

A Feasibility Study of Assimilating European Wind Profiler Data Using the HIRLAM 3D-VAR System

Xiang-Yu Huang^{1*} and Magnus Lindskog²

¹Danish Meteorological Institute, Denmark

²Swedish Meteorological and Hydrological Institute, Sweden

August 13, 2003

Abstract

The HIRLAM 3-dimensional variational data assimilation system has been prepared to handle the data from the European wind profiler (EWP) network. A number of data assimilation experiments have been performed.

In the selected period the amount of the EWP data is rather minor compared with the total data volume. However, if we only take wind profiles into account the EWP data are significant, and on asynoptic hours (0300, 0900, 1500, 2100 UTC) even the major data source.

High data rejection ratios have been found for a few stations, indicating possible problems with the data. Large differences between the observations and the background have been found for a few stations, also indicating a need for bias correction or blacklisting.

Data assimilation experiments are performed for February 2002 without any bias correction or blacklisting. The experiments with 3 h assimilation cycles have shown a marginal positive impact from the EWP data on the analyses and forecasts over the whole month.

* *Corresponding address:* Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, DENMARK. Email: xyh@dmi.dk

Contents

1	Introduction	3
2	European wind profiler data assimilation	4
2.1	HIRLAM variational data assimilation	4
2.2	Processing of the EWP data	5
3	Experiment configuration	7
4	Results	11
4.1	Data usage	11
4.2	Data quality	11
4.3	Data amount relative to other observation types	15
4.4	Data impact on analyses and forecasts	18
5	Conclusions	21
6	Acknowledgements	24
	References	24

1 Introduction

Currently, the wind profiles used by the High Resolution Limited Area Model (HIRLAM) 3-Dimensional VARIational data assimilation (3D-VAR) system are primarily from radiosondes and PILOT balloons. They are mainly observed and reported on the synoptic hours, 0000, 0600, 1200, 1800 UTC. For operational systems using 3 h data assimilation cycles, there is almost no wind profile information for cycles on the asynoptic hours, *i.e.* 0300, 0900, 1500, 2100 UTC.

On the other hand, wind profile observations are being made more or less continuously in time by wind profilers (except for small time lags that may exist due to filtering and quality control). Wind profilers probe the atmosphere and derive the wind profiles from the Doppler shift of the echoes of the transmitted radio waves due to clear air turbulence. The American wind profiler data have been assimilated by HIRLAM 3D-VAR for some years, but for HIRLAM applications they only have influence on a small part of the domain if the domain is chosen to be large enough to reach the United States.

There is also a wind profiler network in Europe. In recent years, the European Wind Profiler (EWP) data have been available in real time. At European Centre for Medium-Range Weather Forecasts (ECMWF), the EWP data have been monitored and shown to have a positive impact on the 4-Dimensional VARIational data assimilation (4D-VAR) system. They have been used in operation since April 2002 (Andersson and Garcia-Mendez, 2002).

The wind profile observations become more important as the resolution of the data assimilation system increases. For HIRLAM applications, the assimilation of the EWP data should be prepared and investigated. All operational implementations of the HIRLAM variational data assimilation system use the 3-dimensional version (Gustafsson *et al.*, 2001; Lindskog *et al.*, 2001). Therefore the EWP data impact on HIRLAM 3D-VAR is assessed here, although the full potential of the EWP data may first be realized when using 4D-VAR.

In this report the HIRLAM 3D-VAR system is briefly described, followed by some details of how the EWP data are handled (Section 2), the experiment configuration is given (Section 3) and preliminary results from data assimilation experiments are discussed (Section 4). Finally, we also attempt to draw some conclusions and to make some suggestions

for further studies (Section 5).

2 European wind profiler data assimilation

2.1 HIRLAM variational data assimilation

The HIRLAM 3D-Var has an incremental formulation and the assimilation consists of finding the model state assimilation increment vector, $\delta\mathbf{x}$, that minimizes the following cost function:

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

Here J_b measures the distance to a background model state \mathbf{x}^b , which is a short range forecast, and J_o measures the distance to the vector \mathbf{y} of the observations. For HIRLAM the model state increment vector $\delta\mathbf{x}$, includes the horizontal wind components, temperature, specific humidity and the logarithm of surface pressure. 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} is the matrix containing the covariances of the background field errors, while \mathbf{R} is a matrix containing the covariances of the errors in the observations.

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

The observation operator H and the tangent-linear observation operator \mathbf{H} are subdivided into a sequence of sub-operators. For the non-linear observation operator we may formally write, in gridpoint space

$$H = H_{spec}I_vP_{calc}I_h, \quad (2)$$

where I_h denotes horizontal interpolation of model data from grid points to the horizontal positions of the observations, P_{calc} calculation of pressures and geopotentials at model full and half 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. At present observation operators have been developed for various conventional data, as well as for data from TOVS (TIROS

Operational Vertical Sounder) and ATOVS (Advanced TOVS) radiances, GPS (Global Positioning System) atmospheric delays, GPS occultations, scatterometers, Doppler radar VAD (Velocity Azimuth Display technique) wind profiles and Doppler radar radial winds. The observation errors are assumed to be uncorrelated for all observation types [except for a possibility to account for correlations between radiance errors in different ATOVS channels (Schyberg *et al.*, 2003)]. With this assumption, the covariance matrix \mathbf{R} for the observation errors becomes a diagonal matrix and only the observation error standard deviations (σ_o) need to be specified (Lindskog *et al.*, 2001). The σ_o values used here are based on extensive observation system experiments at ECMWF and the Swedish Meteorological and Hydrological Institute (SMHI).

A standard minimization software package (Gilbert and Lemaréchal, 1989), based on an iterative technique, is used to find the minimum of J . For each iteration, the gradient of the cost function with respect to the model state increment vector $\delta\mathbf{x}$ is calculated by

$$\nabla_{\delta\mathbf{x}}J = \mathbf{B}^{-1}\delta\mathbf{x} + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^b + \mathbf{H}\delta\mathbf{x} - \mathbf{y}). \quad (3)$$

Here, \mathbf{H}^T is the adjoint, or transpose, observation operator.

2.2 Processing of the EWP data

In the observation handling system associated with the HIRLAM variational data assimilation the EWP data are handled as PILOT balloon observations.¹ A detailed description of the observation handling, with emphasis on conventional types of observations, can be found in Lindskog *et al.* (2001). Here we will focus on the most important quality control checks applied on the EWP data.

First, in the case of multiple EWP reports from the same station, the one closest to analysis time is chosen.

In the background quality control the *a priori* information from the background state is utilized. The two components (u and v) of each individual EWP wind observation, y_i , are jointly checked against their background values, $[[H(\mathbf{x}^b)]_i]_u$ and $[[H(\mathbf{x}^b)]_i]_v$. An observation is passed to the minimization if the inequality

¹The necessary development for HIRLAM 3D-VAR, which separates EWP data from PILOT data and applies different screening and blacklisting to EWP data, has been made.

$$\frac{1}{2} \left(\left[\frac{([H(\mathbf{x}^b)]_i - y_i)^2}{\sigma_{b,i}^2 + \sigma_{o,i}^2} \right]_u + \left[\frac{([H(\mathbf{x}^b)]_i - y_i)^2}{\sigma_{b,i}^2 + \sigma_{o,i}^2} \right]_v \right) \leq L, \quad (4)$$

is fulfilled, where L is the upper rejection limit and $\sigma_{b,i}$ and $\sigma_{o,i}$ denote the background error standard deviation and the observation error standard deviation, respectively, of observation i . In the case of EWP the value 8 is used for L . For the values of $\sigma_{b,i}$ and $\sigma_{o,i}$ we refer to Table 1.

