TECHNICAL MEMORANDUM

632

Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system: the effect of ensemble size and day-to-day variability

Massimo Bonavita[(1)], Laure Raynaud[(2)] and Lars Isaksen[(1)]

Research Department

[1] ECMWF, United Kingdom
[2] CNRM/GAME, Météo-France/CNRS, France

July 2010

Submitted to Quart. J. Roy. Meteor. Soc.

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract
Ensemble of data assimilations (EDA) methods have been shown to be able to provide flow-dependent estimates of analysis and background error statistics. For this reason, they potentially present a way to overcome one of the main limitations of current variational data assimilation systems. However, the limited number of ensemble members which can be realistically run in an operational context and the stochastic nature of the EDA approach lead to high levels of sampling noise in the relevant ensemble statistics. To answer this problem, an "objective filtering" technique of the sample ensemble variances proposed by Raynaud *et al.* (2009) has been implemented at ECMWF. In this paper we present a comparison of the ability of ensemble data assimilation systems of different sizes (10 to 50 members) to represent flow-dependent background-error variances. In particular, the ensemble-based variances are examined in the case of the severe storm Klaus (24 January 2009) over France and in the case of the Atlantic tropical hurricane Ike (1-14 September 2008). Our results show that, while a relatively small ensemble (10 members) can be sufficient to resolve the larger scale error structures connected to an extra-tropical cyclogenesis, a larger ensemble is beneficial to resolve more localised anomalies like those connected with a hurricane. In this sense, the objective filtering technique provides a useful indication of the spatial scales the ensemble is able to resolve in a statistically robust way.
The day-to-day variability of the ensemble statistics and how this affects the objective spatial filtering procedure are also examined. Our conclusion is that a time-independent implementation of the filter based on a climatology of truncation wavenumbers results in more robust ensemble statistics estimates, and ultimately in improved forecast skill scores.

1 Introduction

It is well known that an accurate specification of the statistics of background errors is a fundamental prerequisite of a successful data assimilation scheme. Methods based on an ensemble of perturbed data assimilations (EDA) have been successfully employed in the specification of background error statistics at ECMWF (Fisher, 2003) and Météo-France (Belo Pereira and Berre, 2006). However, these applications have relied on a static, climatological representation of the background-error matrix, thus implicitly renouncing to account for the flow-dependency of covariance errors.

An attractive characteristic of EDA methods is their ability to provide day-to-day background-error statistics that represent the current meteorological situation, thus possibly overcoming one of the main limitations of covariance models currently in use in variational data assimilation systems. Early attempts to use flow-dependent error information from a 6 or 10 member EDA system have been reported in Kucukkaraca and Fisher (2006), Fisher (2007), Isaksen et al. (2007) and, more recently, in Raynaud et al. (2009, 2010).

A critical aspect of any EDA system is the choice of the ensemble size. For operational applications ensemble size and spatial resolution are typically determined by a compromise between the desired accuracy of the background error estimates and the meteorological features that need to be represented. Affordable computational costs and operational time constraints also play an important role. So it is important to assess the possible gains that a larger ensemble would permit in the accuracy of the background error estimates. This is particularly relevant for the EDA systems which are based on the use of perturbed observations and boundary conditions. Being purely stochastic methods, one can expect the accuracy of error estimates sampled from the ensemble to increase proportionally to the square root of the ensemble size. This slow rate of convergence and the limited (O(10-100)) ensemble sizes which are viable in current NWP operational contexts require the use of ad-hoc filtering techniques (Raynaud et al., 2008,2009; Berre and Desroziers, 2010) on the raw ensemble estimates. Common to any filtering technique, however, is the problem that a fraction of the signal of interest is

lost in the filtering process. It is then of interest to quantify this loss of information and to what extent a larger ensemble would ameliorate the problem.

In the practical implementation of the objective filtering technique proposed by Raynaud et al. (2009), the filtering step tries to maximize the signal-to-noise ratio of the sample ensemble variances. This can be done adaptively, or using values computed from a statistical sample. We review the pros and cons of the two approaches and perform assimilation experiments to determine the best configuration for the ECMWF EDA system.

The paper is organized as follows. Section 2 describes the EDA system at ECMWF. A diagnostic study of total and local error variances is then presented in section 3, where the concept of the "effective" ensemble resolution is introduced. In section 4, ensemble-based variances for a severe extra-tropical storm case and an intense tropical hurricane are examined. The day-to-day variability of the sample ensemble variances and its impact on the variance filtering algorithm are investigated in section 5. Conclusions are given in section 6.

2 Experimental setup

ECMWF runs an ensemble variational data assimilation system (Isaksen et al., 2010) based on the explicit perturbation of the assimilated observations, the sea-surface temperature field and the model physics tendencies. The background states are implicitly perturbed because they are evolved from the perturbed analysis fields running a short range (t+12h) forecast with the model error parameterization term ("Stochastically Perturbed Parameterization Tendencies" method, SPPT, Palmer et al., 2009) active. Differences between the background fields of ensemble members provide surrogates for samples of background error (Fisher, 2003). All assimilated observations are randomly perturbed by sampling a Gaussian distribution with zero mean and standard deviation equal to the expected observation error. Error correlations are taken into account for atmospheric motion vectors (Bormann et al., 2003). Sea-surface temperature fields are also perturbed with correlated errors using the methodology currently applied in the ECMWF seasonal ensemble forecasting system (Vialard et al., 2005). The analysis step is performed through the ECMWF 4D-Var with a TL399L91 outer loop (i.e., spectral triangular truncation T399 with linear grid and 91 hybrid vertical levels of the ECMWF model) and two successive inner loops at TL95 and TL159 triangular spectral truncations. The assimilation window is 12 hours. The 10 member ECMWF pre-operational EDA system will provide estimates of analysis uncertainty and initial time perturbations for the Ensemble Prediction System (Buizza et al., 2008). The ECMWF EDA can also be used to provide flow-dependent background-error variances to the operational deterministic 4D-Var, in replacement of the current quasi-static background errors estimates derived from the "randomization" technique (Fisher and Courtier, 1995; Fisher, 2003). The "randomization" errors are essentially computed from random samples of the static B matrix used at the start of the 4D-Var analysis with a small flow-dependent component which results from the application of the non-linear balance equation and the omega equation linearized around the background state. The ensemble variance estimates are used as proxies of the background errors both in the minimization step and in the observation quality control check.

For this study, two 50 member EDA were run for 45 days with the same configuration as the pre-operational 10-member EDA, over both a boreal winter (January-February 2009) and a summer (August-September 2008) period.

3 Sensitivity of sample standard deviations to ensemble size

In this section, a diagnostic study of ensemble-estimated standard deviations is presented, based on the results from the 50-member EDA experiment over the boreal winter period (January-February 2009).

3.1 Total standard deviation

The total standard deviation on each model level (i.e. the global spatial average of the ensemble background standard deviations) is the first global diagnostic to check in order to evaluate the impact of larger ensembles. Figure 1 presents the ratio between the vorticity standard deviations calculated from 20, 30, 40 and 50 member ensembles with respect to those computed from the baseline 10-member ensemble. Results correspond to the particular date of 24th of January 2009 at 0900 UTC. It can be seen that the total standard deviation does not change noticeably with respect to the 10-member ensemble at each level since an increase of only 3% is obtained with the largest ensembles, and this increase tends to saturate from 40 members on.

This result is reassuring because it gives an indication that even a relatively small 10-member ensemble is able to capture the main portion of variability of the underlying probability density function.

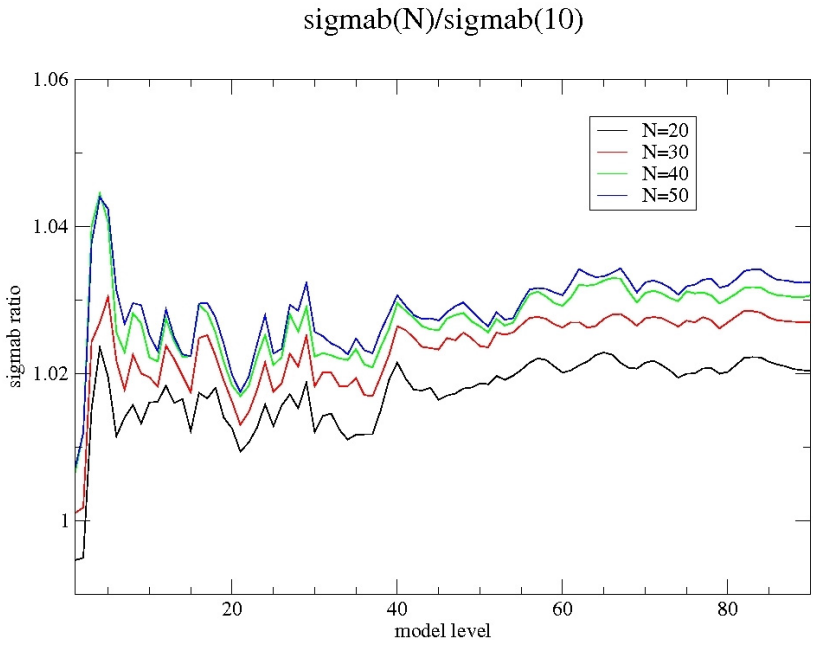

Figure 1: Ratio between horizontal averages of standard deviations estimated with N members (N = 20, 30, 40 and 50) and 10 members, as a function of model level, valid on 24th of January at 0900 UTC.

3.2 Geographical variations

Raw standard deviation maps derived from 10 and 50 members are displayed in Figure 2(a)-(b) for vorticity at model level 64 (close to 500hPa) on 24th of January 2009 at 09 UTC. It can be noticed that the spatial structures (from large to relatively small scales) represented by the 10 and 50-member ensembles are similar, while the associated sampling noise, of smaller scale than the useful signal, is markedly reduced with 50 members. The average Pearson product moment correlation between the two fields is equal to 91%, which points to a good robustness of the 10-member standard deviations. This result basically holds for all model levels and over the entire study period, as it is shown in Figure 3, which presents the time-mean correlation between raw standard deviations estimated from the 10 and 50-member ensembles. The correlation takes values between 90 and 95% in general, except from model levels 34 (70hPa) to 15 (5hPa) for which the correlation is between 80 and 90%. This decrease in correlation in the stratosphere is basically due to smaller correlation values in the Southern Hemisphere (66.6% average value) than globally (close to 80%). This does not reflect any fundamental issue with the underlying EDA system, but it is a consequence of the high level of activity of the Northern hemispheric stratosphere due to the disruption of the polar vortex circulation and stratospheric warming episode taking place during the test period (e.g., Figure 4). The corresponding subdued activity in the Southern hemispheric stratosphere reduces the signal energy spectrum and makes the sampling noise component relatively more important.

