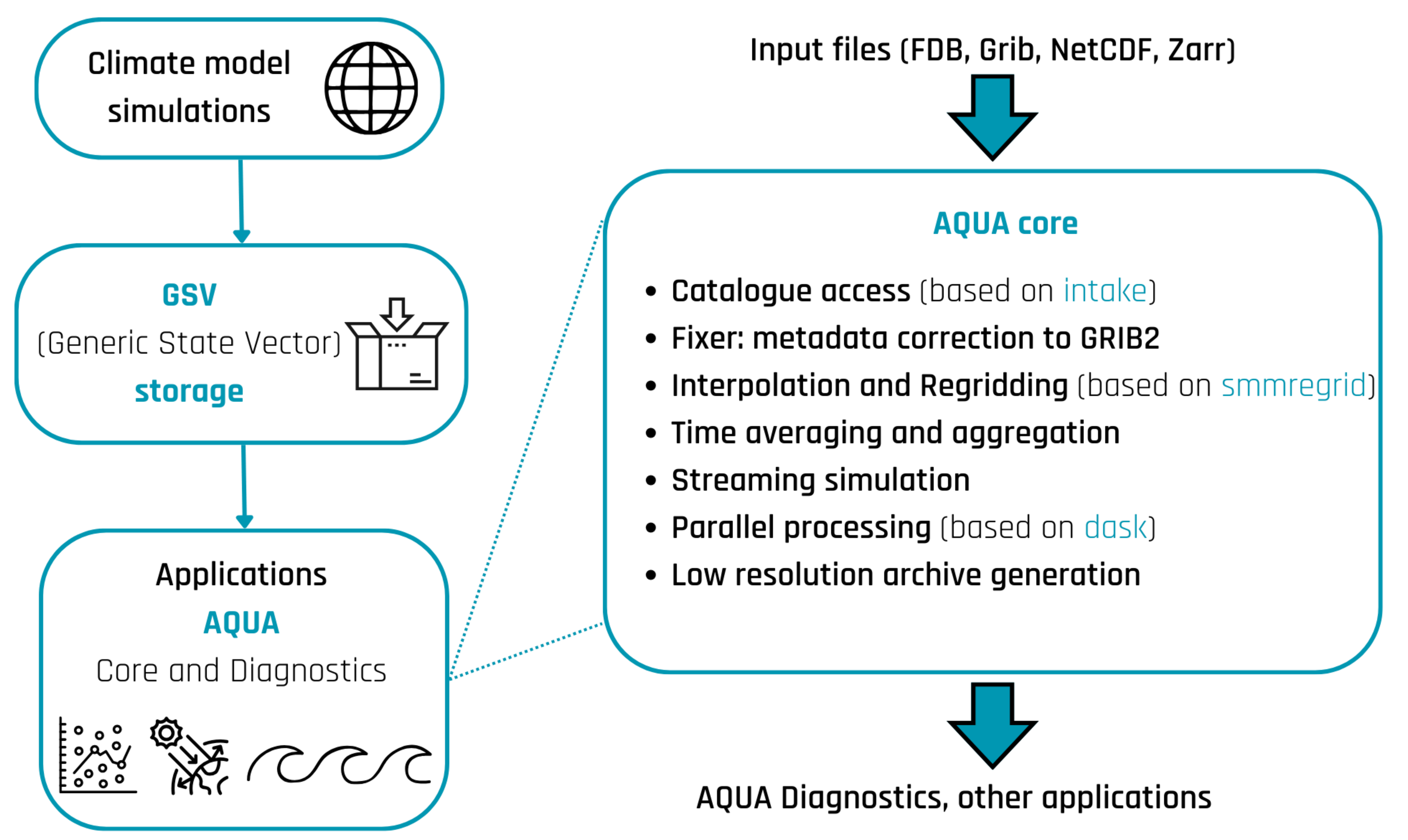
Climate Digital Twin (DE340)
The AQUA diagnostic framework: first results from a novel approach to climate data

The AQUA team

INTRODUCTION

What is AQUA?
AQUA is the **Application for Quality Assessment and Uncertainty quantification** of the Climate Adaptation Digital Twin: it is composed by a **core engine facilitating data access**, combined with a series of **modular and independent diagnostics** to be run continuously to monitor and evaluate the Climate DT simulations.