As a rough first estimate the $\sigma_{o,i}$ for EWP data are the same as used for PILOT balloon reports and, as mentioned in Section 2.1, the values are based on extended experiments at ECMWF and SMHI. In the future the EWP $\sigma_{o,i}$ values need to be revised, utilizing for example innovation statistics from HIRLAM experiments. The $\sigma_{b,i}$ values are the yearly mean horizontally averaged values, obtained with the NMC-method (Parrish and Derber, 1992). Spatially varying background error standard deviations are used to represent the differences in background error statistics resulting from different flow characteristics in different geographical areas, as well as from variations in the station density. Since the horizontal variation (see HIRLAM Newsletter 42) is not fully evaluated we have chosen not to use it in this study. The seasonal variation has been taken into account in this study as described by Lindskog (2000). This slightly reduces the wind $\sigma_{b,i}$'s in summer and increases them in winter, due to a stronger dynamical activity during winter time.

For EWP observations above 700 hPa and for wind speeds greater than 15 m/s, a wind direction check is applied. If more than four consecutive levels of an EWP profile are flagged as suspicious, the whole profile in concern is rejected.

Finally, the minimization itself contains a so-called variational quality control (Lindskog *et al.*, 2001), that accounts for non-Gaussian observation errors in EWP data, as well as in other data. The variational quality control (VarQC) for wind data involves two parameters representing the range of possible values and the probability of gross errors. For EWP, we have initially used the parameter values derived for PILOT balloon reports.

Table 1: Pilot error standard deviation as a function of pressure levels. Unit: m/s.

level	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
σ_o	2.3	2.3	2.5	3.0	3.5	3.7	3.5	3.5	3.4	3.3	3.2	3.2	3.3	3.6	4.5
σ_b	4.3	3.8	3.5	3.5	4.6	5.2	5.1	4.6	4.0	3.5	3.4	3.6	4.0	5.0	7.8

These values are based on investigation of statistics files from the operational OI analysis at the Spanish Meteorological Institute (Gustafsson *et al.*, 1999).

From a practical point of view the introduction of EWP data into the assimilation system involves modifications in MAKECMA and HIRVDA. MAKECMA is the software that converts observation reports from BUFR (Binary Unit For Representation of meteorological data) format into CMA (Central Memory Array) format. It also carries out a number of relatively simple quality control checks and variable conversions. The major changes applied to MAKECMA are the introduction of a BUFR-template defining the expected BUFR format for EWP data and a subroutine converting this data type into CMA format. These components come from the ECMWF observation handling system. In addition the new data type has to be defined in a number of common blocks and namelists. It is defined as a sub-type of PILOT balloon reports, following the procedure for American wind profilers. The modifications in HIRVDA included only small extensions of common blocks and parameters in the HIRVDA attached CMA code, in order to enable recognition of the new data type in HIRVDA.

3 Experiment configuration

The preferred approach to assess impact of a new data type on data assimilation systems consists of 4 steps: 1) develop the necessary observation operator; 2) include the new data in the so-called passive mode, *i.e.*, process the data as all other data but give zero weight to innovations due to the new data in the 3D-VAR minimization iterations; 3) tune observation error statistics based on the passive runs, choose blacklisting and screening strategies; 4) run full data assimilation experiments.

In assessing the EWP data impact we have not gone through steps 2 and 3. ² The EWP data are treated as PILOT balloon data and assumed to have the same error statistics. As will be shown later, the error statistics from the experiments over a month already indicate some differences between PILOT balloon and EWP, and a few problematic profilers. Experiments with properly tuned error statistics, blacklisting and screening should be carried out in the future. However, as it is not possible to perform more experiments using the same experiment configuration on the ECMWF VPP5000 supercomputer, due to its scheduled switching off by the end of March 2003, we now report our preliminary

²Recently DMI has started the routine monitoring of EWP data.

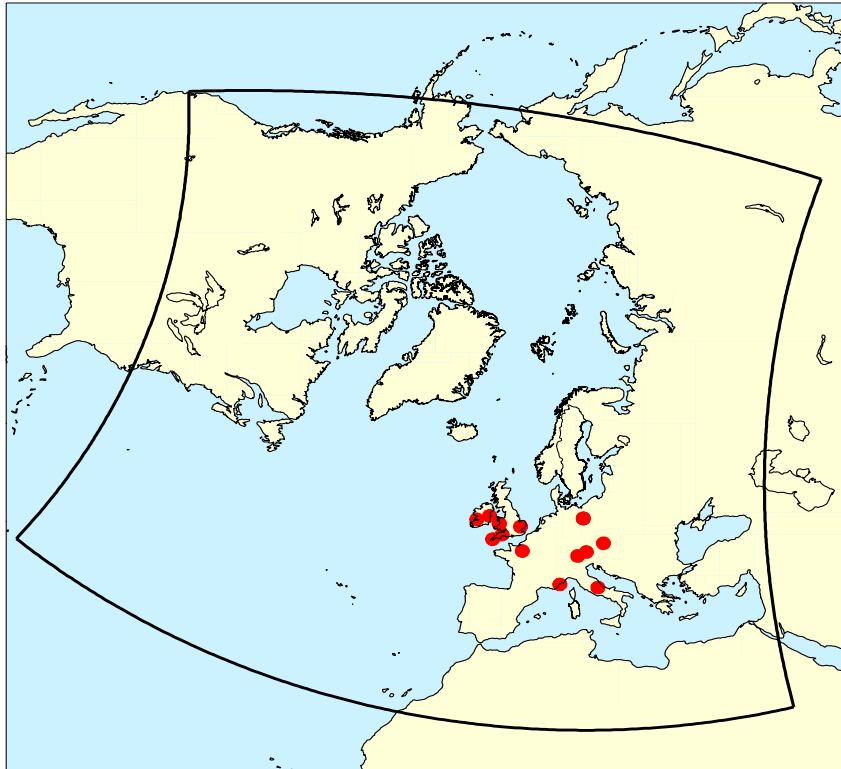


Figure 1: The model domain and the wind profiler locations, from which data are used in this study.

results based on the experiment configuration described below.

The model domain, chosen for the parallel experiments, is shown in Figure 1. The model has 31 levels, 202×190 points at each level, a 0.45 degrees horizontal resolution and on a rotated grid with polar coordinates $(0^\circ, 80^\circ)$ starting at $(-63.725^\circ, -37.527^\circ)$ in the rotated coordinate system.

This model configuration³ was chosen mainly for practical reasons, as it has been used on the ECMWF VPP5000 for more than a year for many other experiments (FGAT, 4D-VAR, etc.). Its computational requirements are reasonable and it can use ECMWF analyses and forecasts as lateral boundaries directly. It would be better to select a higher resolution configuration to test impact of wind observations as most operational HIRLAM systems have resolutions better than 0.45 degree.

The analysis scheme is the HIRLAM 3D-VAR using the increment on the full resolution. Parameters which are not explicitly mentioned in this report are set as the default (Gustafsson *et al.*, 2001; Lindskog *et al.*, 2001).

³The configuration is similar to the operational one in the Danish Meteorological Institute (DMI) operational G-model (DMI-HIRLAM-G).

At the start of the experiments, the very first background is a short range forecast from ECMWF analysis interpolated to the HIRLAM grid. In later assimilation cycles, the background is a short range forecast from the analysis in the previous cycle. The ECMWF analyses are used as lateral boundary conditions.

The data assimilation interval is 3 h or 6 h. Almost all conventional data are assimilated, including synoptic observations (SYNOP), ship observations (SHIP), buoys (BUOY), pilot balloons (PILOT), radiosondes (TEMP) and aircraft reports (AIREP⁴). We extracted these data from the ECMWF MARS archive. However, to have the error statistics consistent with data, the ATOVS AMSU-A radiance data, locally received at DMI, are used, following the current (2002) DMI pre-operational configuration (Amstrup, 2002).

