

Demonstration of the verification tools in the reference HIRLAM system

Kalle Eerola, FMI

1 Introduction

Verification and validation are essential parts of the NWP work, both in operational environment and in research and development work. The HIRLAM reference system contains a verification package, which computes the basic verification scores on daily basis as a part of the analysis/forecast/postprocessing cycle. Verification against observations (at the moment SYNOP and TEMP observations) is done by default for the areas defined by the user, but field verification (verification against analysis) is done only if the user asks for it.

Even if there is a reference verification package, several other verification packages exist in the HIRLAM community. In research and development work the reference verification package is used by many researchers, but in operational work only FMI, INM and KNMI use the reference package. Other institutes running HIRLAM operationally have developed their own verification packages for operational purposes. Hereafter the term "HIRLAM verification package" refers to the verification system included in the HIRLAM reference system.

There are several reasons, why the reference verification system is not used more widely. In the operational environment one reason is that the meteorological institutes want to use continuously the same package to get comparable verification scores from year to year: using the same package and method they get a view of the overall improvement in their operational HIRLAM system. So the software developed originally many years ago before the HIRLAM verification was developed is still used.

Another possible reason for not using the reference package is a very practical one: until now there have been no tools for postprocessing the verification scores, ie. for computing verification statistics over a time period and for presenting the results in graphical form. Several tools have been built in different institutes, but none of them have been installed into the reference system. In the HIRLAM reference version 7.1 the package developed at FMI has been implemented to compute and present summaries of verification scores. The purpose of this report is to demonstrate the possibilities of this package.

The content of this paper is the following. Before demonstrating the new postprocessing possibilities we present, in chapter two, the basic features of the current reference verification system and products from it. After that we discuss, in chapter three, the general features of the package and show several examples of the possibilities of the new tools. There are examples both about verification against observations and verification against analysis (field verification). The last chapter contains a summary of this report.

2 The HIRLAM verification system

The current HIRLAM reference verification system contains two parts

1. Computing the verification statistics on one cycle as a part of the normal analysis/forecast/postprocessing cycle. In this report this is called the "HIRLAM reference verification system".
2. Creating and presenting statistics of the verification scores over the given time period from the data computed in the first part. Hereafter this is called "verification tools".

The second part is a new part in HIRLAM and is the main focus of this report, but before demonstrating its possibilities some comments of the first will be made.

The verification system depends very much on the purpose, for which it has been planned. The HIRLAM verification system is planned to fulfill several requirements. First of all, it is planned to fulfill the need of verification and validation in impact/sensitivity studies. Such studies can be related to physics, data assimilation, system upgrade etc. Secondly, it is planned to be used in regular inter-comparison between different operational HIRLAM systems. Thirdly, the system should fulfill the need of following the progress of the HIRLAM system over long time periods in a measurable way. This requirement deals with long time-series of verification scores and the users can be either developers or the decision makers.

Basically the HIRLAM verification system fulfills these requirements. Although at the moment it uses only SYNOP and TEMP observations, it offers a framework of adding verification against other observation types. On the other hand, there are clearly areas, for which the system is not planned for. It has been built for synoptic scale verification and is not suitable for meso-scale verification. Also it has not been planned to be used in process studies, where exact information about the properties of the process in question is needed. Neither it is planned for verification of ensemble forecasts.

The HIRLAM verification system produces three kinds of output data. Concerning the verification against observations, verification scores of a forecast can be computed for several areas defined by the user. The areas can be polygons defined by the coordinates of the corners or the required stations can be listed in a stationlist. All the forecasts valid at the time of observations are verified and the results are written into the same file, called ve-files with the naming conventions **veyyyyymmddhh**, where *yyyy* is the year, *mm* month, *dd* day and *hh* hour. At the moment supported observation types are SYNOP and TEMP. Concerning SYNOP observations the following parameters are verified: surface pressure, 10-m wind (speed, direction, vector difference), 2-m temperature, 2-m dew point, 2-m rel. humidity, precipitation (6 hr, 12 hr and 24 hr accumulated precipitation depending on observed value) and cloud cover. Concerning TEMP observations (soundings) the verified parameters are: wind (speed, direction and vector difference), temperature, dew point, relative humidity and geopotential on the user-defined constant pressure levels.

For surface observations a separate file is created, containing the identification and coordinates of the station and observed and forecast value on that station. The parameters are the same as mentioned in the previous paragraph. This file can be used in computing the verification statistics on station by station basis. Separate file is created for every parameter and forecast length and the files valid at the same time are collected into zipped tar-files (called often as ZOPS-files), with the naming convention **OFyyyymmdd.hh.tar.gz**.

