Report of WP4 Carpe Diem Project

Sara Riccardo, PROGEA S.L.R. and Magnus Lindskog, SMHI

Summary

Inside the WP4 of the european project Carpe Diem we investigate the Numerical Weather Prediction (NWP) model uncertainty including model errors. In particular a mrthodology for estimation of forecast error variations is developed and implemented into the HIRLAM (HIgh Resolution Limited Area M odelling) 3-dimensional variational data assimilation system.

1 Introduction

Data assimilation in Numerical Weather Prediction (NWP) optimally blends observations with an atmospheric model in order to obtain the spatial distribution of atmospheric variables and to produce the best possible model initial state. The number of observations is generally small compared to the number of degrees of freedom in the initial state of the forecast model. This problem is overcome by introducing some *a priori* information. The *a priori* information is usually a short-range forecast, supplemented with statistical information about its errors, in addition to statistical information about the observation errors.

A number of methods have been applied to estimate static statistics of background error covariaces. One approach is to derive the background error variances from statistics of innovation vectors, under the assumption that the spatial structure of the observation errors are known. This methodology is usually referred to as the Hollingsworth and Lönnberg method (Hollingsworth and Lönnberg,1986). An other approach is to consider statistics of forecast differences as proximation to the forecast error covariances. This corresponds to "the NMC method" (Parrish and Derber, 1992; Rabier et al., 1998a). The theoretical justification for this approach is rather weak. Another method could be to use the time-averaged covariances of an Extended Kalman Filter (EKF) in research mode to specify the static error covariances (Bouttier, 1996), although this is much more costly. Standard deviations of the forecast differences are usually rescaled to match the amplitude of +6 h forecast errors (Parrish and Derber, 1992, Rabier et al., 1998).

Different Kalmanfilterbased techniqes have been used to obtain flow dependent background error covariances, such as the Reduced Rank Kalman filter (Fisher and Courtier, 1995), and the Ensemble Kalman Filter (Houtekamer et al., 1996) . Within this report a methododology for for obtaining flow dependent forecast error covariances based on ideas of Dee (1995, 1999) is implemented into the 3-dimensional variational assimilation (3D-Var) scheme (Gustafsson et al. 2001; Lindskog et al. 2001) developed for the HIRLAM system (Undén et. al, 1998). The methodology is based on Maximun Likelihood theory and on line estimation of forecast errors from innovation vectors. The functionality of the approach is demonstrated in a two week assimilation and forecast experiment.

This report is organized as follows: In Section 2 the basics of the HIRLAM 3D-Var is outlined, followed by an extended description of the methodology for on-line estimation of background errors in section 3. In section 4 the assimilation and forecast experiments and their corresponding results are presented and finally section 5 contains concluding remarks.

2 Hirlam 3D-Var

The Hirlam 3D-Var (Gustafsson et al. 2001; Lindskog et al. 2001) follows the incremental formulation (Courtier et al. 1994) and the assimilation consists of minimizing the cost function

$$J = J_b + J_o = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(H\mathbf{x}^\mathbf{b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(H\mathbf{x}^\mathbf{b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y}). \quad (1)$$

Here J_b measures the distance to a background model state $\mathbf{x}^\mathbf{b}$ and J_o measures the distance to the observations $\mathbf{y}$. The non-linear observation operator H and the tangent-linear observation operator $\mathbf{H}$ transform the background state and assimilation increments, respectively, into the observed quantities. $\mathbf{B}$ and $\mathbf{R}$ are the error covariance matrices for background and observations, respectively. Furthermore, transpose is denoted by T. For HIRLAM the model state increment vector $\delta\mathbf{x}$ includes the horizontal wind components, temperature, specific humidity and the logarithm of surface pressure.

Background error statistics are derived from a sample of differences between forecasts valid at the same time (Parrish and Derber 1992; Rabier et al., 1998) with a non-separable approach, i.e. the vertical variability of horizontal correlations and the dependence of vertical correlations on horizontal scale are represented.

The dynamical couplings embedded in the background error constraint are based on hydrostatic and geostrophic assumptions in order to calculate geostrophic wind increments from temperature and surface pressure increments. Ageostrophic wind is taken into account and it is assumed to be decorrelated with both mass and geostrophic wind. This also implies an assumption of decorrelation between vorticity and divergence, as the geostrophic wind is essentially rotational and spatial covariances of the ageostrophic wind components are modelled in an isotropic way. The humidity is treated in a univariate way, expect for a weak coupling with wind via the hydrostatic equation (which involves *virtual* temperature). A more comprehensive description of the general formulation of HIRLAM 3D-Var and its background error constraint is given in Gustafsson et al. 2001).