It is also interesting to note that a residual, non-negligible sampling noise also affects the 50-member estimates (Figure 2(b)). This suggests that the implementation of a spatial filter remains a necessary step for most practical ensemble sizes (O(10-100)).

These ideas are confirmed by a spectral analysis of the signal. Figure 5 shows the energy spectra of the raw vorticity variances at model level 64 (close to 500hPa) derived from 10 members (in blue) and 50 members (in red), on 24th of January 2009 at 09 UTC. The comparison of the two power spectra reveals that, where the useful signal tends to predominate (large and medium scales, approx. up to total wave number 40), the two variance fields have nearly the same energy. For smaller scales, the variances estimated from the 10-member ensemble have larger energy, linked to a stronger level of sampling noise at these scales (Figure 6).

Figure 2(c) presents the 10-member standard deviations (Fig. 2(a)) after the application of the objective spatial filtering proposed by Raynaud *et al.,* (2009). The sampling noise is removed to a large extent while a large portion of the broad scale signal of interest is preserved. The corresponding energy spectra of the raw and filtered standard deviations (Figure 7) show that the amplitude of small scales is largely reduced, according to the high level of noise there. The amplitude of large and medium scales is also reduced (to a lesser extent): this is consistent with the fact that the whole spectrum is affected by a (more or less) strong sampling noise. One can finally notice that the large-scale structures of the filtered 10-member standard deviations are very close to the filtered 50-member standard deviations (Fig. 2(d)).

(a)

(b)

(c)

(d)

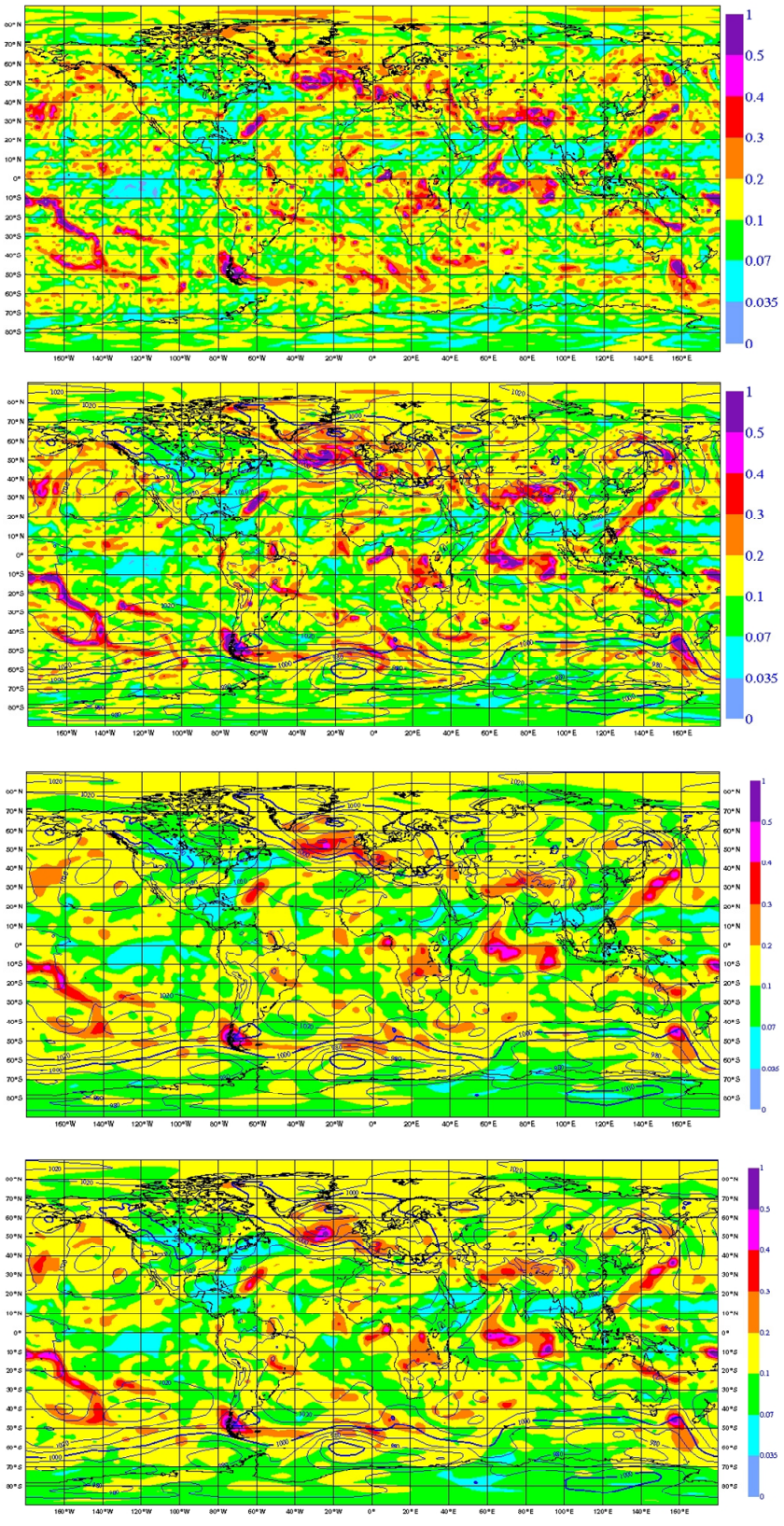

Figure 2: Standard deviations of vorticity at model level 64 (~500hPa), corresponding to 24 January 2009 at 09 UTC. Unit: 5.10^{-5} s^{-1}. (a),(b) Raw estimates from the 10 member and 50 member ensembles, respectively, (c),(d) filtered estimates from the 10-member and 50-member ensemble. The mean sea level surface pressure analysis is overlaid, contour: 10hPa.

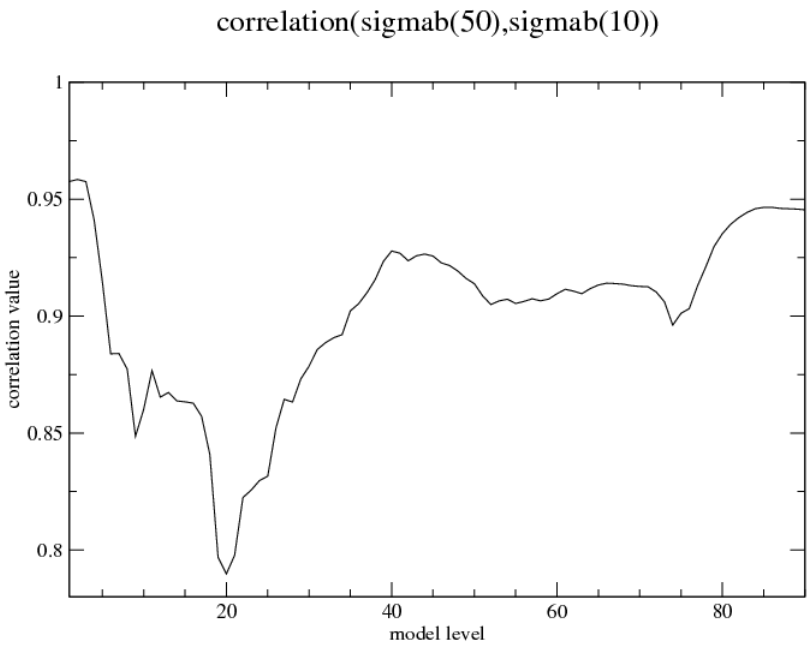

Figure 3: Pearson correlation (averaged over the study period) between the background-error standard deviations of vorticity background fields derived from the 10 and 50-member ensembles, as a function of model level.

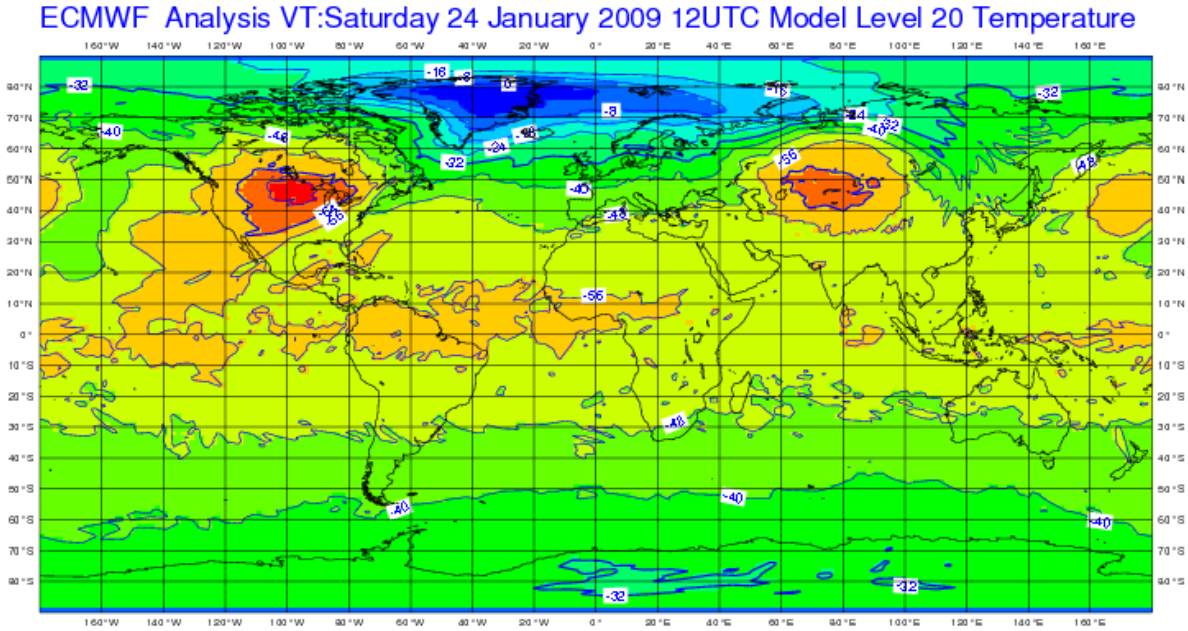

Figure 4: Analysed temperature field at model level 20 (~13hPa), on 24th January 2009 at 12 UTC Unit: C°.

