Evaluation notes: HIRLAM 6.2.2 for a winter episode

Xiaohua Yang
Danish Meteorological Institute

Abstract

A data assimilation experiment is configured on ECMWF computers to validate the recent reference beta release 6.2.2. The experiment is performed on the RCR domain at 0.2 degree resolution and 40 vertical levels, for a half-month winter episode. The results are compared to an earlier validation run for the release 6.1.2/6.2.0, for which a vertical structure of 31 levels were used. Differences in observational verification scores and observation data usage from the parallel results are presented, with additional sensitivity experiments performed for sub-periods, in order to examine the cause of verification features found in the comparisons. The study confirms runs using the new code version (6.2.2) having comparable forecast skills compared to the corresponding ones with the current official version (6.1.2/6.2.0), but also reveals likely problem of extending vertical resolution without corresponding adjustment in assimilation scheme.

1 Main features of the new reference model for validation

As listed in the release notes of 6.2.1 and 6.2.2 (see HIRLAM's HexNet), there has in general been insignificant changes in source code of the HIRLAM system in the latest two beta releases. There are, however, some significant changes in the default set up of the HIRLAM system, as are specified in the reference scripts. Among them, the default model domain is changed from the previous 114x100x31 (0.5 degree in horizontal, 31 level) to 406x324x40 (0.2 degree, 40 level, hereafter referred to as RCR domain). For upper air analysis (3D-VAR), the First Guess at Appropriate Time (FGAT) is adopted as default option. For initialization, the Incremental DFI (IDFI) become now the default scheme.

2 Validation results

The data assimilation experiments are performed on ECMWF platforms (SGI/IBM). The chosen winter episode covers period between 2002011500 and 2002013118. The computational domain is RCR. A 48-hour forecast is made at 6-hour assimilation interval, using ECMWF analyses as lateral boundary conditions. The experiments are run with smoothed orography, as was the case in validation of 6.1.2. Other than that, the whole experiment setup follows those of default ones as specified in the standard scripts, thus including new features such as FGAT, IDFI, and the RCR domain with 40 levels. Only conventional observation types are assimilated. Other aspects of configurations are similar to the previous validation test performed for 6.1.2/6.2.0, except that the latter was made with 31 level.

Observation verifications are performed against EWGLAM station-list for synoptic and sounding data. Figure 1 shows the comparison of verification scores between 6.2.2 (with 40 levels), and 6.1.2, (with 31 levels), for key verification parameters at surface, 850, 500 and 200 hpa levels, valid during the 16-day period (20020115-20020131). As shown in Figure 1, the standard deviation (STD) and bias scores of runs using 6.1.2 (solid line, star) and 6.2.2 (dashed line, circle) are generally similar, with STD scores of 6.2.2 to be slightly worse than for 6.1.2. For surface parameters, 6.2.2 features improved T2m, V10m and RH2m bias, but at the same time marginally worse STD. 6.2.2 has also slightly worse MSLP scores in both STD and BIAS. For precipitation verification, 6.2.2 results are seen to have somewhat reduced total precipitation compared to 6.1.2 (see also Figure 3), and accordingly, reduced positive precipitation bias (not shown here). As shown in the contingency table (Table 3) for 12-hr accumulated precipitation from two selected forecast ranges, the 6.2.2 results are seen to be associated with noticeably improved fit at none-to-small precipitation class, but somewhat worse in the higher classes.

