Tellus
SERIES A
DYNAMIC
METEOROLOGY
AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 385–401
Printed in Singapore. All rights reserved

TELLUS

The ETKF rescaling scheme in HIRLAM

By JELENA BOJAROVA[1]*, NILS GUSTAFSSON[2], ÅKE JOHANSSON[2] and OLE VIGNES[1],
[1]*The Norwegian Meteorological Institute, P.O. Box 43, Blindern, NO-0313 Oslo, Norway;* [2]*The Swedish Meteorological and Hydrological Institute, SMHI, SE-601 76, Norrköping, Sweden*

(Manuscript received 11 January 2010; in final form 4 January 2011)

ABSTRACT

The ETKF rescaling scheme has been implemented into the HIRLAM forecasting system in order to estimate the uncertainty of the model state. The main purpose is to utilize this uncertainty information for modelling of flow-dependent background error covariances within the framework of a hybrid variational ensemble data assimilation scheme. The effects of rank-deficiency in the ETKF formulation is explained and the need for variance inflation as a way to compensate for these effects is justified. A filter spin-up algorithm is proposed as a refinement of the variance inflation. The proposed spin-up algorithm will also act to prevent ensemble collapse since the ensemble will receive 'fresh blood' in the form of additional perturbation components, generated on the basis of a static background error covariance matrix. The resulting ETKF-based ensemble perturbations are compared with ensemble perturbations based on targeted singular vectors and are shown to have more realistic spectral characteristics.

1. Introduction

The Extended Kalman Filter (EKF, Kalman and Bucy, 1961) and the Unscented Filter (UF, Julier and Uhlmann, 1997) are two main strategies in sequential data assimilation for non-linear and/or non-Gaussian models. The EKF is based on a tangent-linear approximation of the non-linear state space system; the Kalman Filter (Kalman, 1960) is applied in this tangent-linear environment. The UF uses a set of discretely, deterministically sampled points (called 'sigma' points), that are related to the covariance eigenvectors, to parametrize the mean and the covariance of the model state. The non-linear development and update of the eigenvectors for the forecast and analysis error covariance matrices are of primary interest in this approach. These two methods cannot in the foreseeable future be applied in meteorological data assimilation directly because they require explicit manipulation of the forecast and analysis error covariance matrices. The huge dimensionality of the model state variable vector that describes the development of the atmosphere ($\approx 10^8$) makes this impossible.

The variational data assimilation techniques (3D-Var, Parrish and Derber, 1992, and 4D-Var, Le Dimet and Talagrand, 1986; Courtier et al., 1994) are successfully implemented at weather services worldwide and provide an approximate solution to the sequential update problem. The observations are collected over 'an assimilation window', the prior Gaussian assumption about the model state, valid at the beginning of this assimilation window, is formulated and the posterior mode is calculated numerically by maximizing the posterior density function of the model state. The scheme is flexible and it treats efficiently even 'complicated' observations that are non-linearly related to the model state and/or have non-Gaussian observation error distributions.

The posterior mode, dynamically integrated forward to the beginning of the next assimilation window, is taken as a prior mean of the model state at the next assimilation cycle. However, in the variational data assimilation there is no algorithm available to update the prior covariance. An identical static full-rank covariance is used at the beginning of each assimilation window. This is clearly unsatisfactory and degrades the performance of the variational data assimilation schemes. Intuitively, one can hope that a Hybrid scheme, which merges variational and sequential data assimilation schemes, will outperform variational data assimilation. The variational scheme could efficiently be used to estimate the model state and the sequential data assimilation scheme could be used to update the uncertainty about the estimate. Hamill and Snyder (2000) suggested that the background error covariance of a hybrid variational ensemble data assimilation could be formed as a linear combination of a static 3D-Var based covariance and a flow-dependent ensemble-based covariance. Lorenc (2003) suggested an alternative hybrid variational ensemble data assimilation algorithm that solves the

*Corresponding author.
e-mail: Jelena.Bojarova@met.no
DOI: 10.1111/j.1600-0870.2011.00513.x

minimization problem by an augmentation of the assimilation control vector and by adding an additional cost function term, representing the weights given to the different ensemble members. Wang et al. (2007b) have proven that the two formulations of the hybrid variational ensemble data assimilation under certain assumptions are equivalent.

The Ensemble Kalman Filters (EnKF, Evensen, 1994) provide a Monte Carlo approximation to the EKF, in which a sample of model states is used to estimate and propagate the mean and the covariance of the model state forward in time, based on a reduced rank approximation. In this approach the Kalman filtering becomes feasible even in the case of large dimensional systems. A Gaussian approximation of the probability distribution for the forecast, model and observation errors as well as a statistical linearization of the state space system underlie this method. The EnKF approach receives increased attention among scientists at present. A large number of different implementations of the EnKF algorithm have been proposed: Perturbed Observations EnKF (Houterkamer and Mitchell, 1998), Resampling EnKF (Anderson and Anderson, 1999; Bengtsson et al., 2003), Reduced Rank EnKF (Verlaan and Heemink, 1997; Pham et al., 1998; Heemink et al., 2001) and Ensemble Square-Root Kalman Filters (EnSRKF, Cohn et al., 1998; Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002; Ott et al., 2004; Sakov and Oke, 2008).

The EnSRKFs build a bridge between the EKF and the UF. Similarly to the UF approach, a set of deterministically selected points is used to model the mean and the covariance of the model state, their update from observations and their propagation forward in time. Furthermore, the selection of these deterministic points (the ensemble members) is based on Kalman Filter theory. The rank of the model state covariance modelled in this way is limited by the amount of independent ensemble members. If the affordable size of the ensemble is much smaller than the dimensionality of the model state variable, the modelled covariance becomes severely rank deficient.

The objective of this study is to investigate the performance of the Ensemble Transform Kalman Filter (ETKF, Bishop et al., 2001) rescaling scheme within the HIRLAM forecasting system (Undén et al., 2002) primarily for operational applications. This requires the ensemble size to be very small. The ETKF is one version of the EnSRKF, where the square-root of the analysis error covariance is obtained by multiplying the square-root of the forecast error covariance from the right by a transformation matrix with the dimension of the ensemble size. In our study the emphasis is on whether a small size ensemble of ETKF perturbations can be used to extract meaningful flow-dependent, inhomogeneous structures of the analysis and forecast errors of the model state. We outline the basic principles of linear filtering, we use a non-standard analysis technique to investigate limitations of the ETKF rescaling scheme imposed by the severe rank-deficiency and we propose a method to handle these limitations.

2. Theoretical basis

2.1. The Kalman Filter recursions

Let $t_0 < t_1 < \ldots < t_i < \ldots < t_n$ denote a sequence of discrete moments in time when observations are collected. In general, the meteorological data assimilation problem can be expressed in a state space model terminology (Durbin and Koopman, 2001). In the following index i refers to time t_i.

$$\begin{aligned} y_i &= \mathcal{H}_i(\mathcal{X}_i) + \epsilon_i \\ \mathcal{X}_i &= \mathcal{M}(i-1,i)\mathcal{X}_{i-1} + T_i\eta_i; \quad 1 \le i \le n, \end{aligned} \tag{1}$$

where y_i is a p-dimensional vector of observations, ϵ_i is a p-dimensional vector of observation errors, $\mathcal{X}_i$ is an m-dimensional vector of a model state and $\mathcal{H}_i$ is a deterministic forward observation operator, which projects the model state to the observations, valid at time t_i, $\mathcal{M}(i-1,i)$ is a deterministic forward model propagator from the previous assimilation time to the current one, η_i is a q-dimensional vector of 'a model error' and T_i is a $m \times q$ transformation matrix, which projects the model error space to the model state space. The numerical discretization, the subgrid variability and unresolved structures of the forecast and analysis errors are possible sources of the model errors. The probability distributions of observation errors, model errors and the model state at the origin, time t_0, are specified. In general, both $\mathcal{H}_i$ and $\mathcal{M}(i-1,i)$ might be non-linear.

Provided (K.1) the forward observation operator and the forward model propagator are linear operators (in this case we denote them H_i and M^i_{i-1}) and provided (K.2) the probability distributions of the model errors, the observation errors and the initial model state are Gaussian, the model state $\mathcal{X}_i$ will have a Gaussian distribution with the mean and the covariance determined via the Kalman Filter recursions. The Kalman Filter recursions describe the sequential update of the first two moments of the model state, $E(\mathcal{X}_i|\cdot)$ and $\mathrm{Cov}(\mathcal{X}_i|\cdot)$, conditional on the whole history of available observations, before time t_i, $\mathcal{Y}_{i-1}$, and including the latest one, $\mathcal{Y}_i$.

$$\begin{aligned} x_i^f &= E(\mathcal{X}_i|\mathcal{Y}_{i-1}) = M^i_{i-1}x^a_{i-1}, \\ B_i^f &= \mathrm{Cov}(\mathcal{X}_i|\mathcal{Y}_{i-1}) = M^i_{i-1}B^a_{i-1}\left(M^i_{i-1}\right)^T + Q_i, \quad i \ge 1, \\ x_i^a &= E(\mathcal{X}_i|\mathcal{Y}_i) = x_i^f + K_i\left(y_i - H_i x_i^f\right), \\ B_i^a &= \mathrm{Cov}(\mathcal{X}_i|\mathcal{Y}_i) = B_i^f\left(I - (K_iH_i)^T\right), \\ K_i &= B_i^f H_i^T\left(R_i + H_iB_i^fH_i^T\right)^{-1}, \end{aligned} \tag{2}$$

where R_i is the observation error covariance and Q_i is the model error covariance. We refer to x_i^f and x_i^a as the forecast and the analysis, and to B_i^f and B_i^a as the forecast error covariance and the analysis error covariances, respectively. Matrix K_i is the so-called Kalman gain and describes the impact of the latest observations y_i on the estimate of the model state $\mathcal{X}_i$. The observation error covariance, R_i, is often modelled as a diagonal matrix containing information on the precision of the observation vector. The parametrization of the model error covariance, Q_i, is not

straightforward. Preservation of the dynamical balances is one of the crucial requirements in meteorological data assimilation. A simplistic approach is to assume Q_i to be proportional to the forecast error covariance, $Q_i \approx \gamma B_i^f$.

2.2. The variational data assimilation

In meteorological data assimilation both the forward model propagator and the forward observation operator are in essence non-linear, therefore the Kalman filter theory is not exactly applicable. One usual approximation of the solution to the problem (eq. 1) is based on the incremental approach in the tangent linear environment, where the Kalman filter theory can be applied. The data assimilation problem is formulated for the increment of the model state $\delta\mathcal{X}_i = \mathcal{X}_i - x_i^b$ (Courtier et al., 1994), where x_i^b is an a priori given estimate of the true model state, often called the 'background state'. The original non-linear state space model is approximated by a tangent linear one, linearizing both the dynamical propagator and the observation operator around x_i^b. Ideally, both the x_i^b and the prior covariance matrix B_i^f should contain impact from previously assimilated observations.

In the tangent linear environment, provided that the conditions (K.1) and (K.2) hold, the analysis state, the expected value of the model state increment given observations, coincides with the mode of the posterior probability density. Thus the analysis state can be obtained numerically by minimizing the corresponding cost function.

