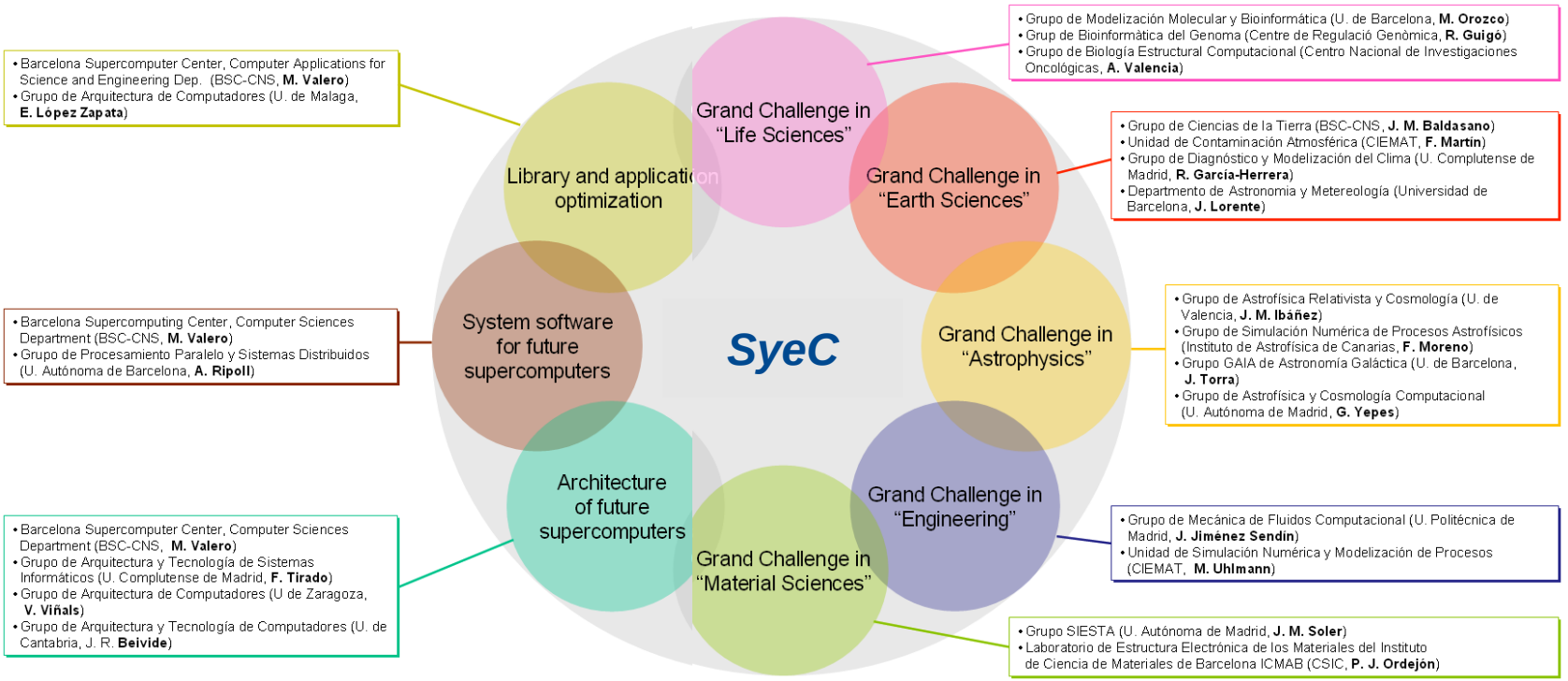
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Ciencias de la Tierra

Francisco J. Doblas-Reyes
BSC Earth Sciences Department

Supercomputing and eScience



- Grand Challenge in Earth Sciences:
  - Characterizing parallel performance of air quality and climate models
  - Adding complexity to models in Earth Sciences
  - Exploit higher resolution configurations

SyeC
Supercomputación y eCiencia

From the Consolider

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

- Improvement of CALIOPE air quality model (BSC-ES, BSC-CS)
- Enhancement of CHIMERE model – new heavy metal and Hg chemistry (CIEMAT)
- Improved urban scale CFD models (CIEMAT)

Performance of air quality models

BSC-ES, CIEMAT

Lessons from the Consolider:

- Further interaction CS-ES communities needed
- Training of ES scientist in CS
- Create stronger links among Spanish ES community
- Need further optimization of codes
- But also, a more rational management of the data

BSC-CS

- Scalability and performance analysis of WACCM (UCM, BSC-CS)
- Analysis of climate simulations with WACCM (UCM, UB)
- Atmospheric dynamics with Alya system (BSC-ES, BSC-CASE)

Scalability, performance, optimization

SyeC

- Enhancement of mineral dust model (BSC-ES)
- Development of full aerosol component for NMMB/BSC-CTM model (BSC-ES)
- Performance of NMMB/BSC-CTM model (BSC-ES, BSC-CS)

BSC-CS BSC-CASE

Scalability of Earth System Models

UCM, UB, BSC-ES

BSC-CS

Development of coupled meteo-chemistry models

BSC-ES

STREAM (Stratospheric and Tropospheric ResEarch And Modeling) group

(1) Dpto. Física de la Tierra II, Universidad Complutense de Madrid
(2) Instituto de Geociencias, CSIC-UCM, Madrid

Stratospheric and Tropospheric Research and Modeling
Stream
group

IGEO
INSTITUTO DE GEOCIENCIAS

VNIVERSITAS COMPLVTENSIS MATRITENSIS

Who we are

Research | About Us | People | Monographs | Publications | Collaborations | Contact

Permanent Staff

Postdoctoral Researches

PhD. Students

Former Members

Permanent Staff

Ricardo García Herrera
Professor
rgarciah(at)fis.ucm.es

Natalia Calvo Fernández
Associate Professor
nataliac(at)fis.ucm.es

Postdoctoral Researches

David Barriopedro Cepero
Researcher
dbarriop(at)fis.ucm.es

Carlos Ordoñez
Researcher

Leo Carro Calvo
Researcher

PhD. Students

Froila María Palmeiro Núñez
fm.palmeiro(at)fis.ucm.es

Maddalen Iza San Juan
m.iza(at)ucm.es

Javier Mellado Cano
jamelladocano(at)gmail.com

Antonio Sánchez

http://stream-ucm.es/

What we do

Research | About Us | People | Monographs | Publications | Collaborations | Contact

Stratospheric and Tropospheric REsearch And Modeling

The Stratospheric and Tropospheric REsearch And Modeling (STREAM) group is a research team of the Universidad Complutense de Madrid devoted to the analysis and modeling of the atmospheric circulation and climate variability. Currently, the group has about ten members, including permanent staff, post-doctoral and PhD students, with an intense international activity and a dense network of collaborations. We have more than 10 years of experience in the field. The research activity of STREAM is organized around four main topics: climate reconstruction in the last 500 years; tropospheric climate variability, dynamics of the middle atmosphere and extreme events.

Stratospheric Dynamics (Middle Atmosphere)

Tropospheric Climate Variability and Change

Climate Reconstruction

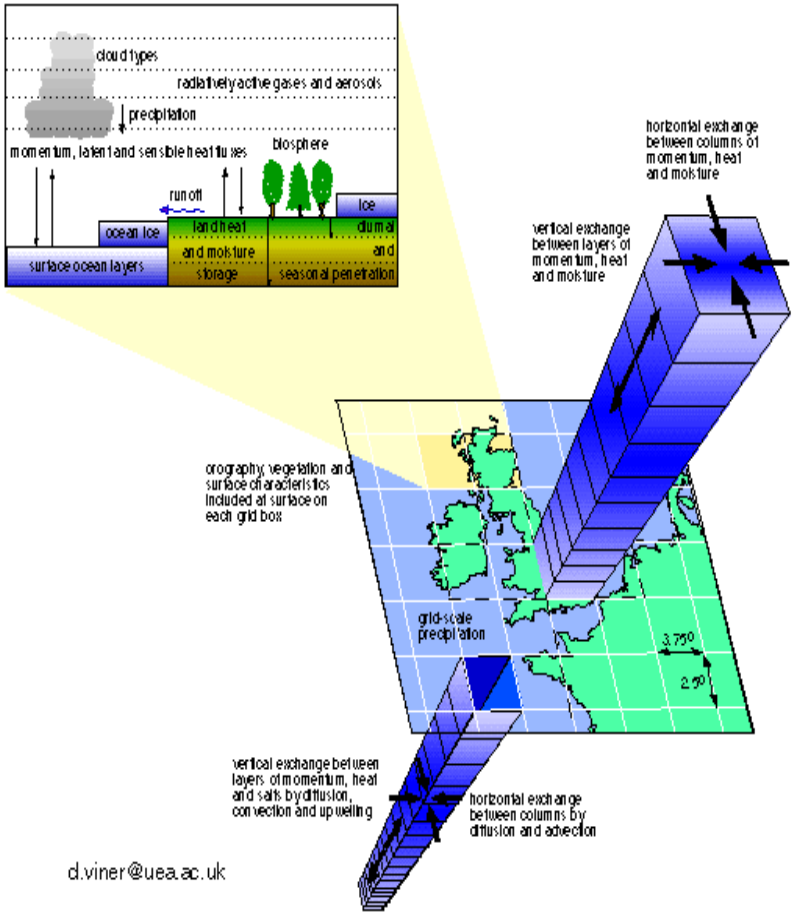
Extreme Events

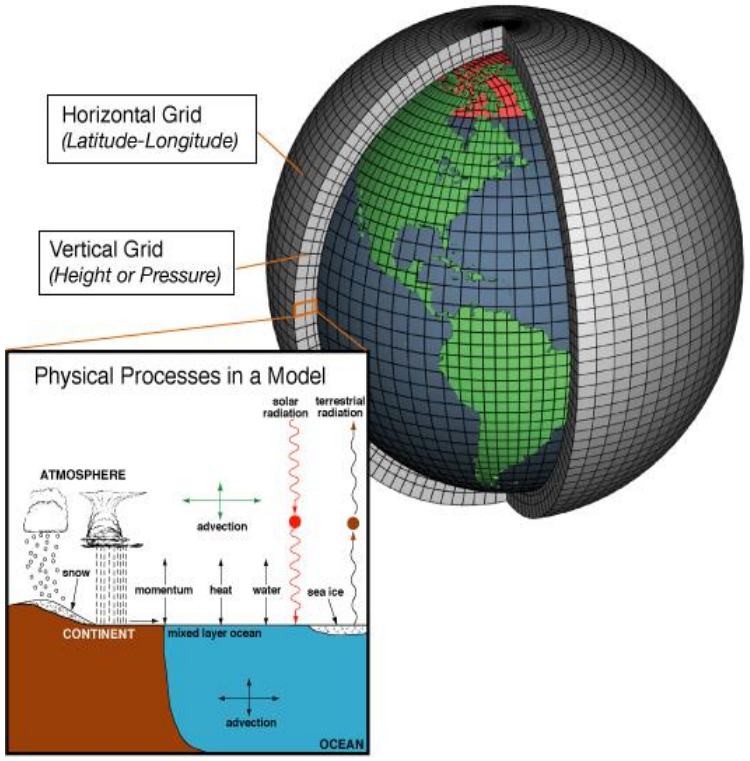
How we do it

- Observational networks and reanalysis products
- Documentary sources and climatic proxies
- Climate Model Simulations

Global Climate models

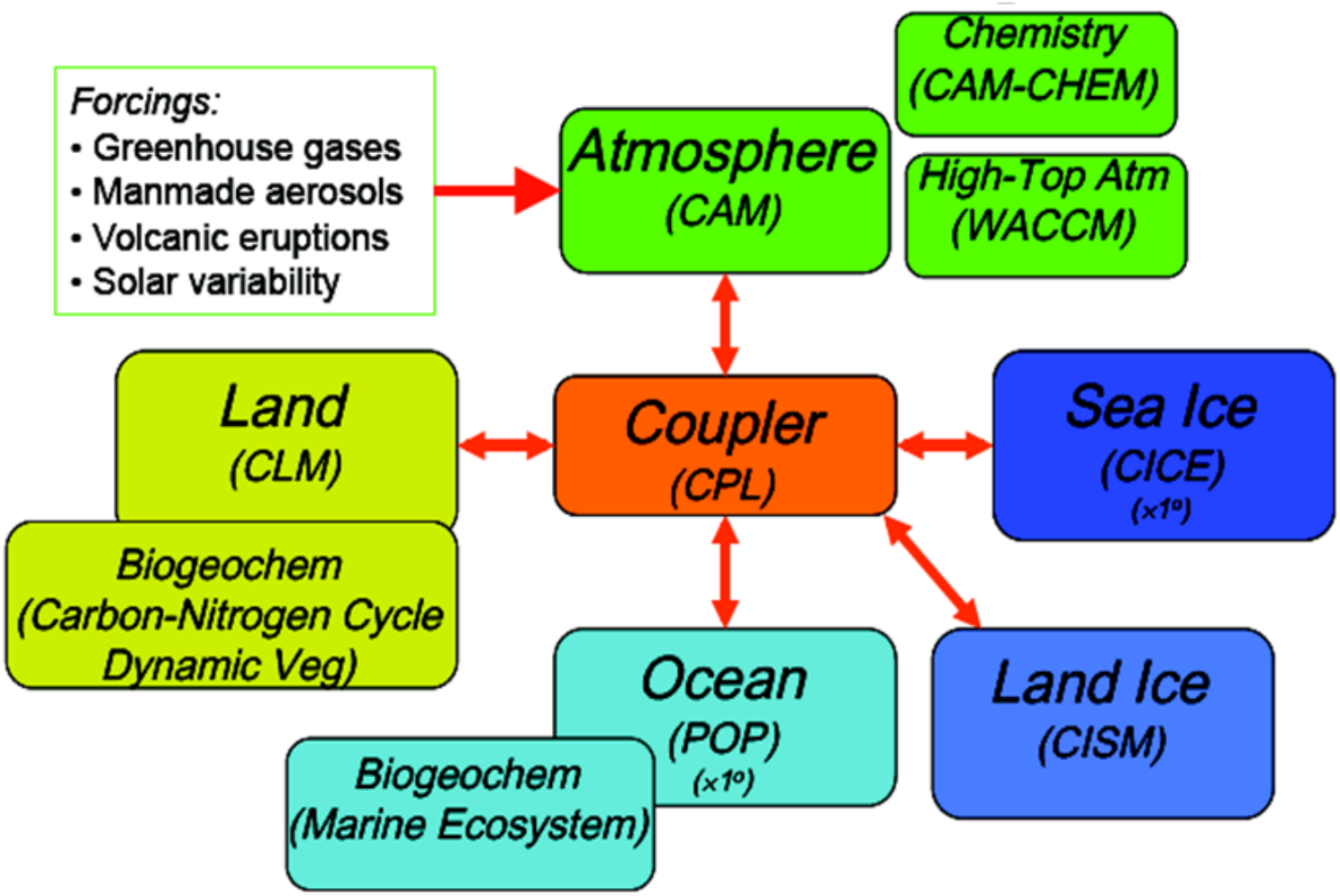
- Mathematical representation of climate and its variability based on:
- i) 3-D thermo-hydrodynamic (primitive) equations that are numerically solved at the gridpoint scale under prescribed boundary and initial conditions;
- ii) parametrizations of the subgrid-scale processes that cannot be explicitly resolved.





