A portable framework to generate initial conditions for IFS/OpenIFS from ERA reanalysis products using OpenIFS

Xavier Yepes-Arbós
Etienne Tourigny
Mario C. Acosta
Francisco J. Doblas-Reyes

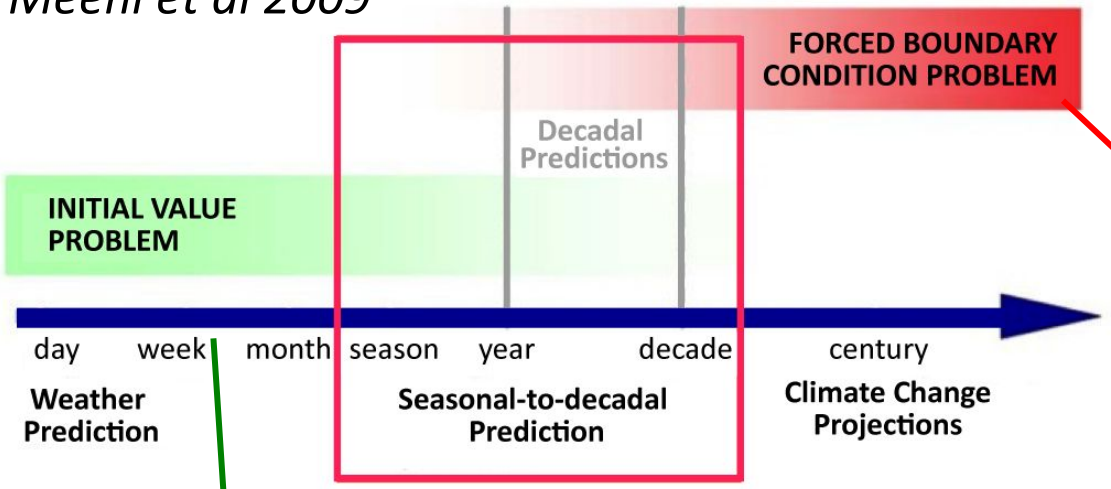
The research leading to these results has received funding from the EU H2020 Framework Programme under grant agreement no. 823988
This material reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains

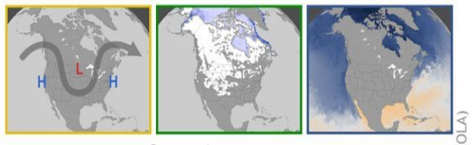
Introduction

- Models used in weather and climate forecasts must initialize the state of the Earth system (atmosphere, ocean, land and cryosphere) from observations

Cornerstones of climate prediction

Meehl et al 2009



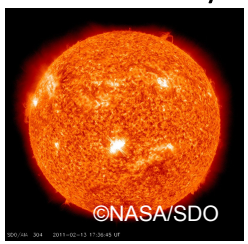


©Paul Dirmeyer (GMU/COLA)