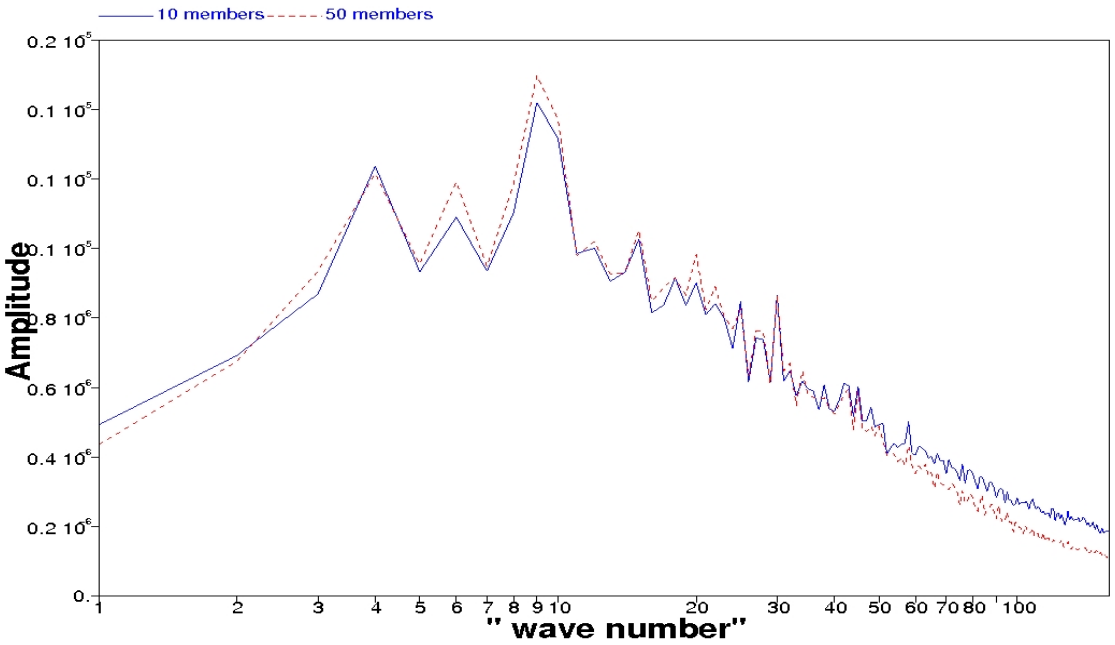

Figure 5: Energy spectra of raw ensemble-based vorticity variances at 500hPa, on 24th January 2009 at 09 UTC: in blue (resp. in red) variances calculated with 10 members (resp. 50 members). Unit: s^{-2}.

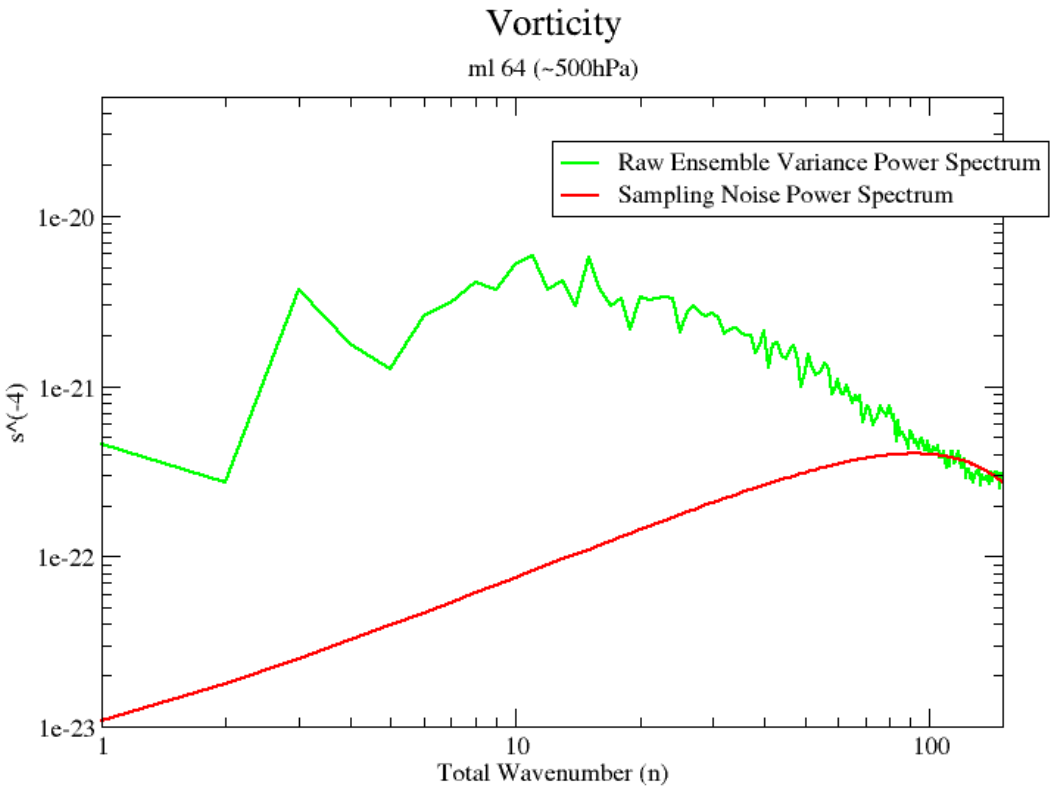

Figure 6: Power spectra of ensemble short range (t+9h) forecast variance (green) and of the sampling noise (red) associated to the estimation of vorticity variances at model level 64 (~500 hPa) with a 10-member ensemble. Unit: s^{-4}.

3.3 Effective ensemble resolution

As it is apparent from figure 6, the objective filtering procedure allows defining a truncation wavenumber beyond which the sampled error estimates are not deemed to be statistically significant. This implies that, at least in terms of forecast error estimation, the effective spatial resolution of the ensemble is typically lower than the nominal resolution at which the ensemble forecasts are actually run (Spectral triangular truncation T399 in the present case).

Figure 8 shows the sample ensemble variance power spectra for temperature at model level 49 (~200 hPa, continuous lines) for a particular date and the corresponding climatological sampling noise spectra (dashed lines) from a 10-member ensemble (red) a 20-member ensemble (blue) and a 50 member ensemble (black). As expected, the portion of the spectrum corresponding to the large scales features (up to wavenumber 40 approx.) is virtually identical for all ensemble sizes. For smaller spatial scales, however, the sampling noise contribution to the total variance spectra is much smaller for the larger ensembles. The climatological noise spectra estimate is also correspondingly smaller, thus allowing the filter to retain finer scale structures from the original signal. In this context, the effective spatial resolution of the statistically significant error estimates that we can compute from the ensemble goes from around T70 for the 10-member ensemble to ~T90 for the 20-member ensemble to >T159 for the 50-member ensemble.

It is noteworthy that even for the 20 member ensemble (20) the effective spatial resolution of the ensemble sample error estimates is far coarser than the nominal ensemble resolution (i.e., ~T90 vs. T399). One needs to use a 50 member ensemble to be able to extract a relatively noise free signal over the whole spectral range shown in the figure (T159). This implies that, in the EDA framework, the most effective way to increase the ensemble resolution is by increasing its size. This conclusion is based on the assumption that sampling error is the dominant form of error in the ensemble estimates of first guess errors. Whenever other error types (i.e., model error) become prevalent, running a larger ensemble will not provide the expected benefits.

The hypothetical advantage of a larger ensemble being able to represent finer spatial details of background error fields should translate into more accurate analysis and forecasts. Preliminary results with a 20-member assimilation experiment show that increasing the sample size from 10 to 20 members has a generally positive impact on forecast scores in the Northern Hemisphere and, more markedly in the Southern Hemisphere (Fig. 9), while the impact is neutral in the Tropics (not shown).

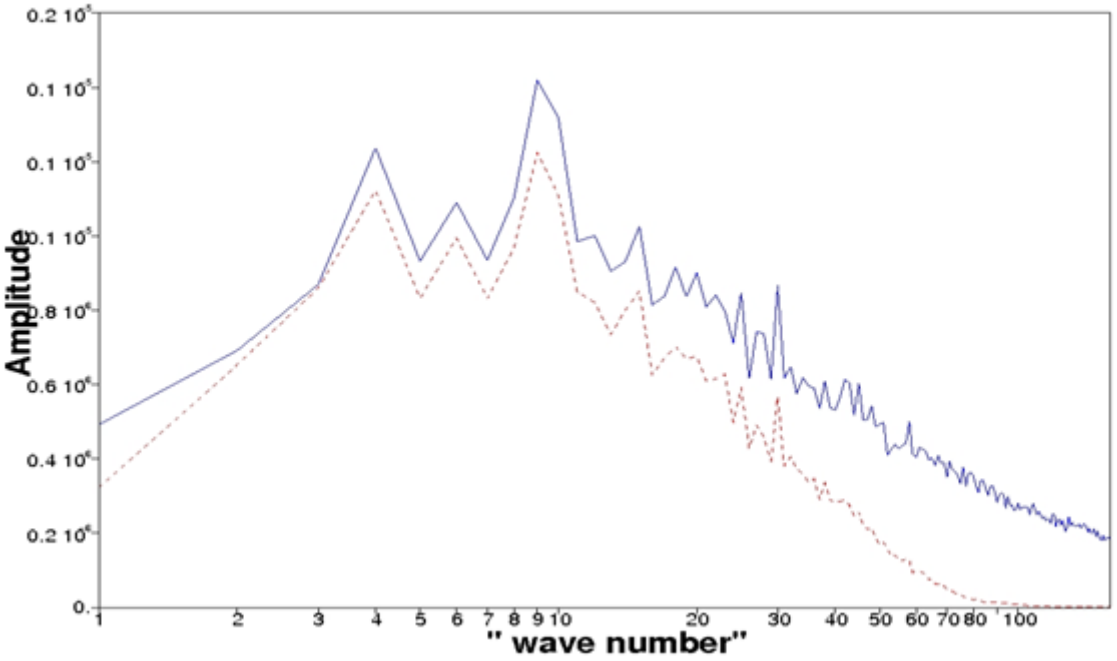

Figure 7: Energy spectra of vorticity standard deviations at 500hPa, on 24th January 2009 at 09 UTC: in blue (resp. in red) raw (resp. filtered) variances calculated with 10 members. Unit: s^{-2}.