However, due to the huge dimensionality of the model state the explicit manipulations with the model state covariance are not feasible at present. Therefore, the prior covariance B_i^f is approximated by an identical full-rank covariance, $B_{\mathrm{3D-Var}}$, for each assimilation window (Parrish and Derber, 1992). The minimization of the cost function in the HIRLAM 3D-Var (Gustafsson et al., 2001) is performed in the 'control vector' space, v_i,

$$v_i = \left(B_{30-\mathrm{Var}}\right)^{-1/2} \delta x_i \tag{3}$$

and the explicit inverse of the large dimensional B_i^f is avoided.

2.3. The Ensemble Transform Kalman Filter (ETKF)

To approximate the update and evolution of the model state covariance, $B_{i-1}^a \rightarrow B_i^f \rightarrow B_i^a \rightarrow B_{i+1}^f$, we have implemented the ETKF, a version of the EnSRKF (Tippet et al., 2003). In the EnSRKF approach, the ensemble of the deterministically selected standardized analysis perturbations,

$$Z_{i-1}^a = \left(z_{i-1,1}^a, \ldots, z_{i-1,N}^a\right),$$

is assumed to represent the square-root of the analysis error covariance matrix,

$$Z_{i-1}^a \left(Z_{i-1}^a\right)^T = B_{i-1}^a.$$

The ensemble Z_{i-1}^a is centred around the mean by construction, namely the analysis perturbations are forced to sum to zero,

$$\overline{z}_{i-1}^a = \sum_{n=1}^{N} z_{i-1,n}^a = 0.$$

This ensemble is propagated forward numerically by the non-linear model

$$z_{i,n}^f = \frac{1}{\sqrt{N-1}}\Big[\mathcal{M}_{i-1,i}\left(x_{i-1}^a + x_{i-1,n}^a\right) - \overline{\mathcal{M}_{i-1,i}\left(x_{i-1}^a + x_{i-1,n}^a\right)}\Big] \tag{4}$$

to reconstruct the prior covariance matrix at the beginning of the next assimilation cycle

$$Z_i^f \left(Z_i^f\right)^T = B_i^f. \tag{5}$$

Then the forecast ensemble Z_i^f is transformed into the ensemble of analysed states Z_i^a, which is considered to represent the square-root of B_i^a, namely

$$Z_i^a \left(Z_i^a\right)^T = B_i^f - B_i^f \bar{H}_i^T \left(\bar{H}_i B_i^f \bar{H}_i^T + R_i\right)^{-1} \bar{H}_i B_i^f. \tag{6}$$

The ETKF uses a right side transform to construct the square-root $(B_i^a)^{1/2}$ from the square-root $(B_i^f)^{1/2}$,

$$Z_i^a = Z_i^f T_i.$$

An overline $\overline{(\cdot)}$ denotes an ensemble mean.

2.4. The unique symmetric transformation matrix T_i

The symmetric transformation T_i, which rescales forecast perturbations to analysis perturbations, is calculated following Bishop et al. (2001).

$$T_i = C_i (G_i + I)^{-1/2} C_i^T, \tag{7}$$

where G_i is a $(N-1) \times (N-1)$ diagonal matrix with non-zero eigenvalues of

$$B_{ens,i}^f = \left(Z_i^f\right)^T \bar{H}_i^T R^{-1} \bar{H}_i Z_i^f$$

on its diagonal; C_i is a $N \times (N-1)$ matrix of the corresponding eigenvectors of $B_{ens,i}^f$, namely

$$C_i G_i C_i^T = \sum_{j=1}^{N-1} \lambda_{i,j} c_i^j \left(c_i^j\right)^T = B_{ens,i}^f, \tag{8}$$

where $B_{ens,i}^f$ is the $N \times N$-dimensional forecast error covariance matrix in observation space, normalized by the observation error variance and transformed into ensemble space; $\lambda_{i,1} > \lambda_{i,2} > \ldots > \lambda_{i,N-1} > 0$ are non-zero eigenvalues of $B_{ens,i}^f$ and c_i^j, $1 \le j \le N-1$, are the corresponding eigenvectors of the $B_{ens,i}^f$, which form columns of C_i. Here we assume that the forecast perturbations are centred by construction, that is $\lambda_{i,N} = 0$. In order to simplify notations we remove the suffix i in the rest of this section since the notations refer to the same time moment t_i, unless stated differently. The transformation matrix T rotates

ensemble perturbations along directions of the eigenvectors of B^f_{ens}, downscales the rotated perturbations with weights proportional to the eigenvalues of B^f_{ens} and rotates the downscaled perturbations backwards.

The backward rotation of the downscaled perturbations, the spherical simplex transform (Wang et al., 2004), assures the centring of the transformed perturbations $Z^a = Z^f T$ around the mean, in such a way that it makes optimal use of the available degrees of freedom, that is, $N-1$. Such a rotation preserves the spread of the ensemble and assures equal likelihood for ensemble members, assuming a deterministic sampling from a Gaussian distribution around the background state.[1] Besides that, the symmetric transformation is unique (Sakov and Oke, 2008) and searches for the Z^a_n which has the largest pattern anomaly correlation with the Z^f_n (Ott et al., 2004).

2.5. The analysis error covariance matrix

Let r denote the rank of the true forecast error covariance matrix in observation space normalized with the observation error covariance, $R^{-1/2}\mathcal{H}(B^f)^{1/2}(R^{-1/2}\mathcal{H}(B^f)^{1/2})^T$. Let us assume that the observation operator is linear and that the ensemble size is larger than the rank r, for example, $N = r+1$ and let B^f_{HX} denote the standardized forecast error covariance matrix in observation space, reconstructed from the ensemble of perturbations,

$$B^f_{HX} = R^{-1/2}\mathcal{H}Z^f(R^{-1/2}\mathcal{H}Z^f)^T.$$

In this case $B^f_{ens} = (Z^f)^T\mathcal{H}^T R^{-1}\mathcal{H}Z^f$ and B^f_{HX} have the same set of r non-zero eigenvalues,

$$\mathrm{diag}(G) = \left(\lambda^r_1, \lambda^r_2, \ldots, \lambda^r_r, 0\right), \quad \lambda^r_1 > \lambda^r_2 > \ldots > \lambda^r_r > 0. \tag{9}$$

After the ETKF rescaling has been applied, the analysis error covariance matrix in observation space normalized with the observation error covariance and transformed to ensemble space becomes

$$\begin{aligned} B^a_{ens} &= (Z^a)^T\mathcal{H}^T R^{-1}\mathcal{H}_i Z^a \\ &= T^T(Z^f)^T\mathcal{H}^T R^{-1}\mathcal{H}Z^f T \\ &= T^T B^f_{ens} T \\ &= C(I_{r\times r}+G)^{-1/2}C^T CGC^T C(I_{r\times r}+G)^{-1/2}C^T. \end{aligned}$$

Here $I_{r\times r}$ denotes an r-dimensional identity matrix. Thus,

$$B^a_{ens} = \sum_{j=1}^{r} \frac{\lambda^r_j}{1+\lambda^r_j} c^j (c^j)^T, \tag{10}$$

where c^j refers to the jth eigenvector of the B^f_{ens}.

A forecast error variability in the non-leading directions, comparable with the variability of the observation error, becomes important for reconstruction of the analysis error covariance matrix. The ETKF update of the uncertainty about the model state estimate illustrates 'the basic principles' of the linear filtering. Note that this linear filtering is applied in observation space.

(i) The analysis error variance never exceeds either the observation error variance or the forecast error variance.

(ii) The uncertainty is reduced at most in the most uncertain directions, $\lambda_j \gg 1$, and the uncertainty is almost untouched in the least uncertain directions, for $\lambda_j \ll 1$.

(iii) The analysis error covariance will have the same rank and the same set of eigenvectors as the forecast error covariance.

(iv) The eigenvectors preserve their rankings: for any $\lambda_j > \lambda_k$, holds

$$\frac{\lambda_j}{1+\lambda_j} > \frac{\lambda_k}{1+\lambda_k}.$$

This means that the leading eigenvector of $B^f_{ens,i}$ will be the leading eigenvector of B^a_{ens}.

The full-rank analysis error covariance has a relatively wide 'white spectrum'. At the same time the model dynamics will develop it to a 'red' forecast error covariance concentrated on a relatively small amount of leading eigenvectors. The experience shows that the sequential application of the ETKF rescaling scheme works as a generalization of the error breeding technique (Toth and Kalnay, 1997) and is able to find directions of the strongest forecast error growth.

2.6. The multiplicative and the additive inflations

In meteorological applications of the ETKF rescaling, the number of perturbations $Z^{f,n}$, $n = 1, \ldots, N$, is significantly smaller than the rank of the true forecast/analysis error covariance matrix in observation space. In this case the ETKF rescaling works in the observation space compressed to ensemble space. Taking the dimensionality of the observation network ($\mathcal{O}(10^5)$) into account, some important information will always be lost or mistreated during this compression, because the operationally affordable ensemble size is quite limited, order of $o(10^2)$. The covariances of the model state reconstructed from such an ensemble of perturbations become rank deficient. The ETKF rescaling scheme must be modified to account for this rank deficiency.

According to the methodology of the ETKF rescaling scheme, the ensemble of analysis states is evolved into the ensemble of the forecast states, the square-root of the forecast error covariance, at the next assimilation time. Therefore the spread of the analysis ensemble should reflect both the initial uncertainty in the directions sampled by the ensemble of the limited size and the impact from the uncertainty in the non-sampled directions.

When real observations are introduced into the numerical model, the model error should be taken into account. The model error originates from a broad range of sources, including among

[1] We refer to selection of quantiles for a certain distribution as to the deterministic sampling from this distribution.

others deficiencies in the numerical scheme, description of dynamical processes, parametrization of physical processes and the modelling of the covariance. We have adopted an oversimplistic method and have parametrized the model error covariance as being proportional to the forecast error covariance, mainly due to our lack of knowledge about the structure of the model error covariance and in order to preserve important dynamical balances. The multiplicative inflation applied to the ensemble of analysis states accounts both for the presence of model error and for the effect of the space compression.

$$Z^a = \Pi Z^f T. \tag{11}$$

We have also employed an additive inflation, which we discuss below in detail, in order to compensate for deficiencies that are introduced with the simple multiplicative inflation. The multiplicative and the additive inflations work in different ways. The multiplicative inflation preserves the eigenvectors of the covariance and changes the spectrum increasing all eigenvalues by the same factor. The additive inflation allows to modify the spectrum in a more elaborated way but it modifies the eigenvectors at the same time.

The trace the covariance matrix is usually used as an overall measure of the spread of the ensemble. In the rank-deficient ETKF rescaling scheme the trace of the full-rank forecast error covariance matrix

$$\text{Trace}\left(B^f_{ens}\right) = \sum_{k=1}^{r} \lambda^r_k \tag{12}$$

will be replaced by the trace of the rank-deficient forecast error covariance matrix

$$\text{Trace}\left(B^f_{ens}\right) = \sum_{k=1}^{N-1} \lambda^{N-1}_k. \tag{13}$$

The superscripts r and $N-1$ denote the dimensionality of B^f_{ens}.

The multiplicative inflation will modify the spectrum of the rank-deficient analysis error covariance matrix such that the trace of the rank-deficient forecast error covariance would be physically meaningful. For the construction of the inflation factor Π_i, we follow the methodology proposed by Wang and Bishop (2003). The inflation factor Π_i is thus defined in a recurrent way,

$$\Pi_i = \sqrt{\alpha_i}\Pi_{i-1},$$

where the parameter α_i is such that

$$\tilde{d}_i^2 := \sum_{j=1}^{p} \left(\frac{y_j - \mathcal{H}_{ji} x_i^b}{\sigma_{o,ji}} \right)^2 \approx \alpha_i \sum_{n=1}^{N-1} \lambda_{ni}^{N-1} + p. \tag{14}$$

The calculation is based on an optimal adjustment of the ensemble spread (the right-hand side of this equality) to the normalized variance of innovations (the left-hand side of this equality), averaged over observation space.

