VACUMM – A PYTHON LIBRARY FOR OCEAN SCIENCE

By S. Raynaud[(1)], G. Charria[(2)], J. Wilkins[(1)], V. Garnier[(2)], P. Garreau[(2)], S. Theetten[(2)]

[1]ACTIMAR, Brest, France
[2]IFREMER, ODE/DYNECO/PHYSED, Brest, France

Abstract

VACUMM is an open-source Python library for processing data from observations and numerical models. The library is now used for several years in research and operational contexts, for instance for producing figures and reports, validating models, converting data, or making simple or advanced diagnostics. In this paper, we introduce how the library is built, and we present two applications of its use: one in an operational context and one in a research context.

A generic library

VACUMM (Validation Analysis Comparisons Multi-Models) is a Python library and collection of scripts that has been mainly designed to validate the MARS3D[1] numerical simulations. However, VACUMM is now used for both processing of ocean and atmosphere data in general, including model and observations. It has a complete dedicated documentation that includes the full list of modules and scripts, and a large collection of tutorials, examples, gallery and test cases: http://www.ifremer.fr/vacumm. It is open-source software with a CeCiLL[2] licence, and is open to new contributions.

VACUMM is written in the Python language, an interpreted language widely used for a long time in the ocean/atmosphere/climate community. This object-oriented language suits both computer experts and scientists, with applications going from the interactive side to the batch or operational sides. The low performance of an interpreted language are generally not a concern with Python since it can be easily interfaced for instance with C or Fortran languages, as it is the case in VACUMM for some interpolation routines. VACUMM has been written as a layer on the top of the CDAT[3] and Matplotlib libraries. CDAT (or UV-CDAT[4]) is used in the climate data analysis community, mainly for facilitating the comparisons between CMIP[5] climate models; it uses the high performance numerical library Numpy, and adds geographical and time informations, regridding capabilities of the ESMF[6], and a collection of climate tools. Matplotlib provides a Matlab® like interface to produce high quality figures. VACUMM typically makes an intensive use of CDAT variables that have a similar structure to those of NetCDF variables, with numerical data, axes and attributes. This makes it easy to exploit them such as when plotting a map of Sea Surface Temperature (SST) calling "`map2(sst)`" with axes, title and units properly formatted, or regridding the same variable "`regrid2d(sst, newgrid)`".

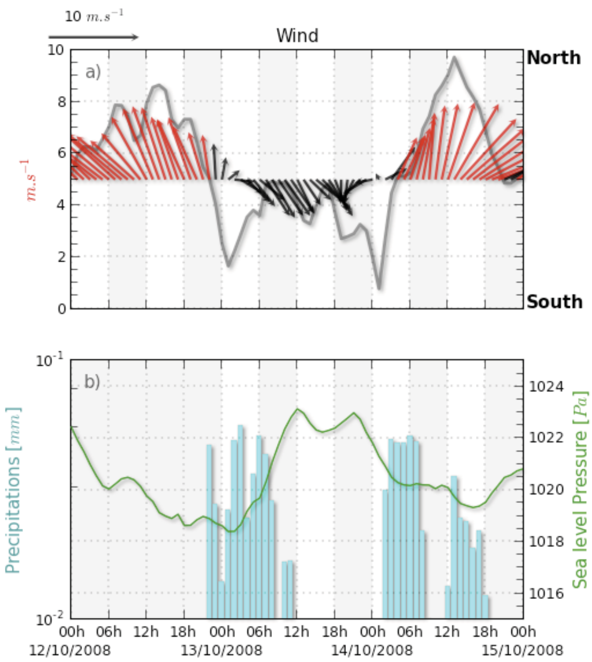

Figure 1: Example of 1D time series of meteorological data including wind vectors, wind intensity (grey line), sea surface pressure (green line), and precipitation bar plot plotted using VACUMM functions (source code: http://www.ifremer.fr/vacumm/tutorials/misc.plot.advanced.meteo.html).

