GANs for integration of deterministic model and observations in marine ecosystem

Gloria Pietropolli[1, 2], Gianpiero Cossarini[2], Luca Manzoni[1]

University of Trieste

National Institute of Oceanography and Applied Geophysics - OGS

31/08/2022

Outline

Outline

Introduction

There exists two ways to monitoring the marine ecosystem:

Introduction

There exists two ways to monitoring the marine ecosystem:

- **observations**
    - accurate
    - insufficient in terms of temporal and spatial coverage

Argo Float.

Satellite.

Introduction

There exists two ways to monitoring the marine ecosystem:

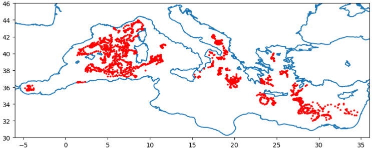
- **observations**
  - accurate
  - insufficient in terms of temporal and spatial coverage

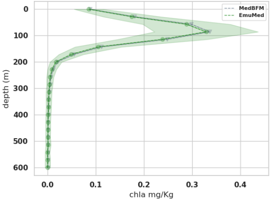
Argo Float.

Satellite.

- **deterministic models**
  - can be inaccurate
  - cover the whole marine ecosystem

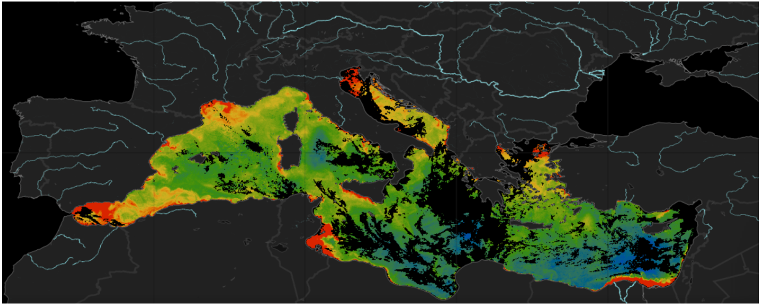
Observations: Argo Float





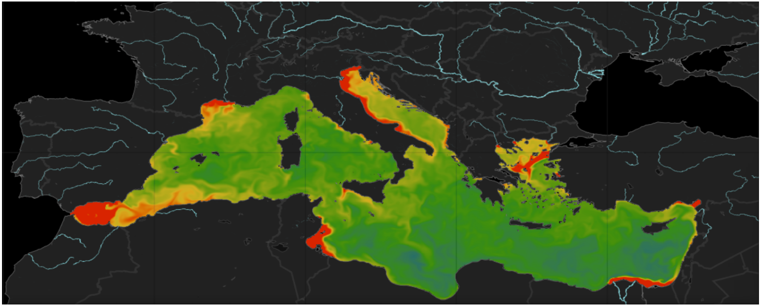
Chlorophyll profiles captured through **Argo Float** instruments.

Observations: Satellite sensor



Chlorophyll surface map captured through **satellite sensors**.

Deterministic physical-biogeochemical Model

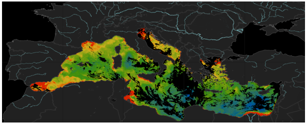


Reconstruction of the distribution of the chlorophyll surface map through the **deterministic model *MedBFM***.

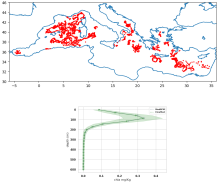
Our approach

Developing a deep learning model to reproduce the biogeochemical variables in the Mediterranean Sea integrating:

- **observations** of the marine ecosystem:
  - high-resolution satellite data from Copernicus
  - in-situ BGC-Argo floats
- output of an existing **deterministic model**



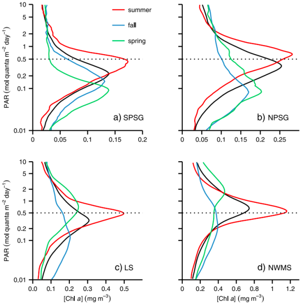
Satellite superficial data



BCG Argo float data

Our Goals - 1

- Infer **vertical profile** related to different time and location

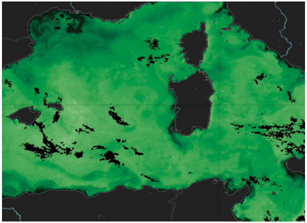


Chlorophyll vertical distribution as a function of light for different regions and over different time periods.[1]

[1] Alexandre Mignot et al. "Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation". In: *Global Biogeochemical Cycles* 28.8 (2014), pp. 856–876.

Our Goals - 2

- Infer **spatial variability**



Surface map of chlorophyll captured by the satellite.

Our Goals - 3

- Build a deep learning model that:
  - **infer the general structure** of the marine variable from the **deterministic model**
  - **embed the information** provided by **observations** to improve the prediction of observed variable
  - **spread of information** provided from the observed variables (temperature, salinity, oxygen, chlorophyll) also to variables that are not possible to directly collect (or hardly observable) via in-situ or satellite measurements (e.g., primary production)

Why a deep learning model?

