Hirlam System: past, present and … future?

Daniel Santos Muñoz
HIRLAM-C Project Leader for System

Outline

HIRLAM System past and present

- Usability and maintenance
- Common code contributions
- Scalability

HIRLAM System future

Hamonie usability and maintenance

Main HIRLAM Objectives:

- Increase Harmonie-Arome **usability**
- **Maintenance** of new system releases

Actions in HIRLAM:

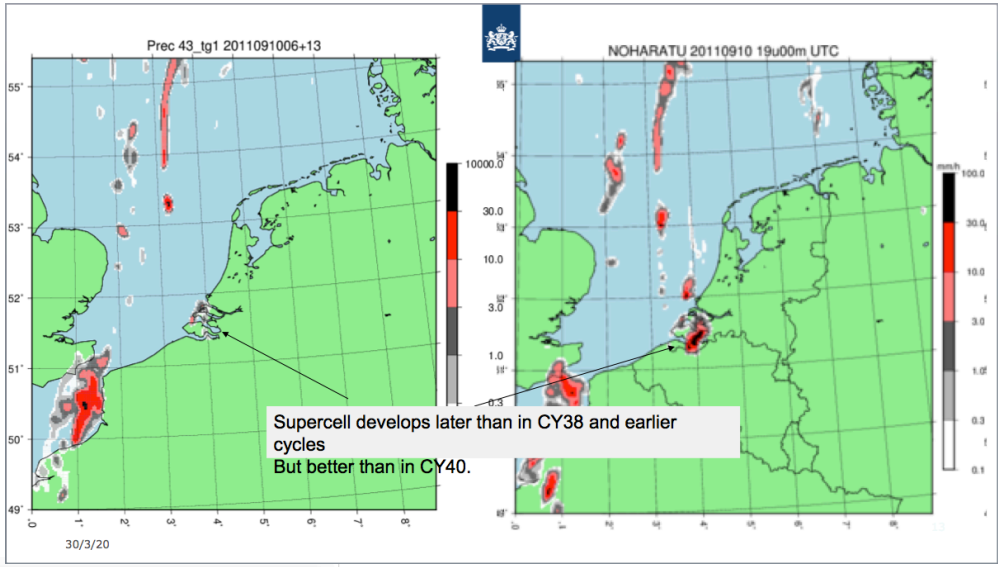
- Explore the possible **revision of the scripting system**: ecFlow, DA flowchart, PrepLAM
- New **modern open source compilation env**: cmake
- Use of more **well-know and modern languages**: python, JSON, TOML
- Use of **git** for repositories: Tutorial, github
- **Training** on: Harmonie, DAskit, Harp, git
- Best **model documentation**: namelist, DA flow chart
- **New tools**: Harp, Titan/GridPP
- **Use of containers and/or virtual box** like MUSC
- **Increase the flexibility of scripting** to cover: Nowcasting, Short-range, EPS, Climate
- **Technical testing and implementation** of new codes: Surfex 8.1 (DA), 4DVar, Phys, SP, Aerosols, HR, Blacklisting, crowd source obs,
- **Common design** of codes: SPP, LAM-OOPS, …

Current Harmonie-Arome development in HIRLAM

Harmonie-Arome release schedule

- Proposed RCR time schedule:
  - **May 1** Tagging harmonie-43h2.1.rc.1
  - **June 10**: RCR reporting about preoperational testing
  - **July 7**: Release harmonie-43h2.1
  - **Sept 15**: Harmonie-43h2.1 operational (det or eps)

    on AEMET and MetCoOp



The following table summarizes the main changes between targets:

Current develop	Target 1	Target 2	Reference cy40h1.1.1
SURFEX			
CROUGH="NONE"	CROUGH="NONE"	CROUGH= "NONE"	CROUGH="NONE"
XRIMAX=0.0	XRIMAX=0.0	XRIMAX=0.0	Not applicable (XRIMAX=0.0)
Maintain Rsmin RSMIN_DECIDUOUS_FACTOR=1 RSMIN_CONIFEROUS_FACTOR=1 RSMIN_C3_FACTOR=1 RSMIN_C4_FACTOR=1	Increase Rsmin RSMIN_DECIDUOUS_FACTOR=1.5 RSMIN_CONIFEROUS_FACTOR=2.75 RSMIN_C3_FACTOR=3.0 RSMIN_C4_FACTOR=1.5	Increase RSMIN RSMIN_DECIDUOUS_FACTOR=1.5 RSMIN_CONIFEROUS_FACTOR=2.75 RSMIN_C3_FACTOR=3.0 RSMIN_C4_FACTOR=1.5	Not applicable (i.e. no factors applied)
OI coefficients for WG2 assimilation (POLYNOMES_ISBA)	Reduced OI coefficients for WG2 assimilation (POLYNOMES_ISBA_MF6)	Reduced OI coefficients for WG2 assimilation (POLYNOMES_ISBA_MF6)	Not applicable (POLYNOMES_ISBA)
ECOCLIMAP_VERSION=2.5_plus	ECOCLIMAP_VERSION=SG	ECOCLIMAP_VERSION=2.5_plus	ECOCLIMAP_VERSION=2.2
XSCALE_H_TREE=1.0	XSCALE_H_TREE=0.658	XSCALE_H_TREE=1.0	Not applicable (i.e. no scaling)
SOIL_TEXTURE_VERSION=FAO	SOIL_TEXTURE_VERSION=SOILGRID	SOIL_TEXTURE_VERSION=SOILGRID	SOIL_TEXTURE_VERSION=FAO
NPATCH=2	NPATCH=2	NPATCH=2	NPATCH=1
LISBA_CANOPY=".FALSE."	LISBA_CANOPY=".FALSE."	LISBA_CANOPY=".FALSE."	LISBA_CANOPY=".TRUE."
SURFEX_LAKES="FLAKE"	SURFEX_LAKES="FLAKE"	SURFEX_LAKES="FLAKE"	SURFEX_LAKES="WATFLX"
SURFEX_SEA_ICE="sice"	SURFEX_SEA_ICE="sice"	SURFEX_SEA_ICE="sice"	SURFEX_SEA_ICE="sice" (default in 40h1.1.1 is none
MICROPHISYCS (Wim)			
STATNW = yes	STATNW = yes	STATNW = yes	Not available
MICROPHISYCS (Karl-Ivar)			
STATNW = yes	STATNW = yes	STATNW = yes	Not available
HGT_QS = no	HGT_QS = no	HGT_QS = no (was intended to be =yes but was wrongly =no)	Not available
VSIGQSAT = 0.02 (relative humidity of 96%)	VSIGQSAT = 0.02	VSIGQSAT = 0.02	Not applicable (i.e. VSIGQSAT = 0.02)
RFRMIN(22)=1.	RFRMIN(22)=1.	RFRMIN(22)=0.5	Not applicable (i.e. RFRMIN(22)=1.)
DATA ASSIMILATION			
Bator adaptations for cy43	Bator adaptations for cy43	Bator adaptations for cy43	Not available
New blacklisting procedure based on ECMWF's	New blacklisting procedure based on ECMWF's	New blacklisting procedure based on ECMWF's	Not available
RCT2SY=10 (T2M rejection limit in CANARI)	RCT2SY=10	RCT2SY=10	RCT2SY=10 (default in 40h1.1.1 is 3.9)

https://hirlam.org/trac/wiki/Harmonie_43h2/Validation_for_tagging_43h2.1

Common Code Contributions

Main HIRLAM Objectives:

- **Increase the number of H contributions**
- **Reduce divergence** between T and H cycles

