



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación



Atmospheric modeling: introduction, some applications and computational demands

Kim Serradell
kim.serradell@bsc.es

Barcelona Supercomputing Center-Centro Nacional de Supercomputación
Earth Sciences Department. Barcelona.

Aules d'Empresa 2011 – Facultat d'Informàtica de Barcelona - 25 de gener 2011

Outline



- **Presentation**
- **Introduction**
- **Models in BSC**
- **Running Models**
- **Examples of work**
 - **Parallelizing**
 - **Optimizing**
 - **Bechmarking**
- **Further directions**

Presentation



- **Made my education in FIB.**
- **Finished on 2005.**
- **Then working in different places.**
- **And two years ago, I went to the BSC and started to work on Earth Sciences Dpt.**

Introduction



- **Atmospheric Model: Mathematical Models with dynamic equations**
- **We need computers to solve these equations**

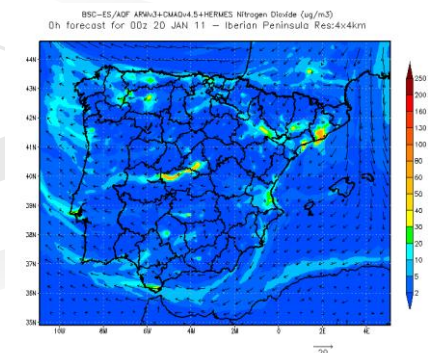
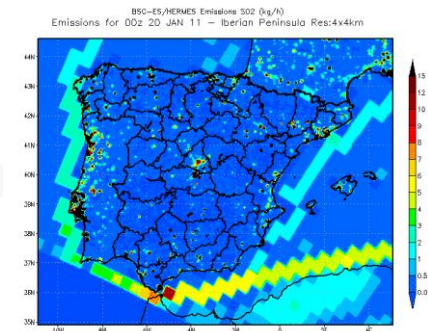
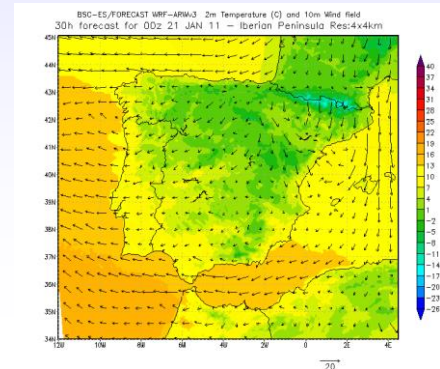


- **Before getting this nice picture, we need to follow a “forecast chain”.**
- **Today we will explain this from the IT point of view.**

Models in ES-BSC

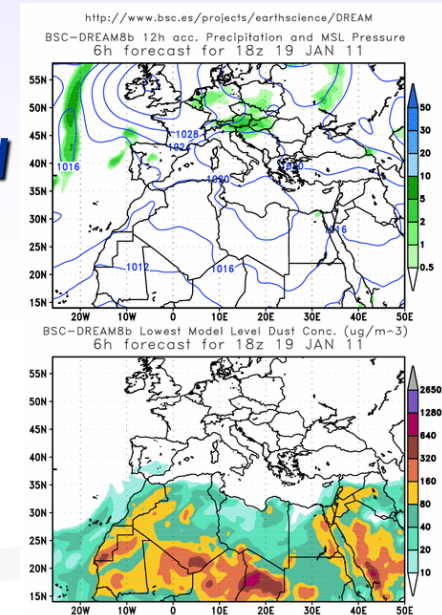


- **Meteorological Modeling**
 - **WRF: Weather Research Forecasting**
 - Fortran Code
 - MPI, OpenMP and CUDA
- **Emissions**
 - **HERMES: High-Elective Resolution Modelling Emissions System**
 - C Code
 - Not parallel
- **Air Quality Forecasting**
 - **CMAQ: Community Multiscale Air Quality**
 - Fortran Code
 - MPI

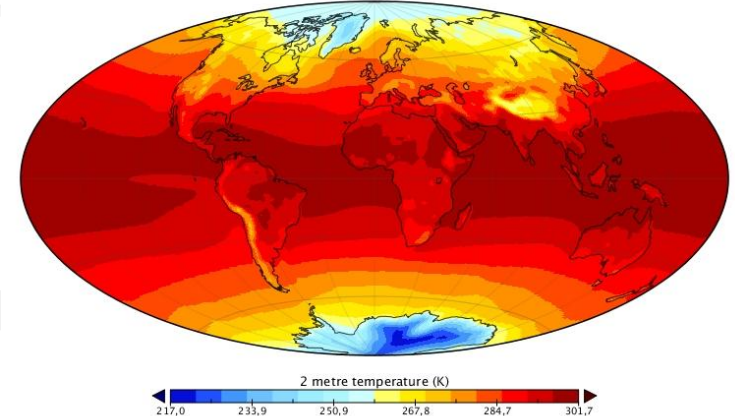


Models in ES-BSC

- **Mineral Dust Modeling**
 - **BSC-DREAM8b: Dust REgional Atmospheric Model**
 - Fortran Code
 - Not parallel
- **Climate Change**
 - **EC-EARTH**
 - Fortran, C
 - MPI, OpenMP



