RMS error (lines with symbols) of 500 hPa geopotential for the operational EPS. Northern Hemisphere mid-latitudes (35N–65N), winter DJF05/06 (red-dashed) and winter DJF06/07 (black-solid).

subspace spanned by the leading 50 singular vectors; details are described by Leutbecher and Palmer (2007). Through the propagator, the initial perturbations are obviously flow-dependent. However, the initial time norm, which should be based on an estimate of the analysis error covariance matrix (Palmer et al. 1998), is not flow-dependent. The ensemble consists of 50 perturbed forecasts and one unperturbed forecast. The integrations are performed with variable resolution (VAREPS, Buizza et al. 2007). The horizontal resolution corresponds to 50 km up to a lead time of 10 d and then drops to 80 km for lead times from 10–15 d. The number of vertical layers is 62. The model tendencies are perturbed with a stochastic diabatic tendency scheme, a.k.a. stochastic physics (Buizza et al. 1999). The net tendencies of the physical parameterisations are multiplied by random numbers drawn from a uniform distribution in the interval $[0.5, 1.5]$. The random numbers are kept fixed in tiles of $10\,^\circ \times 10\,^\circ$ and over a period of 3 h.

The current EPS resolution was introduced in February 2006. Prior to that date, the resolution was 80 km with 40 layers for the vertical discretisation. As the higher resolution system is more active, the amplitude of the initial perturbations based on the evolved singular vectors could be reduced by 33%. This resulted in an improved match between spread and ensemble mean RMS error (Fig. 6).

For the last winter season, DJF06/07, the spread-reliability has been computed for 500 hPa geopotential in the Northern Hemisphere extra-tropics (Fig. 7). The statistics has been computed on a regular $2.5\,^\circ \times 2.5\,^\circ$ grid. The area represented by each grid point has been accounted for by using weights proportional to the cosine of latitude. Variations in spread correspond to significantly smaller variations of the expected ensemble mean RMS error for the shorter lead times 1 d and 2 d while the spread at 5 d is almost perfectly reliable. The high level of spread-reliability is maintained at lead times larger 5 d (not shown).

4 Ensemble forecasts using ensembles of 4D-Var analyses

Now, experimentation with a lower resolution version of the ECMWF EPS is described which investigates potential merits of using initial perturbations based on an ensemble of analyses. These are generated by an en-

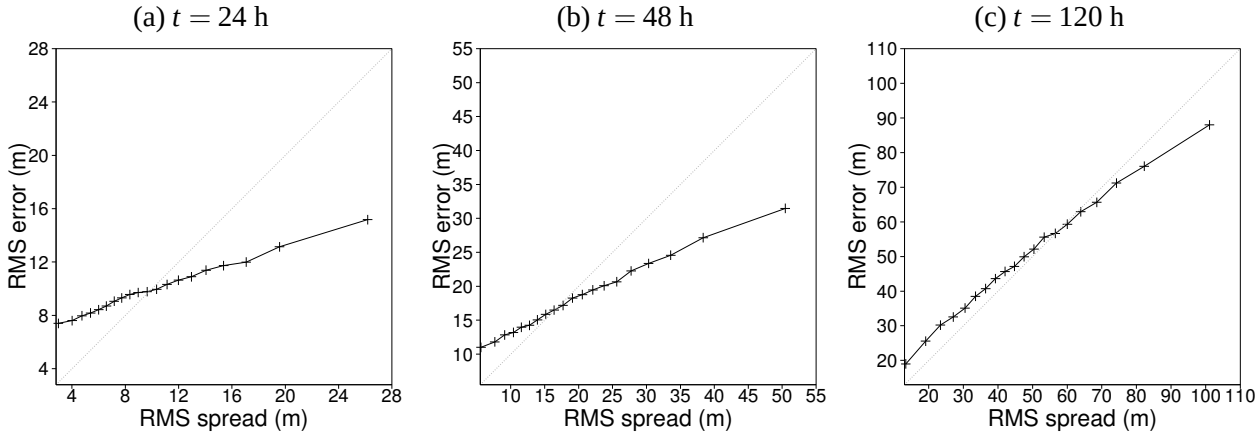

Figure 7: Spread reliability for 500 hPa geopotential at lead time 24 h (a), 48 h (b) and 120 h (c) for operational ECMWF EPS. Northern Hemisphere mid-latitudes (35N–65N), winter DJF06/07.