In field verification the forecast field is compared to the valid analysis gridpoint by gridpoint. For every field, the mean value, bias and rms-error are accumulated on monthly basis.

3 HIRLAM verification tools

The first version of a package for producing summaries of verification scores over a given time period and presenting the results in graphical form is implemented into HIRLAM version 7.0. It is based on the package developed originally at FMI for local purposes. The package is run separately after all the relevant experiments are ready. It supposes that the normal HIRLAM verification has been done as a part of those runs and the verification files described in the previous chapter are available.

The package is run using miniSMS. A special tdf-file (PLAYFILE) called **verify_summary.tdf** is included in the system. So the run is started using the command

```
Hirlam start DTG=...,DTGEND=...,PLAYFILE=verify_summary
```

Also the control of the run is done in the same as in other parts of HIRLAM: a special Env-file called **Env.verif_sum** is used to control the run. By editing this file the user can define, what kind of verification statistics he/she wants, where to find the data, where to put the results etc. By default, the directory \$HL_DATA

is used as a base directory of the output data and working directories. The `Env.verif.sum` included in the system is a good starting point for the user, because it shows also examples of its usage.

The package is written in shell-script and in Perl and gnuplot is the main graphical package. For presenting geographical maps Grads is used.

In the following, examples of the available products are shown and commented. The examples contain products of

- verification against observations
- verification against observations on station basis
- time series of verification scores
- field verification products

3.1 Examples: verification against observations

In verifying against SYNOP and TEMP observations, there are at the moment fixed sets of figures, which can be plotted. Figure 1 shows the verification scores of near-surface parameters against SYNOP observations for the given time period. The setup of the figure is fixed, containing rms-error and bias of mean sea level pressure, 2-meter temperature, 2-meter relative humidity and 10-meter wind speed as a function of forecast length. Several experiments can be plotted into the same figure. The domain can be any of those, for which the daily verification scores have been computed. The user can select, from which daily forecasts the verification scores are computed. This allows, for instance, looking at the daily cycle in the verification scores. For instance, in Figure 1 the forecasts from 00 UTC analysis have been taken into account.

Figure 2 shows the scores computed from soundings (TEMP) for upper-air parameters. The verified parameters are geopotential, temperature, relative humidity and wind speed. At the moment, only the plots for 500 hPa, 850 hPa and 250 hPa pressure levels are supported. Otherwise, the comments given about the Figure 1 are valid.

Separate figures, similar to Figures 1 and 2, showing the number of observations used in computing scores, are also produced (not shown). In this way the user can check that the same number of observations is used in computing the scores in different experiments

For precipitation the contingency tables are printed for the given period, separately for different accumulation periods and forecast lengths. Of course, the accumulation period depends on the observations. In addition, the marginal distributions of observed and forecast precipitation are shown. An example of such a table is shown in Figure 3. In the contingency table the number of cases are also printed in percentages to make the table more readable.

Figure 4 shows an example of verification statistics computed separately for each observing station. Such a figure gives a description of the geographical distribution of errors. It is possible to select any of the surface parameters, for which the scores have been compiled. If several experiments are requested, separate figures are produced for every experiment. In the similar way, separate figures are produced for different valid hours of the day. In this way it is possible to see the errors at daytime and night time separately. Different colors are used to visualize different values. Separate figures are available to show the number of observations used in the computations on each station. Similar figures are plotted showing the rms-error.

The next figure, Figures 5, is an example of time-series of verification scores for the given set of stations. The thick line is the moving average and the user can define the length of the period or skip it totally. Bias and rms-error are plotted in separate figures, but several experiments can be plotted in the same figure. Every parameter and domain, for which the scores have been computed in the basic verification run, can be selected.

3.2 Examples of field verification products

Next we show examples of the products available from the field verification. Figure 6 is an example of the bias map. In this case the error is computed for 48-hour forecasts for January 2007 and the sample contains 31 cases, ie. all forecasts starting from 00 UTC analysis. Similar maps for rms-error can be plotted. The field verification system computes and accumulates mean, bias and rms-error for every field in the given Hirlam files. So it is also possible to plot these scores for any parameter on the constant pressure levels and also on model levels, if the field verification of history files is requested.

As an example of "climatic" fields (mean of the parameter over the given month), Figure 7 shows the mean ice concentration at +48 hours' forecasts in June 2006. Note that the "climate" fields are computed separately for all forecast lengths. This allows of showing also the drift in the forecast model, ie. the mean difference between two forecast lengths.

