ROYAUME DU MAROC

SECRETARIAT D'ETAT AUPRES DU MINISTERE DE L'ENERGIE, DES MINES, DE L'EAU ET DE L'ENVIRONNEMENT CHARGE DE L'EAU ET DE L'ENVIRONNEMENT

NUMERICAL WEATHER PREDICTION IN MOROCCO 2010

ALADIN TEAM

Direction de la Météorologie Nationale, Casablanca, Morocco

Summary of highlights (April 2010/April 2011)

I- New Computing System "IBM HPC"
II- Benchmarking and optimisation
III- New Operational suites
- increasing resolution, switching to Cy36t1 (ALADIN)
- daily run of AROME (North of morocco)
- 3DVAR
some assimilation studies (see second poster)

I- The New Computing Platform

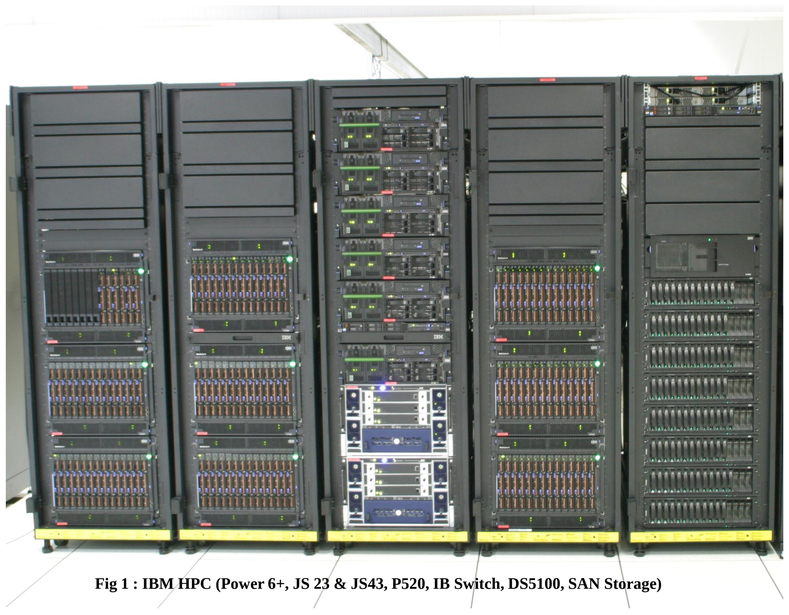


Fig 1 : IBM HPC (Power 6+, JS 23 & JS43, P520, IB Switch, DS5100, SAN Storage)

. **9 Physical Blade Center H :**

114 shared memory nodes : 4 cores eatch, 16GB memory
2 shared memory nodes : 8 cores eatch, 32GB memory
CPU : RISC/UNIX IBM Power6+ @4.2 GHz

. **6 p520 network-I/O nodes, 8 cores, 16GB Memory**
. **2 Switch InfiniBand for I/O and MPI**

~475 core in total
~ 1.95 TB memory,
~ 8.3 Tflops theoretical peak performance for application
~ 52 TB disk space

II- Benchmarking and optimisation

Lots of compilation and linking issues

a- I/O time step problem

In ALADIN model performance beyond 32 nodes decreases. I/O time steps (reading boundary conditions and writing meteorological fields) performance diminishes when we add more nodes. The computation time steps, instead, saw their performance increasing (Figure 2).

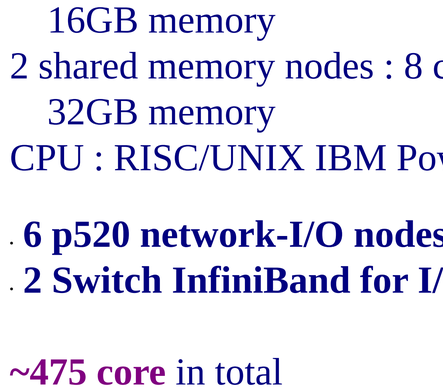


Figure 2 : Real Time during integration

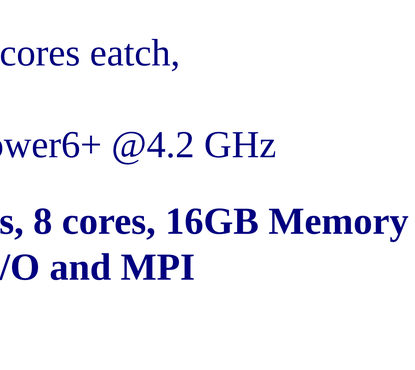


Figure 3 : Real Time if Tuning I/O

If the scalability of I/O in ALADIN is recognized low by most users in massively parallel computers, special features related to IBM GPFS make tough the situation.

