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

Assimilating host model information into a limited area model

By PER DAHLGREN* AND NILS GUSTAFSSON, *Swedish Meteorological and Hydrological Institute, Folkborgsvägen 1, 601 76, Norrköping, Sweden*

(Manuscript received 31 May 2011; in final form 7 December 2011)

ABSTRACT

We propose to add an extra source of information to the data-assimilation of the regional HIgh Resolution Limited Area Model (HIRLAM) model, constraining larger scales to the host model providing the lateral boundary conditions. An extra term, J_k, measuring the distance to the large-scale vorticity of the host model, is added to the cost-function of the variational data-assimilation. Vorticity is chosen because it is a good representative of the large-scale flow and because vorticity is a basic control variable of the HIRLAM variational data-assimilation. Furthermore, by choosing only vorticity, the remaining model variables, divergence, temperature, surface pressure and specific humidity will be allowed to adapt to the modified vorticity field in accordance with the internal balance constraints of the regional model. The error characteristics of the J_k term are described by the horizontal spectral densities and the vertical eigenmodes (eigenvectors and eigenvalues) of the host model vorticity forecast error fields, expressed in the regional model geometry. The vorticity field, provided by the European Centre for Medium-range Weather Forecasts (ECMWF) operational model, was assimilated into the HIRLAM model during an experiment period of 33 d in winter with positive impact on forecast verification statistics for upper air variables and mean sea level pressure.

Keywords: data-assimilation, large-scale constraint, error covariances, limited area, host model

1. Introduction

In order to run a regional Numerical Weather Prediction (NWP) model, an initial condition and a host model providing lateral boundaries are required. In this work we have used the HIRLAM model in which the initial condition is determined by combining observations and Fa short forecast with variational data assimilation, and the host model is usually the operational ECMWF model.

The HIRLAM data-assimilation system is based on the variational technique described in Gustafsson et al. (2001) and Lindskog et al. (2001). In this work it is run in four-dimensional variational analysis (4D-Var) mode, described in Huang et al. (2002) and Gustafsson (2006), and the background error covariances include a statistical balance, Berre (2000).

The variational procedure finds the optimal model state constrained by the models internal balances and observations, i.e. the host model is not taken into consideration. The aim of this paper is to examine a way to include selected information from the host model in the data-assimilation procedure.

One reason for doing this is to make the initial model state more consistent with the host model that will be applied at the lateral boundaries during time-integration. Another reason is that the global models often have more advanced assimilation techniques that use more satellite data, for example, and therefore give a better description of the large-scale flow. Using observations inside a regional domain only, larger scales cannot be described in a proper way, and will be aliased on smaller scales, so by blending the large-scale information into the regional model we can also hope to improve the regional model forecasts.

In Guidard and Fischer (2008), information from the global Action de Recherche Petite Echelle et Grand Echelle (ARPEGE) model analysis was blended into the regional Aire Limitée Adaption Dynamique développement Inter-National (ALADIN) model by adding the ARPEGE analysis as a constraint in the data-assimilation. All upper air variables and surface pressure from ARPEGE were applied with a diagonal error matrix, and the spectral information vector was truncated such that only the large

The review process was handled by Editor-in-Chief Harald Lejenäs
*Corresponding author.
email: Per.Dahlgren@smhi.se

Citation: Tellus A 2012, **64**, 15836, DOI: 10.3402/tellusa.v64i0.15836

scales were affected. In this work we take a different approach by selecting only the vorticity field from the host model, ECMWF, and we let that represent the large-scale flow which we assume is better described due to its extensive use of satellite observations and due its use of a global domain. Furthermore, we use a short forecast from ECMWF, instead of an analysis, to reduce the possibility of error correlations between the host model field and the observations used in the regional model data-assimilation. Also, it is not even possible for us in practice to use the ECMWF analysis in an operational context since it is not available in real time. Finally, the error covariances of the ECMWF vorticity field are not assumed to be represented by a diagonal matrix.

The paper outline is briefly as follows:

Details of the background error constraint, J_b, in HIRLAM and how vorticity is treated is described in Section 2. This information is needed for Section 3, where the formulation of the host model constraint is explained in detail.

Thereafter, in Section 4, we explain how the errors of the host model constraint are modelled and compared with the background error covariances. Section 5 points out some details on the implementation. The impact of the new constraint on an analysis is also investigated with a simple three-dimensional variational analysis (3D-Var) single case test.

The impact on forecasts, using a multi incremental 4D-Var setting, is then shown in Section 6. In Section 7 we summarise and make concluding remarks.

2. Balance constraints and control vector

The following notations will be used throughout the rest of this article:

- Variables in spectral space are denoted: $\hat{\boldsymbol{x}}$
- Subscript *ls* denotes data from the host model (ECMWF), example: $\boldsymbol{x}_{ls}$
- Subscript *b* denotes data from the regional model background (HIRLAM), example: $\boldsymbol{x}_b$
- An increment is the deviation from the regional model background:$\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_b$

The model state increment vector in the HIRLAM assimilation is

$$\delta\hat{\boldsymbol{x}} = \begin{pmatrix} \delta\hat{\boldsymbol{u}} \\ \delta\hat{\boldsymbol{v}} \\ \delta\hat{\boldsymbol{T}} \\ \delta\hat{\boldsymbol{q}} \\ \delta ln\hat{\boldsymbol{p}}_s \end{pmatrix} \tag{1}$$

and the general formulation of the background error constraint, J_b, is

$$J_b = \frac{1}{2}\delta\hat{\boldsymbol{x}}^T\boldsymbol{B}^{-1}\delta\hat{\boldsymbol{x}} \tag{2}$$

where B contains the error covariances of $\hat{\boldsymbol{x}}_b$. All variables in $\hat{\boldsymbol{x}}_b$ are related to each other and therefore the $\boldsymbol{B}$ matrix becomes quite complex. To ensure a fast convergence of the minimisation, a pre-conditioning through a variable transform is introduced. The change of variable is designed as a series of operators that successively removes the correlations between the components in $\delta\hat{\boldsymbol{x}}$ until we end up with a new variable, called $\hat{\chi}$, whose covariance matrix can be assumed to be the identity matrix.

The series of transforms can be written as

$$\hat{\chi} = \gamma_b^{-1/2}\boldsymbol{V}_b^T\lambda_b^{-1/2}\sigma_b^{-1}\boldsymbol{F}\boldsymbol{G}\delta\hat{\boldsymbol{x}} \tag{3}$$

$\boldsymbol{G}$ = calculate vorticity and divergence.
$\boldsymbol{F}$ = balance operators.
σ_b = background error standard deviation. Horizontal average, i.e. one value per vertical level.
λ_b = horizontal spectral density of the background error correlation. *Diagonal matrix*.
$\boldsymbol{V}_b$ = eigenvectors of the vertical background error correlation matrix. *Orthonormal matrix*, i.e. $\boldsymbol{V}_b^{-1} = \boldsymbol{V}_b^T$.
γ_b = eigenvalues of the vertical background error correlation matrix. *Diagonal matrix*.

The balance operators $\boldsymbol{F}$ follows Berre (2000) and the remaining operators are described in Gustafsson et al. (2001).

After the $\boldsymbol{G}$ operator is applied, the state vector contains ($\delta\hat{\zeta}$, $\delta\hat{\eta}$, $\delta\hat{T}$, $\delta\hat{q}$, $\delta ln\hat{p}_s$). Then the $\boldsymbol{F}$ operator removes the cross-correlations between the variables and the state vector is written ($\delta\hat{\zeta}$, $\delta\hat{\eta}_u$, $\delta\hat{T}_u$, $\delta\hat{q}_u$, $\delta ln\hat{p}_s$), where subscript u means that the variables are *unbalanced*. $\delta\hat{T}$ and $\delta ln\hat{p}_s$ are often combined into one variable, and the balance operators described as going from the unbalanced state to the model state. Following the notations in Berre (2000), and also omitting all δ and the *ln* for pressure to make the expression more compact, the balance operators acts as the following

$$\begin{pmatrix} \hat{\zeta} \\ \hat{\eta} \\ (\hat{\boldsymbol{T}}, \hat{\boldsymbol{p}}_s) \\ \hat{\boldsymbol{q}} \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{E} & 0 & 0 & 0 \\ \boldsymbol{MH} & \boldsymbol{E} & 0 & 0 \\ \boldsymbol{NH} & \boldsymbol{P} & \boldsymbol{E} & 0 \\ \boldsymbol{QH} & \boldsymbol{R} & \boldsymbol{S} & \boldsymbol{E} \end{pmatrix}}_{\boldsymbol{F}^{-1}} \begin{pmatrix} \hat{\zeta} \\ \hat{\eta}_u \\ (\hat{\boldsymbol{T}}, \hat{\boldsymbol{p}}_s)_u \\ \hat{\mathbf{q}}_u \end{pmatrix} \tag{4}$$

where $\boldsymbol{E}$ is the identity matrix, $\boldsymbol{M}$, $\boldsymbol{N}$, $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$ and $\boldsymbol{S}$ are vertical balance operators derived by linear regression and $\boldsymbol{H}$ is a similar horizontal balance operator that relates a linearised geopotential with vorticity, including also a horizontally varying Coriolis parameter. It should also be

noted that since $\boldsymbol{F}^{-1}$ is triangular, the inverse is also triangular, *so both* $\boldsymbol{F}^{-1}$ *and F have the same shape*. This means that when we go from model state to the control vector, or from the control vector to the model state as in eq. (4), vorticity is not affected by the balance operators. If we only consider vorticity, the transform from model state to the control vector can be written as

$$\begin{aligned} \boldsymbol{L}_b &= \gamma_b^{-1/2}\boldsymbol{V}_b^T\lambda_b^{-1/2}\sigma_b^{-1} \\ \hat{\chi} &= \boldsymbol{L}_b\delta\hat{\zeta} \end{aligned} \tag{5}$$

but in the minimisation only the inverse and its adjoint are used

$$\begin{aligned} \boldsymbol{L}_b^{-1} &= \sigma_b\lambda_b^{1/2}\boldsymbol{V}_b\gamma_b^{1/2} \\ \delta\hat{\zeta} &= \boldsymbol{L}_b^{-1}\hat{\chi} \end{aligned} \tag{6}$$

3. Formulation of the host model constraint

We believe that the host model is better at describing the large-scale flow of which vorticity is a good descriptor. From the previous section it is also clear that vorticity has very appealing characteristics because it is also a data-assimilation control variable. Therefore we add an extra term to the cost-function that describes the difference between the regional- and host-model vorticity fields:

$$\begin{aligned} J &= J_b + J_o + J_k \\ J_k(\hat{\zeta}) &= \frac{1}{2}(\hat{\zeta} - \hat{\zeta}_{ls})^T\boldsymbol{B}_{ls}^{-1}(\hat{\zeta} - \hat{\zeta}_{ls}) \end{aligned} \tag{7}$$

$\hat{\zeta}_{ls}$ is represented by a short forecast from the host model and $\boldsymbol{B}_{ls}$ contains the error covariances of $\hat{\zeta}_{ls}$. For simplicity we will ignore the cross-correlations between the errors of $\hat{\zeta}_{ls}$ and $\hat{x}_b$ as well as the cross-correlations between the errors of $\hat{\zeta}_{ls}$ and the observation errors. The cross-correlations between the errors of $\hat{\zeta}_{ls}$ and $_b$ are certainly different from zero and will be examined in the next section. Using a host model forecast instead of a host model analysis eliminates the need to consider the cross-correlations between the errors in $\hat{\zeta}_{ls}$ and the observation errors, since the observations used to determine the analysis from which $\hat{\zeta}_{ls}$ originates are at least 6 h old. Also, most regional models do not have the global host model analysis available in real time.