Mariotti et al 2018

Predictability relying on good guess of future changes in the forcing

Solar Activity



Volcanic Aerosols

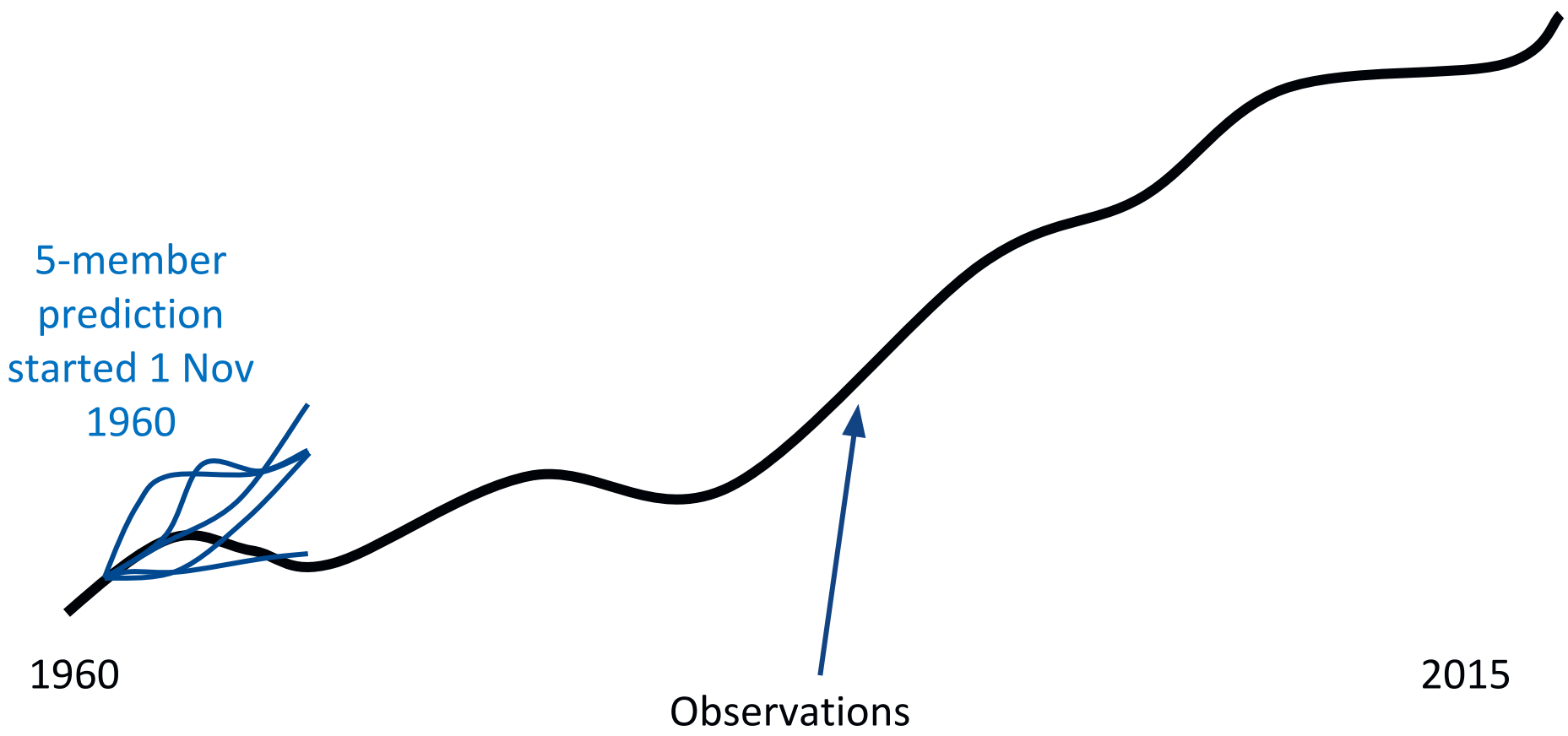
GHGs

Predictability arising from the memory of slow processes/components in the climate system