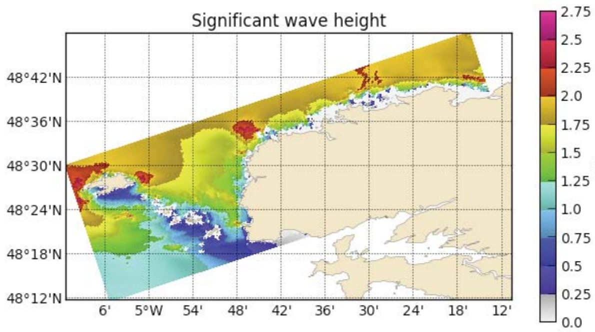

Figure 2: Example of a map of significant wave height generated using VACUMM function (source code: http://www.ifremer.fr/vacumm/tutorials/misc.plot.advanced.swan.html).

[1] MARS3D: coastal ocean numerical model developed in Ifremer - http://wwz.ifremer.fr/mars3d
[2] CeCILL license, compatible with GPL license:: http://www.cecill.info/licences/Licence_CeCILL_V1.1-US.html
[3] Climate Data Analysis Tools - http://uvcdat.llnl.gov/
[4] Ultrascale Visualization Climate Data Analysis Tools - http://uvcdat.llnl.gov/
[5] Coupled Model Intercomparison Project
[6] Earth System Modeling Framework

The library is made of a generic component and several specialized ones. The generic part has tools for dealing with most common 1D and 2D plots (Fig. 1 and 2), regridding (even between curvilinear grids), for interpolating and masking, for time conversions, formatting and splitting, for plotting, for easily reading NetCDF and other formats, for filtering data or performing statistics on huge datasets (Fig. 3). The generic subpackage also offers more advanced tools, for instance for working with remote files, managing advanced configurations, logging, working with xml and namelists. VACUMM also contains specialized interfaces to deal with *in situ* profiles or satellite data (see following section). In addition, a generic interface to specialized gridded data such as outputs from oceanic (MARS3D, NEMO, HYCOM) and atmospheric models (WRF), that facilitates post-processing, has been integrated. Retrieving the mixed layer depth becomes as simple as "`mld=DS('output.nc','mars').get_mld(mode='deltadens')`". This interface uses an extension to CF (Climate and Forecast) conventions that helps discovering variables in NetCDF files and format in-memory variables. The generic interface integrates some advanced diagnostics part of other more thematic subpackages: bathymetry and shorelines, tidal diagnostics, spectral analyses, and physical diagnostics like geostrophic velocity (Fig. 4) or mixed layer depth.

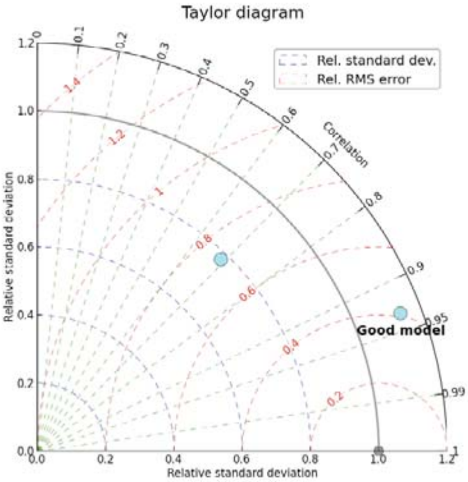

Figure 3: Example of a Taylor diagram to compare standard deviation and correlation between two datasets plotted using VACUMM library (source code: http://www.ifremer.fr/vacumm/tutorials/misc.plot.basic.taylor.html).