Mean Surface Temperature (1990-1999) for EC-EARTH at ES-BSC



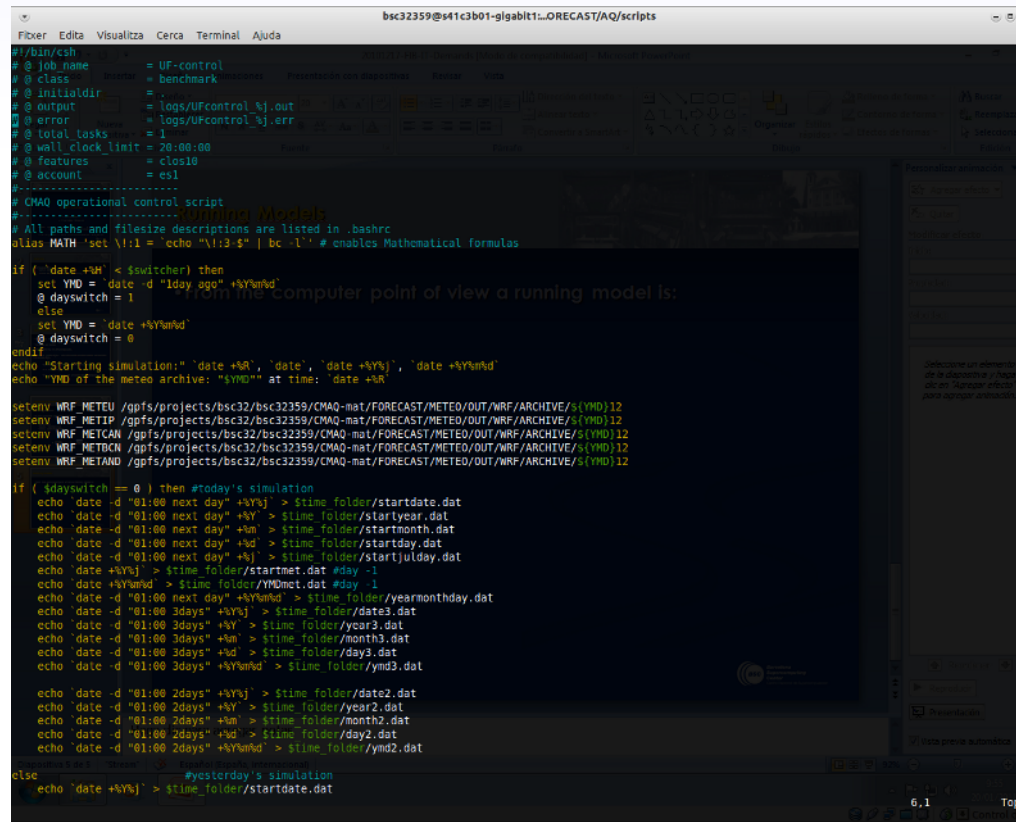
Computational Demands



- Which domains are we simulating ¿?
 - Barcelona
 - Catalunya
 - Spain
 - World
- Which resolution ¿?
 - 1 km² 4 km² 12 km² 50 km²
- Increasing this parameters, increases the system constraints
 - Computation Needs (CPU's, Memory Bandwith...)
 - Data Storage
- Define this parameters in function of your hardware and time to serve forecast.

Running Models

- From the computer point of view a running model is:



```
#!/bin/csh
#@ job_name = UF-control
#@ class = benchmark
#@ initdir = /
#@ output = logs/UFcontrol_%j.out
#@ error = logs/UFcontrol_%j.err
#@ total_tasks = 1
#@ wall_clock_limit = 20:00:00
#@ features = clos10
#@ account = es1

# CMAQ operational control script
#-----Running Models-----
# All paths and filesize descriptions are listed in .bashrc
alias MATH 'set \!1 = `echo "\!2-$" | bc -l` # enables Mathematical formulas

if { `date +%M` < $switcher } then
    set YMD = `date -d "1day ago" +%Ym%d`
    @ dayswitch = 1
else
    set YMD = `date +%Ym%d`
    @ dayswitch = 0
endif
echo "Starting simulation: " `date +%M`, `date`, `date +%Y%j`, `date +%Ym%d`
echo "YMD of the meteo archive: " $YMD " at time: date +%M"

setenv WRF_METEUF /gpfs/projects/bsc32/bsc32359/CMAQ-mat/FORECAST/METEO/OUT/WRF/ARCHIVE/s($YMD)12
setenv WRF_METIP /gpfs/projects/bsc32/bsc32359/CMAQ-mat/FORECAST/METEO/OUT/WRF/ARCHIVE/s($YMD)12
setenv WRF_METCAN /gpfs/projects/bsc32/bsc32359/CMAQ-mat/FORECAST/METEO/OUT/WRF/ARCHIVE/s($YMD)12
setenv WRF_METBCN /gpfs/projects/bsc32/bsc32359/CMAQ-mat/FORECAST/METEO/OUT/WRF/ARCHIVE/s($YMD)12
setenv WRF_METAND /gpfs/projects/bsc32/bsc32359/CMAQ-mat/FORECAST/METEO/OUT/WRF/ARCHIVE/s($YMD)12

if { $dayswitch == 0 } then #today's simulation
    echo `date -d "01:00 next day" +%Y%j` > $time_folder/startdate.dat
    echo `date -d "01:00 next day" +%Y` > $time_folder/startyear.dat
    echo `date -d "01:00 next day" +%m` > $time_folder/startmonth.dat
    echo `date -d "01:00 next day" +%d` > $time_folder/startday.dat
    echo `date +%Y%j` > $time_folder/startnet.dat &day -1
    echo `date +%Ym%d` > $time_folder/YMDnet.dat &day -1
    echo `date -d "01:00 next day" +%Ym%d` > $time_folder/yearmonthday.dat
    echo `date -d "01:00 3days" +%Y%j` > $time_folder/date3.dat
    echo `date -d "01:00 3days" +%Y` > $time_folder/year3.dat
    echo `date -d "01:00 3days" +%m` > $time_folder/month3.dat
    echo `date -d "01:00 3days" +%d` > $time_folder/day3.dat
    echo `date -d "01:00 3days" +%Ym%d` > $time_folder/ynd3.dat

    echo `date -d "01:00 2days" +%Y%j` > $time_folder/date2.dat
    echo `date -d "01:00 2days" +%Y` > $time_folder/year2.dat
    echo `date -d "01:00 2days" +%m` > $time_folder/month2.dat
    echo `date -d "01:00 2days" +%d` > $time_folder/day2.dat
    echo `date -d "01:00 2days" +%Ym%d` > $time_folder/ynd2.dat
else
    #yesterday's simulation
    echo `date +%Y%j` > $time_folder/startdate.dat
```

