

SMHI operational HIRLAM system

HIRLAM ASM 2002-04-03 – 2002-04-05

Copenhagen

Lars Meuller

1. Computer system

HIRLAM at SMHI is run at a SGI 3800 at the National Supercomputer Centre (NSC) at Linköping university. The SGI at NSC has 96 PE's with 96 GB shared memory, 96 Gflops peak performance, IRIX operating system and LSF batch system. For operational Hirlam only 24 PE are used due to bad speed-up on more PE when using MPI for parallellization.

A complete HIRLAM system is also run in a hot backup mode on a CRAY T3E with 272 PE's. For operational use 100 or 110 PE's are used. If the SGI is down the production can easily be switched to the T3E by just issuing one operator command.

Pre- and postprocessing are run at SMHI on high availability DEC/Alpha servers. No real archiving is done from the Unix environment. VAX/VMS tapes are written each day with model and postprocessed fields. A tape robot has been installed and Unix archiving is done on a more ad-hoc basis at present.

Communications between the NSC CRAYs and SMHI Unix system is provided through a dedicated ATM link with 10 MB capacity. As a backup for the communications the university net SUNET with a capacity of 10 MB (100 MB partly) can be used.

Initiating operations on the CRAY system is for security reasons not a straightforward operation. To trigger operational runs of HIRLAM and file transfer from SMHI the cevent facility in NQS is used. For file transfer from NSC to SMHI and for starting e.g. the postprocessing at SMHI UNIX system ordinary ftp and rsh (remote shell) is used.

2. HIRLAM system configuration

In both the operational suite at the SGI and the backup suite at the T3E two domains with different horizontal resolution are run. HIRLAM 44 has a 0.4° resolution over a rotated grid with 202x178 horizontal gridpoints. It uses ECMWF boundary conditions 4 times a day from ECMWF BC project. A nested HIRLAM 22 is run at 0.2° resolution over 162x142 gridpoints and with boundaries from the 44 model. Both versions have the same 31 model levels. Model files are written every hour and sent to SMHI UNIX system where the postprocessing are run.

The BOBA sea ice model and pseudo SST observations from manually analysed Baltic SSTs are used in the surface analysis.

The statistics files are produced operationally and they are converted to formatted files so that they can be used on the workstations for a graphical monitoring tool and for accumulating and plotting RMS observation statistics. These are essential tools for monitoring of the operational runs and for evaluating modifications.

SYNOP, AIREP, AMDAR, BUOY, TEMP and PILOT observations enter the analysis using a version of the ECMWF observation pre-processing system to convert from the WMO alphanumeric code forms used on GTS to BUFR format.

Operational run at SGI

In the operational HIRLAM run at SGI the scripts and forecast model is version 5.0.0 but with Sundqvist scheme and Louis vertical diffusion. MPI parallelisation is used. Semi-Lagrangian integration is used with 12 and 6 min time step for H44 and H22 resp.

The analysis scheme is 3DVAR version 4.4.0 with MPI.

Both the forecast and analysis model are run on 24 PE. It has been difficult to see any speed-up above 16 – 24 PE both with MPI and SHMEM parallelisation.

Backup run at T3E

The forecast model run on the T3E is the version parallelised in a joint effort by SMHI , FMI , INM and CRAY using SHMEM based mainly on 2.5 physics.

The interpolation is Semi-Lagrangian with 7.5 and 4 min time step resp. for HIRLAM44 and HIRLAM22.

The analysis model is the version parallelised by Nils Gustafsson and Deborah Salmond using SHMEM.

The scripts and support libraries are from 4.3.4, the first Y2K proof version.

The interpolation of the boundaries are done inside the forecast model which is a necessity on a T3E due to the weaknesses with slow execution of script commands and slow start of executables.

In order to secure that HIRLAM is run with highest priorities on the machine we also explicitly have to run all HIRLAM executables with mpprun directly so the script Boot is not used on T3E.

Still there is a big disadvantage on the T3E with poor I/O although much work has been done particularly in the analysis code to minimise unnecessary I/O which gave a noticeable speed-up.

Operational schedule

At every 3 hours +0h25m a preliminary H44 analysis is run using mainly SYNOP data. The output is mainly used for automated analysis of weather charts.