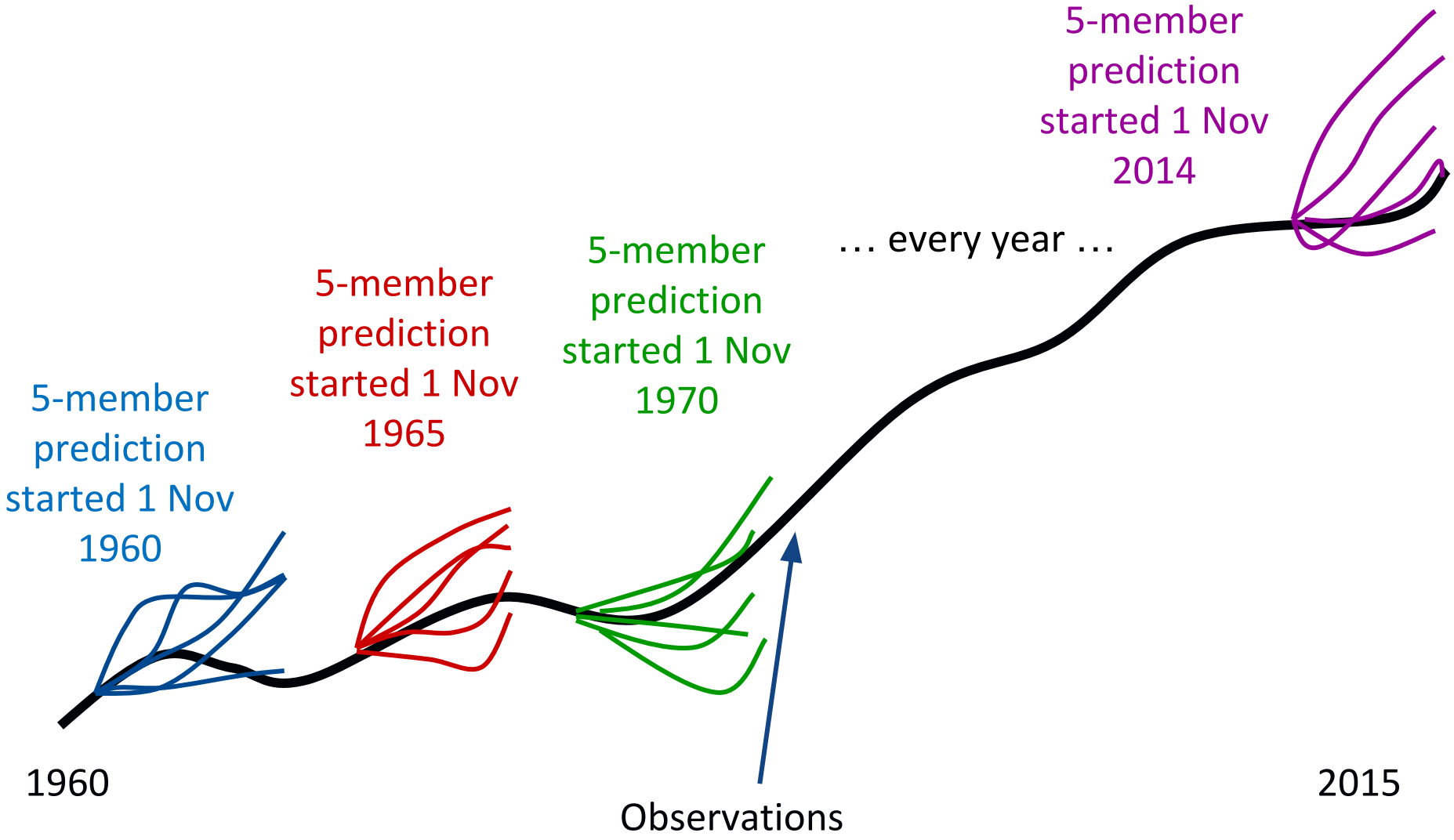
Introduction

- Models used in weather and climate forecasts must initialize the state of the Earth system (atmosphere, ocean, land and cryosphere) from observations
- Retrospective forecasts (or hindcasts) in weather or climate prediction mode are used to evaluate the model's performance and are typically initialized from reanalysis products

