

EGU2020-13614

<https://doi.org/10.5194/egusphere-egu2020-13614>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Speeding-up data analysis: DIVAnd interpolation tool in the Virtual Research Environment

Charles Troupin¹, Alexander Barth¹, Merret Buurman², Sebastian Mieruch³, Léo Bruvry Lagadec⁴, Themis Zamani⁵, and Peter Thijsse⁶

¹University of Liège, GeoHydrodynamics and Environment Research, Astrophysics, Geophysics and Oceanography, Liège, Belgium (ctroupin@uliege.be)

²Deutsches Klimarechenzentrum GmbH (DKRZ), Hamburg, Germany

³Alfred Wegener institute, Bremerhaven, Germany

⁴Fremer, Service Ingénierie des Systèmes d'Information, Plouzané, France

⁵Greek Research and Technology Network (GRNET), Athens, Greece

⁶MARIS BV, Nootdorp, The Netherlands

A typical hurdle faced by scientists when it comes to process data is the installation and maintenance of software tools: the installation procedures are sometimes poorly documented, while there is often several dependencies that may create incompatibilities issues. In order to make easier the life of scientists and experts, a *Virtual Research Environment* (VRE) is being developed in the frame of SeaDataCloud project.

The goal is to provide them with a computing environment where the tools are already deployed and datasets are available for direct processing. In the context of SeaDataCloud, the tools are:

- WebODV, able to perform data reading, quality check, subsetting, among many other possibilities.
- DIVAnd, for the spatial interpolation of in situ measurements.
- A visualisation toolbox for both the input data and the output, gridded fields.

DIVAnd

DIVAnd (Data-Interpolating Variational Analysis in n dimensions) is a software tool designed to generate a set of gridded fields from in situ observations. The code is written in Julia a high-performance programming language (<https://julialang.org/>), particularly suitable for the processing of large matrices.

The code, developed and improved on a regular basis, is distributed via the hosting platform GitHub: <https://github.com/gher-ulg/DIVAnd.jl>. It supports Julia-1.0 since its version 2.1.0 (September 2018).

Notebooks

Along with the source code, a set of jupyter-notebooks describing the different steps for the production of a climatology are provided, with an increasing level of complexity: <https://github.com/gher-ulg/Diva-Workshops/tree/master/notebooks>.

Deployment in the VRE

JupyterHub (<https://jupyter.org/hub>), is a multiple-user instance of jupyter notebooks. It has proven an adequate solution to allow several users to work simultaneously with the DIVAnd tool and it offers different ways to isolate the users. The approach selected in the frame of this project is the Docker containers, in which the software tools, as well as their dependencies, are stored. This solution allows multiple copies of a container to be run efficiently in a system and also makes it easier to perform the deployment in the VRE. The authentication step is also managed by JupyterHub.

Docker container

The Docker container is distributed via Docker Hub (<https://hub.docker.com/r/abarth/divand-jupyterhub>) and includes the installation of:

- The Julia language (currently version 1.3.1);
- Libraries and tools such as netCDF, unzip, git;
- Various Julia packages such as PyPlot (plotting library), NCDatasets (manipulation of netCDF files) and DIVAnd.jl.
- The most recent version of the DIVAnd notebooks.

All in all, Docker allows one to provide a standardized computing environment to all users and helped significantly the development of the VRE.