EGU22-9578
https://doi.org/10.5194/egusphere-egu22-9578
EGU General Assembly 2022

A multivariate convolutional autoencoder to reconstruct satellite data with an error estimate based on non-gridded observations: application to sea surface height

Alexander Barth, **Aida Alvera-Azcárate**, Charles Troupin, and Jean-Marie Beckers
GHER, University of Liège, Liège, Belgium

DINCAE (Data INterpolating Convolutional Auto-Encoder) is a neural network to reconstruct missing data (e.g. obscured by clouds or gaps between tracks) in satellite data. Contrary to standard image reconstruction (in-painting) with neural networks, this application requires a method to handle missing data (or data with variable accuracy) already in the training phase. Instead of using a cost function based on the mean square error, the neural network (U-Net type of network) is optimized by minimizing the negative log likelihood assuming a Gaussian distribution (characterized by a mean and a variance). As a consequence, the neural network also provides an expected error variance of the reconstructed field (per pixel and per time instance).

In this updated version DINCAE 2.0, the code was rewritten in Julia and a new type of skip connection has been implemented which showed superior performance with respect to the previous version. The method has also been extended to handle multivariate data (an example will be shown with sea-surface temperature, chlorophyll concentration and wind fields). The improvement of this network is demonstrated in the Adriatic Sea.

Convolutional networks work usually with gridded data as input. This is however a limitation for some data types used in oceanography and in Earth Sciences in general, where observations are often irregularly sampled. The first layer of the neural network and the cost function have been modified so that unstructured data can also be used as inputs to obtain gridded fields as output. To demonstrate this, the neural network is applied to along-track altimetry data in the Mediterranean Sea. Results from a 20-year reconstruction are presented and validated. Hyperparameters are determined using Bayesian optimization and minimizing the error relative to a development dataset.