semble of 10 independent 4D-Var assimilations. These 4D-Var experiments are described in detail by Isaksen (2007). Observations are perturbed with random numbers drawn from a Gaussian distribution. The standard deviation of the Gaussian is set equal to the estimate of the observation error standard deviation used in the operational unperturbed assimilation. The random numbers are uncorrelated except for atmospheric motion vector winds which are perturbed with horizontally correlated noise. The model tendencies are perturbed with a stochastic backscatter scheme. The scheme originates from the kinetic energy backscatter scheme developed by Shutts (2005). Instead of the cellular automaton it uses a first order autoregressive process in spectral space as pattern generator (Judith Berner pers. comm.). The scheme directly perturbs vorticity tendencies. The perturbation amplitude is modulated by an estimate of the kinetic energy dissipation rate as in the original scheme. The vertical structure is imposed by requiring that the pattern has the same vertical correlation as the background error statistics (Isaksen 2007). In addition, temperature and divergence are perturbed by structures implied by patterns resulting from applying non-linear balance and the omega-equation to the vorticity perturbations. The analysis is computed at a resolution of T_L255 ($\sim$ 80 km) and 91 levels (model top at 1 Pa).

Ensemble forecasts discussed in this section consist of 50 members with a resolution of T_L255 and 62 levels (model top at 5 hPa). The tendencies of the ensemble members are perturbed with the operational stochastic physics scheme. Four different configurations for the initial perturbations are considered

- initial singular vectors and evolved singular vectors (SV i+e)
- initial singular vectors only (SV i)
- perturbations of EnDA members about ens. mean (EnDA)
- EnDA perturbations and initial singular vectors (EnDA+SV i)

The initial perturbations are added to the interpolated operational analysis (T_L799L91). Ensemble forecasts are started every other day for a period in September/October 2006. The sample consists of 20 cases. The forecast range is 10 days.

Figure 8 shows 500 hPa geopotential spread and ensemble mean RMS error versus lead time for the Northern Hemisphere mid-latitudes. Experiment (EnDA) is clearly underdispersive at all forecast lead times. In the other three experiments, the spread is fairly similar and relatively close to the ensemble mean RMS error. Replacing evolved singular vectors by ensemble data assimilation perturbations (SV i+e$\rightarrow$EnDA + SV i) leads

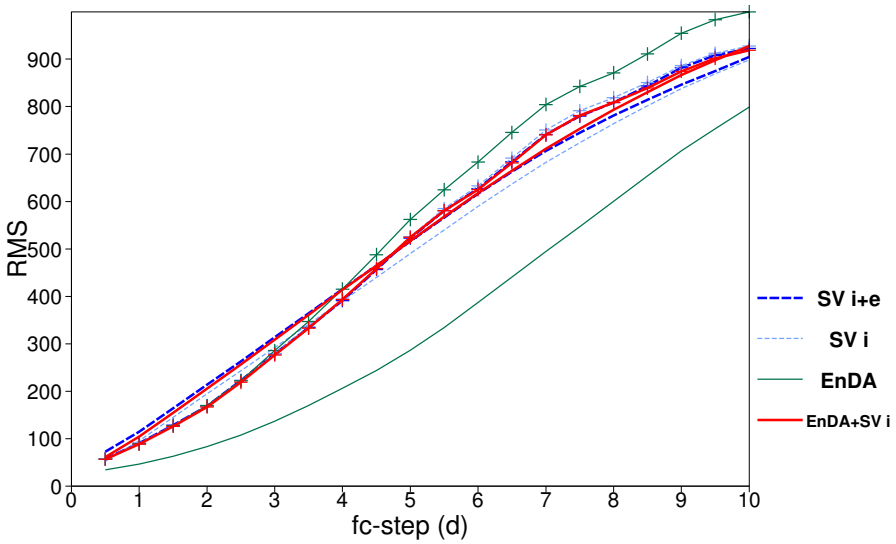

Figure 8: Ensemble standard deviation and ensemble mean RMS error for 500 hPa geopotential in the Northern Hemisphere mid-latitudes (35N–65N) for the ensemble data assimilation (EnDA) experiments. 20 cases in Sep/Oct 2006.

