Machine Learning methods for the Atmosphere, the Ocean, and the Seabed

Valentina Blasone[1], Umberto Di Laudo[1,2], Gloria Pietropolli[1,2], Luca Bortolussi[1], Silvia Ceramicola[2], Gianpiero Cossarini[2] and Luca Manzoni[1,2,*]

[1]*Dipartimento di Matematica e Geoscienze, Università degli Studi di Trieste, Via Alfonso Valerio 12/1, 34127 Trieste, Italy*
[2]*Istituto Nazionale di Oceanografia e Geofisica Sperimentale, Borgo Grotta Gigante 42/c, 34010 Sgonico, Italy*

Abstract
Understanding, observing, and forecasting the environment are all essential step to support a more sustainable interactions between human activities and the environment. Several areas of environmental modelling and classical analysis can be beneficial from the application of novel approach such as Machine Learning techniques. In particular, we are currently working on multiple areas for the development of machine learning techniques to be applied for (1) the modeling of the convection permitting dynamical model for precipitation forecasting, (2) data interpolation for ocean observing systems, in particular using data collected with ARGO floats, and (3) the automatic classification of seabeds for the assessment of geological hazards. Here we detail the current state of the projects in those area and directions for future research.

Keywords
Deep Learning, Digital Twins of the Ocean, Oceanography, Seabed Classification, Geohazards, Disaster Risk Forecasting

1. Introduction

Artificial Intelligence (AI) and machine learning (ML) methods are starting to change the way the environment is modeled, forecasting is performed, and, in general, the way scientific computing will support research in the future. In fact, as proposed in [1], there is a new merging of scientific computing, scientific simulation, and AI that results in what the authors call "simulation intelligence".

The works presented in this abstract are all part of this general movement towards the integration of ML techniques in the scientific field. In particular, here we present three different areas in which there are currently active project:

- Deep learning techniques for the forecasting of the precipitation distribution;
- Deep Learning for data interpolation in ocean observing systems;
- Automatic classification of the seabed.

As it is possible to observe, the three projects covers different aspects of the environment, adding ML as a support for existing simulation models or to manual labeling and analysis of data.

2. Deep Learning for Precipitation and Associated Risk

Every year across the world natural catastrophes due to extreme weather and climate events cause casualties and significant damage to properties and assets. Disaster risk forecasting highly depends on the ability to correctly quantify the phenomena related hazards, specifically at high spatial and temporal resolution. As for the precipitation phenomenon, the classical method of deriving precipitation distribution from simulations of dynamical models is computationally too expensive when it comes to high resolution, limiting its application. ML models can help in this regard, by levering the huge amount of data available from historical records and models simulations. For the precipitation phenomenon few studies have been carried out to date, among them [2] and [3], both based on deep learning techniques.

In this direction, we are developing a novel deep learning framework which represents a first attempt in emulating the convection permitting dynamical models. The main objective is to improve the projection of climatic impact-drivers relevant for risk assessment, in a much more efficient way. In its first version, the framework includes convolutional, recurrent and graph neural networks, to deal with the intrinsic characteristics of the data and the associated challenges. The input data is derived from the ERA5 reanalysis dataset [4], with hourly values

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
valentina.blasone@phd.units.it (V. Blasone); udilaudo@ogs.it (U. Di Laudo); gloria.pietropolli@phd.units.it (G. Pietropolli); lbortolussi@units.it (L. Bortolussi); sceramicola@ogs.it (S. Ceramicola); gcossarini@ogs.it (G. Cossarini); lmanzoni@units.it (L. Manzoni)
0000-0001-5814-8532 (V. Blasone); 0000-0001-7623-8419 (G. Pietropolli); 0000-0001-8874-4001 (L. Bortolussi); 0000-0002-1318-1272 (S. Ceramicola); 0000-0001-7803-8568 (G. Cossarini); 0000-0001-6312-7728 (L. Manzoni)

for temperature, specific humidity, eastward/northward wind components and geopotential atmospheric parameters, on a low-resolution grid of approximately 25 km. In a supervised perspective, the target is represented by the GRIPHO hourly precipitation observations dataset on a high-resolution grid of 3 km [5]. The framework was applied to the northern Italy and successfully trained using a time span of 15 years, from 2001 to 2015. Projections in terms of yearly and seasonal precipitation distribution maps for the year 2016 were derived and compared with observed values, showing a good capacity in resembling the precipitation distribution. Finally the model performance was tested in describing an extreme event.