The background error standard deviations in the reference HIRLAM 3D-Var version are dependent on type of variable and vertical level, in addition to a seasonal dependence. In the upgraded version described in this report a synoptical dependence has also been incorporated.

The observation operator H and the tangent-linear observation operator $\mathbf{H}$ are sub-divided into a sequence of sub-operators. The non-linear observation operator may be formally written in grid point space

$$H \;=\; H_{spec} I_v P_{calc} I_h, \tag{2}$$

where I_h denotes horizontal interpolation of model data from grid points to the horizontal positions of the observations, P_{calc} calculation of pressure and geopotential at hybrid coordinate model levels, I_v vertical interpolation to the levels of the observed data values and H_{spec} any other specific operators for each type of observation. An expression corresponding to (2) exists also for the tangent-linear observation operator. In HIRLAM 3D-Var, observation operator is presently available for conventional *in-situ* data, TOVS (TIROS Operational Vertical Sounder) and ATOVS (Advanced TOVS) radiances, GPS (Global Positioning System) atmospheric zenith delays, GPS occultations, scatterometers, Doppler radar wind data and MODIS winds and humidties. Bias correction is applied to ATOVS data. The observation errors are assumed to be uncorrelated for all observation types (except for inter-correlated ATOVS channels (Schyberg et al. 2003)). With this assumption, the covariance matrix $\mathbf{R}$ for the observation errors is diagonal and only the observation error standard deviations (σ_o) need to be specified (Lindskog et al. 2001).

3 On-line estimation

3.1 Tuning error covariance matrix

¿From an idea suggested by Dee (see....) to estimate error covariance parameters for statistical data assimilation schemes, we applied this approach to sequential data assimilation system which involves an explicit ca lculation of forecast error covariance, like variational data assimilation scheme (3DVAR). The idea is to force the covariances produced by the data assimilation scheme to be compatible with th e actual data being assimilated, based on the following relationship:

$$E[v_k v_k^T] \approx H_k P_k^b H_k^T + R_k \tag{3}$$

where the left-hand of the equation represents the covariance matrix of the innovation v ($Hx_k - z$ the background state projected in observation space by operator H minus observation), while on the right-hand side P_k^b is the background error covariance produced by the assimilation scheme, R is the observation error covariance and k is the time step of analysis. For a correct approach of the problem we investigate the behaviour of covariances treating separately obsversations associated first to the entire domain and secondly for regional subdomain. First we introduced a parameter α to modelling the covariances in the following way

$$E[v_k v_k^T] = \alpha(H_k P_k^b H_k^T + R_k) \tag{4}$$

and we calculated the parameter as follows

$$\alpha = \frac{norm(E[v_k v_k^T])}{norm(H_k P_k^b H_k^T + R_k)} \tag{5}$$

where *norm* is is computed adding all the matrix elements. Secondly the estimation of the covariance of the innovation was involved trought the kriging technique, modelling the innovation's variance to an exponential semivariogram, function of distance between observation points and estimating the sill, the range and the nugget that characterize the function. The estimation of the three parameters is based on the maximum likelihood estimation tecnique. The all passive α parameter estimation will involve a feedback into variational scheme of HIRLAM model (HIRVDA), i.e. the 3DVAR covariance model will be updated on the basis of the parameter estimates and it will investigate the effect on analysis accuracy that this may have.

3.2 Kriging approach

The kriging technique is a geostatical method to estimate the value ($\hat{v}$) and variance ($\hat{\sigma}$) of a spatial variable at a point (x, y) where the true value is unknown. In our problem the variable's variance to estimate is the innovation $v = z - Hx$ defined in data assimilation problem, where z is the observation value, H is a transition matrix (represented linear model) and x the model state variable. This approach is applied for each time step analysis and for that in definitions and equations $_k$ index is omitted. The algorithm is based to the following assumptions:

- the estimated value ($\hat{v}$) is expressed as

$$\hat{v} = v^T \lambda^* \tag{6}$$

  where v is the innovation vector and λ the vector of interpolating weight;