A LARGE SCRIPT

Only the main control has more than 1600 lines.

Three principal steps

- Getting initial data and boundary conditions

- Via FTP
- From other institutions

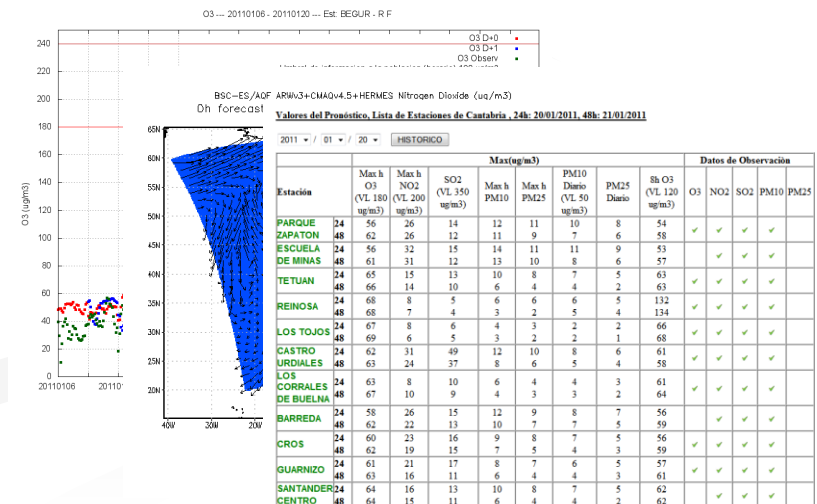
- Running Models

- Intensive calculations

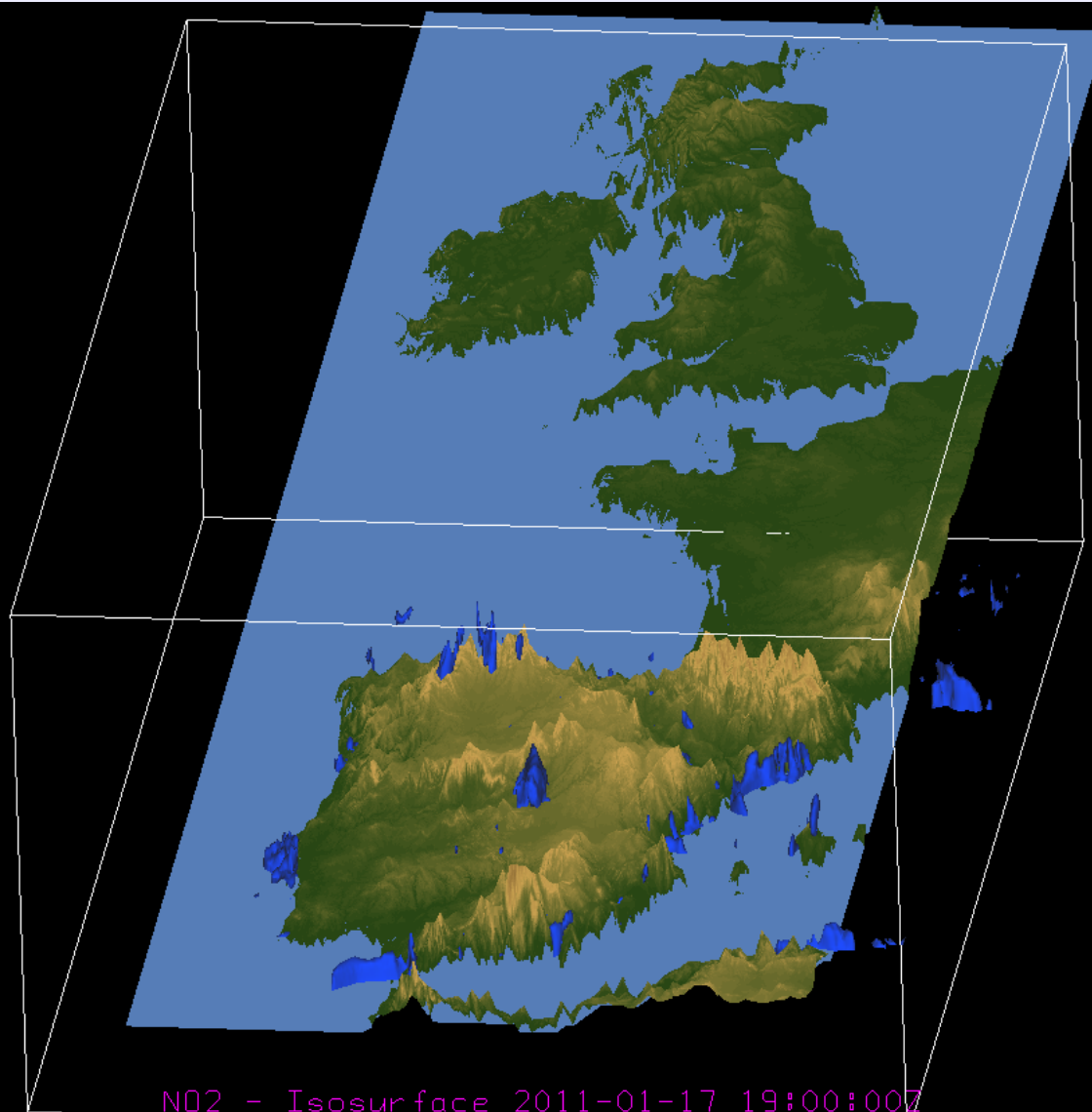
```
bsc32359@s41c3b01-gigabit1~  
Fitxer  Edita  Visualitza  Cerca  Terminal  Ajuda  
ece-hr-2.2.2  bsc32534  Idle  136  8:00:00  Thu Jan 20 08:35:11  
ecearth2.2  bsc32534  Idle  84  1:00:00:00  Thu Jan 20 09:10:36  
mv SUN  bsc32185  Idle  1  1:00:00:00  Thu Jan 20 10:38:19  
hsm_ECE  bsc32534  Idle  1  2:00:00:00  Thu Jan 20 10:55:33  
  
5 blocked jobs  
Total jobs: 20  
bsc32359@login1:~> mng  
  
active jobs-----  
NAME      JOBID      USERNAME      STATE  PROCS  REMAINING      STARTTIME  
FE09-Rest- 3080994    bsc32359      Running 128    2:12:49  Thu Jan 20 08:50:37  