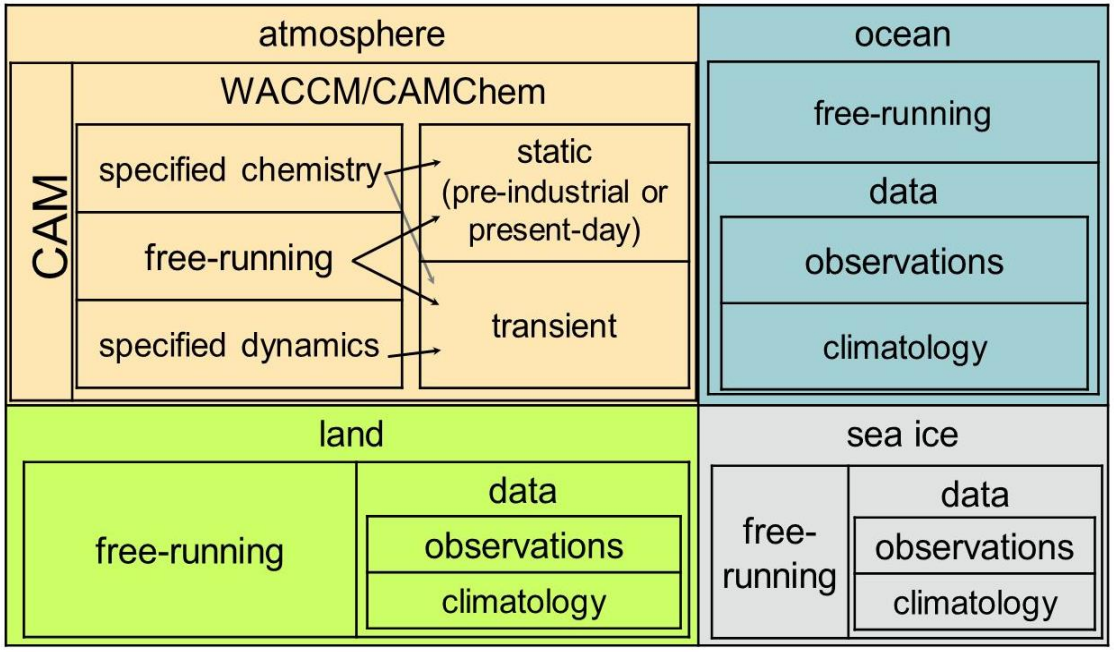
CESM – WACCM model

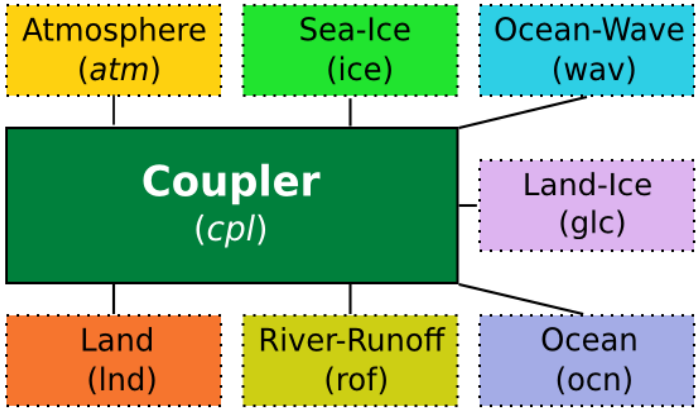
- Based on the “low-top” Community Atmosphere Model (CAM)
- 1.9ºx2.5º spatial resolution. 66 vertical levels (0-140 km). 30 minute time step
- MOZART chemistry package + LMT processes

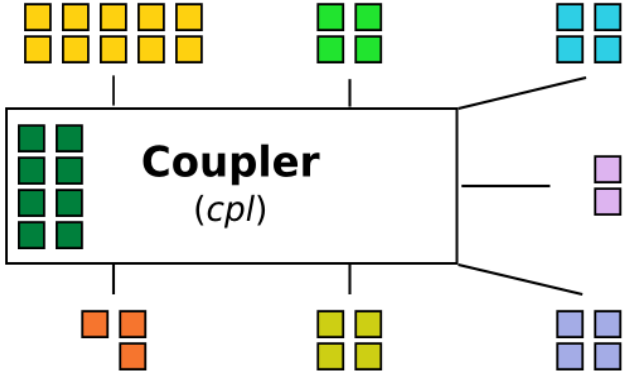


Component configurations

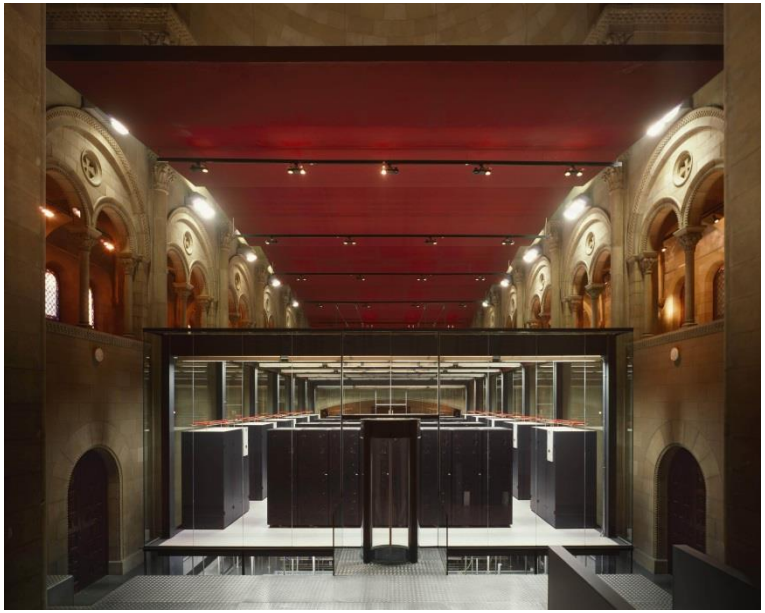
- Design of simulation: control vs forced simulations.
- Mixed sequential/concurrent execution between atm/lnd/ice and ocn components.
- A user can define how many nodes should be allocated for each component.
- MPI and MPI+OpenMP parallelism.





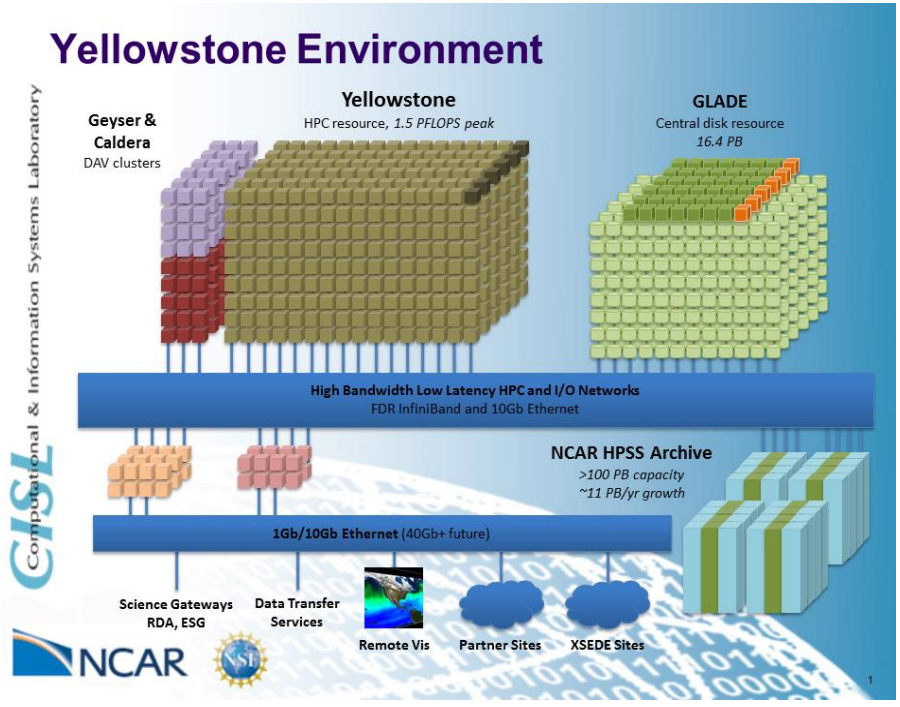


Clusters running CESM-WACCM



MareNostrum	BSC (Barcelona Supercomputing Center)
Ranking Top500 List / Green List	91 / 185
Max. / Peak Performance	0.9 / 1.1 PFLOPS
Arquitecture Computer	Cluster IBM iDataPlex DX360M4
Processor	Intel SandyBridge Xeon E5-2670 8C 2.6GHz
Nodes / Processors	3,056 / 48,896
Memory	100.8 TB
Disk / Archive Storage	2,000 / 8,500 TB
Interconnection	Infiniband FDR10, GbE
OS	SUSE Linux Es

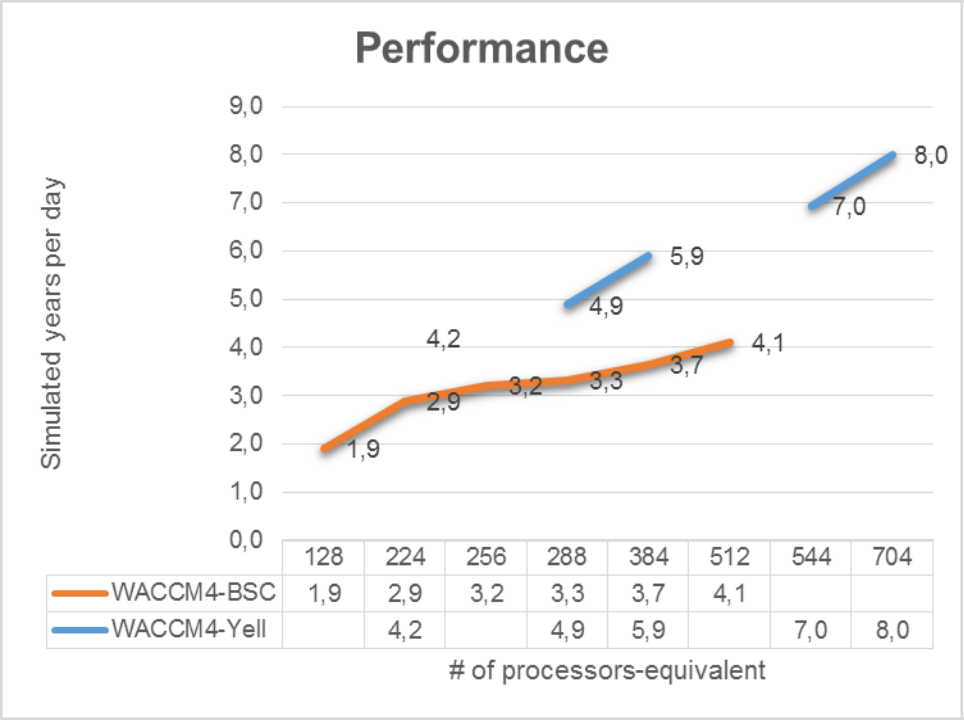
Clusters running CESM-WACCM

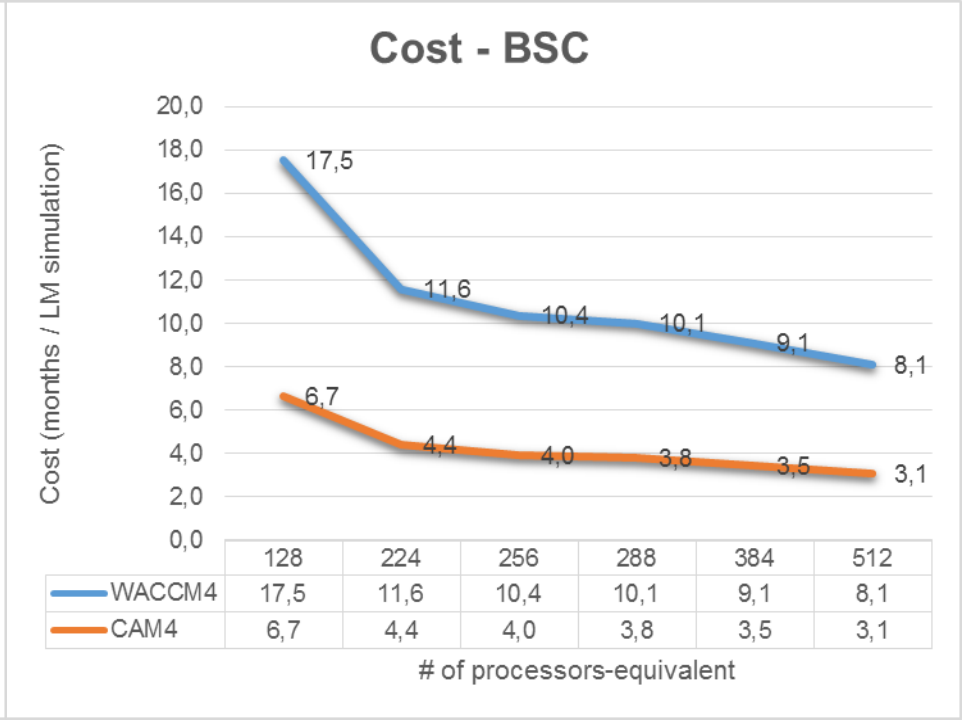


Yellowstone	NCAR (National Center for Atmospheric Research)
Ranking Top500 List / Green List	57 / 198
Max. / Peak Performance	1.3 / 1.5 PFLOPS
Arquitecture Computer	Cluster IBM iDataPlex DX360M4
Processor	Intel SandyBridge Xeon E5-2670 8C 2.6GHz
Nodes / Processors	4,536 / 72,576
Memory	144.6 TB
Disk / Archive Storage	16,000 TB
Interconnection	Infiniband FDR10, GbE
OS	Red Hat Linux

Performance

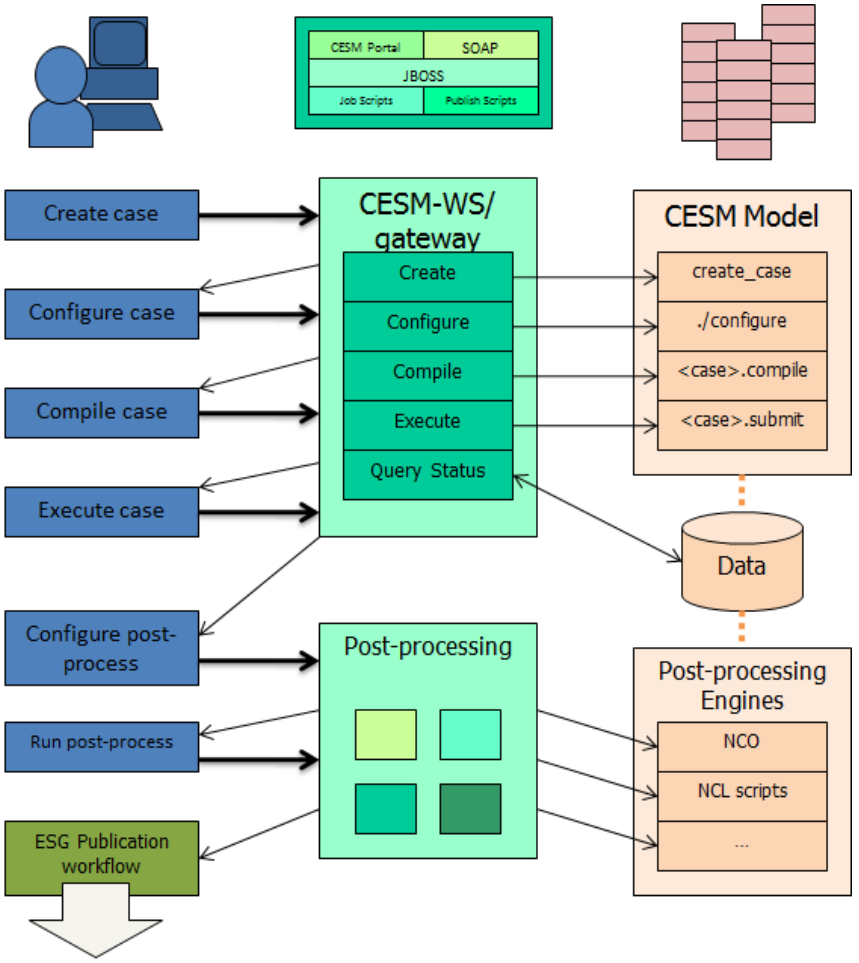
- Model does not scale linearly at MR. With more than 256 cores, the cost increases a lot (due to multiple communication needed between more nodes?)
- The cost is not improved if model is executed with MPI + OpenMP libraries (i.e., the ratio of simulated years / use of nodes is not improved).