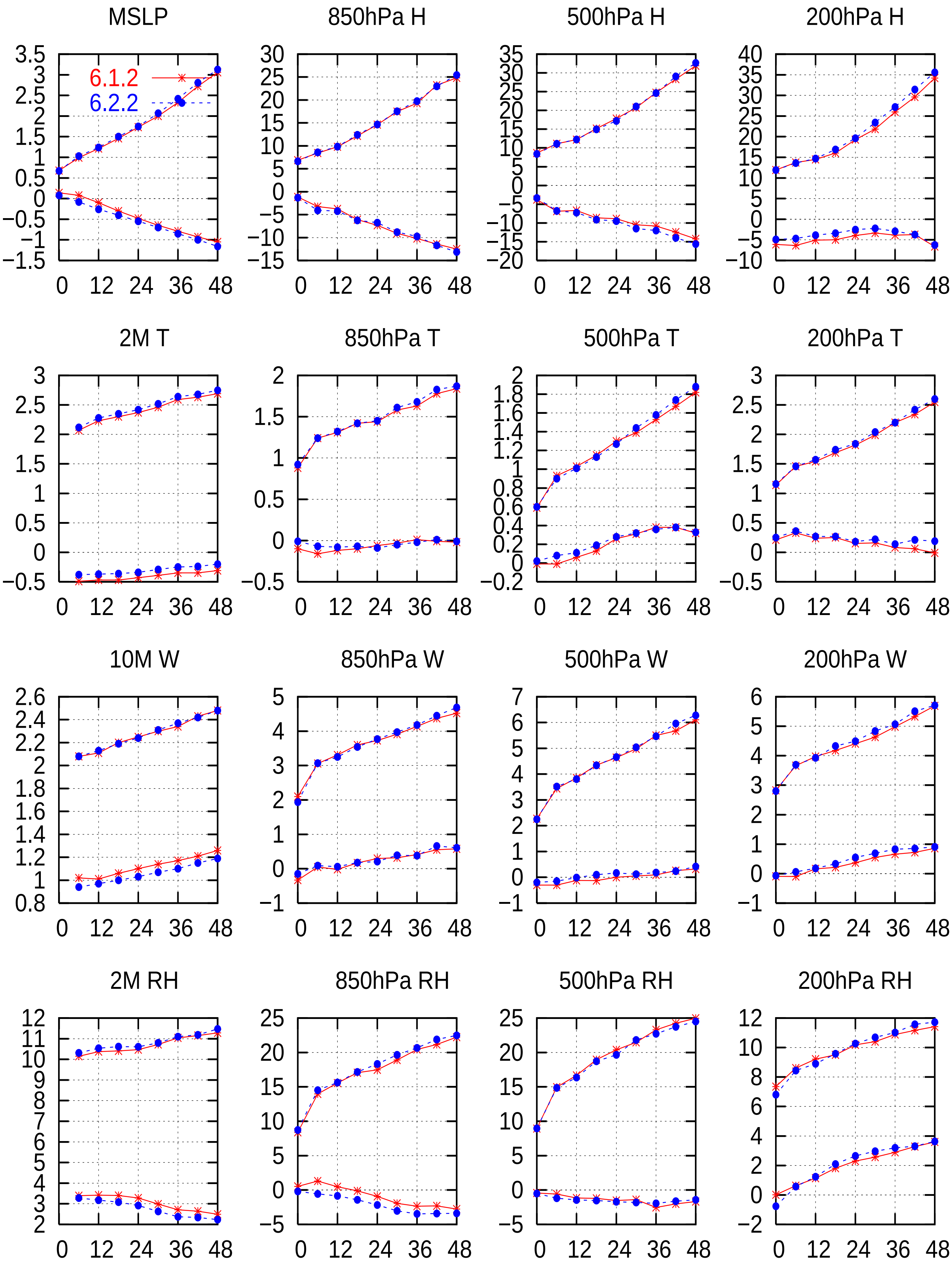

Figure 1: Observation verification against EWGLAM stations for surface (first column), 850 hPa (second column), 500 hPa (third column) and 200 hPa (fourth column) parameters, for period between 2002011500 and 2002013118.