Climate prediction experiments



Climate prediction experiments

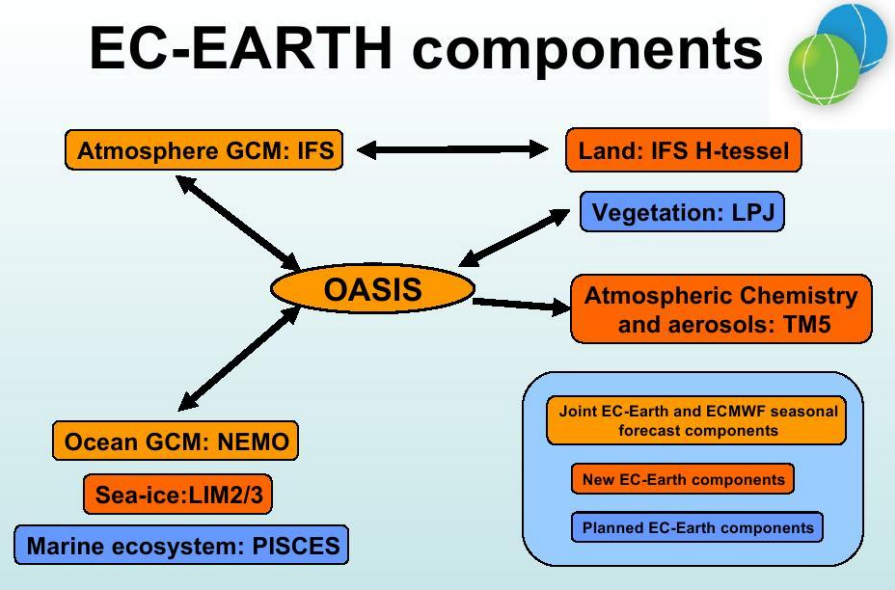


Introduction

- Models used in weather and climate forecasts must initialize the state of the Earth system (atmosphere, ocean, land and cryosphere) from observations
- Retrospective forecasts (or hindcasts) in weather or climate prediction mode are used to evaluate the model's performance and are typically initialized from reanalysis products
- However, the horizontal and vertical resolution of the reanalyses can be different and additional operations are needed to fit the grid model of the hindcast
- This is the case of EC-Earth

EC-Earth

- EC-Earth is a global coupled climate model, which integrates a number of components models in order to simulate the Earth system
- The two main components are IFS as the atmospheric model and NEMO as the ocean model



EC-Earth

- The EC-Earth climate model relies on atmospheric initial conditions derived from ERA-Interim
- These data must be obtained for several start dates (4 each year) for the entire period covered by the reanalysis product

