

EnKF and Hybrid Gain Ensemble Data Assimilation

Mats Hamrud, Massimo Bonavita
and Lars Isaksen

Research Department

September 2014

To be submitted to Monthly Weather Review

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.



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

© Copyright 2014

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

The wish to do detailed comparisons between variational and more scalable ensemble-based data assimilation systems in a semi-operational environment has led to the development of a state of the art EnKF system at ECMWF. A broad description of the ECMWF EnKF is given in this paper, focusing on highlighting differences compared to standard EnKF practice. In particular, a discussion of the novel algorithm used to control imbalances between the mass and wind fields in the EnKF analysis is given. The scalability and computational properties of the EnKF are reviewed and the implementation choices adopted at ECMWF described. The sensitivity of the ECMWF EnKF to ensemble size, horizontal resolution and representation of model errors is also discussed. The performance of the EnKF system has been compared to a 4DVar of similar resolution. It is found that there is not a major difference between the forecast skill of the two systems. However, like the operational hybrid 4DVar-EDA, a hybrid EnKF-Variational system (which we refer to as “Hybrid Gain Ensemble Data Assimilation”, HG-EnDA) is capable of significantly out-performing both component systems. The HG-EnDA has been implemented with little effort following Penny (2014). Results of numerical experimentation comparing the HG-EnDA with the hybrid 4DVar-EDA used operationally at ECMWF are presented.

1 Introduction

The operational data assimilation system at ECMWF is based on a 4DVar algorithm (Rabier et al., 2000) in a hybrid formulation, with flow-dependent background errors and structure functions estimated by an ensemble of lower-resolution, perturbed 4DVar assimilations (Ensemble of Data Assimilations, EDA; Isaksen et al., 2010; Bonavita et al., 2012; Bonavita et al., 2014): this system will be referred to in the following as hybrid 4DVar-EDA. An experimental ensemble Kalman filter (EnKF) data assimilation system for the atmosphere has also been developed at ECMWF. The main motivation for this development is to have an alternative data assimilation system in-house that allows us to do detailed comparisons between the ECMWF 4DVar system and the EnKF. The goal is to highlight the strengths and limitations of both algorithms in a testing environment that uses the same forecast model and observations, and is close to the operational setup. An earlier comparison between the ECMWF 4DVar system and a NOAA EnKF system was presented by Whitaker et al. (2009). This comparison was done in the context of a reanalysis scenario, using only surface pressure observations. The two systems used different forecast models and different ways of computing the observation equivalents. It is more difficult to compare analysis systems with a comprehensive observation network in this way because the differences in forecast models, observation processing and bias correction would severely obscure the comparison findings.

The availability of a complete EnKF system makes it possible to compare the operational hybrid 4DVar-EDA with a hybrid 4DVar-EnKF system. Hybrid variational-EnKF systems have been used in other global NWP Centres (Buehner et al., 2010a,b; Bishop and Hodyss, 2011; Clayton et al., 2013). It is worth noting that the EDA provides flow-dependent covariances for the 4DVar at ECMWF (Bonavita et al., 2014). This is theoretically rigorous and practically effective, but it is also relatively expensive due to the necessity of computing a separate analysis for each ensemble member. On the other hand, the use of a “deterministic” flavour of the EnKF (Tippett et al., 2003) entails the computation of only one Kalman gain for the whole ensemble: this allows substantial computational

savings and the possibility of using a considerably larger ensemble size. Some initial experimental results of a hybrid 4DVar-EnKF will be presented in this paper.

Another important aspect concerns the scalability properties of future data assimilation systems. Current trends in the evolution of supercomputer architectures point to the need of increased levels of parallelism in future assimilation algorithms (Isaksen, 2012). The EnKF algorithm is particularly suited to future, massively-parallel computer architectures, since the forecast stage of the algorithm is intrinsically parallel and the state update is also, at least from a theoretical stand-point, scalable. In practice there are some limitations to the scalability of the algorithm which will be discussed below. The strong-constraint formulation of 4DVar used at ECMWF computes the model trajectory which best fits the background and the observations over the assimilation window by minimizing a cost function measuring the misfit between the atmospheric state and the observational and background information. This minimization requires repeated ($O(30-50)$), sequential integrations of the tangent linear and adjoint models and thus presents itself as a bottleneck for the scalability of the algorithm. It is, however, important to note that the saddle-point formulation has been shown, in simplified systems, by Fisher and Gürol (2014) to improve scalability significantly by introducing a parallel-in-time capability in 4DVar.

A broad description of the EnKF data assimilation system developed at ECMWF is given in Section 2. The description is focused on highlighting differences in the ECMWF implementation compared to standard EnKF practice. In particular, a more in depth discussion of the algorithm used to control imbalances between the mass and wind fields in the EnKF analysis is given.

In Section 3 we discuss the scalability and computational properties of the EnKF. It is widely accepted that some implementations of the EnKF are, in principle, perfectly scalable. However, the scalability of practical implementations of the EnKF will tend to be limited by input/output constraints.

In the initial implementation of the EnKF, a number of experiments were conducted to determine the appropriate values for various parameters of the filter. In the majority of cases, results have been quite similar to what has been already reported in the literature. We therefore only discuss a small subset of experiments designed to test the sensitivity of our implementation of the EnKF to: a) ensemble size, b) ensemble spatial resolution and c) covariance inflation and model error parameterizations. Results of this experimentation are presented in Section 4.

The performance of the EnKF system has been compared to a similar resolution 4DVar using the same observational information. It is found that, similar to Buehner et al. (2010b), there is not a major difference between the forecast skill of the two systems. However, a hybrid EnKF-Variational system, which we refer to as “Hybrid Gain Ensemble Data Assimilation”, HG-EnDA, is capable of outperforming both systems; this will be described in detail in Section 5.

The adoption of the HG-EnDA system for the estimation of analysis and background errors and even for the estimate of the initial conditions for NWP offers a range of exciting new possibilities for further development. Some of the possible avenues of research are discussed in Section 6.

2 Description of the ECMWF EnKF system

The EnKF system developed at ECMWF is to a large extent based on operators and work flow developed and used by the operational hybrid 4DVar-EDA data assimilation system, especially with

the EDA component (ECMWF, 2014; Isaksen et al., 2010). The model propagating the ensemble is the same and the operators calculating the observation equivalents are the same, in both cases using the ECMWF Integrated Forecasting System (IFS) code base. The observation Quality Control (QC) and data selection procedure are also similar, using the same kind of first-guess checks and data thinning. For the EnKF these decisions are taken on the basis of the control forecast (the forecast from the mean analysis, in our case) and the estimate for the background uncertainty used by the QC comes from the spread of the short-range (3-hour) ensemble forecast.

In the operational analysis we employ Variational Quality Control (VarQC) in conjunction with the use of the Huber norm (Tavolato and Isaksen, 2014). In the context of 4D-Var this is easy to implement and comes at a small computational cost. VarQC has been shown to have a small but clear positive impact on the analysis quality (Anderson and Järvinen, 1999; Tavolato and Isaksen, 2014). We have implemented an Adaptive Quality Control in the EnKF, using the Huber norm and the same parameter settings as in the operational 4DVar quality control, which in theory should achieve similar results (Jeff Whitaker, pers. comm.). This Adaptive Quality Control in the EnKF context comes, however, with a substantial computational cost as it involves iteratively performing an analysis in observation space and adjusting the observation errors. Because of this cost, most of the EnKF experimentation has been performed with this feature turned off (this also implies using approximately 10% higher observation errors than in operations, as the operational observation errors are tuned to be used in conjunction with Huber norm VarQC). More recent experimentation has been done with Adaptive Quality Control enabled as it facilitates comparisons with 4DVar experiments (as a similar number of observations are retained in both analyses); also at higher analysis resolutions, its relative cost decreases.

Apart from the features described above, the EnKF system also utilizes large parts of the technical infrastructure of the operational system at ECMWF. The observations are stored within ECMWF's Observational Data Base (ODB, Saarinen, 2004; Fouilloux, 2010), field data are exchanged in WMO GRIB format and the task scheduling is performed in a similar way to that of the EDA suite, using the ECMWF Supervisor Monitor Scheduler (SMS, ECMWF, 2011). The post-processing of the analysis, including producing pressure level data, is performed by the IFS.

Two main flavours of deterministic EnKF have been implemented: the Ensemble Square Root Filter (EnSRF, Whitaker and Hamill, 2002) and the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007). Early experimentation showed that, in terms of time-averaged skill scores, there was no statistically significant difference in performance between the two schemes, so that other considerations (i.e., scalability characteristics) need to be made for choosing one algorithm versus the other. Some differences between the two schemes were noted in single-observation experiments, but these were traced to the different types of covariance localization techniques that need to be employed in the two algorithms; i.e. covariance localization through a Schur product of the empirical covariance with a decaying function of distance for the EnSRF (**B** localization; Houtekamer and Mitchell, 2001), and domain localization (Evensen, 2004; Ott et al., 2004) in the LETKF, where the observations are further weighted according to their distance to the analysed grid point (**R** localization; Hunt et al., 2007). In both schemes the distance-dependent decorrelation function is implemented using the compact piecewise polynomial approximation of a Gaussian localization function of Gaspari and Cohn (1999). As noted by other authors (Greybush et al., 2011), the effective localization length scale in the **B** localization is smaller than for the **R** localization for the same localization function, leading to broader, smoother increments for the LETKF in single observation tests. While these results would

point to the need for broader decorrelation functions in the EnSRF, experimental results with the full observing system (detailed in Section 3) show a remarkable insensitivity to the variation of the horizontal decorrelation length scale in a rather broad range. It has also been found useful, in the LETKF context, to perform an implicit covariance localization by limiting the number of observations affecting the analysis of any grid point. This number is currently set at 30 for each combination of different report type and observation variable, where the observations spatially closest to the analysed grid point are selected. This was found to produce a significant advantage in terms of forecast scores in the northern hemisphere (between 3% to 5% in the forecast anomaly correlation metric in the troposphere) compared to using all available observations in the analysis update; at the same time this allows massive computational savings. Results, once again, are remarkably insensitive to the precise number of selected observations (i.e., doubling or halving the number of locally selected observations has a marginal impact on analysis skill). As in most ENKF implementations two complementary types of covariance inflation have been implemented, i.e. multiplicative and additive inflation (Whitaker and Hamill, 2012). The multiplicative inflation uses relaxation towards the background variance. Experiments have shown that the EnKF skill does not appear to be very sensitive to the relaxation factor; normally a factor of 0.9 has been used in the experimentation. The additive inflation is based on the difference between the 48-hour and 24-hour IFS forecasts verifying at the same time sampled over a 6 years period retrieved from ECMWF's operational archive. A random sample of these forecast differences is added to each ensemble member at each analysis cycle. The scaling factor applied is 0.25, unless otherwise stated; this converts the forecast difference to a nominal 6 hour difference.

Most satellite observations and their observation operators are affected by non-negligible biases of which data assimilation algorithms need to be aware (Dee, 2005). This is achieved in variational schemes by defining a simple, adaptive model of the observational bias whose parameters are estimated together with the model state (Derber and Wu, 1998). In the ECMWF implementation, the bias model is typically a linear function of air-mass and scan position, different for each satellite sensor (Dee, 2004) and its parameters are part of the 4DVar control vector. The main difference with respect to a non-bias-aware 4DVar is the presence of the bias correction term in the observation part of the cost function. In terms of computational cost, the overhead of variational bias correction is essentially insignificant in terms of the whole 4DVar cost. In the EnKF context, an equivalent scheme can be implemented but at a non-negligible computational cost (Fertig et al., 2009; Miyoshi et al., 2010). We have not implemented such a scheme; instead we have relied on the stored bias corrections produced by the 4DVar system. As mentioned in the introduction, the aim of this work was not to produce a pure EnKF system to replace the current 4DVar system; in the context of the hybrid system presented in Section 4 the bias correction can be performed by the 4DVar component of the system.

The analysed atmospheric variables in the ECMWF EnKF are temperature, wind vector components (u,v), specific humidity and surface pressure, plus surface pressure tendency. Except for the addition of surface pressure tendency, whose rationale is explained below, this is a fairly standard set of analysis variables in the EnKF: some limited experimentation has also been performed on alternative sets of wind variables (i.e., streamfunction and velocity potential; vorticity and divergence) following ideas from Kepert (2009). Results, however, have been generally negative in terms of analysis and forecast accuracy.

It is well documented (Lorenc, 2003; Kepert, 2009) that EnKF analyses tend to have problems in representing the model dynamical balance constraints, thus leading to the excitation of spurious

inertia-gravity waves in the short-range forecast. The cause of this is normally attributed to the effect of covariance localization disrupting the mass-wind field geostrophic balance. Diagnostics within the ECMWF EnKF system do show evidence of quite significant inertia-gravity wave noise in forecasts made from the EnKF analysis with respect to the 4DVar analysis (Fig. 2). To counter this problem, EnKF systems often apply some form of initialization (Houtekamer and Mitchell, 2005), normally some form of Digital Filter Initialization (DFI; Lynch and Huang, 1992). The IFS model has the capability of performing DFI both in stand-alone mode and as an additional penalty term in the 4DVar cost function (Gauthier and Thépaut, 2001). Experimentation with different versions of the stand-alone DFI applied to the EnKF analysis did in fact reduce the imbalances but always at the cost of some loss of analysis quality as evaluated by forecast scores. One possibility would be to make use of incremental DFI; i.e. applying DFI to the analysed increments instead of to the full analysed fields, which has been shown to provide better results (Polaparavu et al., 2004). We have opted instead for a new, physically-based, method aimed at adjusting the horizontal wind divergence in each vertical column. To achieve this we compute the surface pressure tendency of the background forecast by integrating the mass divergence in the vertical, i.e.:

$$\frac{\partial p_s}{\partial t} = - \int_0^1 \nabla \cdot \left(V \frac{\partial p}{\partial \eta} \right) d\eta$$

where η is the vertical hybrid coordinate of the IFS. Then we analyse the surface pressure tendency as any other state variable. Finally we regard the analysed surface pressure tendency as the “truth” and adjust the divergence to correspond to the analysed surface pressure tendency. Obviously the system is completely underdetermined in how to distribute the integrated adjustment in the vertical. The choice has then been made to distribute the divergence increment according to the uncertainty of the wind analysis as diagnosed by the standard deviation of the ensemble. This “divergence adjustment” is performed for all ensemble members. The wind increments caused by this procedure are normally very small (much smaller than the analysis increment) and the details of how the distribution in the vertical is achieved seem not to be very important. The very significant change in the initial pressure tendency achieved by this procedure is illustrated in Fig. 1 (note that the values in the right-hand panel have been multiplied by a factor of 10 to be able to use the same contour interval). This method effectively imposes an additional dynamical balance constraint to the analysis and thus removes a large portion of the excess inertia-gravity waves from the short-range forecast. This is visible in Fig. 2, where we show the effect of applying the divergence adjustment method to the EnKF control (member 0) and member 25 on the evolution of the absolute surface pressure tendency in the subsequent forecast. From this plot there is also an indication that the EnKF control (which in our case coincides with the forecast from the EnKF mean analysis) is more “balanced” than an arbitrary EnKF member.