Now we set $\hat{\zeta} = \hat{\zeta}_b + \delta\hat{\zeta}$ in eq. (7):

$$J_k(\delta\hat{\zeta}) = \frac{1}{2}(\hat{\zeta}_b + \delta\hat{\zeta} - \hat{\zeta}_{ls})^T\boldsymbol{B}_{ls}^{-1}(\hat{\zeta}_b + \delta\hat{\zeta} - \hat{\zeta}_{ls}) \tag{8}$$

where $\hat{\zeta}_b$ is the regional model background.

We now also denote $\hat{d}_k = \hat{\zeta}_b - \hat{\zeta}_{ls}$ which gives

$$J_k(\delta\hat{\zeta}) = \frac{1}{2}(\hat{d}_k + \delta\hat{\zeta})^T\boldsymbol{B}_{ls}^{-1}(\hat{d}_k + \delta\hat{\zeta}) \tag{9}$$

B_{ls} is not a diagonal matrix which makes this formulation difficult to use. Therefore a pre-conditioning is introduced using a similar transform as for the background error covariances $\boldsymbol{B}$, but using the error characteristics of $\boldsymbol{B}_{ls}$:

$$\boldsymbol{L}_{ls} = \gamma_{ls}^{-1/2}\boldsymbol{V}_{ls}^T\lambda_{ls}^{-1/2}\sigma_{ls}^{-1} \tag{10}$$

Projecting both $\hat{d}_k$ and $\delta\hat{\zeta}$ through $\boldsymbol{L}_{ls}$ will give a state vector whose error covariance matrix is the identity matrix.

$$J_k(\delta\hat{\zeta}) = \frac{1}{2}(\boldsymbol{L}_{ls}\hat{d}_k + \boldsymbol{L}_{ls}\delta\hat{\zeta})^T(\boldsymbol{L}_{ls}\hat{d}_k + \boldsymbol{L}_{ls}\delta\hat{\zeta}) \tag{11}$$

This pre-conditioning hopefully also helps assuring a good minimisation towards the host model vorticity field.

As a last step, eq. (6) is used to express J_k as a function of the control vector.

$$J_k(\hat{\chi}) = \frac{1}{2}(\boldsymbol{L}_{ls}\hat{d}_k + \boldsymbol{L}_{ls}\boldsymbol{L}_b^{-1}\hat{\chi})^T(\boldsymbol{L}_{ls}\hat{d}_k + \boldsymbol{L}_{ls}\boldsymbol{L}_b^{-1}\hat{\chi}) \tag{12}$$

The derivative of $J_k(\hat{\chi})$ with respect to $\hat{\chi}$ gives a gradient that can be directly added to the total gradient, that now becomes

$$\begin{aligned} \frac{\partial J}{\partial\hat{\chi}} = &\underbrace{\hat{\chi}}_{\nabla J_b} + \underbrace{\boldsymbol{L}_b^{-T}\boldsymbol{H}^T\big(H(\mathbf{x}_b) - \boldsymbol{y} + \mathbf{H}\mathbf{L}_b^{-1}\hat{\chi}\big)}_{\nabla J_o} \\ &+ \underbrace{\boldsymbol{L}_b^{-T}\boldsymbol{L}_{ls}^T(\boldsymbol{L}_{ls}\hat{\mathbf{d}}_k + \boldsymbol{L}_{ls}\boldsymbol{L}_b^{-1}\hat{\chi})}_{\nabla J_k} \end{aligned} \tag{13}$$

4. Error covariances

To implement the host model constraint J_k we need the error covariances of the ECMWF model as seen in the HIRLAM model geometry. To achieve this, ECMWF forecasts were interpolated to the HIRLAM geometry and then processed with the software used to determine the HIRLAM background error statistics. Both HIRLAM and interpolated ECMWF fields were processed simultaneously which allowed for cross-covariances between the errors of two models to be calculated.

The statistics presented will focus on displaying contents of importance for this work. For example, the contents of the $\boldsymbol{L}_b^{-1}$ and $\boldsymbol{L}_{ls}$ operators that will be used in the assimilation, eqs. (6) and (10).

4.1. Regional model geometry

All calculations are carried out on the HIRLAM reference area, called Regular Cycle with the Reference (RCR), Fig. 1. The horizontal grid is expressed in a rotated lat/lon geometry defined by shifting the South pole. Model top is at 10 hPa and there are 60 vertical levels. Model levels are numbered from the top and down, i.e. model level 1 is at

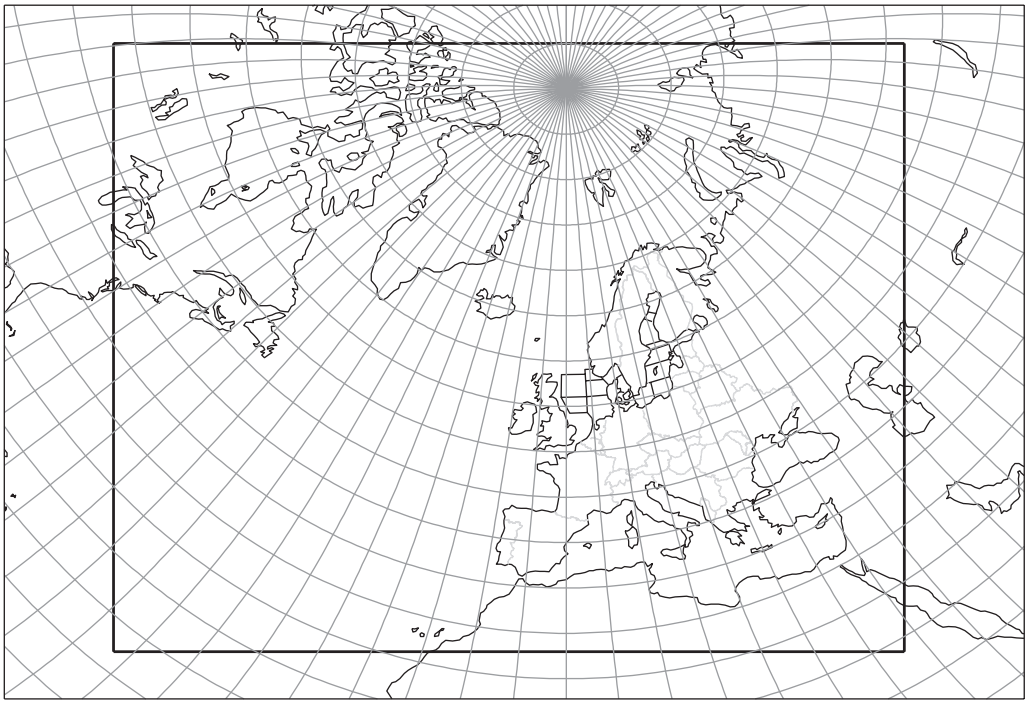

Fig. 1. Geographical coverage of the HIRLAM reference, called RCR, domain which is used in these experiments.

the top and level 60 is closest to ground. The data-assimilation calculations are carried out in spectral space and therefore an area extension, following Haugen and Machenhauer (1993), is used to ensure periodic variations in both horizontal dimensions. The horizontal resolution and number of gridpoints are presented in Fig. 2.

On each vertical level, an arbitrary variable, $\boldsymbol{a}(x,y)$, is represented by a bi-Fourier series:

$$\boldsymbol{a}(x,y) = \frac{1}{\sqrt{NXLNYL}} \sum_k \sum_l \hat{\boldsymbol{a}}_{kl} e^{2\pi i\left(\frac{kx}{NXL}+\frac{ly}{NYL}\right)} \tag{14}$$

So the variable $\boldsymbol{a}$ is in spectral space defined by the *complex coefficients* $\hat{\boldsymbol{a}}_{kl}$. An elliptical truncation

$$\left(\frac{k}{K_{\max}}\right)^2 + \left(\frac{l}{L_{\max}}\right)^2 \leq 1 \tag{15}$$

is used to obtain an isotropic and homogeneous resolution over the whole area. Each wave-number pair (k, l) is linked to a total wave-number k^* by

$$k^* = N_s \sqrt{\left(\frac{k}{NXL}\right)^2 + \left(\frac{l}{NYL}\right)^2} \tag{16}$$

where N_s is a scaling factor that is equal to the length of the extended domain divided by a map factor.

4.2. Host model geometry

The ECMWF data used in the forecast experiments in this work originate from a spectral *T*799 model with 91 vertical levels and the model top at 0.01 hPa. *T*799 corresponds to a gridpoint resolution of about 25 km in a reduced Gaussian grid.

It should be noted that the ECMWF forecasts used for the derivation of the host model error statistics in this study were retrieved from the ECMWF archives in a horizontal resolution of 0.5 latitude–longitude. These forecast fields were subject to a bi-linear horizontal interpolation followed by a vertical interpolation to the HIRLAM geometry. Due to the simple interpolation procedure, one should not trust statistical information concerning horizontal waves shorter than approximately 150 km.