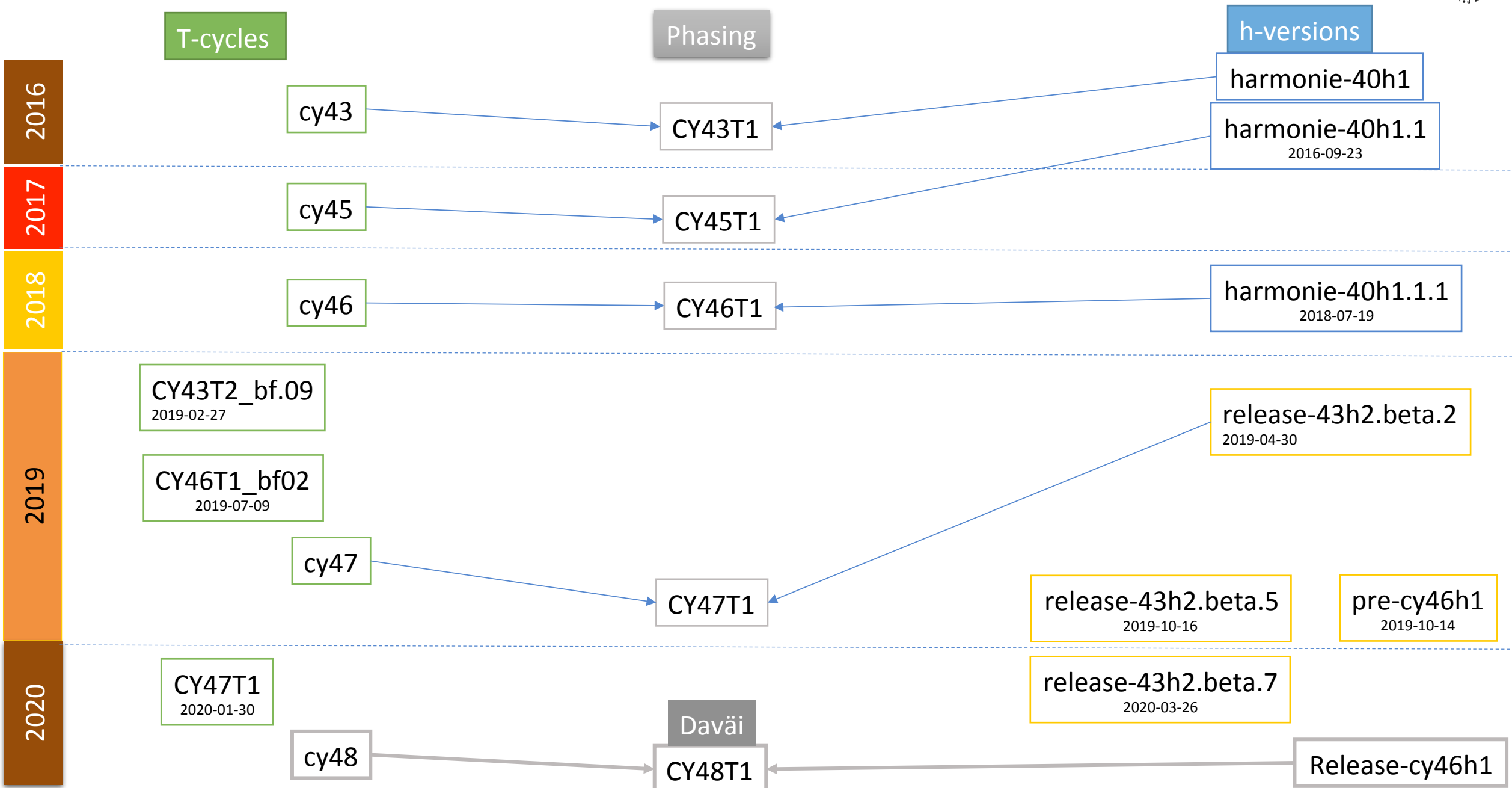
Actions in HIRLAM:

- Change **svn to git**:

  - **Training** and explore **more usable tools:** github, gitlab …
  - **Repository design and connection**: scripting and code in common repo or not

- Work on **more than one cycle**: cy43 op and cy46 next op and phasing ad
- **Forward phasing exercis**e: cy46 DNF and now in cy46h to cy48T
- Prepare **list of contributions** and discuss with MF
- Create **per topic branches and test them**: in cca or in MF computers
- Increase the knowledge on **testing tools**: Mitraillette and Daväi:
- Basic **exercise about testing CMC** under mitraillette (Daan)

Current Harmonie-Arome development in HIRLAM



Scallability

Main HIRLAM Objectives:

- Increase the **knowlege** about **model computer performance**
- Be **prepared** for **new architectures, model configurations:** Grid Point Solver, 4DVar, EPS, Nowcasting, Climate

Actions in HIRLAM:

- Test harmonie-43h2 performance and **HPC placement on ECMWF and in local HPC**.
- **BSC scalability project** financed by HIRLAM.
- **ARM testing in KNMI**
- Performance enhancement on ***Cray computers from FMI and DMI***
- **Single precision** compilation and running on harmonie-43h2.1

Scaling on CCA

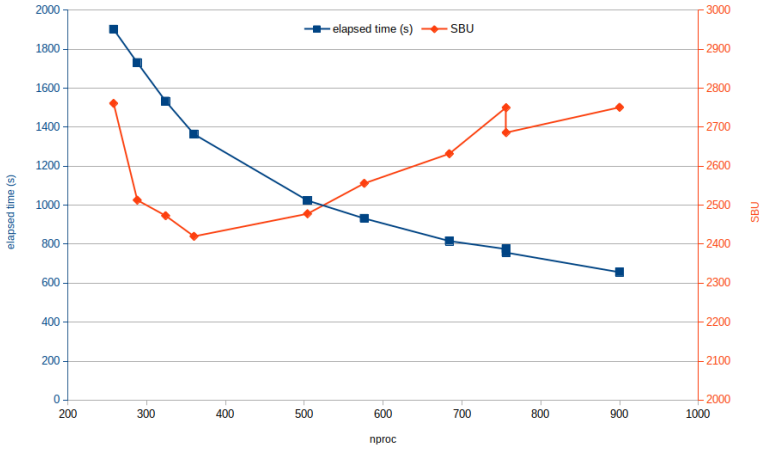
Some settings:

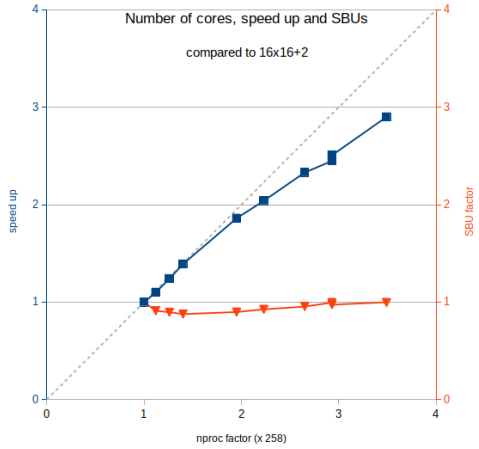
version	/home/ms/spsehlam/hlam/harmonie_release/git/branches/harmonie-43h2.1.target.2 (3c619ce5a8bc5efc238bd822f6f694eced1977a4)
platform	ECMWF ecgb/cca
domain	NETHERLANDS (800x800@75s)
DTG	2018112000 (cold start)
LL	12
NPROMA	-30
MBX_SIZE	increased to 928000000 in harmonie_namelists.pm

The table below shows the configurations tested for task Forecast and the results as derived from the epilogue of task Forecast for logfile HM_Date_2018112000.html. Note that results are from a single run per configuration.

name	nprocx	nprocy	nproc_io	nproc	nnodes	x/y ratio	elapsed (s)	queued (s)	SBU	cores factor	nodes factor	speed up	SBU factor
16x16+2	16	16	2	258	7.17	1.00	1902	234	2761	1.00	1.00	1.00	1.00
15x19+3	15	19	3	288	8.00	0.79	1731	72	2513	1.12	1.00	1.10	0.91
16x20+4	16	20	4	324	9.00	0.80	1533	52	2473	1.26	1.13	1.24	0.90
17x21+3	17	21	3	360	10.00	0.81	1364	153	2420	1.40	1.25	1.39	0.88
20x25+4	20	25	4	504	14.00	0.80	1024	37	2478	1.95	1.75	1.86	0.90
22x26+4	22	26	4	576	16.00	0.85	932	39	2556	2.23	2.00	2.04	0.93
26x26+8	26	26	8	684	19.00	1.00	816	24	2632	2.65	2.38	2.33	0.95
26x29+2	26	29	2	756	21.00	0.90	775	34	2750	2.93	2.63	2.45	1.00
25x30+6	25	30	6	756	21.00	0.83	757	59	2686	2.93	2.63	2.51	0.97
28x32+4	28	32	4	900	25.00	0.88	656	173	2751	3.49	3.13	2.90	1.00

Up to 25 nodes (max tested) the elapsed time decreases. Although saturation starts to show if more than 14 nodes are used, the amount of SBUs remains less or similar to the default 16x16+2 configuration. In the graphs below the main results of the scaling exercise can be seen.