Why AQUA?
Though many available diagnostics suites already exist to analyse data from global climate models, **AQUA has been specifically developed to deal with heavy km-scale datasets**, with the goal of unifying and simplifying climate data access for all Climate DT users. At the same time, using a selected range of focused diagnostics **aims at providing responsive feedback to climate model developers** to improve the quality of climate simulations, through the integrations within the Climate DT workflow. In the future, it will support more process-oriented diagnostics and uncertainty quantification measures.



AQUA KEY FEATURES

AQUA represents a portable and modular **Python3 application**, which builds on top of community open-source tools such as **Xarray, Dask, CDO and Intake**, providing fast and lazy data access to multiple sources (**GRIB, NetCDF, FDB, Zarr**) in the simplest way possible. The philosophy is to allow the user to specify only what is strictly needed to identify the data, **hiding in ready common configuration files the difficult technical details**. AQUA includes regridding and averaging functionalities, the capability to deal with streamed data and the possibility to convert to a common metadata vocabulary different datasets or experiments.

- **Catalog access**: a 3-level hierarchy to identify model experiments and different sources has been introduced building on intake. Extra capabilities has been added to process fdb-data.
- **Regridding**: flexible and fast interpolation to regular grids from any type of grids is supported, building on CDO interpolation weights and on the *smmregrid* package capabilities.
- **Fixing**: variable names and related metadata from each dataset can be converted to a GRIB2 standard with a series of "fixes" based on *eccodes* and configurable *yaml* files.
- **Spatial averaging**: the AQUA reader incorporates area for any type of grids, and this is used for doing spatial averaging operations. Areas are computed using CDO.
- **Temporal aggregation**: efficient wrapper to *xarray* built-in functions deal with time averaging.
- **Parallel processing**: building on *dask,* AQUA can scale efficiently on large datasets.

CODE SAMPLE

The **core of AQUA** is the *Reader()* class, which must be instantiated to provide access to the data and to AQUA features. Data can be loaded by specifying the 3-level hierarchy model-experiment-source on which the AQUA catalogue is built. In this example we will access IFS data, specifying that interpolation should be performed toward a 1-degree grid and data averaging on monthly timescale:

```
from aqua import Reader
reader = Reader(model="IFS", exp="control-1950-devcon",
                source="hourly-native", freq='monthly', regrid="r100")
```

Once the *Reader()* is operative, we can select one variable with the *retrieve()* method and then remap it with the *regrid()* method. **All the operations are done lazily with dask** so that no actual computation is done until necessary.

```
data_2t = reader.retrieve(vars='2t')
data_2t = reader.regrid(data_2t['2t'])
```

We can then **aggregate the data** to the monthly frequency with *timmean()*, compute the **area-weighted average** (AQUA stores cell areas in the reader) with *fldmean()* and then plot the resulting timeseries:

```
monthly = reader.timmean(data_2t)
```

```
global_mean = reader.fldmean(monthly)
global_mean.plot()
```

With few lines of code, we can process any data available in the catalog.

AQUA DIAGNOSTICS

The AQUA framework incorporates several **state-of-the-art diagnostic tools**, designed for evaluating the general performance of the models and to identify model biases and drifts:

- ***Performance indices** (PIs):* an objective assessment of model biases compared to trusted observational sources;
- ***Radiation budget**:* quantification of the radiation budget at the top of atmosphere (see Figure 2);
- ***Atmospheric global mean timeseries and biases***: timeseries of temperature, precipitation, or any variable required;
- ***Teleconnection indices**:* computation of NAO and ENSO indices and their pattern correlations;
- ***Ocean circulation evaluation**:* evaluation of anomalies of oceanic variables such as temperature, salinity, density and mixed layer depth;
- ***Tropical cyclones detection, tracking and zoom in**:* detection and tracking of tropical cyclones. Possibility to save a set of prescribed variables (e.g., precipitation, wind, wind gust etc...) in the vicinity of TCs centres at original model resolution.

A selection of these diagnostics is planned to be run within the workflow of Climate DT so that it can provide monitoring and evaluation feedback to modellers.

On top of that, the AQUA framework features **frontier diagnostics**, which facilitate the analysis of high-resolution, i.e., km-scale hourly data:

- ***Sea surface height (SSH) variability**:* useful to evaluate surface ocean dynamics, including mesoscale eddies (see Figure 1);
- ***Tropical rainfall**:* assessment of extreme precipitation in the tropics (see Figure 3).

