Filling data gaps through interpolation: innovative analysis tools for oceanography

Charles Troupin,ULiège/GHER (Belgium), c.troupin@uliege.be
Aida Alvera-Azcárate, ULiège/GHER (Belgium), a.alvera@uliege.be
Alexander Barth, ULiège/GHER (Belgium), a.barth@uliege.be
Matjaž Ličer, National Institute of Biology (Slovenia), matjaz.licer@nib.si
Dimitry Van der Zande, Royal Belgian Institute of Natural Sciences (Belgium), dvanderzande@naturalsciences.be
Jean-Marie Beckers, ULiège/GHER (Belgium), jm.beckers@uliege.be

Spatial interpolation of in situ and remote-sensing observations has been performed for decades in oceanography, with many applications ranging from simple data visualisation to initialisation of numerical models. The large amount of data to process, the presence of physical boundaries (coastlines, land) and processes (advection, diffusion), combined with the variety of spatial and temporal scales all contribute to making the interpolation task not trivial. Furthermore, while satellite observations usually provide a better coverage than in situ data at the sea surface, sensors working in the optical and infrared bands are affected by clouds, which obscure part of the ocean underneath. We present here three open-source software tools that have in common to provide a gridded field as an output and to take into account the numerous specificity of the oceanographic data.

Tool	**Meaning**	**Source code (language)**
DIVAnd	Data-Interpolating Variational Analysis - n-dimensional	https://github.com/gher-ulg/DIVAnd.jl (Julia)
DINEOF	Data-Interpolating Empirical Orthogonal Functions	https://github.com/aida-alvera/DINEOF (Fortran)
DINCAE	Data-Interpolating Convolutional Auto-Encoder	https://github.com/gher-ulg/DINCAE (Python, Julia)

DIVAnd is an analysis tool designed for the generation of gridded data products from in situ observations. DIVAnd extends the 2D capabilities of the DIVA tool and allows the interpolation of observations on curvilinear, orthogonal grids in an arbitrary high dimensional space by minimizing a cost function. This cost function penalizes the deviation from the observations, the deviation from a first guess and abruptly varying fields based on a given correlation length (potentially varying in space and time). Physical constraints can be added to this cost function, such as an advection constraint, diffusion or source terms. One major advantage of the method is that it naturally decouples basins that are not connected and where water masses often have very different properties.

DIVAnd was rewritten from scratch using generalized mathematical formulation, in the programming language Julia which allows a high-level programming style but is compiled to machine code for better performance. There are also functions implemented to directly query online the following databases: World Ocean Database, Copernicus Marine Environment Monitoring Service (CMEMS) and EMODnet Physics. The bathymetries, needed to delimit the domains of interpolation can be extracted from EMODnet Bathymetry or from GEBCO.

Application: new practical application in the context of SeaDataCloud for surface currents acquired high-frequency radar data, EMODnet Physics data and EMODnet Biology data with initial test of DIVAnd coupled with a neural network will be given.

DINEOF

DINEOF is an EOF-based method to fill in missing data from geophysical fields, typically because of the presence of clouds. The method has been applied on a wide variety of parameters, including sea surface temperature (SST), sea surface salinity (SSS), chlorophyll concentration or suspended particulate matter (SPM). DINEOF also allows the analysis of 2 or more variables jointly through a multivariate approach, for instance SST and wind. The consistency between successive fields is enhanced by filtering the temporal covariance matrix. Recent applications have been dealing with very-high resolution images from Sentinel-2 missions. This gave rise to new issues such as the effect of the cloud shadows, which were not affecting the measured variables at lower resolution. An advanced algorithm has been designed to detect and remove the cloud shadows prior to the analysis with DINEOF. Another aspect under development is the analysis of observations of a common variable but from sensors with different spatial and temporal resolutions, for instance Sentinel-2, Sentinel-3 and SEVIRI. As an application, a reconstruction of SPM in the North Sea will be presented.

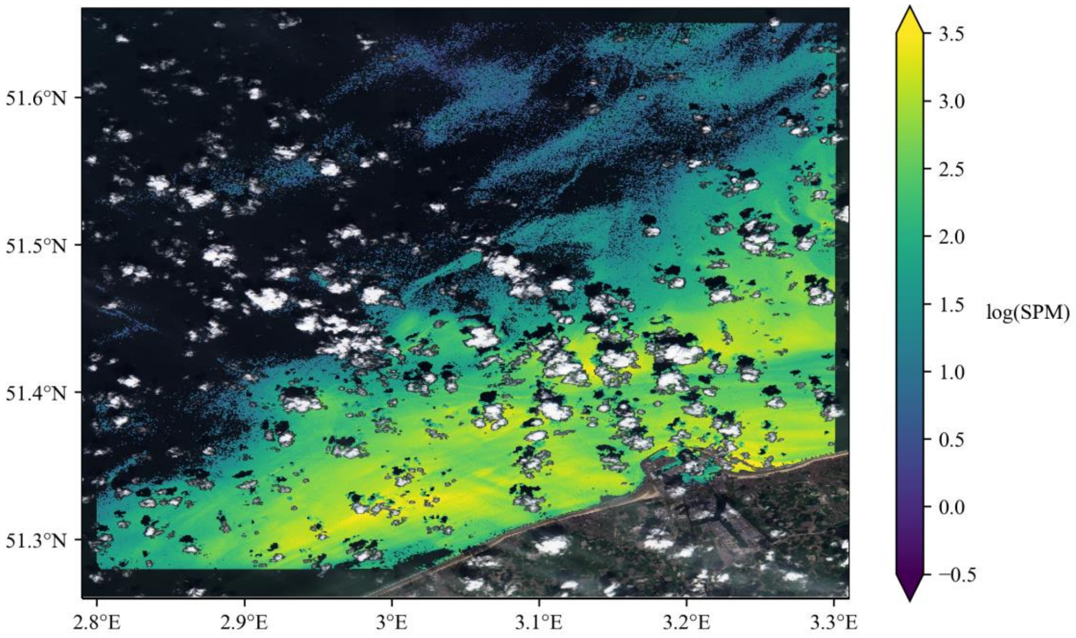

Figure 1: A typical Sentinel-2 SPM field, displaying small scale features and affected by clouds and their shadow.

DINCAE

DINCAE is a method to reconstruct missing data using a neural network. A neural network with the structure of a convolutional auto-encoder is developed to reconstruct the missing data based on the available cloud-free pixels in satellite images. Contrary to standard image reconstruction with neural networks, this application requires a method to handle missing data (or data with variable accuracy) in the training phase. DINCAE introduces a consistent approach which uses the satellite data and its expected error variance as input and provides the reconstructed field along with its expected error variance as output. The neural network is trained by maximizing the likelihood of the observed value. The approach is so far tested to 25-year time-series of Advanced Very High Resolution Radiometer (AVHRR) SST data and compared to DINEOF. The reconstruction error of both approaches is computed using cross-validation and in situ observations from the World Ocean Database. DINCAE results have lower error, while showing higher variability than the DINEOF reconstruction.