- v is an intrinsic random function whose system of ordinary Kriging equations can be written as

$$\gamma = \Gamma\lambda \tag{7}$$

  where Γ is a $[n+1, n+1]$ matrix defined as

$$\Gamma = \begin{pmatrix} \Gamma^* & \vdots & u \\ \dots & \dots & \dots \\ u & \vdots & 0 \end{pmatrix} \tag{8}$$

  where Γ^* is the [n,n] simmetric matrix of variogram values $\gamma_{i,j}$, describing the spatial dependence among the n measurements points, u is the unit vector and λ is a vector defined as:

$$\lambda = \begin{pmatrix} \lambda^* \\ \dots \\ \mu \end{pmatrix} \tag{9}$$

  with μ a lagrangian multiplier.

- the experimental semi-variogram (in classical way) for each lag distance is

$$var_{i,j} = 0.5(v_i - v_j)^2 \tag{10}$$

  and it is assumed isotropic.

The improvement that we set is to estimate on-line (for each time step analysis) parameters of the variogram γ, defined as exponent ial form

$$\gamma_{i,j} = \begin{cases} p + \omega(1 - e^{(-h_{i,j}/d)}) & i \neq j \\ 0 & i = j \end{cases} \tag{11}$$

where $h_{i,j}$ is distance from each couple of observations. With tecnique of maximum likelihood estimation we estimate "on line" the following parameters: d the range, p the nugget and ω the sill, to modelling in sense of best fit the experimental semi-variogram. The elements of thecovariogram matrix V are defined as following

$$V_{i,j} = \begin{cases} p + \omega e^{(-h_{i,j}/d)} & i = j \\ \omega e^{(-h_{i,j}/d)} & i \neq j \end{cases} \tag{12}$$

and represents the innovation covariance defined as $E[v, v^T]$.

3.3 Maximum likelihood estimation

Even if the observation variable z is not a Gaussian random field, we assume that x and v are indipendet and gaussian distributed: $x \sim \mathcal{N}(\mu, P)$ and $v \sim \mathcal{N}(0, V)$. The joint probability density function of the n innovations can be expressed as:

$$f(v, \theta) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} exp[-\frac{1}{2} v^T V^{-1} v] \tag{13}$$

where v is the vector of innovation and θ denotes the vector of semivariogram parameters (namely d, p and ω). The vector of parameters θ will be determined through maximising the likelihood function

$$L(\theta, v) = f(v, \theta) \tag{14}$$

or by minimizing the negative log-likelihood function defined as:

$$\mathcal{L}(\theta, v) = -lnL(\theta, v) \tag{15}$$

3.4 Application

The data set for winter period ($1 - 14$ January 2002) consist on analysis data (innovation vectors, observation error variances, background error covariance matrices in observation space) for the entire HIRLAM domain. Firstly we devided the data set for each type (SYNOP, TEMP, AIREP) and sub-type (LAND and SHIP for SYNOP) for each level domain. Then we had to trasform the lat long coordinates of the observations in kilometers (North Polar Stereograghic coordinates) because we have to know the scale of the problem to set the initial parameters of the algorithm.

As we show in figure we choosed these observations because they are well spatially distribuited and numerically significant. Provided the data set, we

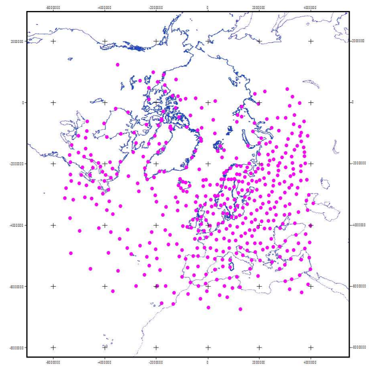

Figure 1: Spatial distribution of geopotential observations on surface level.

tested the algorithm for geopotential SYNOP and DRIBU observations on land surface. Then we run program to estimate variogram parameters that implement the maximum likelihood estimation. After that we calculated the α factor of correction to be applied in data assimilation process for geopotential observations. Figures belowed show some results.