Info-extra 3081082    bsc32359      Running 1      3:36:05  Thu Jan 20 10:43:53  
DRS_FILL  3080698    bsc32359      Running 4      4:04:23  Thu Jan 20 00:12:11  
DRS_FILL  3080859    bsc32359      Running 4      4:05:28  Thu Jan 20 05:13:16  
control200 3078969    bsc32359      Running 1      1:22:26:57 Wed Jan 19 09:34:45  
HSMcctmEUC 3081029    bsc32359      Running 1      2:22:45:14 Thu Jan 20 09:53:02  
  
6 active jobs      139 of 10124 processors in use by local jobs (1.37%)  
                    2516 of 2531 nodes active (99.41%)  
  
eligible jobs-----  
NAME      JOBID      USERNAME      STATE  PROCS  WCLIMIT      QUEUETIME  
  
0 eligible jobs  
  
blocked jobs-----
```

- Publishing results

- Maps
- Plots
- KMZ
- Data to analyze...



Or even 3D...



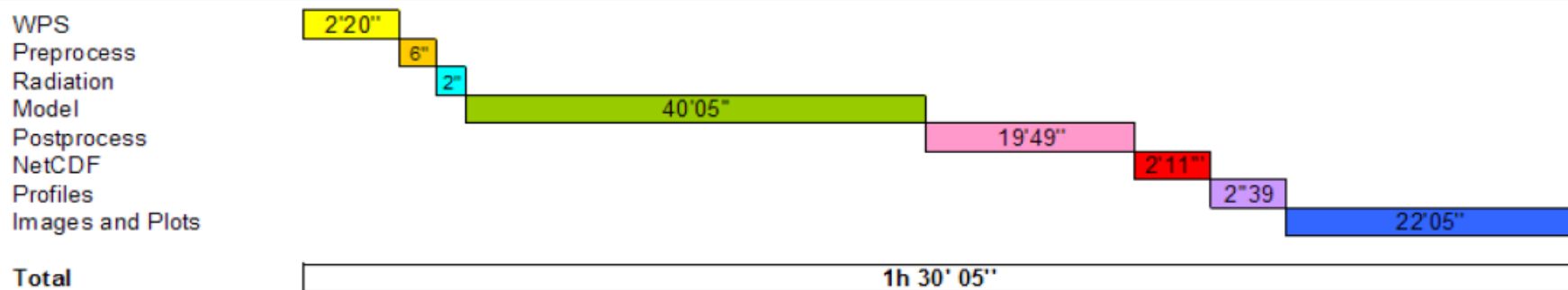
N02 - Isosurface 2011-01-17 19:00:007
Altitude - Color-Shaded Image As Topography