When the multiplicative inflation is applied, the eigenvalues $\lambda^N_j,\ 1 \le j \le N-1$ of B^f_{ens}, corresponding to the rank-deficient ensemble, $N < r$, will overestimate the corresponding eigenvalues $\lambda^r_j,\ 1 \le j \le N-1$ of the full rank covariance. This type of property is discussed by Keppert (2004) and Wang et al. (2007a), for example. We may speculate that this overestimation is stronger for smaller λ^r_j. Such an overestimation will lead to a flatter spectrum for the analysis error covariance.

When inflation Π is applied, the set of $N-1$ non-zero eigenvalues of B^a_{ens} or B^a_{HX} will be

$$\left(\Pi^2 \frac{\lambda^N_1}{1+\lambda^N_1}, \Pi^2 \frac{\lambda^N_2}{1+\lambda^N_2}, \ldots, \Pi^2 \frac{\lambda^N_{N-1}}{1+\lambda^N_{N-1}} \right). \tag{15}$$

The larger the λ^N_j is the closer the ratio $\frac{\lambda^N_1}{1+\lambda^N_1}$ approaches 1. As a result, the patterns of the analysis error in directions corresponding to smaller eigenvalues are amplified relative to the patterns in the leading directions. This subject is extensively discussed in Wang et al. (2007a). They proposed an improved derivation of the ETKF rescaling matrix, which results in a less 'white' spectrum of the analysis error covariance matrix. Their scheme is based on an estimation of the true rank of the forecast error covariance in observation space and it requires the history of innovations.

Notice that the analysis error variance in the direction of the eigenvector c^j is equal.

$$c^j B^a_{ens} (c^j)^T = \Pi^2 \frac{\lambda_j^{N-1}}{1+\lambda_j^{N-1}} = \Pi^2 \frac{1}{\frac{1}{1} + \frac{1}{\lambda_j^{N-1}}}. \tag{16}$$

This estimate is not consistent. The forecast error variance of the rank-deficient system in the direction c^j, $\lambda^{N-1}_{j,i}$, reflects the compression from the r-dimensional observation space to the $N-1$-dimensional ensemble space. To be consistent, the observation error variance must be increased correspondingly, by a factor $\kappa = \frac{r}{N-1}$, to account for the space compression. There is no straight-forward procedure to estimate the rank r of the observation space.

The corrected spectrum of the analysis error variance is

$$\text{diag}\left(B^{\text{CORR},a}_{ens,i}\right) = (\Pi^*)^2 \left(\frac{\kappa\lambda_1^{N-1}}{\kappa+\lambda_1^{N-1}}, \frac{\kappa\lambda_2^{N-1}}{\kappa+\lambda_2^{N-1}}, \ldots, \frac{\kappa\lambda_{N-1}^{N-1}}{\kappa+\lambda_{N-1}^{N-1}} \right), \tag{17}$$

where Π^* is an inflation factor accounting for the impact of unresolved structures of the analysis error covariance on the structures of the forecast error covariance. The corrected analysis error covariance has a more realistic spectrum with a relatively larger spread in the direction of the leading eigenvectors.

In this case the rescaling matrix T_i (eq. 7) should be modified to

$$T^* = C\Pi^* \left(I_{(N-1)\times(N-1)} + \frac{1}{\kappa} G \right)^{-1/2} (C)^T. \tag{18}$$

This corrected rescaling matrix is similar to the improved rescaling matrix proposed by Wang et al. (2007a). The inflation factor can provide some guidance for estimating the parameter

κ, the rate of compression. However, tuning must be applied. Too small values of κ will leave the trace(B^a_{ens}) unchanged. Too large values of κ will reduce the spread in directions corresponding to small eigenvalues too strongly. This can result in a further increase of κ.

We propose an alternative scheme to 'spin-up' the filter, that is, to provide the analysis error covariance with a larger spread in the directions of the leading eigenvectors. This scheme is based on an advanced additive inflation having different impact along different eigenvectors c^j. First, an empirical downscaling matrix $A_{\mathrm{DSC}} = \mathrm{diag}(a_1, a_2, \ldots, a_{N-1})$ is applied to the raw ETKF rescaling perturbations in order to prohibit the amplification of the analysis error in the non-leading directions.

$$X^{a,\mathrm{DSC}} = X^f C(G + I_{(N-1)\times(N-1)})^{-1/2} A_{\mathrm{DSC}}(C)^T.$$

The construction of these downscaling weights is based on ad hoc considerations ($0.99 = a_1 > a_2 > \ldots > a_{N-1} = 0.5$). The weights are set up in such a way that the eigenvalues of B^a_{ens} employing the multiplicative inflation and the downscaling do not exceed the corresponding eigenvalues of B^f_{ens}, namely

$$\Pi^2 \frac{\lambda_j^{N-1} a_j}{1 + \lambda_j^{N-1}} \leq \lambda_j^{N-1}.$$

Secondly, in order to conserve the trace of the analysis error covariance, the corresponding part of the spectrum

$$\left(\Pi^2 \frac{\lambda_1^{N-1}}{1 + \lambda_1^{N-1}} \left(1 - a_1^2\right), \ldots, \Pi^2 \frac{\lambda_{N-1}^{N-1}}{1 + \lambda_{N-1}^{N-1}} \left(1 - a_{N-1}^2\right) \right)$$

is replaced with the structures of the static error covariance matrix $B_{\mathrm{3D-Var}}$. The resulting analysis ensemble is an additive combination of two ensembles

$$X^{a,B\mathrm{3D-V}} = X^{a,\mathrm{DSC}} + Y^{a,B\mathrm{3D-V}}, \tag{19}$$

where $Y^{a,B\mathrm{3D-V}}$ is the contribution from the structures of the static forecast error covariance $B_{\mathrm{3D-Var}}$,

$$Y^{a,B\mathrm{3D-V}} = \gamma Y C_Y \left(I_{(N-1)\times(N-1)} - A^2_{DSC}\right)^{1/2} (C)^T.$$

Here Y is a centred N-dimensional ensemble of random perturbations, $Y_{n,} \sim \mathcal{N}(0, B_{\mathrm{3D-Var}})$, $1 \leq n \leq N$, such that $\sum_{n=1}^{N} Y_n = 0$. C_Y is a matrix of the $N-1$ non-zero eigenvectors of

$$B^{f,\mathrm{3D-Var}}_{ens} = \frac{1}{N-1} (\mathcal{H} Y^T R^{-1} \mathcal{H} Y),$$

and γ is a scalar downscaling of the forecast error variance to the analysis error variance.

This additive combination of ensembles is constructed such that impact from the structures of the $B_{\mathrm{3D-Var}}$ is largest in the least observable directions, and is restricted to be small in the direction of the leading dynamical instabilities. We consider that such a merging allows to exploit efficiently the ETKF rescaling scheme in determining the flow-dependent structures. Besides that, the merging of the ETKF perturbation with the random structure of $B_{\mathrm{3D-Var}}$ injects 'fresh blood' from a full-rank system each assimilation time, that may help to capture the growth of uncertainty in the directions not included in a small ensemble originally.

3. Configuration of the experiment

An ensemble of $N = 8$ initial perturbed states $(x_{0,1} + x^b_0, \ldots, x_{0,N} + x^b_0)$, centred around the background state x^b_0, was created by sampling randomly from structures of the static forecast error covariance $B_{\mathrm{3D-Var}}$ (Berre, 2000). This forecast error covariance is full rank and contains statistically derived balances and structures, assuming homogeneity and isotropy. The numerical forward propagator $\mathcal{M}(i-1, i)$ is the reference HIRLAM (Undén et al., 2002), which runs on a domain in a rotated lat–lon grid [the South Pole Coordinates are (-40°LAT, 22°LON)] with $N_l = 40$ vertical levels and a horizontal resolution of 0.20°. The model state variable is $(u_{ijk}, v_{ijk}, T_{ijk}, q_{ijk}, ln(ps)_{ij})$, $1 \leq i \leq N_x$, $1 \leq j \leq N_y$, $1 \leq k \leq N_l$, $N_x = 302$, $N_y = 260$. The lateral boundaries are perturbed using the EuroTEPS forecast perturbations, initialized each analysis time. The EuroTEPS perturbations form the basis for the Ensemble Prediction System (EPS) applied at the Norwegian Meteorological Institute (Frogner and Iversen, 2002). EuroTEPS perturbations are obtained by applying a targeting procedure (Buizza, 1994) to the ECMWF's global EPS in order to construct perturbations with maximum growth of dry total energy over a specified area, the Northern Europe. The EPS is based on singular vectors with an optimization time of 48 h. The initial EuroTEPS and the evolved EuroTEPS perturbations, valid at the same time, are combined together to obtain patterns which better represent structures of the analysis error. Evolved singular vectors have been subject to non-linear model development and they therefore better represent more mature stages of real atmospheric developments. The ETKF rescaling is sequentially applied during the period 1200 UTC 12 August–0000 UTC 24 August 2007 each 6 h, when the 3D-Var data assimilation is performed on the background state. Only conventional observations (TEMP, PILOT, SHIP, AIREP, DRIBU and SYNOP) are used in the assimilation, no satellite information is included. This period was selected due to presence of strong precipitation events on 21 August 2007 in Finland. The total number of assimilated observations throughout the experiment as well as the model domain and orography are shown in Fig. 1. The ETKF perturbations are relaxed towards the EuroTEPS perturbations on the lateral boundaries and in the stratosphere (above 100 hPA), where the rescaling performs poorly due to the lack of conventional observations.

In the validation of the ETKF rescaling scheme we will compare the ETKF rescaling perturbations with the Targeted EPS perturbations. All illustrations in this paper correspond to the multiplicative-additive inflation algorithm (eq. 19), if not stated differently.