4.3. Statistics

Short-range forecast errors were simulated by applying an ad hoc assumption that differences between forecasts of different length valid at the same time can be taken as a proxy for such forecast errors. Thus, we applied what is referred to as the 'NMC method'. Pairs of +36 h and +12 h forecasts, valid at the same time, were obtained from the ECMWF and HIRLAM RCR archives for 87 situations, all at 00UTC, during the period September–November 2008. Forecast differences were relaxed towards

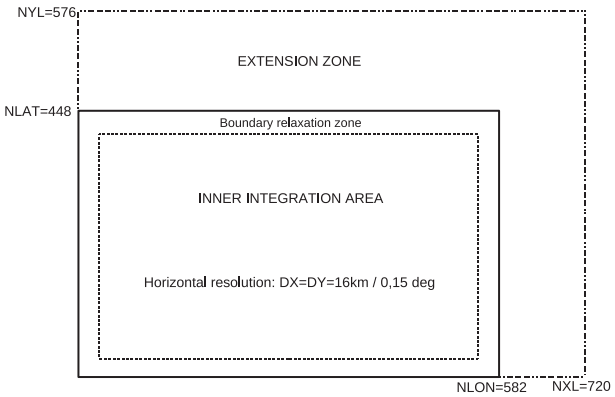

Fig. 2. Integration area of the spectral HIRLAM tangent linear and adjoint model applied in 4D-Var, together with the horizontal resolution and the number of gridpoints for the setup used in the experiments presented in this work.

zero in the lateral boundary relaxation zone of the forecast model (Fig. 2), in order to avoid aliasing effects through the extrapolations to the area extension zone that are applied in order to get bi-periodic variations over the extended domain. All calculations of statistics were carried out in spectral space ignoring covariances between spectral component coefficients of different horizontal wave-numbers, thus implicitly assuming horizontal homogeneity with respect to the covariances in gridpoint space.

The horizontal variance power spectra for vorticity, $P_z(k^*)$, is an output entity from the NMC statistics software that shows the contribution to the total variance for different wavelengths. $P_z(k^*)$ is plotted on three different vertical levels in the left column of Fig. 3, while the cross correlation between the HIRLAM and ECMWF vorticity fields is plotted in the right column of the same figure.

We see that HIRLAM has larger values throughout the whole spectrum. Since the NMC statistics are based on forecast differences, one reason could be that HIRLAM is more active and has a larger spread due to its higher resolution. Another reason may be difference between the data-assimilation methods applied in the forecasting systems producing the forecast fields that enter into the statistical calculations. ECMWF applied 4D-Var, while the HIRLAM fields during autumn 2008 were produced with 3D-Var. It is known that introduction of 4D-Var reduces the 'jumpiness' in the time-series of forecast evolutions, and the difference between +12 h and +36 h forecasts valid at the same time is essentially a measure of jumpiness. Also note that the horizontal resolution used in the retrieval of the ECMWF forecast fields from the archives makes any comparison of covariance spectra meaningless for horizontal wave-numbers greater than 64 (wave length 160 km).

The cross-correlation between ECMWF and HIRLAM vorticity forecast differences is strongest for the longest scales, right column in Fig. 3. Since longer scales are more predictable, and since these larger scales are to a large extent controlled by the host model through the lateral boundary relaxation, it is likely that both models produce somewhat similar large-scale structures which make the shape of the statistics reasonable. It seems like it is safe to ignore the cross-correlations for shorter scales while it may be more questionable for longer scales.

The horizontally averaged standard deviations, Fig. 4, also show that the HIRLAM values exceed the ECMWF ones on all levels. They both have rather similar shapes from the top and down to level 30 with a maximum around model level 15 ($\approx$350 hPa) which corresponds to the jet-stream level. Below, HIRLAM has another maximum around level 43 ($\approx$850 hPa) which is not present in the ECMWF statistics. These horizontally averaged standard deviations are applied in the assimilation weights $\boldsymbol{L}_b^{-1}$ and $\boldsymbol{L}_{ls}$ but are adjusted with an empirical scaling factor. The scaling factor for σ_{ls} is used to adjust the weight of the host model term, see Section 5.2.

Figure 5 shows the horizontal spectral densities, $\lambda_z(k^*)$, as calculated by the NMC statistics software. $\lambda_z(k^*)$ and $P_s(k^*)$ are related via a simple expression:

$$P_z(k^*) = \lambda_z(k^*)\left(2\pi k^* \frac{NXLNYL}{N_s^2}\right)$$

$\lambda_z(k^*)$ is the input to the HIRLAM data-assimilation, but before applied in assimilation they are scaled such that their sum, divided with the number of gridpoints, is equal to 1. It should be noted that the result of that scaling is resolution dependent. So what we see in Fig. 5 is not exactly what is used in the $\boldsymbol{L}_{ls}$ and $\boldsymbol{L}_b^{-1}$ operators, i.e. $\lambda_{ls}^{1/2}$ and $\lambda_b^{1/2}$. In Fig. 5 we see that the ECMWF spectrum is sharper than HIRLAM on all levels. At the same time, the ECMWF values are larger for large scales (small wave-numbers) and vice versa for short scales. For ECMWF in relation to HIRLAM, this means that large scales contribute more than short scales to the horizontal correlation which gives broader horizontal structure functions for ECMWF (not shown). This is natural due to the difference in resolution.

Next we take a look at the eigenmodes of the vertical error covariance matrix, Figs. 6 and 7. Each eigenmode is represented by an eigenvector and a corresponding eigenvalue.

The eigenvector and eigenvalue for mode n and wave-number k^* can be denoted $\boldsymbol{V}_{n,k}^z \,.\, z = 1$, Number of vertical LEVels ($NLEV$) and $\gamma_{n,k}$. The sum of eigenvalues over all modes, for a specific wave-number k^*, is constant

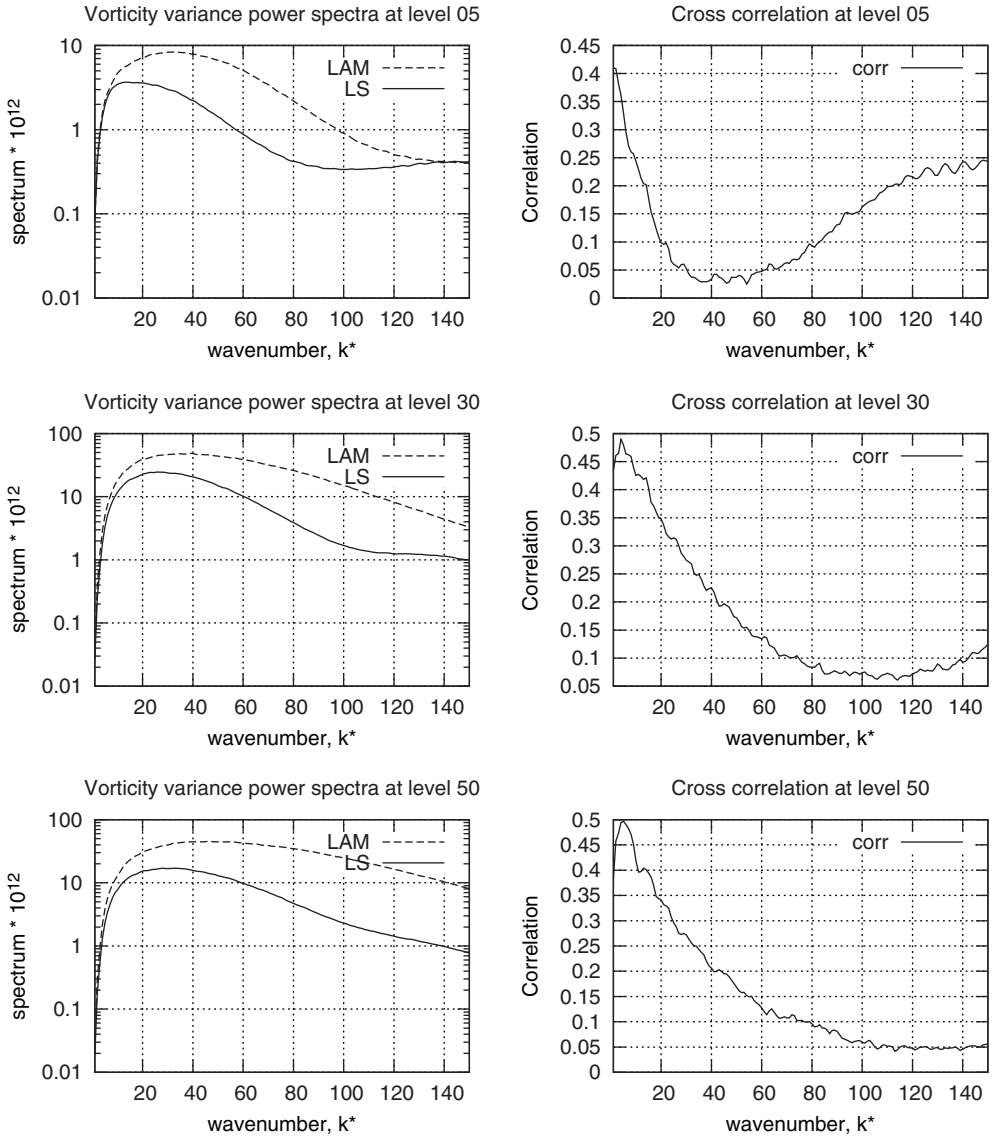

Fig. 3. Left column: Horizontal variance power spectra, $P_z(k^*)$, for vorticity forecast differences at model levels 5 (≈90 hPa), 30 (≈630 hPa) and 50 (≈920 hPa), full line = ECMWF and dotted line = HIRLAM. Right column: cross-correlation of vorticity forecast differences between HIRLAM and ECMWF.

and equal to the number of modes, which is the same as the number of vertical levels, thus:

$$\sum_{n=1}^{Nmod} \gamma_{n,k^*} = NLEV$$

and the ratio

$$\frac{\gamma_{n,k^*}}{\sum_{n=1}^{Nmod} \gamma_{n,k^*}}$$

tells us how much of the total variance that is explained by mode n. How the variance is distributed vertically for eigenmode n is shown by the eigenvector $V^z_{n,k}$.