Challenges of the IT



- **Assuring daily execution on the model**
 - Crash recovery
 - Monitoring the Model
 - Assuring transfers
- **Timing the execution**
 - Results have to be on time
- **Data Storage**
 - Huge size of data
 - Storing and cleaning
- **Helping researchers in modeling/running/optimization**
- **And many more...**

Example: Model Chronogram BSC-DREAM8b

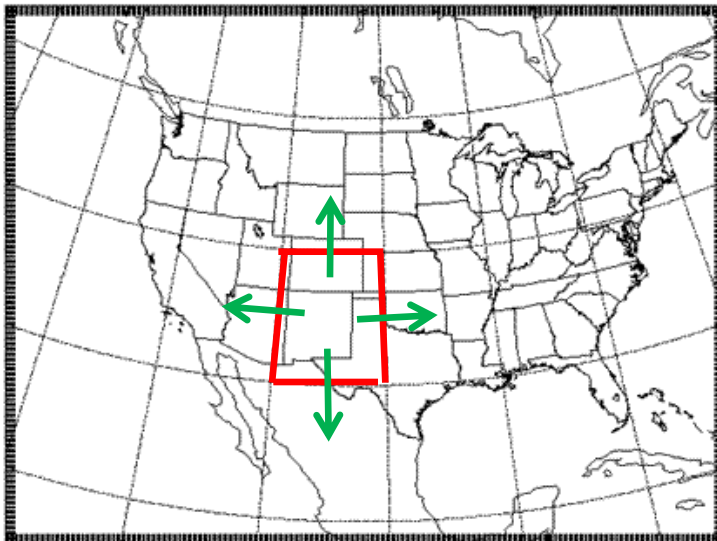


- Different steps involving different executables, data...
- Many failures points
- Data transfers
- Fixed time to run

Parallelizing Atmospheric Models



- We need to be able to run these models in Multi-core architectures.
- What's the way to do it?
- Model domain is decomposed in patches
- Patch: portion of the model domain allocated to a distributed memory node.