Two configurations were run on 21 nodes: 26x29+2 and 25x30+6. The latter one is slightly faster, while it uses less cores for calculations. This may indicate 2 IO cores is too few when using many nodes, but it may also be down to chance. This will also greatly depend on the requested output and system IO-performance.
Queueing times don't seem to increase, though most runs were done over the weekend.
The ICMSHHARM+0012 and ICMSHSELE+0012.sfx files of the all the runs were compared against those of the default 16x16+2 set up with xtool_grib_api and for all tested configurations the results were identical, i.e. xtool_grib_api reported no differences.
So, the Forecast task can be run a 2.9x faster, perhaps even more, without using more SBUs or affecting model results. Perhaps ECMWF will start to complain if everybody starts using 25 nodes for the Forecast, but using 10 nodes (17x21+3) seems a more optimal default than 16x16+2.

https://hirlam.org/trac/wiki/Harmonie_43h2/Validation_for_tagging_43h2.1

ARM testing

Timing

On the Skylake machine, a 3 hour forecast run (cycle 40h1, no figures available for cycle 43) uses 2 hours. The (non-radiation) time steps take between 40 and 45 seconds.
The equivalent run on the ARM machine took at least 3 hours (but this timing was marred by swapping, as the Cycle 43 code apparently does not always fit in 128 Gbyte). The (non-radiation) time steps take between 67 and 68 seconds.

Scalability

Duration of a non-radiation time step (seconds)

`nprocx` * `nprocy`	No OpenMP	# OMP THR = 2	# OMP THR = 4	# OMP THR = 8
8 * 6	67			
6 * 6	74			
6 * 4	90	68		
4 * 3	150		68	
3 * 2	243			72

Power Consumption

The Intel Skylake machine takes 200 W of power to run the HIRLAM forecasts versus 59 W when idle.
The ARM ThunderX machine takes 175 W of power to run the HIRLAM forecasts versus 118 W when idle.
The difference in “idle” power consumption is probably a reflection of the different roles blades (server farms) and workstations (office space) have.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

BSC-HIRLAM collaboration: HARMONIE code profiling current status

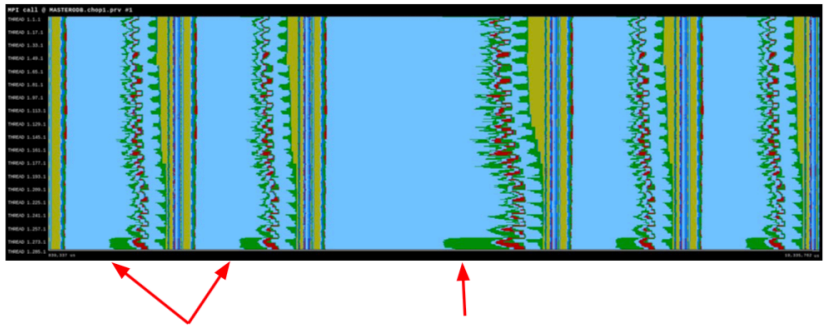
Xavier Yepes-Arbós
Mario C. Acosta
Kim Serradell

26/11/2019

HARMONIE System Working Week, MET Norway

Structure of HARMONIE

- Cut trace containing 5 time steps from the profiled scenario
- Original trace is get from a 1-hour run -> about 30 GB



Regular time steps

Larger time step (it is probably computing radiation)