Figure 6 shows the eigenvalues for the first six leading modes. At the bottom, the combined effect of these modes is shown. We see that these modes include 90% of the variance for the longest scales and then gradually decrease as the scales get shorter. This is important to remember for the ECMWF data because we choose to use only the six first, or leading, modes of $\boldsymbol{V}_{is}$ and γ_{ls} in the operator $\boldsymbol{L}_{ls}$, eq. (10). $\boldsymbol{L}_{ls}$ then acts as a vertical filter that removes part of the small scale variance from the host model, including also vertically small scale noise introduced by the vertical interpolation procedure. In the same figure we also see how the higher order modes contain less large-scale information. Fig. 7 shows the structures of some

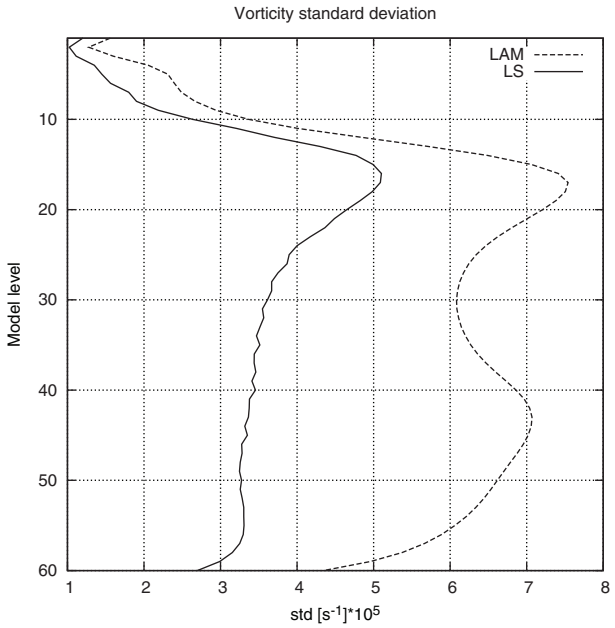

Fig. 4. Vertical profiles of horizontally averaged standard deviation for vorticity forecast differences. These are both scaled in the minimisation to adjust the weight in the data-assimilation. Full line = ECMWF and dotted line = HIRLAM.

eigenvectors. For each mode, the wave-number that is associated with the largest eigenvalue is chosen. The shapes are quite similar with a slight shift of phase for modes 5–8. It should however be mentioned that the shape of the eigenvectors for a specific mode varies a little between the wave-numbers.

5. Implementation

5.1. Low resolution increments

A short ECMWF forecast, with the geometry described in the first paragraph of Section 4.2, is interpolated to the HIRLAM geometry described in Section 4.1. The interpolated ECMWF wind-field is thereafter transformed to spectral space. It is important to mention this because it means that a 25 km grid is interpolated to a 16 km grid, and the ECMWF field is thus oversampled and contains false high-resolution information.

In practice, however, this is never a problem for us due to two reasons:

First, the actual minimisation is a carried out on a grid mesh with lower resolution than the original grid (16 km), this incremental formulation is also described in Gustafsson et al. (2001). One of the reasons for this is to reduce the computational cost. In this work, all analyzes are carried out in a resolution either three or six times lower than the original resolution, i.e. a grid mesh of 48 km or 96 km. This means that the oversampled part of the ECMWF vorticity field is never used.

Second, The J_k term itself is truncated to an even lower resolution as described in Section 5.2.

5.2. Tuning and truncation of the J_k term

In Section 4.3, a scaling of the standard deviation σ_{ls} in the operator $\boldsymbol{L}_{ls}$ [eq. (10)] was mentioned. The purpose of this scaling is to adjust the weight for the J_k-constraint in relation to the other terms in the cost-function. In the HIRLAM variational analysis code, a similar scaling is applied to the background error standard deviation to adjust the weight of the J_b term in relation the observation term J_o. The tuning factor, here called α, is applied as such:

$$\boldsymbol{\sigma}_{ls}^{\text{scaled}} = \sqrt{\alpha}\boldsymbol{\sigma}_{ls}$$

and the actual tuning procedure in this study is rather ad hoc. A few different values of scaling-factors, $\alpha = 1$, 5, 10, 100, were applied on a 3D-Var test case with $J = J_b + J_o + J_k$ and the behaviour of the cost-function, and the individual terms separately, during minimisation, were studied. $\alpha = 100$ gave J_k no weight at all, while $\alpha = 1$ made J_k too dominating. With $\alpha = 5$ and $\alpha = 10$ the J_k and J_o terms both minimised smoothly. $\alpha = 5$ however gave a more distinct minimisation of the J_k term without losing the fit towards observations. Therefore $\alpha = 5$ was used throughout all experiments in this study.

The J_k term is also truncated at $k^* = 45$, wave length around 230 km, to make sure only large scales are penalised. In terms of low-resolution increments, described in Section 5.1, $k^* = 45$ corresponds to a resolution about seven times lower than the original resolution.

5.3. 3D-Var single case test

As a first step, we perform a 3D-Var analysis with only the background and J_k as constraints, i.e. $J = J_b + J_k$. The date of the single case test was 22 October 2008 at 06UTC, and both $\boldsymbol{x}_b$ and ζ_{ls} are +06 h forecasts from the previous cycle.

The analysis converges in four iterations, not shown, and the resolution was 48 km, i.e. three times lower than the original resolution but the J_k term was truncated at $k^* = 45$ as mentioned in the previous section. Both the first guess $\boldsymbol{x}_b$ and analysis $\boldsymbol{x}_a$ are then compared with the host model field x_{ls}. This is done by computing Root Mean Square (RMS) values in gridpoint space for $(\boldsymbol{x}_b - \boldsymbol{x}_{ls})$ and $(x_a - x_{ls})$ for the upper air variables ζ, $\boldsymbol{u}$, υ, $\boldsymbol{T}$, $\boldsymbol{q}$. The differences are computed for every vertical level and are presented as profiles in Fig. 8.

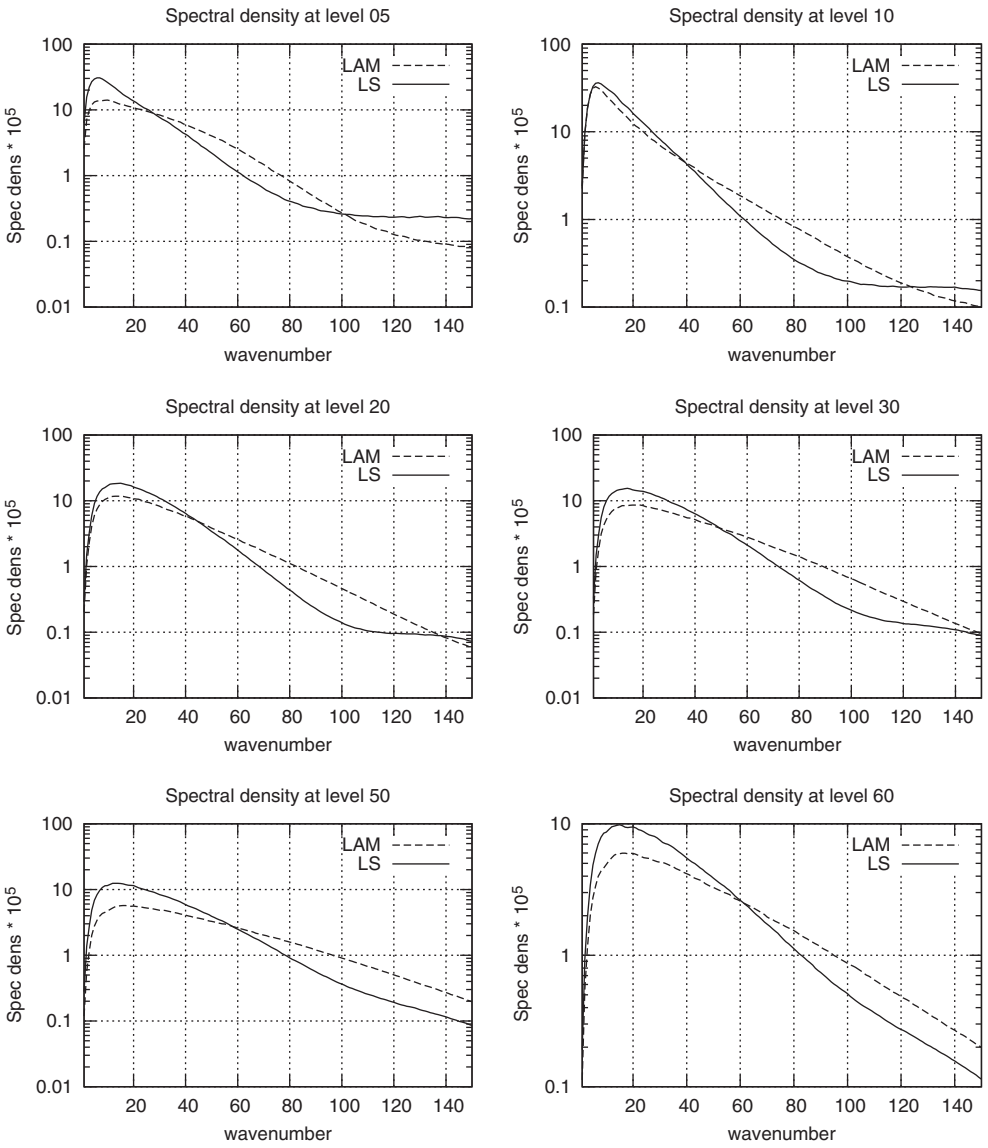

Fig. 5. Horizontal spectral densities for vorticity forecast differences, γ. Full line = ECMWF, dotted line = HIRLAM. Before these are applied in the assimilation they undergo a scaling so that their sum, divided with the number of gridpoints, is equal to 1.

Because J_k contains vorticity only, with no cross-correlations to any other variables, the analysed vorticity field must be closer to ECMWF than the first guess, otherwise, there would be something wrong with the implementation.

We now see in the upper left plot in Fig. 8 that the analysed vorticity field is clearly closer to ECMWF compared with the first guess. So the J_k implementation seems to do what it is supposed to. Differences in the other variables are produced by HIRLAM's internal balances via the B matrix and do therefore not have to be closer to ECMWF.