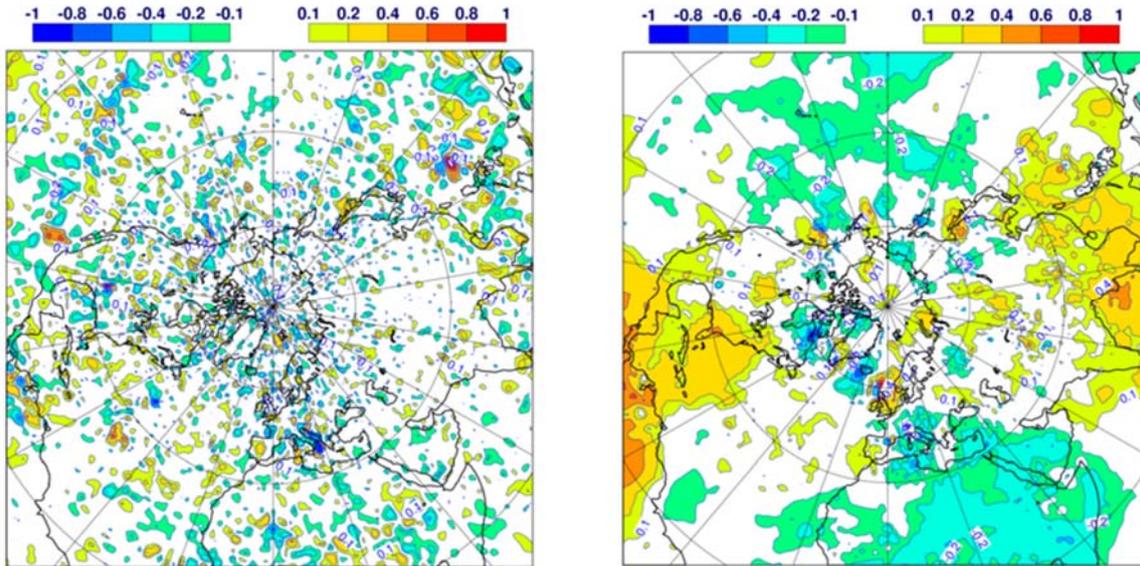


Figure 1: Initial surface pressure tendency over the northern hemisphere for the mean analysis. The left-hand panel shows the surface pressure tendency field before applying the divergence adjustment, the right-hand pane is from the final adjusted analysis. Note that the unit in the left-hand panel is [Pa/s] whereas in the right-hand panel it is [0.1 Pa/s], a difference of a factor of 10.

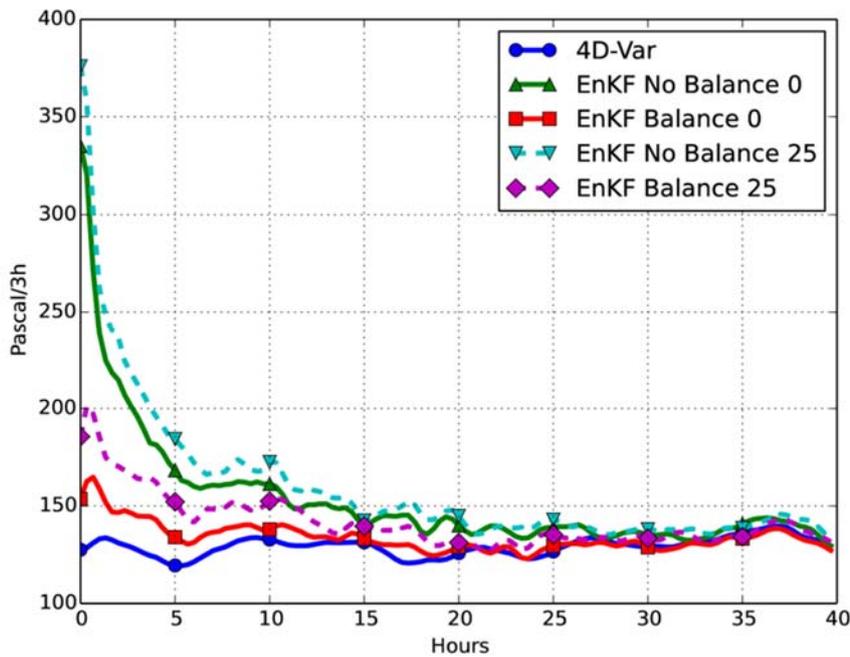


Figure 2: Evolution of the globally averaged absolute pressure tendency for members 0 (ensemble mean) and 25 of the EnKF before and after the application of the divergence adjustment procedure. The absolute pressure tendency for 4DVar is also shown for reference.

The impact on the analysis quality of this “divergence adjustment” is significant and positive. In the 6-hour forecast fit to observations we see an improvement in the fit for land-bases surface pressure observations of around 5% and for DRIBU surface pressure observations of around 10%. The effect

on the observation fit of other variables is negligible. As can be seen from Fig. 2, the ECMWF model quickly dampens inertia-gravity waves; this can mainly be attributed to the semi-implicit time stepping formulation of the model. This feature of the model explains why the impact of the “divergence adjustment” on the observation fit is not as dramatic as it could be inferred from Fig. 2.

3 Scalability and technical implementation

As mentioned earlier, there are no serious algorithmic constraints for the scalability of an EnKF data assimilation. At the forecast stage the limit is set by the scalability of the individual forecast members. The computation of the observation equivalents can again be done independently for all ensemble members; in the ECMWF EnKF system this is done by running forecasts over the observation window and storing the observation equivalents in the ODB. The EnKF analysis itself is local; in principle the limit of independent pieces of work is the size of the forecast state.

There exist however a number of practical limits to the scalability of the EnKF system. One issue is the Input/Output (IO) bottleneck. It is advantageous from both the scalability and resilience point of view to run the ensemble forecasts as separate jobs. In the analysis stage, however, all forecasts are needed to describe the background state, thus they have to be read in and distributed. At the end of the analysis all the analysis states have to be written out to provide the initial states for the next analysis cycle’s forecasts. In addition, all the observational data, including the observation departures for all ensemble members, have to be read in. This IO load can be distributed over tasks, which is what we do in the ECMWF EnKF implementation, but at scale the overall throughput of the IO subsystem still limits the scalability.

The practical scalability issues for the actual analysis stage are quite different for the EnSRF and LETKF analyses. For an EnSRF with a small number of observations it is quite easy to create a system with good load balance and no communications within the actual analysis stage by randomly distributing the state variables over processors and duplicating the observations on all processors. The update in observation space is then duplicated on all processors, this only makes sense if the number of observations is much smaller than the size of the state. For a large number of observations, in the ECMWF EnKF we typically assimilate about three million observations in a 6-hour window; this is not possible and the observations need to be distributed and the update in observation space done in a distributed fashion. The way to achieve good load balance is again to distribute the observations randomly. This leads to a sequential algorithm where the observation information with the updated observation model equivalents need to be broadcast from the processor that “owns” the observation to all other processors before the observation is ingested by the analysis. With three million observations, leading to three million broadcasts, this leads to an implementation severely constrained by message passing latency. In the ECMWF system we have tried to improve this by processing batches of observations before broadcasting them thus reducing the number of communication stages. Other options that could be explored to improve the scalability could include using asynchronous communications or finding truly independent tasks by a graph colouring method (see also Houtekamer et al., 2014, for a discussion of scalability issues in the context of a perturbed observation EnKF).

For the implementation of the LETKF, where the state is distributed in contiguous areas and the observations are distributed so that each task has local access to the observations it needs, the problem becomes that of load balancing. The cost of analysing a grid point in state space is dependent on the

number of close observations (Szunyogh et al., 2008) and as the observation density is not homogeneous, the cost varies for different parts of the globe. This means that a distribution with equal number of state points on each processor would lead to a poor load balance. To prevent this in our implementation, we measure the cost of performing the analyses of each state column as a function of the number of close observations and cycle this information within the analysis suite. Then, in subsequent analysis steps, we find out the number of close observations for each state profile in the global view and combine this information with the cycled empirical relation between observation density and computational cost to compute weights for the distribution of the state space variables. The grid points are then distributed so that each processor ends up with approximately the same computational weight rather than with the same number of grid points. The areas are still contiguous in space. We find that this method normally gives a reasonable load balance, with a ratio of the fastest to the slowest analysis time on an individual processor typically around 0.85.

For an operational centre like ECMWF it is not only the scalability of the analysis system that is of interest but also the total computational cost. The natural comparison for an EnKF implementation is ECMWF's EDA system, based on multiple independent copies of 4DVar (Isaksen et al., 2010). Here an EnKF implementation has an obvious advantage; the cost per member of an EnKF system is considerably lower than for the EDA mainly due to the fact that the Kalman Gain is only computed once in the EnKF.

4 Sensitivity experiments with EnKF system

While EnKF systems are relatively easy to set-up and tend to produce acceptable results with a broad range of "reasonable" choices of the main filter parameters, they have been found to be harder to tune for optimal performance than variational data assimilation systems. This is in part due to the cost of exploring the parameter space for realistic data assimilation configurations, and partly due to the self-adjusting properties of the filter itself. The sensitivity of the filter performance to a range of covariance localization parameters has been found to be small for values up to 50% different from the ones we are currently using (3000 km is the distance of zero impact of an observation on an analysis grid point due to spatial localization in the horizontal, and 2.5 scale heights in the vertical). Similarly, the precise values used in the tuning of the multiplicative and additive covariance inflation schemes were found to have a relatively small effect on global deterministic forecast skill.

Other parameters of the system were however found to have a measurable impact on the filter performance and are discussed in the following. All experiments discussed below have used the LETKF version of the EnKF, a 6-hour assimilation window and the Adaptive Quality Control for observations turned off.

a) Ensemble Size

Ensemble size is one of the most important configuration parameters that need to be defined in an EnKF implementation. Given the obvious importance of controlling sampling errors in the background error covariance estimates, the goal is to have as large an ensemble size as it is practical given the computational and operational constraints and the competing demands on the computational resources from other aspects of the EnKF configuration (i.e., spatial resolution, observation usage, etc.). It is thus of interest to have an idea of the sensitivity of the filter to the ensemble size within a realistic range of ensemble dimensions. A set of experiments was carried out in which the ensemble size was

set to 60, 120 and 240. The experiments were run with a triangular spectral truncation of 159 for the IFS model (~120 km grid spacing) and 91 vertical levels. All operationally available observations were used, with the exception of radiances from satellite imagers (SSM/I/SSMIS, AMSR-E, TRMM-TMI) and scatterometer winds, which, at the time of the experiments, had not yet been implemented in the EnKF.

Figure 3 gives examples of how the increase in ensemble size helps to better characterize sampled error correlations from the EnKF. The top panels of Fig. 3 show, as function of horizontal and vertical distance, the azimuthally-averaged sampled absolute correlations between 3-hour forecast errors of 500 hPa temperature and AMSU-A (Advanced Microwave Sounding Unit A; NASA, 2002) brightness temperature model equivalents for the 60 member EnKF (panel (a)) and the 120 member EnKF (panel (b)). The bottom panels of Fig. 3 show the azimuthally-averaged sampled absolute correlations between 500 hPa zonal wind 3-hour forecast errors and AMSU-A brightness temperature model equivalents for the 60 member EnKF (panel (c)) and the 120 member EnKF (panel (d)). In both instances it is apparent how the structure of the spatial correlations is much better defined in the 120 member ensemble, which also presents lower values of spurious long-range correlations. The reduction of the level of spurious long range correlations can also be appreciated from Fig. 4, where we present the absolute correlation of the short-range (3-hour) temperature forecast at 500 hPa with the radiosonde temperature observations at the same pressure level, as a function of distance. Note how the magnitude of the long-range spurious, random correlations decreases approximately proportionally to $N_{\text{ens}}^{-1/2}$, confirming that they are due to sampling error.

The improved representation of the background error covariances due to the reduction in sampling noise has a direct effect on the EnKF analysis and forecast accuracy. This can be appreciated in Fig. 5 where we present the forecast skill scores (Anomaly Correlation for 500 hPa geopotential height) of the EnKF mean forecast for a 60 member ensemble (solid line), a 120 member ensemble (dashed line) and a 240 member ensemble (dotted line) in the northern hemisphere (NHem., left panel) and the southern hemisphere (SHem., right panel). The improvement going from a 60 member to a 120 member ensemble is consistent and statistically significant throughout the forecast range (with a relative average improvement of the order of 5% in the NHem., and of 10% in the SHem., not shown). On the other hand, the gains are more limited when moving from a 120 member ensemble to 240 members. It could be argued that a larger ensemble size would make it advantageous to use broader covariance localization parameters, and this should be beneficial for the larger ensemble sizes. This is confirmed by the results given in Fig. 6, showing the relative improvement in forecast skill of a 240 member EnKF using a vertical covariance localization tapering to zero at two times the length scale used in the standard configuration and allowing the use of twice as many observations in the update of each analysis grid point. These changes are clearly beneficial, at least in the early part of the forecast range, but, given the size of the differences, longer and considerably more expensive experimentation would be required to confirm positive impact at larger forecast ranges. In general we appear to be achieving diminishing returns from increasing the ensemble dimension once a size $O(100)$ is reached and other limiting factors beyond sampling errors seem to become relatively more important in limiting the skill of the EnKF. This is not dissimilar to the experience of other operational Centres (Houtekamer et al., 2014).