Future research will focus on improving the framework, extending its applicability in spatial-temporal terms. The medium-term goal is to replace the input reanalysis data with data from simulations of dynamical models for the same atmospheric parameters. The long-term goal is to produce risk maps (e.g., floods) for the coming decades, integrating hydrological and vulnerability information.

3. Deep Learning for Ocean Observing Systems

Improving the capability of monitoring and predicting the status of the marine ecosystem has important implications, also considering the changes caused by human activities. In fact, marine ecosystem health is impacted by human activity: during the last decades, the Ocean has been increasingly affected by global changes (e.g. acidification) caused by the exponential augmentation of human assets. Among the many applications of AI in the oceanographic field, we deal with the automated real-time production of short-term forecasts of the state of the sea, with the aim of increasing the reliability of modeling predictions by correcting model results based on real-time observations of physical, chemical, geological and biological processes in the seas and oceans. Investigating marine ecosystem evolution and variability can be based on observations and modelling. In general, observations are accurate but limited and sparse both in time and space, and, most importantly, unevenly available among different variables. On the other hand, models reproduce ecosystem dynamics and cover different spatial and temporal scales of the processes but they can be inaccurate due to several source of uncertainties.

Integrating models and observations is widely used to provide optimal (e.g., in statistical sense) estimates of the state of the oceans. Classically, it is done through data assimilation approaches (e.g. Kalman filter, variational approaches). Here, we propose two novel approaches using AI techniques to advance the model/observation integration. The first application consists on the integration of Deep Learning model and data assimilation to improve the forecast skill of a marine model forecast system of the Mediterranean Sea [6, 7]. Specifically, the deep learning model is set to generate relationships between high-frequency sampled variables and low-frequency ones with the aims to generate reconstructed observations for the data assimilation. We consider as a dataset to train the deep learning model the collection of measurements collected by the so-called ARGO profiling float. Existing applications based on a feed-forward model (e.g., multilayer perceptron) are unaware of the typical shape of the profiles of biogeochemical variables that they try to infer. To overcome this issue we tested an innovative approach based on convolutional deep learning architecture to reconstruct nutrient profiles. The underpinning idea is that the typical shape of the vertical profiles of a variable is a constraint that has to be learned during the training. Preliminary experimental results confirm that the curves produced by the convolutional architecture guarantee the generation of smoother – and thus, more similar to those of the real world – profiles. A second application consists of the direct integration though a deep learning approach of observations and the deterministic model output to predict 3D fields of biogeochemical variables in the Mediterranean Sea by integrating observations and the output of an existing deterministic model, that is MedBFM. The deep learning architecture that we exploited for the aforementioned task is based on the *inpainting* [8], a computer vision technique developed to fill missing pixels of a considered image. Here, the resulting ML model can instead be used to fill the gaps between the observations in a way that "corrects" the output of the determinstic model. We successfully developed and trained this ML model on a portion of the Mediterranean Sea, obtaining promising results, and now we are working on the extension of this to the whole Mediterranean area.

4. Automated classification of the seabed

Geological hazards or geohazards are the result of natural, active geological processes. They include volcanic eruptions, earthquakes, tsunamis, landslides and several types of mass wasting phenomena. Geohazards can endanger and cause damage mainly in coastal areas where people live and important economic infrastructures (harbors, highways, airports, and so on) are located.

The assessment of geohazards is the basis for carrying out susceptibility and risk assessment and to apply a sustainable management of the seafloor and coastal areas as indicated in several European and international directives. Assessing geohazards in marine environments is a very time consuming and costly exercise as it implies using research vessels, geological and geophysical expertise and expensive tools. In addition geological interpretation and seabed mapping can be very subjective to experience and time available.

Having the help of human-supervised AI could be strategic when dealing with huge databases and to reduce the inevitable human subjectivity when interpreting seabed data, thus providing a uniform interpretation. The goal of this project is to construct a model that performs an automated classification for the seabed and that in particular, is able to find and recognize in automated way (or better automated human supervised) features indicative or prone of to be geohazards.

First, we want to construct a model that automatically map the seabed, using a bathymetric map as input. For this purpose, we are using a model [9] that consist in two Region-based Convolutional Neural Network (R-CNN) [10]. GIS data (that specify the depth, the slope and the curvature of the seabed) are used as input and treat them as an RGB image. The image is cut in small windows and for each window they apply the Selective Search algorithm in order to localize features and then train the networks [11]. The first CNN is basically a binary classification and tell us if there is or not a feature indicative of hazard in some region; the second one classifies features into different classes. Transfer learning from VGG19 is used to train the networks. The overall objective is to improve the existing model described above in order to have a best efficiency and less time-consuming predictions.