The modular environment of AQUA is planned to be used to extend diagnostics, both frontier and state-of-the-art, in the following phases of Destination Earth.

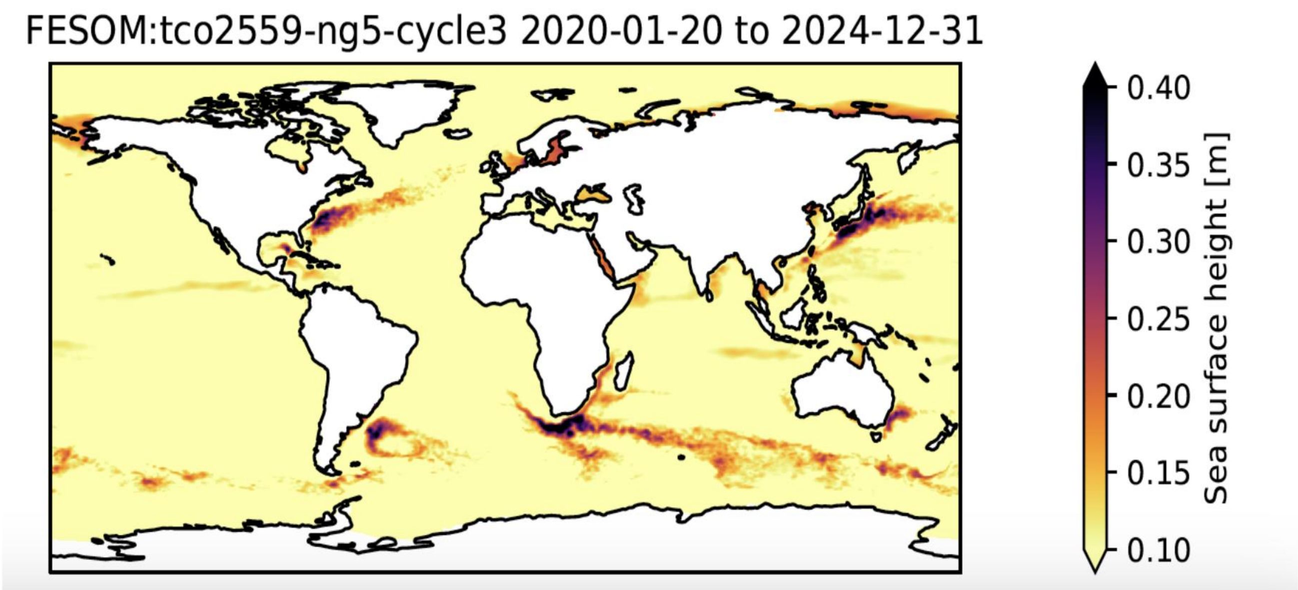


Fig 1: Sea surface height (SSH) variability (in m) in NextGEMS Cycle3 FESOM-tco2559-ng5, re-gridded into a 25-km grid.