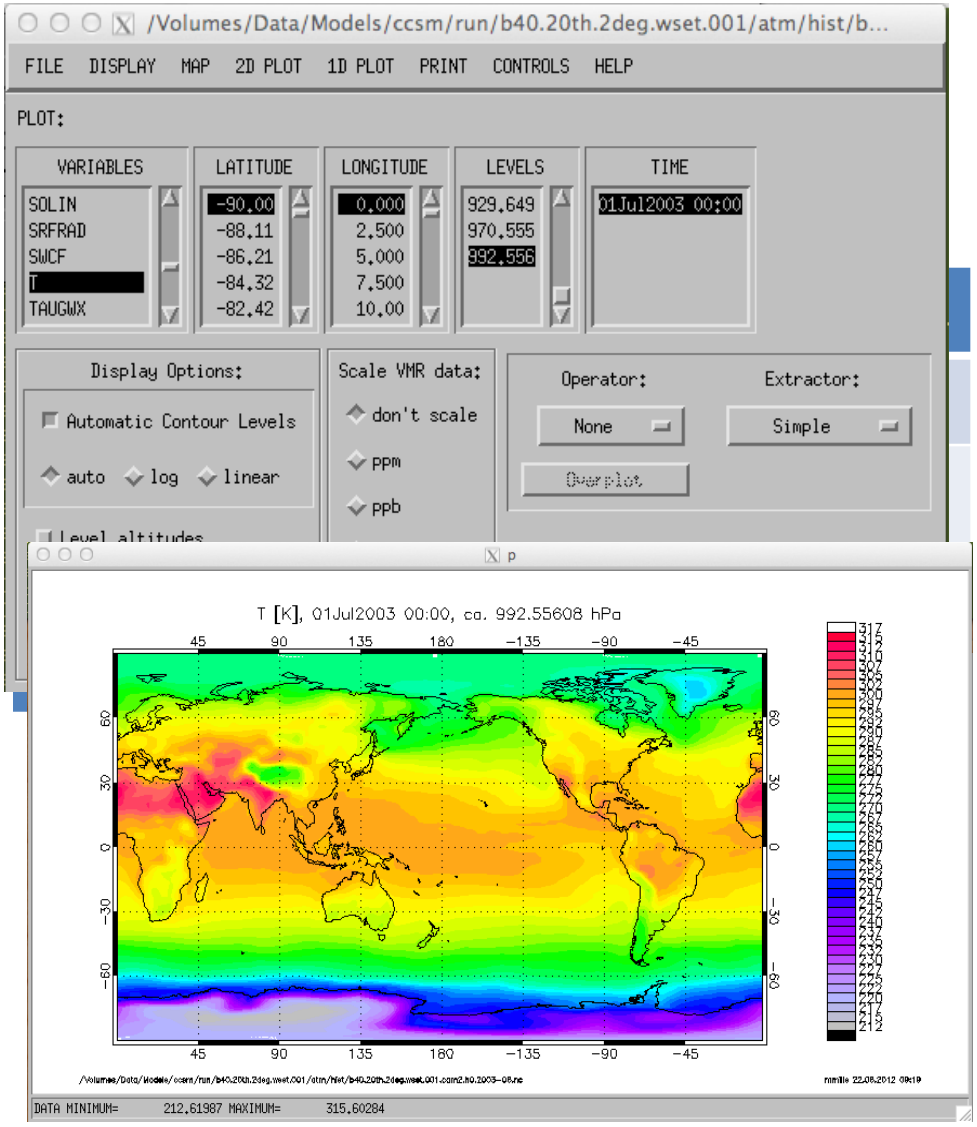




Post-processing data analysis

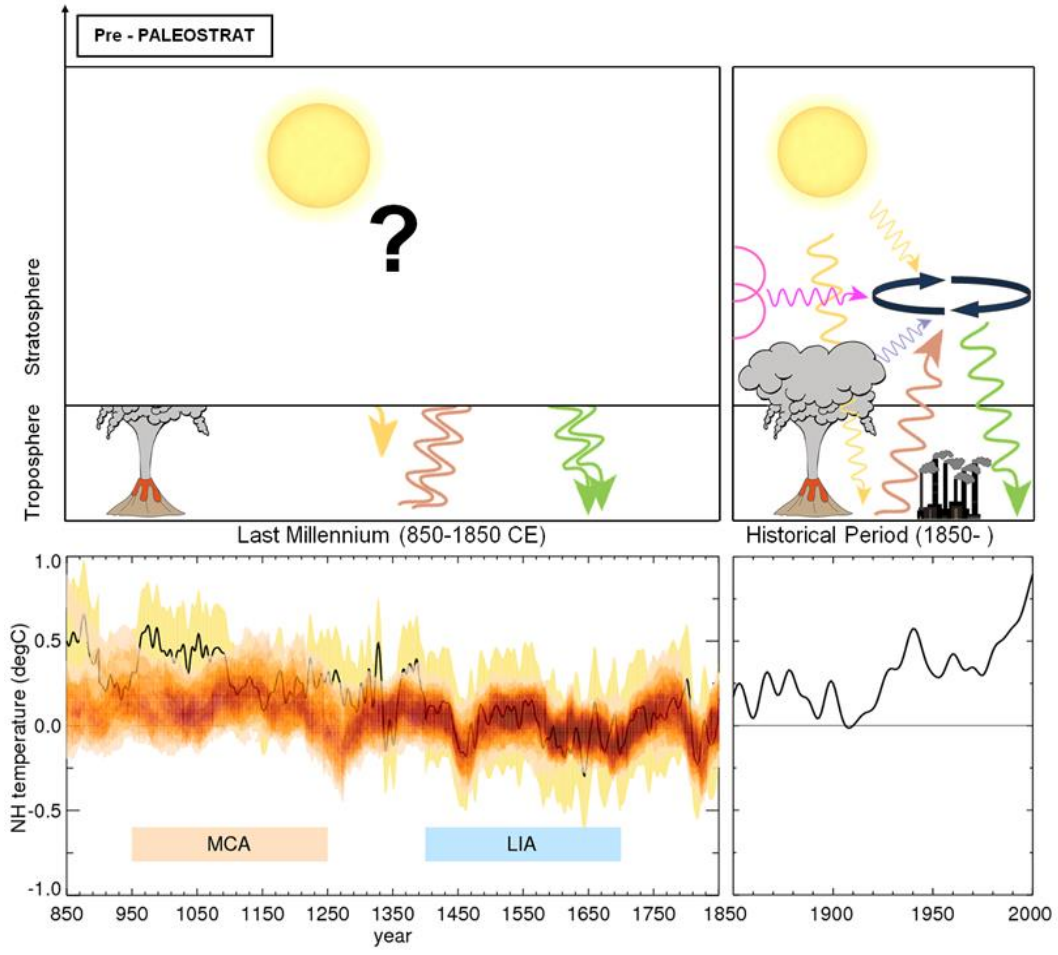
- CESM output files are (x,y,z,t) in netCDF format, and may be analyzed with standard analysis tools, including NCO, NCL, Matlab, IDL and diagnostic packages (GEOV).
- Typically output involves ~60 GB per simulated year (i.e., 0.6 TB per 100-yr simulation).
- Massive storage is required.





Upcoming needs

- PALEOSTRAT will aim to investigate the impact of the stratosphere on the climate using Last Millennium (LM) simulations.
- Objective 1: to better understand the stratospheric variability its responses to external and internal forcings.
- Objective 2: to characterize the role of the stratosphere in the climate of the LM, addressing the discrepancies between models and reconstructions.



Issues

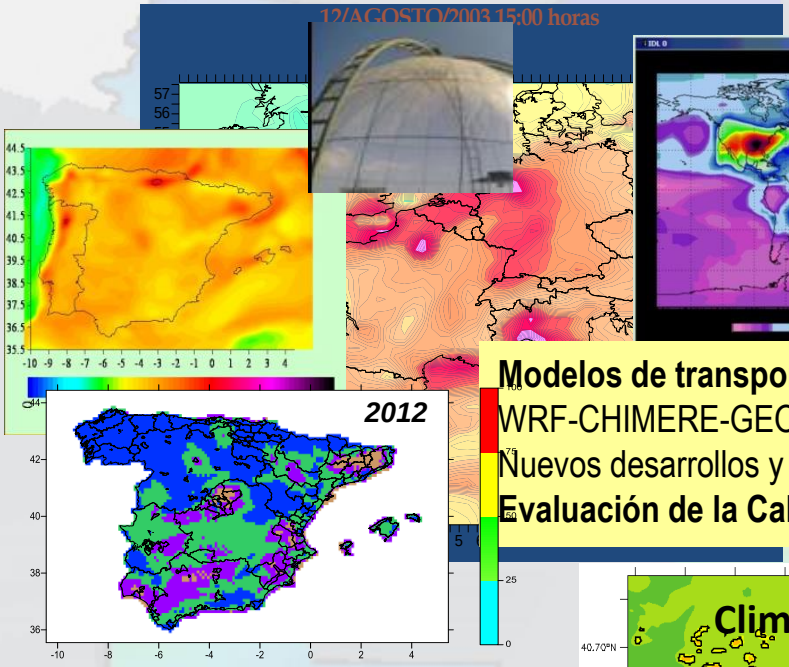
- Highly valuable
  - Technical support (changes in compiler, etc.)
  - # of accepted proposals
- Recommended
  - Increasing demand involves doubling the requirements in terms of CPU time and # of cores.
  - Optimization (reducing the cost by improving parallel processing, scalability, etc.)
- Highly recommended
  - Massive storage system.

Modelización Atmosférica y Supercomputación
SyeC
Grupo de MCA - CIEMAT

Fernando Martín
División de Contaminación Atmosférica
CIEMAT

Reunión SyeC, Barcelona Abril 2017

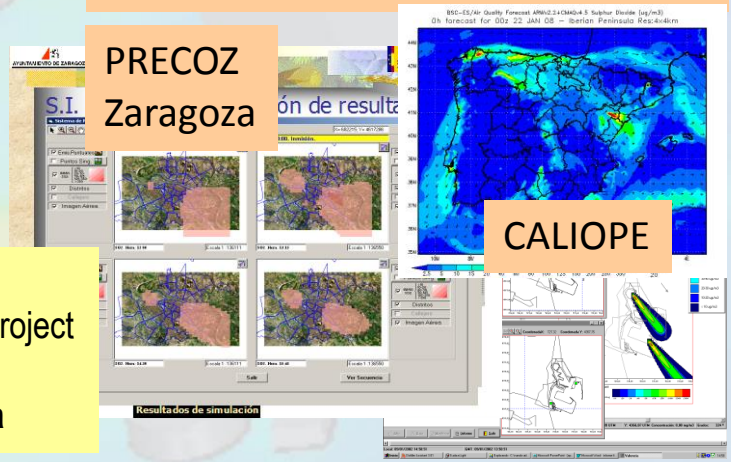
Grupo de Modelización Contaminación Atmosférica (MCA – CIEMAT)

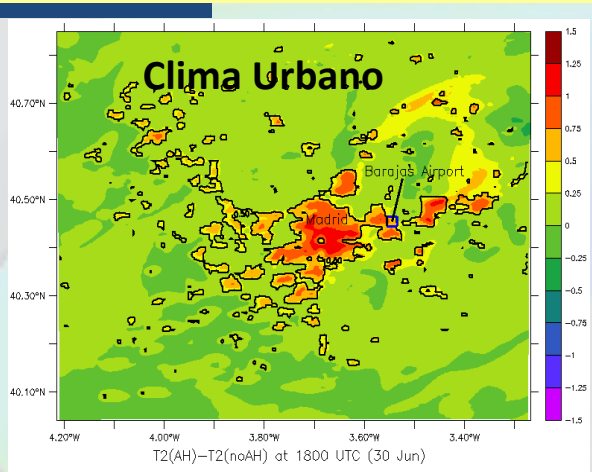


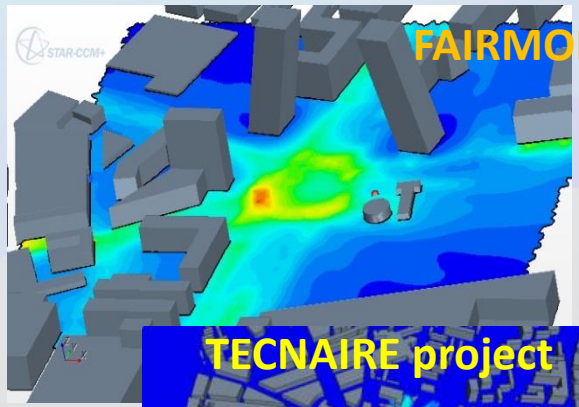
Modelos de transporte y Química.
WRF-CHIMERE-GEOSCHEM. EURODELTA project
Nuevos desarrollos y validación
Evaluación de la Calidad del aire en España

Predicción de calidad del aire

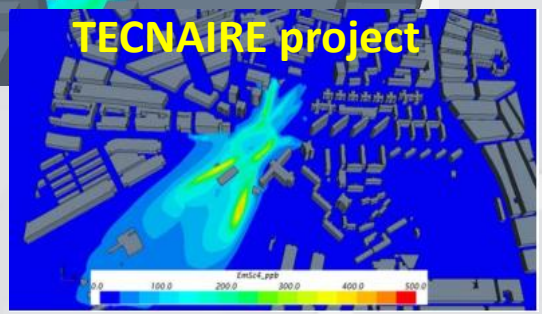
PRECOZ Zaragoza

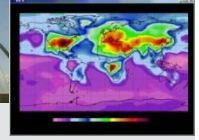
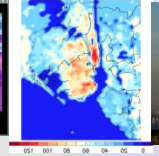


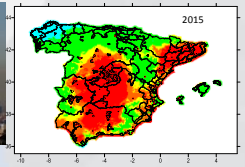
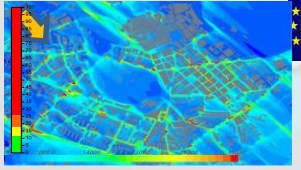
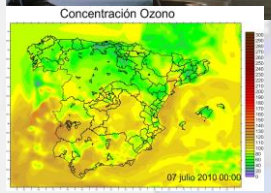


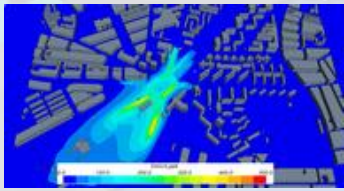


Modelización de atmósfera urbana
Modelos CFD-Street-canyon
Parametrizaciones capa límite. Modelo WRF





Lineas de I+D

DESARROLLOS

- Modelización a macro y mesoscala.

Modelo GEOSCHEM, WRF/CHIMERE

- Modelización a escala urbana

Parametrizaciones urbanas, modelo WRF/BEP/BEM

- Modelización en calles

Modelos CFD – Street Canyon

- Modelización química atmosférica

Mejoras en mecanismos químicos (SOA), metales pesados, benzopireno, etc. Instalación EUPHORE

- Modelización emisiones

Emisiones de partículas focos difusos

- Combinación modelos y mediciones

Métodos geo-estadísticos, Kriging

APLICACIONES

Investigar procesos atmosféricos

- Episodios de contaminación

Episodios de contaminación (ozono, NOx, PM)

- Clima urbano

Isla térmica urbana, olas de calor, efecto aire acondicionado vegetación, tejados verdes, etc

Apoyo a administraciones

- Evaluación de calidad del aire

Cumplimiento de legislación sobre calidad del aire

- Representatividad de estaciones y diseño de redes.
- Mejora de la calidad del aire

Implantación de BATs en industrias, efecto vegetación, tráfico, calefacciones, pavimentos fotocatalíticos,

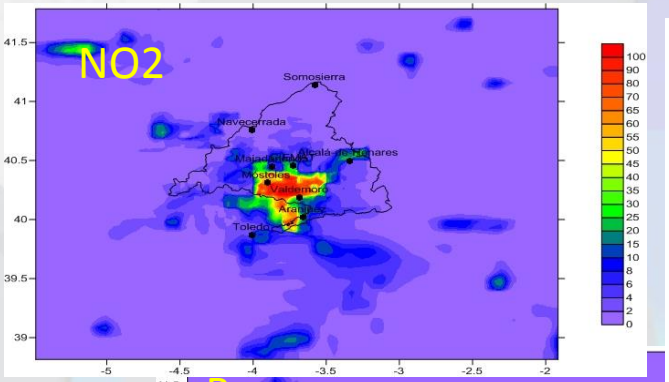
- Predicción de la calidad del aire

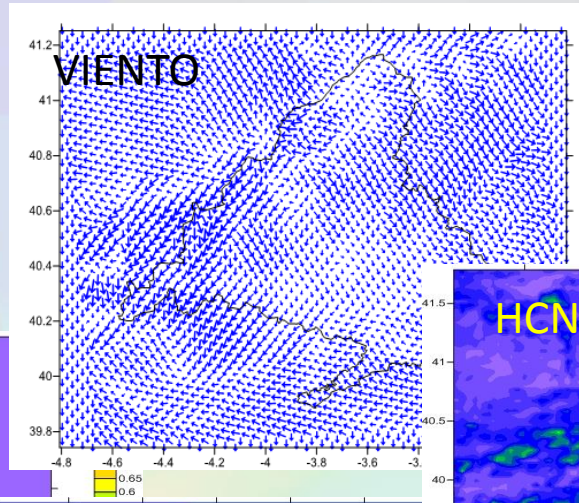
CALIOPE, PRECOZ

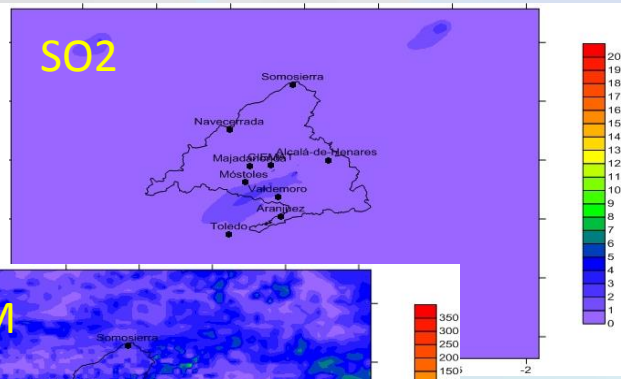
- Emergencias

Ataques terroristas (M. Defensa), accidentes (Incendio de Seseña), erupciones volcanicas (Eyjafjallajokull, Islandia).