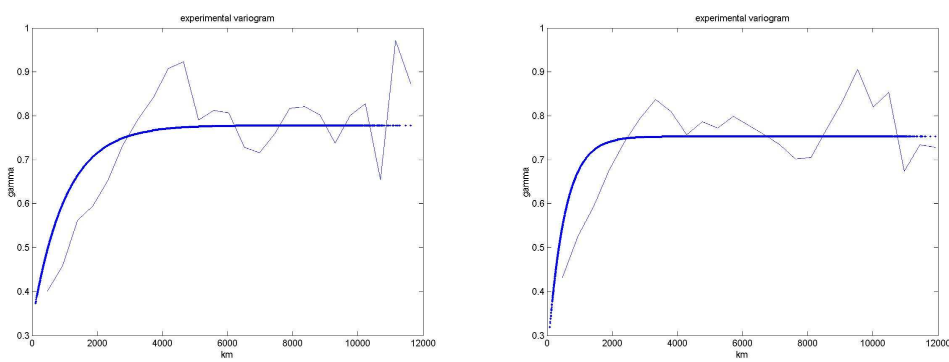

Figure 2: Example of esperimental variogram and estimate variogram from two different time step analysis.

The α factors calculated for geopotential innovations from SYNOP and SHIP observations and for one single domain containing the HIRLAM model integration area were used to calculate a time-dependent global scaling factor $hifac$. The scale factor for assimilation cycle i, within the two week period, is calculated with the following expression:

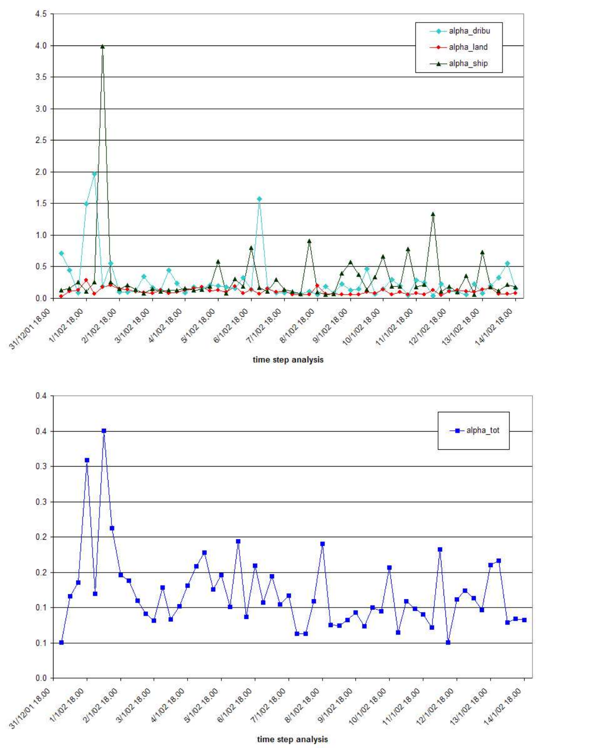

Figure 3: α parameter for land, ship and dribu for geopotential observations (rigth), α parameter for entire domain (left).

$$hifac(i) = 0.75\frac{alpha_{SYNOP}(i)}{\sum alpha_{SYNOP}(i)} + 0.25\frac{alpha_{SHIP}(i)}{\sum alpha_{SHIP}(i)} \qquad (16)$$

where the summation is over all assimilation cycles during the two week period (4 assimilation cycles per day) Finally to filer variations that were considered to be of too high frequency a 1.25 day time filter was applied to the scalefactors, as given by (x). The resulting time series of scalefactors (time-filtered $hifac$) is illustrated in Figure (ml1). The surface geopotential at a position and its variations is a result of the status in the entire air column above. Therefore it was consider an approperiate first approximation to apply the scaling factor not only to surface pressure, but also to the upper air temperatures and winds.

4 Assimilation and Forecast Experiments

4.1 Experimental design

A two week parallel assimilation and forecast experiment, extending from 1 to 14 January, 2002, has been performed over an area covering Northern Europe and the Northern Atlantic. Synoptically the period is characterised by several cyclones passing over the Northern Atlantic and through the Baltic Sea area. The model domain is illustrated in Figure ml2.

The two parallel data assimilation experiments are characterized as follows:

- The repressentation of background error standard deviations in the data assimilation is constant in time and in accordance with the values assumed for January in the HIRLAM 3D-Var reference.
- The repressentation of background error standard deviations in the data assimilation has a synoptical variation through the application of a global scaling factor based on on line estimation.

For both of the configurations as described above, data was assimilated in a 6 h assimilation cycle. The observations were retrieved from the ECMWF (European Center for Medium Range Forecasts) archive. After each analysis, a digital filter initialisation was applied (Lynch and Huang, 1992), followed by a 48 h forecast. For the lateral boundary conditions, 3-hourly ECMWF analyses were used.