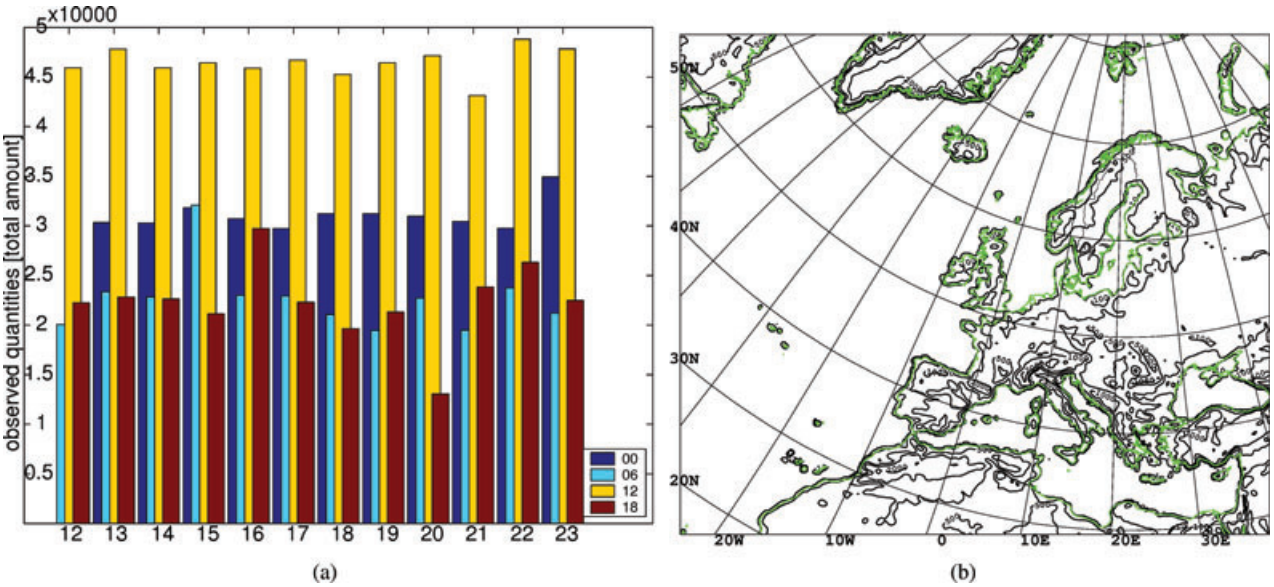

Fig. 1. Left plot: The variation of the total amount of assimilated observed quantities throughout the rescaling experiment 0600 UTC 12 August 2007–1800 UTC 23 August 2007. Right plot: The model domain and orography for the data assimilation experiments

4. Basic properties of the ETKF rescaling algorithm

We will illustrate some basic properties of the ETKF rescaling algorithm by showing an example of the +6 h forecast error variance, $Z_i^f(Z_i^f)^T$, and the analysis error variance, $Z_i^a(Z_i^a)^T$, reconstructed from the ensemble of the ETKF perturbations ($N = 8$), for the u- and v-wind model state components at model level 10 (≈210 hPa), valid at the assimilation time $t_i = 1200$ UTC 18 August 2007, an arbitrary chosen time. As a background we show the 300 hPa geopotential and wind fields of the +6 h background forecast, $x_i^f = \mathcal{M}_{i-1,i}x_{i-1}^a$, in Fig. 2. Figures 3 and 4 illustrate the performance of the ensemble transformation. The uppermost left plots in both figures show the variance of the u-wind (green isolines) and v-wind (blue isolines) components of the model state at the model level 10 (≈210 hPa), before application of the rescaling matrix (Fig. 3) and after application of the rescaling matrix (Fig. 4). There is a clear correspondence between strong spatial gradients for the model state component of the background forecast (jet streams or fronts) and a large variance of the corresponding forecast error component, estimated from the perturbations. Remaining plots in Figs 3 and 4 show squared projections of these fields on the non-zeroing eigenvectors c^j, $j = 1, \ldots, 7$, of the B_{ens}^f (eq. 8).

The perturbations in the projection on the eigenvector c^j are downscaled from $\sqrt{\lambda_j}$ to $\Pi\frac{\sqrt{\lambda_j}}{\sqrt{1+\lambda_j}}$ (see eq. 15), thus the downscaling is strongest in the leading direction and more moderate in non-leading directions. Notice that the leading eigenvector of the B_{ens}^f does not explain the major part of the variability in model space. A large part of the variability is projected on non-leading eigenvectors c^3, c^4 and especially c^5, which reflect the structures not so well observed by the observation network. The analysis error variance will be determined by both the dynamically unstable and the poorly observable structures.

The EnSRKF works in the tangent linear Gaussian environment and is an approximation to the EKF. However, there is an essential difference between these two methods. The covariances B_i^a and B_i^f are full-rank matrices in the EKF approach. Opposite to that, in the EnSRKF the rank of $Z_i^f(Z_i^f)^T$ and $Z_i^a(Z_i^a)^T$ is limited by the ensemble size. In case of large dimensional systems the covariances, reconstructed from ensembles, can be severely rank deficient.

5. Application of the multiplicative and the additive inflation to stabilize the performance of the ETKF rescaling scheme

In the examples shown in Figs 3 and 4, the set of non-zero eigenvalues of B_{ens}^f (eq. 9), valid at $t_i = 1200$ UTC 18 August 2007, is (10041.2, 9071.9, 8379.8, 7422.9, 7025.9, 6521.7, 5797.2). The trace of the forecast error in observation space is $\text{trace}(B_{ens}^f) = 54260.6$ and the amount of assimilated observations is $p = 45505$. This means that the compression of the whole observation space into seven directions is quite problematic and that a huge amount of important directions, explaining the variability of the forecast error, are not represented in the ensemble.

We illustrate the effect of the compression from observation space to ensemble space in Fig. 5. Only the multiplicative inflation is applied for the construction of this figure. Notice the very flat spectrum of the analysis error perturbations in comparison with the spectrum of the forecast error perturbations. As

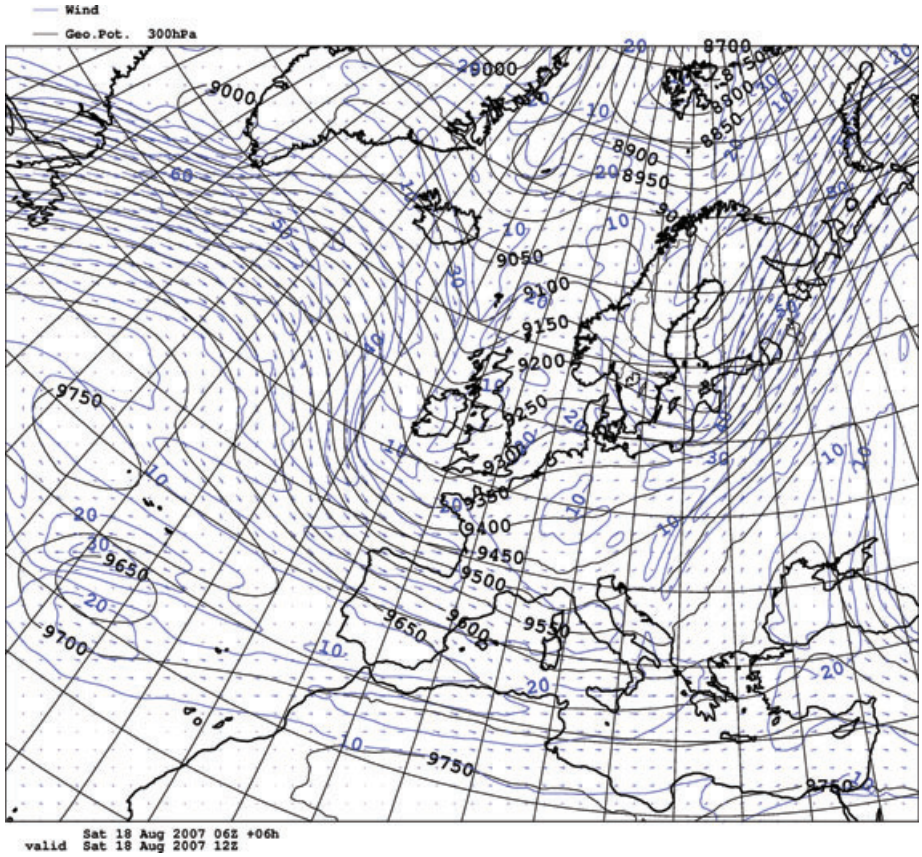

Fig. 2. 300 hPa geopotential height (black isolines), wind vectors (blue arrows) and wind speed (blue isolines) of the control +6 h forecast $x_i^f = \mathcal{M}_{i-1,i} x_{i-1}^a$ valid at t_i = 1200 UTC 18 August 2007.

explained above, this is a direct result of the rank deficiency. The N random centred perturbations, with structures of the full-rank static forecast error covariance matrix $B_{\mathrm{3D-Var}}$, $\{Y_n,\ n = 1, \ldots, N\}$, such that $Y_n \sim \mathcal{N}(0, B_{\mathrm{3D-Var}})$, $\sum_{n=1}^{N} Y_n = 0$, are transformed to observation space and normalized with the observation error variance. The $B_{ens}^{f,\mathrm{3D-Var}}$, constructed from such an ensemble, has $N-1$ non-zero eigenvalues in the range between 17 000 and 26 000. These huge eigenvalues result from the compression of the full-rank system ($p = 45\,505$) to the $N-1 = 7$-dimensional ensemble space. Notice that the trace of $B_{ens}^{f,\mathrm{3D-Var}}(\mathrm{trace}(B_{ens}^{f,\mathrm{3D-Var}}) = 149015)$ is several times larger than the trace of the B_{ens}^{f}, because the former one contains homogeneous structures with possibilities for the perturbations to be placed anywhere within the model domain. For the EuroTEPS perturbations, that have non-homogeneous structures of the forecast error covariance, the trace of $B_{ens}^{f,\mathrm{TEPS}}$ is comparable with the trace of $B_{ens}^{f}(\mathrm{trace}(B_{ens}^{f,\mathrm{TEPS}}) = 16674 + 12258 + 9989 + 8308 + 1751 + 1357 + 1025 = 51\,362)$. The rescaling factor ΠT, with $\Pi = 57.24(\mathrm{trace}(B_{ens,i}^{a}) = 6.993 \times 57.24^2 = 22993$), is used in the example of Fig. 5.

Because we apply the ETKF rescaling scheme in the High Resolution Limited Area Model for Numerical Weather Prediction, the inflation factor should not change too abruptly from one assimilation time to another. In order to prohibit undesirable oscillations of the inflation factor Π_i, we adjust the ensemble spread to the averaged variance of innovations (eq. 14) only at 1200 UTC and we apply a time filter on the raw inflation factor,

$$\Pi_i^{flt} = \Pi_{i-1} + s(\Pi_i - \Pi_{i-1}) + s(\Pi_{i-2} - \Pi_{i-1}), \qquad (20)$$

where s is a filter parameter ($s = 0.3$ in our application). Observation networks at 0000 UTC and 1200 UTC are not completely compatible with the observation networks at 0600 UTC and 1800 UTC. They are not only different by the amount of the observed quantities (Fig. 1) but they are different by observation types as well. Therefore the inflation factor is calculated based on the most comprehensive observation network (1200 UTC) and empirically rescaled to be consistent with observation networks at 1800 UTC, 0000 UTC and 0600 UTC.