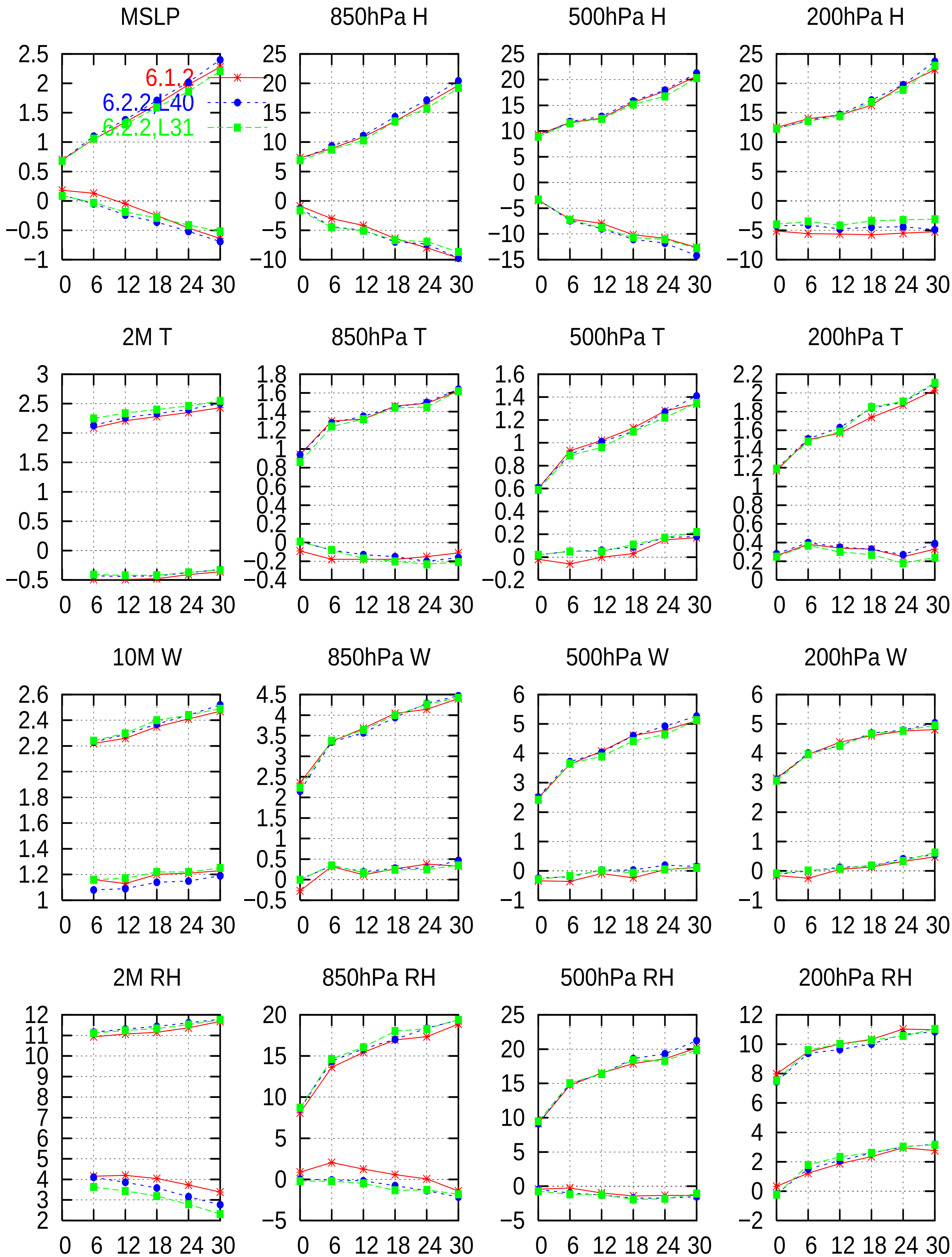

Figure 2: Observation verification against EWGLAM stations for surface (first column), 850 hPa (second column), 500 hPa (third column) and 200 hPa (fourth column) parameters, for period between 20020122 and 20020131.

Table 1: Contingency table for 12-hour accumulated precipitation, 20020115-013118. The precipitation events are grouped in 4 classes, between 0-0.1, 0.1-2, 2-10 and above 10 mm, respectively.

	12-24 h				36-48 h			
obs→ / ↓ for	O1	O2	O3	O4	O1	O2	O3	O4
F1, 6.1.2	2423	115	50	14	2604	210	98	18
F1, 6.2.2	2691	163	52	12	2778	250	112	23
F2,6.1.2	2261	1239	317	29	2036	1121	345	37
F2,6.2.2	1986	1221	361	29	1850	1100	367	37
F3,6.1.2	157	424	926	46	202	416	873	62
F3,6.2.2	165	397	905	63	213	396	846	62
F4, 6.1.2	3	10	82	62	5	14	68	44
F4, 6.2.2	2	7	57	47	6	15	59	39