It is also interesting to investigate how the host model constraint acts on different scales. For simplicity, we only look at the vorticity field and compare the complex spectral coefficients of the differences $\left(\hat{\zeta}_b - \hat{\zeta}_{ls}\right)$ and $\left(\hat{\zeta}_b - \hat{\zeta}_{ls}\right)$. As explained in Section 4.1, these differences are represented as bi-Fourier series as shown in eq. (14), and each wave-number pair (k, l) is linked to the total wave-number k^* [eq. (16)].

To be able to visualise these difference fields they have been expressed as 1-D arrays over k^*. Each element is the average of the absolute values of all spectral coefficients with wave-number pairs (k, l) that are linked to the same k^*.

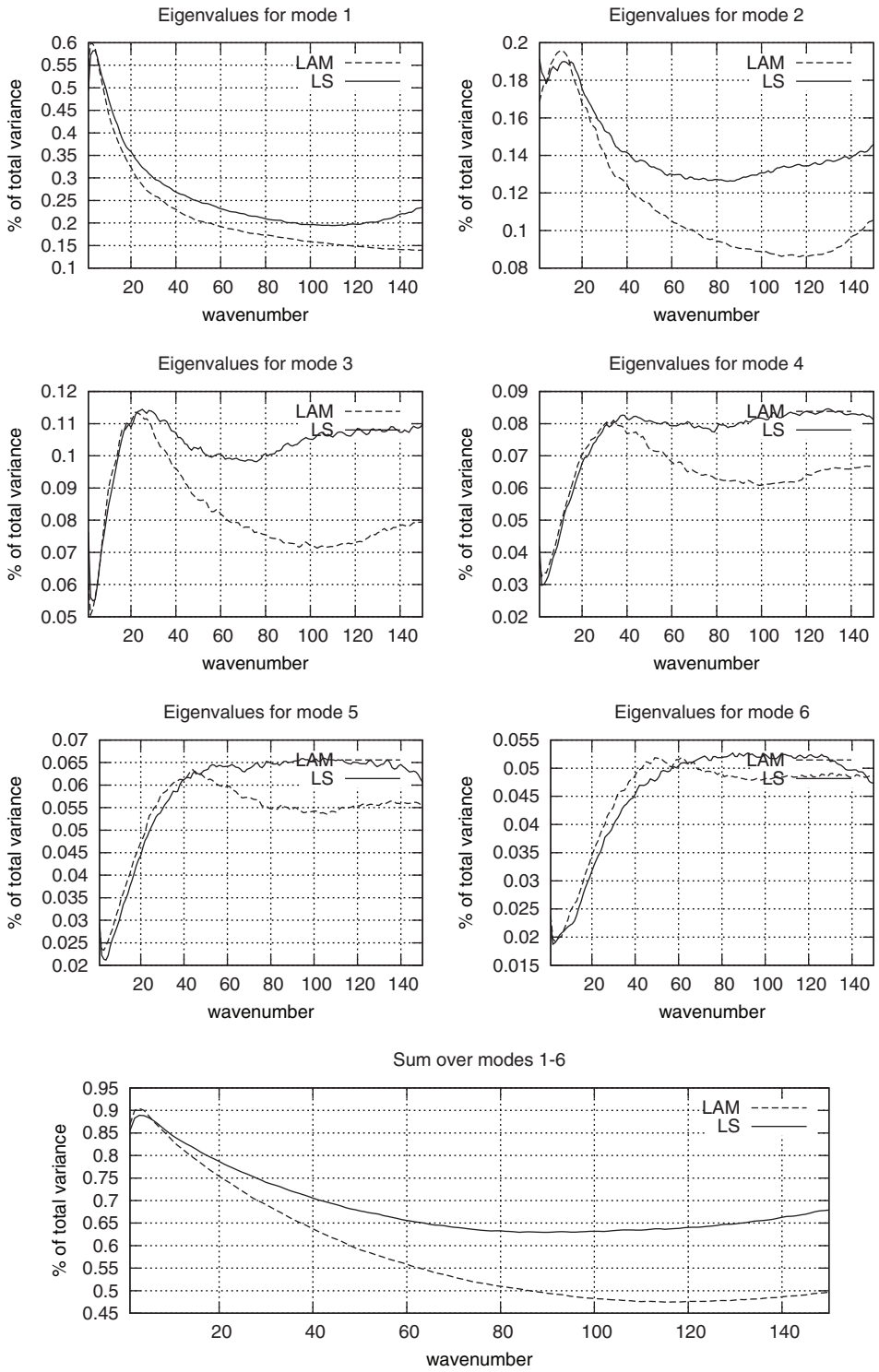

Fig. 6. Eigenvalues for the vorticity forecast difference vertical covariance matrix, the first six leading modes. Mode number 1 is the leading mode and so on. The sum over the six modes are presented in the bottom figure.

In Fig. 9, the top figure shows the averaged 1-D spectrum of the difference $\left(\hat{\zeta}_b - \hat{\zeta}_{ls}\right)$ as a full line and $\left(\hat{\zeta}_b - \hat{\zeta}_{ls}\right)$ as a dotted line. It can now be seen that the analysed vorticity is drawn towards ECMWF on all scales where J_k is active.

The size of the increments is also shown in the middle of Fig. 9. This plot shows the average absolute value of the

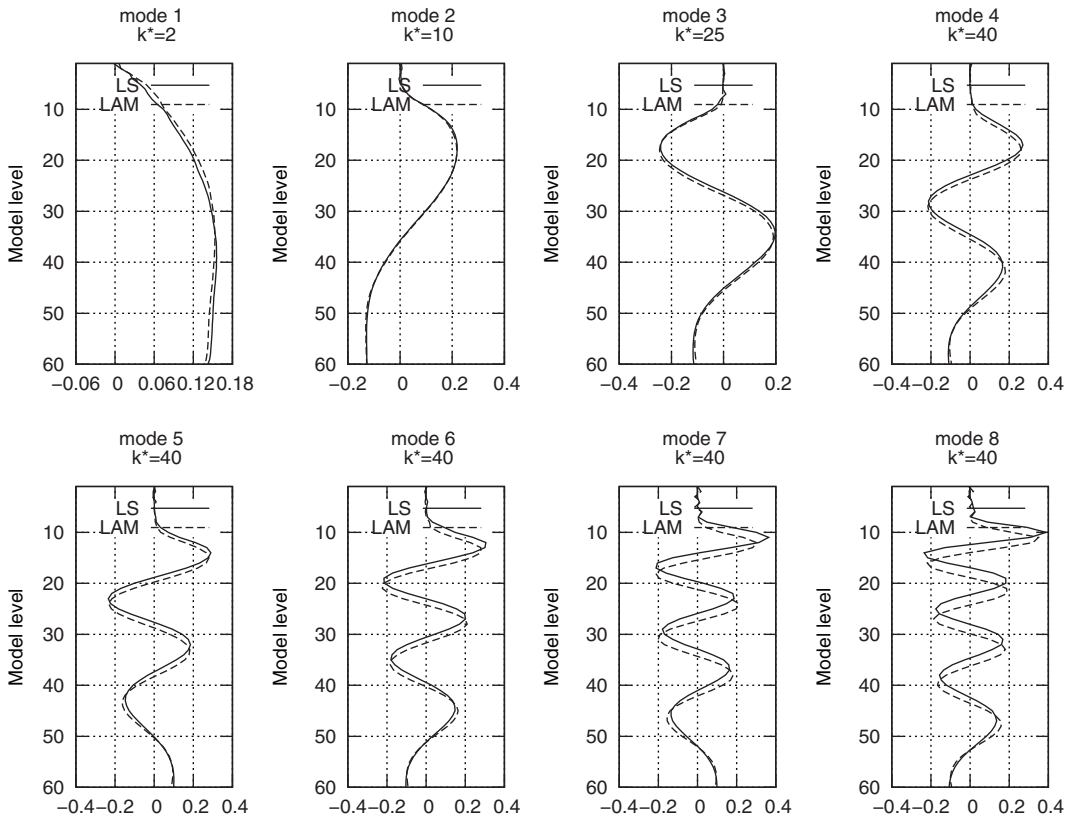

Fig. 7. Eigenvectors for the first eight leading modes for the vorticity forecast difference vertical covariance matrix. The wave-number k^* for each mode is the one that has the largest eigen-value, see Fig. 6.

vorticity increments $\delta\hat{\zeta}$ at the last iteration which corresponds to the difference between the full and dotted line in the top plot in Fig. 9.

The largest increments are introduced at $k^* = 2$–8, peaking at 5 and thereafter slowly decreases until $k^* = 45$ where it becomes zero.

However, the relative size of the increments is rather uniformly distributed over the spectrum as shown in the bottom plot of Fig. 9.

Now we have verified that the implementation of the J_k term draws the analysis closer to the host model (ECMWF) vorticity field, and that the largest increments are introduced on the longest scales.

6. Forecast impact

6.1. *Analysis setup*

A 4D-Var setup with two outer loops was chosen for the forecast impact experiment in order to be close to what is used operationally at SMHI.

Two outer loops in HIRLAM means that the inner loop begins in a coarse resolution and performs a number of iterations. Then, as an intermediate step, the control vector is saved and a temporary analysis file written from which a new, non-linear trajectory is calculated with the forecast model to make a re-linearization of the observation operators and for the tangent-linear model. The second inner loop then continues in a higher resolution to finish the analysis.

In HIRLAM the 4D-Var time-window is ± 3 h around the nominal analysis time and the first outer loop uses the $+03$ h forecast from the previous cycle as first guess. Since the control vector is updated and the cost-function gradient is calculated at the beginning of the time-window it is appropriate to also use a $+03$ h forecast for the host model constraint.

The first loop is carried out with 96 km resolution (six times lower than original) for a maximum of 60 iterations, and the second loop is carried out in 48 km resolution for a maximum of 30 iterations.