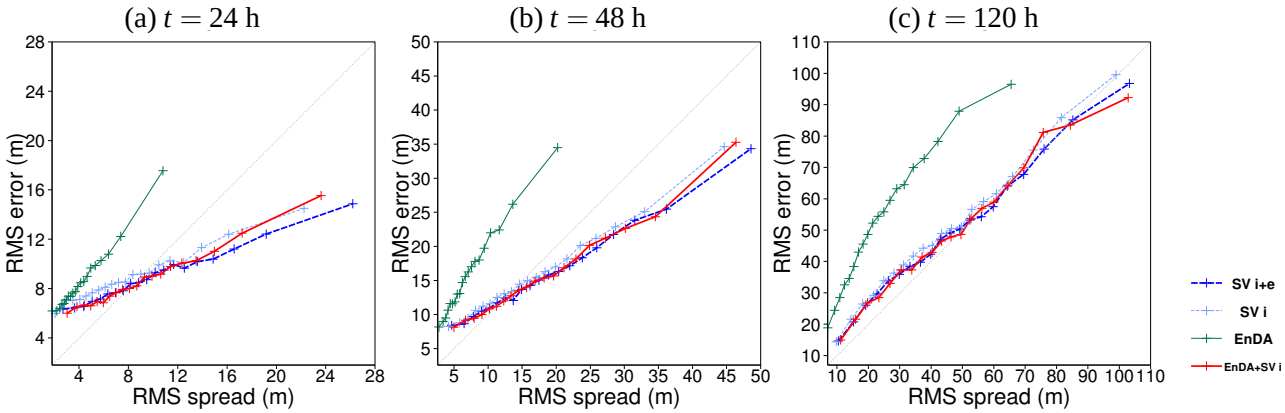

Figure 9: Spread reliability for 500 hPa geopotential at lead time 24, 48 and 120 h over the Northern Hemisphere mid-latitudes (35N–65N) for the EnDA experiments. 20 cases in Sep/Oct 2006.

to a moderate reduction in spread in the early forecast range ($\leq$ 3 d) and a moderate increase in spread in the late forecast range ($\geq$ 7 d). This is beneficial for the spread-skill relationship.

The spread-reliability statistics in Fig. 9 also emphasises the underdispersive characteristics of experiment (EnDA). The spread-reliability of the three other configurations is almost identical for lead times $\geq$ 2 d. At a lead time of 1 d, experiment (EnDA+SV i) appears slightly superior to Exp. (SV i+e) as the ensemble standard deviation in the three bins with largest spread is lower in the former configuration, thus, reducing somewhat the over-dispersion.

The verification of 500 hPa geopotential focuses on the large-scale aspects of the flow. Therefore, it is of interest to complement it with the verification of another parameter which is dominated by smaller spatial scales. Here, we present the zonal wind component at 850 hPa. Figures 10 and 11 show the spread and ensemble mean RMS

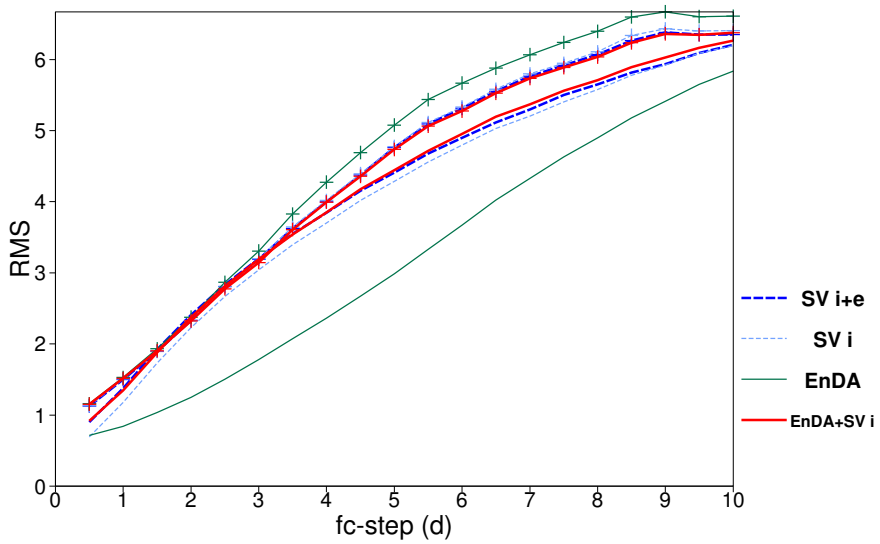

Figure 10: Ensemble standard deviation and ensemble mean RMS error for the zonal wind component at 850 hPa in the Northern Hemisphere mid-latitudes (35N–65N) for EnDA experiments. 20 cases in Sep/Oct 2006.