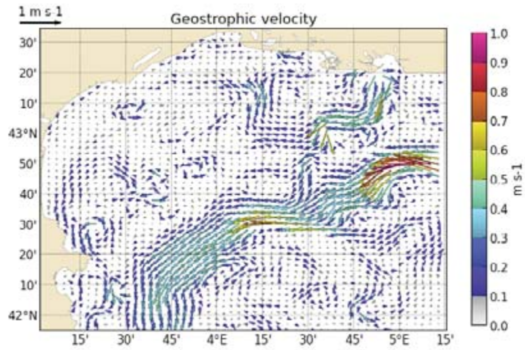

Figure 4: Example of geostrophic velocities computed from MARS3D modelled sea surface height using VACUMM library (source code: http://www.ifremer.fr/vacumm/tests/test_dataset_get_uvbt_menor.html).

Application for operational oceanography: modeled sea surface temperature validation and real time cruise support

Modeled sea surface temperature validation

One way to illustrate the VACUMM library is to show examples of applications. For example, in the frame of the PREVIMER[7] project, we developed an application to validate the modeled Sea Surface Temperature (SST) using remotely sensed observations. This tool is mainly dedicated and designed for operational oceanography but it has also been used and applied on longer interannual simulations in the frame of research projects.

The algorithm follows a classical scheme divided in 4 steps:
- Getting model and observation data (*e.g.* from remote data server),
- Collocation of model and observations,
- Statistic computation (*e.g.* spatial and temporal mean, standard deviation, correlation, root mean square, …),
- Plotting and generating validation report (*e.g.* figures, web page).

Based on SEVIRI Sea Surface Temperature remotely sensed data (METEOSAT SST provided by OSI-SAF belong to EUMETSAT), PREVIMER Bay of Biscay and Northwestern Mediterranean simulations are qualified on a daily basis. Results available on the PREVIMER website (http://www.previmer.org/produits/qualifications) display temporal average of the modeled and observed sea surface temperature, the time-averaged bias (Fig. 5), the time series of the spatial average temperature (observed and modeled) with the amount of cloud free observations (Fig. 6) to sustain our interpretation of qualification results. The algorithm is also able to provide Root Mean Square, Standard Deviation and Correlation statistics. Furthermore, the algorithm design allows adding new diagnostics quickly and applying analysis on a different dataset (*e.g.* the tool has already been used for surface chlorophyll concentrations).

[7] http://www.previmer.org

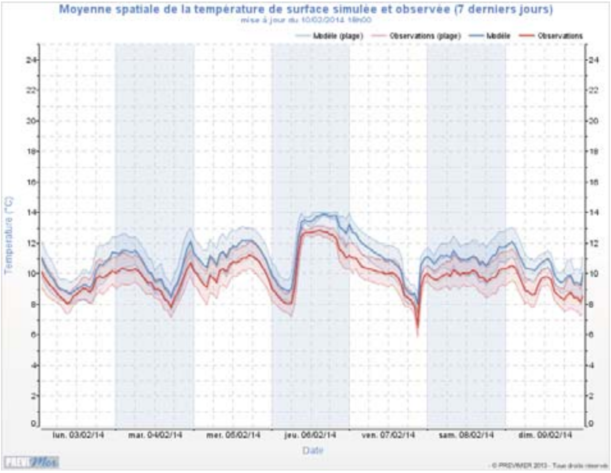

Figure 5: Spatial mean of observed and modeled Sea Surface Temperature from the 3rd to the 9th February 2014.

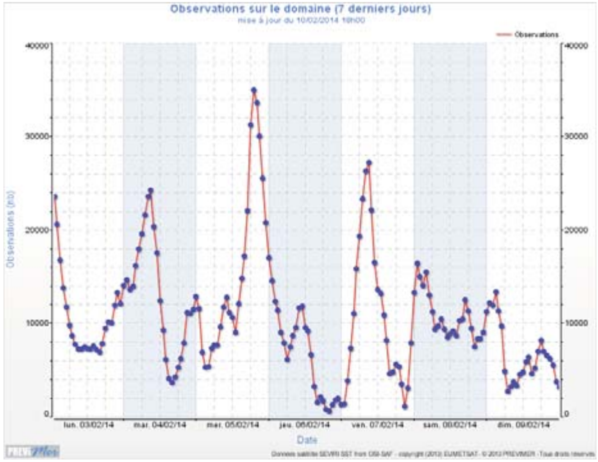

Figure 6: Number of Sea Surface Temperature cloud free pixels from the 3rd to the 9th February 2014.