An example of systematic difference is shown in Figure 8. It shows the difference of the "climates" of two experiments. In this example we see that the experiment ECM701 produces, in the mean, lower surface pressure than experiment ECM71 on the northern sea areas, supposing that the analysis are similar. Such difference maps can be at the moment produced only if the horizontal areas of experiments are the same

4 Summary

The purpose of this report was to show examples of the products available at the moment in the HIRLAM "verification tools". This toolbox contains at the moment tools to show basic verification scores from:

- verification against observations
- verification against observations on station basis
- time series of verification scores
- field verification products

At the moment quite a limited and fixed set of products is available. The basic verification scores can be presented from different experiments. The present version of the toolbox, implemented into HIRLAM 7.0, is the first version and especially the field verification part contains still several weak points.

It is also worth mentioning that recently we have noticed some weak points in the current reference verification system. They are all related to the difficulties in guaranteeing that exactly the same observations are used to compute the verification scores when comparing different experiments. After realizing there is a plan to develop the verification system further. After this work is ready there will be an option to do verification in such a way that the user can be sure of using exactly the same observations in different experiments. So the verification and comparison of different models/experiments is done scientifically in a sound way.

In addition to that we shortly discussed the "HIRLAM reference system". It is well suited for synoptical scale verification, but is not planned for mesoscale verification or process studies or verification of EPS products. Also we mentioned that it contains some drawbacks when used in intercomparison/validation studies. The most important aspect is that it is very difficult to be sure that exactly the same observations are used when comparing verification scores from different experiments/models. In this respect there are plans to correct this in the coming versions.

Verification against observations EXP: V644 FMI71rc1
 Time: 2007010100 - 2007013118 Domain: EWG Forecast from 00

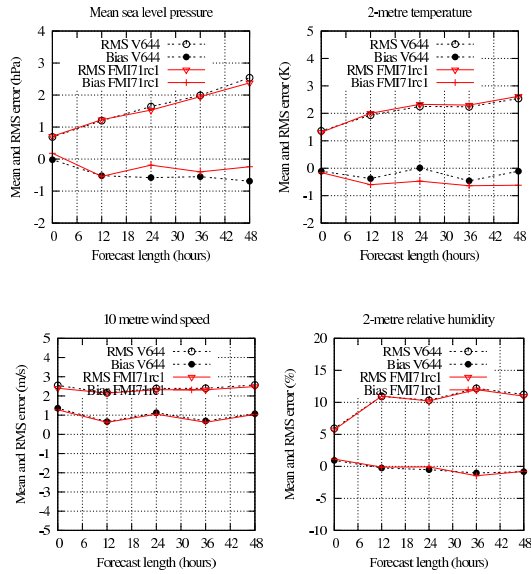


Figure 1: An example of verification scores against surface observations. Scores from two experiments (V644 and FMI71rc1) are shown for January 2007. Domain is so-called EWGLAM stationlist and only forecasts from 00 UTC analysis are used.

Verification against observations EXP: V644 FMI71rc1
 Time: 2007010100 - 2007013118 Domain: EWG Forecast from 00

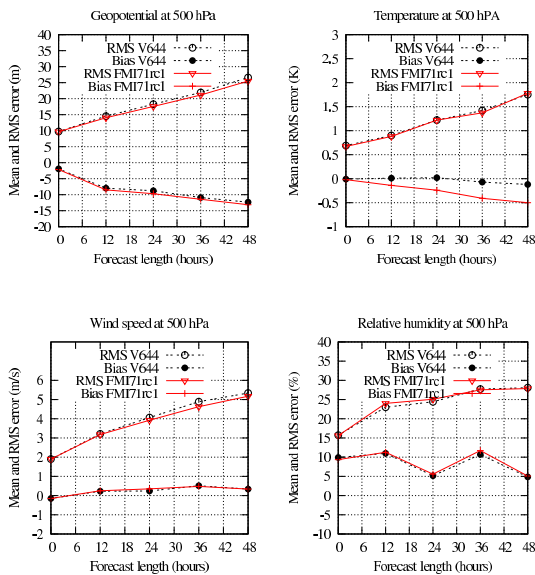


Figure 2: As Figure 1, but verification scores computed against sounding (TEMP) observations are shown.