The EWP data, which are the main concern of this study, are also extracted from the MARS archive. During February 2002, there are observation reports from 16 European wind profilers, which are listed in Table 2 together with their characteristics. The locations of these wind profilers are marked in Figure 1 by dots. It is obvious that the EWP network only covers a small part of the model domain, which is already covered by the radiosonde network, the PILOT balloon network and the surface synoptic observation network. With this in mind, we should not expect much from the additional EWP data.

The period chosen for the experiments is February 2002. It was just a normal winter month in which a few cyclones crossed the Atlantic. The weather situation over Europe was not very interesting that month. The period was chosen for the practical data availability reasons. To assess the impact of EWP data it could be better to select more intense weather systems with smaller scales.

Four experiments have been performed. The experiment names are listed in Table 3 together with some short descriptions. Most discussions will be concentrated on the two experiments with 3 h cycling, T3W and T3R.

The spectral HIRLAM (Gustafsson and McDonald, 1996) is used with HIRLAM 4.3 physics. It includes the turbulence scheme by Holtslag and Bovine (1993), the STRACO (Sass *et al.*, 1999) scheme for the clouds and condensation parameterisation, and the Savijärvi (1990) radiation scheme. The Eulerian semi-implicit time step is 4 min. A

⁴Aircraft Meteorological Data Relay (AMDAR) and Automatic Computerized Aircraft Reporting System (ACARS) are included.

Table 2: List of stations from which data are available from the ECMWF MARS archive. In the table, the vertical range are indicated by three types: boundary layer (BL), troposphere (TROP) and lower stratosphere (LST).

stn id	lat	lon	Station name	Country	Elevation (m)	Vertical range
03500	52.42	-4.00	Aberystwyth	UK	50	BL/TROP
03501	52.42	-4.00	Aberystwyth	UK	50	TROP/LST
03591	52.07	0.58	Wattisham	UK	87	BL/TROP
03807	50.13	-5.10	Camborne	UK	88	BL/TROP
03840	50.87	-3.23	Dunkeswell	UK	253	BL/TROP
03962	52.70	-8.93	Shannon	IRL	26	BL/TROP
03969	53.43	-6.24	Dublin	IRL	100	TROP
06601	-	-	Mobile system	CH	-	BL
07112	48.61	0.87	La Ferte Vidame	F	245	TROP/LST
07690	43.66	7.19	Nice	F	4	BL/TROP
10391	52.17	14.12	Lindenberg	D	70	BL
10394	52.21	14.13	Lindenberg	D	107	BL/TROP/LST
11036	48.10	16.60	Vienna	A	227	BL
11120	47.55	7.58	Innsbruck	A	593	TROP
11150	47.47	13.00	Salzburg	A	430	BL
16228	42.40	13.40	L'Aquila	IT	1000	TROP

Table 3: List of experiments.

Name	Short description
T3R	Control (reference) run with 3 h assimilation cycles and no EWP data.
T3W	Same as the control run but with EWP data.
T6R	Same as the control run but with 6 h assimilation cycles.
T6W	Same as T6R but with EWP data.

48 h forecast is produced from assimilation cycles at synoptic hours (0000, 0600, 1200, 1800 UTC). Comparisons are made between T3W and T3R using both subjective and objective evaluations. The former are based on case studies and the latter on verification of analyses and forecasts against radiosonde and surface observations using the EWGLAM list of stations (Hall, 1987).

4 Results

4.1 Data usage

The EWP observation usage and rejection statistics for the time period 2-28 of February 2002 is presented in Table 4. As an observation we refer to an observation at one level, not a whole profile. It can be seen that the EWPs provide a significant amount of data. However the data amount and quality are clearly station dependent. The rejection ratio is defined as the sum of the number of observations rejected in either the background check or the VarQC, divided by the total number of observations before the two types of checks. An observation is considered to be rejected in VarQC when it is marked with a flag larger than 1 (Lindskog *et al.*, 2001). It can be seen that for 8 out of 16 stations the rejection ratio exceeds 10%, and for 4 stations it even exceeds 20%. A high rejection ratio is an indication of poor data quality.

4.2 Data quality

The highly station dependent data quality indicated in Table 4 is confirmed by Table 5, which shows observation minus background (OmB) and observation minus analysis (OmA) bias and root-mean-square error (rms) values for all wind vector observations assimilated during the period 1-28 February 2002. The absolute values of the OmB bias vary from 0.08 to 3.06 (m/s). The absolute values of the OmA bias vary from 0.01 to 0.93 (m/s). The OmB rms values vary from 2.95 to 6.34 (m/s). The OmA rms values vary from 1.59 to 4.27 (m/s).

For some stations the high OmB rms value can be explained by a significant OmB bias. Figure 2 shows OmB rms and bias as function of height for two stations suffering from large biases. Evidently, the bias contributes to the large rms values significantly. The bias problem of the two stations is perhaps even more evident in Figure 3, in which

Table 4: Total number of observations before quality controls (second column), number of rejections by background check (third column), VarQC flags (columns 4 to 7) and rejection ratio (column 8) for different stations (first column). Data are for the time period 2-28 February 2002.

Station ID	Total number of observations	Background rejection	Observations with VarQC flag				Rejection ratio (%)
			1	2	3	4	
03500	4899	171	4727	0	0	1	3.5
03501	31460	6988	24398	25	21	28	22.4
03591	3105	28	3067	2	2	6	1.2
03807	5754	209	5523	7	2	13	4.0
03840	3411	143	3255	2	2	9	4.6
03962	986	127	808	7	0	44	18.1
03969	1791	271	1422	8	9	81	20.6
06601	5161	1327	3772	8	5	49	26.9
07112	5329	129	5168	3	4	25	3.0
07690	4014	452	3557	1	0	4	11.4
10391	4195	3	4184	4	0	4	0.3
10394	2860	154	2676	2	3	25	6.4
11036	5846	719	5005	11	12	99	14.4
11120	1373	279	1050	5	2	37	23.0
11150	4112	726	3314	3	3	66	13.7
16228	1542	94	1441	0	0	7	6.5

Table 5: Total number of observations used in the statistics (column 2), the OmB bias (column 3), the OmA bias (column 4), the OmB rms (column 5) and the OmA rms (column 6), for different stations (column 1). The statistics are for data from all heights and for all observations that passed the quality control checks. The unit is m/s.

Station ID	Total number of observations	Bias		RMS	
		OmB	OmA	OmB	OmA
03500	5325	-0.85	-0.16	2.95	1.59
03501	27330	-1.14	-0.16	4.29	2.65
03591	3524	0.82	0.30	3.04	1.95
03807	6059	0.74	0.21	3.33	1.86
03840	3609	0.11	-0.01	3.53	1.93
03962	1002	-0.38	-0.56	6.34	4.20
03969	1726	0.39	0.26	5.93	4.27
06601	4242	0.08	0.01	4.78	2.35
07112	5677	-0.78	-0.40	3.71	2.58
07690	3973	-0.80	-0.11	3.23	1.23
10391	4517	1.66	0.51	3.72	2.00
10394	2956	0.58	-0.13	3.49	1.97
11036	5877	1.22	0.32	4.66	2.46
11120	1517	-3.06	-0.93	5.69	3.02
11150	3932	1.01	0.18	4.57	1.95
16228	1470	-1.55	-0.32	4.38	1.91