One should recognize that the ETKF rescaling scheme is efficient in finding the most unstable directions for the forecast error growth but it fails in determining the amplitude of the perturbations. At the same time a physically meaningful amplitude of the forecast ensemble perturbations is crucial for the efficient performance for the ETKF rescaling scheme. We have noticed that the optimally adjusted inflation factor shows a statistically significant positive correlation with the amount of observed quantities

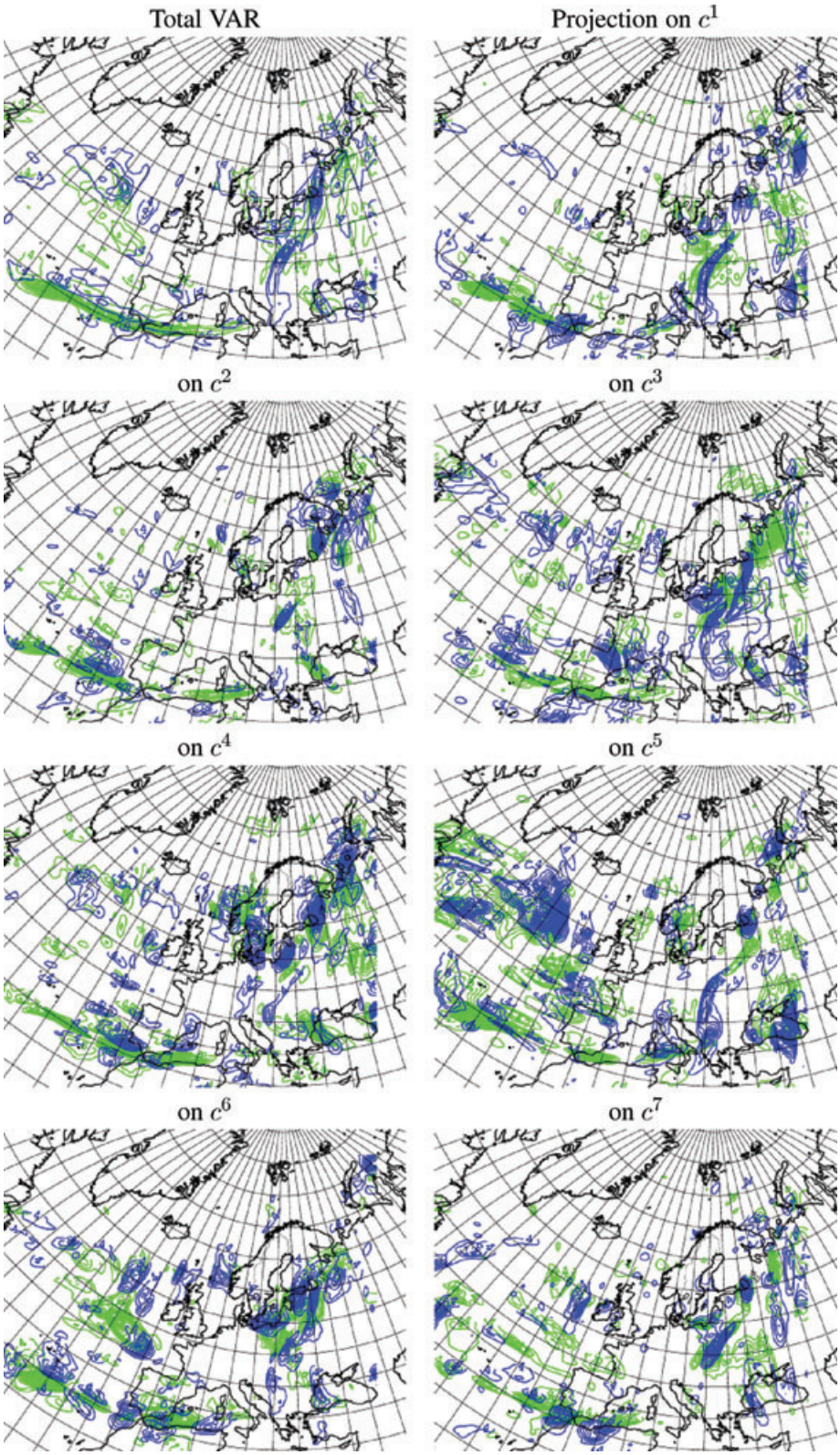

Fig. 3. The ensemble estimate of the variance for *u*- ('green' isolines) and *v*- ('blue' isolines) wind components at model level 10 and the squared projection of the perturbations on the eigenvectors c^j, $1 \leq j \leq N-1 (N = 8)$ before the rescaling is applied, valid at $t_i = 1200$ UTC 18 August 2007

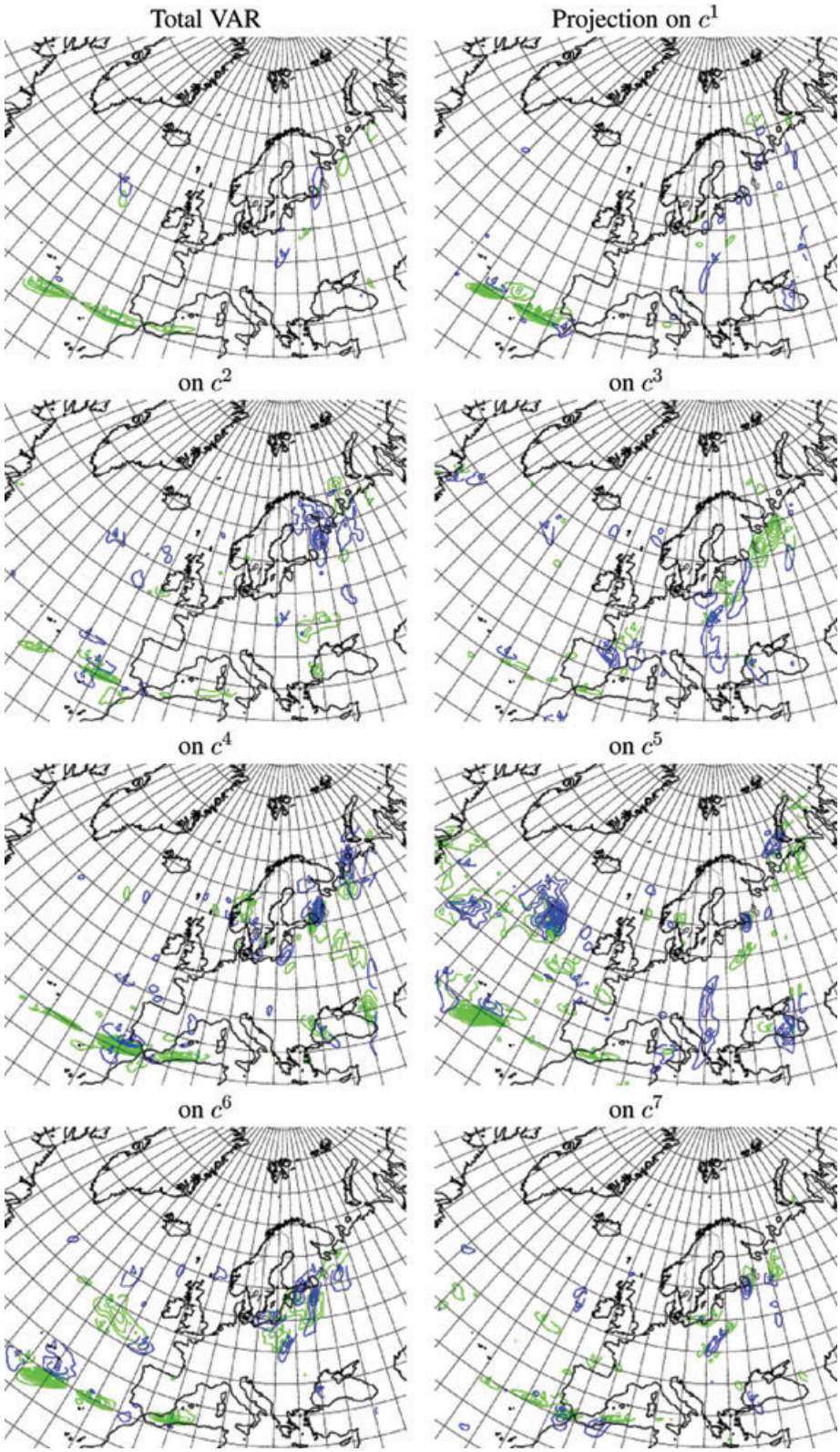

Fig. 4. The ensemble estimate of the variance for *u*- ('green' isolines) and *v*- ('blue' isolines) wind components at model level 10 and the squared projection of the perturbations on the eigenvectors c^j, $1 \leq j \leq N-1 (N=8)$ after the rescaling is applied, valid at $t_i = 1200$ UTC 18 August 2007.

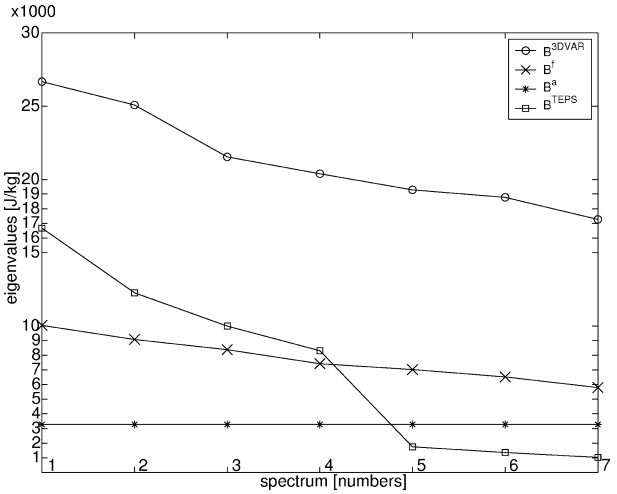

Fig. 5. The spectrum of the forecast and analysis error covariance matrices B^f_{ens} (crosses) and B^a_{ens}(stars) ($t_i = 1200$ UTC 18 August 2007), calculated from the ETKF perturbation. The spectrum of the $B^{f,\mathrm{TEPS}}_{ens}$ (squares), calculated from the EuroTEPS perturbations and the spectrum calculated from the random perturbations with structures of $B_{3\mathrm{D-Var}}$ (circles) are given for comparison.

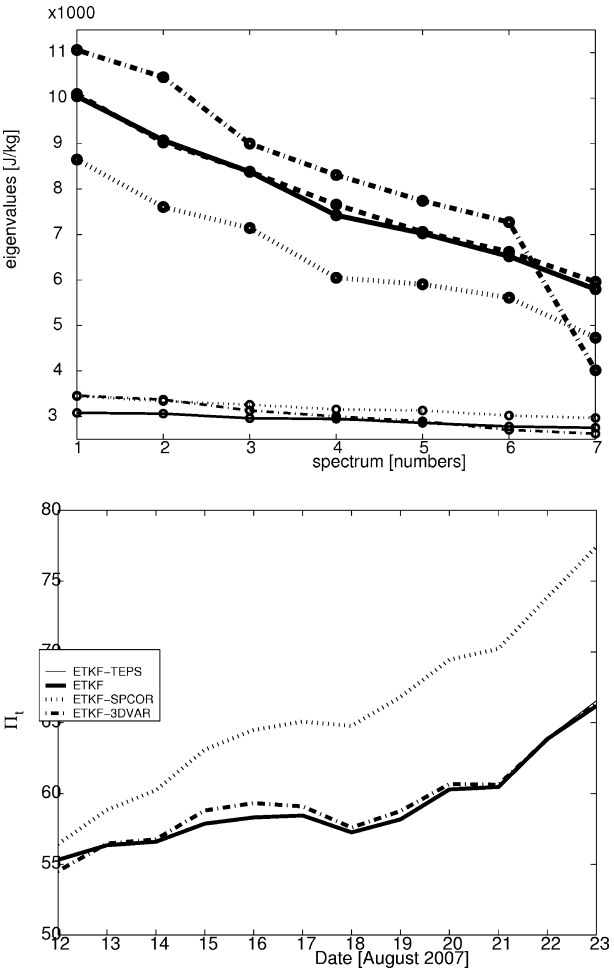

Fig. 6. Upper graph: The spectrum of $B^f_{ens,i}$ (before rescaling, thick curves) and $B^a_{ens,i}$ (after rescaling, thin curves), valid at $t_i = 1200$ UTC 18 August 2007, for different inflation schemes: the multiplicative inflation only (solid curve), the corrected spectrum ETKF-CORR (dotted curve), the merged ETKF-3DVAR (dash-dotted) and the merged ETKF-TEPS (dashed). Lower graph: The time evolution of the inflation factor for different inflation schemes: the multiplicative inflation only (solid curve), the spectrum corrected ETKF-CORR (dotted), the merged ETKF-3D-Var (dash-dotted) and the merged ETKF-TEPS (dashed).

at the current assimilation cycle and a statistically significant negative correlation with the amount of observed quantities at the next assimilation cycle. Indeed, the larger the amount of the assimilated observations is, the larger is the trace of the forecast error covariance in observation space. Therefore, a larger amount of assimilated observations leads to a stronger downscaling and a stronger inflation is required to compensate for the downscaling (eq. 15). The variance inflation is applied in order to obtain a physically meaningful spread of the forecast perturbations at the next assimilation cycle. With a larger amount of observed quantities at the next assimilation cycle, the compression of the observation space into the ensemble space partially compensates the need for inflation.