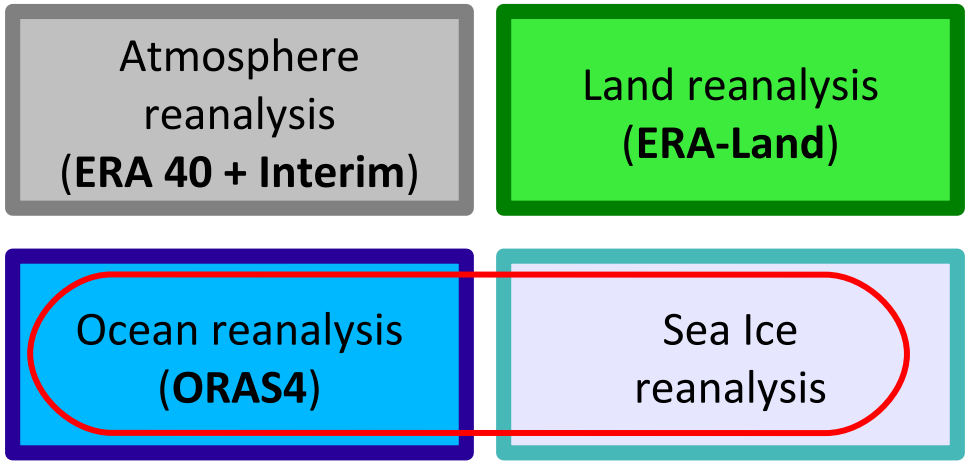
Initial conditions produced at BSC

ATM:
Interpolated to model grid with IFS + prepIFS

Now at BSC with OpenIFS + Autosubmit

Etienne Tourigny / Xavier Yepes

Initial Conditions



produced in-house

LAND:
Offline land-surface simulation with corrected fluxes from ERA-Interim

Emanuel Dutra / Etienne Tourigny

ocean

sea ice

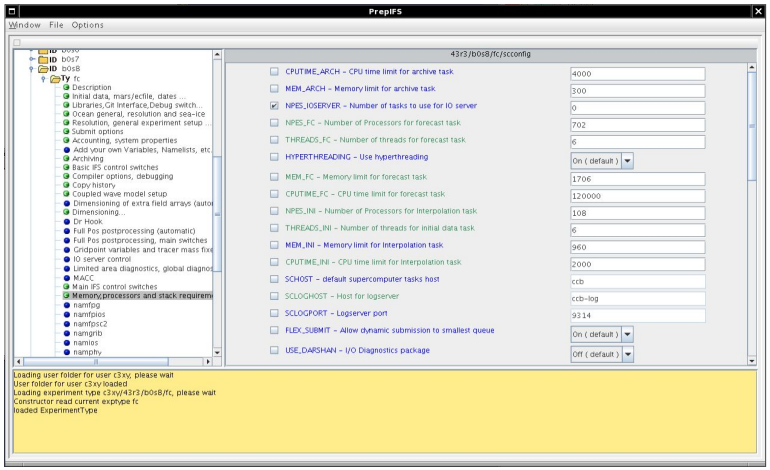
soil moisture

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

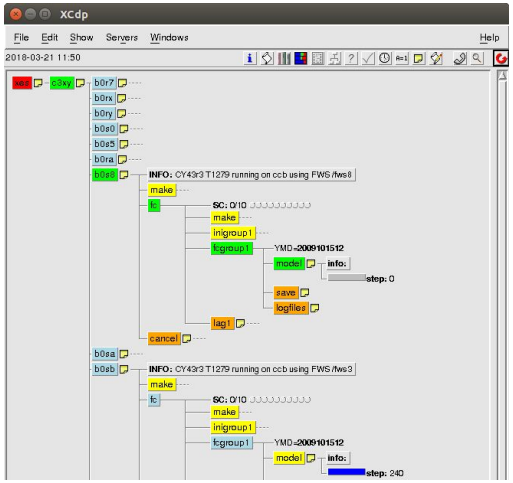
esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Current procedure

- The current procedure to get initial conditions derived from ERA-Interim is using the ECMWF HPC infrastructure
- It runs an IFS experiment created with the prepIFS tool and managed with the XCdp manager
- These two softwares are ECMWF-dependent



experiment submission



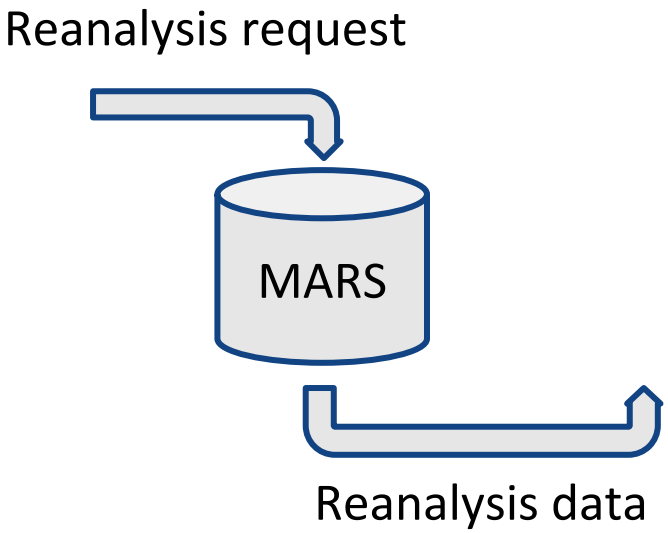
prepIFS

XCdp

Current procedure

The three main steps are:

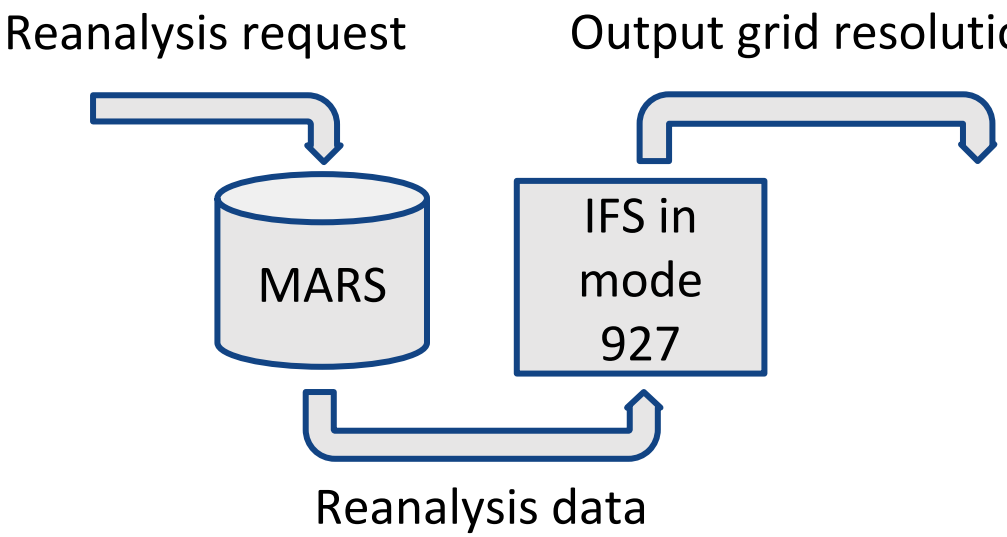
- Download the reanalysis data at its native resolution using MARS requests



Current procedure

The three main steps are:

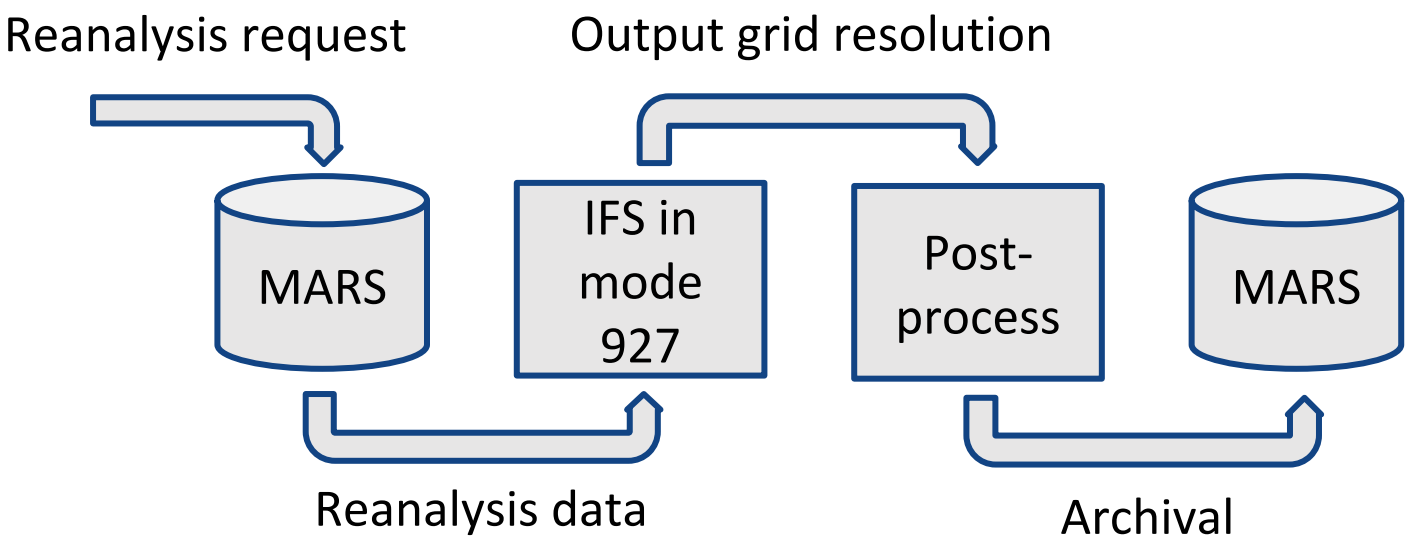
- Download the reanalysis data at its native resolution using MARS requests
- Running IFS in mode 927 to make use of FullPos, a powerful post-processing package that performs horizontal and vertical interpolations