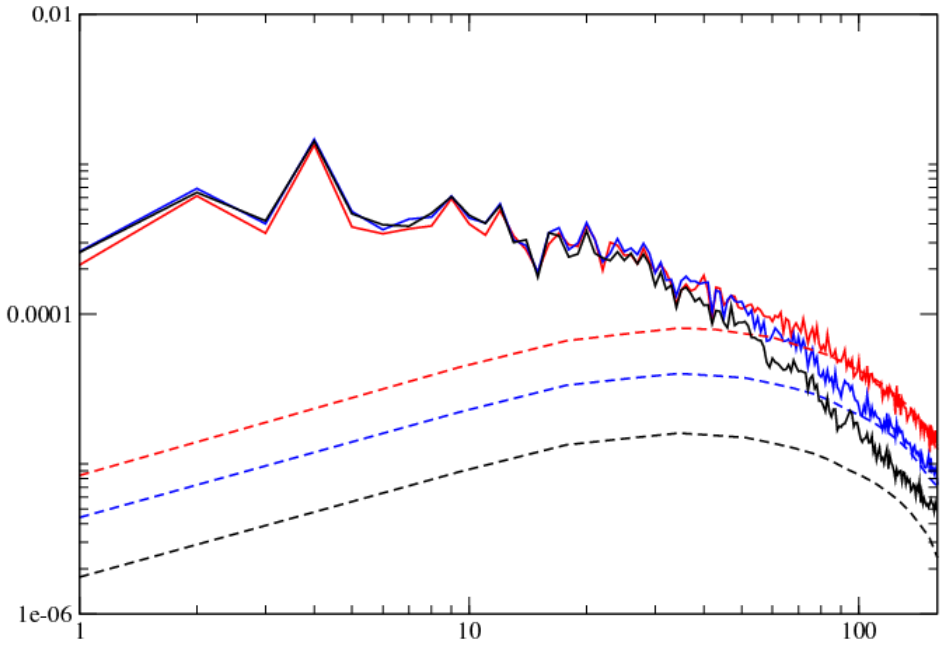

Figure 8: Power spectra of ensemble first guess variance valid on 21 Jan 2009 9UTC (continuous lines) and of the sampling noise (dashed lines) associated to the estimation of temperature variances at model level 49 (~200 hPa) for a 10-member ensemble (red), a 20-member ensemble (blue) and a 50 member ensemble (black). Unit: $°C^4$.

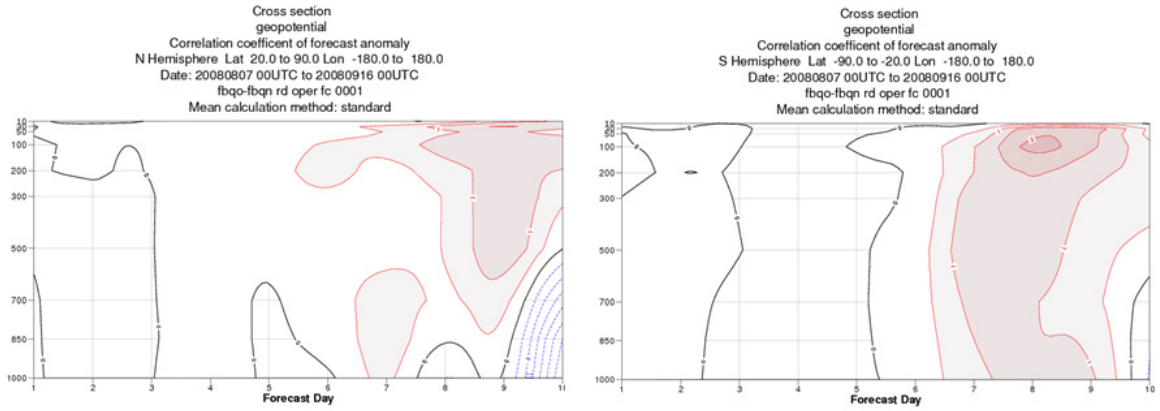

Figure 9: Forecast skill scores difference (Anomaly correlation, Geopotential) of forecasts from deterministic 4D-Var analyses with background error variances estimated from a 20 member EDA versus using variance estimates from a 10 member EDA. Scores are averaged over the period 20080807–20080916. Positive differences contoured in red, negative in blue (contour interval=0.5).

4 Case studies

In this section we examine two different severe weather events. The first is a severe winter storm in the northern extra-tropics (the Klaus storm in January 2009), the second a tropical hurricane (Hurricane Ike in September 2008) which heavily affected the Caribbean Islands and the southern USA. These two case studies will allow us to a) examine the error variances estimated with EDA systems of different size and their link with the prevailing meteorological conditions; b) highlight the impact of using cycled EDA variances in the ECMWF deterministic analysis cycle; c) evaluate the strengths and weaknesses of the objective filtering technique which is currently used to reduce the impact of sampling noise on the EDA variances and d) qualitatively assess the possible benefit of using a larger EDA system.

4.1 The Klaus storm (24th January 2009)

After having caused serious damages in Spain, the severe storm Klaus hit large parts of central and southern France. In the early morning on the 24th of January, the storm reached the Atlantic coast and moved eastwards towards the south-east coast throughout the day (Figure 10). Its intensity was comparable to storm Lothar in December 1999. Wind gusts were recorded of intensity stronger than 36 m/s inland (e.g. 45 m/s in Bordeaux), while they reached 47 m/s along the Atlantic coast and exceeded 53 m/s in the Mediterranean Sea.

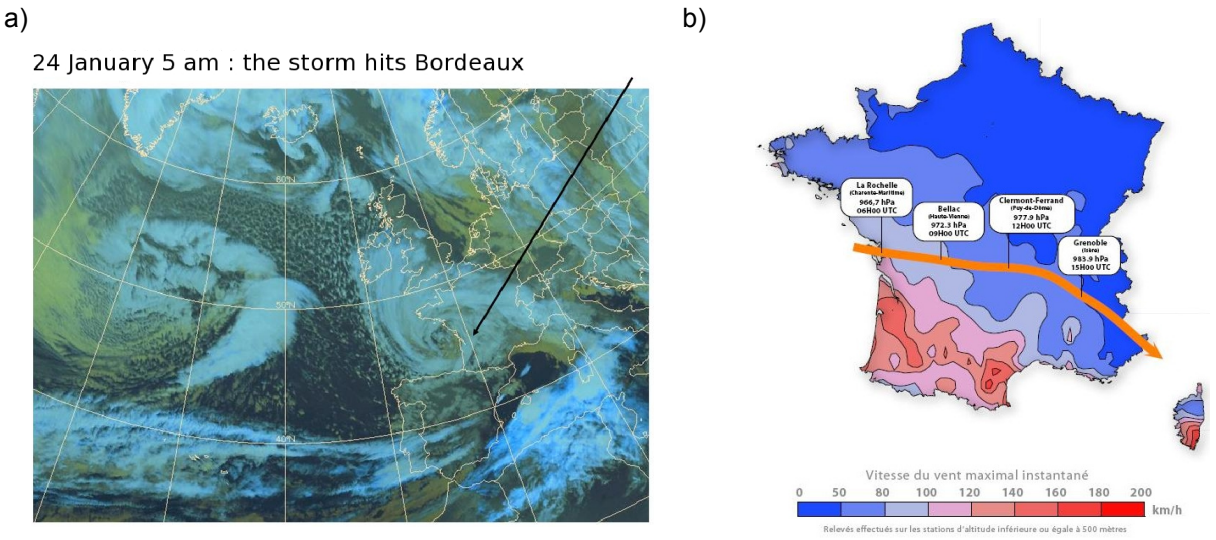

Figure 10: Storm Klaus the 24th January 2009. (a) Satellite image at 5 am UTC and (b) trajectory of the storm and maximum intensity of observed wind. (Data supplied by Météo-France/Dclim).

Flow-dependent background error standard deviations can potentially have a significant impact in the analysis and forecast of severe storm events (Kucukkaraca and Fisher 2006) through an improved representation of the uncertainty in the background state in dynamically active areas.

Figure 11 presents ensemble-based standard deviations of vorticity near the surface, valid on the 24th January 2009 at 09 UTC over Europe and the near Atlantic. Panels (a)-(e) correspond to the raw estimates obtained with 10, 20, 30, 40 and 50 members respectively. The similarity of the different maps and the connection with the underlying dynamical situation are apparent. The uncertainty associated to the low over the Atlantic Ocean is well captured, and all the estimates show a pattern of high standard deviation values in the southern and central parts of France, with local maxima located along the Mediterranean coast. This suggests that an ensemble of only 10 members is thus sufficient to represent fairly accurate (in terms of both position and intensity) estimates of background error associated to this extra-tropical storm event. The objective spatial filtering applied to the 10-member estimates (panel (f)) tends to smooth the geographical variations while keeping an area of large values over the regions most affected by the storm.

A more quantitative assessment of the impact of using flow-dependent background-error estimates in the data assimilation process can be obtained from an examination of the forecast skill of an analysis/forecast experiment using EDA variances. For that purpose, an impact experiment has been performed over a one-month period in January-February 2009 with the ECMWF operational assimilation and forecast systems (IFS). This experiment was run using 4D-Var at resolution T799, and filtered background error variances estimated from a 10-member ensemble run at T399 resolution were used both in the first-guess check and in the background error term of the cost function.