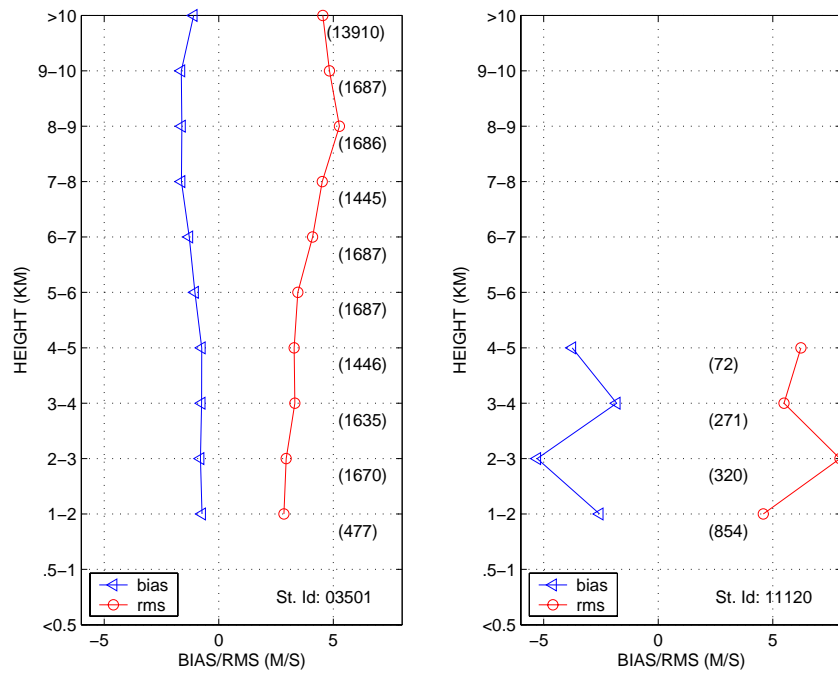


Figure 2: OmB vector background departure bias and rms for stations 03501 (left) and 11120 (right), as function of height. Unit: m/s. The numbers inside parenthesis represent the number of observations, at each level, on which the statistics are based.

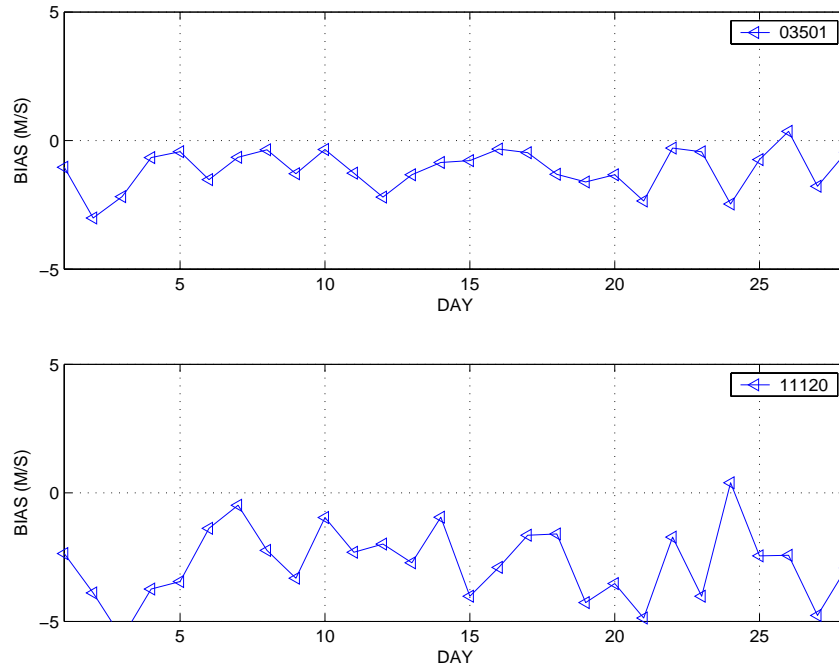


Figure 3: Vertically averaged OmB bias as a function of day within the period of the assimilation experiment (for 1-28 February 2002). The time series are for stations 03501 (upper) and 11120 (lower). Unit: m/s.

the time series of the OmB biases for the two stations are displayed as function of time for the one month period of the assimilation experiment.

For some other stations, like the ones illustrated in Figure 4, the OmB rms is high, although the bias is low.

Most stations provide data of high quality. An example of high quality data is shown in Figure 5. It can be seen that the data are relatively unbiased with respect to the forecast model and that the OmB rms is relatively small. Furthermore, it can be seen that the assimilation draws reasonably much for the EWP data. The OmA rms is much smaller than the OmB rms. Figure 5 indicates that the assumed EWP observation errors are reasonable as compared to background errors.

It seems that the poor quality of the data from some stations, like station 11120, can be handled by introduction of a station dependent bias correction procedure. However high rms scores, not accompanied by a particularly high bias, like in station 03969, is impossible for correct for. One needs to introduce a blacklist for these stations until the problems are solved.

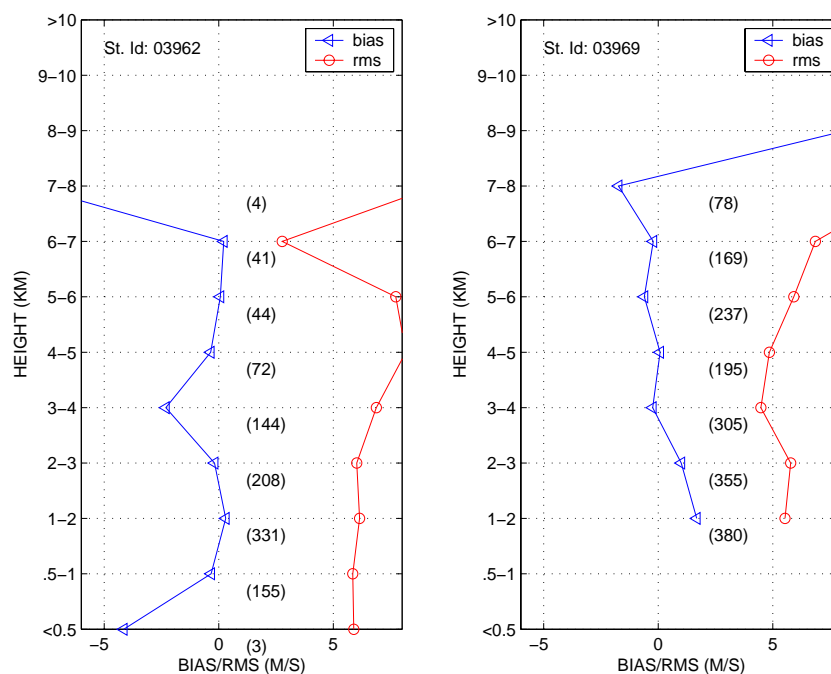


Figure 4: OmB bias and rms for stations 03962 (left) and 03969 (right), as function of height. Unit: m/s. The numbers inside parenthesis represent the number of observations, at each level, on which the statistics are based.

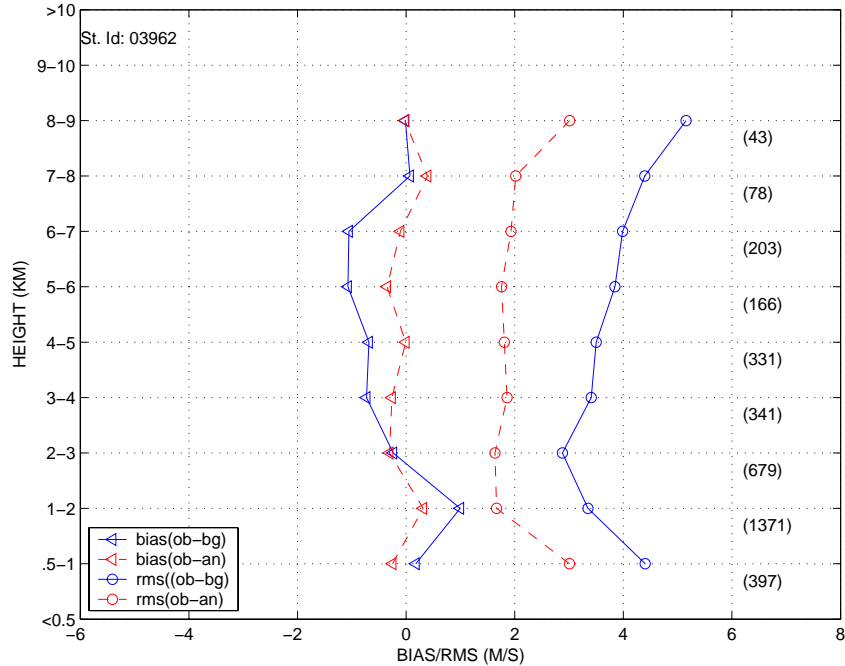


Figure 5: OmB bias and rms, OmA bias and rms for station 03840, as function of height. Unit: m/s. The numbers inside parenthesis represent the number of observations, at each level, on which the statistics are based.