The experiments were run with the HIRLAM reference system (version 6.2.0) on a SGI Origin 3800 parallel computer using 16 processors; 202×178

horizontal grid points with a resolution of 0.4° and 40 vertical levels were used for these runs. The HIRLAM reference model physics uses turbulence scheme based on turbulent kinetic energy (Cuxart et al., 2000). For the clouds and condensation parameterisation, the STRACO scheme is applied (Sass et al., 1999). The main features of this scheme include sub-grid scale condensation with a statistical distribution of cloud condensate, a smooth transition between stratiform and convective regimes and microphysics of condensation according to Sundqvist (1993). The radiation scheme is based on ideas of Savijärvi (1989). Semi-imlicit time integration and a fourth order implicit horizontal diffusion scheme were used in all model integrations.

To evaluate the relative quality of the analyses and subsequent forecasts of the different parallel experiments, we verified them against observations in the list of (radiosonde and SYNOP) stablished by the European Working Group on Limited Area Models (EWGLAM). The verification was done for weather parameters, at the surface level, and at the vertical levels of 850, 500 and 300 hPa. The model data used in the statistics were the analyses and the 6, 12, 18, 24, 30, 36, 42 and 48 h forecasts.

4.2 Results

Bias and Root Mean Square (RMS) scores as function of forecast length indicate that there is a neutral or sligtly positive impact of the on-line forecast error estimation on forecast quality. This is illustrated for temperature and wind at the 300, 500 and 850 hPa levels in Figure ml3. As can be seen the slight positive impact can be seen specially at longer forecast ranges.

One should keep in mind that there is a large time variability in the forecast quality. This is true both for the absolute quality and also the relative quality of the control forecasts and the forecasts based on on-line estimation. This is illustrated in Figure ml4, which shows the RMS sores for verification of 48 h Mean Sea Level Pressure forecasts, as function of assimilation cycle within the two week period. It can be seen that for sometimes the control run forecast show better scores and sometimes forecasts based on one-line estimation appears better. On the average, over the 14 day period, the 48 h MSLP scores scores are however rather neutral.

5 Concluding Remarks

A methodology for on-line estimation of the synoptical variation of forecast errors has been developed and incorporated into the HIRLAM NWP system. The methodology include elements of maximum likelihood theory and Krieging. To demonstrate the functionality of the system for on-line estimation a two week parallel assimilation and forecasting experiment has been carried out with the HIRLAM NWP system. Results for verification against observations indicate a netral to slightly positive impact of the new developments on HIRLAM forecasts. The impact is higly variable in time so that more extended experiment are needed to confirm the promising results obtained here.

In addtion there are room for further refinements of the methodology for estimation of the on-line estimation of the forecast errors. These iprovements include repressenting also the spatial variation of of the forecast errors for each synotical situation. Furthermore, by including dynamical balance constraints between wind, temerature and geopotential it would be possible to have an improved estimate forecast errors.

6 References

Courtier, P, Thepaut J-N and A Hollingsworth, 1994: A Strategy for operational Implementation of 4D-VAR, using an incremental approach. *Q.J.R. Meteorol. Soc.*, **120**, 1367–1387.

Cuxart, J., Bougeault, P., and Redelsperger, J.-L., A turbulence scheme allowing for mesoscale and large eddy simulations., *Q. J. R. Meteorol. Soc.*, *126*, 1-31, 2000.

Dee, D.P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev., 123, 1128-1145

Dee, D.P., and A.M. da Silva, 1999: Maximum likehood estimation of forecast and observation error covariance parameters. Part I: Methodology. Mon. Wea. Rev., 124,1822-1834

Derber, J. and Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195-221.

Fisher, M. and Courtier, P., 1995: Estimating the covariance matrix of analysis and forecast errors in variational data assimilation. ECMWF Res. Dep. Technical Memorandum, No. 220.

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

Hollingsworth, A. and Lönnberg, P., 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.

Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H.L., 1996: A System Simulation Approach to Ensemble Prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Lindskog, M, Gustafsson, N, Navascués, B, Mogensen, K S, Huang, X-Y, Yang, X, Andræ, U, Berre, L, Thorsteinsson, S and Rantakokko, J, 2001: Three-dimensional variational data assimilation for a limited area model. Part II: Observation handling and assimilation experiments. *Tellus*, **53A**, 447–468.