Current procedure

The three main steps are:

- Download the reanalysis data at its native resolution using MARS requests
- Running IFS in mode 927 to make use of FullPos, a powerful post-processing package that performs horizontal and vertical interpolations
- Post-processing the result and archival on MARS



New procedure

- The current procedure is robust and relatively fast
- However, there are two main drawbacks:
  - It relies on ECMWF infrastructure and support
  - It is vulnerable to changes in the HPC platform used
- To overcome them, the original scripts have been simplified to run on other HPC platforms
- The new framework is made of three scripts:
  - Get initial data
  - Initial data interpolation
  - Initial data post-processing

Get initial data

- Retrieves reanalysis data from MARS at the native resolution:
  - Land surface and upper-air fields from ERA-Interim (T255L60), ERA5 (T639L137), etc
  - Optionally, surface fields from ERA-Land like experiments. Ongoing work to generate these surface fields with EC-Earth OSM.
- Runs on ecgate with some MARS commands, for efficiency and ease of porting
- Result can be automatically uploaded to a remote HPC machine (e.g. MareNostrum4 at BSC)

Get initial data

- Retrieves reanalysis data from MARS at the native resolution:
  - Land surface and upper-air fields from ERA-Interim (T255L60), ERA5 (T639L137), etc
  - Optionally, surface fields from ERA-Land like experiments. Ongoing work to generate these surface fields with EC-Earth OSM.
- Runs on ecgate with some MARS commands, for efficiency and ease of porting
- Result can be automatically uploaded to a remote HPC machine (e.g. MareNostrum4 at BSC)

However, it could be adapted to run on a remote HPC machine by using the CDS API instead of MARS client

Initial data interpolation

- OpenIFS 40r1 can be deployed and set up on a remote HPC machine to run on mode 927 (e.g. MareNostrum4 at BSC):

```
mpirun -n $NPES ./master.exe -v ecmwf -e $EXPVER \
      -t1. -ft0 -aeul 2>stderr.lst >stdo.lst
```

- Mode 927 uses FullPos to interpolate input GRIB files to a given output resolution
- It is necessary to adjust the OpenMP stack size from 128M to 512M
- Several output resolutions have been successfully tested, such as T255, T511 and T1279

