A Neural Network-Based Observation Operator for Coupled Ocean-Acoustic Variational Data Assimilation

Andrea Storto[1,2] G. De Magistris[1] S. Falchetti[1] P. Oddo[1]

[1]Centre for Maritime Research and Experimentation, La Spezia, Italy [2]Institute of Marine Sciences, National Research Council, Rome, Italy

Contact: andrea.storto@cnr.it

Summary

Deep learning algorithms are now widely spread in a diverse range of fields to help solving automatic classification and regression problems. Here, we present and assess a strategy aimed at introducing an observation operator based on neural networks in variational data assimilation. Linearization of such operator, required by variational schemes, is also implemented. The methodology is applied to the coupled oceanic-acoustic data assimilation problem, and provides promising results. Our approach may be extended in the future to assimilate any remotely sensed type of observations. **More details at https://doi.org/10.1175/MWR-D-20-0320.1**

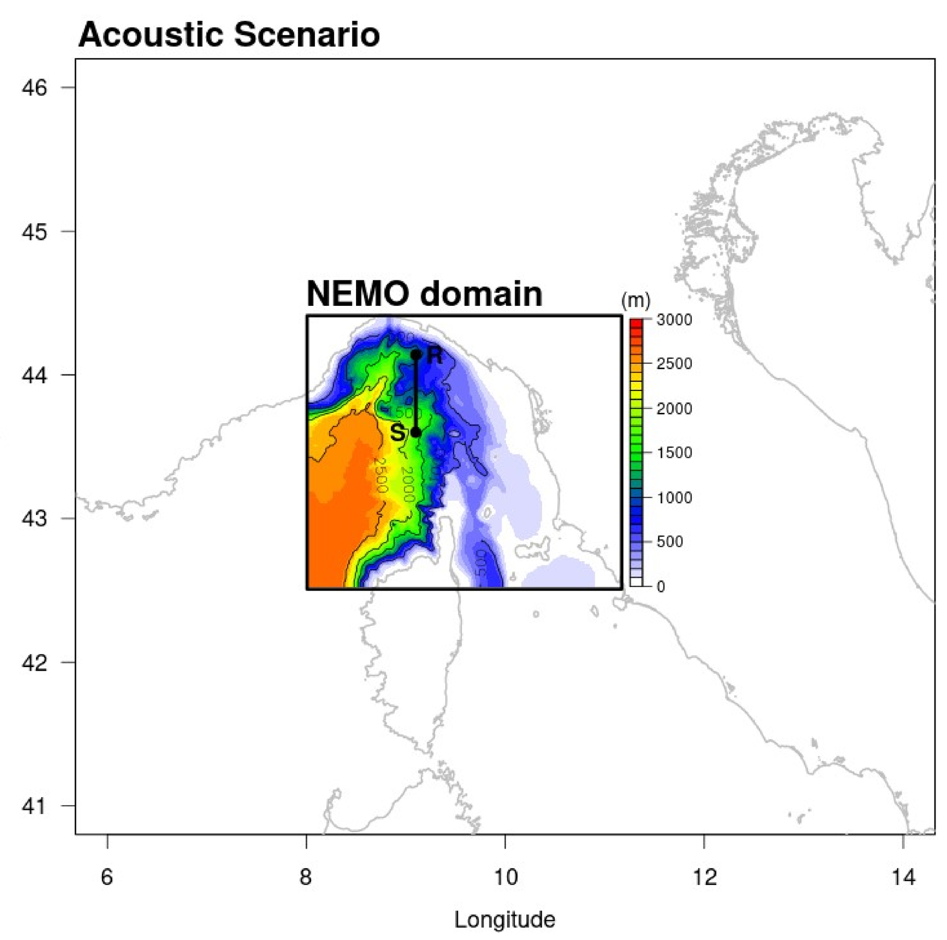
Motivation and observing scenario

An increasing number of observations in empirical and/or non-linear relationship with the data assimilation control variables may benefit from algorithms inherited from deep learning (for instance artificial neural networks).

As an application example, we investigate the feasibility of a data-driven approach for the assimilation of underwater acoustic measurements in oceanographic models.