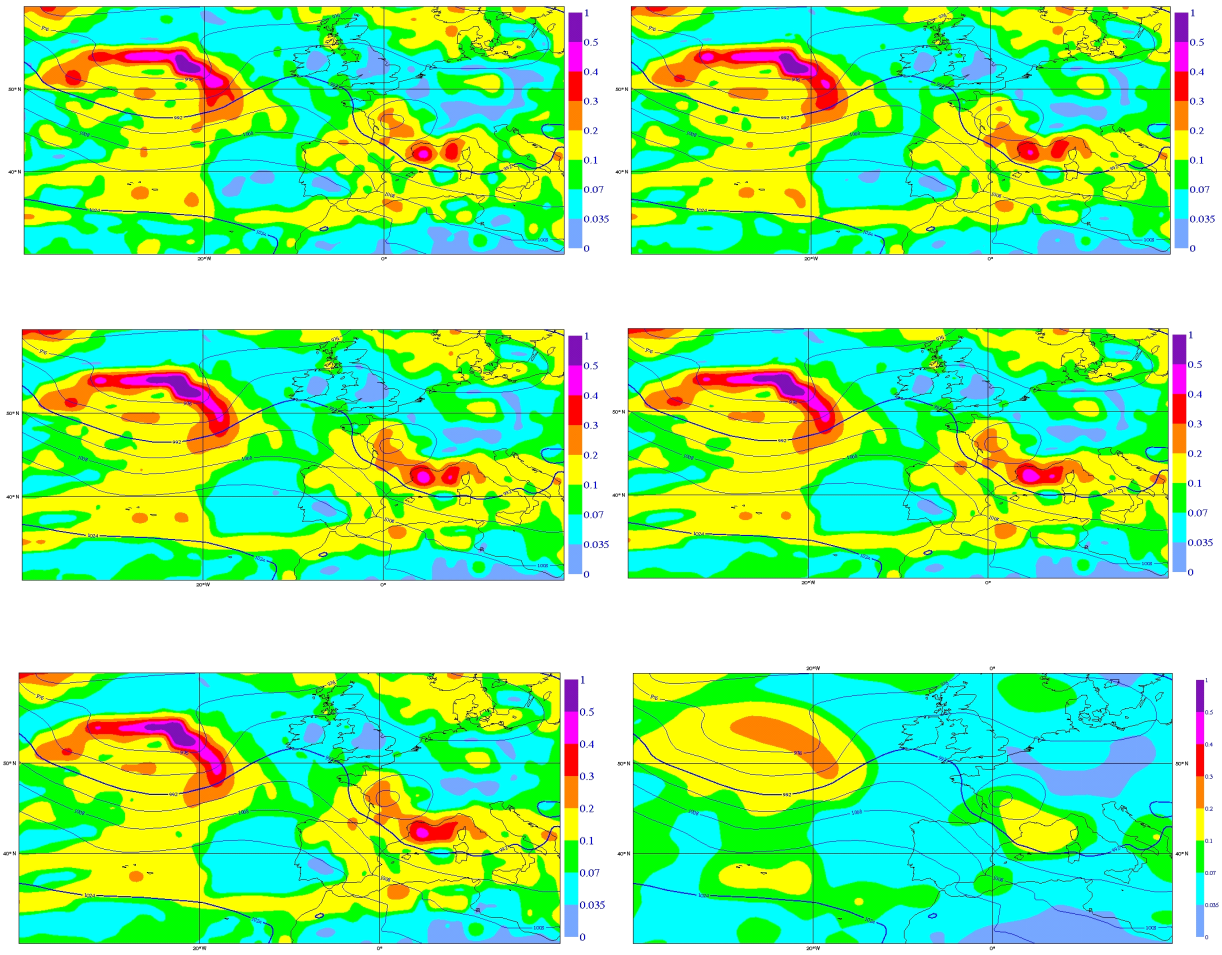

Figure 11: Standard deviations of vorticity near the surface, corresponding to 24 January 2009 at 09 UTC. Unit: 5.10^{-5} s^{-1}. Raw estimates calculated with : (a) 10 members (b) 20 members (c) 30 members (d) 40 members and (e) 50 members. (f) Filtered estimates from the 10-member ensemble. The mean sea level surface pressure analysis is overlaid, contour: 10hPa.

Figure 12 presents the RMS error of t+48h forecasts of surface pressure over Europe for the period 20/01/2009-24/01/2009. Scores are calculated for the control experiment, which uses the operational "randomised" errors, and for the experiment using flow-dependent variances. One can first notice that the EDA variance experiment is improved upon the control experiment over the whole period, with a reduction of RMS error up to around 0.3 hPa. The largest improvement is obtained for the 48h forecast starting from 22 January at 00 UTC and valid on 24 January at 00 UTC.

The temporal evolution of the storm forecasts is presented in Figure 13 for both the control run and the EDA variance experiment (ENS10), along with the analysis valid on 24th January at 00 UTC. The forecasts are issued from analyses valid on 20, 21, 22 and 23 January at 00 UTC. It can be noticed that the two experiments produce quite similar 96h forecasts that already provide a relevant warning signal of the storm. The location of the storm, as well as the position and intensity of strong 10m winds, are coherent with the verifying analysis. The EDA variance experiment somewhat outperforms the control run in forecasting a deeper cyclone and a more accurate area of strong winds. However, both forecasts predict a considerably higher central pressure than the analysed one. The central pressure reaches 983.5 hPa in the control run and 980.8 hPa in the ensemble variance run, while the analysed value is 967.6 hPa. The 72h forecasts tend to confirm the development of the storm. The intensity of associated

winds is more accurately forecasted compared to the verifying analysis, but the central pressure is still higher (982,8hPa for the control run and 982,3hPa for the experimental run). The two 48h forecasts do not suggest such a clear scenario since both experiments seems to lose track of the developing cyclone. There is still an indication of a surface pressure minimum along the Spanish coasts, but the central pressure has increased and the winds are weaker. The EDA variance run is still better at this lead time, with a better positioning of the 992hPa pressure isoline, a stronger pressure gradient and a deeper central pressure (988,4hPa compared to 992hPa in the control run). 24 hours before the storm reaches Spain, the two experiments produce very similar forecasts. The low pressure system is accurately located and the intensity of the winds is quite well captured. The central pressures, although still too high (975hPa in the control and 975,5hPa in the experiment), are shallower than in the previous forecasts. Table 1 summarizes the results of the control and the EDA variance experiments in terms of storm central pressure and maximum wind intensity.

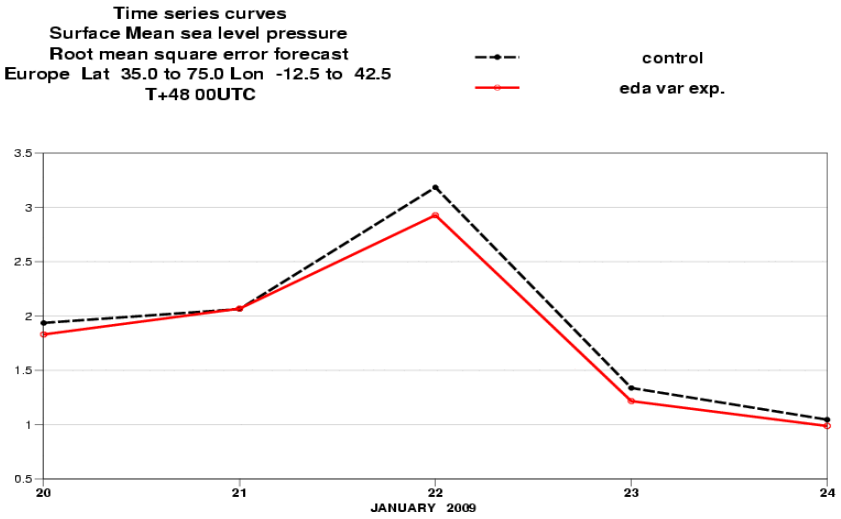

Figure 12: Skill scores of t+48h forecasts of mean sea level pressure over Europe for the period between 20090120 and 20090124, for the control run and for the experiment which used background-error variances derived from the 10-member EDA.

Time range/ Experiment	Ctrl	ENS10
96h forecast	983.5 hPa - 24.6 ms^{-1}	980,8 hPa - 24.6 ms^{-1}
72h forecast	982.8 hPa - 25.6 ms^{-1}	982,3 hPa - 25.4 ms^{-1}
48h forecast	992.0 hPa - 16.7 ms^{-1}	988,4 hPa -17.0 ms^{-1}
24h forecast	975.0 hPa - 24.9 ms^{-1}	975,5 hPa - 24.9 ms^{-1}

Table 1: Central pressures of the storm Klaus, forecasted by the control run and the 10-member EDA variance experiment, for 96h to 24h time ranges.

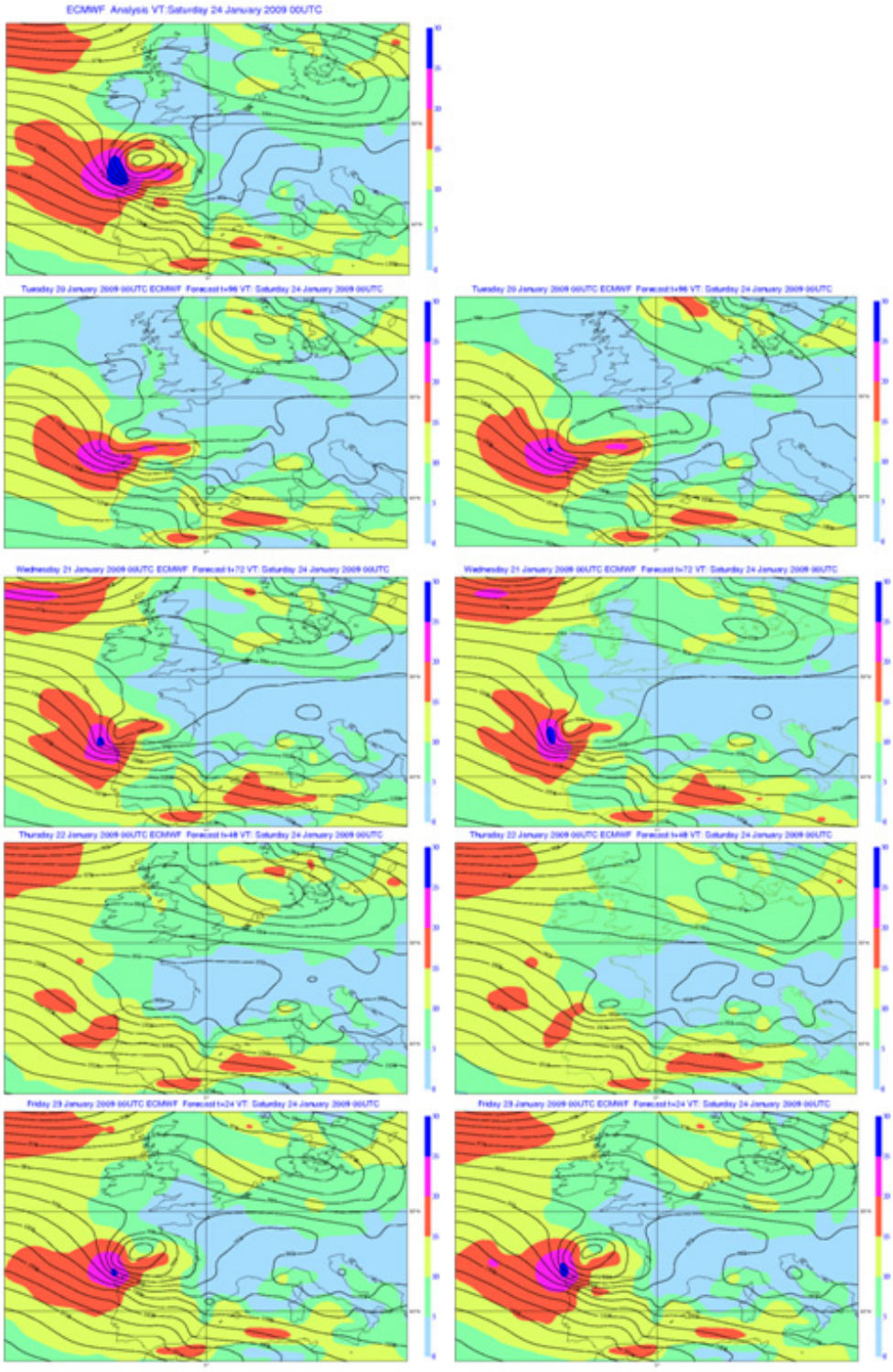

Figure 13: (a) Operational ECMWF analysis of MSLP (contour interval: 4hPa) and 10 metre wind speed ($m.s^{-1}$) valid on 24 January at 00 UTC, central pressure of the low is 967,6hPa and the maximum wind speed is 29.5 $m.s^{-1}$. (b) - (i) Forecasts of MSLP and 10 metre wind speed valid on 24 January at 00 UTC: control run (left panels) and 10 member EDA variance run (right panels). (b) Ctrl 96h-forecast : 983.5 hPa, 24.6 $m.s^{-1}$; (c) ENS10 96h-forecast : 980.8 hPa, 24.6 $m.s^{-1}$; (d) Ctrl 72h-forecast : 98.,8 hPa, 25.6 $m.s^{-1}$; (e) ENS10 72h-forecast : 982.3 hPa, 25.4 $m.s^{-1}$; (f) Ctrl 48h-forecast : 992 hPa, 16.7 $m.s^{-1}$; (g) ENS10 48h-forecast : 988.4hPa, 17 $m.s^{-1}$; (h) Ctrl 24h-forecast : 975 hPa, 24.9 $m.s^{-1}$; (i) ENS10 24h-forecast : 975.5hPa, 24.9$m.s^{-1}$.