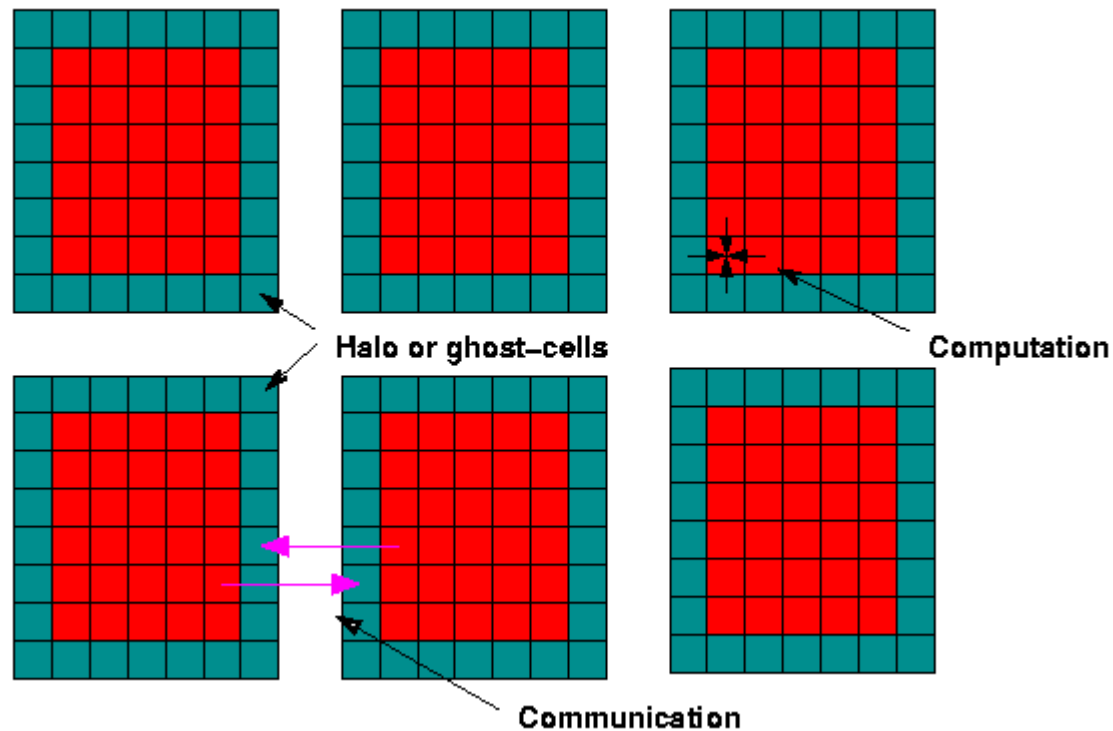


Patch



MPI Communication
with neighbours

Parallelizing Atmospheric Models

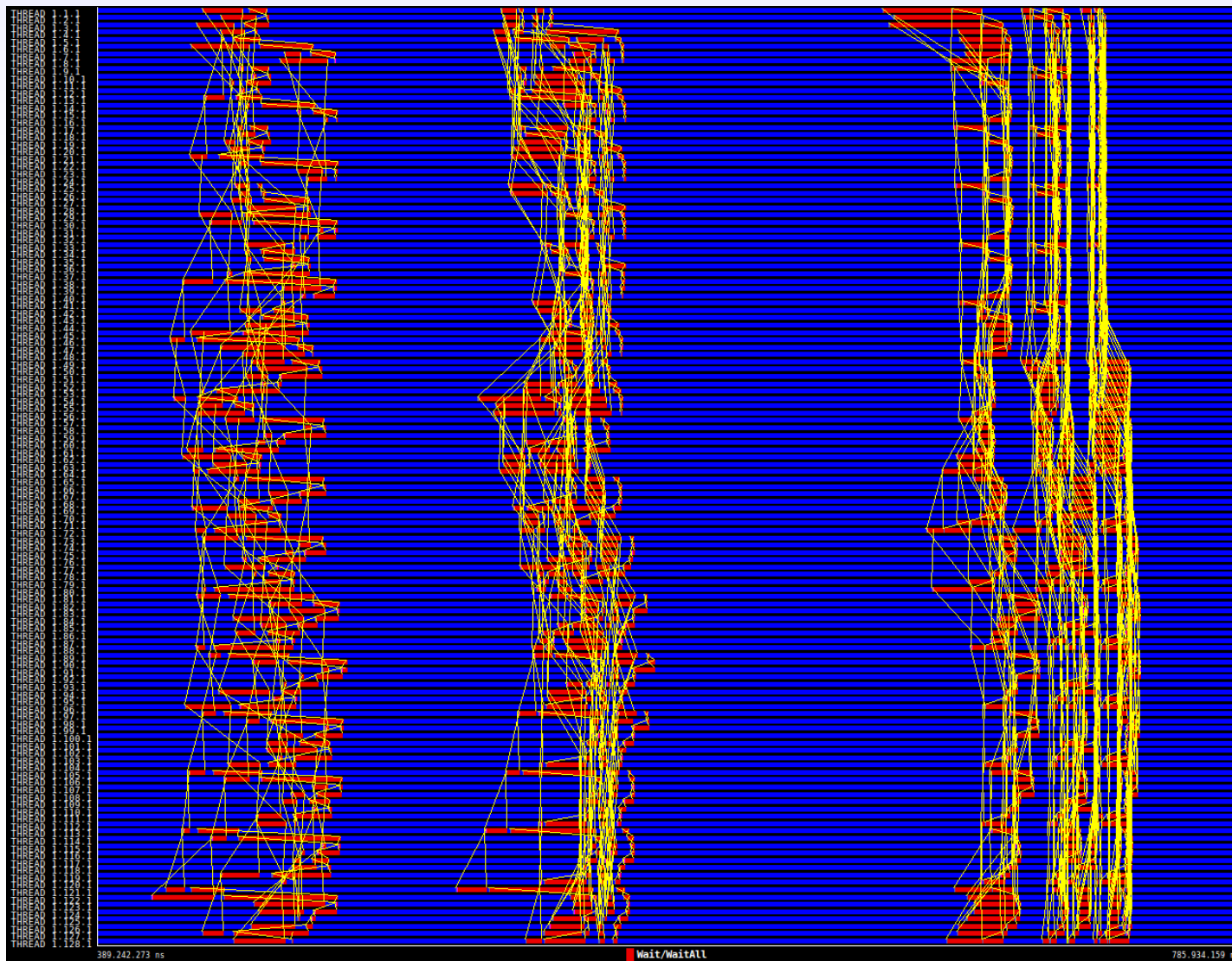


Optimizing Applications



- **Multicore applications**
- **Applications sometimes not scaling correctly.**
- **We need to find the causes and correct it.**
- **But how to do it, in a 128 Cores Run ¿?**
- **These optimizations are highly hardware dependent.**
- **Also, compiler optimizations.**
 - **Which flags are the best for our machine ¿?**
 - **Do I have accurate results ¿?**

Optimizing Applications



Idle
Running
Not created
Waiting a message
Blocking Send
Thd. Synchr.
Test/Probe
Sched. and Fork/Join
Wait/WaitAll
Blocked
Immediate Send
Immediate Receive
I/O
Group Communication
Tracing Disabled
Others
Send Receive