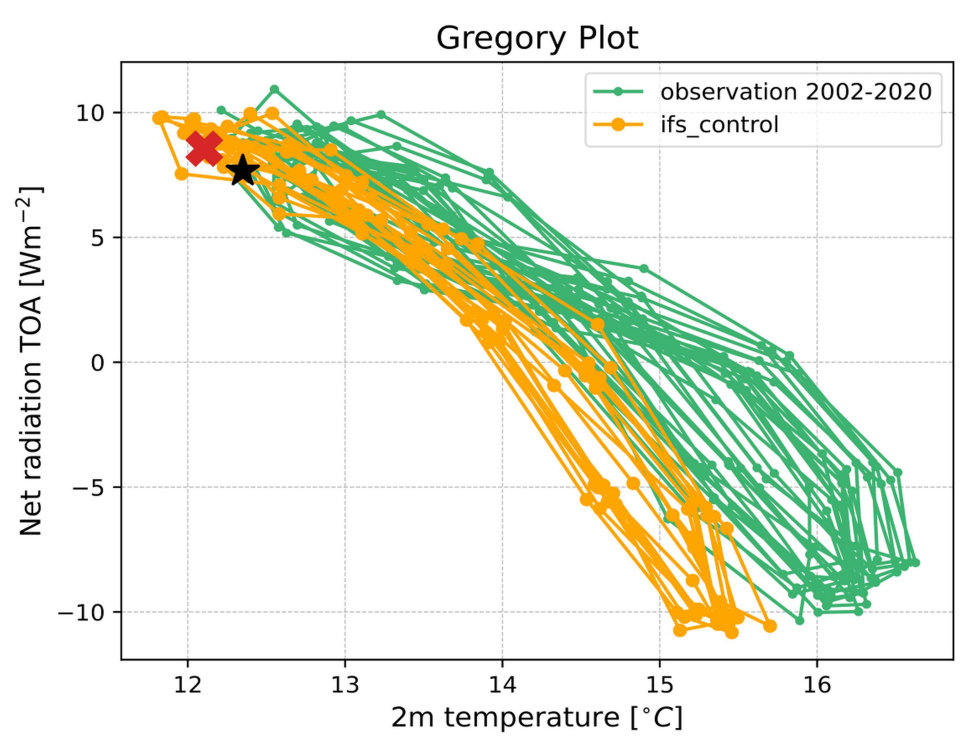


Fig. 2: Gregory plot for the Climate DT IFS-NEMO control-1950 simulation (orange) comparing the worldwide averaged TOA net radiation (Wm−2) and 2m temperature (°C). The black star and red cross represent the initial and last time series values. Comparison is against observational/reanalysis data, ERA5 2m temperature, and CERES TOA net radiation (green).

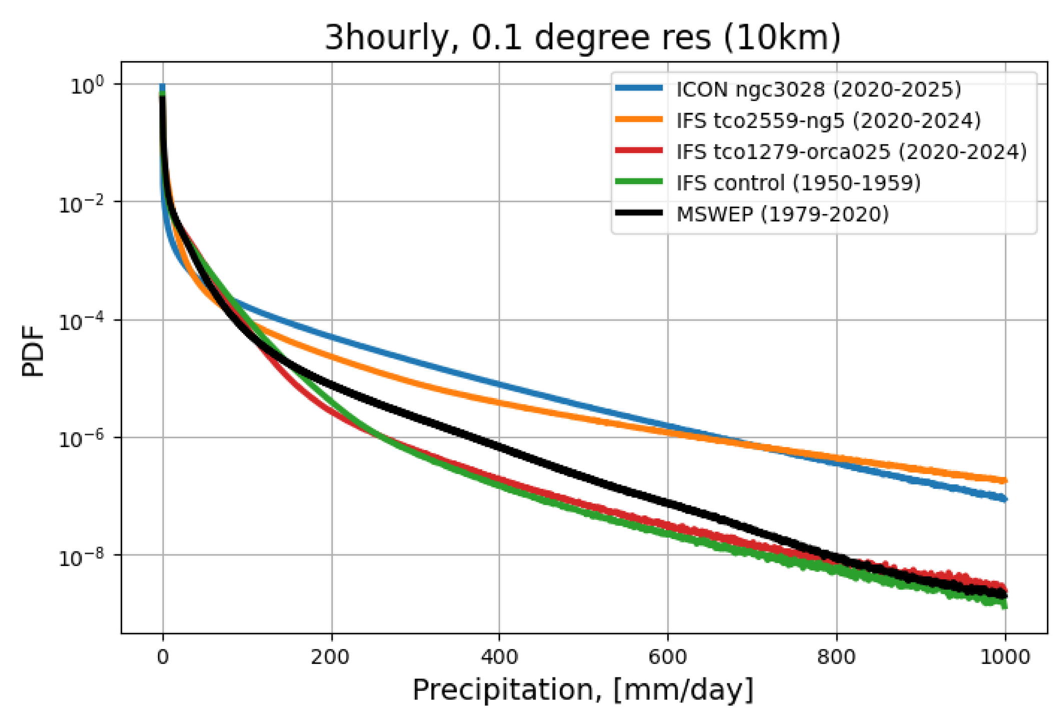


Fig 3: PDFs of tropical precipitation, between 15°S—15°N from 3-hourly, 0.1 resolution data in NextGems Cycle3 ICON-ngc3028 (blue line), NextGems Cycle3 IFS-FESOM-tco2559-ng5 (orange line), NextGems Cycle3 IFS-NEMO-tco1279-orca025-cycle3 (red line), Climate DT IFS-NEMO control-1950 (green line), MSWEP (black line) and ERA5 (grey line).

Funded by the European Union
Destination Earth
implemented by ECMWF esa EUMETSAT