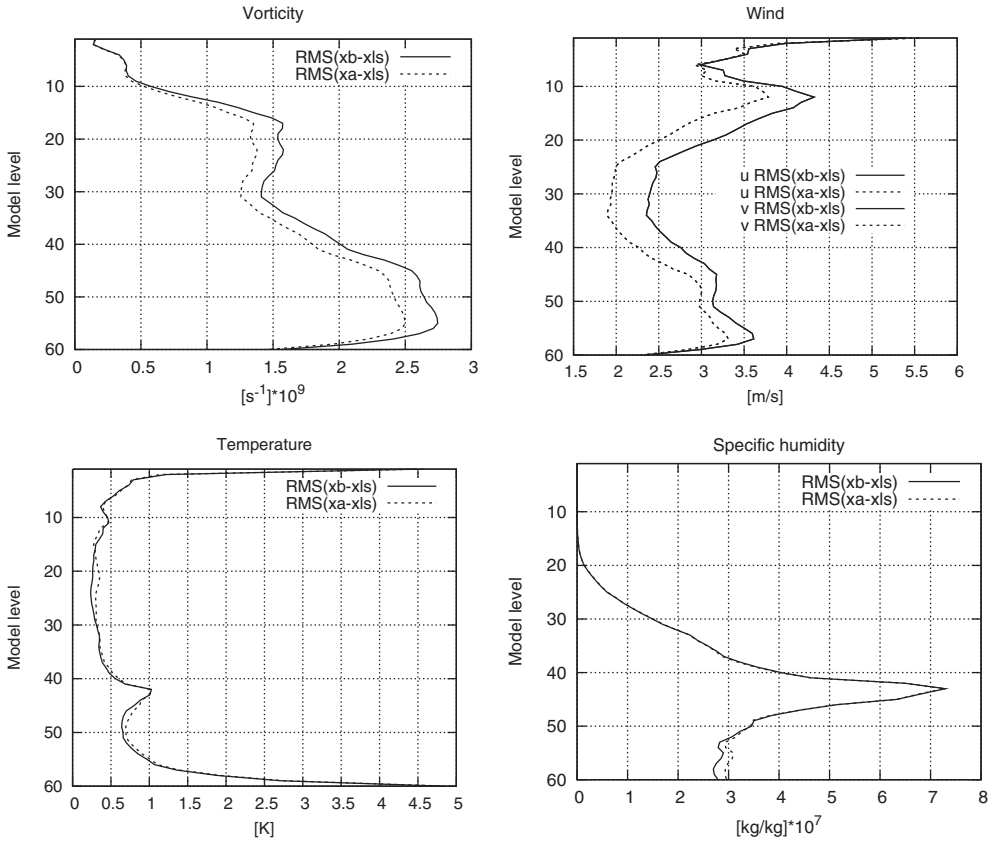

Fig. 8. Analysis impact on upper air fields for a test case using 3D-Var and only the background and host model constraints, i.e. $J = J_b + J_k$ and thus no observations. Full line: RMS difference between x_b and x_{ls}. Dotted line: RMS difference between x_a and x_{ls}. x_b is the background and first guess HIRLAM field and x_{ls} is the ECMWF $+06$ h forecast used as constraint in J_k. Notice that the inclusion of J_k draws the analysed vorticity field closer to ECMWF.

6.2. Experiment period and observation usage

The model was run up to $+48$ h every 6th hour from 24 January 00UTC to 25 February 2009 at 18UTC, i.e. 33 d. For simplicity, only conventional observations were used in the data-assimilation. Conventional observation types include SYNOP, SHIP, BOUY, TEMP, PILOT and AIRCRAFT.

Two sets of assimilation experiments were carried out throughout the 33-d period:

- *ref* – reference run. Only observations used in the analysis.
- *jk_active* – same as reference but with the host model constraint activated in data-assimilation.

6.3. 4D-Var minimisation case

As in the 3D-Var singe case test, it is interesting to see the cost-function behaviour for a 4D-Var case with observations. One date in the impact experiment was chosen arbitrarily, 24 January 2009 at 12UTC, and studied. The cost-function contains three terms J_b, J_o and J_k and each part is plotted separately, together with the total value J_{tot} in Fig. 10.

It is important to note that the cost-function values of the J_b and J_k terms are resolution dependent. J_b is the scalar product of the control vector, $J_b = \hat{\chi}^T\hat{\chi}$, and similarly the J_k value is the scalar product described in eq. (12). These vectors are thus longer in the 2nd outer loop. At the same time, the horizontal spectral density, λ_b and λ_{is}, undergoes a resolution-dependent scaling described in Section 4.3, as well as the coefficients of the spectral fields. These three factors combined, cause the cost-function values of J_b and J_k to make a discontinuous jump between the 1st and 2nd outer loops. Because of this discontinuity, the decrease of the cost-function value should be compared with the initial value of each loop.

In Fig. 10 we see that the total cost-function value, as well as the observation term J_o, decreases and converges in both loops, while J_k decreases and converges, already in the

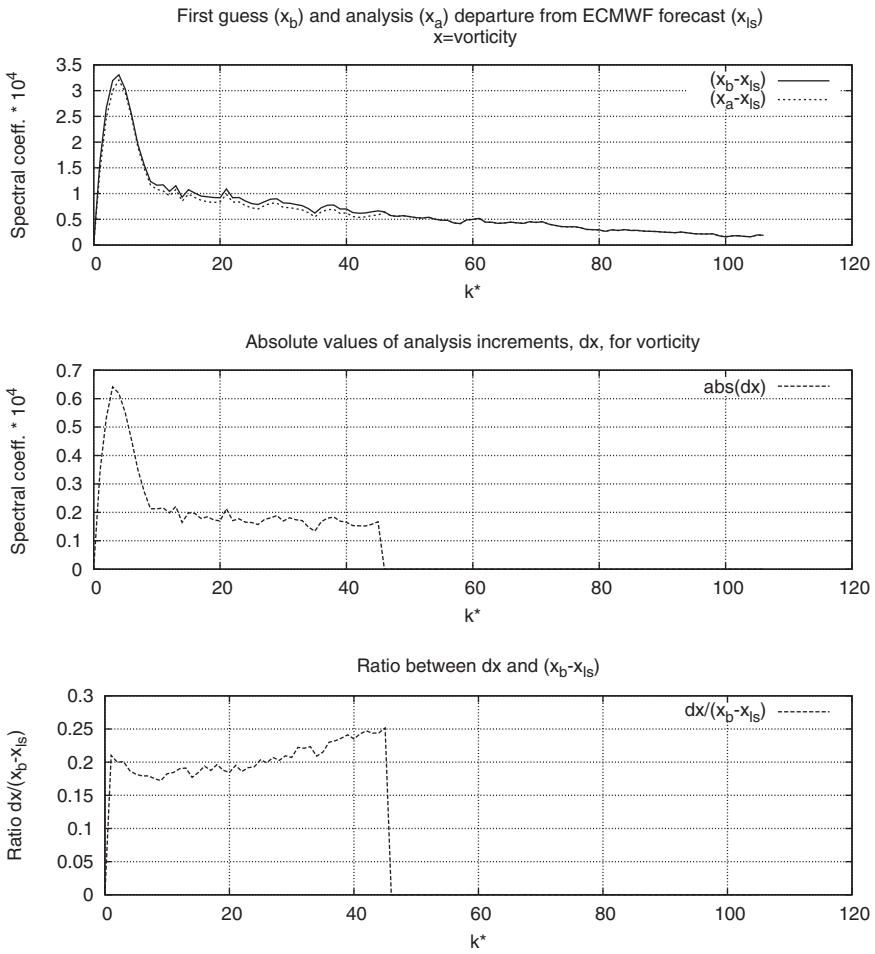

Fig. 9. Spectrum of analysis impact on vorticity for a test case using 3D-Var and only the background and host model constraints, i.e. $J = J_b + J_k$ and thus no observations. Top figure, full line: absolute value of spectral coefficients for the difference $(\hat{\zeta}_b - \hat{\zeta}_{ls})$. Top figure, dotted line: absolute value of spectral coefficients for the difference . Middle figure: absolute value of the analysis increments for vorticity, $\delta\hat{\zeta}$ which corresponds to the difference between the full and dotted line in the top plot. Bottom figure: size of analysis increments relative to the first guess and host model vorticity fields, i.e. $\delta\hat{\zeta}/(\hat{\zeta}_b - \hat{\zeta}_{ls})$.

1st loop, and is then almost constant in the 2nd. It is possible that the reason for this is that J_k only acts on large scales which converges faster. The main conclusion from Fig. 10 is that the analysis draws closer to both observations and J_k simultaneously.

Root Mean Square differences for this particular case, as described in Section 5.3 and Fig. 8, also show that the analysed vorticity is clearly closer to ECMWF (not shown).

6.4. Observation verification results

Forecasts were verified by computing mean and RMS errors against all available SYNOP and radiosonde stations in the area. In this case it adds up to around 2000 SYNOP and 175 radiosonde stations.

For surface parameters like 2 m temperature, humidity and 10 m wind the impact was neutral. Mean sea level pressure showed a clear positive impact, Fig. 11. Upper air parameters, computed against radiosondes, also had a clearly reduced RMS error for the *jk_active* experiment for temperature, wind, geopotential and specific humidity. Fig. 12 shows vertical profiles of mean and RMS errors of temperature calculated at midnight, 00UTC and noon, 12UTC.

6.5. One case for illustration

Before making the concluding remarks, we also compare one case in the forecast experiment period and see what the differences looks like. The date was chosen from a time-series of RMS errors of +24 h forecasts of surface