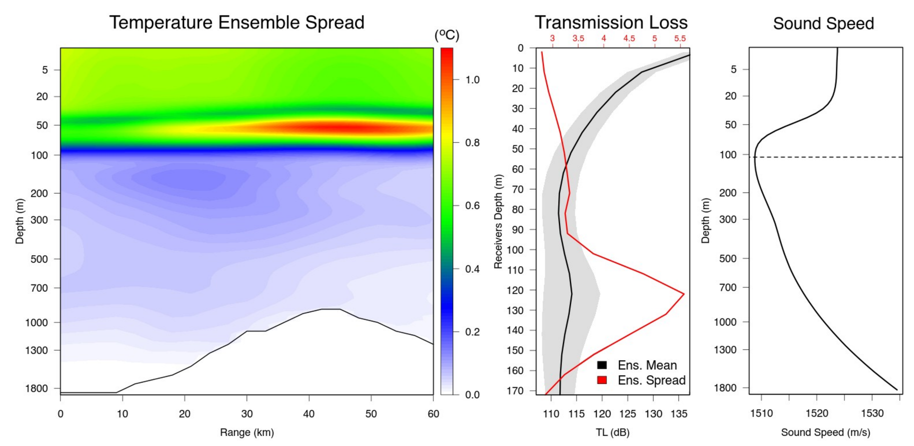
Acoustic propagation from the source (**S**) to the receiver (**R**) at a given frequency is subject to a certain transmission loss (**TL**), which depends also on the sound speed fields along the propagation transect, hence on the underlying seawater temperature.

Here, we study in particular the possibility to assimilate TL data relative to a 60 km propagation path over the Ligurain Sea (western Mediterranean Sea) at a frequency of 75 Hz (typical of ship noise). We consider an hydrophone tower with 18 receivers (upper 200m).



Modelling suite

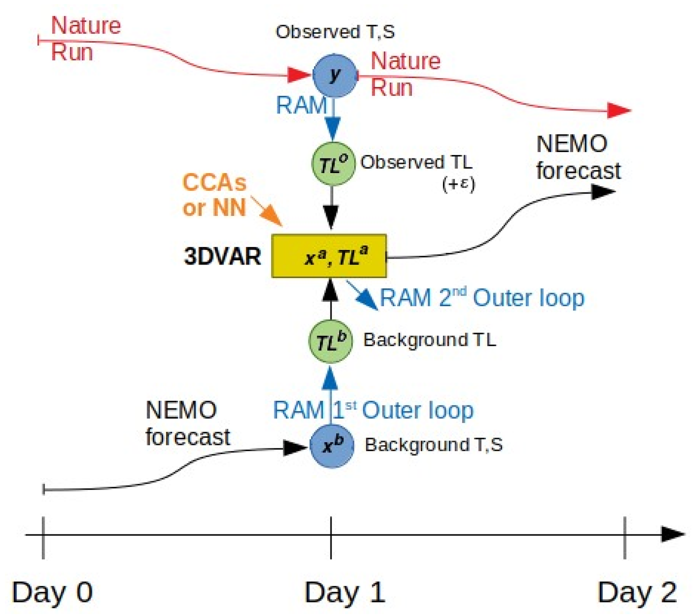
The **ocean model** is NEMO 3.6, implemented with a resolution of about 1.8 km and 91 depth levels with partial steps. It is forced at the surface by the ECMWF operational analyses and forecast (with the CORE bulk formulas) and at the lateral boundaries by the CMEMS Mediterranean Sea forecasting system.



The **acoustic model** is RAM (Range-dependent acoustic model), which is a parabolic equation model using the split-step Padè approximation.

Input sound speed fields for RAM are provided by the NEMO ocean outputs, while geoacoustic characterization is taken from the NOAA Deck41 database. Top right panels show the variability of temperature across the propagation path and that of transmission loss at the receiver, in an ensemble of simulations.

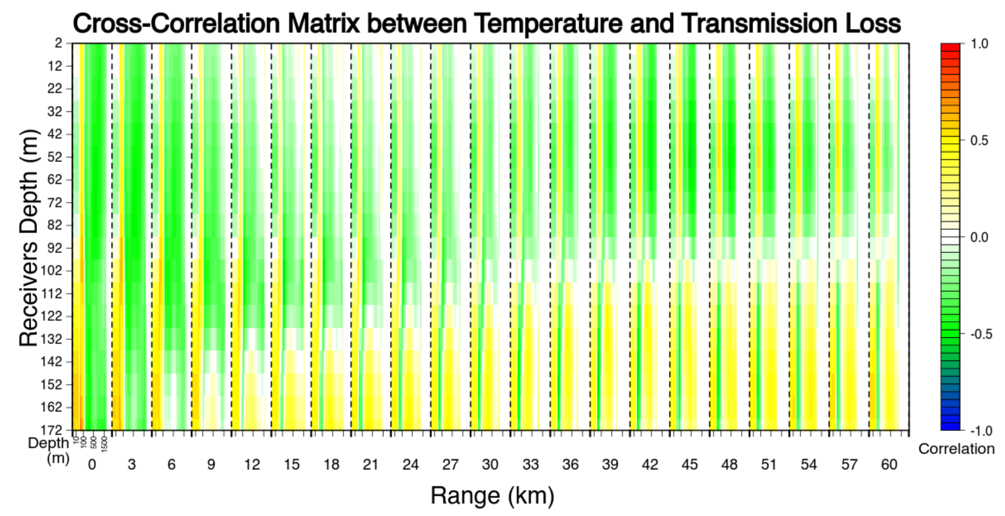
The data assimilation formulation is a three-dimensional variational scheme (**3DVAR**) with incremental formulation and control variable transformation. The control variables in physical space are temperature, salinity and sea surface height. The 3DVAR scheme uses multi-variate EOFs for the vertical covariances and a recursive filter to model horizontal correlations.