3 Additional experiments for sub-episode

While the verification results presented above are not catastrophic for the latest version (6.2.2), it is natural to ask why no obviously improved scores with 6.2.2 using the current RCR settings with features such as FGAT, IDFI as well as an increased vertical resolution. A sensitivity run using DFI initialization with 6.2.2, as was used in runs of 6.1.2, are conducted for a 7-day episode within the test period, and the verification scores showed marginally worse scores with DFI compared to the run with IDFI, thus excluding possibility of degraded scores due to possible implementation error of IDFI (results not shown here). In another experiment, the above 6.2.2 run is redone, using 31 levels, for a sub-period covering 20020122 and 20020131, which featured several stormy episodes over northern Europe including heavy precipitation events. Comparing the thus obtained 6.2.2 results with those from the 6.1.2 (31 level) and 6.2.2 (40 level) runs for the same period, it is found that the 31 level 6.2.2 runs have slightly better scores than both 6.2.2 using 40 level and 6.1.2 with 31 level, as shown in Figure 2. Thus it appears that, at least for the tested winter episode, the 6.2.2 version is a slight improvement over the version 6.1.2, using more comparable configuration. On the other hand, there seem to be a slight degradation of forecast skills associated with change of vertical resolution from 31 level to 40 level. This will be discussed further in next section.

4 Some aspects of the 3D-VAR analysis

The HIRLAM 6.2.2 includes HIRVDA 6.2.1 as upper air analysis scheme. When used for assimilation of conventional data, HIRVDA 6.2.1 has no significant changes comparing to previous version, 6.1.2, as used in HIRLAM 6.1.2/6.2.0. In 6.2.2, the default scheme is now changed to FGAT. By comparing observation values to the closest first guess in time, the FGAT has been demonstrated to be able to use observation data more properly and results in slight gain in forecast skills. Furthermore, with the current DMR/RCR, the horizontal resolution is 0.2, which is about twice the resolution of the original Swedish operational model from which the currently used statistical background error structure function was derived. For that consideration, the 3DVAR minimization in 6.2.2 for RCR is performed at a reduced resolution (0.4 degree). This is also believed to result in a slightly more efficient minimization. Finally, when running 40 level RCR, the previously derived 31 level background error structure function from SMHI historical data archive is interpolated to the new vertical coordinate with 40 level. This is an ad hoc approximation due to lack of forecast error statistics directly derived from the new 40-level model.

In Table 2, the statistics of active and unused reports are summarized for 6.1.2 (31 level) and 6.2.2 runs (40 and 31 levels), averaged over all assimilation cycles in the sub-period between 20020122 and 20020129. The unused reports include those rejected by gross first guess check, as well as those erroneous or redundant reports, etc. Interestingly, the Table 2 shows a clearly reduced data usage in

Table 2: Average number of observation report per assimilation cycle, based on statistics from the 3D-VAR data screening module, for period 22–29 January 2002. The numbers shown are sums of active or unused observation reports.

Observation	6.1.2, 31 level			6.2.2, 40 level			6.2.2, 31 level		
type	active	unused	total	active	unused	total	active	unused	total
SYNOP	2183	5298	7481	2118	3407	5525	2118	3407	5525
AIREP	2529	3272	5801	2523	3263	5786	2523	3263	5786
BUOY	115	526	641	113	488	601	113	488	601
TEMP	98	1	99	97	0	97	97	0	97
PILOT	13	12	25	13	12	25	13	12	25
SHIP	228	273	501	195	202	397	195	202	397

the 3D-VAR procedure in 6.2.2 runs comparing to the previous run with 6.1.2.

The reduced total amount of observation data entering 3D-VAR procedure is caused solely by an error correction in the 6.2.2 script for observation data handling (MakeCMA). Previously, the observation data window for an analysis cycle centered at a certain synoptic hour is defined to be [T - CYCINT/2, T + CYCINT/2] (in minute), where T is the analysis time, CYCINT is the assimilation interval. The data window is now corrected to be [T - CYCINT/2, T + CYCINT/2 -1]. This is to avoid repeated use of those observation data with nominal sampling time at precise start (and end) of the intermittent assimilation data window. As a consequence, this subtle differences in the data window definition result in a noticeable reduction of input data entering 3D-VAR analysis scheme, most significantly for SYNOP/SHIP data, as is shown in the Table 2. A repeated use of same observation data in consecutive assimilation cycles is theoretically unjustifiable, thus the reduced data amount entering each assimilation cycle in 6.2.2, in comparison to 6.2.1, is healthy.