4.3 Data amount relative to other observation types

Including additional data to an observing system may lead to impact on data usage of other observation types. The numbers of different observations which passed data screening for the first three data assimilation cycles on 0600, 0900, 1200 UTC 1 February 2002 are given in Table 6. The difference in PILOT observations between T3W and T3R is obvious, *i.e.*, there are 15 EWPs in T3W run. Other PILOT observations may also be American wind profilers as well as the ordinary PILOT balloons. A noticeable impact on AIREP screening on 0900 UTC is unexpected. There are 40 more AIREP observations rejected by the screening module in T3W. Similar screening results are also found in other cycles on 0900, 1500, 2100, 0300 UTC (we refer to them as asynoptic hours in this report). The differences in AIREP numbers are smaller in cycles on 0600 and 1800 UTC. They are negligible in cycles on 0000 and 1200 UTC. For other observation types, the screening results from T3R and T3W are almost identical.

The numbers of different observation types in Table 6 should not be directly compared.

Table 6: The numbers of observations which passed data screening at 0600, 0900 and 1200 UTC 1 Feb 2002.

Observation type	0600 UTC		0900 UTC		1200 UTC	
	T3R	T3W	T3R	T3W	T3R	T3W
SYNOP	2563	2563	2152	2151	2572	2572
AIREP	2425	2415	2209	2169	3356	3357
BUOY	120	120	122	123	130	130
TEMP	29	29	3	3	218	218
PILOT	23	38	4	19	15	30
ATOVS	347	347	0	0	241	240
SHIP	230	230	135	135	242	242

Observed variables (*e.g.* u and v) and levels of each observation should also be taken into account. It is a general belief that observations with information of vertical distributions of atmospheric quantities are relatively more important than those without. ATOVS radiances contain vertical information of temperature. Their importance relative to TEMP is still to be assessed for the HIRLAM system. AIREP observations also provide vertical information during the aircraft ascending and descending. However, it is difficult to estimate how many AIREP observations provide vertical information. If we assume TEMP and PILOT are the major sources for vertical information of the atmosphere, we may state that the EWP data increase the observed vertical profiles significantly on asynoptic times (from 3+4 to 3+19, $\sim 200\%$ increase), marginally on 0600 and 1800 UTC (from 29+23 to 29+38, $\sim 30\%$ increase), and negligibly on 1200 and 0000 UTC (from 218+15 to 218+30, $\sim 6\%$ increase).

Similar remarks can be made for the observations after VarQC. The numbers of different observations which passed the VarQC (with the VarQC flag 1) for the first three data assimilation cycles on 0600, 0900 and 1200 UTC 1 February 2002 are given in Table 7. The major difference in the number of quality controlled observations are in PILOT. The unexpected impact on the AIREP screening extends also after the AIREP VarQC. The differences between T3W and T3R in the number for other observation types are very small.

The noticeable difference (~ 40) in AIREP screening at 0900 UTC (Table 6) affects both winds and temperatures. This indicates that the reason of the difference is the data thinning, that is applied on report level (one AIREP report usually containing both

Table 7: The numbers of observations which passed VarQC (with flag 1) at 0600, 0900 and 1200 UTC 1 Feb 2002.

Observation		0600 UTC		0900 UTC		1200 UTC	
type	variable	T3R	T3W	T3R	T3W	T3R	T3W
SYNOP	<i>height</i>	2249	2254	1986	1984	2278	2280
BUOY	<i>height</i>	89	89	90	88	90	90
SHIP	<i>height</i>	217	217	127	127	224	224
AIREP	<i>temperature</i>	2394	2384	2184	2144	3310	3309
	<i>wind - u</i>	2395	2386	2153	2116	3286	3284
	<i>wind - v</i>	2395	2386	2153	2116	3286	3284
TEMP	<i>temperature</i>	1233	1233	63	64	7405	7415
	<i>wind - u</i>	1241	1243	63	63	6364	6356
	<i>wind - v</i>	1241	1243	63	63	6364	6356
	<i>humidity</i>	1069	1070	54	54	6147	6154
PILOT	<i>wind - u</i>	357	753	130	489	228	548
	<i>wind - v</i>	357	753	130	489	228	548
ATOVS	<i>radiance</i>	3470	3470	0	0	2410	2400

wind and temperature information). The final numbers of VarQC checked AIREP wind observations, and that of the temperature observations in T3W and T3R, are clearly different at 0900 UTC, but differences in T , u and v remain at the same level as 40 (Table 7).

If we compare wind profiles from TEMP and PILOT, we could also state that the EWP data increase the data volume significantly on asynoptic times (from $2 \times (63 + 130)$ to $2 \times (63 + 489)$, $\sim 190\%$ increase), marginally on 0600 and 1800 UTC (from $2 \times (1245 + 375)$ to $2 \times (1245 + 753)$, $\sim 23\%$ increase), and negligibly on 1200 and 0000 UTC (from $2 \times (6377 + 228)$ to $2 \times (6376 + 548)$, $\sim 5\%$ increase).

Just from the above simple comparisons, we may argue that when the data assimilation experiments are performed with 6 h cycles, as T6W and T6R, the impact of the EWP data will be much smaller than with 3 h cycles.

Finally, we should stress here that even when the EWP data increase the volume of observations significantly they only contribute to part of the wind data. To assess the relative value of the EWP data other aspects of the experiments have to be examined.

4.4 Data impact on analyses and forecasts

To assess the impact of EWP data on the analyses and forecasts, observation verifications are made using the radiosonde and synoptic observations from stations on the EWGLAM station list. Figure 6 shows observation verification for T3W and T3R. In each panel, the two upper curves are for rms and the two lower curves are for bias. Variable selections are made according to the operational practice.

From Figure 6 we see a neutral or marginally positive impact from EWP data on most variables at different levels: Mean Sea Level Pressure (MSLP), 10 m winds, 850 hPa, 500 hPa, 200 hPa geopotential height, temperature and winds. As discussed earlier, we need a reasonable monitoring statistics, a blacklisting and screening strategy before we can properly conduct impact studies. Without them the full potential of the EWP data cannot be explored. Erroneous data could lead to detrimental failures in data assimilation experiments. With that in mind we can only regard T3W as preliminary and we may expect larger positive impact when problematic stations are properly treated.

The largest impact from EWP is found on MSLP (Figure 6). As analyses on asynoptic hours are not included in the verification, no difference between T3W analyses and T3R analyses can be seen. The impact increases as the forecast length increases.

In order to see whether this impact comes from a few cases, we have looked at the daily variation of the above scores. Figure 7 shows observation verification scores (rms and bias) of MSLP for analysis (a) and 12 h (b), 24 h (c), 48 h (d) forecasts for T3R and T3W. In general the EWP data only change the scores slightly on a daily basis. There are two periods over which differences can easily be seen. The EWP data have a negative impact over the first period, around 1800 UTC 18th, and a positive impact over the second period, around 0600 UTC 27th.

The differences in observation verification scores are generally quite small. However, over the periods with relatively large impact, we are able to see the differences reflected on weather maps.