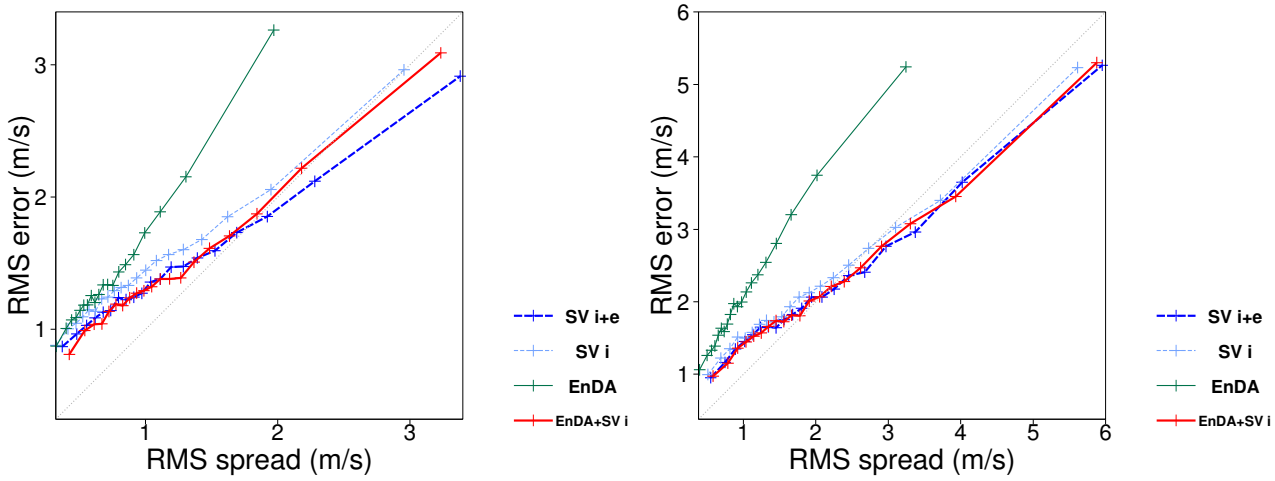

Figure 11: Spread reliability for the zonal wind component at 850 hPa and lead time 24, 48 h over the Northern Hemisphere mid-latitudes (35N–65N). EnDA experiments, 20 cases in Sep/Oct 2006.

error versus lead time and spread-reliability statistics, respectively. The results agree with those for 500 hPa geopotential in a qualitative sense. Figure 12 shows the Ranked Probability Skill Score (RPSS) for the mid-latitudes and the Northern Hemisphere extra-tropics for the four experiments. The RPSS has been computed using ten climatologically equally likely bins. Experiment (EnDA) is clearly inferior to the three experiments which use singular vectors. The latter three are quite similar in terms of their probabilistic skill. For the extra-tropics, Exp. (EnDA+SV i) is slightly superior to Exp. (SV i+e), i.e. the operational configuration, at all forecast ranges.

In Exps. (SV i) and (SV i+e), the initial conditions are unperturbed in large parts of the tropical region. In contrast, the initial conditions are perturbed in the entire tropical belt in Exps. (EnDA) and (EnDA+SV i),