Forecast length: 42 Accumulation period: 12hr Domain: EWG

Identification: EWG_V644_2007010100-2007013118_CY_00

Marginal distributions: 8

| Limits | <0.1 | <0.3 | <1.0 | <3.0 | <10.0 | <30.0 | <100.0 | >100.0 |
|----------|------|------|------|------|-------|-------|--------|--------|
| Forecast | 2574 | 983 | 1759 | 1659 | 955 | 114 | 2 | 0 |
| In % | 32 | 12 | 22 | 21 | 12 | 1 | 0 | 0 |
| Observed | 4777 | 699 | 1145 | 713 | 590 | 115 | 7 | 0 |
| In % | 59 | 9 | 14 | 9 | 7 | 1 | 0 | 0 |

Contingency table: obs--> forec |

| Limits | <0.1 | <0.3 | <1.0 | <3.0 | <10.0 | <30.0 | <100.0 | >100.0 |
|--------|------|------|------|------|-------|-------|--------|--------|
| <0.1 | 2441 | 81 | 32 | 11 | 9 | 0 | 0 | 0 |
| <0.3 | 773 | 90 | 94 | 18 | 7 | 1 | 0 | 0 |
| <1.0 | 987 | 262 | 362 | 104 | 40 | 4 | 0 | 0 |
| <3.0 | 448 | 218 | 461 | 347 | 166 | 19 | 0 | 0 |
| <10.0 | 119 | 45 | 190 | 229 | 315 | 54 | 3 | 0 |
| <30.0 | 8 | 3 | 6 | 4 | 53 | 37 | 3 | 0 |
| <100.0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| >100.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 3: An example of contingency table of precipitation. Contingency table and marginal distributions of observed and forecast number of cases for 12 hour accumulated precipitation are shown.

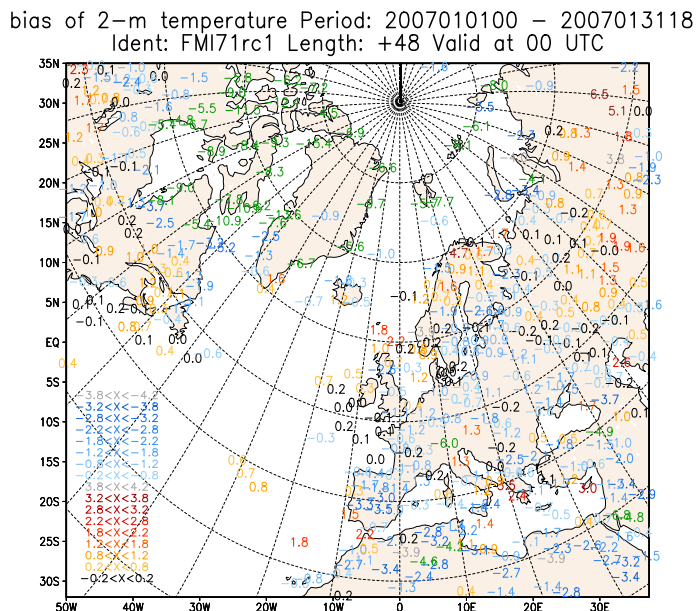


Figure 4: An example of verification scores on stations: bias of 2m-temperature on observing stations. The length of forecast is 48 hours and all the forecasts are valid at night time (00 UTC).

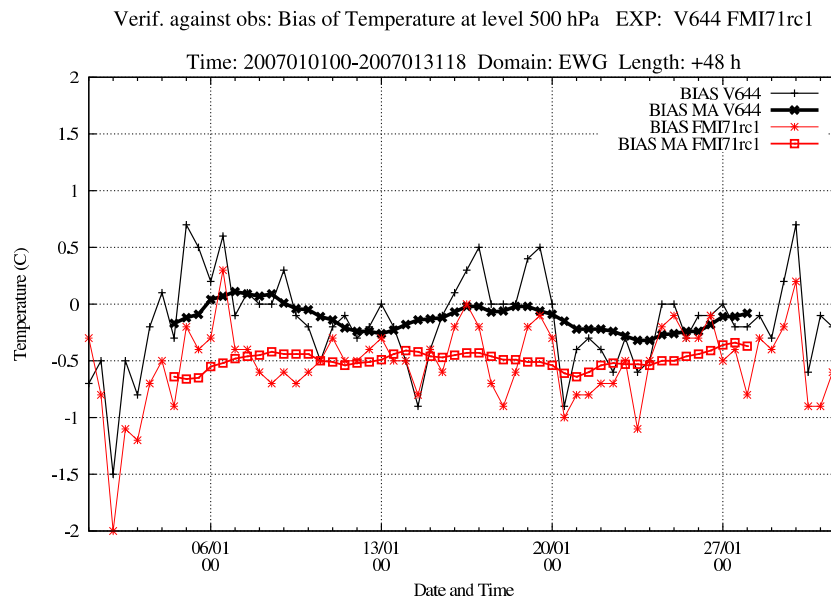


Figure 5: An example of time-series of scores: bias of temperature on 500 hPa level. The forecast length is 48 hours and scores from two experiments, V644 and FMI71rc1, are plotted. The scores are computed on so-called EWGLAM stations.

