EGI-ACE Open Call no.1

Checkpoint meeting with Shepherds

Large sample testing of high-resolution distributed hydrological model

Miguel Caballer/UPV
Albrecht Weerts/Deltares & Wageningen University

Dissemination level:

Disclosing Party:

Recipient Party:

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

Outline

- *Background about the scientific use case*
- *Ambition, Impact and Challenges*
- *Integration Support*
- *Capacity Requirements*
- *Timeline*

Background about the scientific use case

The objective is to perform large sample testing of high-resolution distributed hydrological model to identify where process description or used datasets (forcing, geofabric, etc) fall short and need improvement. At the same time, we hope to be able to identify where changes in discharge regimes occur around the globe either from changes in climate, the environment or society.

The starting point of this work is to focus on locations where we have long enough observational records (i.e. discharge, volume estimates) like CAMELS (USA, Brazil, Chile, UK), EWA/GRDC databases, locally available discharge networks (Poland, Norway, Sweden, Australia) and the GRanD database for upstream reservoirs (~5000).

We will make use of available datasets (ERA5 reanalysis or locally available high-resolution gridded rainfall datasets) to force the distributed hydrological model. Given the impact of model parameter uncertainty, for each location models will be run multiple parameter sets to allow for uncertainty analysis.

TEAM: WUR (The Netherlands) , Deltares (The Netherlands)

Ambition, Impact, Challenge(s)

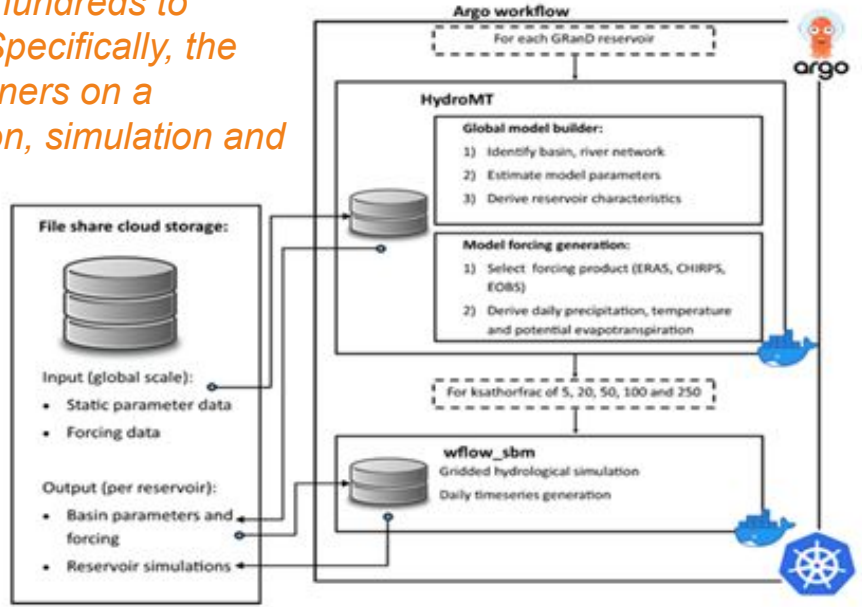
The benefit of the cloud in comparison to traditional HPC environments is that it allows to run the hydrological models in parallel on hundreds to thousands of CPUs (depending on resources available). Specifically, the Argo workflow manager, using different Dockerized containers on a Kubernetes cluster, allows for effective scheduling, creation, simulation and evaluation of hydrological model data.

For the current project, we will use this approach for the datasets described in the previous slide to:

1. *Generate model forcing and parameters,*
2. *run the model,*
3. *perform uncertainty analyses,*
4. *evaluate results for 40-50 years of historical data simulation (1970-present),*
5. *if available and time allows, cover the future period to assess the impact of future climate change*

Resource providers:

- *CESGA: Already in use..*



Integration Support

We want to deploy our ARGO workflow (previous slide) using the EGI Cloud Compute resources. This workflow was tested on Azure Kubernetes cluster before.

We need:
Experts to deploy Kubernetes clusters.

- *IM used to deploy K8s cluster with ARGO Workflow system.*

Experts to integrate EGI-Datahub for external data access.

Capacity Requirements

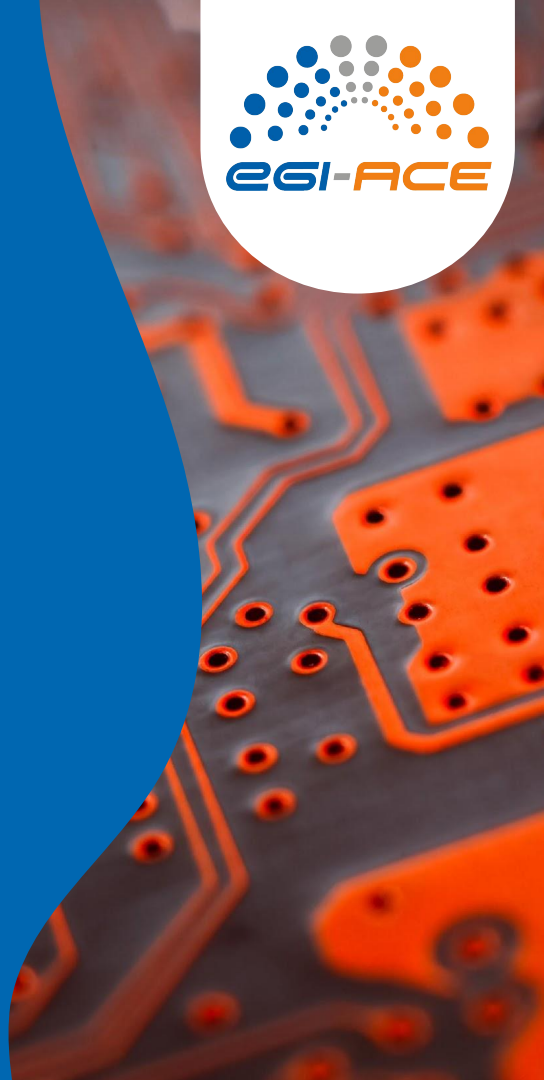
Per application we would need a CPU with about 8 Gb of memory and 25-200 Gb of temporary storage to enable running of the workflow and generate results.

Total capacity needed:1000 CPU (has been reduced to 200) with 4GB memory / CPU and with 25-100 TB disk in total.The amount of storage per CPU will vary depending on the size of the catchments and the amount and form (grids/timerseries) of output that is being saved (therefore the suggested reduction of 200Tb to 100Tb if this is a problem).

In addition we would need a storage bucket or file share where static data is being stored for generating the models and model forcing: 1 fileshare / bucket 15Tb.

Timeline

- Current Status:
  - Orders created in EOSC Marketplace
  - K8s cluster with 25 WNs (8 CPUs, 32 GB of RAM) and 5 TB of block storage created at CESGA
    - Deployed using IM.
    - Input data manually ingested.
  - Big run in progress (~15000 jobs).
  - Preparing next run.
- Next planned actions:
  - Evaluating the usage of EGI DataHub for data storage.



Thank you!

Contact: **egi-ace-po@mailman.egi.eu**
Website: **www.egi.eu/projects/egi-ace**

EGI Foundation

@EGI_eInfra

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.