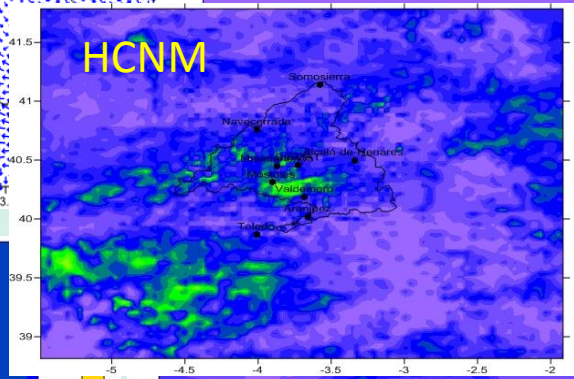
Episodios de contaminación

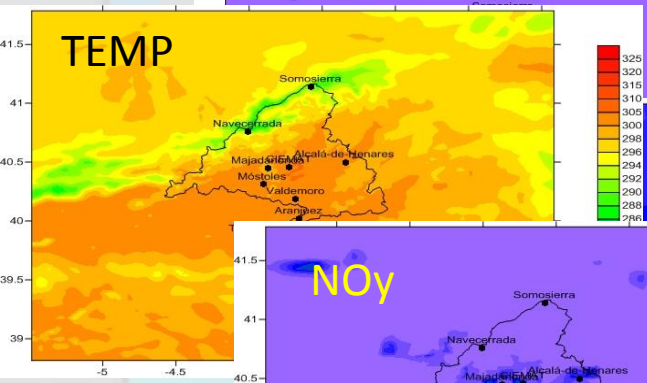


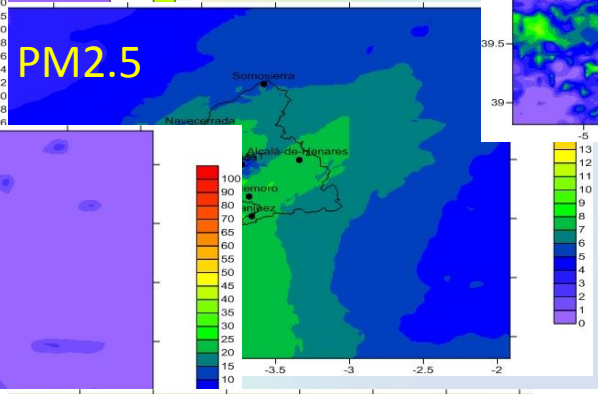


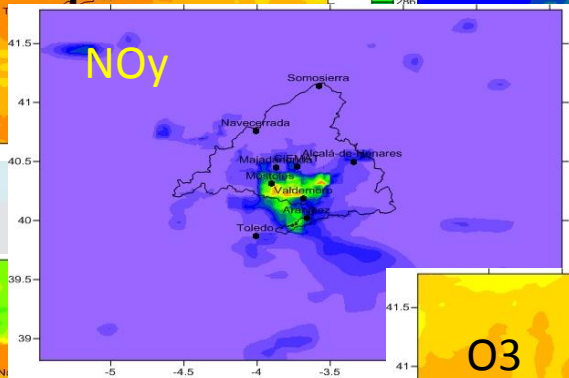


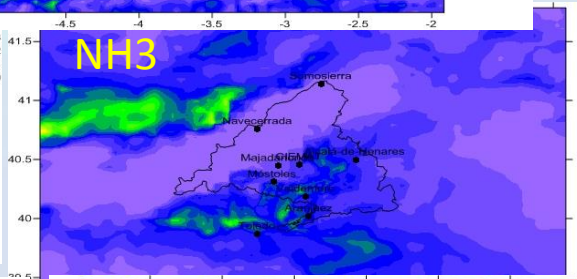
Benceno

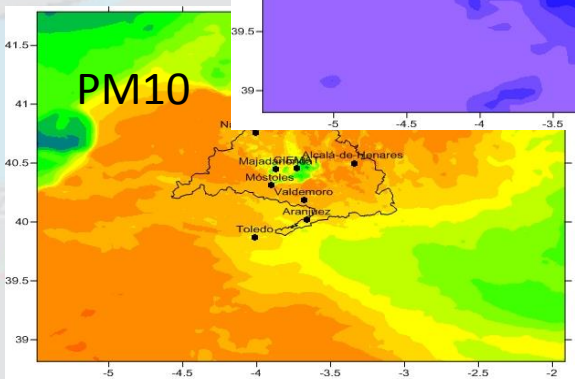


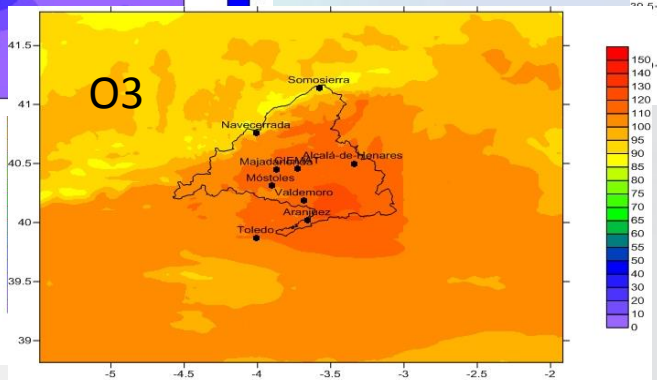


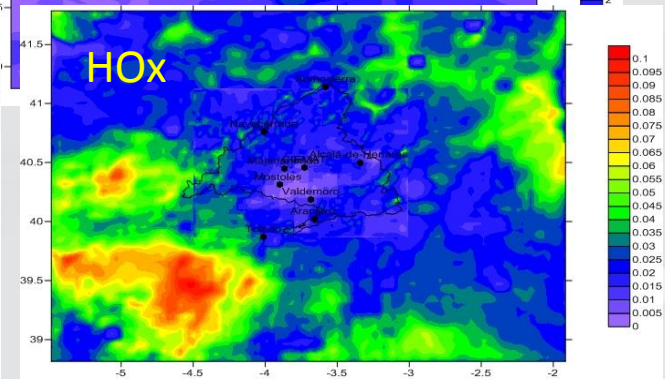






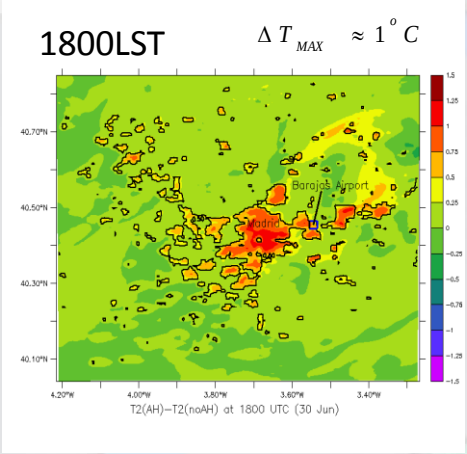


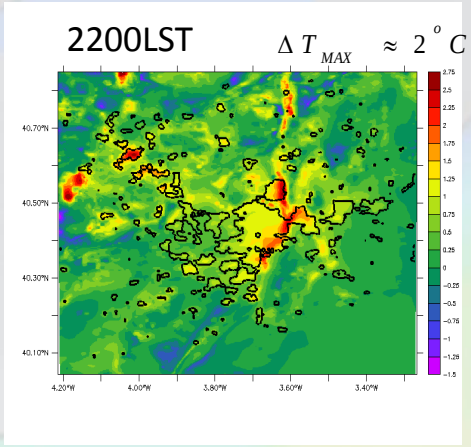




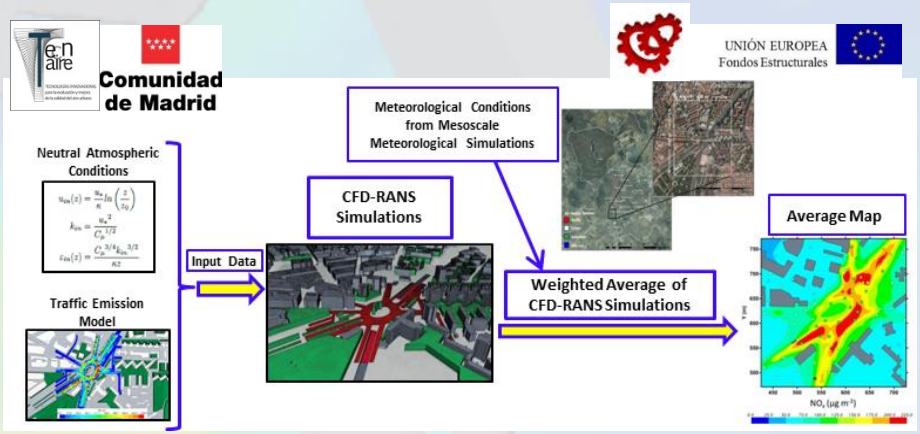
Atmósfera urbana: clima y calidad del aire

Isla de Calor urbana de Madrid, Efecto de Aire Acondicionado





Programa TECNAIRE, calidad del aire urbana, herramientas de monitorización y modelización

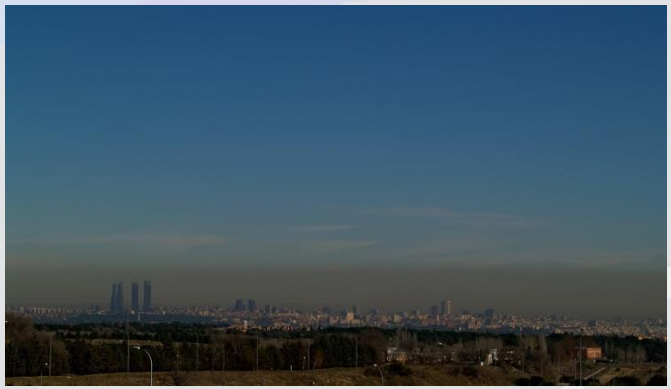


Tejados verdes - New York City

Sin green roofs

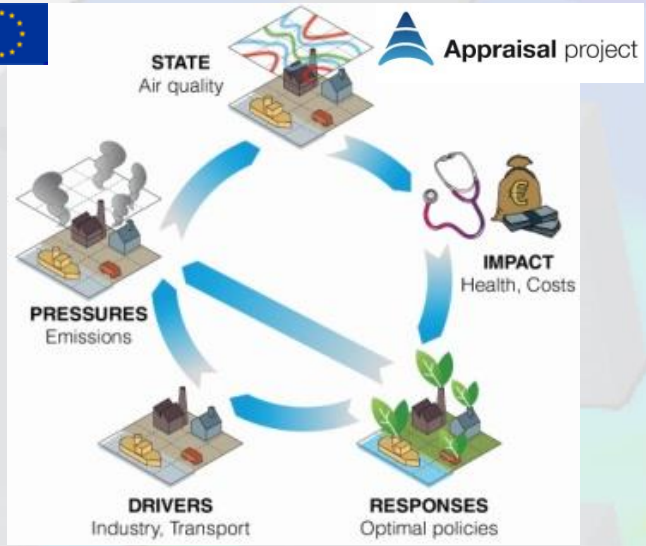
Con green roofs

Proyecto EXCLUR PE-Retos
Olas de Calor
Fuertes inversiones térmicas
Desarrollo de modelos específicos



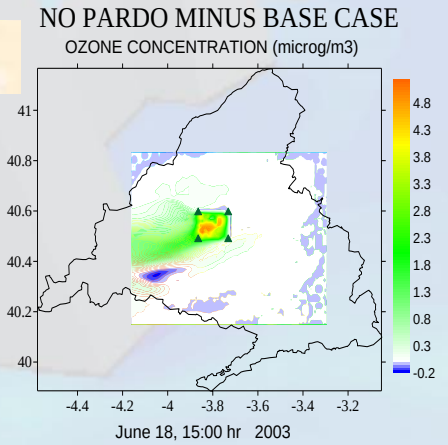
Mejora de calidad del aire

FP7 EU APPRAISAL Project

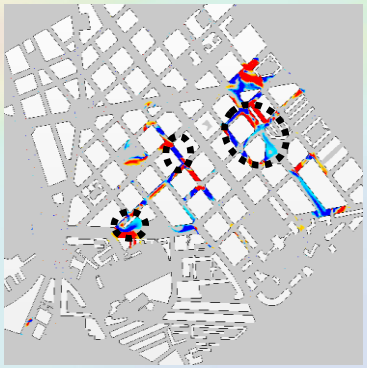


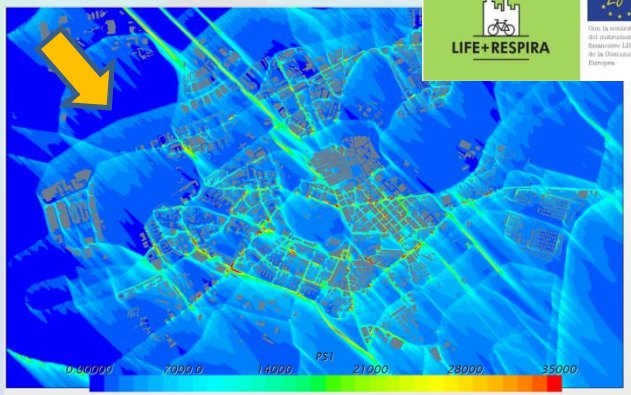
Efecto de vegetación en ozono

FP7- HEREPLUS
Impacto de Monte de El Pardo en ozono

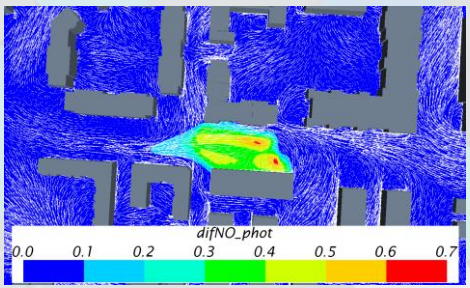


Efecto de vegetación urbana

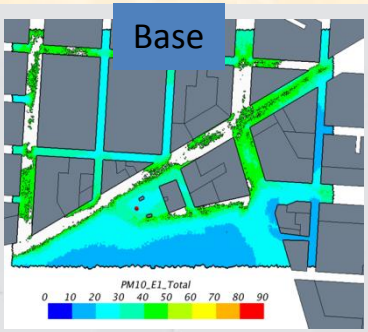


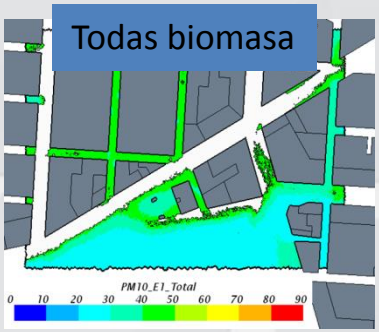


LIFE MINOx Street
Efecto de Pavimentos Fotocatalíticos



Impacto de combustibles de calefacciones





EG-MAGRAMA
Impacto de Directiva NEC

Líneas futuras – MCA

- **Modelización Atmosférica a escala europea/global: Modelos Macroescala**
  - Concentración y depósito de contaminantes, Intercomparación de modelos, BenzoPireno, tendencias y efecto del cambio climático. (Proyecto EURODELTA, Convenio de Ginebra (TFMM) y EG-MAPAMA)

- **Modelización Atmosférica a escala nacional:**: **WRF/CHIMERE y CALMET/CALPUFF**
  - Episodios de contaminación
  - Emergencias.
  - Evaluación de la calidad del aire. (EG MAPAMA) ➔ Modelos WRF/CHIMERE y Combinación de modelos y mediciones (mejora de técnicas geoestadísticas)
  - Mejora de la Calidad del Aire (EG MAPAMA).
    - Impacto de Directiva de Techos Nacionales de Emisión en Calidad del Aire.
    - Efectos de medida de mejora, impactos en salud y costes/beneficios.

- **Modelización Atmosférica en ciudades/calles: WRF/CHIMERE y CFD-Street Canyon**
  - Parametrizaciones flujos de energía para modelos (**WRF**) ➔ Clima urbano (isla de calor, olas de calor, confort térmico) y episodios de contaminación (inversiones térmicas), tejados verdes, datos de temperatura con smartphones (EXCLUR y NCAR-USA)
  - Mejora de modelos CFD (química, enlace con modelos a mesoscala, etc)
  - Representatividad espacial de estaciones de medida de la calidad del aire ➔ **modelos CFD** (FAIRMODE).
  - Exposición de la población a la contaminación (TECNAIRE, LIFE RESPIRA)
  - Predicción de la calidad del aire (PRECOZ)
  - Impacto de medidas de mejora de calidad del aire en ciudades y distritos (TECNAIRE)
  - Modelización de impacto de vegetación urbana en calidad del aire (LIFE RESPIRA)

Visión de Modelización Atmosférica y Supercomputación

- Simulaciones con **modelos atmosféricos con grid 3D** buscando siempre:
  - **dominios espaciales mayores**, (más grid cells, más operaciones):
    - mesoscala (100 km ➔ 1000 ➔ 10000 km)
    - Microescala urbana (1 km ➔ 10 km ➔ 100 km)
  - **más resolución**, (más grid cells, más operaciones):
    - mesoscala (1 km ➔ 200 m)
    - microescala urbana (10 m ➔ ≤ 1 m)
  - **simular más procesos físico-químicos** (más operaciones):
    - Mesoscala: más contaminantes (PAHs, metales pesados, POP, etc), química más detallada, capa límite urbana, etc
    - Microscala urbana: procesos químicos en CFD, enlaces con modelos emisines tráfico alta resolución, enlace con modelos mesoscala, etc
  - **técnicas más precisas** (exigen más resolución y más operaciones en muchos casos):
    - Modelos CFD (RANS ➔ LES ➔ DNS).
- Todo va en la línea de **mayor carga computacional**.
- Es **necesario**:
  - **Más potencia de cálculo**, se deberá multiplicar por 10 o más en próximos 10 años.
  - **Mayor velocidad de proceso** para al menos mantener los tiempos de calculo actuales con los dominios mayores, mayor resolución, más procesos y técnicas más precisas.