The assimilation of acoustic observations is tested through OSSE experiments, as shown in the above scheme. A (perturbed) nature run of NEMO-RAM provides synthetic observations, which are then assimilated in the 3DVAR scheme by means of the observation operator presented hereafter.

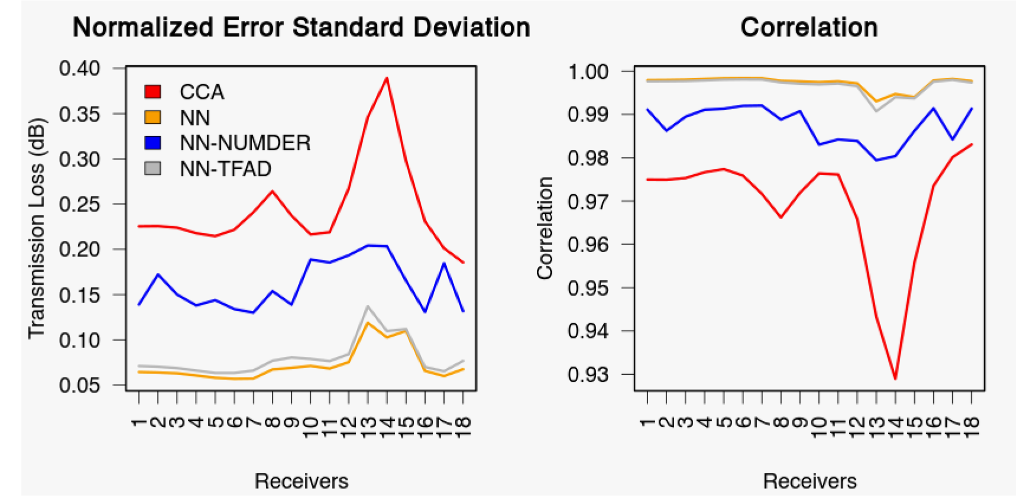
Observation operators

Assimilating acoustic data in oceanographic models requires the formulation of the acoustic observation operator. Here, we follow a data-driven approach, given the highly non-linear regime of the acoustic propagation models, which questions the validity of its tangent-linear approximation.



CCA observation operator: Canonical correlation analysis (CCA) is used to provide a linear data-driven observation operator. CCA finds the transformation matrices that project input and output datasets onto maximally correlated matrices.

NN observation operator: Artificial neural networks (NN) are used to formulate the second data-driven observation operator. As variational schemes require the tangent-linear version of any operator, we assess the impact of two linearization strategies: numerical derivation (**NUMDER**) through Richardson extrapolation or reverse-mode automatic differentation provided by the Tensorflow package (**TFAD**).



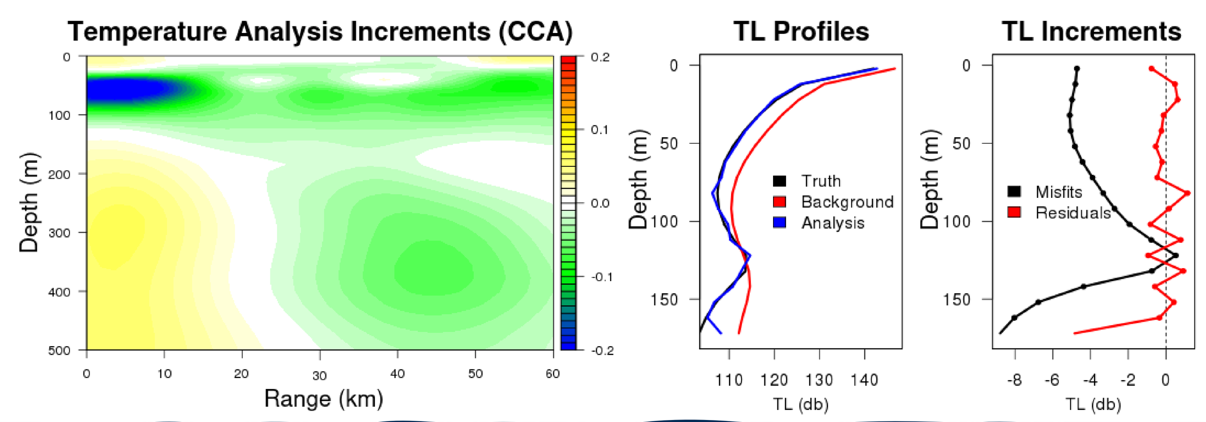
Training data (temperature and transmission loss data) are provided by an ensemble of NEMO runs (with stochastic physics), coupled to underwater acoustic propagation simulations performed by the RAM model.