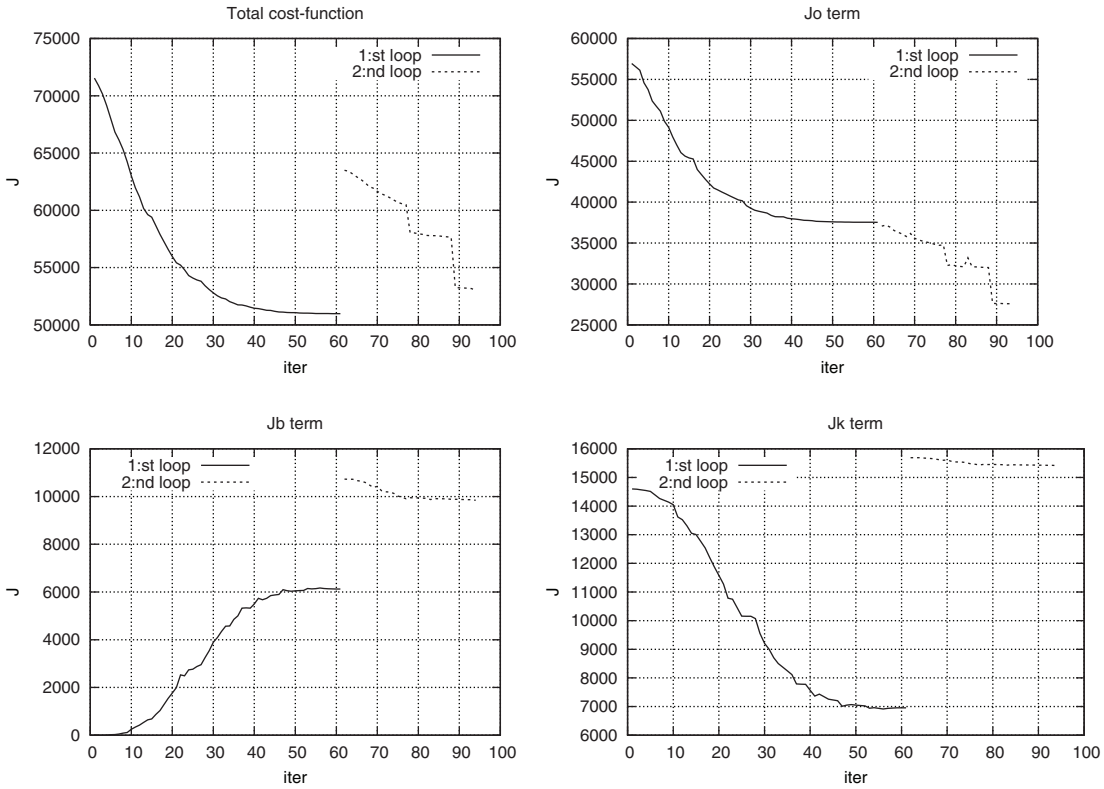

Fig. 10. Cost-function behaviour for one case in the impact experiment. Here, a 4D-Var setting with two outer loops is used. As mentioned in Section 6.3, the resolution dependency of J_b and J_k is clearly shown here.

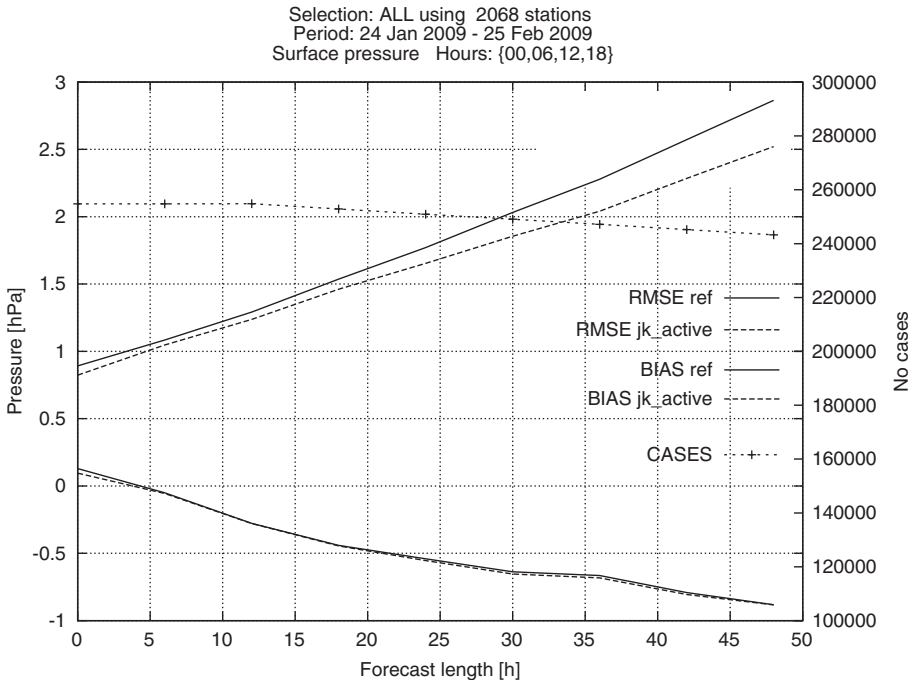

Fig. 11. Verification of forecasts compared with observations of mean sea level pressure (MSLP). Upper lines are the Root Mean Square (RMS) error and the lower lines show the mean error, or bias. Full line: reference experiment (i.e. the host model constraint is not activated). Dotted line: same as reference except that J_k is now actively used. The amount of observations used in the statistics is also shown using the right hand y-axis.

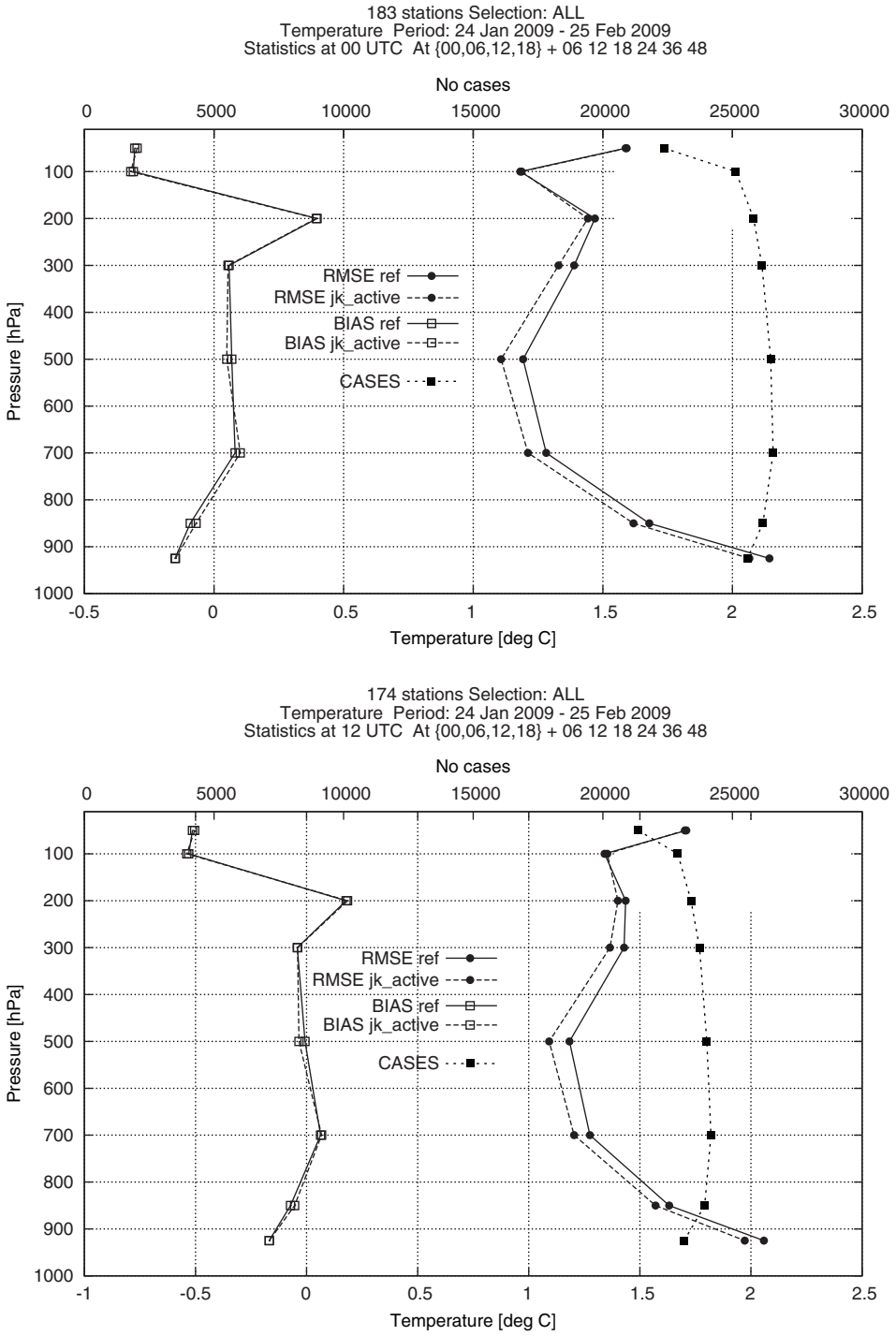

Fig. 12. Verification of temperature forecasts compared with observations from radiosondes. To the right we have the Root Mean Square (RMS) error and to the left the mean error, or bias, averaged of all forecast lengths, valid at 00UTC (upper figure) and 12UTC (lower figure) are shown. Full line: reference experiment (i.e. the host model constraint is not activated.) Dotted line: same as reference except that J_k is now actively used. The amount of observations used in the statistics is also shown using the upper x-axis.

pressure, Fig. 13. Around 30 January 2009, the RMS errors of +24 h forecasts differed with approximately 1.0 hPa between the experiments, with advantage to the *jk_active* experiment. Therefore the 12UTC run from 29 January was chosen for the study since the +24 h forecast from that time is valid on 30 January. We are thus

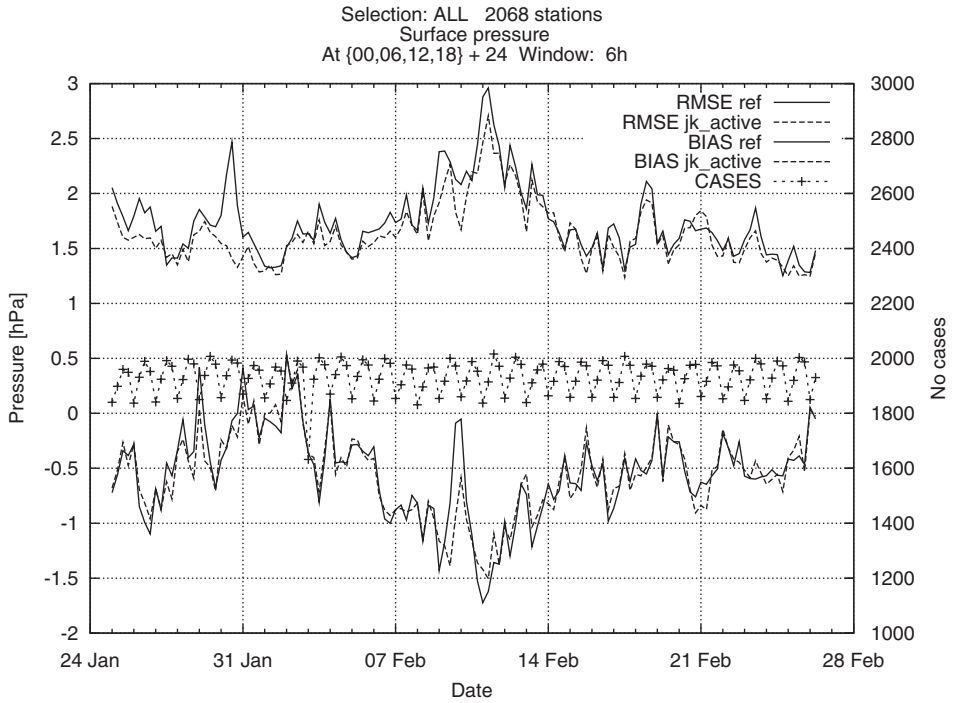

Fig. 13. Time-series of surface pressure +24 h forecast errors compared with observations. The two lines at the top show the RMS errors, and the ones at the bottom the mean error. Full line: reference experiment (i.e. the host model constraint is not activated. Dotted line: same as reference except that J_k is now actively used.