In the example, we show the time-averaged bias for the week between the 3rd and the 9th February 2014 (Fig. 5). During this period, the model (covering the Bay of Biscay and Channel region - Berger *et al.*, 2014) overestimates the sea surface temperature compared with satellites observations. The bias is around 1°C. The standard deviation is in agreement with observations. Strong variations appear during the period and have to be cross-analyzed with the data availability (Fig. 6). Indeed, lowest temperatures are linked with a smaller number of observations and, then, the average is limited on specific regions.

Based on a similar approach, a tool has been developed to compare model simulations with *in situ* vertical profiles (*e.g.* from RECOPESCA program - not shown).

Real time cruise support

In 2013, the Operational Center of the SOP2 experiment (February-March 2013 - Hydrological cycle in Mediterranean Experiment program) asked PREVIMER to provide synthetic "quick-looks" describing the hydrological situation of the Northwestern Mediterranean sea and the development of deep convection and dense water formation. They focused on two areas - NWMED [0-12°E; 38.1-44.5°N] and GOL [2-8°E ; 41-43.8°N] - and were of different types: surface-300m-1000m depth 2D snapshots, hovmöller diagrams or time series graphs at key points and glider sections. The documented fields were atmospheric (radiative and turbulent fluxes, wind stress, evaporation, precipitation) and hydrodynamical (temperature, salinity, density, currents). The "quick-look" specification was extremely well precise in terms of colorbar definition, scaling, units, as well as criteria to extract diagnostics (*e.g.* the mixed layer depth was defined according to 3 different ways: turbulent, density or temperature criteria).

"Quick-look" scripts have been developed from the VACUMM library and allowed to complete the type of available graphs available, to improve the graphical parameterization and to validate the genericity of the model output reading and vertical interpolation. During the experiment, the PREVIMER system routinely created and sent to the SOP website more than 2000 snapshots a day (see Figure 7 as an example), extracted from MARS (PREVIMER configuration) or NEMO (MONGOOS configuration in which MENOR is embedded) outputs.

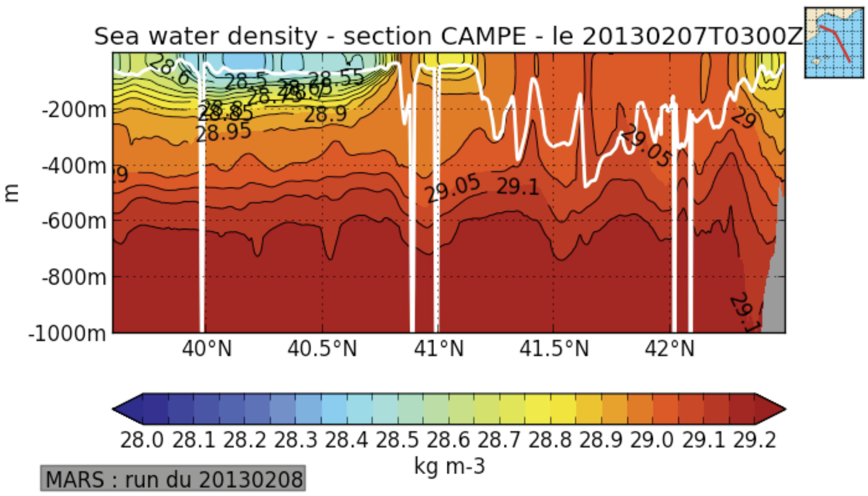

Figure 7: Example of density vertical section from MARS3D numerical simulations for the 7th February 2013 (http://sop.hymex.org/mars.php?current=20130207&nav=MarsSection&expected=DENS_CAMPE). The white line represents the mixed layer depth. Section position is displayed as a red line on the small map.