The 48 h surface pressure forecast error (forecast - analysis) for T3R and T3W, the 48 h surface pressure difference between T3W and T3R forecasts, and the mean sea level pressure analysis (from T3R, but that from T3W looks almost identical), all valid at 1800 UTC 18 February 2002, are shown in Figure 8.

This is a case, in which the EWP data have a negative impact on the 48 h forecast.

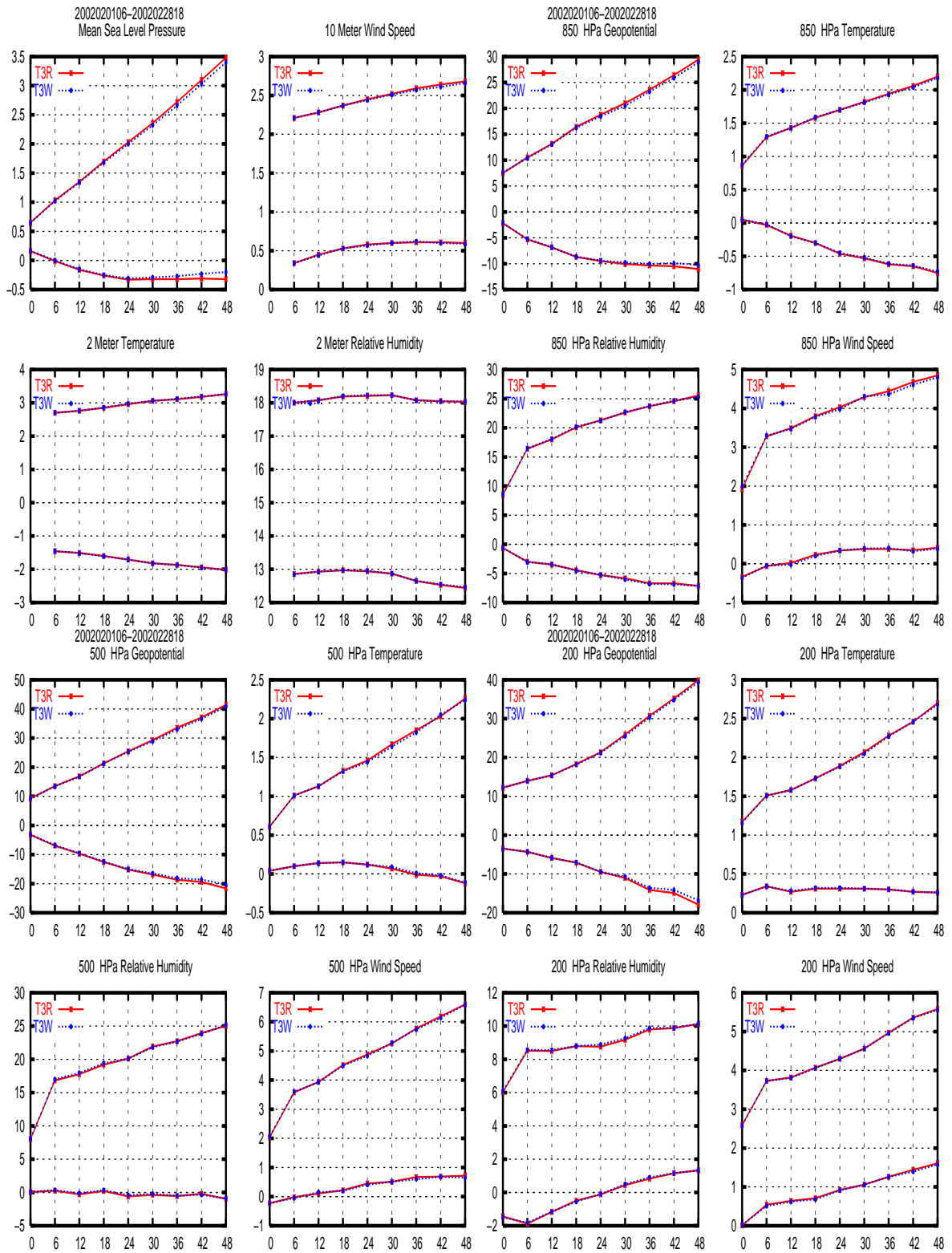


Figure 6: Observation verification for T3W and T3R over February 2022 as functions of forecast length. In each panel, two upper curves are for rms and two lower curves are for bias.

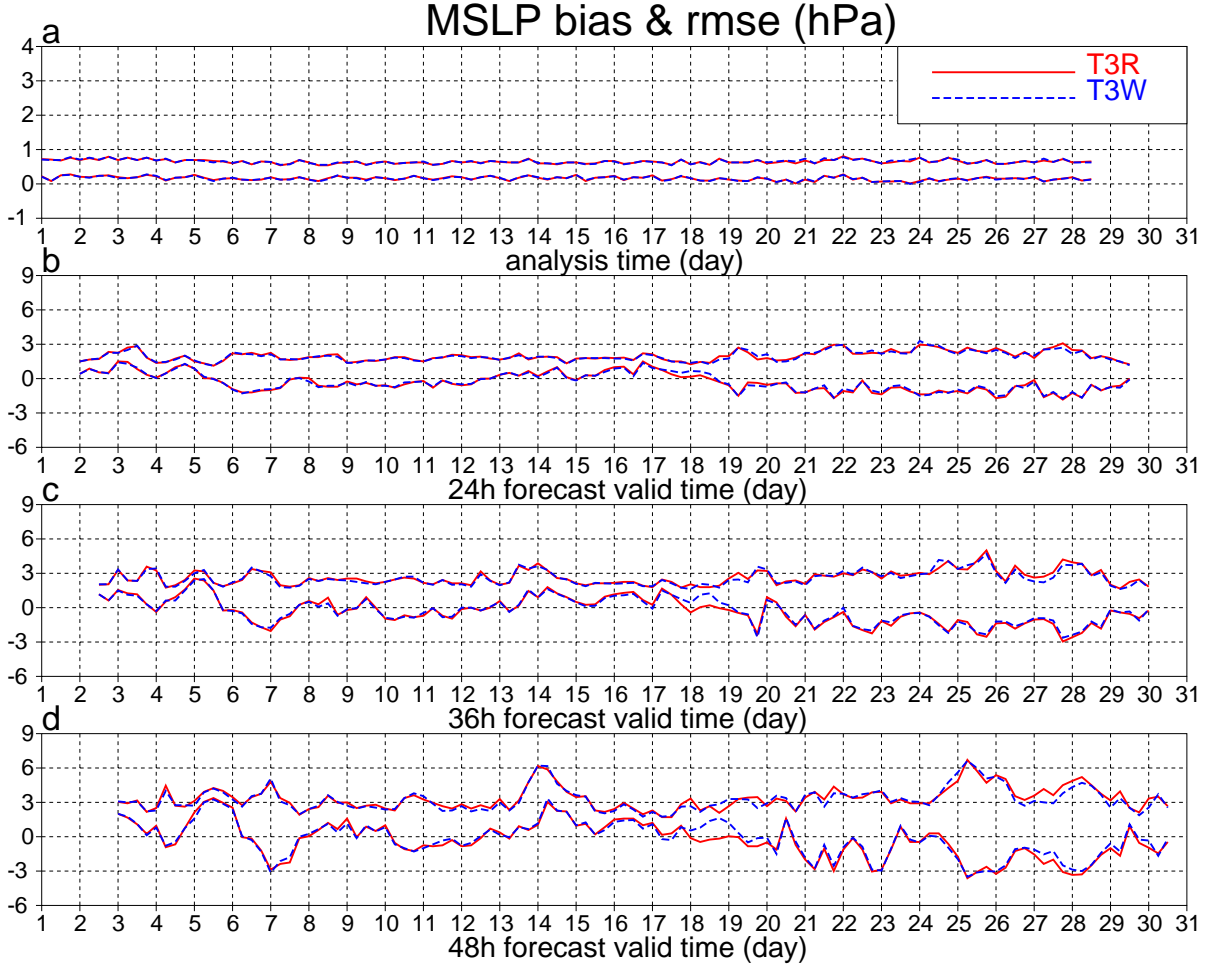


Figure 7: Daily observation verification scores of mean sea level pressure as functions of their valid time for T3W and T3R. (a) analyses. (b) 12 h forecasts. (c) 24 h forecasts. (d) 48 h forecasts.