To resolve this we make minor changes in the code to enable eatch MPI task to read and write its own files fromits own directory and not from a shared one.

As we can see in Figure 3 when moving from 32 to 54 nodes, we get a very good acceleration calculations. However, the explosion of I/O time destroys the gain of parallel performance. We also see that simple modifications to the directories of MPI tasks allowed us to keep the I/O time substantially constant.

b- Improvement

For more improvement investigation has been led on some part of ALADIN dedicated to treating inputs/outputs. We found that when the dedicated tasks (NSTRIN/NSTROUT) exchanged their data with other computing tasks, the observed rate of exchange was often satisfactory (between 500 and 7000 MB/s). However, few messages got their speeds very low (0.01MB/s !). Those rare messages led to a dramatic increase in time spent in processing input/output.

The idea is to give more CPU cycles to the I/O tasks by switching from SMT (Simultaneous Multi Threading) mode to ST (Single Threading). Performance time steps of input/output has been greatly improved, confirming the thesis of a lack of CPU resources during sessions exchange of messages.

The solution then appeared : ALADIN running in mixed mode (MPI + OpenMP=4), we adjusted the variable XLSMPOPTS = spins = 0: yield = 0 to be **XLSMPOPTS = spin = 1: yield = 1.**

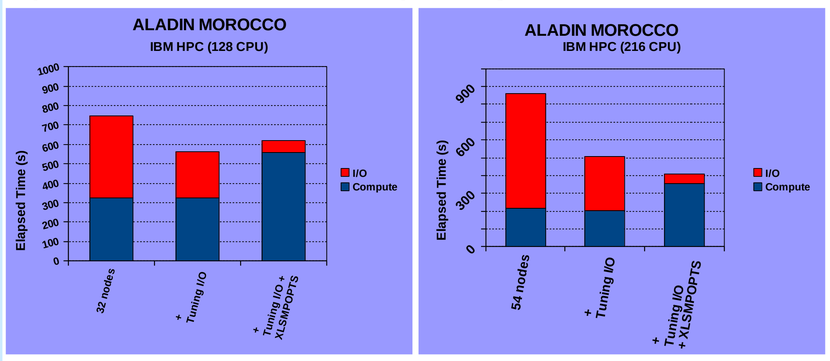


Figure 4 : XLSMPOPTS of the OpenMP run-time greatly improve the performance

Variable XLSMPOPTS of the OpenMP run-time allows us to adjust the way how inactive threads are waiting to be assigned the work.
With spin = 0, the inactive OpenMP threads are in a mode called "busy wait". It spin on a lock while consuming hardely the CPU. This mode is the most reactive because once the lock is released by another thread, the thread "busy wait " catch the lock without waiting.
The other modes are "yield" when the thread releases the CPU voluntarily and "sleep"when the thread is inactif. These modes save the CPU but require slightly more work to wake sleeping threads. By switching to spin = 1 and yield =1 we order threads to sleep when they have nothing to do, which allows the thread 0 of each task to have more CPU resources to work (for example to send and receive messages). The result is spectacular, as can be seen in the figure 5 below.

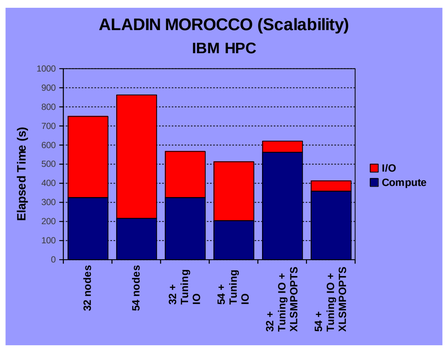


Fig 5 : Results of introducong I/O Tuning and variable XLSMPOPTS modified

We applied the same changes to ALARO and AROME and we have obtained very substantial gains.

C- Conclusion

1- Two minor changes have pushed the limits of scalability :
- each MPI task reads and writes its own files from a propre directory
- Choice for the variable XLSMPOPTS promoting the release of CPU resources by threads waiting for work.

2- These changes are even more important when the number of nodes involved is high.