Validation (see the Figure) of the two operators (performed over data independent from the model fitting) shows the significantly higher accuracy of the NN observation operator. While using numerical derivation leads to detrimental skill scores, the adoption of the automatic differentiation provides an accuracy as high as the non-linear NN.

Inversion of TL

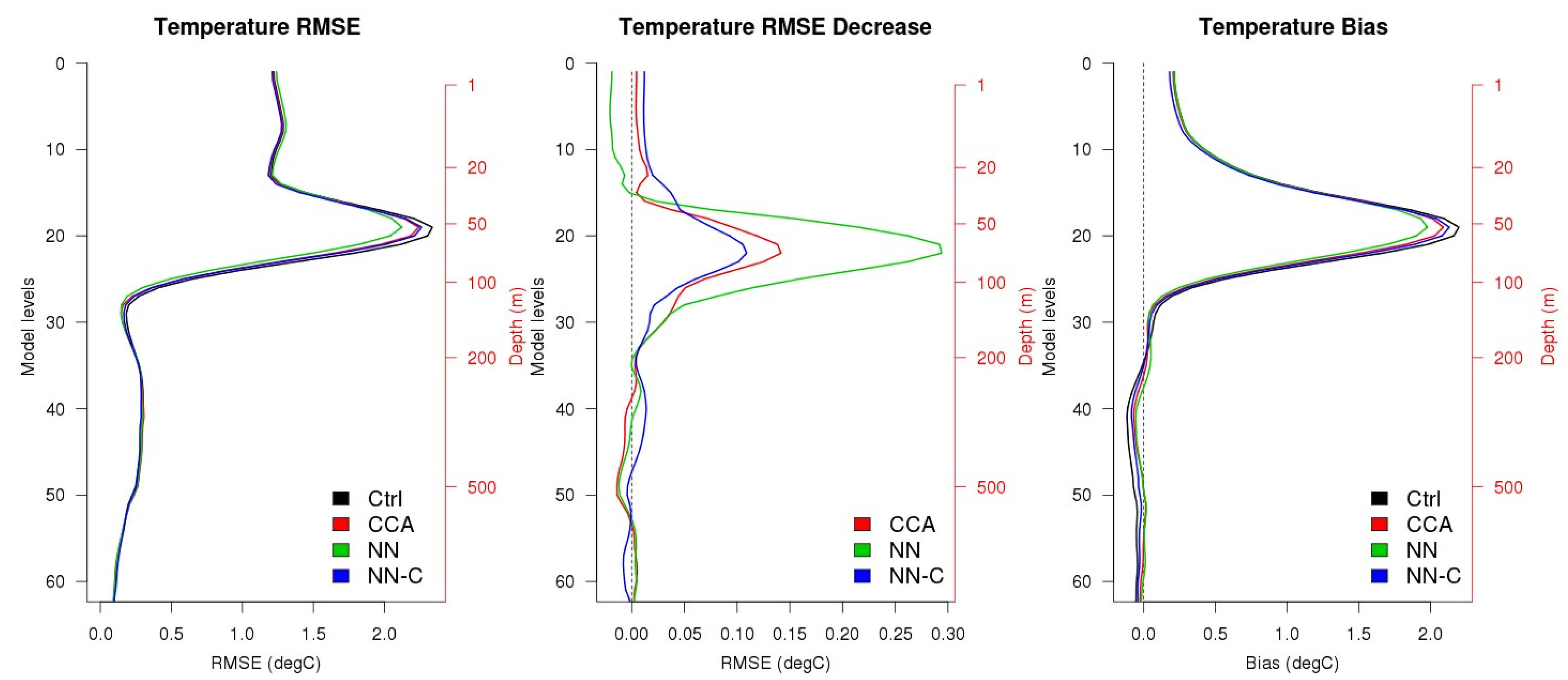
The observation operator provides of way to invert TL acoustic data onto temperature fields.

The adjoints of the **CCA** or **NN** operators and the background-error covariances can map a profile of transmission loss innovations onto the cross-section of temperature (underwater acoustic propagating path).



Results

OSSE experiments within the Ligurian forecasting system show that the **NN**-based observation operator outperforms both the **Ctrl** (no assimilation) and the **CCA**-based observation operator, reducing the temperature bias near the thermocline. Non-adaptive linearization of the observation operator (**NN-C**: linearization around climatology) provides detrimental RMSE results.



REFERENCE: Storto, A., De Magistris, G., Falchetti, S., & Oddo, P. (2021). A Neural Network-Based Observation Operator for Coupled Ocean-Acoustic Variational Data Assimilation, Monthly Weather Review, https://doi.org/10.1175/MWR-D-20-0320.1