Every 6 hours, HIRLAM 44 is run with a cut off of 1h55m. The forecast is run out to 48 hours and then HIRLAM 22 is started, and the forecast is run to 48 hours.

The clock time for the complete H44 run is c:a 48 min on T3E and c:a 21 min on SGI while the H22 run is completed in 38 min (T3E) and 18 min (SGI).

3. HIRLAM supervision

During 1999 SMHI implemented a new system KARO for supervising operations at SMHI. KARO is commercial software from a Swedish company PRONYX AB and from summer 1999 it is used for supervising the operational HIRLAM operations.

It is used to help the operators give early warnings if some parts of both the software and hardware of the rather complex operational environment is not working properly. All essential events and existence of files on all platforms and data amounts are monitored throughout the 24 hours.

Thus the reception of boundary data from ECMWF, the result of the preprocessing of observations as well as the running of the BOBA sea ice model are supervised on SMHI UNIX system and so are the reception of these data at NSC CRAY system. The start, end and eventual failure of the HIRLAM executions on T3E are monitored.

The reception of HIRLAM modelfile output at SMHI UNIX system and the output of postprocessed files at SMHI and also the creation of certain products that are generated from HIRLAM data are supervised in the KARO system.

In addition to supervising the output of the system, the system itself, the T3E and SMHI UNIX system and the amount of available discspace are supervised together with the the communication link.

4. Events

During last year only minor changes were made to the operational HIRLAM system at SMHI. In June 2001 the 3-D HIRVDA became operational and in the end of the year SMHI decided to use the SGI 3800 as the main machine and let the T3E act as backup machine. This was done when the HIRLAM version on SGI was upgraded from 4.7.3 to 5.0.0 to be able to utilize boundaries from ECMWF BC project that only is available as 'frames'. This work was completed in Nov 2001.

5. Coming work

For a long time HIRLAM has suffered from rather bad verification scores in comparison with other models as the ECMWF model for parameters like screen temperatur and humidity in particular during spring and early summer. SMHI still uses rather old physics in certain areas (Sundqvist) and there has at SMHI been some concern about when results from the HIRLAM research and development could become useful in operational model runs.

In the autumn 2001 a project, Hirlam X, was started and some considerable resources were allocated to the subject of upgrading the SMHI operational HIRLAM model.

Much work was done during the autumn/winter and after testing, considerations and discussion we decided on a system that should undergo thorough testing and validation during 4 different months under all 4 seasons. The testmonts had been decided together with FMI and a common domain has also been decided that should be used in the future. The domain has 0.2 degree resolution and is 438 x 310 gridpontos with 40 levels that in the future should replace the current nested system.

Due to, at present, limited computer resources the monthly test runs has been done with a different area, 278 x 238 gridpoints with 0.3 degree resolution. These runs are now evaluated and should lead to a decision if further modifications to the system is needed.

The system that has been decided to be validated is basically HIRLAM 5.1.4 with:

- hirvda version 5.0.3
- Kain-Fritsch/Rasch-Kristiansson cloud physics
- 40 vertical levels that differs from the reference system. In particular, the lowest level is higher up, c:a 28 meter
- changes to ISBA. Soil freezing removed from land type 4 and 5 (low veg. and forest)
- changes to the cloud fraction routine CLDFRC.f
- Filtered orography

Initial tests has been encouraging. The T2m bias during the spring conditions has been removed and humidity forecasts has been significantly better but the validation continues at present.

In addition to the 4 selected monts, the Hirlam X system has also been put in a semi-operational mode to run with the same observational and boundary conditions as the operational suite and is continually validated.

There is also technical changes that will involve work to be done at SMHI. NSC has decided to replace the CRAY T3E with a PC-Cluster they are going to built themselves. A first test cluster has now been built and work is going on to put up the HIRLAM system and get it running. This first cluster consists of 4 nodes each with 2 processors and a front-end node also with 2 PE. The processors are ADM Athlon MP 2000+ with 1.67 MHz and the network connecting the nodes are Fast Ethernet and SCInet. Myrinet will also be tested. Libraries for communication are ScaMPI, Mpich and LAM. The cluster are using PBS batch system. This cluster will theoretically be able to replace the T3E for the Hirlam backup system and if it works well NSC can close down the T3E. Later a bigger cluster will be built for operational Hirlam exclusively that will be sufficient for the next setup with higher resolution and 4DVAR.