BSC Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

What

Environmental forecasting

Why

Our strength …

- research
- operations
- services
- more than 60 people working together

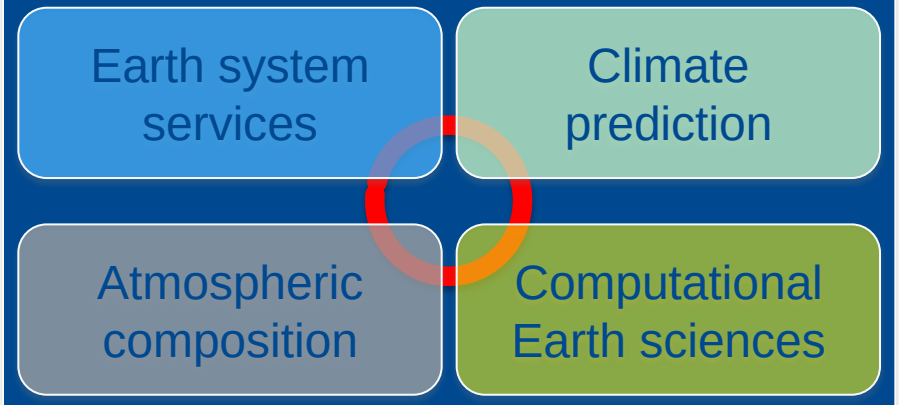
How

Develop a capability to model air quality processes from urban to global and the impacts on weather, health and ecosystems

Implement a climate prediction system for subseasonal-to-decadal climate prediction

Develop user-oriented services that favour both technology transfer and adaptation

Use cutting-edge HPC and Big Data technologies for the efficiency and user-friendliness of Earth system models

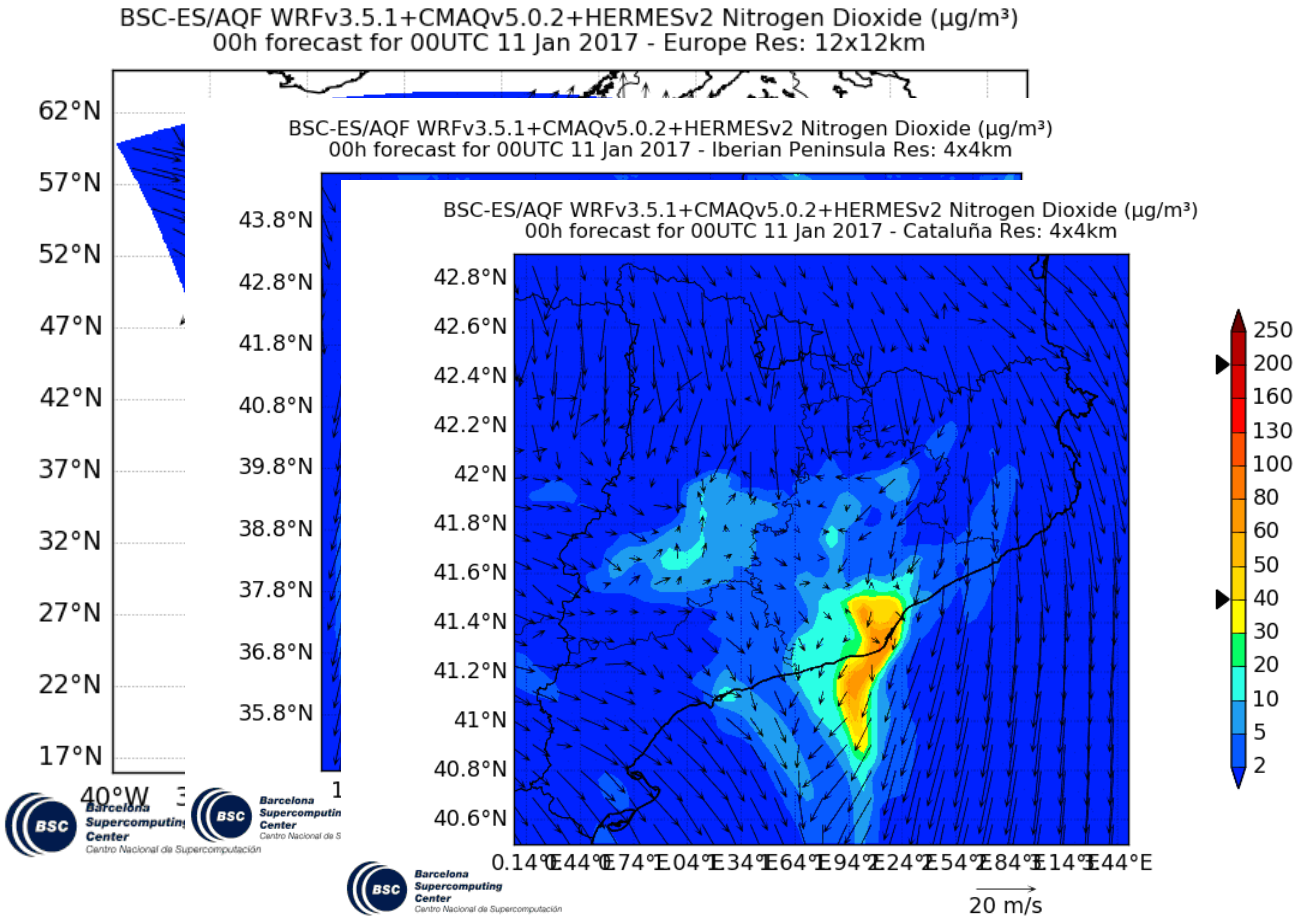


CALIOPE real-time air quality forecasts

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Provides air quality related information for the coming days and for the application of short term action plans for air quality managers.



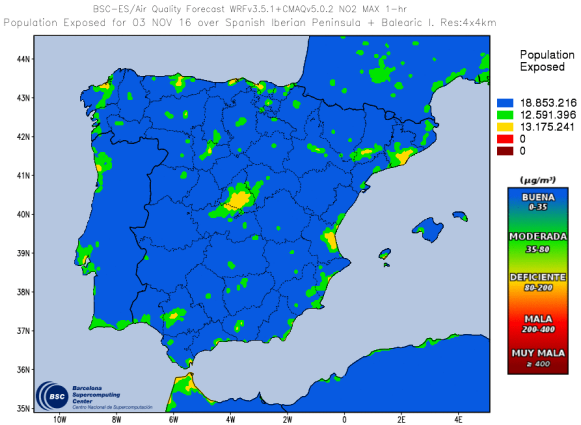
Information is delivered using both online or custom applications:

www.bsc.es/caliope

Smart city platform

vodafone IBM

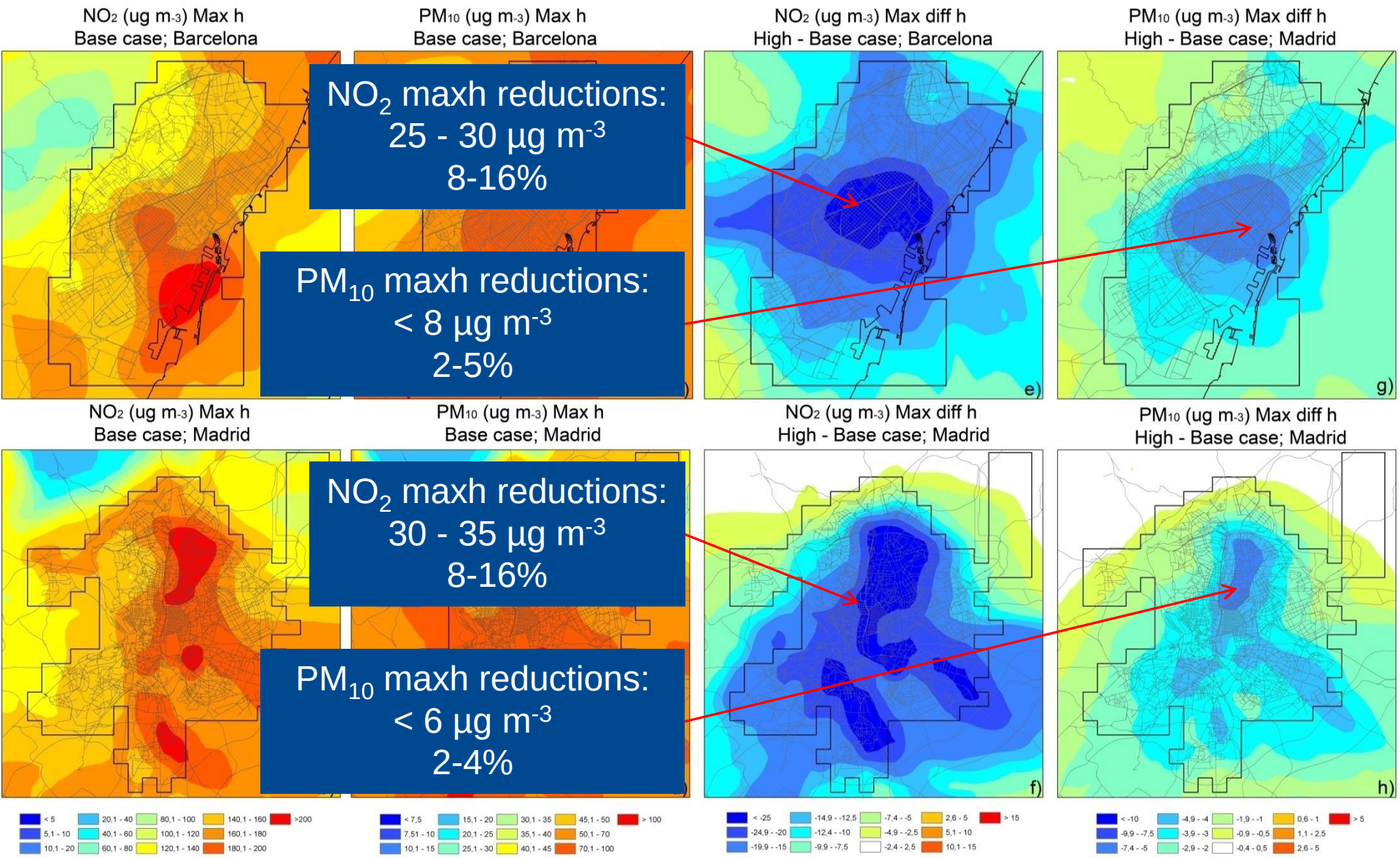
Air quality index & population exposed



Pay et al. 2014, GMD

GOBIERNO DE ESPAÑA · MINISTERIO DE AGRICULTURA, ALIMENTACIÓN Y MEDIO AMBIENTE · JUNTA DE ANDALUCIA · Gobierno de Canarias

Air quality planning: Electrification



NO_2 maxh reductions:
25 - 30 µg m^{-3}
8-16%

PM_{10} maxh reductions:
< 8 µg m^{-3}
2-5%

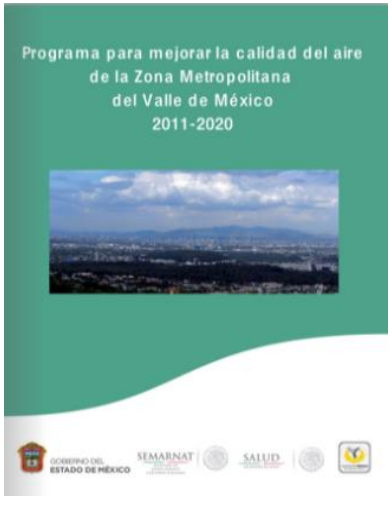
NO_2 maxh reductions:
30 - 35 µg m^{-3}
8-16%

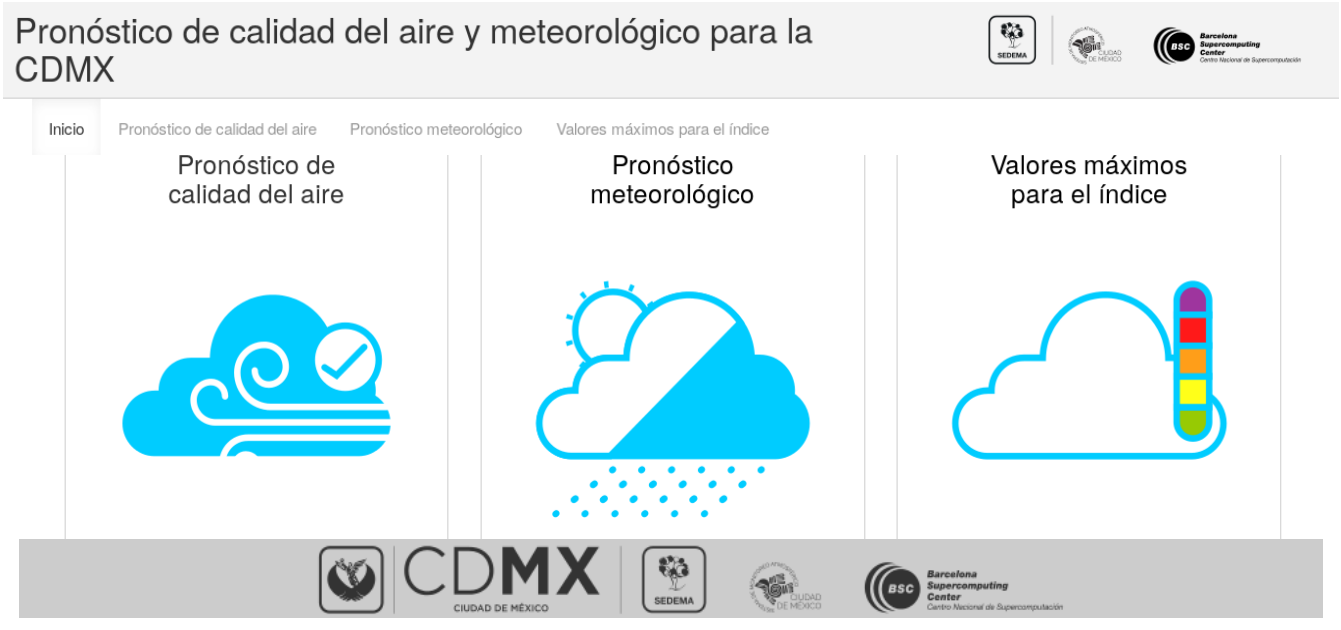
PM_{10} maxh reductions:
< 6 µg m^{-3}
2-4%

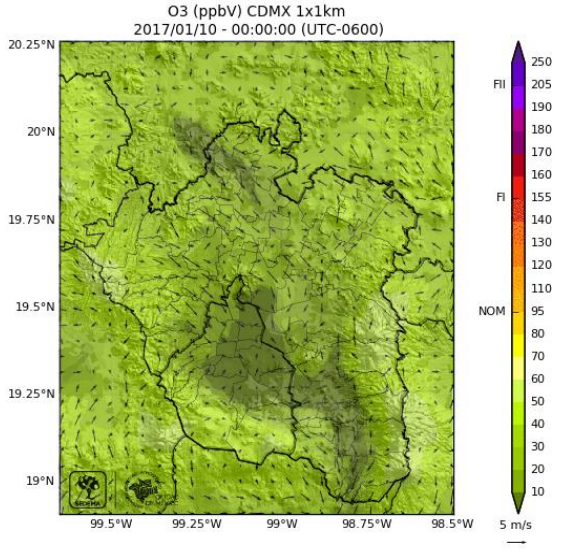
Soret et al. 2014, Atm. Env.