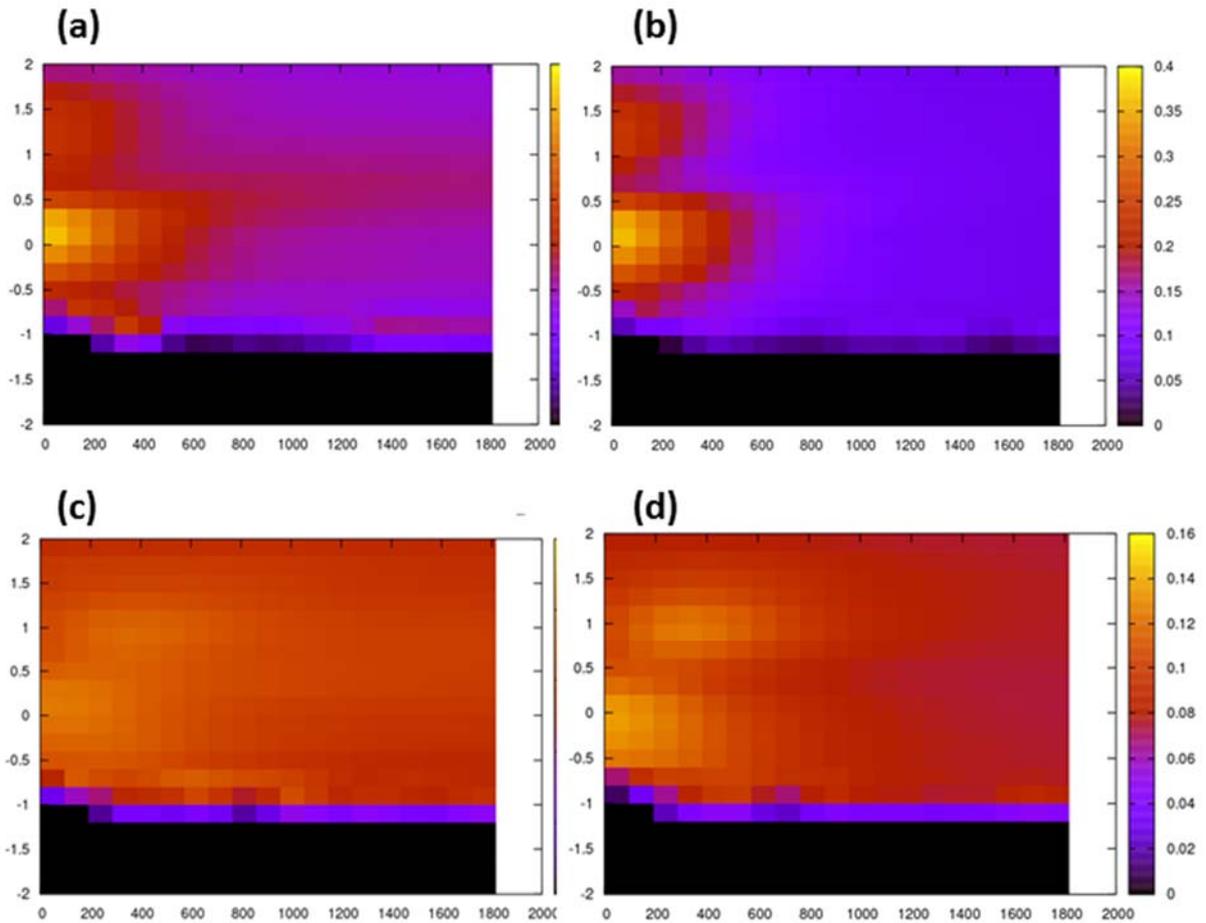


Figure 3: Top row: Azimuthally-averaged absolute correlation between short-range (3-hour) forecast of temperature at 500 hPa and AMSU-A brightness temperature model equivalents for a 60 member EnKF (panel a) and a 120 member EnKF (panel b). Bottom row: Azimuthally-averaged absolute correlation between short-range (3-hour) forecast of zonal wind at 500 hPa and AMSU-A brightness temperature model equivalents for a 60 member EnKF (panel c) and a 120 member EnKF (panel d). Vertical coordinate in scale height, horizontal coordinate in km.

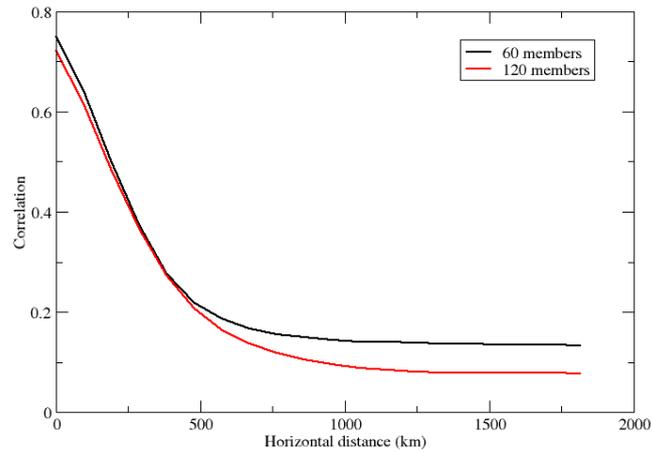


Figure 4: Absolute correlation of the short-range (3-hour) temperature forecast at 500 hPa with the radiosonde temperature observations at the same pressure level, as a function of horizontal distance (Units: km).

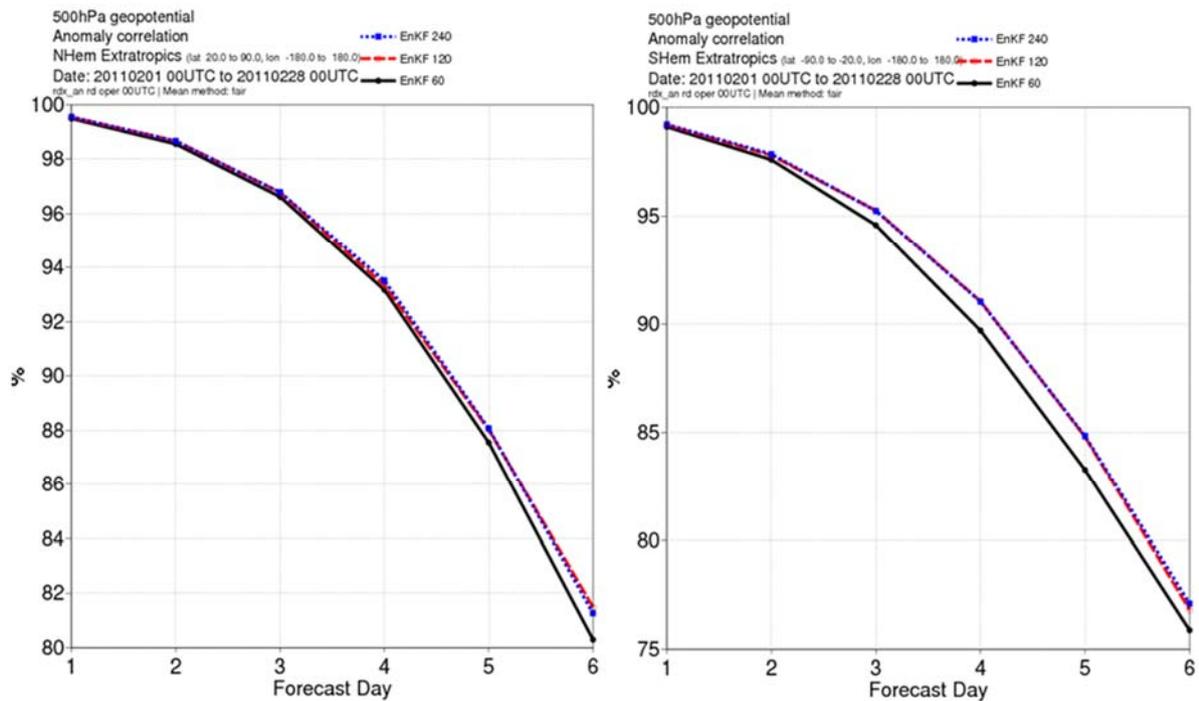


Figure 5: 500 hPa geopotential forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) for the 60 member EnKF (solid line), 120 member EnKF (dashed line) and 240 member (dotted line). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2011-02-01 to 2011-02-28 period.

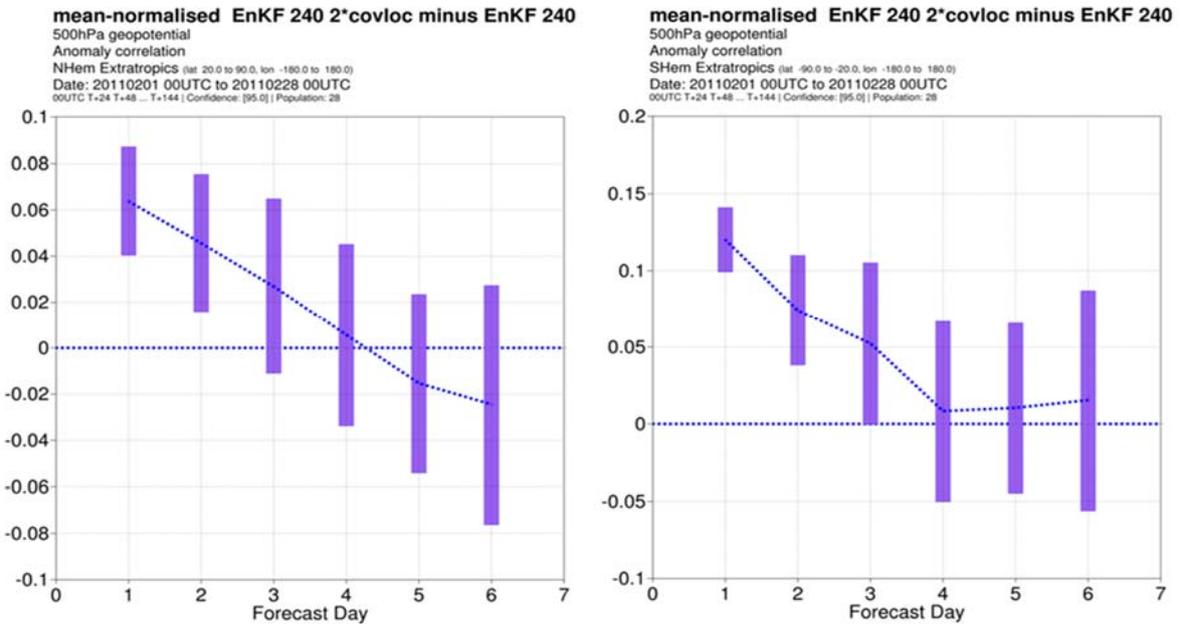


Figure 6: Normalised difference of the 500 hPa geopotential forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) between the 240 member EnKF with standard covariance localization (EnKF 240) and the 240 member EnKF experiment (EnKF 240 2xcovloc) with double vertical covariance localization and observation counts. Positive values indicate larger forecast anomaly correlation for the EnKF 240 2xcovloc experiment. Scores are computed with respect to ECMWF operational analysis and are cumulated over the 2011-02-01 to 2011-02-28 period. Error bars represent 95% confidence levels.

b) Ensemble resolution

The spatial resolution at which the EnKF ensemble is run is also one of the main aspects of the EnKF configuration that needs to be investigated. When computational resources are increased, the relative benefits of running a larger ensemble versus running the ensemble at higher resolution need to be identified.

Higher resolution is expected to be beneficial for the EnKF as the short-range forecast fields have a more realistic behaviour in the high wavenumber part of the spectrum and benefit from a more accurate representation of physiographic constraints. This in turn will lead to more accurate model equivalents, in particular near complex orography, and thus allow a better use of observational information available in those areas. This ansatz is borne out by results shown in Fig. 7, where we present the forecast skill scores for 60 member EnKF assimilation experiments run at three different resolutions: spectral truncations T159 (approximate grid point spacing 120 km; continuous line), T319 (approximate grid point spacing 60 km; dashed line) and T639 (approximate grid point spacing 30 km; dotted line). The positive impact of increasing horizontal resolution is quite evident going from T159 to T319, while diminishing returns are seen going from T319 to T639 (note, however, that T639 skill scores presented in Fig. 7 are still better than T319 scores at the 95% confidence level for the first 6 days of the forecast range at 1000 hPa in both hemispheres). Note also how the largest impact is seen in the NHem. 1000 hPa geopotential height scores, supporting the argument presented above.

The increase in horizontal resolution produces more active ensemble background fields with larger spread (Fig. 8). This in turn has the desirable effect of reducing the reliance of the EnKF on the adaptive covariance inflation algorithms (relaxation to prior variance, in our case) that are generally needed to prevent the filter collapse. Similar results are found for the ECMWF EDA (Bonavita et al., 2014).

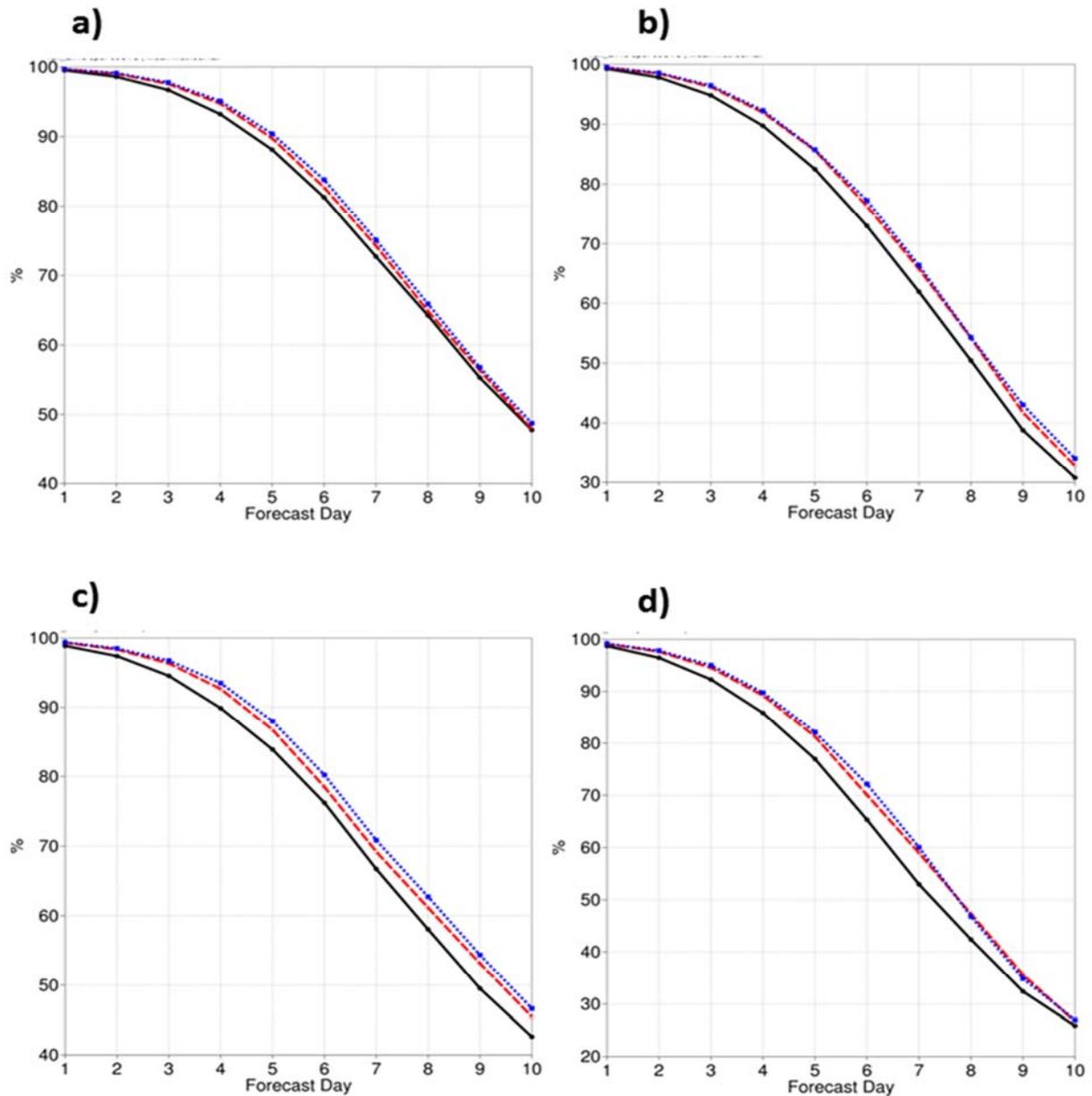


Figure 7: Geopotential forecast anomaly correlation scores for (a) 1000 hPa northern hemisphere, (b) 1000 hPa southern hemisphere, (c) 500 hPa northern hemisphere and (d) 500 hPa southern hemisphere. Curves refer to the T159 EnKF (continuous line), the T319 EnKF (dashed) and T639 EnKF (dotted). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2011-02-01 to 2011-03-31 period.