Figure 6 illustrates the performance of the different inflation schemes, described in Section 2: the multiplicative inflation only (ETKF, eq. 11, solid curve), the spectrum corrected multiplicative inflation (ETKF-CORR, eq. 18, dotted curve), the multiplicative and additive inflation by merging with structures of $B_{3\mathrm{D-Var}}$ (ETKF-3DVAR, eq. 19, dash-dotted curve) and the merging of the ETKF and EuroTEPS perturbations (ETKF-TEPS, eq. 21, dashed curve). In the ETKF-TEPS configuration of the experiment the analysed perturbation is a weighted average of the ETKF and EuroTEPS perturbations in the inner area of the domain.

$$X^{a,\mathrm{TEPS}} = \beta X^a + (1. - \beta)W. \tag{21}$$

In the example shown in Fig. 6, $\beta = 0.8$. The spectrum of the B^f_{ens}, calculated from the forecast perturbations (thick curves), and the spectrum of B^a_{ens}, calculated from the analysis perturbations (thin curves) valid at $t_i = 1200$ UTC 18 August 2007, are shown on the upper graph in Fig. 6. The time evolution of the inflation factor Π, corresponding to these four inflation schemes, during the extent of the experiment, is shown on the lower graph in Fig. 6. One can see that the analysis ensemble, corresponding to the ETKF-3D-Var experiment, has a larger spread in the directions of the leading eigenvectors of B^a_{ens} and a smaller spread in a poorly observable directions. As a result, the forecast perturbations have a larger projection on the leading uncertainty eigenmodes. Hopefully, this may lead to a more efficient use of the observation network to reduce uncertainty about the model

estimate. The merging of the ETKF and the low-rank EuroTEPS structures has a small impact on the spectrum of B^f_{ens} and B^a_{ens}.

The performance of the ETKF-CORR experiment is less satisfactory. In this experiments the value of parameter κ (eq. 18, kappa = 1048.8) seems to be too small to make the desirable correction of the spectrum of B^a_{ens}, which remains flat. At the same time the value of the parameter κ seems to be too strongly penalizing the energy in the non-leading directions of B^f_{ens}. The trace of B^f_{ens} is reduced which results in a continuous growth of the inflation factor Π. Probably the corrected spectrum algorithm is inefficient due the extremely small ensemble size. Still we would like to remark that the merging of the ETKF and the full-rank 3D-Var structures corrects the spectrum of B^a_{ens} with almost no increase of the inflation factor.

Eight ensemble members is clearly a too small size ensemble in attempting to model the flow-dependent structures for mesoscale numerical weather prediction. One needs the computational power to describe and to resolve important processes for the mesoscale forecasting at the same time as the forecasting problem is probabilistic in nature. For the coming few years and for operational forecasting, we will therefore have to live with small ensemble sizes. This is why we have selected an 'extremely' small ensemble size to investigate the problems in this 'realistic operational' scenario. Besides that, we are aware that 2 weeks is a too short period to make definite conclusions about the performance of the ETKF rescaling algorithm. Two weeks is of the order of the time scale for weather regimes. The performance can be strongly influenced by a particular situation, as is the case in our example, since during the second part of the time period a severe storm situation developed over Northern Europe. In fact, the deterministic data assimilation system with the static forecast error covariance was not able to utilize observations efficiently during the second part of the time period and the background state diverged from the observations. This lead to an increase of the inflation factor for the corresponding period (eq. 14 and Fig. 6).

6. Validation of the performance of the ETKF rescaling scheme

We will first discuss one example of the ensemble based +6 h forecast variance fields, representing the ensemble spread at +6 h, valid at 21 August 2007 1200 UTC. The corresponding control +6 h forecast from 21 August 0600 UTC is illustrated for 300 hPa geopotential and wind in Fig. 7 (left-hand side) and for 850 hPa temperature in Fig. 7 (right-hand side). In the upper levels we may notice a strong meridional circulation with a meandering jet-stream over the North Atlantic. Associated with this is a high pressure ridge along 20W towards Iceland and a low pressure system (trough) centred over France. Also note a small-scale cut-off low at 35N 30W. In the 850 hPa temperature we may notice strong gradients (fronts) associated with the cut-off low over the Atlantic and also in the area of the upper-air trough over Western Europe.

The ensemble based +6 h forecast variance fields, associated with the +6 h forecast model state illustrated in Fig. 7, are presented in Fig. 8. The variance fields for u-wind (green isolines) and v-wind (blue isolines) components at model level 10 (approximately 210 hPa) are shown in Fig. 8 (left-hand side) and temperature at model level 30 (approximately 850 hPa) in Fig. 8 (right-hand side).

One can notice the strong inhomogeneity of these variance fields contrasting radically the assumptions on homogeneity in variational assimilation. Furthermore, areas of large variance

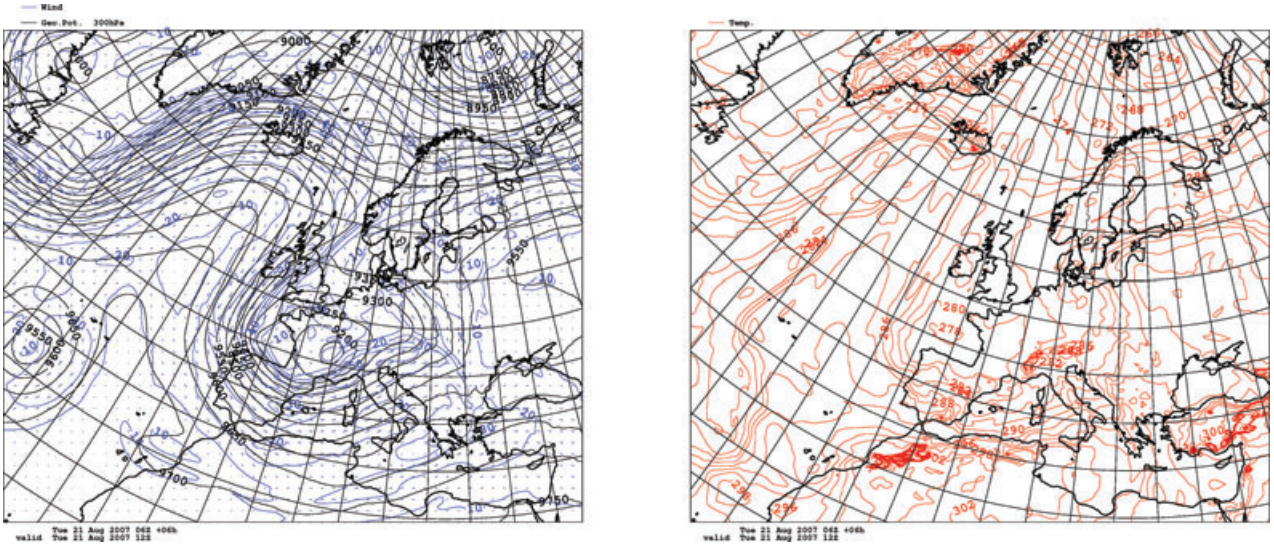

Fig. 7. 300 hPa wind and geopotential (left-hand side) and 850 hPa temperature (right-hand side) taken from the background model state at 21 August 2007 06UTC +6 h.

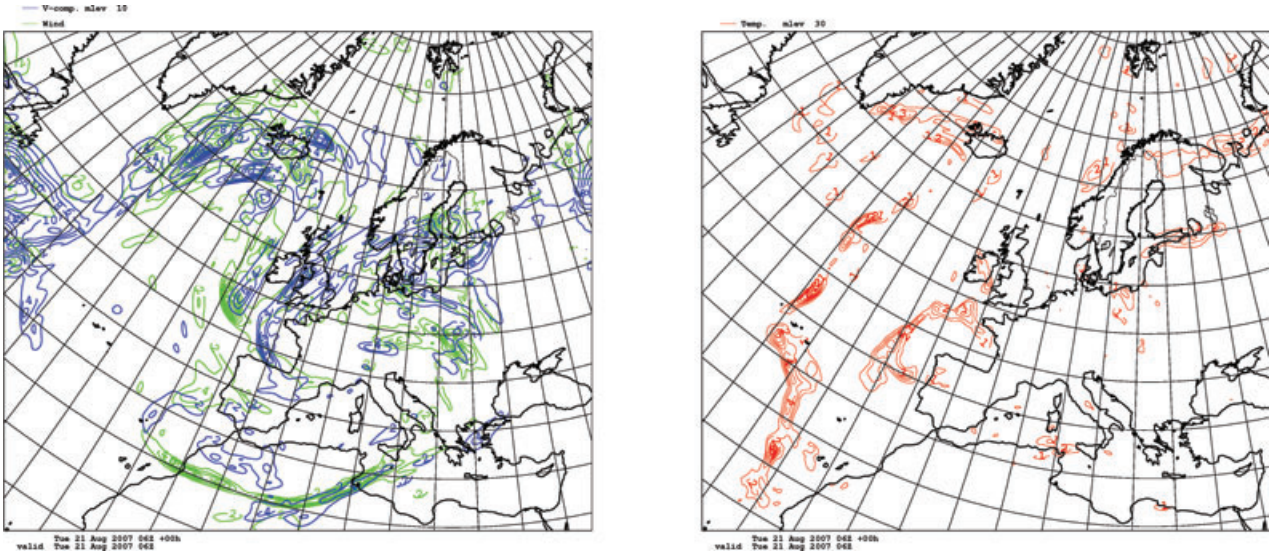

Fig. 8. Examples of estimated background error variances based on the ensemble of +6 h forecast valid at 1200 UTC 21 August 2007. Wind components at model level 10 (left-hand side) and temperature at model level 30 (right-hand side).

are clearly related to features in the control forecast model state, providing support to the core idea of this study, to introduce flow-dependent background error statistics into the data assimilation process. To mention a few details, large uncertainties seem to be associated with the position and the strength of the meandering jet-stream over the North Atlantic, while the corresponding variances are estimated to be smaller when approaching the more data dense areas over the European continent. The association of large variances with the jet-stream is easily understandable. The horizontal wind shear close to the jet stream is large and even small perturbations in the forecast horizontal position of the jet-stream will give rise to large absolute differences, measured in metres per second. Furthermore, the jet stream is often associated with a large vertical wind shear and dynamical instability. To handle this dynamical source of uncertainty in upper-air data assimilation, one does not necessarily need to introduce the utilization of ensembles, it was shown by Lindskog et al. (2006) that a diagnostic interpretation of instability can be applied successfully.

It is interesting to notice that large values of the forecast variance for low-level temperature (Fig. 8, right-hand side) occur in horizontally elongated (anisotropic) structures. Comparing with the +6 h control forecast model state for 850 hPa temperature in Fig. 7 (right-hand side), it is clear that these large forecast variances are associated with the occurrence of atmospheric fronts. One may in particular notice the frontal systems in the vicinity of the upper-air cut-off low at 35N 30W.

It is evident that the ETKF rescaling perturbations contain physically meaningful structures, which are influenced by dynamical as well as observation network effects. We now present some further diagnostic quantities, which qualitatively summarize the behaviour of the ETKF rescaling perturbations.