Bias of Temperature Level 850 hPa Ident: FMI71rc1 NO: 31
First date: 2007010100 Init. time 00 Length +48

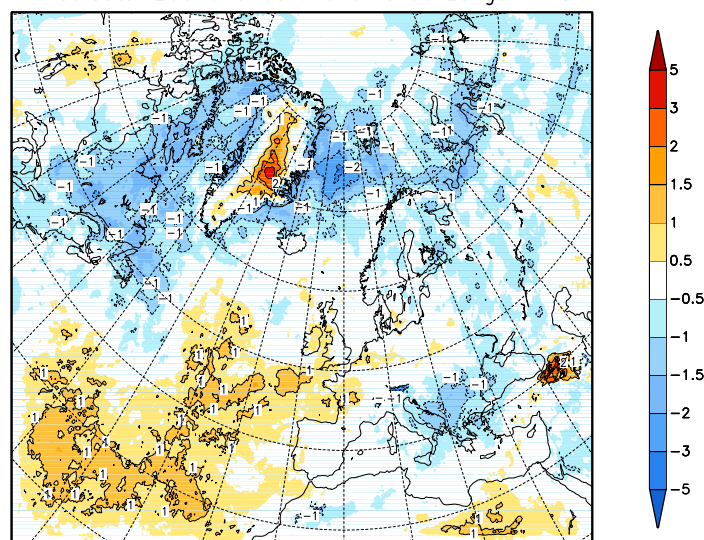


Figure 6: An example of the field verification product: bias of temperature on 850 hPa level in January 2007.

Climat. of Fraction of ice % Ident: ECM701 NO: 28
 First date: 2006060100 Init. time 00 Length +48

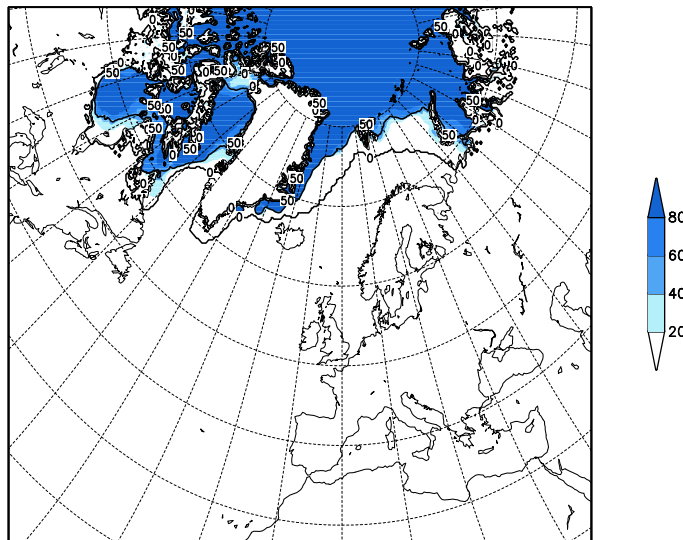


Figure 7: An example of the field verification product: mean ice concentration in +48 hours' forecasts in June 2006.

Climate diff: Surface Pressure hPa Ident: ECM701-ECM71
 NO: 28 First date: 2006060100 Init. time 00 Length +48

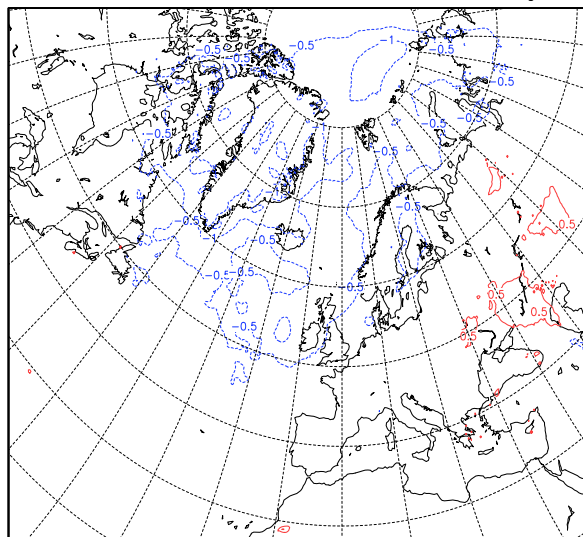


Figure 8: An example of the field verification product: the difference in mean surface pressure of +48 hour forecasts, showing the difference drift of surface pressure in the two experiments.