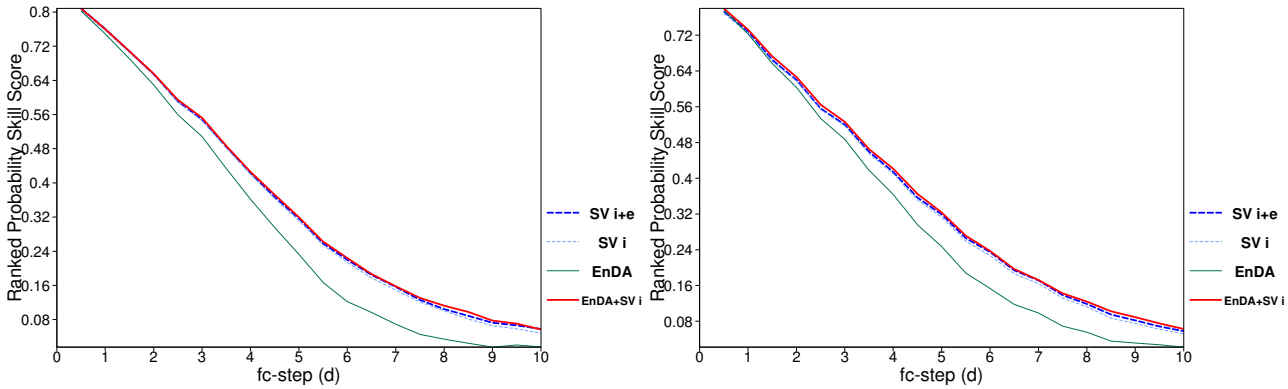

Figure 12: Ranked probability skill score for the zonal wind component at 850 hPa for the EnDA experiments. (a) Northern Hem. mid-latitudes (35N–65N), (b) Northern Hem. extra-tropics (20N–90N).

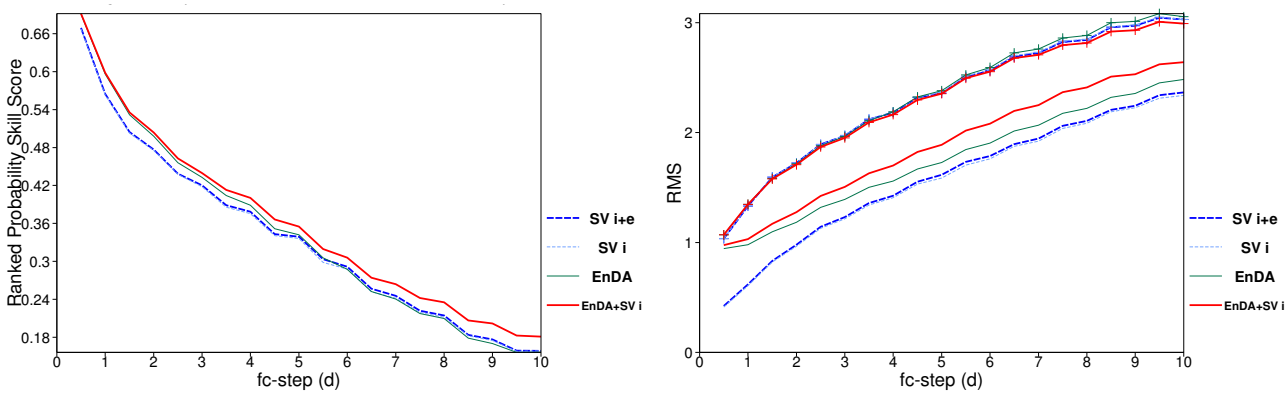

Figure 13: Ranked probability skill score (a) and ensemble standard deviation and ensemble mean RMS error (b) over the tropics (20S–20N) for the zonal wind component at 850 hPa.

which use perturbations from the ensemble of analyses. This has a significant positive impact on the RPSS (Fig. 13a). The operational configuration, Exp. (SV i+e), has significantly lower probabilistic skill than Exp. (EnDA+SV i) at all lead times. The spread in the latter experiment is the largest of all four configurations and the most consistent with the RMS error of the ensemble mean (Fig. 13b). Initially, Exp. (EnDA) has a similar spread as Exp. (EnDA+SV i). However, from a lead time of 48 h onwards, the spread is noticably lower in the former experiment indicating the importance of dispersion in the extra-tropics for maintaining a realistic spread in the tropics.

5 Discussion

The results presented in the previous section indicate that the skill of the EPS could be improved by replacing the initial perturbations based on evolved singular vectors by initial perturbations obtained from an ensemble of analyses. These results also indicate a promising avenue for future research as it is anticipated that further improvements in probabilistic skill can be achieved if the level of underdispersion in the ensemble of analyses is reduced. This work may eventually lead to an EPS that uses only initial perturbations from an ensemble of analyses. This would make the EPS more similar to the Canadian ensemble prediction system which is based

on an Ensemble Kalman filter which uses perturbed observations (Houtekamer et al. 1996, 2005). However, at present, the initial singular vectors are still an essential component of the ECMWF EPS as they efficiently compensate for missing sources of uncertainty or sources of uncertainty that are not adequately represented in the ensemble of analyses.