The ETKF ensemble has proven to be able to span the unstable directions of the forecast error growth (Wang and Bishop, 2003). The direction of the fastest linear error growth over the forecast length h in the subspace spanned by the ensemble is defined via a linear combination b_h of the analysis perturbations, $X_i^a b_h$, where X_i^a is the ensemble of analysis perturbations valid at time t_i and b_h is a vector whose elements are N coefficients of the linear transformation, that maximizes

$$\frac{b_h^T \left(X_{i+h}^f\right)^T S X_{i+h}^f b_h}{b_h^T \left(X_i^a\right)^T S X_i^a b_h}, \tag{22}$$

where S is an appropriate norm, the total dry energy norm in this case, and X_{i+h}^f is the ensemble of forecast perturbations valid at time $t_i + h$.

A total dry energy norm $\langle x, Ex \rangle$ for a perturbation x (Talagrand, 1981; Barkmeijer et al., 1999; Errico, 2000) is defined as follows:

$$\langle x, Ex \rangle = \sum_l \left(u_l^T D_l u_l + v_l^T D_l v_l + \frac{c_p}{T_r} T_l^T D_l T_l \right) + R_d T_r P_{\text{ref}} \ln(P_s)^2, \tag{23}$$

where l is a vertical model level, D_l is the vertical integration weight, $D_l = P_{l+1} - P_l$. u_l, v_l, T_l and P_l are wind components, temperature and pressure at the model level l, P_s is the surface pressure of perturbation x, $c_p = 1004$ J K^{-1} kg^{-1} is the specific heat for dry air at constant pressure, $T_r = 273$K is the reference temperature, $R_d = 287$ J mol^{-1} K^{-1} is the gas constant for dry air and $P_{\text{ref}} = 800$ hPa is a reference pressure.

In fact, the linear combination b_h (eq. 22) is a leading eigenvector of $A^{-1/2} B A^{-1/2}$, where $A = (X_i^a)^T X_i^a$ and $B = (X_{i+h}^f)^T X_{i+h}^f$ (Wang and Bishop, 2003). The corresponding

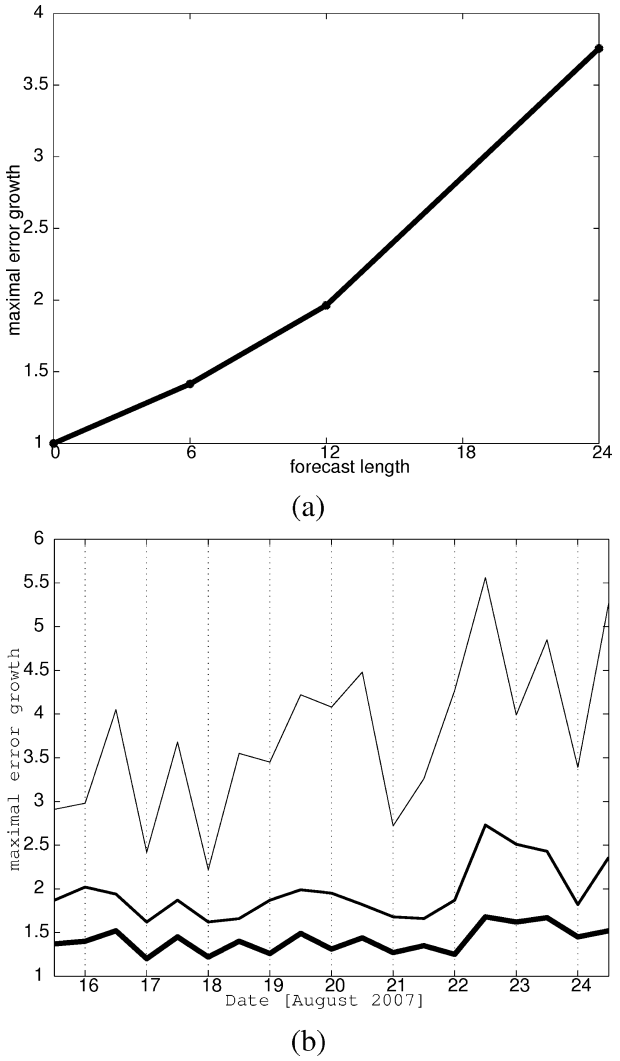

Fig. 9. (a) The time averaged fastest energy growth as a function of the forecast length. (b) The time evolution of the fastest energy growth per 6 h (thick curve), 12 h (normal curve) and 24 h (thin curve).

eigenvalues for $h = \{06, 12, 24\}$ h forecast length are shown in Fig. 9(b) for the period 1200 UTC 15 August 2007–1200 UTC 24 August 2007. The time average of the fastest error growth as a function of forecast length is shown in Fig. 9(a).

The total dry energy of the perturbations increases on average by factor $\approx$2 per 12 h and by $\approx$3.75 per 24 h. The time evolution of the fastest error growth spanned by the ETKF ensemble, shown in Fig. 9(b), agrees with the mean squared distance between the background state and the observations. The former increases for all forecast lengths towards the end of the period, from the 21 August 2007 0000 UTC and later on, when a severe storm situation developed over Northern Europe. The time evolution of the fastest error growth is consistent with the time evolution of the inflation factor Π (Fig. 6, lower panel), the calculation of which is based on the mean squared distance between the background state and observations.

The ETKF rescaling scheme, being a linear one, preserves balances and structures, created through the dynamical forward propagation. The horizontal spectral density of the analysis/forecast error variance for the vorticity component of the flow, averaged over space and time, is shown in Fig. 10. Plot (a) shows the spectral density estimated from the ETKF ensemble of perturbations. Plot (b) shows the spectral density estimated from the EuroTEPS ensemble, for comparison. The spectral density of both the analysis and the forecast error variances, estimated from the ETKF ensemble (Fig. 10a), has a nice linear slope at higher wavenumbers. The analysis error variance of vorticity has slightly more energy on larger scales compared to the forecast error variance, but agrees well in the total shape. Numerical noise is observed on the scale of numerical truncation only, and it increases slightly with forecast length.

The horizontal spectral density of the analysis error variance, estimated from the EuroTEPS perturbations (Fig. 10b), has a totally different structure compared to the horizontal spectral density of the forecast error variance. In fact, the resolution of the ECMWF global model used to calculate the EuroTEPS initial perturbations is T_L159, which corresponds to wavenumbers 20–30 in the present HIRLAM domain. The kink in the curve of the horizontal spectral density corresponding to the EuroTEPS analysis is seen around these wavenumbers. It may be that a too slow decrease of the horizontal spectral density for large horizontal wavenumbers reflects misrepresentation of the global model regular lat–lon grid by the HIRLAM rotated lat–lon grid.

The presence of strong numerical noise in the stratosphere can be seen by investigating structures of vertical correlations between the model state components. These correlations are investigated by the aid of the same algorithms that are used to construct data assimilation background error constraints (for details see Berre, 2000). The vorticity perturbation is treated as the basic model state variable, and from the vorticity perturbation the balanced part of the divergence perturbation is derived by regression in the vertical direction. Then, in a second step, balanced temperature and balanced surface pressure increments are estimated using the vorticity increments and the unbalanced divergence increments as predictors. From Fig. 11 we may conclude that the unbalanced divergence perturbations (plot a) have a vertical length-scale of the order of five vertical model layers, with exception for the planetary boundary layer due to boundary layer turbulence and due to the increased vertical resolution of the model. The vertical correlations for unbalanced temperature (plot b) exhibit a particular structure with strong anticorrelations in the vicinity of the tropopause (warm upper troposphere, high-level tropopause and a cold lower stratosphere, for example). The vertical cross-correlations between un-balanced temperature and un-balanced surface pressure (plot c) mainly illustrate the surface pressure response to an un-balanced thermal forcing. A presence of the negative correlation between un-balanced

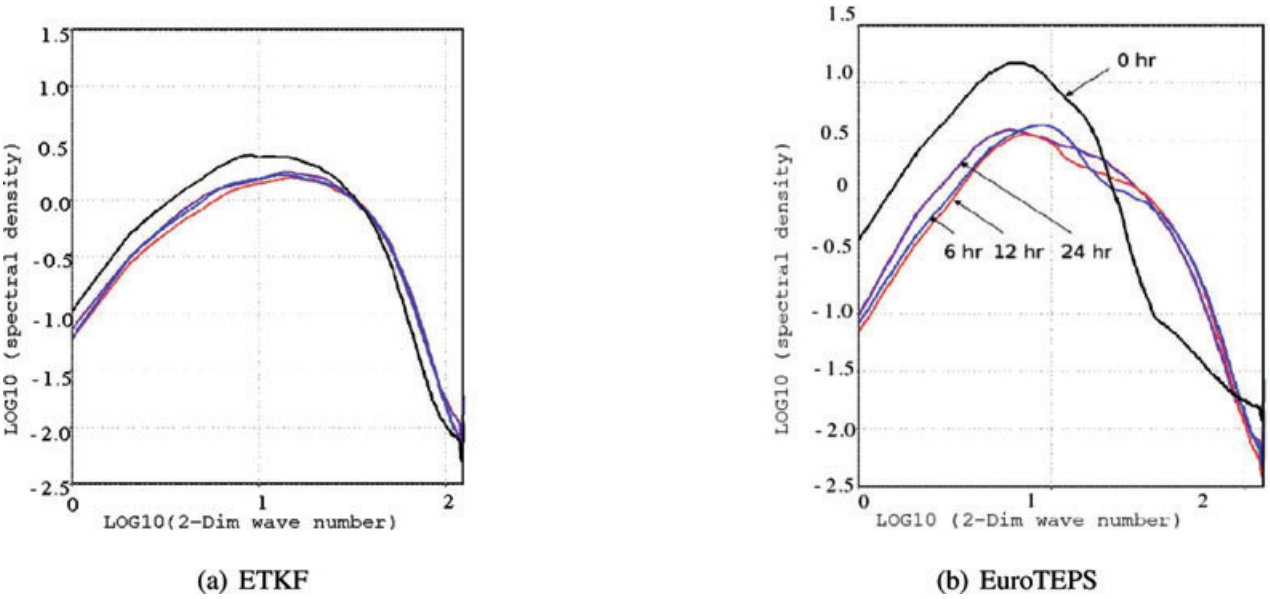

Fig. 10. The power spectrum of the forecast error variance of vorticity at 400 hPa for different forecast lead times, estimated from the ETKF perturbations (a) and from the EuroTEPS perturbations (b). The curves corresponding to different forecast lead times are colour coded as follows: +00 h (black), +06 h (blue), +12 h (red) and +24 h (purple).