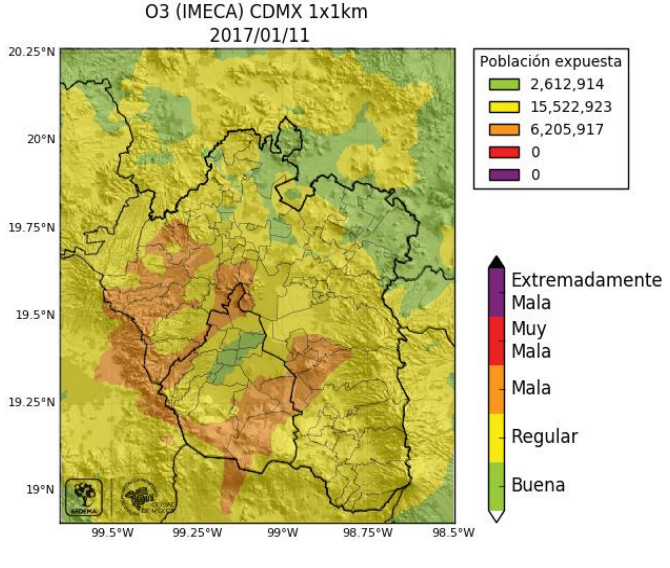
Air Quality System for México DF

- ✓ Complement the public information service provided by the monitoring network
- ✓ Know in advance the possibility that air pollution episodes occur
- ✓ Contribute to the development and evaluation of air quality plans (ProAire)





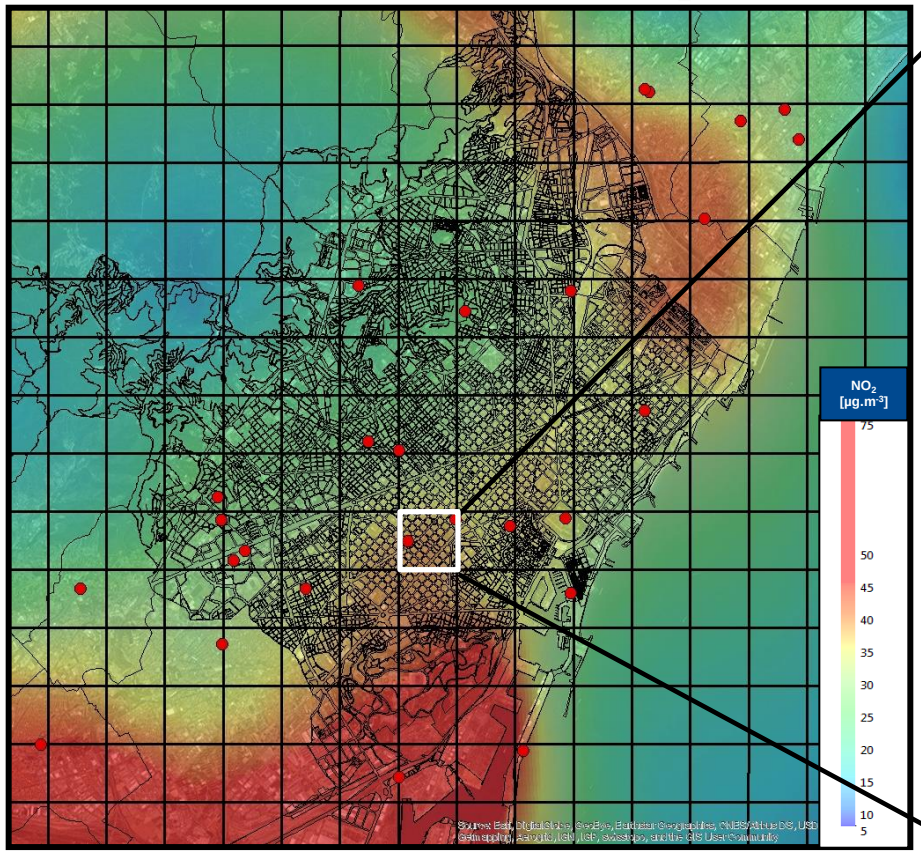




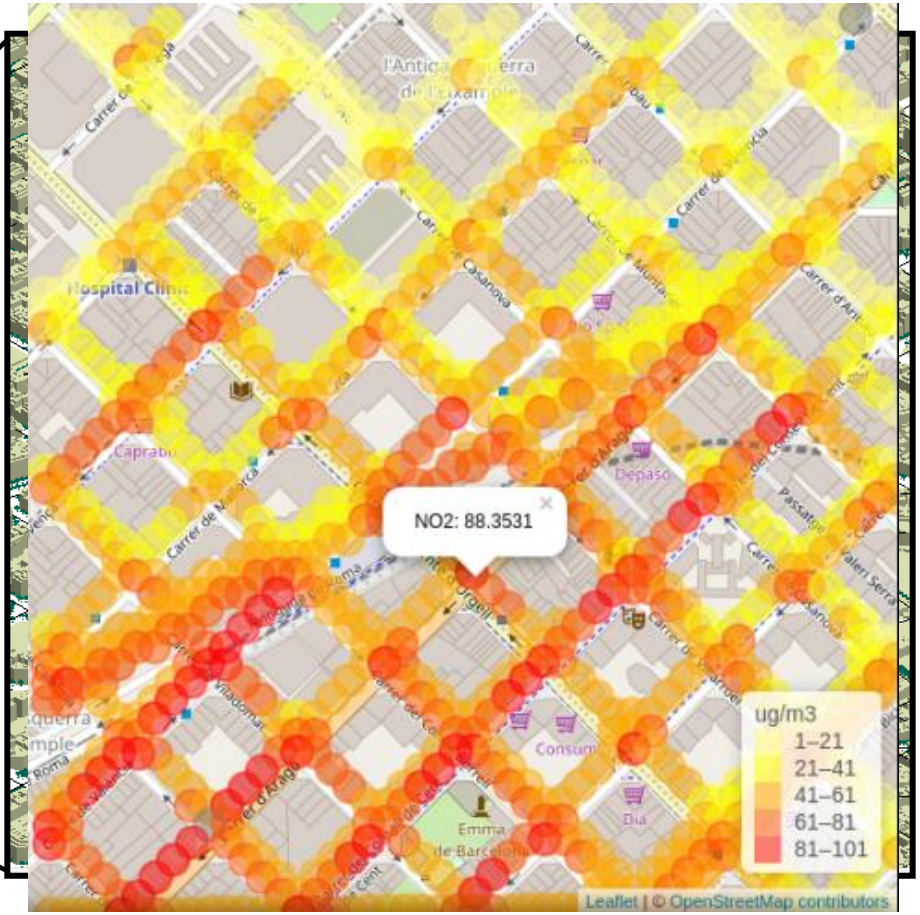
http://www.aire.cdmx.gob.mx/privado/pronostico/

The future of urban air quality

Where we are now

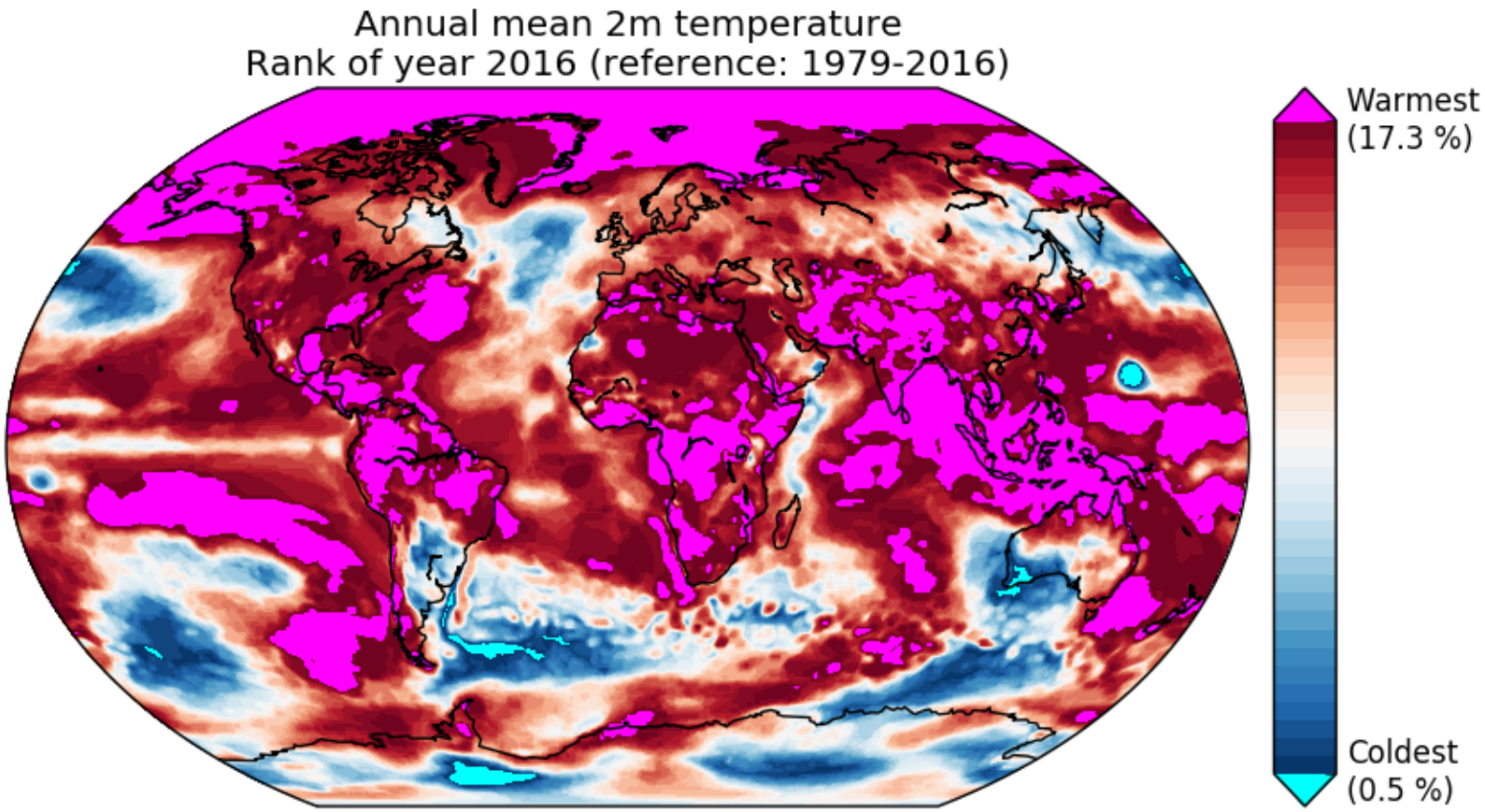


Where we want to go



The type of signal to be predicted

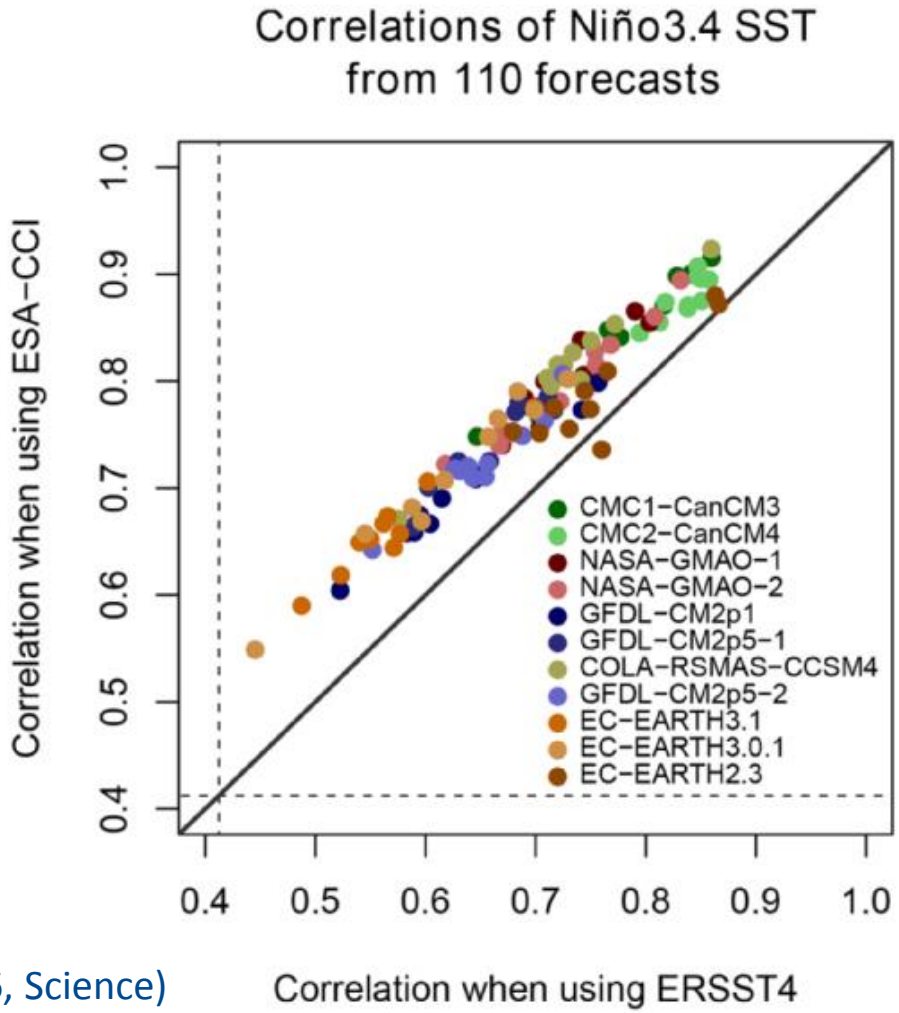
Ranking of the 2016 annual mean temperature over the last 37 years from ERA Interim.



Data: ERA-Interim. Figure: F. Massonnet - BSC

Verifying with uncertain observations

Need to take into account the large observational uncertainty in the forecast quality estimates. Models can also be used to estimate the quality of observational estimates.

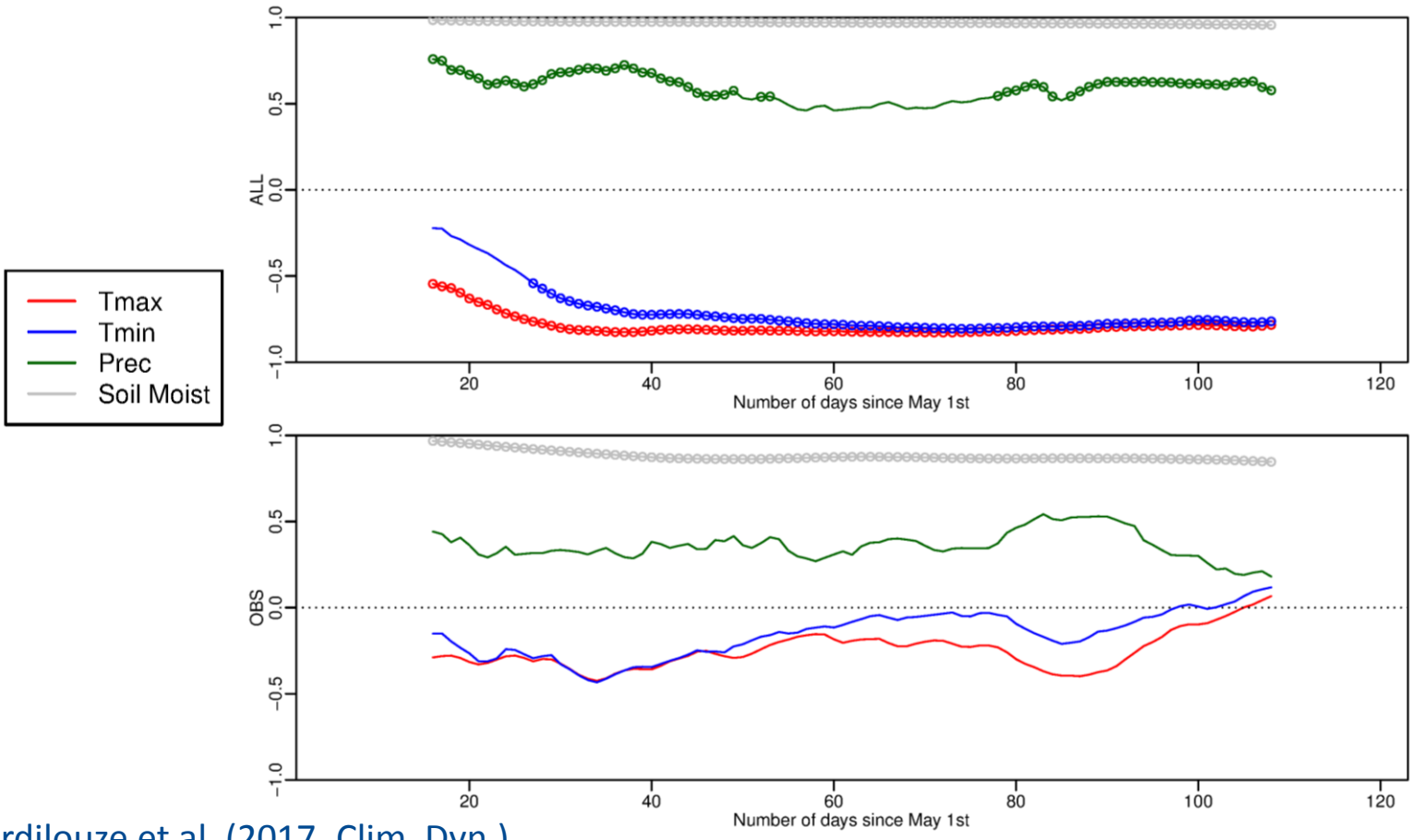


Massonnet et al. (2016, Science)

Drift helps uncovering model errors

Correlation between 1st of May total soil water content and 31-day running mean of variables from the SPECS multi-model seasonal forecast (top) and ERAInt (bottom) over North American Great Plains.

