SMHI operational HIRLAM system

HIRLAM ASM 2001-05-07 – 2001-05-09
Reykjavik

Lars Meuller

1. Computer system

HIRLAM at SMHI is run on the CRAY T3E at the National Supercomputer Centre (NSC) in Linköping. The T3E at NSC has 272 PEs and SMHI finances 23% of the available resources at NSC. About 9% are used for operations. The operational forecast uses 100 or 110 PEs. A complete HIRLAM system is also run in a hot stand-by mode on the SGI 3800 with 96 PE, 96 Gflops, 96GB shared memory, IRIX operating system and LSF queuing system. If the T3E is down the production can easily be switched to the SGI by just issuing one operator command. The T3E is rather stable, but there have been a few occasions of switching to SGI for a run or two.

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 T3E and the backup suite at the C90 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 from both the 00UTC and 12UTC runs. The 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 alphamuneric code forms used on GTS to BUFR format.

Operational run at T3E

The forecast model run operationally 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.

Backup run at SGI

In the backup version of HIRLAM run at SGI the forecast model is version 4.7.3 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.

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 36 hours.
The clock time for the complete H44 run is c:a 43 min on T3E and c:a 25 min on SGI while the H22 run is completed in 35 min (T3E) and 22 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

The main task during 2000 was to install the backup version of HIRLAM on the new SGI 3800 computer that NSC had chosen to replace the CRAY C90. This task was completed 2001-03-05.
One final task for the installation to be complete is to replace the present 400 MHz processors with 500 MHz and that will be done in June 2001.

From 2000-10-29 the 3DVAR analysis is operational on the backup HIRLAM at SGI.

5. Plans

Work is going on to introduce 3DVAR as the analysis model in the HIRLAM on the T3E and will hopefully take place in June 2001.

The next HIRLAM version to be installed and tested will be the version with the ISBA scheme and the new surface analysis.

If that proves to be an improvement we plan to make SGI to be our operational machine and let the T3E act as backup computer.