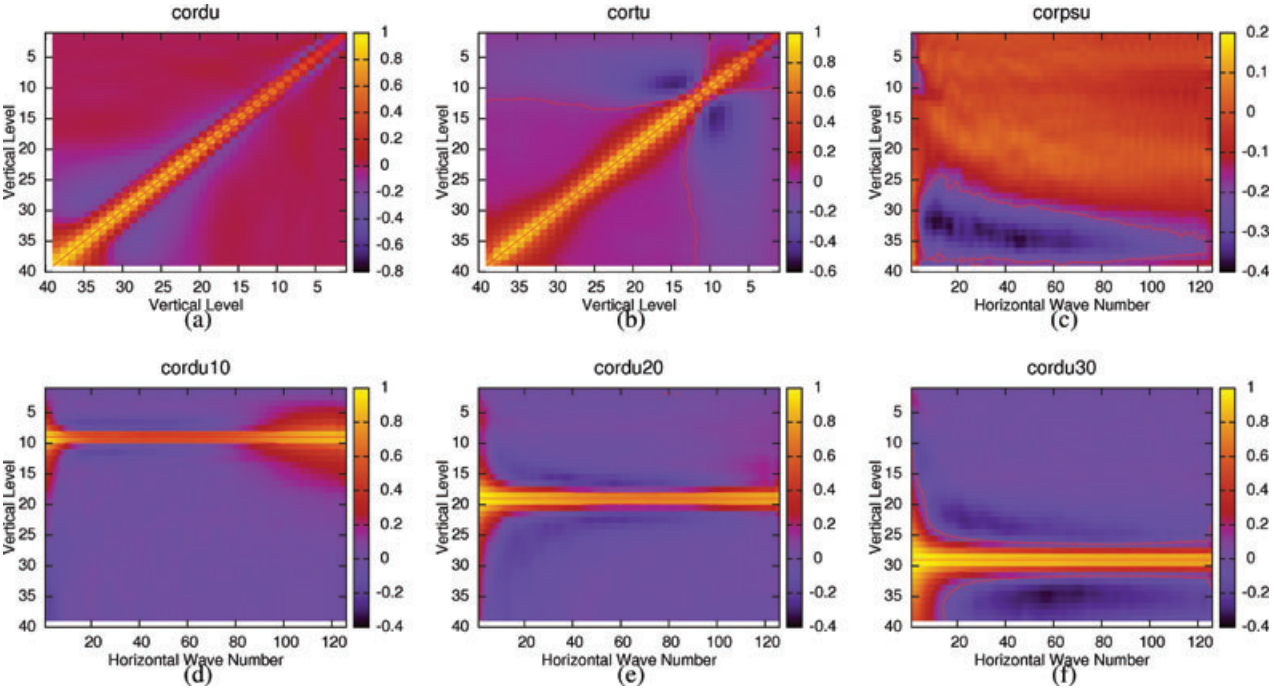

Fig. 11. Upper row: Temporal and horizontal average of vertical autocorrelations for unbalanced divergence (left panel) and for unbalanced temperature (middle panel), temporal average of cross-correlations between unbalanced temperature and unbalanced surface pressure as a function of horizontal wavenumber (right panel). Lower row: Temporal average of vertical autocorrelations for vorticity as a function of horizontal wavenumber with level 10 as the reference (left panel), with level 20 as the reference (middle panel) and with level 30 as the reference (right panel). All correlations are estimated from the time series of ETKF 12 h forecast ensemble perturbations over the period of the experiment.

surface pressure and unbalanced stratospheric temperatures for the largest horizontal scales probably indicates a model deficiency. With exception for this feature, the presented statistics indicate that the ETKF perturbations are physically reasonable and similar to corresponding statistics derived from differences between forecasts of different lengths valid at the same time. More alarming are the strong vertical correlations over deep vertical layers for horizontally small-scale vorticity perturbations at stratospheric levels (plot d), possibly indicating noise originating from vertically propagating gravity waves. The vertical correlations for vorticity perturbations at lower tropospheric levels exhibit a quite natural behaviour, with larger vertical scales associated with larger horizontal scales.

7. Summary

The ETKF rescaling scheme has been implemented into the HIRLAM variational data assimilation scheme in order to estimate the uncertainty about the model state estimate. The sensitivity to the observation network is built into the ETKF rescaling scheme. Besides that, the ETKF rescaling scheme iteratively estimates the leading directions of the forecast error variability. By sequentially applying the ETKF rescaling scheme, an ensemble of perturbations with flow-dependent structures, which reflect the density of the observation network, can be created.

The forward numerical integration of each ensemble member is the main computational burden of the ETKF rescaling scheme, and this is the reason why we have selected an extremely small ensemble size for the investigations here. The forecast and analysis error covariances, reconstructed from the ETKF ensemble, are severely rank deficient with an affordable ensemble size. In the rank-deficient implementation of the ETKF rescaling scheme, a variance inflation must be applied, as an ad hoc procedure to account for the impact from non-sampled directions of the forecast error variability and to account for the compression from the r-dimensional observation space to the $N-1$-dimensional ensemble space, in which the downscaling is indeed performed. Variance inflation by a scalar factor leads to an unreasonably flat spectrum of the analysis error covariance. To improve the spectrum of the analysis error covariance an additive inflation is employed. The original ETKF perturbations are merged with random structures of the static forecast error covariance matrix used in the variational assimilation scheme of HIRLAM. The resulting perturbations receive 'fresh blood' injections each assimilation time and this also prohibits filter collapse.

A preliminary validation confirms that the sequential application of the ETKF rescaling scheme produces an ensemble of perturbations, which contains dynamically feasible structures. Horizontal spectral densities of the perturbations have been estimated and compared with the spectral densities of ensemble perturbations based on targeted singular vectors. It has been shown that the spectra of ETKF perturbations are the more realistic ones, in particular at analysis time and for short forecast lengths. The amplitude of the ETKF perturbation clearly reflects both the density of the observation network and the areas of dynamical instabilities.

The proposed filter spin-up algorithm, the additive inflation using structures of the full-rank static forecast error covariance matrix, has the effect of making the spectra of the analysis perturbations more dominated by the leading eigenmodes and the analysis perturbations will have less energy in poorly observable structures. As a result of that, the forecast uncertainty will have larger projection on the leading uncertainty eigenmodes. Assuming that Gaussianity is valid, this also means that the observation network will be used more efficiently to reduce the uncertainty of the model estimate. Furthermore, since the proposed spin-up algorithm inserts random structures on spatial scales of dynamical instabilities, it may help to recover important uncertainties not included in a small ensemble originally.

Remaining deficiencies of the ETKF scheme, as presently identified, are the too simplistic variance inflation algorithm, the less satisfactory performance close to the lateral boundaries and the presence of noise in the stratosphere. The lateral boundary problem and the stratospheric noise problem are known deficiencies of the HIRLAM forecast model, but it seems that the ETKF rescaling scheme acts as an amplifier of these problems.

Ongoing developments of the ETKF rescaling scheme includes introduction of satellite observations, with more stratospheric observations and with a more even distribution of observations in time, that should contribute to a stabilization of the scheme. The first results from a hybrid variational ensemble data assimilation, based on HIRLAM 3D-Var and on an augmentation of the assimilation control varaiable as proposed by Lorenc (2003), have also been obtained and will be the subject of future publication.

8. Acknowledgments

This research is partially funded by the eVITA programme of the Research Council of Norway, under contract 178894. The constructive suggestions and comments by four anonymous reviewers, which have significantly improved the structure and readability of our manuscript, are very much appreciated.

References

Anderson, J. L. 2001. An ensemble adjustment filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903.

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**, 2741–2758.

Barkmeijer, J., Buizza, R. and Palmer, T. N. 1999. 3D-Var Hessian Singular Vectors and their potential use in ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.* **125**, 2333–2351.

Bengtsson, T., Snyder, C. and Nychka, D. 2003. Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.* **108**(D24), 8775, doi:10.1029/2002JD002900.

Berre, L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Mon. Weather Rev.* **128**, 644–667.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter, Part I: theoretical aspects. *Mon. Weather Rev.* **129**, 420–436.

Buizza, R. 1994. Localisation of optimal perturbations using a projection operator. *Q. J. R. Meteorol. Soc.* **120**, 1647–1681.

Cohn, S. E., da Silva, A. M., Guo, J., Sienkiewich, M. and Lamich, D. 1998. Assessing the effect of data selection with the DAO Physical-space Statistical Analysis System (PSAS). *Mon. Weather Rev.* **126**, 2913–2926.

Courtier, P., Thepaut, J. N. and Hollngswirth, A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

Durbin, J. and Koopman, S. J. 2001. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series, Oxford University Press, Oxford.

Errico, R. M. 2000. Interpretation of the total energy and rotational energy norms to determination of singular vectors. *Q. J. R. Meteor. Soc.* **126**, 1581–1600.

Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10 143–10 162.

Frogner, I.-L. and Iversen, T. 2002. High-resolution limited area ensemble predictions based on low-resolution targeted singular vectors. *Q. J. R. Meteorol. Soc.* **128**, 1321–1341.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M. and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model, Part I: general formulation and the background error constraint. *Tellus* **53A**, 425–446.

Hamill, T. M. and Snyder, C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Weather Rev.* **128**, 2905–2919.

Heemink, A. W., Verlaan, M. and Segers, A. J. 2001. Variance reduced ensemble Kalman filtering. *Mon. Weather Rev.* **129**, 1718–1728.

Houterkamer, P. L. and Mitchell, H. L. 1998. Data Assimilation using ensemble Kalman filter technique. *Mon. Weather Rev.* **126**, 796–811.

Julier, S. J. and Uhlmann, J. K. 1997. A new extension of the Kalman filter to non-linear systems. *Int. Symp. Aerospace/Defense Sensing, Simulations and Controls* **3**, doi:10.1117/12.280797.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45.

Kalman, R. E. and Bucy, R. S. 1961. New results in linear filtering and prediction. *J. Basic Eng.* **83D**, 95–108.

Keppert, J. D. 2004. On ensemble representation of the observation error covariance in the ensemble Kalman Filter. *Ocean Dyn.* **54**, 561–569.

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithm for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**, 97–110.

Lindskog, M., Gustafsson, N. and Mogensen, K. S. 2006. Representation of background error standard deviations in a limited area model data assimilation system. *Tellus* **58A**, 430–444.

Lorenc, A. 2003. The potential of the ensemble Kalman filter for NWP: a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**, 3183–3203.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Parrish, D. F. and Derber, J. C. 1992. The National Meteorological Center's statistical-interpolation analysis system. *Mon. Weather Rev.* **120**, 1747–1763.

Pham, D. T., Verron, J. and Roubaud, M. C. 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Mar. Syst.* **16**, 323–340.

Sakov, P. and Oke, P. R. 2008. Implication of the form of the ensemble transformation in the ensemble square root filters. *Mon. Weather Rev.* **136**, 1042–1053.

Talagrand, O. 1981. A study of the dynamics of four-dimensional data assimilation. *Tellus* **33**, 43–60.

Tippet, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Weather Rev.* **131**, 1485–1491.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the Breeding Method. *Mon. Weather Rev.* **125**, 3297–3319.

Undén, P., Rontu, L., Järvinen, H., Lunch, P., Calvo, J. and co-authors. 2002. HIRLAM-5 Scientific Documentation. http://hirlam.org/ (Available from SMHI, SE-61767 Norrköping, Sweden). Accessed 7 Feb 2011.

Verlaan, M. and Heemink, A. W. 1997. Tidal flow forecasting using reduced-rank square root filters. *Stoch. Hydrol. Hydraulics* **11**, 349–368.

Wang, X. and Bishop, C. H. 2003. A comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecasting Schemes. *J. Atmos. Sci.* **60**, 1140–1158.

Wang, X., Bishop, C. H. and Julier, S. J. 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Weather Rev.* **132**, 1590–1605.

Wang, X., Hamill, T. M., Whikaker, J. S. and Bishop, C. H. 2007a. Comparison of hybrid Ensemble Transform Kalman Filter: Optimum Interpolation and Ensemble Square Root Analysis Schemes. *Mon. Weather Rev.* **135**, 1055–1076.

Wang, X., Snyder, C. and Hamill, T. M. 2007b. On the theoretical equivalence of differently proposed ensemble-3DVar hybrid analysis schemes. *Mon. Weather Rev.* **135**, 222–227.

Whitaker, J. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**, 1913–1924.