We see the forecast errors are in general larger in the T3W forecast (Figure 8b) than in the T3R forecast (Figure 8a). The difference between the two forecasts (Figure 8c) can be as large as 8 hPa. The two verifying analyses are very similar. Unlike the observation verification, which mainly provides error estimates over observation dense areas (here Europe), the verification against analyses, which is also referred to as field verification, has the ability to reveal problems over data sparse areas. There are large forecast errors over the ocean in both forecasts. A closer look at the error distribution reveals that over a large part of Europe, the EWP data actually have a positive impact on the forecast (the T3W forecast has smaller errors) for this case.

The 48 h surface pressure forecast error (forecast - analysis) for T3R and T3W, the 48 h surface pressure difference between T3W and T3R forecasts, and the mean sea level

pressure analysis (from T3R, but that from T3W looks almost identical), all valid at 0600 UTC 27 February 2002, are shown in Figure 9.

This is a case in which the EWP data have a positive impact on the 48 h forecast. The two analyses are again very similar. We see the forecast errors are smaller in the T3W forecast (Figure 9b) than in the T3R forecast (Figure 9a). The improvement in the T3W forecast is mainly over the European continent. The T3W forecast errors over the ocean are even slightly larger than the T3R forecast errors. The differences between the forecasts (Figure 9c) also confirm that there are large differences over Europe and the T3W forecast is better.

We have also tried to go through the above comparisons between the two experiments with 6 h cycles, T6W and T6R, but found almost no impact (neutral impact) from the EWP data.

5 Conclusions

The HIRLAM 3D-VAR system has been enhanced to handle data from the EWP network. Four data assimilation experiments have been performed for February 2002.

During this month the EWP network provided a significant amount of data. The data amount and quality are clearly station dependent. High data rejection ratios have been found for a few stations, indicating possible problems with the data. Large bias and rms values between observations and the background states have been found for a few stations, also indicating a need for bias correction or blacklisting.

Compared to the total data volume, the contribution of the EWP data is rather minor. However, compared to the wind profile data volume, mainly from radiosondes and PILOT balloons, the EWP data are significant and, at asynoptic hours, even the major data source.

Data assimilation experiments are performed without any bias correction or blacklisting. The experiments with 3 h assimilation cycles show that small positive impacts from the EWP data on the analyses and forecasts dominate over the one month period.

The necessary development for HIRLAM 3D-VAR, which separates EWP data from PILOT data and applies different screening and blacklisting to EWP data, has recently been made. DMI has also stated the routine monitoring of EWP data.

The selected experiment period and model resolution are not ideal for assessing the

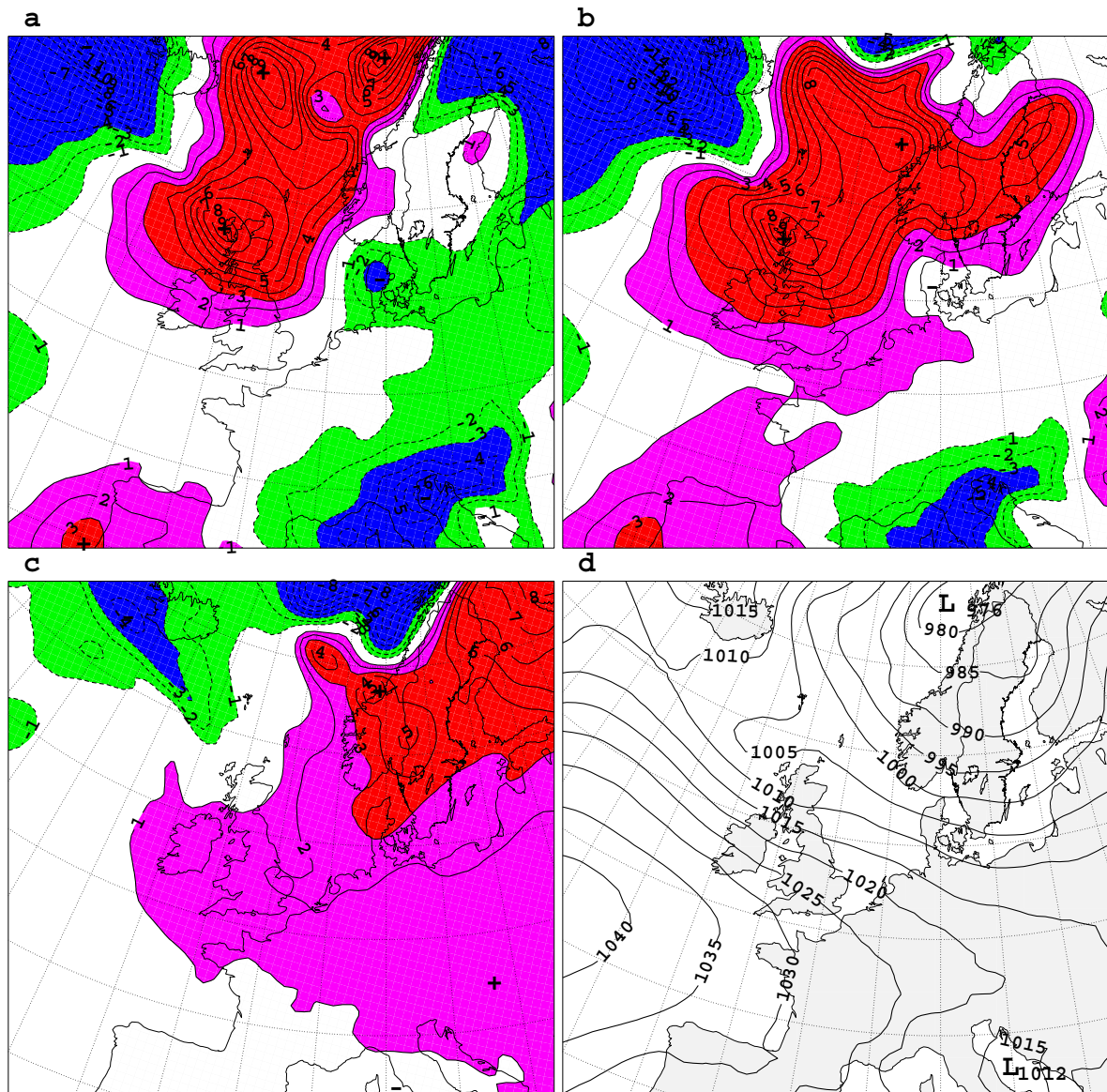


Figure 8: The surface pressure difference and the mean sea level pressure, both valid at 1800 UTC 18 Feb. 2002. a) 48 h forecast error for T3R; b) 48 h forecast error for T3W; c) 48 h forecast difference between T3W and T3R; d) verifying T3R analysis.