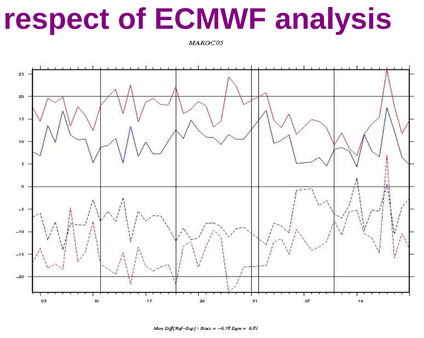
III- New operational suites

a- ALADIN-MOROCCO :

Cycle: cy 36t1
Characteristics :
Hydrostatic
Semi-implicit semi-lagrangian two-time-level scheme; DT=450s
2 runs / day 00, 12 : 72 hrs forecast range
Boundary conditions from ARPEGE (3 hrs coupling frequency)
domain : 250x250 points, Dx=10Km (Lambert Projection – linear grid)
60 vertical levels

Scores 2009/2010 suites in respect of ECMWF analysis

Z at 500 hpa : 12H



Z at 850 hpa : 36 H

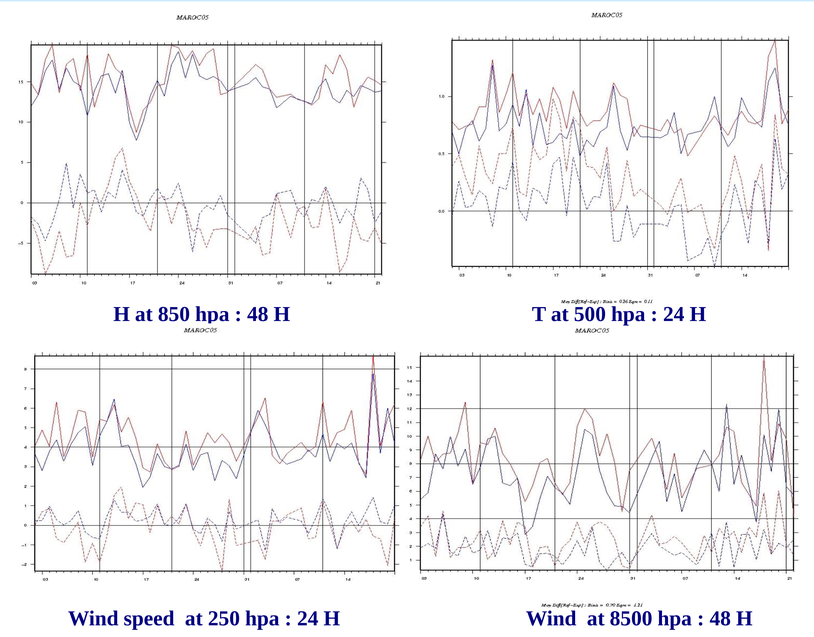


Figure 6 : 48 Cases (Jan-Fed 2010), **---- EQM ----, ...Biais....., ------ Cy36t1, ----- Cy29t2**

b- AROME (North of morocco, test suite) :

Cycle: cy 36t1
Characteristics :
NON-Hydrostatic
Semi-implicit semi-lagrangian two-time-level scheme; DT=60s
2 runs / day 00, 12 : 24 hrs forecast range
Boundary conditions from ALADIN-MAROC (1 hrs coupling frequency)
domain : yyyxyyy points, Dx=2.5Km (Lambert Projection – linear grid), 60 vertical levels

AROME-NORDM (Rabat : 15 Sept 2009)