5 d into the experiment period and the two experiments have evolved separately through a number of data-assimilation cycles and forecast integrations. One cannot expect large-scale differences only just because J_k acts on large scales. The non-linear forecast integration can generate differences on all scales.

Figure 14 shows the analyzes of the mean sea level pressure (MSLP) fields, from the experiment (*jk_active*) and reference (*ref*) runs, and the difference between them. The overall large-scale synoptic features are quite similar with an intense low at the coast of Newfoundland, and a high pressure system to the south-east, over the Atlantic. A large trough stretches from Greenland to the south-east with a low-pressure system just to the West of Ireland. There is also another low-pressure minima near Iceland. Then there is a massive high pressure over northern Russia with a ridge that stretches to the south-west into Scandinavia. The difference map reveals that the most noticeable differences are associated with the low-pressure system to the West of Ireland. A closer look at the MSLP fields shows that it is the shape of the low-pressure system that is different, and that the low is bit deeper in the experiment (*jk_active*).

When this is integrated forward in time we see that it is a blocking high pressure over Russia/Scandinavia, and the low to the West of Ireland moves straight to the North while the low outside of Newfoundland moves to the East. After 24 h, Fig. 15, the low originally to the West of Ireland is located near Iceland. The differences have grown in both spatial scale and magnitude and are still associated with the low-pressure system near Iceland. If we compare the MSLP fields we see that it is the shape and the depth of the low that have changed. The low in the reference run is now deeper and more intense than the low in the (*jk_active*) experiment. This low has thus developed differently in the experiments. In the reference run, the low moved northwards and intensified while the low in (*jk_active*) experiment moved to the North and weakened a bit.

7. Summary and concluding remarks

In this paper, we have described and tested a method to assimilate the host model information (ECMWF) into a regional NWP model (HIRLAM). The vorticity field from the host model was chosen to act as a constraint in the regional model variational analysis. For simplicity, the host model vorticity constraint was assumed to be univariate, i.e. no cross-correlations to any regional model variables or observations. This, along with the characteristics of vorticity in the statistical balance, led to an error covariance matrix that could be made diagonal by using the horizontal spectral densities and vertical eigenmodes.

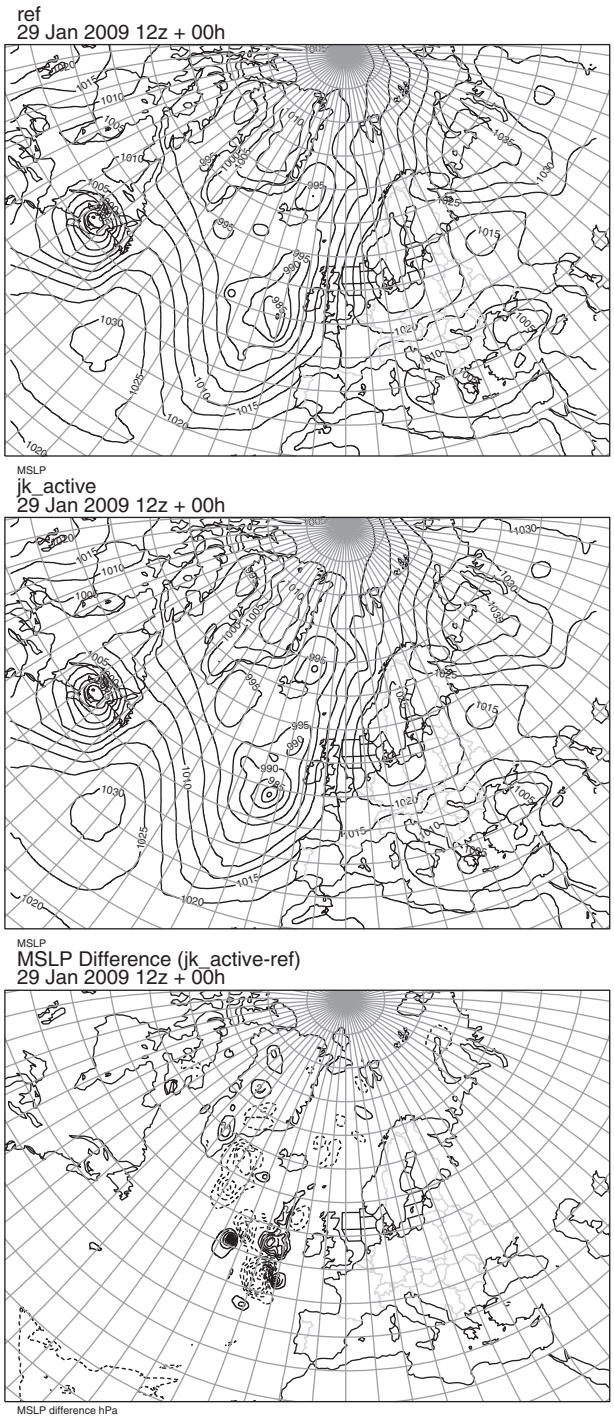

Fig. 14. Analyzes of mean sea level pressure (MSLP) on 29 January 2009 12UTC. *ref* (top) refers to the run in which J_k is not used, *jk_active* (middle) is the run that uses J_k. The difference map (bottom) is the difference between *jk_active* and *ref*.

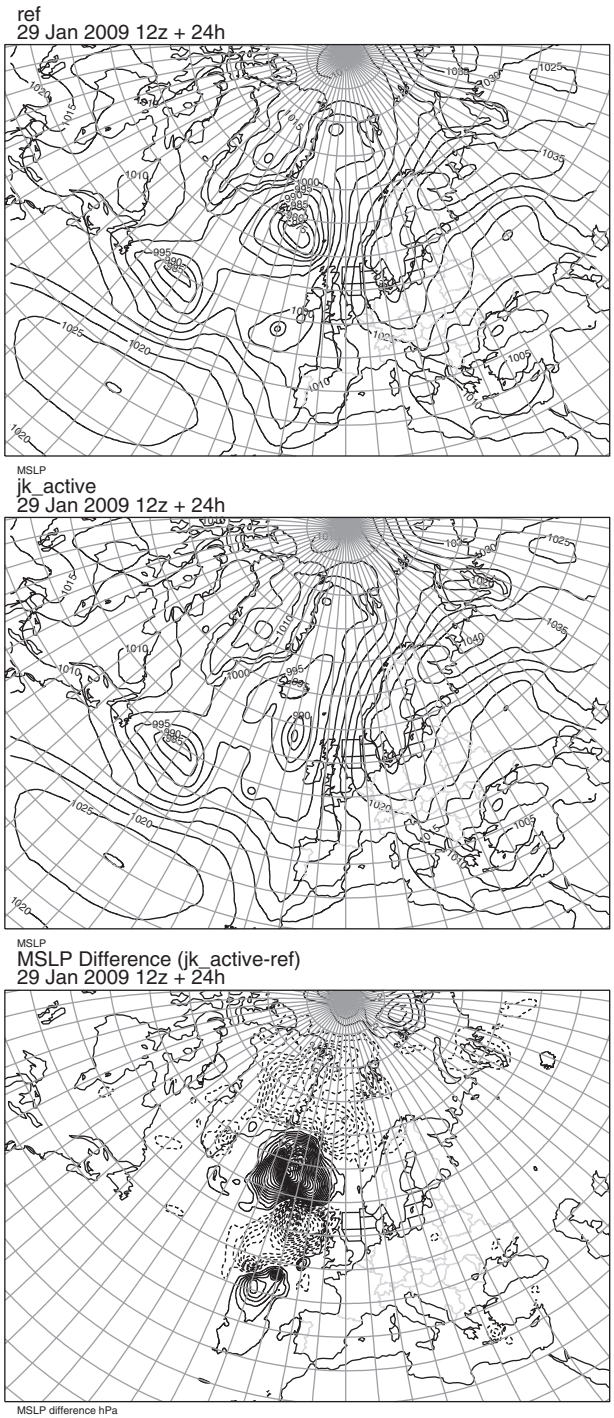

Fig. 15. Mean sea level pressure fields for +24 h forecasts from 29 January 2009 12UTC, thus valid at 30 January 2009 12UTC.

A forecast experiment of 33 d has given very promising results with reduced RMS errors of upper air forecast fields compared with radiosonde observations as well as mean sea level pressure forecast fields compared with synop observations.

The study clearly indicates that the host model information is important to include when determining the initial condition.

8. Acknowledgements

Constructive comments from Xiang-Yu (Hans) Huang, NCAR and one anonymous reviewer helped us to improve this manuscript and these contributions are acknowledged.

References

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

Guidard, V. and Fischer, C. 2008. Introducing the coupling information in a limited-area variational assimilation. *Q. J. R. Meteorol. Soc.* **134**, 723–736.

Gustafsson, N. 2006. Status and performance of HIRLAM 4D-VAR. *HIRLAM Newslett.* **51**, 10pp.

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

Haugen, J.-E. and Machenhauer, B. 1993. A spectral limited-area model formulation with time-dependent boundary conditions applied to the shallow water equations. *Mon. Weather Rev.* **121**, 2618–2630.

Huang, X.-Y., Yang, X., Gustafsson, N., Mogensen, K.-S. and Lindskog, M. 2002. Four-dimensional variational data assimilation for a limited area model. *HIRLAM Tech. Rep.* **57**, 10pp.

Lindskog, M., Gustafsson, N., Navascués, B., Mogensen, K. S., Huang, X.-Y. and co-authors. 2001. Three-dimensional variational assimilation for a limited area model. Part II: observation handling and assimilation experiments. *Tellus* **53A**, 447–468.