A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations (DINCAE)

Alexander Barth¹, Aida Alvera Azcárate¹, Matjaz Licer², and Jean-Marie Beckers¹
¹University of Liege, AGO/GHER, Liege, Belgium (a.barth@uliege.be)
²National Institute of Biology, Marine Biology Station, Piran, Slovenia

A method to reconstruct missing data in satellite data using a neural network is presented. Satellite observations working in the optical and infrared bands are affected by clouds, which obscure part of the ocean underneath. In this paper, 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. However, it is unclear how to handle missing data (or data with variable accuracy) in a neural network when using incomplete satellite data in the training phase. The present work shows a consistent approach which uses essentially the satellite data and its expected error variance as input and provides the reconstructed field along with its expected error variance as output. The approach is motivated by the way models and observations are combined in the frame of data assimilation. The neural network is trained by maximizing the likelihood of the observed value. The corresponding error variances are estimated during training and do not need to be known a priori. The approach, called DINCAE (Data-Interpolating Convolutional Auto-Encoder) is applied to a relatively long time-series of Advanced Very High Resolution Radiometer (AVHRR) sea surface temperature data and compared to DINEOF (Data Interpolating Empirical Orthogonal Functions), a method to reconstruct missing data based on an EOF decomposition. 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. The resulting error estimates are also validated using the cross-validation data and they follow closely the expected Gaussian distribution.