Lorenc, A., 1997: Development of an operational variational assimilation scheme. *J. Meteorol. Soc. Japan*, **75**, 339-346.

Lynch, P. and Huang, X.-Y., 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019-1034.

Parrish, D.F. and Derber, J.C., 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. Mon. Wea. Rev., 120, 1747-1763.

Rabier, F. Mc Nally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A., Bouttier, F., 1998a: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part II: Structure functions. *Q. J. Roy. Meteorol. Soc.*, **124**, 1809-1830.

Sass, B.H, Nielsen, N.W., Jørgensen, J.U. and Amstrup, B., 1999: The Operational HIRLAM System at DMI. *DMI Tech. Rep.* no. **99-21**, 42 pp., October 1999. Available from DMI, Lyngbyvej 100, Copenhagen.

Savijärvi, H., 1989: Fast Radiation Parameterization Schemes for Mesoscale and Short-Range Forecast Models. *J. Appl. Meteor.*, **29**, 437–447.

Schyberg, H., T. Landelius, S. Thorsteinsson, F. T. Tveter, O. Vignes, B. Amstrup, N. Gustafsson, H. Järvinen, and M. Lindskog, 2003: Assimilation of ATOVS data in HIRLAM 3D-VAR System. HIRLAM Tech. Rep., 60, 69 pp.

Sundqvist, H., 1993: Inclusion of ice phase of hydrometers in cloud parameterization for mesoscale and large scale models. *Beitr. Phys. Atmosph.* , **66**, 137-147.

Undén, P., Rontu, L, Järvinen, H., Lynch, P., Calvo, J., Cats, G, Cuxart, J., Eerola, K.,Fortelius, C., Garcia-Moya, J. A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascues, B., Woetman Nielsen, B., Ødegaard, V., Rodriguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Hansen Sass, B., Savijärvi, H., Wichers Schreur, B., Sigg, R., The, H., and Tijm, S., 2002: HIRLAM-5 Scientific Documentation, HIRLAM-5 Project, c/o Per Undén SMHI, SE-601 76 Norrköping, Sweden.

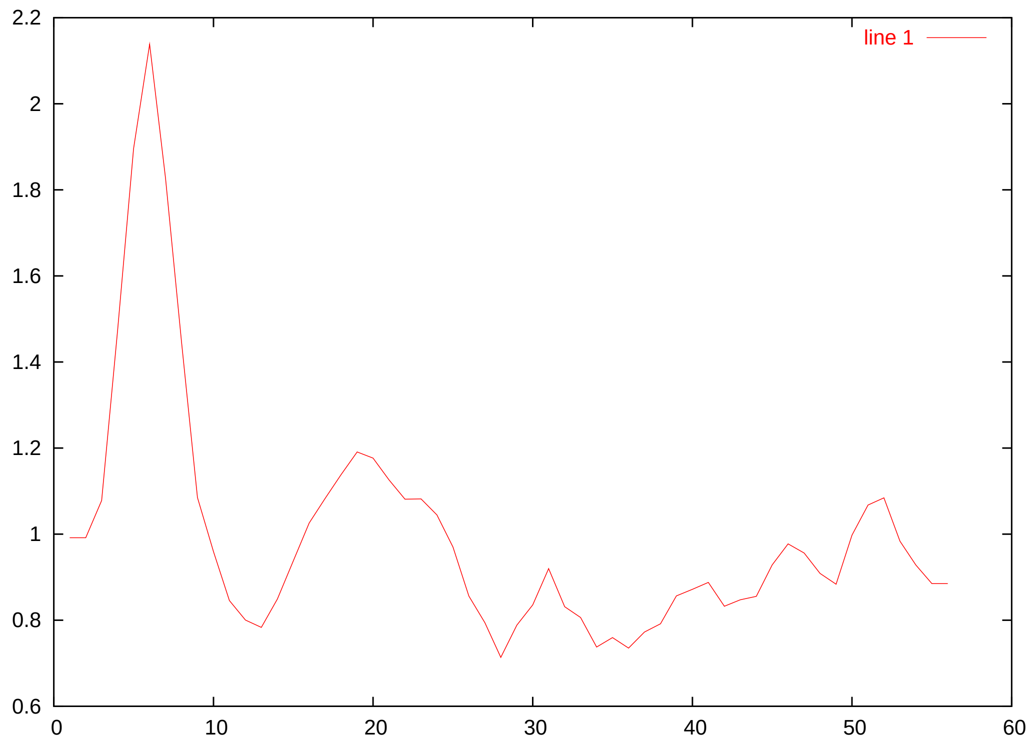

Figure 4: Time Variation of scaling factor.