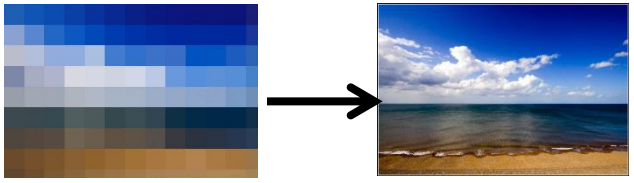
The model shifts quickly to excessive land-atmosphere coupling.



Ardilouze et al. (2017, Clim. Dyn.)

Effect of increasing the resolution

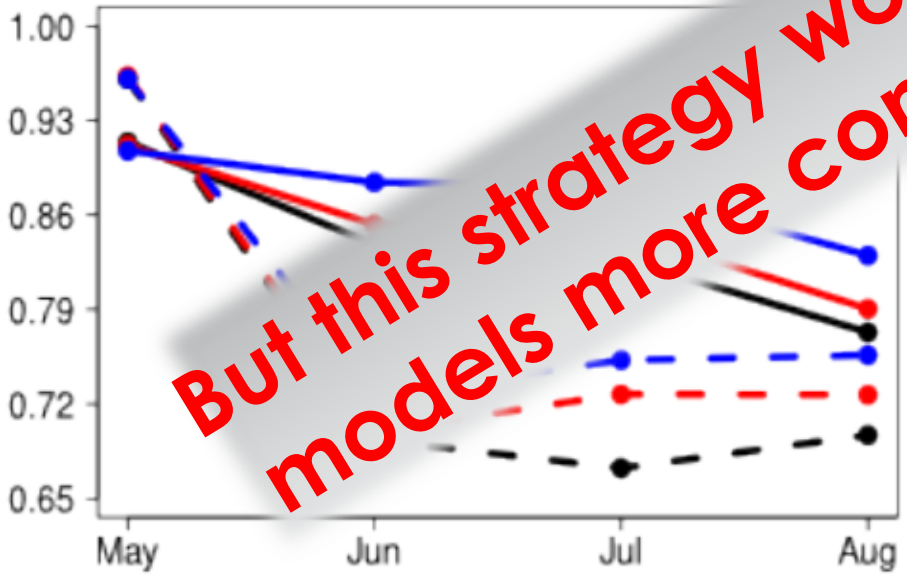
Forecast quality from EC-Earth3.1 seasonal hindcasts (1993-2009, Glorys2v1, ERAInt and ERALand initial conditions). Solid for ESA-CCI and dashed for ERSST. Blue for high resolution ocean and atmosph[illegible]ed for high resolution ocean, black for standard resolution.

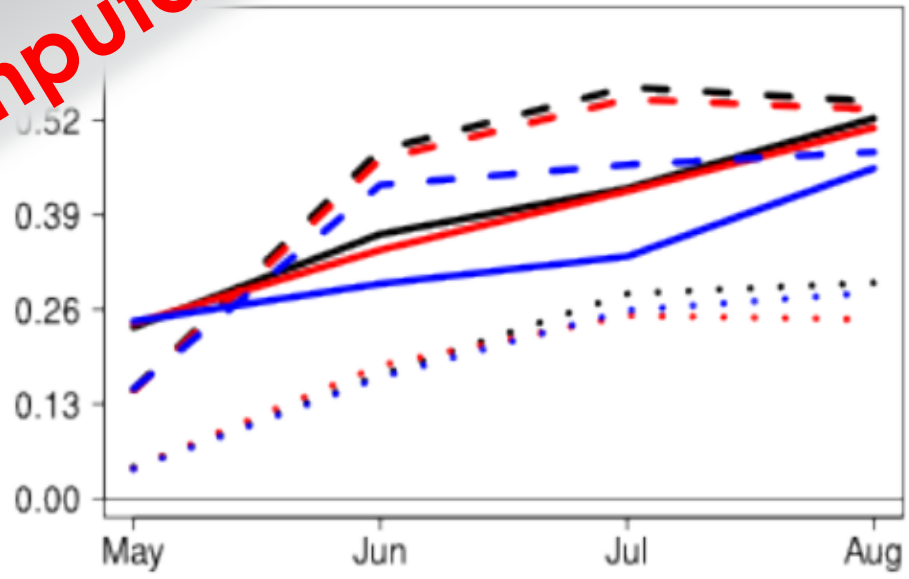


May star[illegible]

SRes
IRes
HRes

a) Correlation [illegible]ead and RMSE





But this strategy works better making the models more computationally efficient

Prodhomme et al. (2016, J. Climate)

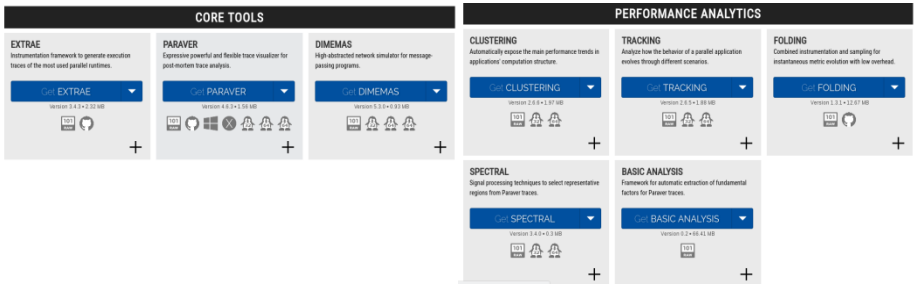
Performance activities at BSC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

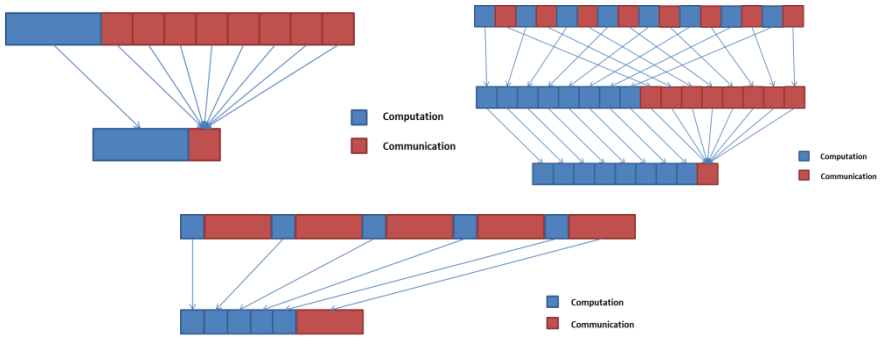
Collaboration with computer sciences department

BSC performance tools



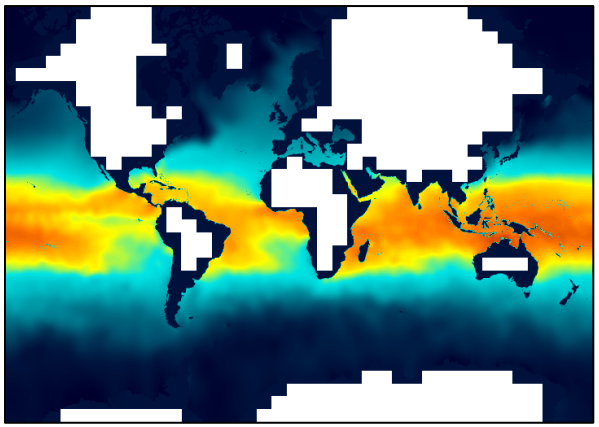
MPI communications optimizations

Reducing p2p and collective communications overhead



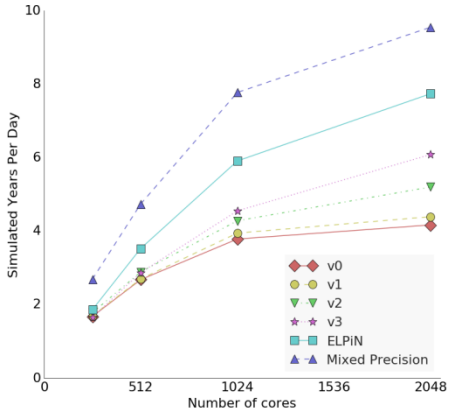
Exclude land processes in NEMO

Finding an optimal domain decomposition



Explore mixed precision

Which precision is needed in NEMO?



Potential overall improvement:

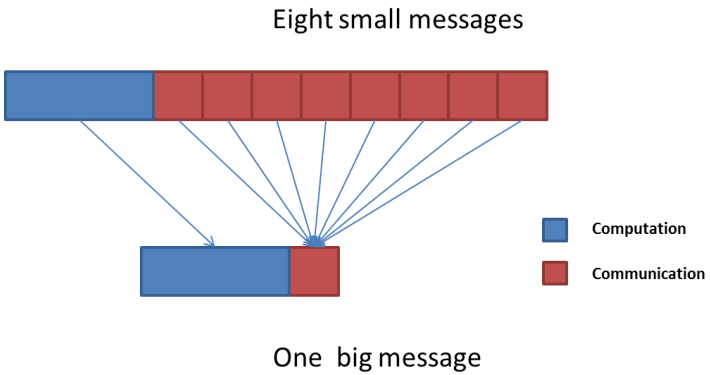
2x

miguel.castrillo@bsc.es

NEMO optimisation

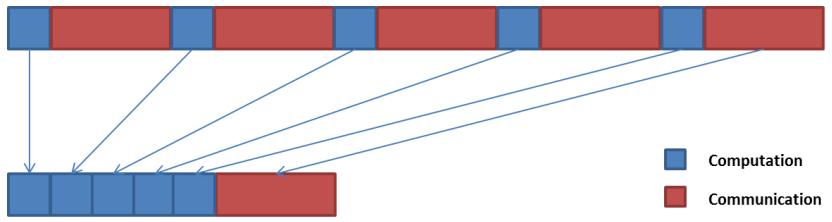
MPI message packing

Taking in account that NEMO is really sensitive to latency, messages aggregation is the best way to reduce the time invested in communications. Therefore, consecutive messages have been packed wherever the computational dependencies allow to do so.



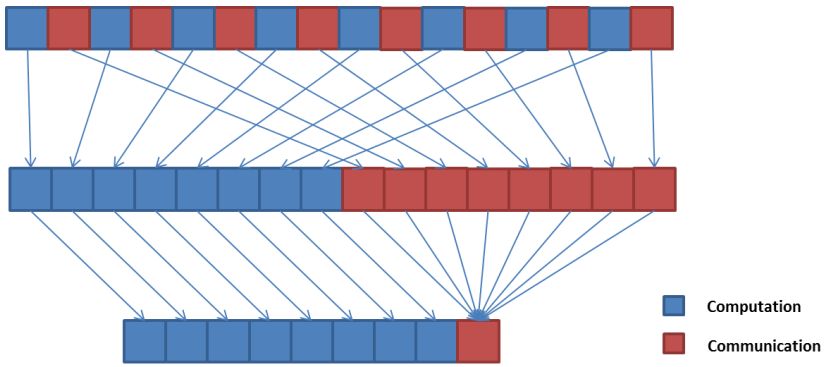
Convergence check reduction

Some routines use collective communications to perform a convergence check in iterative solvers. The cost of this verifications is really high, reaching a 66% of the time. Wherever the model allowed it, we reduced the frequency of this verifications in order to increase parallel efficiency.



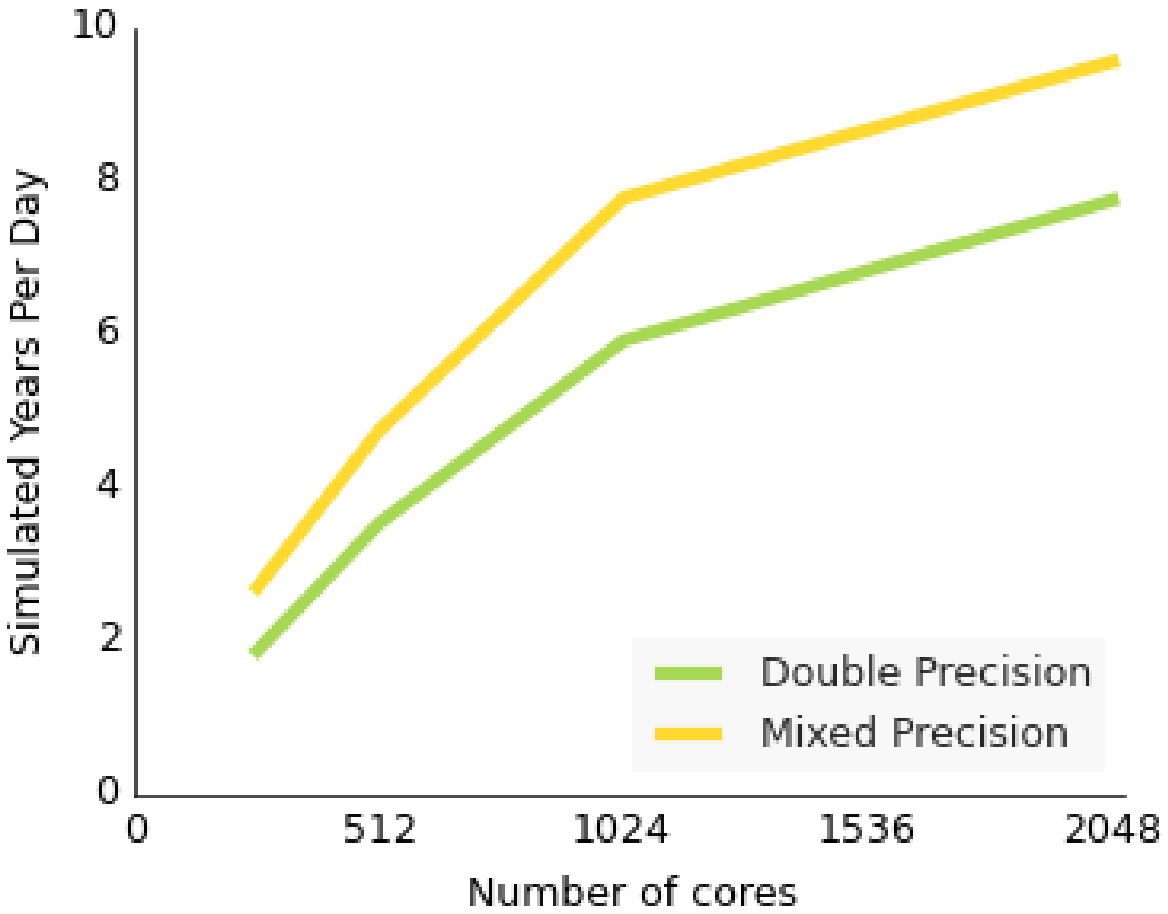
Reordering

In order to apply the message packing optimization to as many routines as possible, it was necessary to rearrange some computation and communication regions, taking into account the dependencies between them, to reduce the number of messages. This way it was possible to compute (and communicate) up to 41 variables at the same time, resulting in a dramatic reduction of the granularity.