4.2 Hurricane Ike (1-14 September 2008)

Hurricane Ike was the most intense hurricane of the 2008 Atlantic hurricane season and one of the costliest to make landfall in the continental USA, with maximum sustained winds of 64 m/s and an estimated central pressure of 935 hPa. Starting as a tropical disturbance off the coast of West Africa, Ike intensified to become a tropical depression on the 1st of September 2008 and further strengthened to hurricane status on the 3rd of September. When it made landfall in Cuba Ike had reached category 4 status, while it was down to a category 2 hurricane on landfall in Galveston, Texas (Fig. 14, 15).

In figure 16 three snapshots of the first-guess EDA standard deviations for vorticity at model level 64 (~500 hPa) are shown. They capture three different stages of the hurricane development: hurricane formation in the Mid-Atlantic (4/09/08, 0900 UTC), landfall in Cuba (7/09/08, 0900 UTC) and landfall in Texas (12/09/08, 0900 UTC). It is apparent that even the 10-member ensemble (left column) is able to identify and track the main error features of the hurricane. The most striking difference with the 50-member ensemble (right column) is the considerably higher level of sampling noise present in the 10-member ensemble. As a consequence, the spatial filtering algorithm needs to suppress smaller scale features much more aggressively in the 10-member variances in order to retain the statistically robust signal. This can be seen in figure 17, where the objectively filtered EDA standard deviations are shown for the same verification times. It is apparent that the filtered 10-member EDA (left column) is able to resolve only the larger scale features of the hurricane error structures, while the errors' magnitude is severely reduced. On the other hand, the filtered 50-member EDA (right column) allows a much more precise characterization of the relevant error structures.

A general conclusion that may be drawn from the discussion of the two case studies is that the objective filtering technique of Raynaud *et al*., 2009, allows for the statistically significant part of the EDA variance signal to be extracted from the raw ensemble first-guess forecasts. However, the higher level of sampling noise of a small ensemble leads the filter to being much more scale selective, thus retaining only the larger scale features of the raw EDA variances. This approximation is still acceptable when modelling relatively large scale synoptic features linked to extra-tropical cyclogenesis, but it is clearly inadequate if we aim to represent smaller scale meteorological events such as tropical hurricanes. In this sense, the objective filtering technique should not be seen as a substitute for employing as large an ensemble size as operational constraints allow.

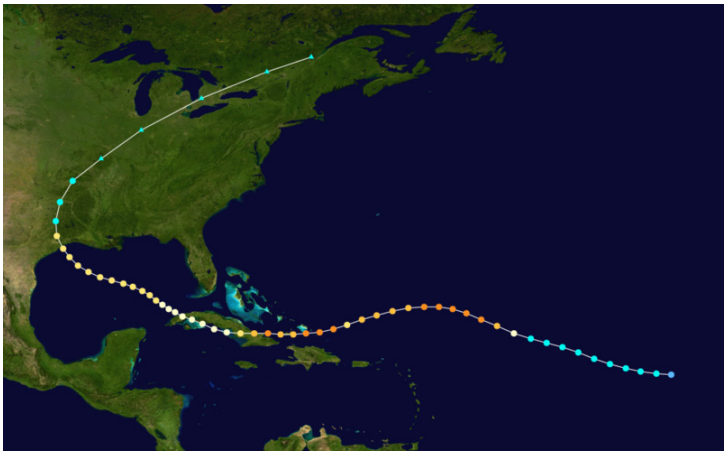

Figure 14: Hurricane Ike (1-14 September 2008) track. Colour code scheme from the Saffir-Simpson Hurricane Scale. Source: Wikimedia Commons from data from NASA and National Hurricane Center.

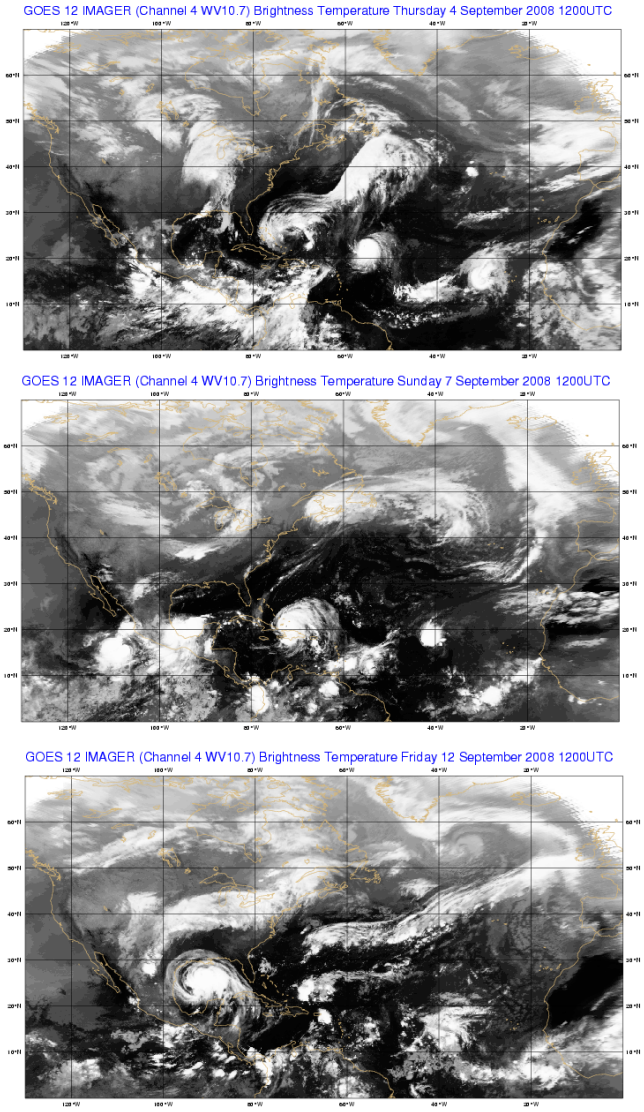

Figure 15: Hurricane Ike in the mid-Atlantic (4/09/2008, 12 UTC), at landfall in Cuba (7/09/2008, 12 UTC), at landfall in Texas (12/09/2008, 12 UTC). Data from NOAA.

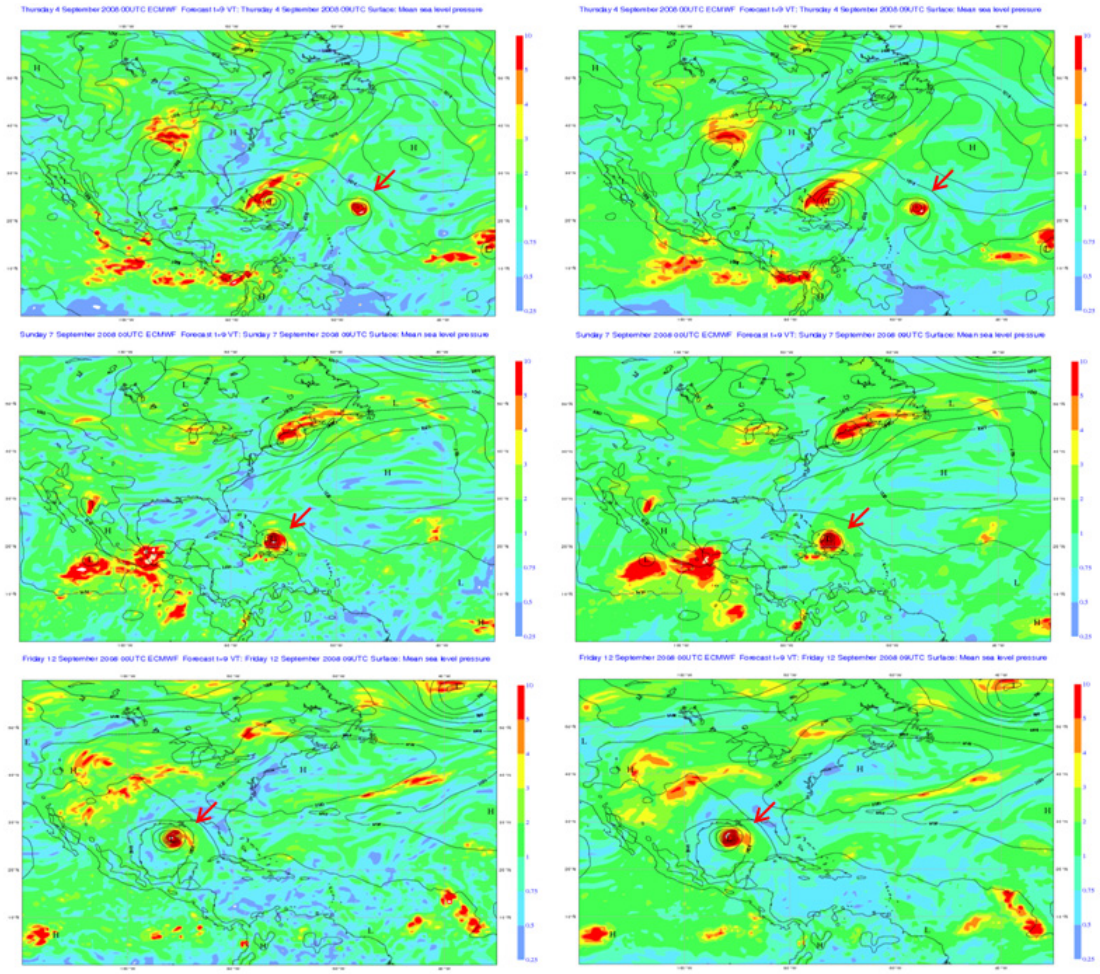

Figure 16: Standard deviations of vorticity at model level 64 (close to 500 hPa) valid on 4 September 2008, 09 UTC (first row); 7 September 2008, 09 UTC (second row); 12 September 2008, 09 UTC (third row). Unit: 5.10^{-5} s^{-1}. Raw estimates calculated with: 10-member ensemble (first column); 50-member ensemble (second column). The mean sea level surface pressure analysis is overlaid,5 hPa contour. Position of Hurricane Ike is highlighted by red arrow pointer.