Strong scaling

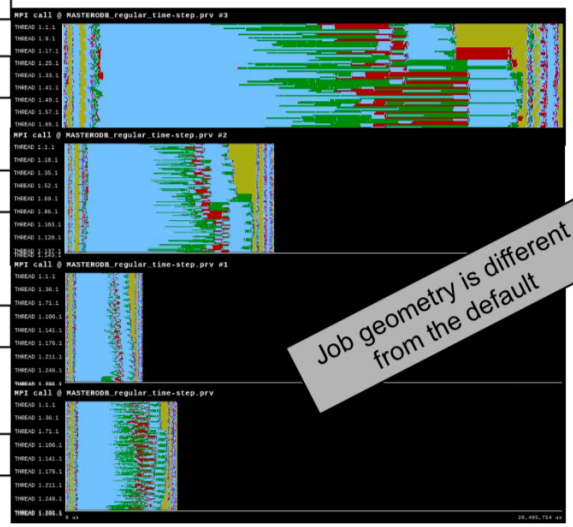
nx*ny*th

8*9*4 = 288

11*13*2 = 286

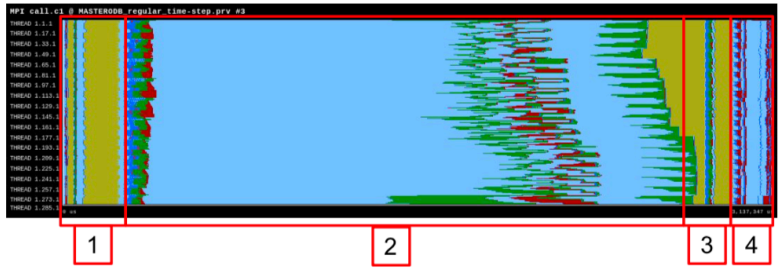
15*19*1 = 285
(default)

15*19*6 = 1710



Job geometry is different from the default

Load balance of the time step (2)



1 - Inverse transformations - good load balance (90%) ✓
2 - Grid-point computations - some load imbalance (84%) ✗
3 - Direct transformations - load balance is ok (87%) ✓
4 - Spectral computations - load balance is ok (87%) ✓


LACE
together
community
portable
training
common
joint
HIRLAM
consolidated
source code
optimization
git
tools
collective
working environment
scalability
scripting
usable
cooperative
repositories
united
ALADIN
repository
applications

Thanks

System progress

Harmonie-43h2

- Proposed RCR time schedule:
  - April 1: Tagging harmonie-43h2.1.rc.1
  - June 10: RCR reporting about preoperational testing
  - June 15: Release harmonie-43h2.1
  - Sept 15: Harmonie-43h2.1 operational (det or eps)

- Harmonie-43h2.1.beta.6 retagged Feb 25 2020 - Daniel
  - Surfex tunning - Patrick
  - HGT_QS =yes
  - OpenMP fixes - Ole
  - Single Precision from Xmas - Ole

- Documentation:
  - MF documentation recovered from Staphane Martinez – Daniel
  - Commented namelist on going (cross checks needed) – Daniel Patrick
  - Open wiki page about harmonie-43h2.1 – Daniel

- Fixing EPS and vfld contents: RH, cloudbase … - Ulf
- DA obs testing – Roger
- Testing cmake – Yurii
- HCLIM on new git repo connected with cy43 dev - Bert & Kai
- Scalability and perforrmance
  - BSC report templeate and problems related with IO server and OpenMP
  - ARM performance and scalability – Toon (isn't really optimized for (low) power consumption)
  - Porting DMI/FMI improvements on Cray to ECMWF before rc1 - Niko
  - Musc on Virtual Machine

System progress

pre-46h1

- https://hirlam.org/trac/wiki/HarmonieSystemDocumentation/PhasingWithGit/Cycle46t1_bf.02
- Pre-46h1 compiles and runs but not correctly – Eoin, Ulf
- cy46T2_bf.02 compilation under makeup works – Daniel, Ulf
- Problems to run mitraillette in cca - Ulf
  - Move to MF computers – Niko Toon
  - There is a CY46T1 compiled in cca (contact point Stéphane Martinez) -Claude
- Merge develop with pre-cy46h1
- The next export version will be based in CY46T1_bf* and 3DVar is under test by Daväi – Claude

cy48

- Prepare the list of code to discuss with MF – Daniel & Roger
- CY47 is not installed in cca - Cluade
- The pre-CY48 version will be sent to ECMWF in March and also will be tested under Daväi -Claude
- Traning on Daväi – Alexandre and Hirlam System Core

AOB

- Github exercise – Roel
- Unit testing by Travis -Roel
- PrepLAM – Roel