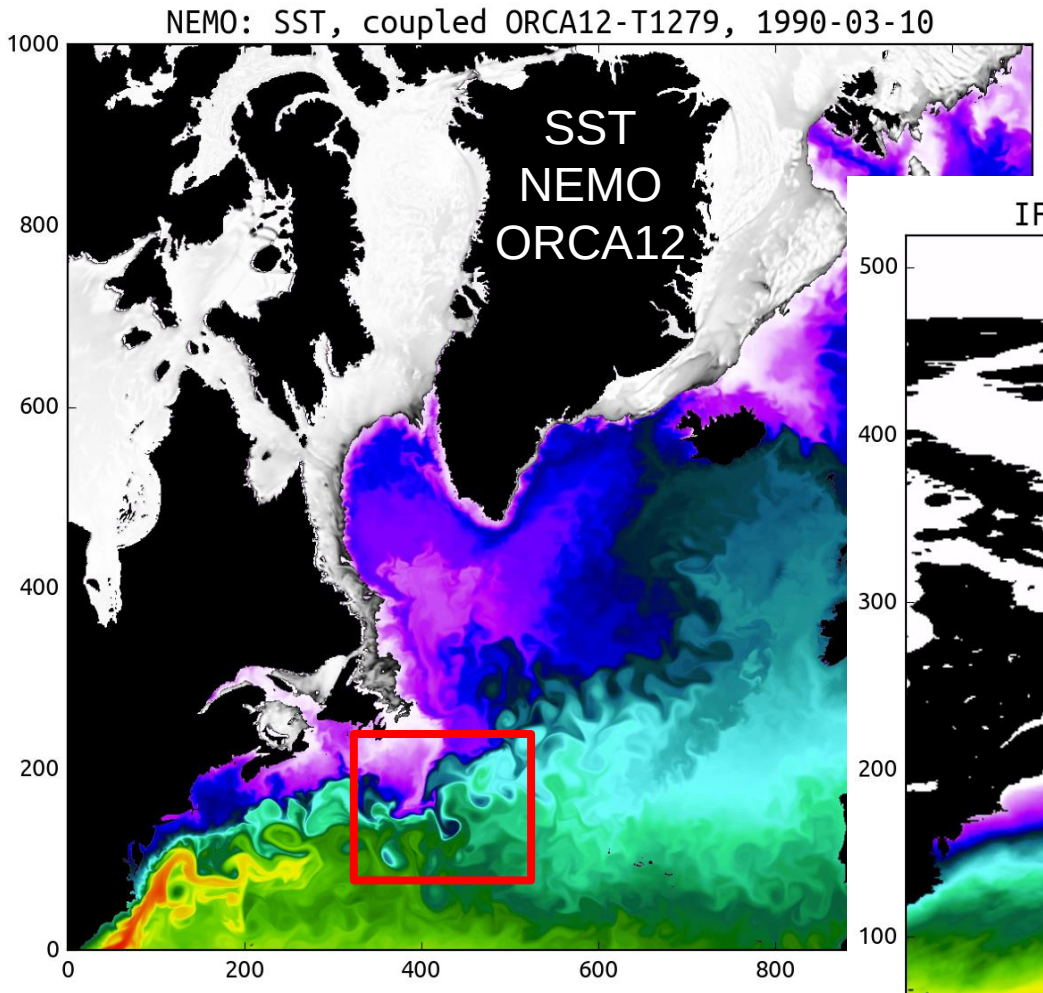


NEMO optimisation

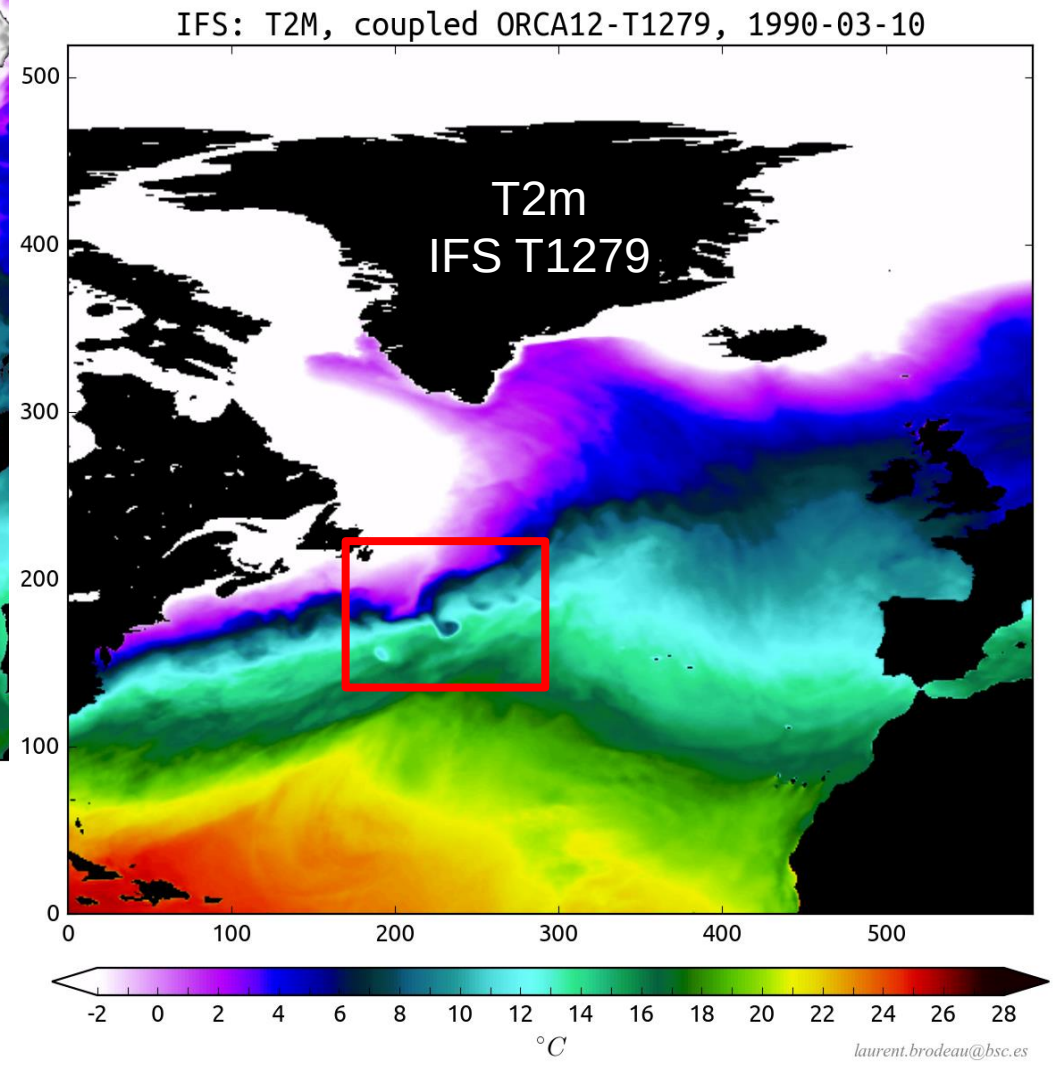
Speed up of the NEMO3.6 (ORCA025L75) code when switching some parts of the code from double to single precision.



NEMO-related activities at BSC



EC-Earth GLOBAL ORCA12-T1279
(ocean and atmosphere at ~15 km!)



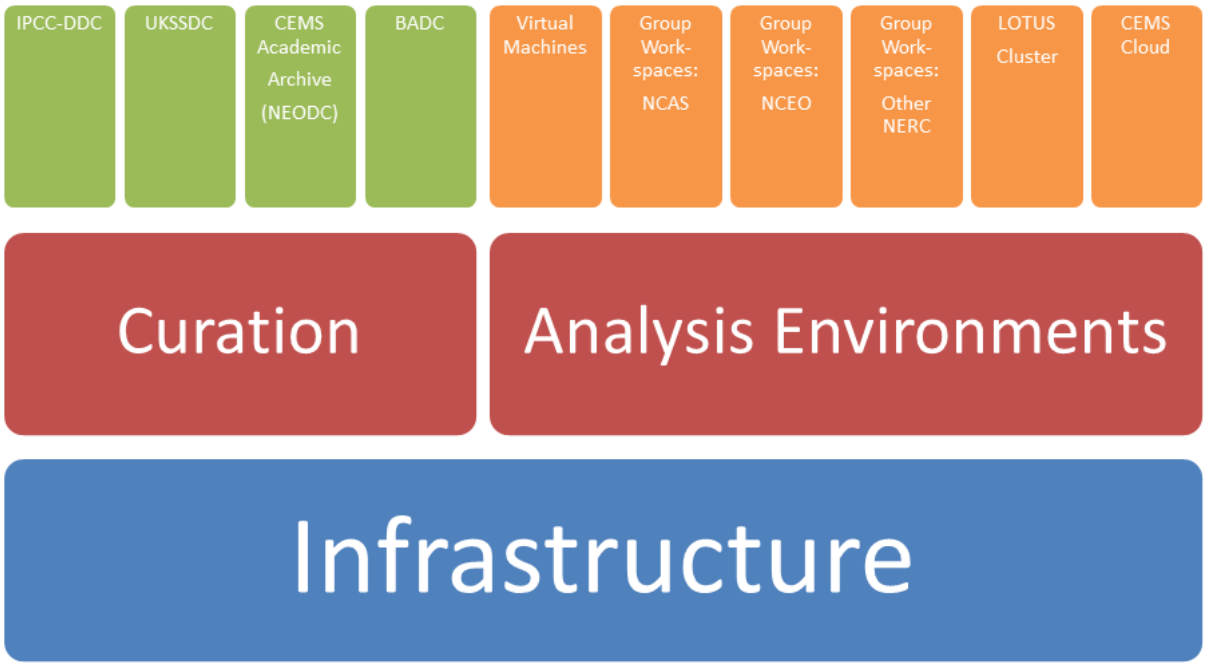
laurent.brodeau@bsc.es

Some ideas for a future activity

- **Earth sciences has struggled to benefit from the network**: many more actors exist at the Spanish level, need to work together.
- **Adapting to the evolving HPC technology is a challenge**: it proves almost impossible to make the most of a rapidly evolving technology (heterogeneous nodes, software, mobile data capture, visualisation, etc.) without dedicated resources.
- **A new paradigm has come to stay, user-driven research**: opportunities appear in a context where research and services are closer together in an area of clear social impact like Earth sciences.
- **Heterogeneity**: link to and merge our data with communities with larger impact (urban, arts, social).
- **Data, storage and standards**: the community has no mean to efficiently share the large volumes of data (observations and simulations) generated (close to 500 TB per year of public data only at the BSC) and provide basic analytics services.

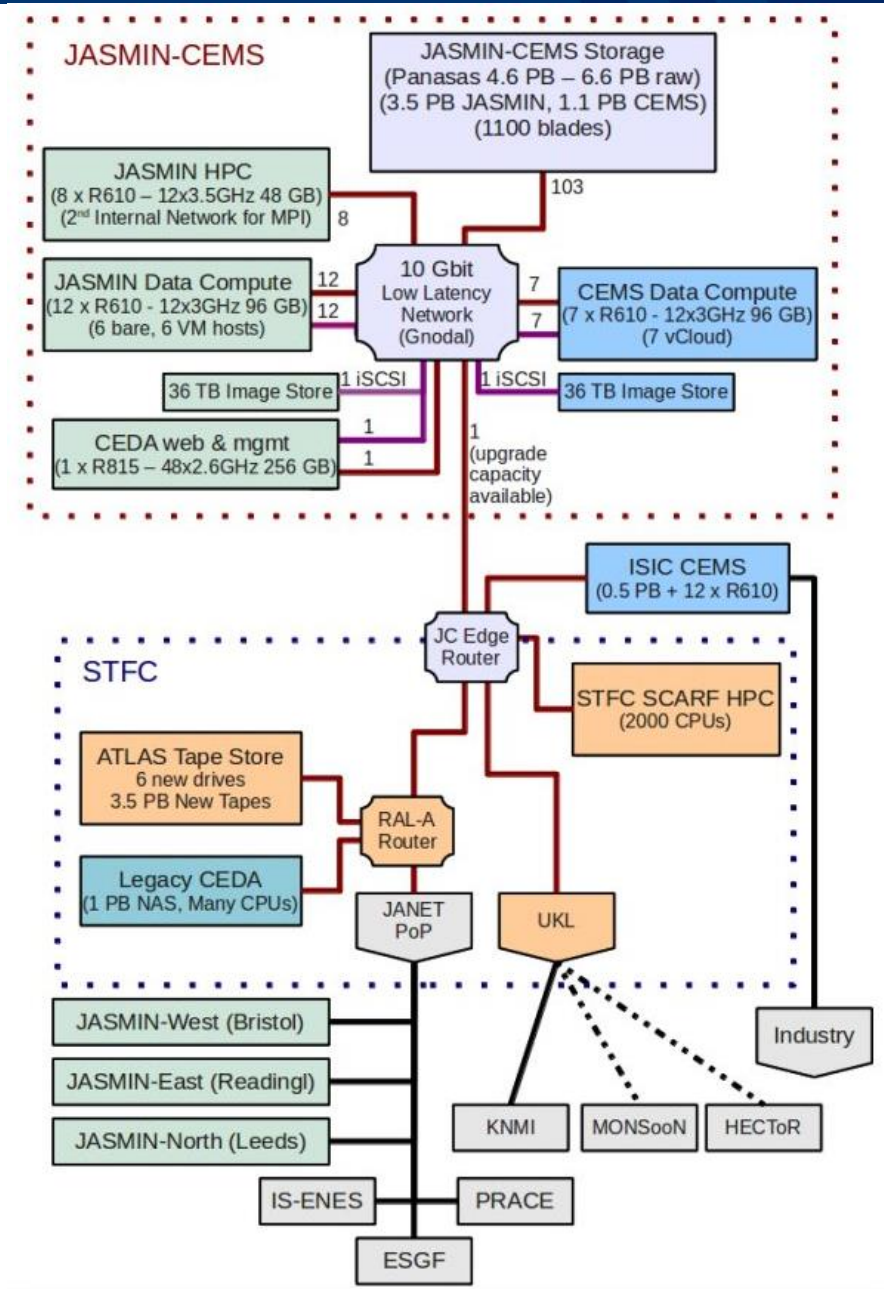
Jasmin: a national storage solution

- A Petascale storage and cloud computing for Big Data challenges in environmental science offering data-intensive computing (bringing compute to the data) and flexible data access.
- Nodes in the UK connected with 1 Gbit/s links.
- National funding: e-Infrastructure capital investment by the UK Government (NERC and NCAS).

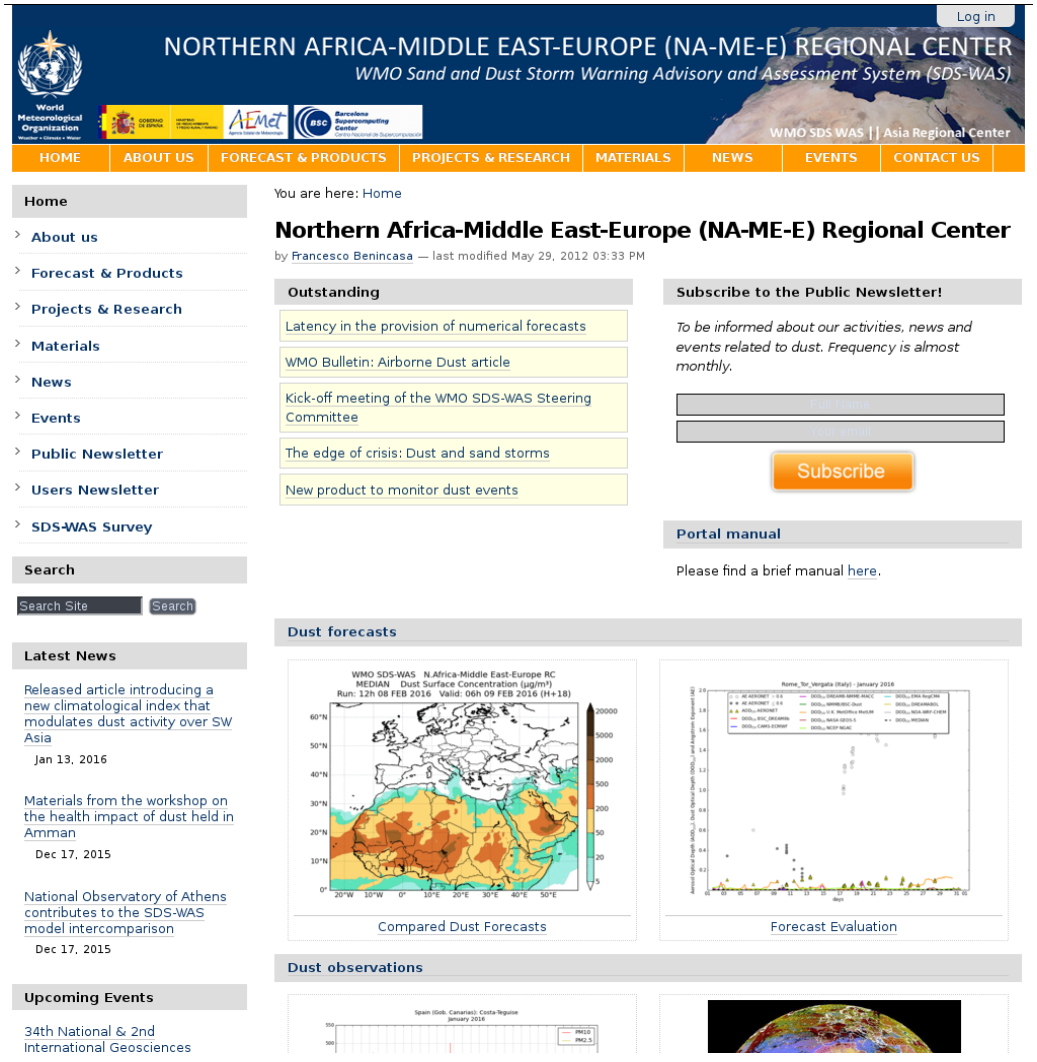


Jasmin: a national storage solution

- 9.3 PB of storage divided in archive and scratch according to the projects
- More than 370 computing cores provide local computation
- Thousands of users from English universities and European projects.
- Hosts data (observations and model outputs) from a large variety of projects
  - CMIP5/6 model outputs
  - Satellite data
  - In situ observations



SDS-WAS: some things are done



http://sds-was.aemet.es

- *Research and development* ***warning advisory & assessment system***
- *Ensemble of models provided by various international partners*
- *Provides:*
  - *. Model inter-comparison*
  - *. Multi-model products*
  - *. AERONET evaluation*
  - *. Satellite (MODIS, …) evaluation*
  - *. Numerical scores (BIAS, …)*
  - *. Datasets download*

The system is managed by a consortium of AEMET and BSC in Barcelona, Spain