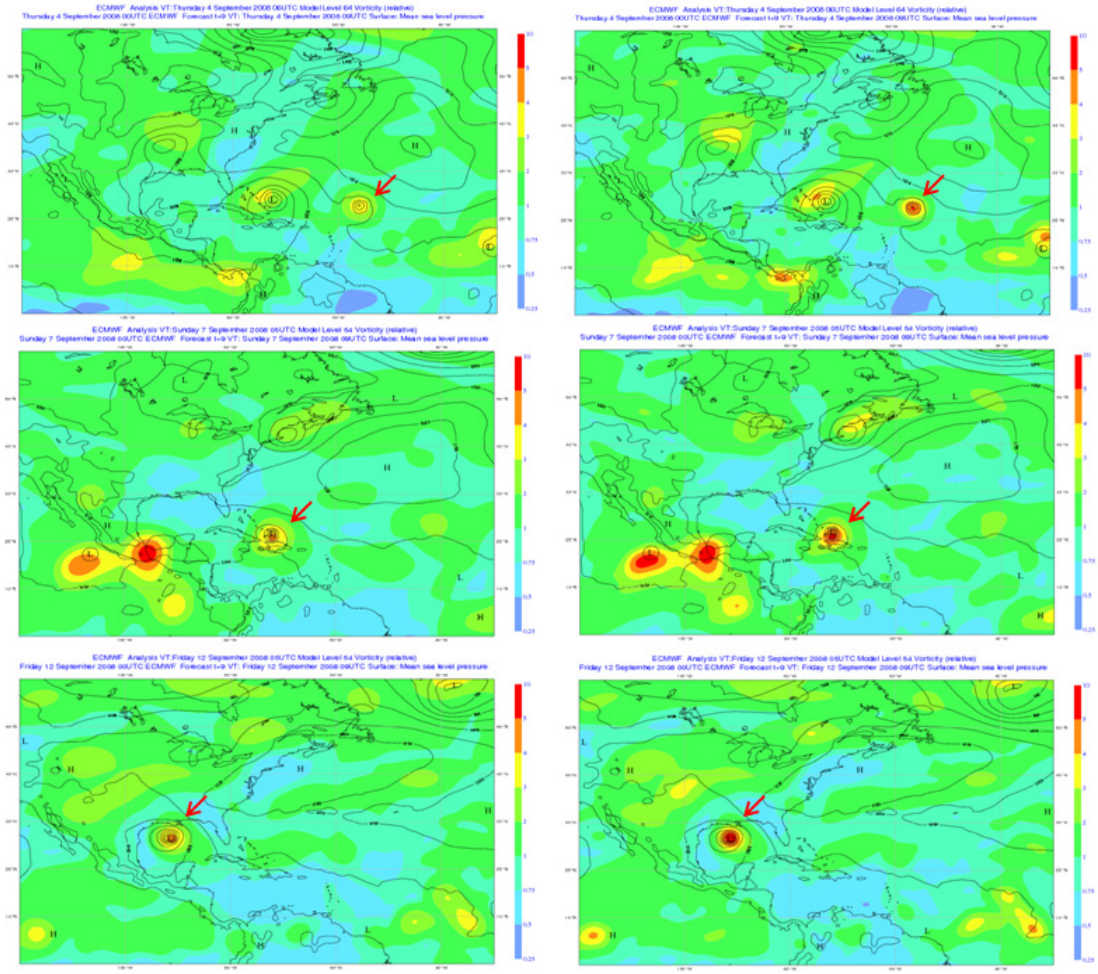

Figure 17: Standard deviations of vorticity at model level 64 (close to 500 hPa) valid on 4 September 2008, 09 UTC (first row); 7 September 2008, 09 UTC (second row); 12 September 2008, 09 UTC (third row). Unit: 5.10^{-5} s^{-1}. Filtered estimates calculated with: 10-member ensemble (first column); 50-member ensemble (second column). The mean sea level surface pressure analysis is overlaid,5 hPa contour. Position of Hurricane Ike is highlighted by red arrow pointer.

5 Daily variability of the ensemble-based variance filtering

The objective filtering technique of Raynaud et al. (2009), which has been implemented in the ECMWF EDA system, is performed with the application to the raw spectral EDA variance fields of the spectral low-pass filter defined by:

$$\rho(n) = \left[\cos\left(0.5 * \pi * min\left(n, N_{trunc}\right) / N_{trunc}\right)\right]^2 \qquad (1)$$

where n is the total spectral wavenumber and N_{trunc} is the truncation wavenumber of the filter. The application of this low-pass filter in spectral space can be seen as a weighted spatial averaging technique in grid-point space, which enables the large-scale signal of interest to be extracted while filtering out the small-scale sampling noise. The essence of the algorithm is in the computation of N_{trunc}, which is generally a function of model variable and vertical model level. This objective truncation wavenumber is determined according to the estimated signal-to-noise ratios of the ensemble first-guess raw variance fields. The sampling noise energy spectrum is based on the Legendre transform of a time- and space-averaged estimate of the ensemble raw covariances (i.e., Equ. (6) in Raynaud et al., 2009), while the signal power spectrum is computed on the sampled EDA raw variances. This time-adaptive filtering can thus adjust the truncation in real time to take into account the spatio-temporal evolution of the sampled variances. However, the daily variations of the sampled variances power spectrum can lead in some particular cases to an erratic behaviour of the filter. An example of this is shown in figure 18, which presents the energy spectra of temperature variances near the surface for two consecutive ensemble runs, the 20th of January at 0900 UTC (in black) and 2100 UTC (in red), along with the climatological noise spectrum. While the two variance spectra look rather similar, the algorithm for computing the truncation wavenumber (based on the idea of fitting the analytical filter defined by equation (1) to the calculated filter, which is derived from spectral signal-to-noise ratios, Raynaud et al., 2009) ends up producing two very different truncation numbers, respectively 160 and 71. This is a clear indication that a simple, adaptive computation of the truncation wavenumber can lead to unstable estimates for parameters and model levels where day-to-day variability is non negligible (this is another kind of sampling noise, this time with respect to the estimation of the signal power spectrum).

To clarify this issue, the objective filtering technique has been applied to a one-month long sample of EDA first-guess error variances valid at 0900 and 2100 UTC. In this way, we obtain a 60-member statistical sample of truncation total wavenumbers for each model variable and vertical level. Figure 19(a) shows the vertical profile of the sample median truncation values for vorticity, along with the first and third quartiles. First, similarly to Raynaud et al. (2010), one can notice that the truncation tends to decrease with altitude, apart from a small increase in the lower stratosphere (model level 39 approximately corresponds to 100 hPa). The sample variability of the truncation appears to be rather limited, especially up to model level 23 (20hPa). The case of temperature (Figure 19(b)) is somewhat different: the truncation is subject to strong daily variations in the boundary layer and between model levels 56 (300hPa) and 34 (70hPa) with an interquartile range than often exceeds 100. Moreover, this large variability does not seem to have a plausible physical explanation, but rather to be an artefact of the way we estimate the truncation wavenumber.

We are thus faced with the problem of finding a more robust estimator of the truncation wavenumbers. Recomputing the sample truncation wavenumbers over a summer (August-September 2008) case did

not show noticeable seasonal variations to the estimates in figure 19(a,b). This indicates that the monthly-averaged values provide sufficiently robust estimators of underlying sample statistics. This idea has been verified by performing two analysis/forecast experiments using 10-member EDA variances, over a January-February 2009 period. Experiment ENS10/FT uses daily varying filters for each model variable and vertical level, whereas the experiment ENS10/FM uses time-stationary filters, which are calculated using the median values of the monthly truncation wavenumbers' distribution. A control run using the currently operational randomization technique (Fisher and Courtier, 1995) to compute the initial background-error variances was also performed. Both experiments using EDA variances show some improvement over the control, which suggests that there is an advantage (more pronounced for the Southern Hemisphere) in using a constant filter instead of a time-dependent filter. For example, forecast scores over the Southern Hemisphere for 500hPa geopotential height (Figure 20) show an improvement with the constant filter, especially in the latter part of the forecast range.

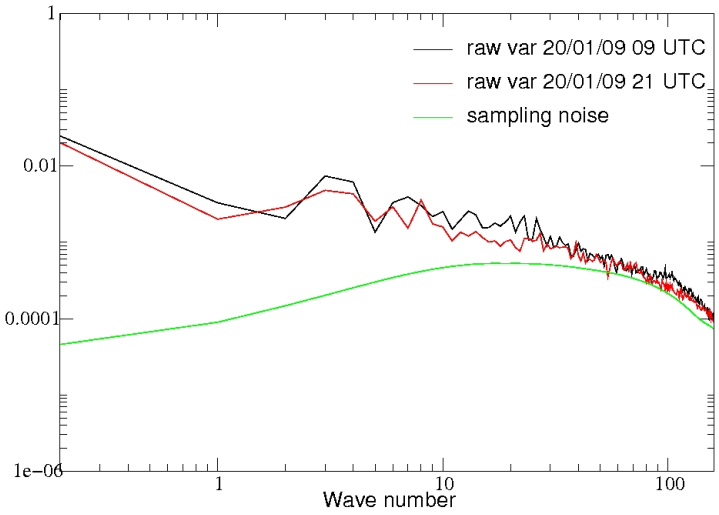

Figure 18: Energy spectra of temperature variances near the surface calculated with a 10-member ensemble. The verification date is 20th January 2009 at 09 UTC (in black) and 21 UTC (in red). The associated sampling noise spectrum is shown in green. Unit: K^{-2}.

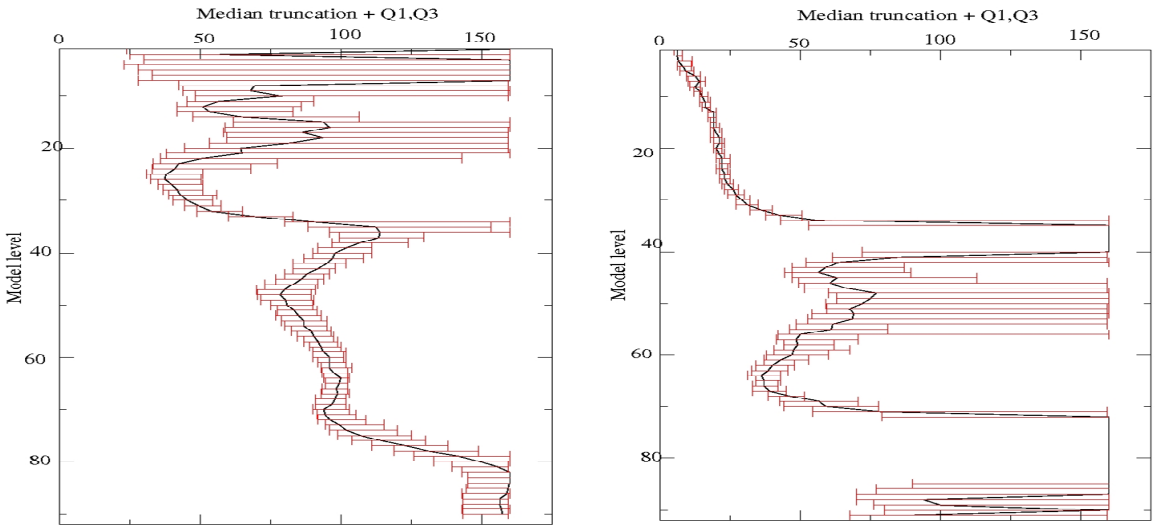

Figure 19: Vertical profiles of objective truncation for the filtering of raw ensemble-based variances of (a) vorticity and (b) temperature. In black: median truncation, in red: first and third quartiles.