Application for research experiment: model validation from glider section

Taking advantage of the possibility to manage both ASCII files (Microsoft Excel or CSV) through Python modules and "xyzt" coordinates from Netcdf's outputs of ocean numerical model through UV-CDAT, VACUMM library is particularly designed to facilitate the collocation of *in situ* data and gridded results. It is now straightforward to compare, for example, glider data stored in Coriolis database (http://www.coriolis.eu.org/) as a succession of profiles with 3 hourly outputs of numerical models. During the IMEDIA cruise (March 2012) a glider recorded salinity and temperature along the French Provencal coast in the Mediterranean Sea monitoring a density current flowing along the continental slope from the Ligurian Sea to the Balearic Sea. This current called "North Current" exhibits warmer and fresher water along the coast. On the density field, the North Current is characterized by a lower density vein (until 400m depth) when the glider is close to the coast (Fig. 8). This current is also present in the model. The range in density is correct; nevertheless the vertical mixing seems to be overestimated in the numerical model leading to an important mixed layer depth offshore. Its clearly highlights that the observation vertical resolution is higher than the model one !

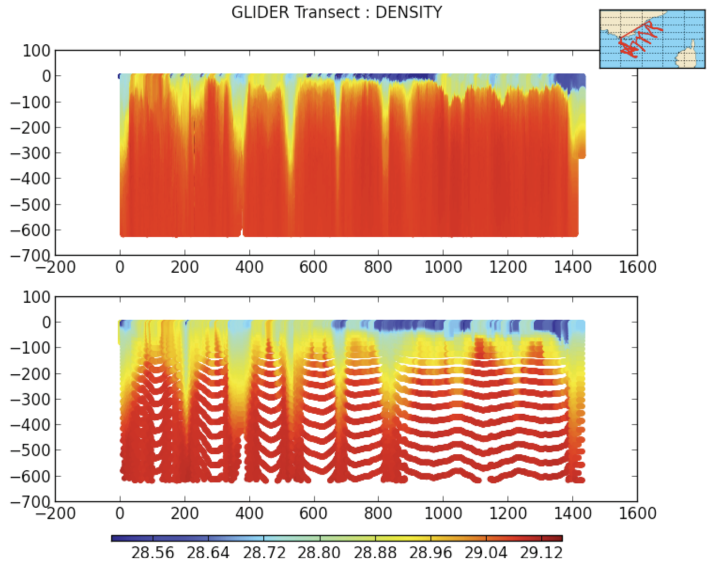

Figure 8: Density profile along a glider trajectory during IMEDIA cruise experiment (Mach 2012). The glider profile read in Coriolis data base (top) is compared to a co-localised numerical model results (MARS3D-MENOR configuration)(bottom).

Conclusion

The open-source VACUMM library appears as an efficient toolbox to process ocean data (numerical models and observations) in the frame of research project or to sustain operational systems. Development prospects could for example include an improved interfacing with usual oceanographic and atmospheric models and datasets, or the implementation of new physical or numerical diagnostics, algorithms or schemes. The library is open to new contributors to ensure that it meets at least the requirements of a wide oceanographic community.

Acknowledgements

The authors wish to thank Fabrice Lecornu and the PREVIMER project that allowed developments and improvements of the VACUMM library. A special thanks to all other junior or advanced developers: Mickaël Garo, Julie Gatti, Pierre Garreau, Serguei Skrypnikov-Laviolle, Sébastien Petton, Jean-François Le Roux. Remotely sensed Sea Surface Temperature data are provided by OSI-SAF (http://www.osi-saf.org/) belongs to EUMETSAT.

References

Berger, H., F. Dumas, S. Petton, and P. Lazure, Evaluation of the hydrology and dynamics of the operational MARS3D configuration of the Bay of Biscay : MANGAE4000, Coriolis - Mercator Ocean Quarterly Newsletter, this issue, 2014.