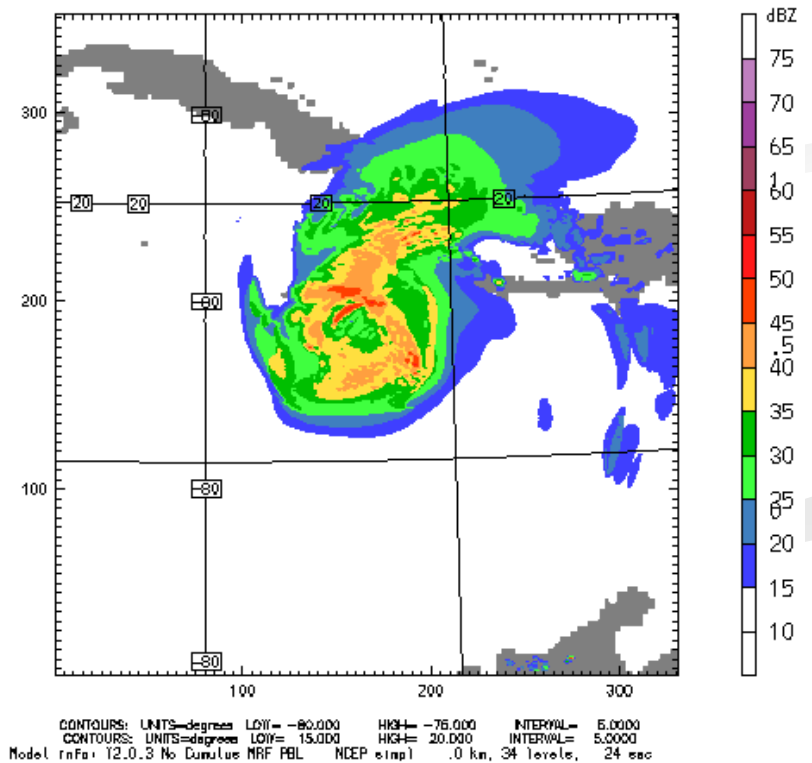
Benchmarking



- Compare our execution with other architectures.
- Test cases are defined:
 - Cases with intense activity
- We want to evaluate:
 - Execution time (speed) in our architecture.

Ivan 12-km WRF
Fcst: 1 h
LAND MASK (1 FOR LAND
Max Reflectivity
Latitude
Longitude

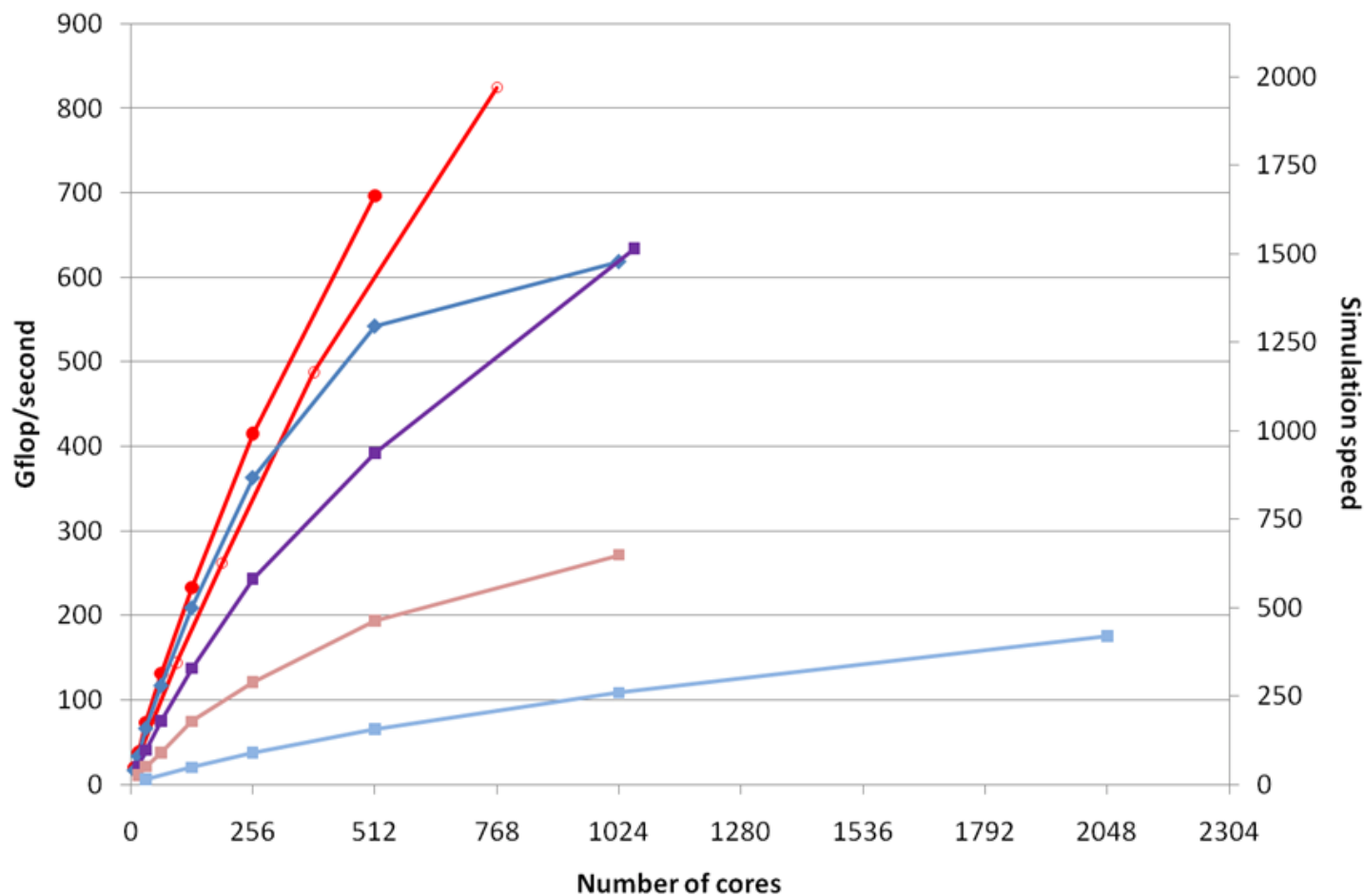
Init: 00 UTC Sat 11 Sep 04
Valid: 01 UTC Sat 11 Sep 04 (19 MDT Fri 10 Sep 04)