Further work is planned to examine the sensitivity of the results to the configuration of the ensemble of variational assimilations. The main aspects that will be considered are

- inner and outer loop resolution in 4D-Var
- the number of analysis members
- the selection of observations, in particular satellite radiances
- the representation of correlated observation error in the perturbations applied to the observations
- the representation of model uncertainty in the analysis ensemble. Revised versions of the backscatter algorithm will be considered as well as optimal tendency perturbations, also known as forcing singular vectors (Barkmeijer et al. 2003).

When this work is more mature, further tests will be performed with ensemble forecasts run at the resolution of the operational EPS (T_L399L62).

Should further experimentation confirm the need to keep initial singular vectors in the EPS, it seems worth to investigate the impact of using uncertainty information from the ensemble of analyses in the singular vector initial norm. Currently, the singular vectors are computed with a total energy norm at initial time which can be viewed as a computationally cheap approximation of an analysis error covariance based norm (Palmer et al. 1998). A more sophisticated and computationally more costly approximation of the analysis error covariance norm is provided by the Hessian of the variational cost function (Barkmeijer et al. 1998). Singular vectors computed with this norm are referred to as Hessian singular vectors. Barkmeijer et al. (1999) and Lawrence et al. (2007) have compared the skill of ensemble forecasts using initial perturbations based on Hessian singular vectors with ensemble forecasts using total energy singular vectors. Both studies found that the impact of using Hessian singular vectors on probabilistic skill was close to neutral overall. This might be due to the fact that the Hessian singular vectors are computed with a static background error covariance term and, therefore, the implied initial uncertainty estimate may lack important flow-dependent variations. Ensembles of analyses offer the prospect of computing singular vectors with a flow-dependent analysis error (co)-variance metric following earlier work by Gelaro et al. (2002) and Buehner and Zadra (2006). It is possible to obtain these analysis error (co)-variance singular vectors as solution of a symmetric ordinary eigenproblem. Thus, they could be computed at roughly the same cost as the operational total energy singular vectors.

6 Conclusions

One of the outstanding challenges in predictability and data assimilation is the estimation of initial condition uncertainty. Due to the flow-dependence of error growth, such estimates of initial uncertainty are expected to vary from day to day. The probabilistic skill of an ensemble prediction system is expected to increase with the realism of the initial uncertainty representation. However, it is difficult to obtain accurate estimates of initial uncertainty because some of the relevant sources of uncertainty are poorly known themselves. We expect that initial uncertainty estimates are sensitive to, for instance, the representation of model uncertainty and to the representation of observation error correlations.

In principle, data assimilation techniques can provide accurate flow-dependent estimates of initial uncertainty when all sources of uncertainty are precisely known and represented. This is demonstrated here within the framework of a 40-variable Lorenz-95 system using an extended Kalman filter. A benchmark ensemble forecasting system is considerd which samples initial conditions from the multi-variate Gaussian distribution predicted by the extended Kalman filter. The benchmark ensemble is then compared with other ensembles that start from inferior initial uncertainty estimates but the same unperturbed analysis. The key diagnostic employed here is a *spread-reliability* statistic. It considers, pairs of ensemble mean RMS error and ensemble standard deviation for individual locations and forecast start dates. These pairs are stratified by the ensemble standard deviation and then the RMS error and spread are computed over subsets corresponding to a spread-interval. This statistic focuses on the ability of an ensemble prediction system to capture variations in forecast uncertainty. Then, several ensemble prediction systems are considered that use flow-independent estimates of initial uncertainty. The results indicate that the skill of ensemble forecasts benefits from accurate, flow-dependent estimates of analysis error covariances at lead times shorter than ~ 5 days. Furthermore, the experiments suggest that only gross systematic errors in representing initial uncertainty, like a systematic over- or underestimation of the initial error variance over a large region, appear to be capable of deteriorating the ensemble forecasts at longer lead times (> 5 d).