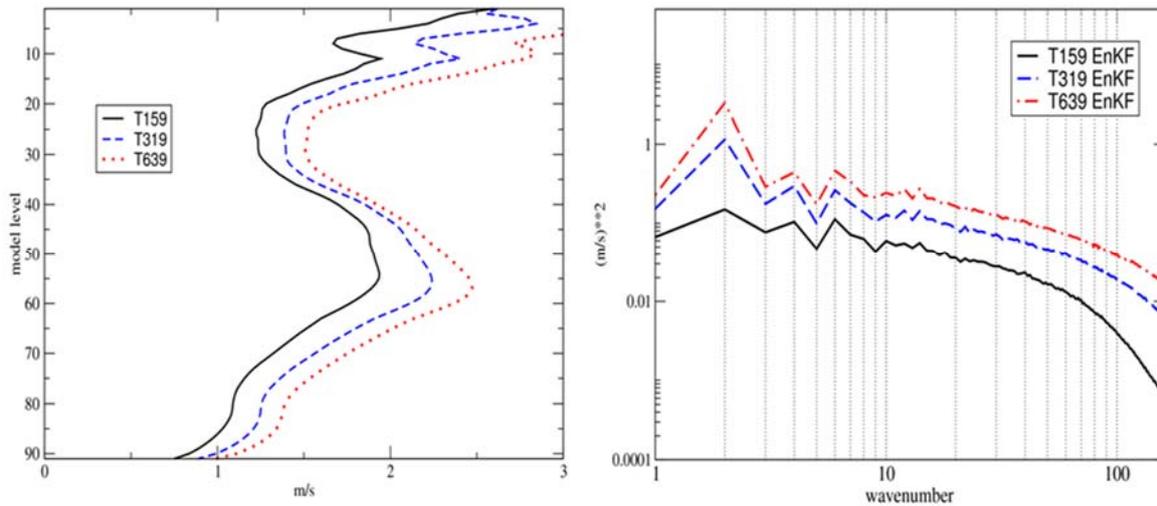


Figure 8: Left panel: Standard deviation profiles of background (6-hour) forecast of zonal wind for T159 EnKF (solid line), T319 EnKF (dashed line) and T639 EnKF (dotted line). Global average valid on 2011-03-30 at 00 UTC. Right panel: power spectra of zonal wind background forecast variance at model level 56 (300 hPa) for different EnKF resolutions (T159, solid; T319, dashed; T639, dash-dot), averaged over March 2011.

c) Covariance inflation and representation of model error

The simple use of the EnKF update equations usually results in the posterior ensemble having a spread smaller than optimal (Miyoshi et al., 2010, and references therein). This is due to many reasons (model error, nonlinearities in error evolution and observation operators, sampling errors, etc.) and can potentially lead to the filter giving too low weight to new batches of observations and eventually decoupling from reality (“filter divergence”). It is a crucial and still open issue in the EnKF how to best tackle this variance underestimation. Simple multiplicative covariance inflation in the state-dependent form proposed by Whitaker and Hamill (2012) has been implemented in the ECMWF EnKF. This has been found to provide a robust and effective baseline solution. The work of Whitaker and Hamill (2012) suggests that multiplicative inflation is well suited to account for sampling errors and errors linked to observational distribution and density. This is confirmed by Fig. 9, where we present an example of the multiplicative inflation factor that is diagnosed and applied on the temperature field at model level 49 (~200 hPa) by the relaxation to prior variance algorithm: the signature of the radiosonde network and of the more trafficked aircraft routes is clearly discernible.

However, an important component of the unrepresented background error is given by the model error. In both the EnKF and the ensemble prediction literature various solutions have been proposed in order to represent model error. The methods involve using different prognostic models (multi-model approach), different physical parameterizations available in the same model (multiple parameterizations approach) or running the different ensemble members’ forecasts perturbing some of the key parameters in the physical parameterizations of the model (perturbed parameter approach). All these methods, however, are typically difficult to tune and to maintain, especially in an operational environment. Other common approaches to deal with model error in the EnKF context include a) the use of an additive inflation of structured climatological noise fields (Mitchell and Houtekamer, 2000) and b) the use of a stochastic model error parameterizations in the prognostic model. Option a) is easy to implement and has been proved to be effective (Whitaker and Hamill, 2012, and references therein) while option b) is attractive because of the ready availability of the stochastic model error

parameterization code in the IFS and the long history of use and development of this code in the ECMWF Ensemble Prediction System (EPS; Palmer et al., 2009; recently renamed to the ensemble forecast, ENS) and in the EDA. A limited set of experiments has been performed to evaluate the relative contribution of these last two methods using an EnKF system with multiplicative covariance inflation (relaxation to prior variance) as the baseline. The additive inflation uses a climatology of forecast differences (48-hour minus 24-hour) verifying at the same time which have been linearly rescaled to represent 6-hour error growth (i.e., they have been multiplied by 0.25). These climatological perturbations are applied to the analysis ensemble, so that the forecast step has a chance to dynamically condition the error structures that have been introduced (this was found to give a marginally positive impact under idealized conditions in Whitaker and Hamill, 2012). The stochastic model error parameterization uses a configuration similar to the one used in the operational ECMWF EPS, which employs both the Stochastically Perturbed Parameterization Tendencies formulation (SPPT, Buizza et al., 1999; Palmer et al., 2009) and the Stochastic Kinetic Energy Backscatter (SKEBS; Shutts and Palmer, 2004; Shutts, 2005; Berner et al., 2009). The only important difference is that the EPS uses the three-time scales version of the SPPT algorithm (i.e., stochastic errors are generated satisfying three different sets of spatial and temporal decorrelation scales: 6 hours/500 km, 3 days/1000 km and 30 days/2000 km), while in the EnKF the one-time scale version of the algorithm (6 hours/500 km decorrelation scales) is used. This is similar to how the SPPT is used in the EDA.

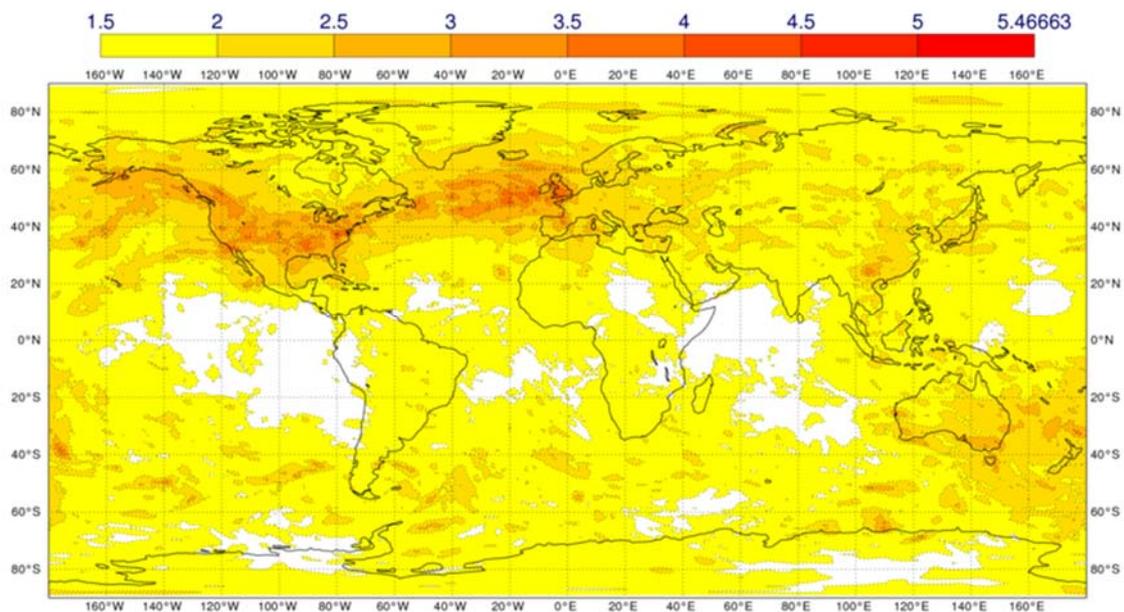


Figure 9: Multiplicative inflation factor for temperature at model level 49 (~200 hPa). T319L91 EnKF valid on 20110330 at 00 UTC

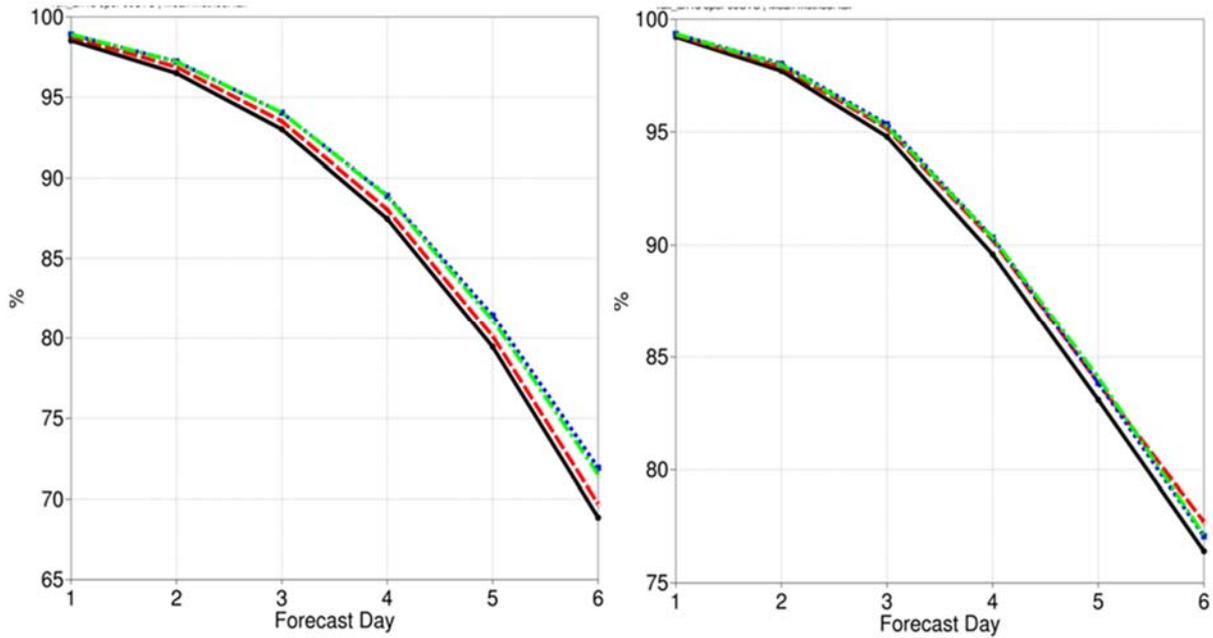


Figure 10: Geopotential forecast anomaly correlation in the northern hemisphere (left column) and southern hemisphere (right column) at 500 hPa. Curves refer to a T159 EnKF using multiplicative inflation (continuous line); multiplicative inflation plus stochastic model error parameterizations (dashed line); multiplicative inflation plus additive inflation (dotted line); multiplicative inflation plus additive inflation plus stochastic model error parameterizations (dash-dot line). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2010-07-01 to 2010-07-31 period.

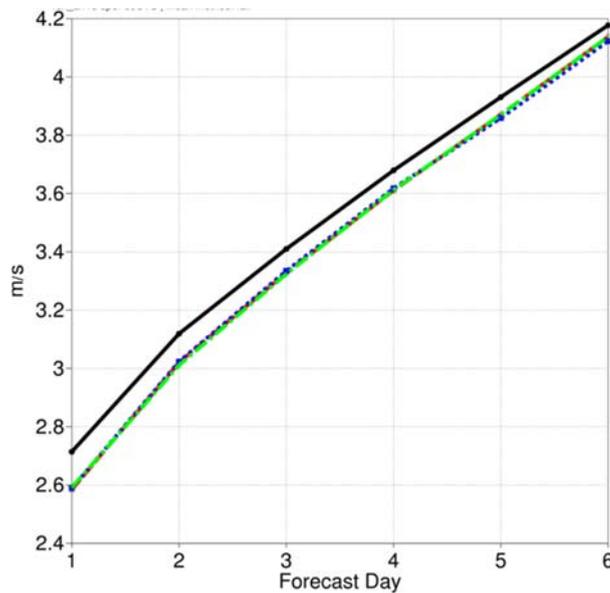


Figure 11: Wind vector root mean square forecast error in the tropics (20°S - 20°N) at 850 hPa. Curves refer to a T159 EnKF using multiplicative inflation (continuous line); multiplicative inflation plus stochastic model error parameterizations (dashed line); multiplicative inflation plus additive inflation (dotted line); multiplicative inflation plus additive inflation plus stochastic model error parameterizations (dash-dot line). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2010-07-01 to 2010-07-31 period.

A small sample of the results of these experiments is presented in Figs. 10 and 11. Some tentative conclusions can be drawn from this limited experimentation. In the first place we note that both additive inflation and stochastic parameterizations are able to improve ensemble mean forecast skill compared to the baseline system with multiplicative inflation: this confirms that both methods accounts for different types of errors than the multiplicative inflation. Secondly, additive inflation in combination with multiplicative inflation improves forecast scores more than stochastic parameterizations in combination with multiplicative inflation. Finally the addition of stochastic model error parameterizations to an EnKF already running with multiplicative and additive inflation does not lead to any statistically significant improvement in terms of ensemble mean forecast skill. These last two results are in agreement with the conclusions presented by Whitaker and Hamill (2012), who also showed, albeit in an idealized environment, that it is difficult for physically based model error parameterizations to improve on a simple combination of multiplicative and additive inflation in the context of ensemble data assimilation.

5 Hybrid Gain Ensemble Data Assimilation

The development activities described in the previous chapters have led to a mature EnKF-based data assimilation system which has an overall performance close to that of a comparable low resolution 4DVar system which uses static, climatological background error and covariance estimates. This can be appreciated in Fig. 12 where we present the 500 hPa geopotential height forecast skill scores of a T399, 137 levels, 100 member EnKF together with those of a T399 static 4DVar with T95-T159 inner loops (which is the control, unperturbed member of a standard EDA). Both experiments use the full observational data set used in operations with the exception of the EnKF which is not using rain-affected radiances (Bauer et al., 2010) and ground-based precipitation composites (Lopez, 2011).