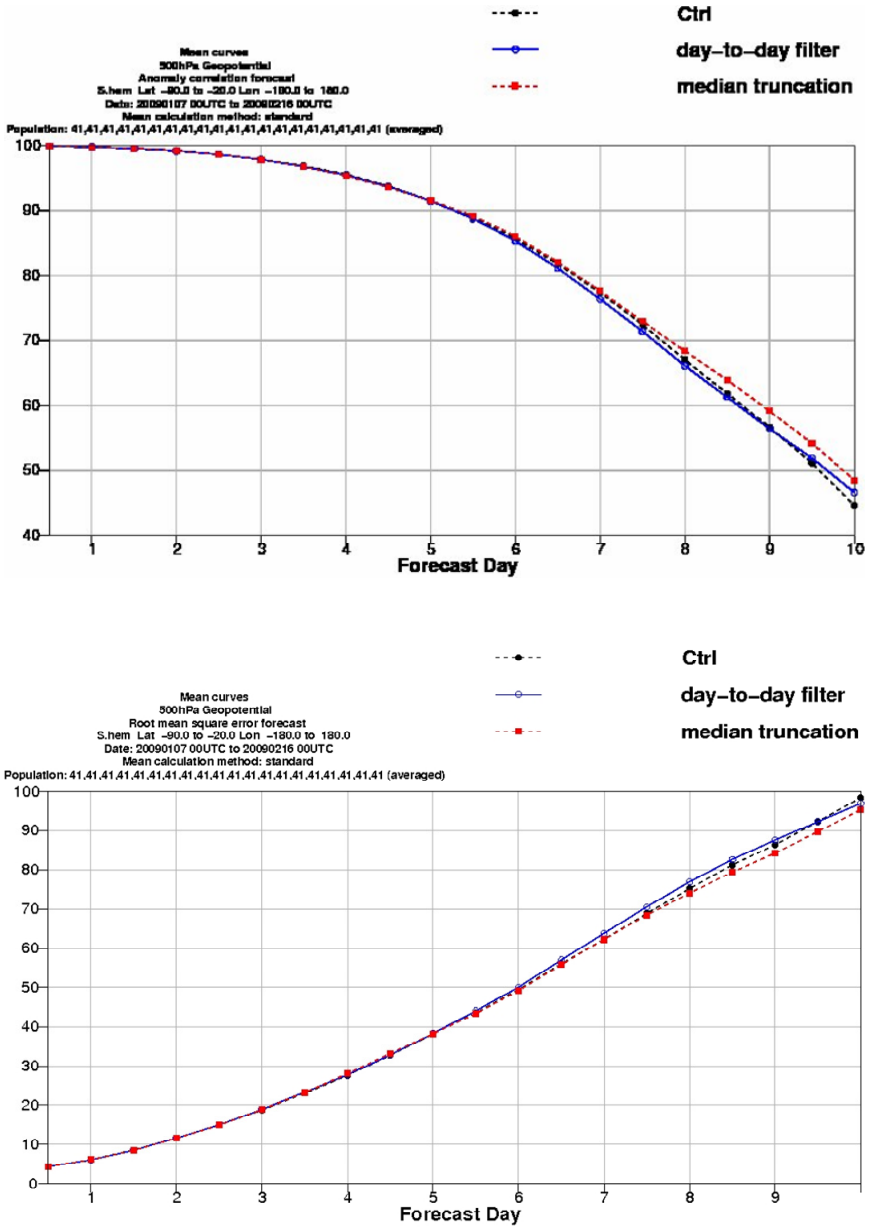

Figure 20: Forecast scores for 500hPa Geopotential height over the Southern Hemisphere for the winter period between 20090107 and 20090216 for the experiments which used ensemble-based background-error variances from a 10-member ensemble with (ENS10/FT, in blue) day-to-day variations of the filter, and (ENS10/FM, in red) a median truncation. The control run, which uses the error variances calculated with the current operational method, is shown in black. (a) Anomaly correlation and (b) rms error forecast.

6 CONCLUSIONS

The main goal of this study was to assess the robustness of the background error estimates that are diagnosed from the forecast variances of an ensemble of perturbed data assimilations. Two aspects have been evaluated: a) the sensitivity of the error estimates to the ensemble size and b) the sensitivity of the filtering procedure to the variances' daily variability.

As regards the first aspect, the results presented here broadly confirm the validity of conclusions from previous work on the subject by Kucukkaraca and Fisher (2006), Isaksen et al. (2007) and Raynaud et al. (2008, 2009). Namely, it was shown that a fairly accurate estimate of background-error variances can be obtained with a limited number of ensemble members (10 in the present case). In particular, the uncertainty in the background state associated with an extra-tropical severe storm event was seen to be well captured by an ensemble of only 10 members.

However, our investigation has shown that there are some caveats attached to the use of small ensembles. The filtering technique employed to extract the statistically significant signal from the ensemble variances drastically reduces the effective spatial resolution of the ensemble with respect to the nominal one (for vorticity at model level 64, effective truncation wavenumber is around T80, versus the T399 nominal resolution). Furthermore, this aggressive filtering inevitably affects the spectrum of the signal we would like to retain. These much coarser filtered ensemble variances may still have enough detail to be able to resolve the large scale uncertainties in the extra-tropical flow. On the other hand, when we are faced with more localised anomalies such as those associated with a tropical disturbance, it can be argued that a filtered 10-member ensemble is only partially able to resolve the error features of interest and the advantage of a larger ensemble is more noticeable. From a more quantitative perspective, the plot in figure 1 would seem to imply that some saturation of the useful signal that can be extracted from the ensemble starts to appear with a 40-member ensemble, in terms of global amplitude. On the other hand, the discussion in section 2 on the effective spatial resolution of the ensemble points to further possible gains with larger ensemble sizes, in terms of local error characterization. Assimilation experiments with a 20-member EDA seem to confirm the benefit on forecast skill scores of using a larger ensemble to characterize background errors. Another possible approach to deal with error structures of different scales, and in particular to extract localised patterns of useful signal may be to develop a spatial filter that depends on the geographical position. This could be achieved with a filter in grid-point or wavelet space for instance.

If the benefit of using a larger ensemble is apparent in the estimation of background-error variances, a fortiori this must be the case for the much larger-dimensional problem of finding robust estimates of background-error correlations. Our expectations are that in this second case the role of the filtering procedure will become even more crucial. In this context a promising path of research would envisage the use of the wavelet formulation of the background error matrix (Fisher, 2003).

The second part of the study has addressed the filter ability to deal with the day-to-day variations in the EDA variances. It was shown that the basic filter configuration could not do a good job at handling relatively small changes in the EDA variances' power spectra, resulting in significantly different truncation wavenumbers in consecutive analysis cycle. A possible solution was identified in the use of the median truncation values from their monthly climatological distribution. This was implemented in the ECMWF EDA system and it was shown to have a largely positive impact on forecast skill scores.

ACKNOWLEDGMENTS

Part of this work has been conducted during a monthly stay of Laure Raynaud at ECMWF. Laure Raynaud acknowledges Météo-France and ECMWF for their financial support, Jean-Noël Thépaut and Lars Isaksen for their help in organizing the stay and Massimo Bonavita for his help with the ECMWF EDA system and for useful discussions.

REFERENCES

Belo Pereira, M. and L. Berre, 2006: The use of an ensemble approach to study the background-error covariances in a global NWP model. *Mon. Weather Rev.*, **134**, 2466–2489.

L. Berre and G, Desroziers, 2010: Filtering of background error variances and correlations by local spatial averaging: a review. Accepted for publication in *Monthly Weather Review*.

Bormann, N., S. Saarinen, G. Kelly and J.-N. Thépaut, 2003: The spatial structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data, *Mon. Weather Rev.*, **131**, 706–718.

Buizza, R., M. Leutbecher and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051-2066.

Fisher, M., 2003: Background error covariance modelling. *Proceedings of the ECMWF Seminar on recent developments in data assimilation for atmosphere and ocean*, ECMWF, pages 45–63. (Available from: http://www.ecmwf.int/publications/)

Fisher, M., 2007: The sensitivity of analysis errors to the specification of background error covariances. *Proceedings of the ECMWF Workshop on flow-dependent aspects of data assimilation* ECMWF, pages 27–36. (Available from: http://www.ecmwf.int/publications/)

Fisher, M. and Courtier, P., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Technical Memorandum* 220. (Available from: http://www.ecmwf.int/publications/)

Isaksen, L., M. Fisher and J. Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. *Proceedings of the ECMWF Workshop on flow-dependent aspects of data assimilation*. ECMWF, pages 65–86. (Available from: http://www.ecmwf.int/publications/)

Isaksen, L., J. Haseler, R. Buizza and M. Leutbecher, 2010: The new Ensemble of Data Assimilations. *ECMWF Newsletter* No. 123, pages 22–27. (Available from: http://www.ecmwf.int/publications/)

Kucukkaraca, E., and M. Fisher, 2006: Use of analysis ensembles in estimating flow-dependent background error variances. *ECMWF Technical Memorandum*, 492. (Available from: http://www.ecmwf.int/publications/)

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A Weisheimer, 2009: Stochastic Parametrization and model uncertainty, *ECMWF Technical Memorandum*, 598. (Available from: http://www.ecmwf.int/publications/)

Raynaud, L., L. Berre, and G. Desroziers, 2008: Spatial averaging of ensemble-based background-error variances. *Quart. J. Roy. Meteor. Soc*., **134**, 1003–1014.

Raynaud, L., L. Berre, and G. Desroziers, 2009: Objective filtering of ensemble-based background-error variances. *Quart. J. Roy. Meteor. Soc*., **135**, 1177–1199.

Raynaud, L., L. Berre, and G. Desroziers, 2010: An extended specification of flow-dependent background-error variances in the Météo-France global 4D-Var system. Submitted *to Quart. J. Roy. Meteor. Soc.*

Vialard J., F. Vitart, M. A. Balmaseda, T. N. Stockdale and D. L. T. Anderson, 2005: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Mon. Weather Rev*., **133**, 441–453.