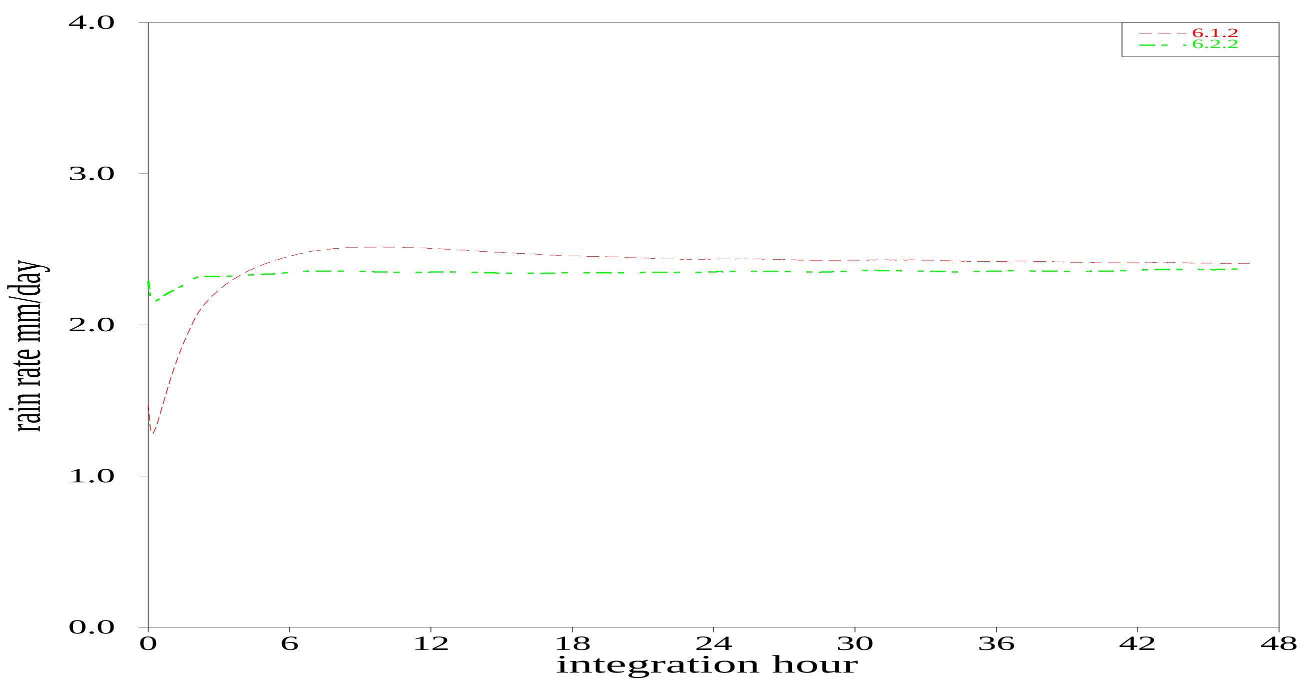

Figure 3: Time series of domain averaged total precipitation rate, averaged for all forecast runs between 20020115 and 20020126.

In Table 3, the corresponding statistics of data usage based on records of 3D-VAR variational

Table 3: Results of the variational quality control module, averaged from periods 22 — 29 January 2002. The numbers shown are sum of observations with different VarQC flags (1 - correct; 2 - probably correct; 3 - probably incorrect; 4 - incorrect).

		6.1.2,31 level				6.2.2, 40 level				6.2.2, 31 level			
TYPE	VAR	1	2	3	4	1	2	3	4	1	2	3	4
SYNOP	height	1936	17	9	26	1888	15	7	21	1890	15	6	21
BUOY	height	83	0	0	1	82	0	0	1	82	0	0	1
SHIP	height	210	1	0	4	183	0	0	2	183	0	0	2
AIREP	temperature	2480	2	1	1	2449	1	0	33	2481	1	0	2
	U/V-wind	2459	9	5	16	2450	6	3	39	2481	6	4	9
TEMP	temperature	3509	9	3	7	3433	9	4	63	3506	8	5	8
	U-/V- wind	3541	7	4	23	3458	9	5	69	3522	9	5	26
	Humidity	2787	1	0	1	2719	1	0	57	2787	1	0	2
PILOT	U-/V- wind	118	0	0	0	114	0	0	4	118	0	0	1