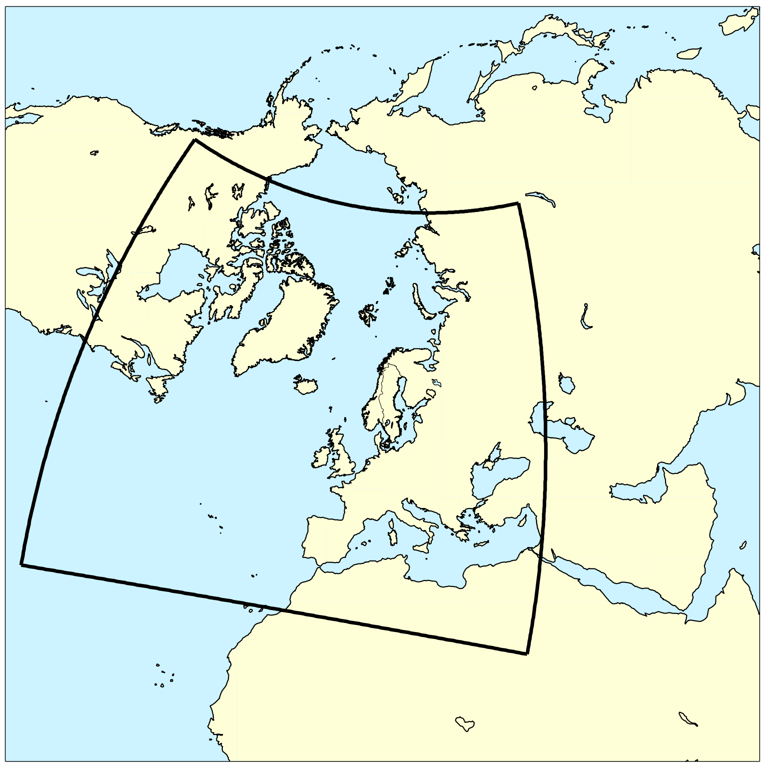

Figure 5: Model integration area (indicated by frame) for parallel assimilation and forecast experiment with the HIRLAM system.