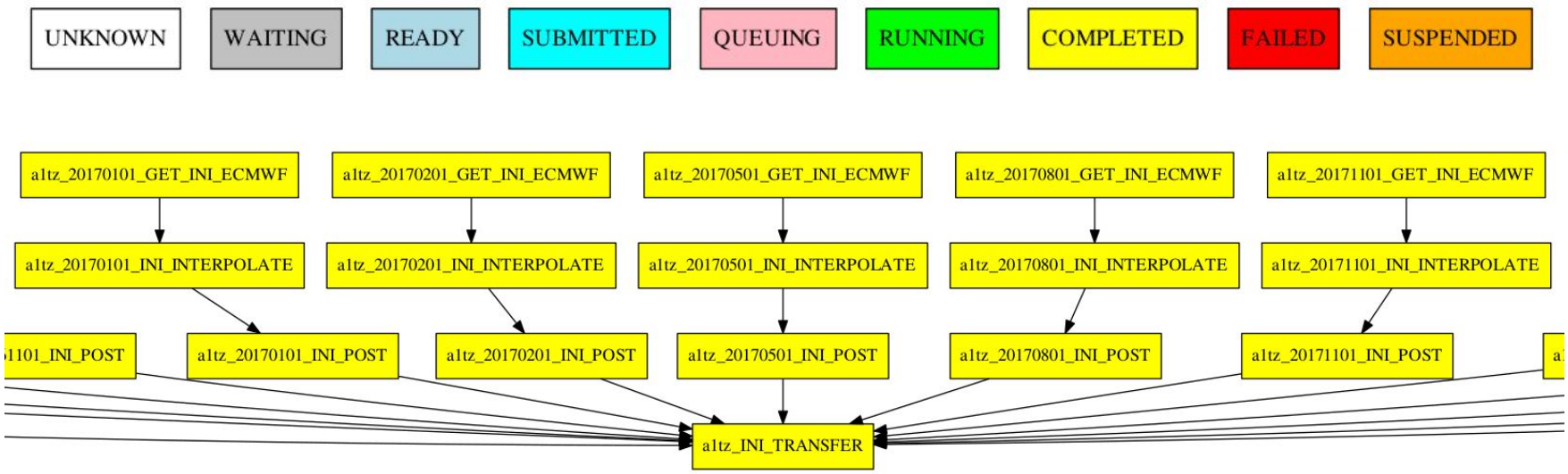
Initial data post-processing

- Performs post-processing such as:
  - Re-ordering the fields of the output GRIB files (replacing MARS get/put commands of the previous procedure)
  - Replace some land fields from the ERA-Land data
  - Perturb the 3D temperature field using Python (grib_api + random) for generation of additional members (used for seasonal prediction)
- Runs on a remote HPC machine (e.g. MareNostrum4 at BSC)
- Initial conditions data is ready to be used for IFS/OpenIFS
- Packaged in a .tar file compatible with the Autosubmit workflow manager for initializing EC-Earth historical/seasonal prediction experiments

Automation with Autosubmit

- Autosubmit is a python-based tool to create, manage and monitor experiments by using computing clusters, HPCs and supercomputers remotely via ssh
- It is an open source software developed by BSC and publicly available on the PyPI repository
- The three previous scripts are automated with Autosubmit to:
  - Minimize user intervention
  - Minimize errors
  - Optimize some processes

Automation with Autosubmit



Results

Initial conditions were created using ERA-Interim:

- For years 1979-2018 (last year available)
- Start of each calendar year (Jan 1st)
- Each seasonal prediction start date (Feb 1st / May 1st / Aug 1st / Nov 1st)
- At T255 resolution (a1tz) and T511 resolution (a1uc)
- ERA-Land conditions from existing experiments (gbg4, gbg6)

Results

Performance:

- Bottleneck in the get_inidata: can take from 5 minutes to one hour per start date, depending on MARS status, 24h sufficient to download entire ERA-Interim period with 5 dates/year
- Interpolation and post-processing take each < 1 minute per start date
- Using the Autosubmit workflow manager allows to do tasks in parallel: the interpolation and post-processing are done as soon as data is available for each start date
- Queuing times on MN4 much longer than computation times
- Improvements can be done by grouping several interpolation/post-processing tasks in one scheduler task

Results

Validation:

- Results were compared to ICs from those generated using prepIFS (b0q0, b0q2): very few differences were found
- Differences in climatology between 2 sets of ICs:
  - GG SFC fields: 100% identical
  - GG ML fields: differences < 1% in scarce grid points for cc and clwc
  - SH lnsp systematic differences < 0.00013243% , due to minor differences in orography
  - SH ML: {t, vo, d} small differences in scarce grid points, much smaller than observational errors, probably due to differences in compilers
- A seasonal prediction hindcast was done for all start years (1979-2018), and 4 start dates:
  - No instabilities were found
  - Results statistically similar to those initialized with ICs from prepIFS

Conclusions

- In the EC-Earth community there is a need to generate initial conditions to run climate prediction experiments
- The current procedure is not manageable for large periods and relies on ECMWF infrastructure
- To overcome these shortcomings, a framework to generate initial conditions for IFS/OpenIFS has been implemented being:
  - Portable across other HPC platforms rather than ECMWF
  - Easy to use
  - Uses OpenIFS to interpolate reanalysis data
  - Interpolations are fast, including configuration T1279L137
  - Results have almost no differences between the two procedures
  - Everything is automated with Autosubmit

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

xavier.yepes@bsc.es