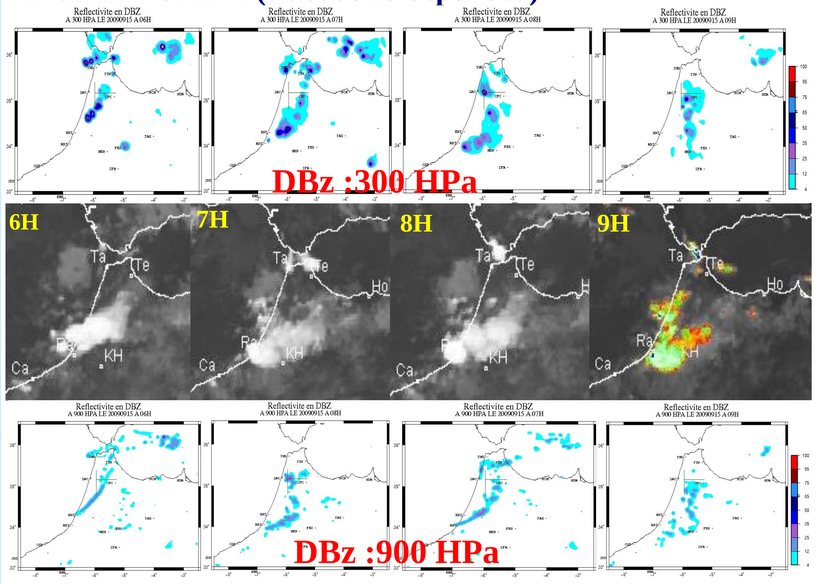


Fig 7 : AROME Radar simulator : Dbz at 300 Hpa and at 900 Hpw

(Casablanca 09-10 Sept 2009)

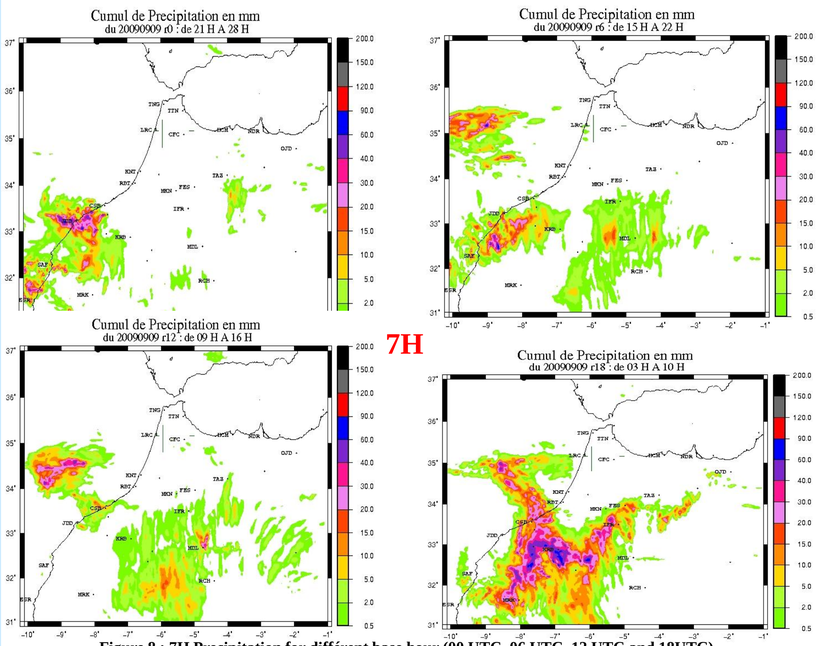


Figure 8 : 7H Precipitation for différent base hour (00 UTC, 06 UTC, 12 UTC and 18UTC)

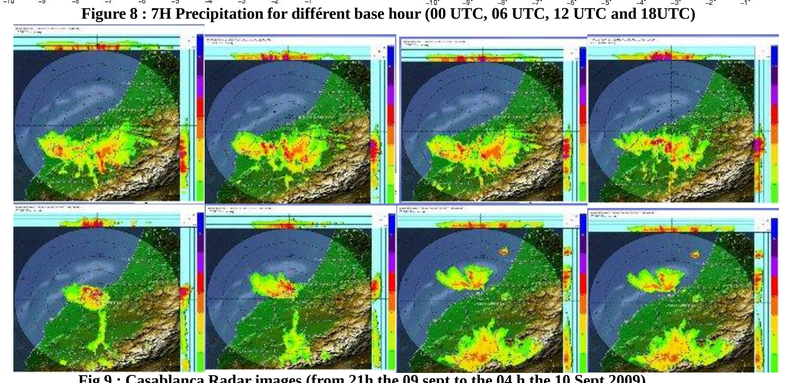


Fig 9 : Casablanca Radar images (from 21h the 09 sept to the 04 h the 10 Sept 2009)

Conclusion and perspective :

- The new HPC paltforme has allowed MAROC METEO to :
  - significantly improve the scores of the operational model ALADIN MOROCCO.
  - run the NH model AROME twice a day with a resolution of 2.5 Km
  - To conduct research activities in variational assimilation (See the 2nd post)
- Engage in more NWP developments in ALADIN consortium