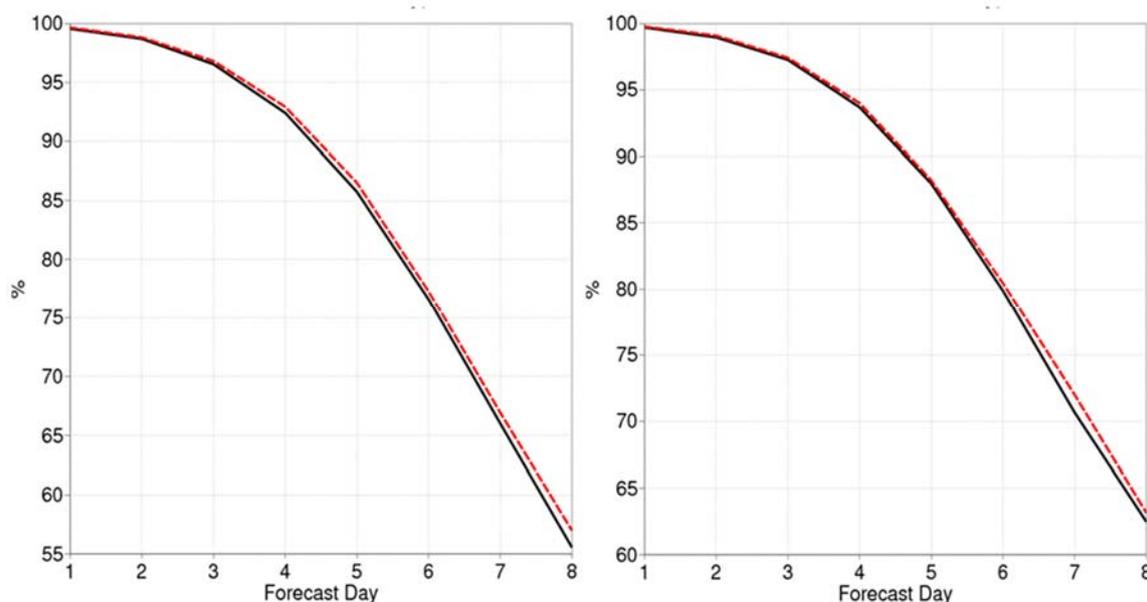


Figure 12: Geopotential height forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) at 500 hPa. Curves refer to a T399 100 member EnKF (continuous line); and a T399 (T95/T159 inner loops) 4DVar using static background errors and covariances (dashed line). Scores are

computed with respect to ECMWF operational analysis and are averaged over the 2012-06-02 to 2012-07-31 period.

The natural question then is: is it possible to combine these two data assimilation systems and improve on the performance of both? From a 4DVar perspective the answer is certainly positive and is based on introducing flow-dependent background error information from the EnKF or EDA in the standard 4DVar algorithm (Barker, 1999; Hamill and Snyder, 2000; Lorenc, 2003; Buehner, 2005, 2010a,b; Wang, 2010; Bonavita et al., 2012, Bonavita et al., 2014). From an EnKF perspective less work has been done. The basic limitation of the pure EnKF lies in the reduced dimensionality of the represented background error subspace, which limits the amount of observational information that can be extracted from non-local observations and, possibly, in regions with high density of observations. On the other hand, the full rank of the modelled background error covariance matrix used in variational assimilation makes 3-4DVar algorithms immune to this problem. This suggests that exploiting this property of the variational approach in an EnKF context would be advantageous. Recent work by Penny (2014), in an idealized setup, suggested a new hybrid variational/EnKF formulation that involves taking a weighted mean of the EnKF and variational analysis and using this “control” state to re-centre the EnKF analysis ensemble. This “hybrid gain” approach was also shown to be functionally equivalent, in the limit of linear estimation theory, to the standard “hybrid B covariance” approach commonly used in variational implementations, but it also offers some significant advantages. At ECMWF the method has been implemented re-using available, well-tested components of the IFS. The hybrid-gain approach also allows making use in the EnKF framework of components of the 4DVar system (such as variational bias correction) that are simpler and cheaper to run in a variational context. On the other hand, error covariance localization needs only be applied in the EnKF analysis, where it is a natural extension of the EnKF algorithm: thus, the use of a static 4DVar in the re-centring step of the hybrid gain analysis can be seen as a regularization procedure of the EnKF analysis with two potentially important benefits: a) reducing the effect of sampling noise and covariance localization, and b) introducing climatological information on the background error covariances, which is important to reduce the effect of model biases in the analysis and it is not easy to do in a standard EnKF. On a more fundamental level, it allows independent development of the EnKF and 4DVar systems which will both have a beneficial influence on the resulting hybrid system.

Following these ideas we have implemented the Hybrid Gain Ensemble Data Assimilation (HG-EnDA) system shown in the schematic of Fig. 13, where the forecast from the ensemble mean analysis is also used as background and starting linearization trajectory for an incremental 4DVar analysis. The analysis fields from the EnKF and 4DVar are then linearly combined to produce a “hybrid gain” analysis around which the EnKF analysis ensemble is re-centred.

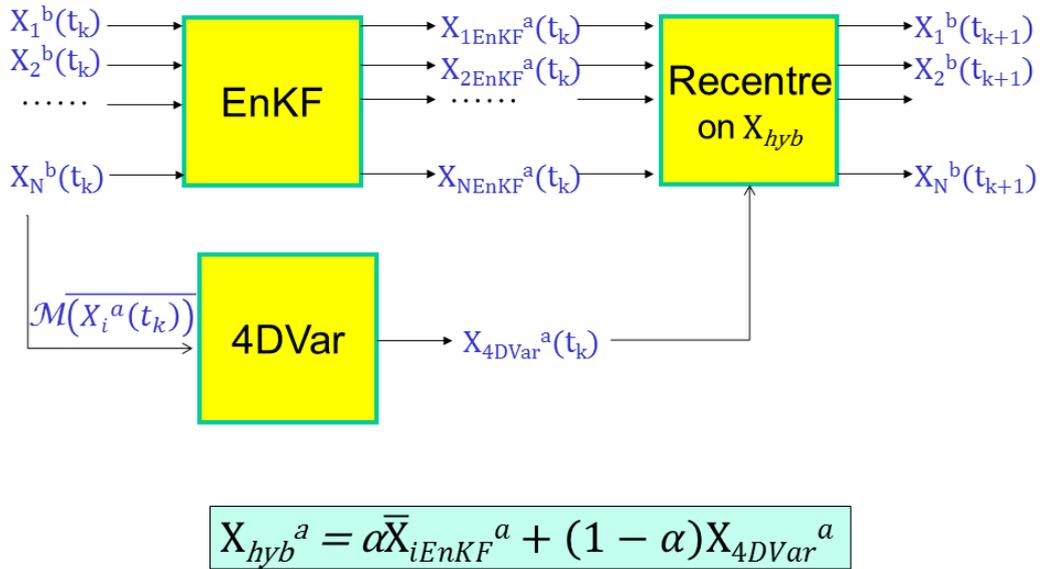


Figure 13: Schematic of the hybrid gain EnDA.

Given the comparable skill of the EnKF and static 4DVar systems, a hybrid gain EnDA experiment has been run by giving equal weight to the EnKF and 4DVar analysis in the hybrid, i.e. $\alpha=0.5$. The 4DVar analysis inside the hybrid algorithm is an exact replica of the standalone static 4DVar (T399 outer loop with two T95/T159 inner loops) with the only notable exception being that it makes use of the background error standard deviation estimates produced by the EnKF short range forecast.

The introduction of the hybrid gain formulation has a clear positive impact on the analysis and forecast skill of the system. In Fig. 14 we show the vertical profiles of the root mean square (RMS) of background departures for the three systems for zonal wind and temperature. The hybrid system has distinctly reduced RMS departures for both variables over most of the globe and pressure levels, with the only notable exception of the boundary layer winds in the southern hemisphere. This advantage in analysis skill is also visible for the other main model variables (including specific humidity) and translates into a corresponding advantage in forecast skill (Fig. 15), which is larger in the southern extra-tropics and the tropics.

The ability of the hybrid gain EnDA to outperform its two component systems (i.e., a static 4DVar and an EnKF) is an encouraging starting point. A more stringent test, however, is comparing the performance of the proposed algorithm to that of the state of the art ECMWF hybrid 4DVar-EDA. The ECMWF hybrid 4DVar-EDA assimilation system (Bonavita et al., 2012; Bonavita et al., 2014) makes use of flow-dependent errors and structure functions estimated from a set of EDA short-range forecasts. This hybrid system has been shown to bring a clear improvement in analysis and forecast skill with respect to a 4DVar using static, climatological estimates of background error covariances. In Fig. 16 we compare the 500 hPa geopotential forecast anomaly correlation for a T511 100 member hybrid gain EnDA and a T511 (T95-T159-T255 inner loop resolution) hybrid 4DVar-EDA. The plots

on the top row are computed using the ECMWF operational analysis as the truth, while the bottom plots use the experiments' own analyses as the verifying dataset. Apart from the first 2–3 days, which are more influenced by the verifying analysis (since the operational analysis used for verification is a similar hybrid 4DVar-EDA system run with T1279 outer loop resolution), we can see that the two systems have comparable skill (note, however, that the advantage of the HG-EnDA over the early forecast range is likely to be real, as observation departures for the HG-EnDA have consistently smaller variances than those of the hybrid 4DVar-EDA, i.e. Fig. 17). This conclusion holds for most of the extra-tropical tropospheric scores. In the stratosphere (Fig. 18) the HG-EnDA has superior performance in the northern hemisphere, while hybrid 4DVar-EDA performs better in the southern hemisphere: this could be due to seasonal effects. In the tropics, as usual, the skill of two systems with similar level of performance is more difficult to assess due to the inherent, larger uncertainties in the verifying analysis. Verification against the sparse radiosonde network generally shows a similar level of skill (Fig. 19).

Since the hybrid gain EnDA control analysis is basically a blend of an EnKF and a 4DVar analysis, it is also of interest to try to identify the similarities and, more importantly, the differences between the two component analyses.

Globally the EnKF analysis increments (i.e., analysis minus background forecast differences) are about 30% to 40% larger than the corresponding 4DVar increments (Fig. 20). This is also apparent in Fig. 21 where we present the zonally-averaged vertical profiles of the analysis increments of zonal wind for the two systems. Apart from the difference in magnitudes, the EnKF increments show more variability and sharper gradients than the 4DVar increments. This feature is confirmed by the analysis of the spectral characteristics of the analysis increments. In Fig. 22, left panel, we show the power spectra of the 4DVar analysis increments (for the control member and the average of the perturbed members) and the EnKF analysis increments (for the ensemble mean and the average of the ensemble members) for geopotential at 500 hPa. It is apparent that the EnKF increments (mean and members) have more energy than the corresponding 4DVar increments throughout the spectrum, with the exception of the first 4–5 wavenumbers. This reduced energy of the EnKF increments for the longest waves is more evident at the surface and tends to disappear above the tropopause (not shown). Its likely explanation is that this is the result of the localization applied to the EnKF increments, which dampens long range correlations, and the inhomogeneous distribution of conventional observations, which makes this phenomenon more apparent in the troposphere. In the stratosphere, on the other hand, the analysis is dominated by the satellite component of the observing system, which provides a global, homogeneous cover and thus makes the EnKF increments relatively insensitive to the horizontal localization. On the other hand the larger energy of the EnKF increments at synoptic and smaller scales is a combination of the effects of a) the higher resolution of the EnKF analysis (the 4DVar increments are only computed at T95 and T159 resolutions in the EDA), and b) the impact of an excessively high level of additive noise used in the experiment (see below for further discussion).

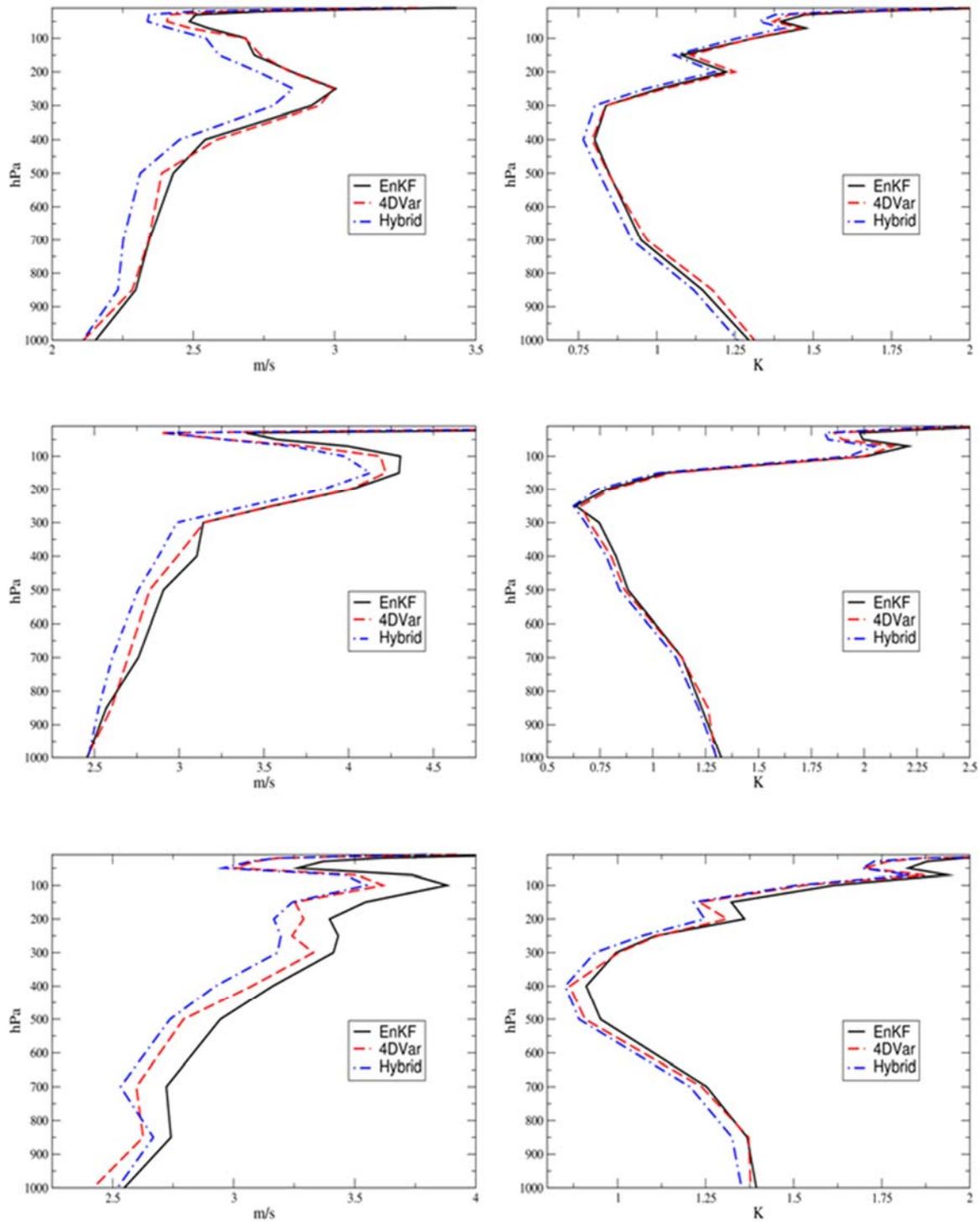


Figure 14: Root mean square background forecast observation departures for radiosonde zonal wind (left column) and radiosonde temperature (right column) for the northern hemisphere (top row), tropics (middle row) and southern hemisphere (bottom row). Curves refer to T399 EnKF (continuous line), T399 static 4DVar (dashed line), T399 Hybrid gain EnDA (dash-dot line) and present statistics computed over the 2012-7-01 to 2012-7-15 period.