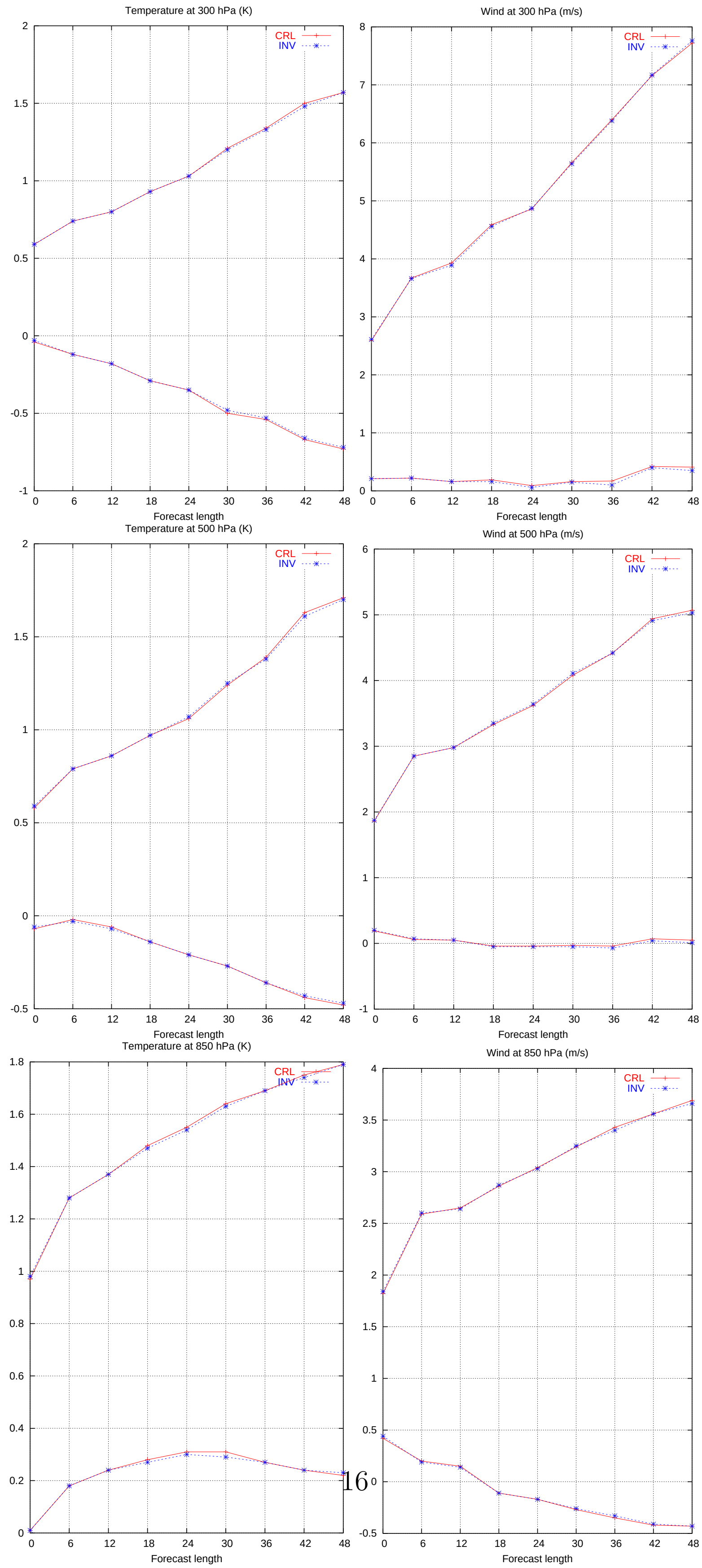

Figure 6: Bias and RMS scores for verification of control (red with ,,) and on-line estimation based forecasts against observations as functions of forecast length in hours. Temperatures (left, unit: K) and wind speeds (right, unit

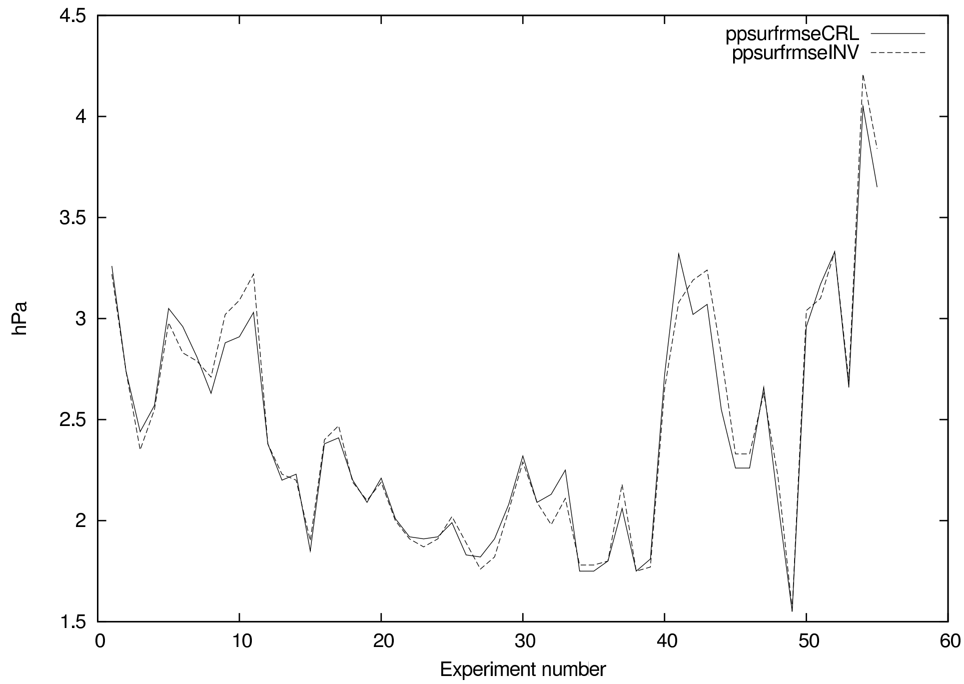

Figure 7: RMS scores for 48 h Mean Sea Level Pressure forecasts againt observations as function of assimilation cycle within the the period 20020101 00 UTC- 20020101 18 UTC (unit hPa). The scores are for control (red with ,,) and on-line estimation based forecasts.