Benchmarking



WRFV3 12KM CONUS (Performance per core)

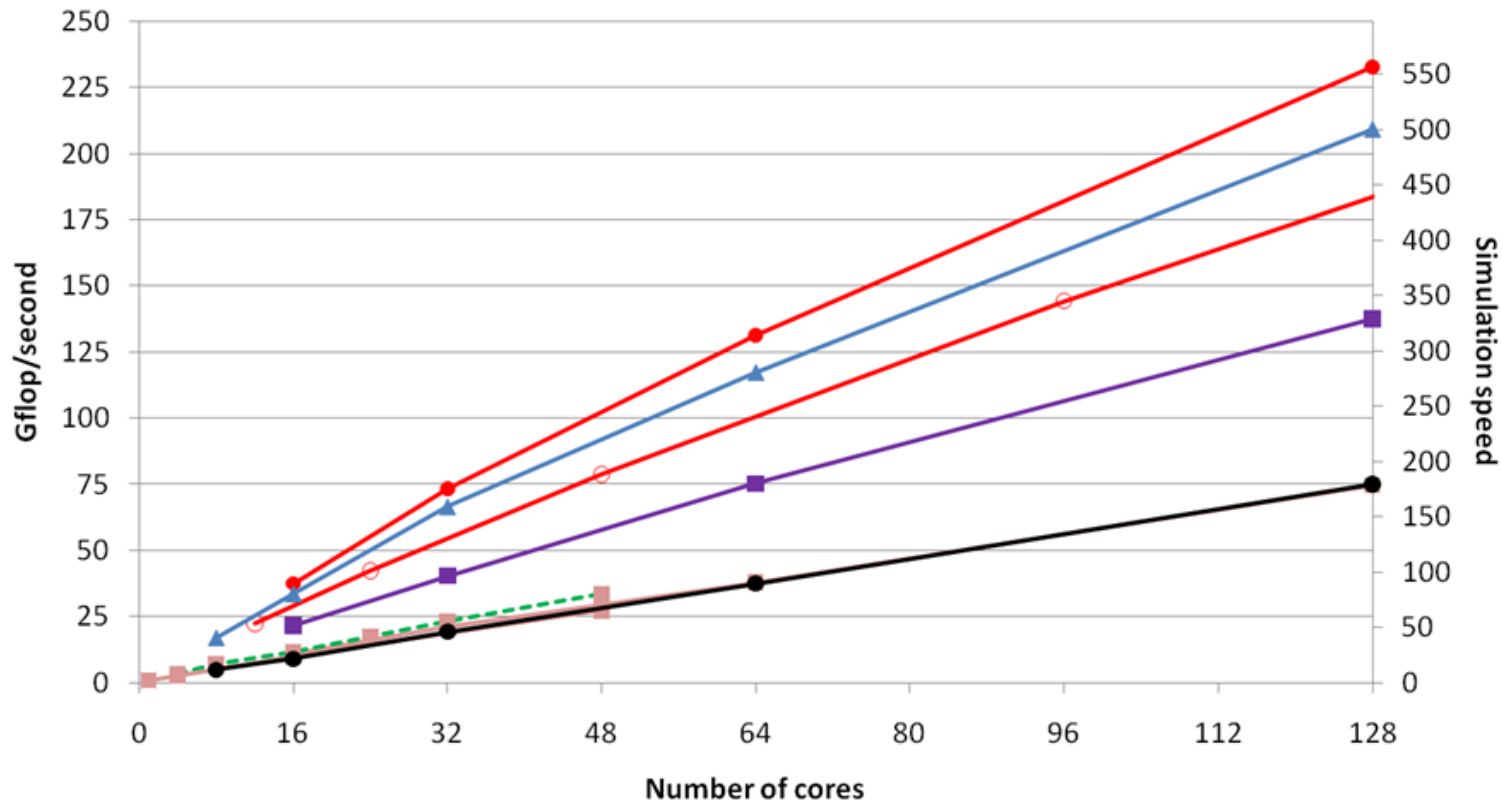


Intel Xeon 5670 (6 core) Intel Xeon 5560 (4 core) IBM Power6 Cray XT5 Sun/AMD (Ranger) BG/P (Surveyor)

Benchmarking



WRFV3 12KM CONUS (Performance per core) -- ZOOMED



Further Directions

- **Petascale supercomputers to Earth Sciences**
- **GAEA: The Fastest Supercomputer for Earth Sciences**
- **In June 2010, installation concluded for a 260-teraflop Cray XT6 system with 2,576 AMD “Magny-Cours” 12-core, 2.1 GHz processors.**
- **In June 2011, a 720-teraflop Cray XE6 system will be added to Gaea. It will employ the next-generation AMD Interlagos 16-core processor.**





- **Using GPU's on Earth Sciences**

- **Graphics processing units are highly parallel, multi-threaded, many-core processors with a very high computational power and memory bandwidth.**
- **Porting code to this new technology.**
- **WRF Community started last year (1% code translated can give 20% of speedup).**





GRÀCIES !