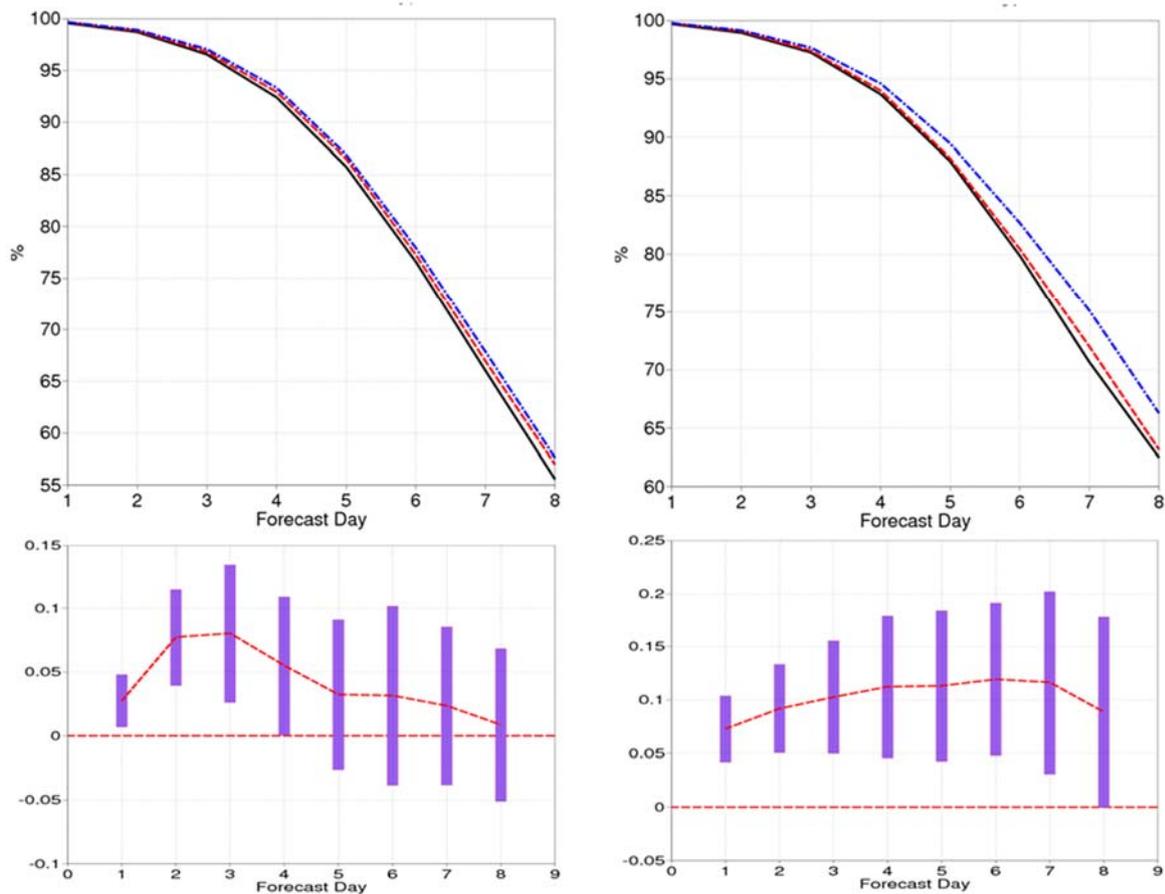


Figure 15: Top row: Geopotential height forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) at 500 hPa. Curves refer to a T399 100 member EnKF (continuous line); a T399 (T95/T159 inner loops) 4DVar using static background errors and covariances (dashed line); and a T399 100 member HG-EnDA (dot-dash line). Bottom row: Normalised difference of the 500 hPa geopotential height forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) between the HG-EnDA and the static 4DVar (positive values indicate superior skill of the HG-EnDA). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2012-06-02 to 2012-07-31 period.

Another interesting feature revealed by Fig. 22 is that the EDA members throughout the spectral range are homogeneously more energetic than their unperturbed control; on the other hand, the EnKF members have, on average, a much larger energy than their control in the synoptic scales (wavenumber range 10–50). The main reason lies in the spectral distribution of the additive errors that are used in the EnKF (Fig. 22, right panel). These climatological perturbations are constructed as scaled differences of 48 to 24-hour forecasts verifying at the same time, which, at this forecast range, tend to be dominated by synoptic-scale errors. It is evident that, in this experiment, the model error formulation has a large influence on the shape of the EnKF analysis spectrum. It is worth bearing in mind that the value of 0.25, which is applied to the forecast differences, was found to be satisfactory in the context of experimentation conducted at T159 resolution and is probably not suitable for a T511 set-up. In later experiments, we have in fact tested running the EnKF with a further halving of the weight given to the additive inflation perturbations (i.e., rescaling the climatological forecast

differences by 0.125 instead of 0.25). This change has led to marked improvements in the forecast scores, especially in the southern hemisphere (Fig. 23) and the tropics (not shown). This confirms that the additive inflation was clearly overdone in previous experiments and that increases in resolution greatly reduce the need to rely on covariance inflation techniques to maintain ensemble spread: this is ultimately beneficial in terms of accuracy and reliability of the resulting ensemble.

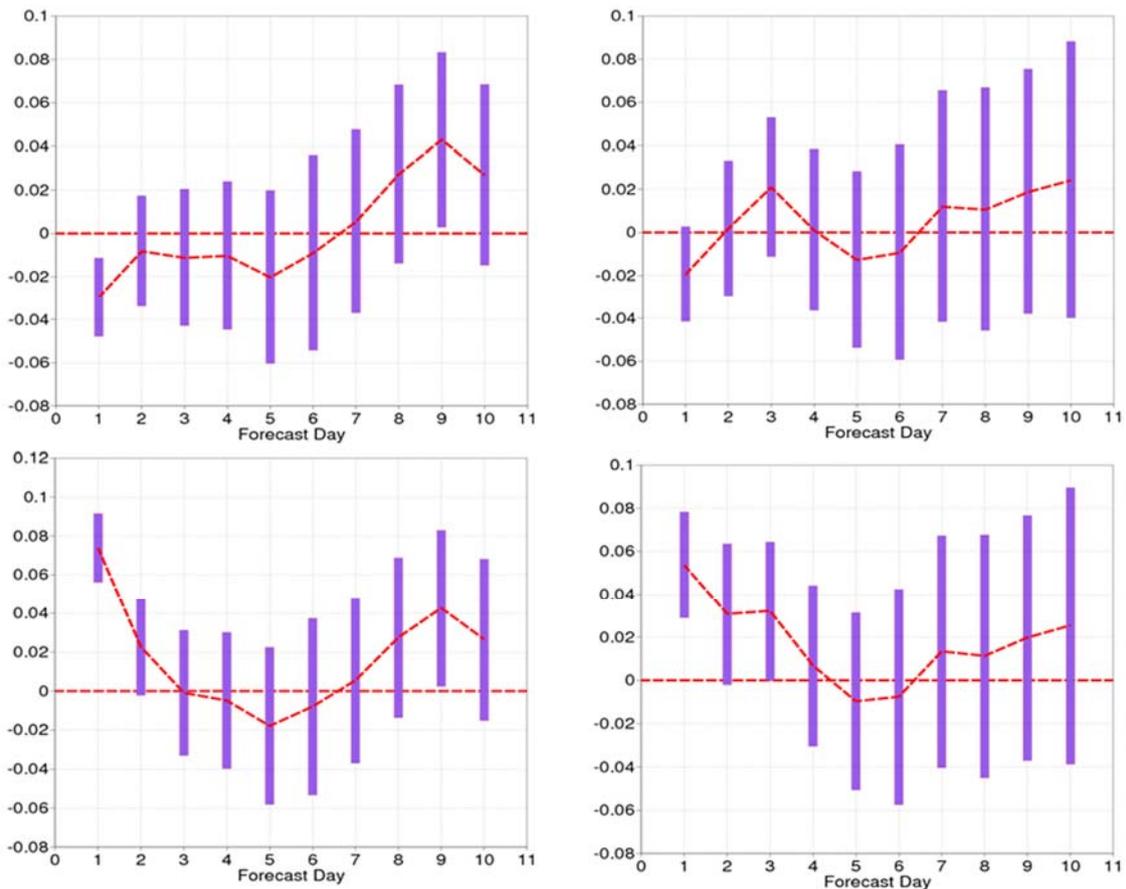


Figure 16: Normalised difference of the 500 hPa geopotential forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) between the hybrid gain EnDA and the hybrid 4DVar-EDA (positive values indicate superior skill of the hybrid gain EnDA). Scores are computed with respect to ECMWF operational analysis in the top row plots; with respect to own analysis in the bottom row plots; and are averaged over the 2012-06-02 to 2012-09-10 period.

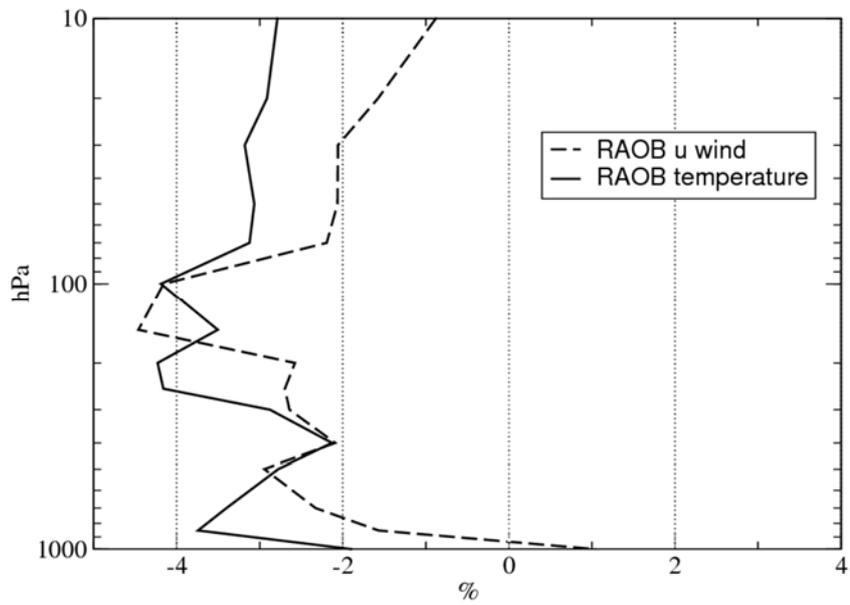


Figure 17: Normalised difference of RMS background forecast observation departures for radiosonde zonal wind observations (dashed line) and radiosonde temperature observations (continuous line) between the HG-EnDA and the hybrid 4DVar-EDA (negative values indicate better fit of the hybrid gain EnDA). Statistics computed over the 2012-7-01 to 2012-7-31 period and averaged over the globe.

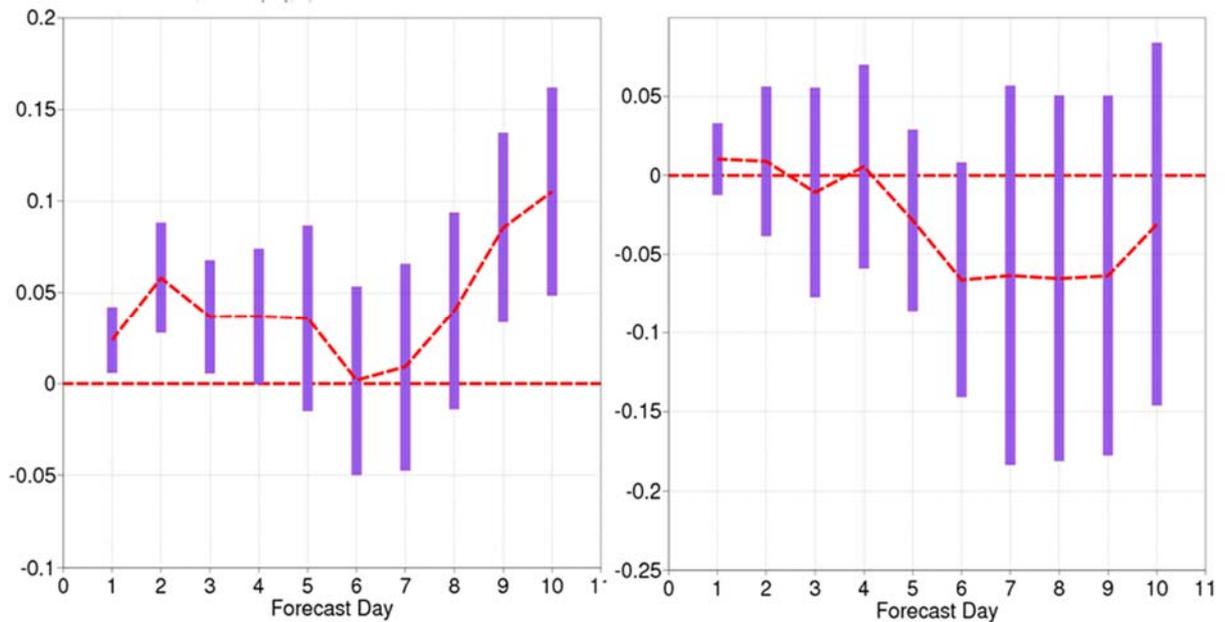


Figure 18: Normalised difference of the 50 hPa geopotential forecast anomaly correlation in the northern hemisphere (left panel) and southern hemisphere (right panel) between the HG-EnDA and the hybrid 4DVar-EDA (positive values indicate superior skill of the HG-EnDA). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2012-06-02 to 2012-09-10 period.