Verification of the operational ECMWF EPS ($T_L399L62$) for the last winter season (DJF06/07) shows that the ensemble standard deviation matches fairly closely the ensemble mean RMS error for 500 hPa geopotential when considering spatially aggregated statistics. However, the verification in terms of *spread-reliability* reveals that, for the early forecast ranges ($\leq$ D+3), the EPS is systematically over-dispersive at locations and times when the spread is large and systematically under-dispersive when the spread is low. The spread-reliability improves with lead time. By 5 days, the spread is almost perfectly reliable. The gradual improvement of the spread-reliability with lead time was also observed for the Lorenz-95 system and may be a generic property of nonlinear systems which sensitively depend on initial conditions.

Results of first tests of initial perturbations from an ensemble of 4D-Var analyses in a $T_L255L62$ version of the EPS were presented. The ensembles of 4D-Var analyses use perturbed observation values and a representation of model uncertainty based on a stochastic kinetic energy backscatter algorithm. The analysis ensembles are discussed in detail by Isaksen (2007). Four different configurations for the ensemble prediction system were considered: *(i)* an ensemble using initial perturbations from the ensemble of analyses, *(ii)* an ensemble using initial perturbations based on initial and evolved singular vectors (the operational configuration), *(iii)*, an ensemble using initial perturbations based on initial singular vectors only, and *(iv)* a hybrid configuration using initial perturbations based on perturbations from the ensemble of analyses plus perturbations based on initial singular vectors. Results are based on 20 cases in September/October 2006. The ensemble using initial perturbations from an ensemble of analyses exhibits a lack of dispersion at all forecast ranges up to 10 days. It is concluded that initial singular vectors are still required to represent the effect of unknown aspects of the initial error distribution on ensemble dispersion. However, despite the underdispersiveness of the ensemble of analyses, experiments indicate that replacing the operational initial perturbation configuration by the hybrid configuration *(iv)* which uses perturbations from the ensemble of analyses instead of the evolved singular vector perturbations has a moderately positive impact in the extra-tropics and a significantly positive impact in the tropics on the probabilistic skill of the EPS. This positive impact cannot be obtained by simply using only initial singular vectors in the EPS.

References

Barkmeijer, J., R. Buizza, and T. N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2333–2351.

Barkmeijer, J., T. Iversen, and T. N. Palmer, 2003: Forcing singular vectors and other sensitive model structures. *Quart. J. Roy. Meteor. Soc.*, **129**, 2401–2423.

Barkmeijer, J., M. Van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Roy. Meteor. Soc.*, **124**, 1695–1713.

Buehner, M. and A. Zadra, 2006: Impact of flow-dependent analysis-error covariance norms on extratropical singular vectors. *Quart. J. Roy. Meteor. Soc.*, **132**, 625–646.

Buizza, R., J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, T. N. Palmer, and F. Vitart, 2007: The new ECMWF VAREPS (variable resolution ensemble prediction system). *Quart. J. Roy. Meteor. Soc.* (in press, manuscript available at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/401-500/tm499_rev.pdf).

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Buizza, R. and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.

Fisher, M., M. Leutbecher, and G. A. Kelly, 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3235–3246.

Gelaro, R., T. E. Rosmond, and R. Daley, 2002: Singular vector calculations with an analysis error variance metric. *Mon. Wea. Rev.*, **130**, 1166–1186.

Houtekamer, P., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.

Isaksen, L., 2007: Use of analysis ensembles in estimating flow-dependent background error variances. In *These proceedings*, ECMWF.

Lawrence, A. R., M. Leutbecher, and T. N. Palmer, 2007: Comparison of total energy and Hessian singular vectors: Implications for observation targeting. *Quart. J. Roy. Meteor. Soc.* (submitted, available at http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/501-600/tm517.pdf)).

Leutbecher, M. and T. N. Palmer, 2007: Ensemble forecasting. *J. Comp. Phys.* (in press, doi:10.1016/j.jcp.2007.02.014, available also as ECMWF Tech. Memo. 514).

Lewis, J. M., 2005: Roots of ensemble forecasting. *Mon. Wea. Rev.*, **133**, 1865–1885.

Lorenz, E. N., 1995: Predictability: A problem partly solved. In *Seminar on Predictability*, volume Vol. I, ECMWF, Reading, UK, 1–18.

Lorenz, E. N., 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587.

Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677.