quality control module is shown, for runs using 6.1.2 and 6.2.2, the latter with both 40 and 31 levels. Such statistics is normally a good measure of relative quality of analyses used in different experiment. However, due to significantly different amount of observation data entering 3D-VAR procedure in 6.1.2 and 6.2.2 as discussed above, it is not straightforward to compare the relative performance of 3D-VAR procedure in 6.1.2 and 6.2.2. Of more interest is the relative performance between 6.2.2 runs using 40 and 31 levels, as revealed in the table. It is seen that, contrary to the natural expectation, the 31-level runs are seen to be able to use noticeably more observation data, especially those from upper air observation, than the 40-level runs, despite of better vertical resolution in the latter. Also interesting is the better use of wind data with AIREP in 6.2.2 with 31 level compared to 6.1.2. For AIREP data, the total amount that enters 3DVAR minimization is slightly higher in 6.1.2 than in 6.2.2 (Table 2), due to the reason discussed in Section 4. However, more wind data is actually used in 6.2.2 (2481) than in 6.1.2 (2459) according to Table 3. This is most likely due to the FGAT scheme used in 6.2.2.

5 Initial spin-up feature with IDFI

Incremental DFI initialization is adopted as default scheme in 6.2.2, replacing the previous normal DFI scheme. The incremental approach performs DFI filtering to the first guess and analysis successively, and uses the difference of the filtered prognostic variables as ”initialized analysis increment” to add to the first guess fields. One of the main motivations of the incremental scheme is to reduce the spin-up behavior in the cloud and moisture fields, as well as those in the surface fluxes.

The improvement of spin-up features with IDFI is confirmed when examining time series of domain averaged moisture quantities such as cloud water, cloud cover, rain rate and surface fluxes. Figure 3 show, e.g., the time series for total precipitation rate along forecast length, for part of the winter episode. The spin-up feature seen in the precipitation curve for 6.2.2 (IDFI) look much better than the one for 6.1.2 (DFI). Time series for other traditional ”spin-up” quantities feature similar behaviors (not shown here), thus it is believed that the implemented IDFI, as is used in 6.2.2, successfully alleviates the moisture ”spin-up” problem associated with DFI.

6 Summary

Recent reference HIRLAM beta-release, version 6.2.2, contains generally minor upgrade in the source code, but at the same time a change on several important aspects concerning standard configurations, such as the default model domain with increased horizontal and vertical resolution, 3DVAR-FGAT and incremental DFI scheme. Validation of HIRLAM 6.2.2 for a two-week winter period, performed on ECMWF SGI/IBM computation platform, indicates comparable performance in forecast quality when compared to equivalent runs using 6.1.2. The new features of FGAT option in 3DVAR and incremental DFI in initialization procedure have been confirmed to behave normal. However, results with increased vertical resolution from 31 to 40 levels show slight degradation in some observation verification scores. The problem is also reflected in slightly reduced amount of observation data effectively used in 3DVAR upper air analysis. It is suggested that one possible reason of the unsatisfactory performance with increased vertical resolution may be related to the lack of forecast error statistics derived directly from the 40-level model. The statistical comparison of forecast skills with varying vertical resolution need to be examined further, partly through more careful monitoring of model results by objective and subjective means, partly through similar parallel experiment for more periods covering different seasons. The experiment reported here should also be repeated, when the new version of the forecast error structure function, derived from statistics for RCR runs using 40 levels, become available. [1]

Acknowledgment Kristian Sten Mogensen, DMI, offered useful comments on the observation data usage in 3DVAR.

[1] Recently, Beatriz Navascues at INM discovered erroneous wind fields at level 33 and level 36 in RCR 40-level runs performed at the ECMWF computers. Further examination indicate the problem as being introduced during the lateral boundary interpolation of ECMWF BC frames, performed on ECMWF/SGI platform. No similar problem has yet been found in the parallel runs reported in this report, for which ECMWF analysis, instead of BC frames, are used as lateral boundaries. However, the incidence serves as a reminder that one can not exclude possibility of unexpected technical errors which could cause degradation of model results as observed here.