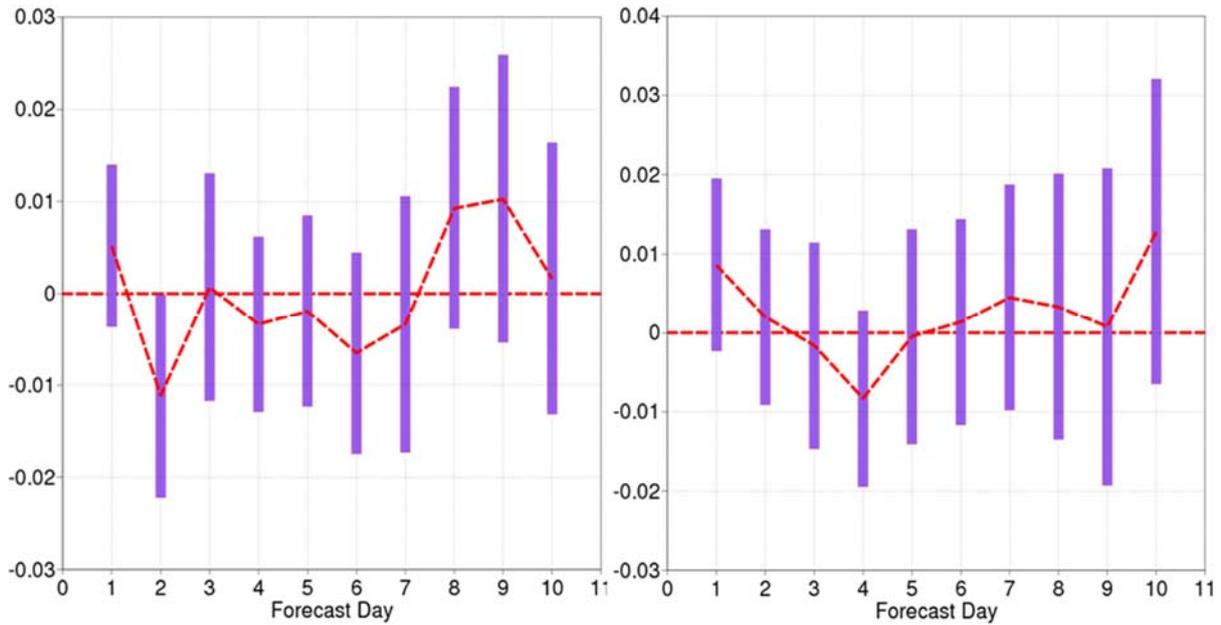


Figure 19: Normalised difference of the 850 hPa (left panel) and 200 hPa (right panel) wind vector forecast root mean square error in the tropics between the HG-EnDA and the hybrid 4DVar-EDA (positive values indicate superior skill of the HG-EnDA). Scores are computed with respect to available radiosonde stations and are averaged over the 2012-06-02 to 2012-09-10 period.

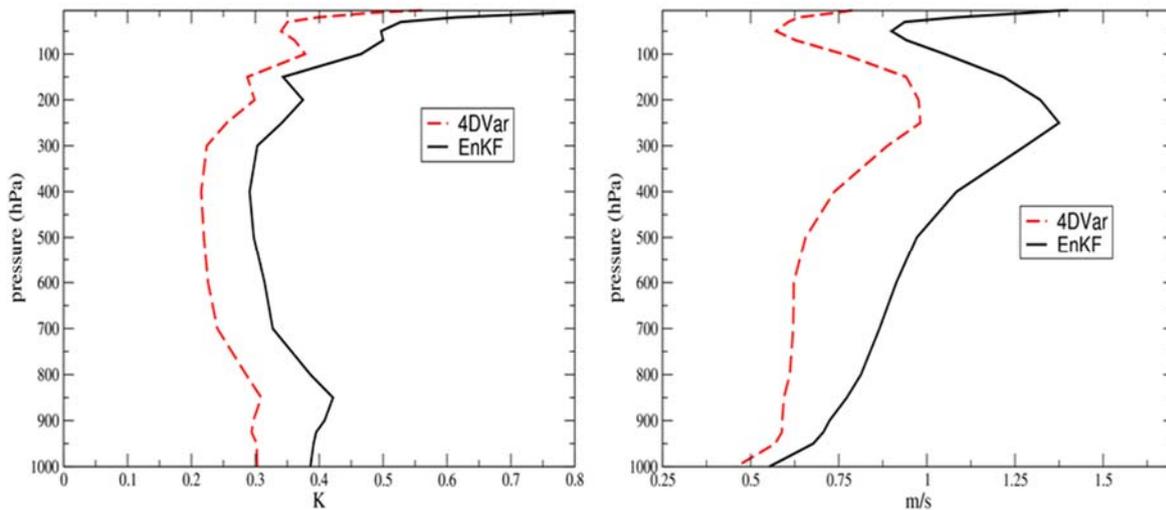


Figure 20: Vertical profiles of the standard deviation of the temperature (left panel) and zonal wind (right panel) analysis increments (analysis minus background forecast differences) of the EDA unperturbed control 4DVar (dash line) and of the EnKF mean (dotted line). Values averaged over 1 month (July 2012). Vertical units: hPa; range: 10–1000 hPa. Horizontal units: Kelvin (left panel) and m/s (right panel).

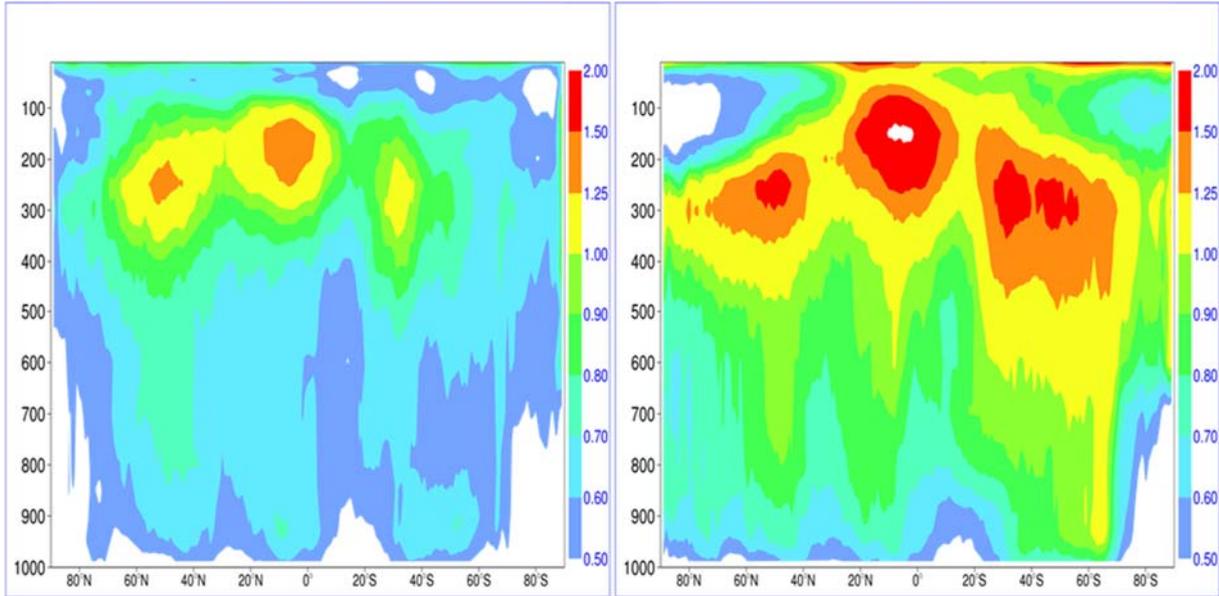


Figure 21: Zonal averages of the standard deviation of the zonal wind (right panel) analysis increments (analysis minus background forecast differences) of the EDA unperturbed control 4DVar (left panel) and of the EnKF mean (right panel). Values averaged over 1 month (July 2012). Vertical units: hPa; range: 10–1000 hPa. Legend in m/s.

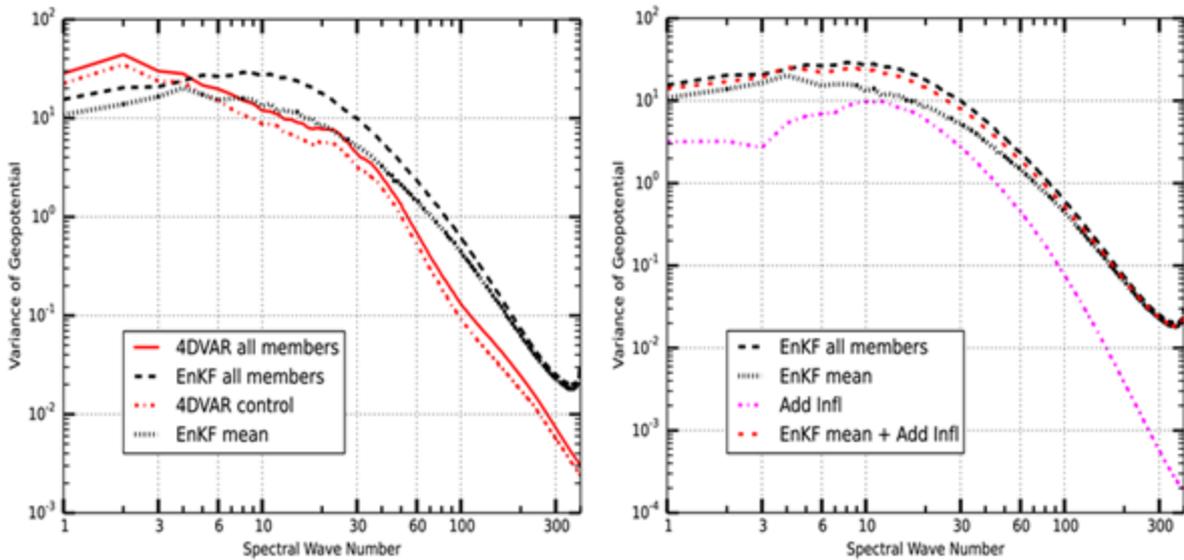


Figure 22: Left panel: Power spectrum of the 500 hPa geopotential analysis increments (analysis minus background forecast differences) averaged over all EDA perturbed members (continuous line); EDA unperturbed control (dash-dot line); EnKF mean (dotted line); average over all EnKF members (dashed line). Right panel: Dotted line and dashed line like left panel. Dash-dotted line shows power spectrum of the climatological perturbations which are added to the EnKF members in the additive inflation algorithm. All analysis increment curves show averages over one month (July 2012) at 00 UTC and 12 UTC.

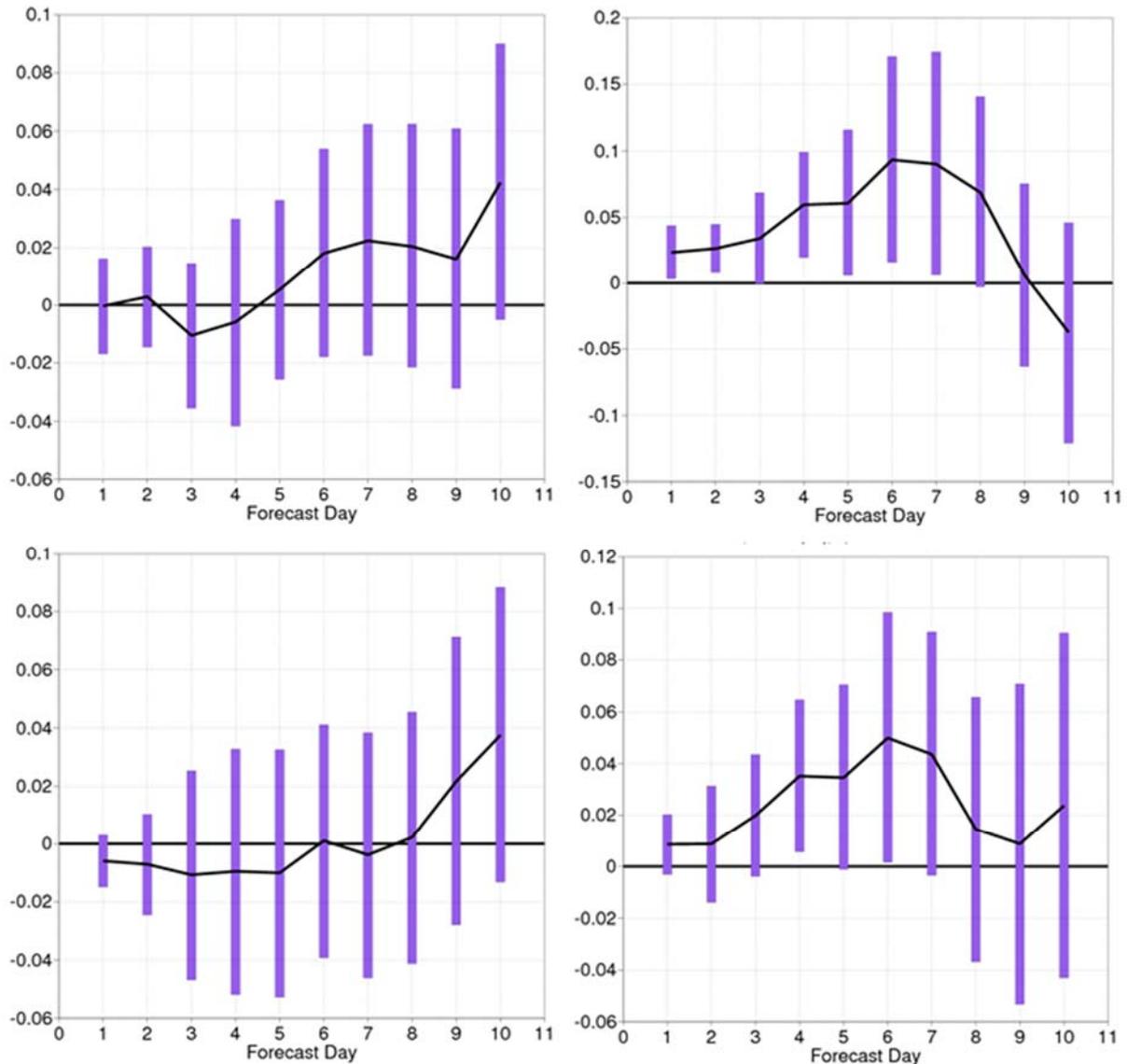


Figure 23: Normalised difference of the 100 hPa geopotential forecast anomaly correlation (top panels) and 500 hPa geopotential forecast anomaly correlation in the northern hemisphere (left column) and southern hemisphere (right column) between T399 EnKF experiments with a rescaling factor of 0.25 and 0.125 of the climatological 24-hour forecast differences used for additive inflation (positive values indicate superior skill of the 0.125 experiment). Scores are computed with respect to ECMWF operational analysis and are averaged over the 2012-06-01 to 2012-08-20 period.

6 Discussion

The wish to do detailed comparisons between variational and more scalable ensemble based assimilation systems in a semi-operational environment have led to the development of a state of the art EnKF system at ECMWF. The human resource cost of these activities has been relatively small, partly because of the intrinsic simplicity of the EnKF algorithm, but primarily because it was possible to reuse large parts of the ECMWF IFS codebase and scripts, control structures, observation and field

data bases developed for the operational EDA at ECMWF. This paper presents some results from the experimentation done with the ECMWF EnKF, and a new, hybrid variational-EnKF assimilation system which we call hybrid gain EnDA.