Once we have a model that, given unlabelled data, returns us the overall structure of the seabed, it comes the complex step of assessing geohazards. Indeed, the fact that a seabed feature (i.e., a submarine canyon) is prone or not to hazard, depends from several factors: its depth, its geological activity, its proximity to the coast etc. It is easy to understand that it is not trivial to develop such a model but would be a significant support to practitioner to enable the reduction of vulnerability and to enhance community resilience to disasters. For these reasons it is important for the resulting ML-driven method to be able to estimate the likelihood that seabed features may represent potential hazards. Given that this step is currently not automate, such a model would provide a signficant step forward in geohazards assessment for huge regions such as for the Mediterranean sea.

The main goal of the project is to construct an AI tool that is able to label a seabed map with all the possible features indicative of geohazards. An further step would be to provide an explainable model, providing a justification for the choice the model makes about the possibility of having or not an hazard. In other words developing an AI model with human supervision able to provide information and, possibily, explanation, in the assessment of geohazards in marine environment would be extremely useful to the coastal community dealing with hazards and disasters.

5. Final Remarks

As it is possible to observe, ML model are being developed for a large number of applications for climate, sea, and seabed. In all cases it is necessary to have a close collaboration with the domain experts in order to understand the requirements for a model and to evaluate it. Some advantages of machine learning in those areas are the ability to automate expensive human activities, e.g., manual labeling of geohazards, to integrate information from multiple sources even if the relation between them has not already be formalized (i.e., there is no "classical" model explaining the relation between two variables), and to capture spatial and temporal relations only from the data. One obstacle is, however, that a ground truth is not always present, but usually only sparse data and deterministic models (which are themselves approximations of the reality). Hence, ML models should be able to deal with uncertainty in the input data – and possibly provide a corresponding estimation of uncertainty in the output.

References

[1] A. Lavin, D. Krakauer, H. Zenil, J. Gottschlich, T. Mattson, J. Brehmer, A. Anandkumar, S. Choudry, K. Rocki, A. G. Baydin, C. Prunkl, B. Paige, O. Isayev, E. Peterson, P. L. McMahon, J. Macke, K. Cranmer, J. Zhang, H. Wainwright, A. Hanuka, M. Veloso, S. Assefa, S. Zheng, A. Pfeffer, Simulation intelligence: Towards a new generation of scientific methods, 2022. arXiv:2112.03235.

[2] J. Wang, Z. Liu, I. Foster, W. Chang, R. Kettimuthu, V. R. Kotamarthi, Fast and accurate learned multiresolution dynamical downscaling for precipitation, Geoscientific Model Development 14 (2021) 6355–6372. doi:10.5194/gmd-14-6355-2021.

[3] C. Wang, P. Wang, P. Wang, B. Xue, D. Wang, A spatiotemporal attention model for severe precipitation estimation, IEEE Geoscience and Remote Sensing Letters 19 (2021) 1–5. doi:10.1109/LGRS.2021.3084293.

[4] H. Hersbach, W. Bell, P. Berrisford, A. Horányi, M.-S. J., J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, D. Dee, Global reanalysis: goodbye era-interim, hello era5, 2019. URL: https://www.ecmwf.int/node/19027. doi:10.21957/vf291hehd7.

[5] A. Fantini, et al., Climate change impact on flood hazard over italy, 2019.

[6] S. Salon, G. Cossarini, G. Bolzon, L. Feudale, P. Lazzari, A. Teruzzi, C. Solidoro, A. Crise, Marine

ecosystem forecasts: skill performance of the cmems mediterranean sea model system (????).

[7] G. Cossarini, L. Feudale, A. Teruzzi, G. Bolzon, G. Coidessa, C. Solidoro, V. Di Biagio, C. Amadio, P. Lazzari, A. Brosich, et al., High-resolution reanalysis of the mediterranean sea biogeochemistry (1999–2019), Frontiers in Marine Science 8 (2021) 1537.

[8] G. Pietropolli, G. Cossarini, L. Manzoni, Gans for integration of deterministic model and observations in marine ecosystem, in: Progress in Artificial Intelligence: 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31–September 2, 2022, Proceedings, Springer, 2022, pp. 452–463.

[9] R. Soemantoro, M. Gaimann, Machine Learning How to Sea: Recognition and classification of subsea structures using artificial intelligence and high-performance computing, PRACE: Partnership for Advanced Computing in Europe, 2022, pp. 35–37.

[10] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2014, p. 580–587.

[11] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M. Smeulders, Selective search for object recognition, in: International Journal of Computer Vision, Springer, 2013, pp. 154–171.