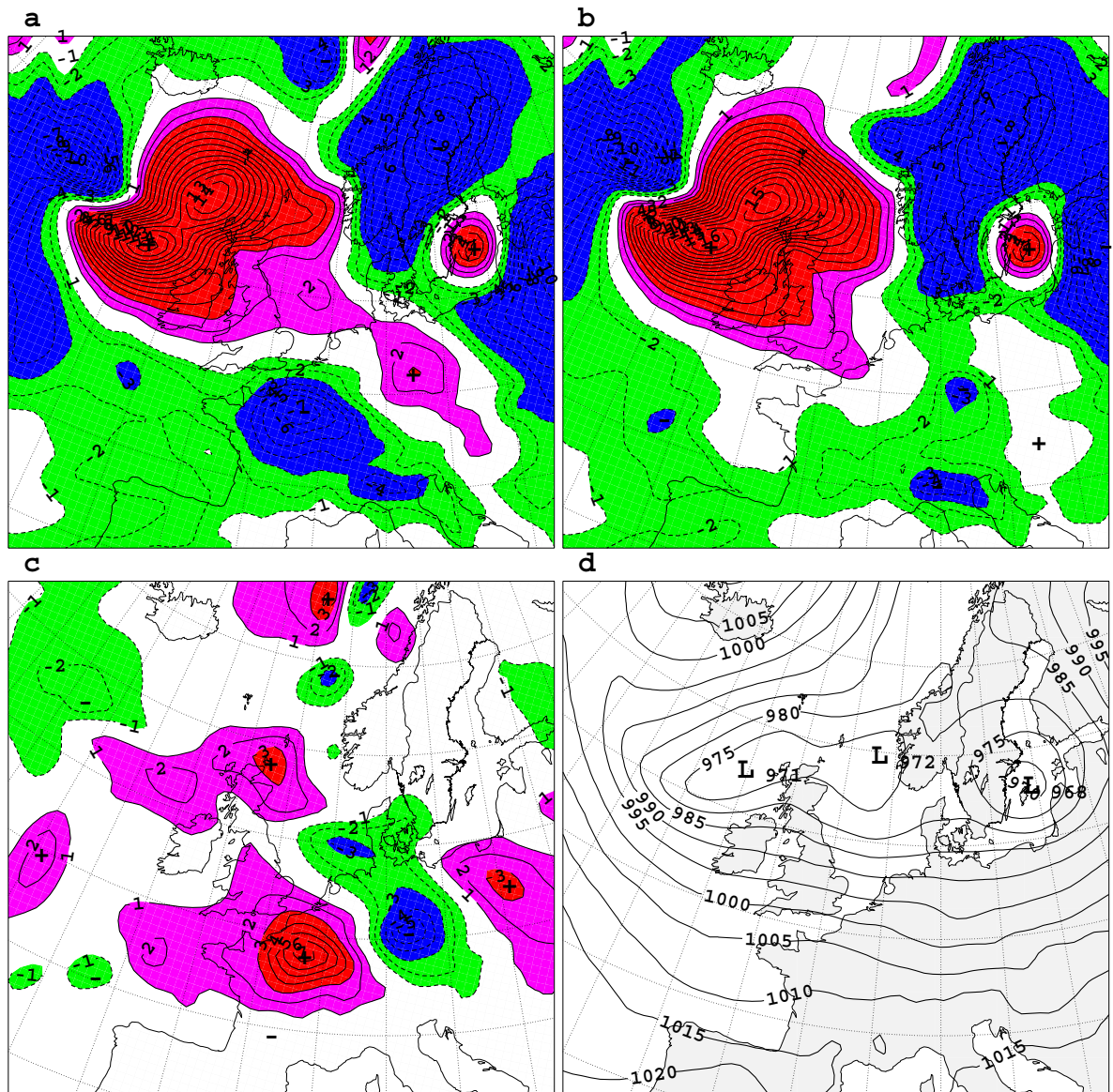


Figure 9: The surface pressure difference and the mean sea level pressure, both valid at 0600 UTC 27 Feb. 2002. a) 48 h forecast error for T3R; b) 48 h forecast error for T3W; c) 48 h forecast difference between T3W and T3R; d) verifying T3W analysis.

EWP data impact. The lack of bias correction or blacklisting probably also reduces the benefit from the data. All these aspects will be taken into account in further studies.

There are HIRLAM monitoring and impact studies on Doppler radar VAD wind profiles (Lindskog *et al.*, 2002). It would be interesting to combine the efforts spent on VAD profiles and EWP profiles.

It is worth mentioning that the EWP data are assimilated by the ECMWF 4D-VAR at 30 min interval (Andersson and Garcia-Mendez, 2002). The EWP data volume used in their experiments is 6 or 12 times that used in our experiments. To further explore the potential of the EWP data the HIRLAM 4D-VAR should also be considered.

6 Acknowledgements

The authors would like to thank Erik Andersson (ECMWF) for providing us with their suggestions and programs, Bjarne Amstrup (DMI) for providing us with DMI ATOVS data and ATOVS related files, Henrik Vedel (DMI), Per Undén and the anonymous reviewer for useful comments.

References

- Amstrup, B. 2002. Impact of ATOVS AMSU-A radiance data in the DMI-HIRLAM 3D-Var analysis and forecasting system - February 2002. *Pages 68–76 of: HIRLAM Workshop on Variational Data Assimilation and Remote Sensing, 21-23 January 2002, Finnish Meteorological Institute, Helsinki, Finland. Available from HIRLAM-5, c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.*
- Andersson, E. and Garcia-Mendez, A. 2002. *Assessment of European wind profiler data, in an NWP context.* ECMWF Tech. Memo. 372, 14pp. Available from the European Centre for Medium Range Weather Forecasting, Shinfield Park, Reading, Berks. RG2 9AX, UK.
- Gilbert, J.C. and Lemaréchal, C. 1989. Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Prog.*, **B25**, 407–435.
- Gustafsson, N. and McDonald, A. 1996. A comparison of the HIRLAM gridpoint and spectral semi-Lagrangian models. *Mon. Wea. Rev.*, **124**, 1008–2022.

- Gustafsson, N., Hörnquist, S., Lindskog, M., Berre, L., Navascués, B., Thorsteinsson, S., Huang, X.-Y., Mogensen, K.S. and Rantakokko, J. 1999. *Three-dimensional variational data assimilation for a high resolution limited area model (HIRLAM)*. HIRLAM Tech. Rep. 40. Available from HIRLAM-5, c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.
- 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.
- Hall, C. 1987. A common verification scheme for limited area models. *EWGLAM Newsletter*, **15**, 144–147.
- Holtstlag, A. A. M. and Boville, B. A. 1993. Local versus nonlocal boundary layer diffusion in a global climate model. *J. Climate*, **6**, 1825–1842.
- Lindskog, M. 2000. An estimate of the seasonal dependence of background error statistics in the HIRLAM 3D-Var. *HIRLAM Newsletter*, **35**, 71–86. Available from HIRLAM-5, c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.
- Lindskog, M., Gustafsson, N., Navascués, B., Mogensen, K.S., Huang, X.-Y., Yang, X., Andrae, 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*, **353A**, 447–468.
- Lindskog, M., Järvinen, H. and Michelson, D.B. 2002. *Development of Doppler radar wind data assimilation for the HIRLAM 3D-Var*. HIRLAM Tech. Rep. 52, 22pp. Available from HIRLAM-5, c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.
- Parrish, D.F. and Derber, J.C. 1992. The national meteorological center's global spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Sass, B.H., Nielsen, N.W., Jørgensen, J.U. and Amstrup, B. 1999. *The operational HIRLAM system at DMI - October 1999*. DMI Technical Report 99-12. Available from Danish Meteorological Institute, Lyngbyvej 100, 2100 Copenhagen Ø, Denmark.

Savijärvi, H. 1990. Fast radiation parameterization schemes for mesoscale and short-range forecast models. *J. Appl. Meteor.*, **29**, 437–447.

Schyberg, H., Landelius, T., Thorsteinsson, S., Tvetter, F.T., Vignes, O., Amstrup, B., Gustafsson, N., Järvinen, H. and Lindskog, M. 2003. *Assimilation of ATOVS data in the HIRLAM 3D-Var system*. HIRLAM Tech. Rep. 60, 67pp. Available from HIRLAM-6, c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.