The sensitivity of the EnKF to a rather large subset of parameters has been investigated. Results are generally similar to what has been found by other studies. What is remarkable is the robustness of the EnKF performance to relatively large changes in various parameters, such as covariance localization lengths and covariance inflation parameters. On the other hand a clear sensitivity has been found to the resolution at which the ensemble is run and to the ensemble size. The advantage of running the EnKF at higher resolution is clear going from T159 to T319 spectral truncation. A small additional improvement of skill is obtained from going up to T639 (~30 km grid spacing), which was the highest resolution tested. This is assumed to be the result both of better model accuracy and also of larger and more realistic levels of model activity which result in bigger background forecast spread and thus reduced reliance on covariance inflation adjustments. Also in terms of ensemble size, larger ensembles produce more accurate analysis and forecasts. This is evident going from a 60 member to a 120 member ensemble; a small, but statistically significant advantage is also seen going from a 120 to a 240 member ensemble. The reasons for this improvement are clear: covariances sampled from a larger ensemble have an intrinsic lower level of sampling noise, which makes it possible to get better estimates of weak error covariances (such as vertical covariances and mass-wind cross-covariances, i.e. Fig. 3) and at the same time relax the localization constraints. It is shown that broader vertical covariance localization significantly improves the 240 member ensemble results.

A limited amount of experimentation has also gone into defining the best way to represent model error in the EnKF system and to combat its under-dispersiveness. Results confirm findings from other scientific studies; i.e. a combination of multiplicative covariance inflation and additive inflation using climatological forecast differences is, currently, the most effective way to tackle the problem and has been adopted as the standard configuration. The use of physically-based model error parameterizations was not found to give further improvements with respect to the standard setup. It appeared that the model error formulation had a large influence on the shape of the EnKF analysis increments' spectrum. Further experimentation has however shown that this feature was an artefact of the excessively large weight given to the forecasts perturbations used in the additive inflation scheme in experiments which were run at higher spatial resolution. In general, increasing the resolution of the EnKF and HG-EnDA was found to significantly reduce the need for model error parameterizations.

While the standard EnKF has proven to be a remarkably robust and skilful assimilation algorithm, it has not shown any improvement in forecast skill compared to a standard 4DVar with static background error covariances when the full set of operationally available observations are used. We speculate that this is due to some fundamental limitations of the EnKF: a) The ensemble size which is currently (and in the near future) practical to run in operational environments will require some form of localization of the sampled error statistics, thus limiting the amount of information that is possible to extract from non-local observations; b) the sensitivity of the analysed state to the observational information comes only through the sampled background error statistics, while in standard 4DVar this information is implicitly propagated in a dynamically consistent way by the linearized and adjoint model and observation operators: when accurate linear models and their adjoints are available, as is the case at ECMWF, it is therefore beneficial to make use of them; c) Standard EnKF background errors and covariances are only estimated from the latest set of background forecasts: this can be sub-optimal when these forecasts are affected by non-negligible state dependent model errors (i.e., biases)

and it might be preferable (as it has also been found in the hybrid 4DVar-EDA case, Bonavita et al., 2014) to reduce these state dependent biases in the analysis by blending climatological information to the online estimates; d) Difficulty within the EnKF system of assimilating observations with strongly non-linear observation operators. We have so-far not been able to include rain-affected radiances in the EnKF assimilation and ambiguous winds from scatterometers use only the wind direction closest to the first guess. We note that these limitations also apply, partly or totally, to algorithms like 4D Ensemble Var (4D-En-Var, Liu et al., 2008) that mimic the EnKF in a variational context: this is possibly one of the main reasons why these algorithms have, so far, been unable to show any significant performance improvement with respect to a standard EnKF (Buehner et al., 2010b; see also Fairbairn et al., 2014, for a comparison in a toy model environment).

At ECMWF the operational hybrid formulation is using a sophisticated, modelled representation of the background error matrix (wavelet \mathbf{B} , Fisher, 2003) and an ensemble of perturbed 4DVar assimilation cycles (EDA) to provide an online estimate of the background errors (Bonavita et al., 2012; Bonavita et al., 2014). This strategy, together with a continuing effort to improve the \mathbf{B} model itself, has proven very successful.

The other hybrid approach that has been adopted in some global NWP Centres has been to combine static estimates of \mathbf{B} used in 3-4DVar systems with online estimates directly sampled from an EnKF (“hybrid \mathbf{B} covariance”, Clayton et al., 2013; Wang et al., 2013). The methodology of the hybrid gain EnDA presented in this paper is instead to combine the Kalman gain matrices of an EnKF and a static 4DVar. This approach can be shown to be functionally equivalent to the hybrid \mathbf{B} covariance method (Penny, 2014), and has some significant advantages. It has been simple to implement at ECMWF as it involves re-using available components of the IFS and it allows to make use in the EnKF framework of components of the 4DVar system (such as variational bias correction) that are simpler and cheaper to run in a variational context. The use of a static 4DVar in the re-centring step of the hybrid gain EnDA can be seen as a regularization procedure of the EnKF analysis which has two main benefits: a) reduces the effect of sampling noise and localization; and b) introduces climatological information on the background error covariances, which is important to reduce the effect of model biases in the analysis and it is not easy to do in a standard EnKF. On a more fundamental level, it allows independent development of the EnKF and 4DVar systems which will both have a beneficial influence on the resulting hybrid system. This is quite relevant for research work, where running multiple instances of the EnKF or the hybrid system can be impractical. The weight given in the hybrid to each of the analysis components will reflect their perceived accuracies and will thus be the main tuning parameter. It is remarkable that the HG-EnDA, without any specific tuning, is already competitive with a low resolution version (T511) of the hybrid 4DVar-EDA which is run operationally at ECMWF.

In terms of scalability and computational costs the ECMWF EnKF (and HG-EnDA) is very competitive. The scalability of the algorithm, especially in the LETKF formulation, is only limited by the throughput of the IO operations needed to read in the ensemble background and the observations at the start of the analysis; and the IO operations needed to write out the analysis ensemble at the end. In terms of computational costs the EnKF is more efficient than the ECMWF EDA configuration, so it would enable ECMWF to run an ensemble with a larger size and/or higher resolution.

One aspect that has not been reported in this work is the performance of the hybrid gain EnDA as an error cycling system for the high-resolution 4DVar analysis cycle: this will be described in detail in a

forthcoming paper. Preliminary results show, however, that also in this respect the HG-EnDA is competitive with the current operational configuration at ECMWF.

All the experiments presented in this work have been run with 6-hourly assimilation windows. This is different from the operational analysis cycle at ECMWF, which runs with a 12-hour assimilation window. This does not affect, however, the validity of the presented results, as internal experimentation at ECMWF has recently shown that the two cycling systems produce analyses and forecasts of comparable skill. In its current implementation, the HG-EnDA uses the ensemble mean as its control state. This limits the resolution of the control analysis to the resolution at which the whole ensemble can be practically run. It is possible to run a control member at higher resolution than the rest of the ensemble, thus having a unified high-resolution 4DVar - EnKF system able to concurrently cycle estimates of the atmospheric state and its error covariances: this will be one possible avenue for further development.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of Jeff Whittaker in the initiation of this project and for letting us benefit from his extensive experience in EnKF data assimilation. Also we would like to thank Tony McNally, Sean Healy and Alan Geer for providing the means of vertically localizing satellite radiances. George Mozdzyński provided us with a method for distributing the grid points in a weighted fashion. Anne Fouilloux made it possible for us to use the ODB in the EnKF. We have also benefited greatly from discussions with many colleagues at ECMWF in the development of the EnKF system

References

- Andersson, E. and H. Järvinen (1999): Variational quality control. *Q. J. R. Meteorol. Soc.*, **125**, 697–72.
- Barker, D. M. (1999). Var scientific development paper 25: The use of synoptic-dependent error structure in 3DVAR. *Met Office Tech. Rep.*, 2 pp. [Available from Met Office, FitzRoy Rd., Exeter, Devon EX1 3PB, United Kingdom].
- Bauer, P., A. J. Geer, P. Lopez and D. Salmond (2010): Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Q. J. R. Meteorol. Soc.*, **136**: 1868–1885.
- Berner, J., G. J. Shutts, M. Leutbecher and T. N. Palmer (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, *J. Atmos. Sci.*, **66**, 603–626.
- Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background error covariance Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, **131**, 1013–1043.
- Bishop, C. H. and D. Hodyss (2011). Adaptive Ensemble Covariance Localization in Ensemble 4D-VAR State Estimation. *Mon. Wea. Rev.*, 139, 1241–1255

- Bonavita M., L. Isaksen and E. Holm (2012). On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.*, **138**, 1540–1559.
- Bonavita M., E. Holm and L. Isaksen (2014). Hybrid Data Assimilation with EDA perturbations in the ECMWF 4D-Var. In preparation.
- Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He (2010a). Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Wea. Rev.*, **138**, 1550–1566.
- Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He (2010b). Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations. *Mon. Wea. Rev.*, **138**, 1567–1586.
- Buizza, R., M. Miller and T. N. Palmer (1999). Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2887–290.
- Clayton, A. M., A. C. Lorenc and D. M. Barker (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.*, **139**, 1445–1461.
- Dee, D. P. (2004). Variational bias correction of radiance data in the ECMWF system. In Proc. *ECMWF Workshop on the Assimilation of High Spectral Resolution Sounders in NWP*, 28 June–1 July 2004, Reading, UK, 97–112.
- Dee, D. P. (2005). Bias and data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3323–3343.
- Derber, J. C. and W.-S. Wu (1998). The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299.
- ECMWF (2011). SMS users guide. (Available from: <http://www.ecmwf.int/publications/manuals/sms/documentation/index.html>).
- ECMWF (2014). Integrated Forecasting System Documentation, Cycle 40R1, 2014. (Available from: <http://www.ecmwf.int/research/ifsdocs/CY40r1/>).
- Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF .No.: 54 . p.: 539-560 . Ocean Dynamics
- Fairbairn, D., S. R. Pring, A. C. Lorenc and I. Roulstone (2014), A comparison of 4DVar with ensemble data assimilation methods. *Q. J. R. Meteorol. Soc.*, **140**: 281–294.
- Fertig, E. J., S.-J. Baek, B. R. Hunt, E. Ott, I. Szunyogh, J. A. Aravéquia, E. Kalnay, H. Li and J. Liu (2009). Observation bias correction with an ensemble Kalman filter. *Tellus A*, **61**, 210–226.
- 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. and S. Gürol (2014). Parallelisation in the time dimension of four-dimensional variational data assimilation, submitted to *Q. J. R. Meteorol. Soc.*
- Fouilloux, A. (2010) ODB (Observational DataBase): past, present and future. In Proc. *Fourteenth Workshop on Use of High Performance Computing in Meteorology*, 1–5 November 2010, ECMWF, Reading, UK (Available from http://www.ecmwf.int/newsevents/meetings/workshops/2010/high_performance_computing_14th/)
- Gaspari, G. and S. E. Cohn (1999). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723–757.
- Gauthier, P. and J.-N. Thépaut (2001). Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.*, **129**, 2089–2102.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide and B. R. Hunt (2011). Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.*, **139**, 511–522.
- Hamill, T. M. and C. Snyder (2000). A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- Houtekamer, P. L. and H. L. Mitchell (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Houtekamer, P. L. and H. L. Mitchell (2005). Ensemble Kalman filtering. *Q.J.R. Meteorol. Soc.*, 131: 3269–3289. Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek, N. Gagnon (2014). Higher Resolution in an Operational Ensemble Kalman Filter. *Mon. Wea. Rev.*, **142**, 1143–1162.
- Houtekamer, P. L., Bin He and H. L. Mitchell (2014). Parallel Implementation of an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **142**, 1163–1182.
- Hunt, B. R., E. J. Kostelich and I. Szunyogh (2007). Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.
- Isaksen, L. (2012). Data Assimilation on future computer architectures. In Proc. *Seminar on Data Assimilation for Atmosphere and Ocean*, 6–9 September 2011, ECMWF, Reading, UK (Available at <http://www.ecmwf.int/publications/>)
- Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher and L. Raynaud (2010). Ensemble of data assimilations at ECMWF. *ECMWF Tech. Memo. No. 636*. (Available from: <http://www.ecmwf.int/publications/>)
- Kepert, J. D. (2009). Covariance localisation and balance in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.*, **135**, 1157–1176.
- Liu, C., Q. Xiao and B. Wang (2008). An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 3363–3373.

- Lopez, P. (2011): Direct 4D-Var Assimilation of NCEP Stage IV Radar and Gauge Precipitation Data at ECMWF. *Mon. Wea. Rev.*, **139**, 2098–2116.
- Lynch, P. and X.-Y. Huang (1992). Initialization of the HIRLAM model using a Digital Filter. *Mon. Wea. Rev.*, **120**, 1019-1034.
- Lorenc, A.C. (2003). The potential of the ensemble Kalman filter for NWP: a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, **129**, 3183–3203.
- Mitchell, H. L. and P. L. Houtekamer (2000). An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433.
- Miyoshi T., Y. Sato and T. Kadowaki (2010). Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System. *Mon. Wea. Rev.*, **138**, 2846–2866.
- NASA, (2002). Advanced Microwave Sounding Unit-A (AMSU-A) Instrument Guide. Available at http://disc.gsfc.nasa.gov/AIRS/documentation/amsu_instrument_guide.shtml
- Ott, E., B. H. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich and co-authors (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.
- 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 Tech. Memo. No. 598*, ECMWF, Shinfield Park, Reading RG2-9AX, UK.
- Penny, G. S. (2014). The Hybrid Local Ensemble Transform Kalman Filter. *Mon. Wea. Rev.*, **142**, 2139–2149.
- Polavarapu, S. M., S. Ren, A. Clayton, Y. Rochon and D. Sankey (2004). On the relationship between incremental analysis updating and incremental digital filtering. *Mon. Wea. Rev.*, **132**, 2495–2502.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons (2000). The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143–1170.
- Saarinen, S. (2004). ODB user guide. Available at http://old.ecmwf.int/research/ifsdocs/CY28r1/pdf_files/odb.pdf
- Shutts, G. and T. Palmer (2004) The use of high resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In Proc. *ECMWF/CLIVAR Workshop on Simulation and Prediction of Intra-seasonal Variability with Emphasis on the MJO*, 3–6 November 2003, 83–102.
- Shutts, G. J. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Q. J. R. Meteorol. Soc.*, **131**, 3079–3102.

- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield and J. A. Yorke (2008). A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60**, 113–13.
- Tavolato, C. and L. Isaksen, (2014). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. Accepted for publication in *Q. J. R. Meteorol. Soc.*
- Tippett, M. K., J. L. Anderson, G. H. Bishop, T. M. Hamill and J. S. Withaker, (2003). Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Wang, X. (2010). Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework. *Mon. Wea. Rev.*, **138**, 2990–2995.
- Wang, X., D. Parrish, D. Kleist and J. S. Whitaker (2013). GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.
- Whitaker, J. S. and T. M. Hamill (2002). Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Whitaker, J. S., G. P. Compo and J.-N. Thepaut (2009). A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations. *Mon. Wea. Rev.*, **137**, 1991–1999.
- Whitaker, J. S. and T. M. Hamill (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089.