To answer this question it is necessary to introduce the deep learning model exploited.

Outline

Inpainting Architecture

Inpainting

In computer vision a general **inpainting architecture** consists of the training of a **generative network** to "fill-in" in the most realistic way possible an image with one (or even more) parts of it masked.

Example of the application of an inpainting network.[2]

[2] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and locally consistent image completion". In: *ACM Transactions on Graphics (ToG)* 36.4 (2017), pp. 1–14.

Inpainting Architecture

The inpainting architecture exploited is based on Generative Adversarial Network (**GAN**).

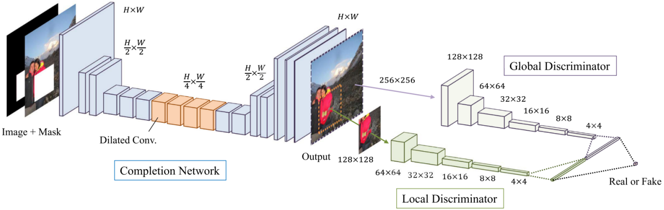


Illustration of the GAN-based inpainting architecture used.[3]

[3] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and locally consistent image completion". In: *ACM Transactions on Graphics (ToG)* 36.4 (2017), pp. 1–14.

Inpainting Architecture

The architecture is composed of 3 interacting Convolutional Neural Network (CNN):

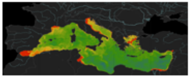
- the **completion network** used to complete the image
- the **global discriminator network**
- the **local discriminator network**

The completion and the discriminators compete in a two-player game.

- the **completion** learns how to fill the holes in a realistic and coherent way
- **discriminators** are trained to understand whether or not the provided input has been completed.

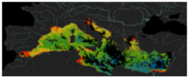
Inpainting adapt to the marine framework

Image

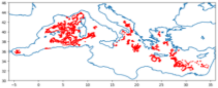


Horizontal map

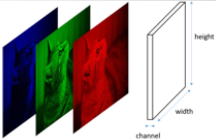
Image with holes

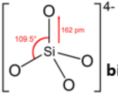


Observations



RGB channels





Marine physical and biogeochimical variable

Inpainting adapt to the marine framework

We will introduce two deep learning model to merge these information:

Inpainting adapt to the marine framework

We will introduce two deep learning model to merge these information:

- **EmuMed** which **learns spatial and temporal relationship** among the marine ecosystem variables starting from the deterministic model MedBFM.

Inpainting adapt to the marine framework

We will introduce two deep learning model to merge these information:

- **EmuMed** which **learns spatial and temporal relationship** among the marine ecosystem variables starting from the deterministic model MedBFM.
- **InpMed** **adds observations** to EmuMed while maintaining the same architecture.

Why a deep learning model?

- **Convolutional-based architecture**
  $\rightarrow$ naturally suitable for dealing with **spatial data**.

Why a deep learning model?

- **Convolutional-based architecture**
  $\rightarrow$ naturally suitable for dealing with **spatial data**.

- **Inpainting architecture**
  $\rightarrow$ Dealing with in-situ measurements comport **insufficient spatial coverage** of information.

Why a deep learning model?

- **Convolutional-based architecture**
  → naturally suitable for dealing with **spatial data**.

- **Inpainting architecture**
  → Dealing with in-situ measurements comport **insufficient spatial coverage** of information.

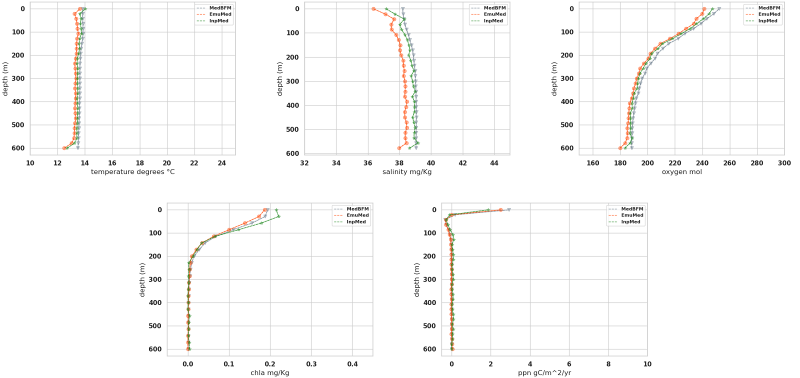
- **RGB channels** are strongly interrelated and dependent on each other
  → **introduce intrinsic relation** between marine ecosystem variables
  → **pass this information** from observed variable to the non-observed one

Experimental Settings

- **area**: Western Mediterranean Sea ($36°$ N - $44°$ W)
- **depth**: $0 - 600$ m
- **year**: 2015
- **channel**:
  - temperature (model, float)
  - salinity (model, float)
  - oxygen (model, float)
  - chlorophyll (model, float, satellite)
  - primary production (model)

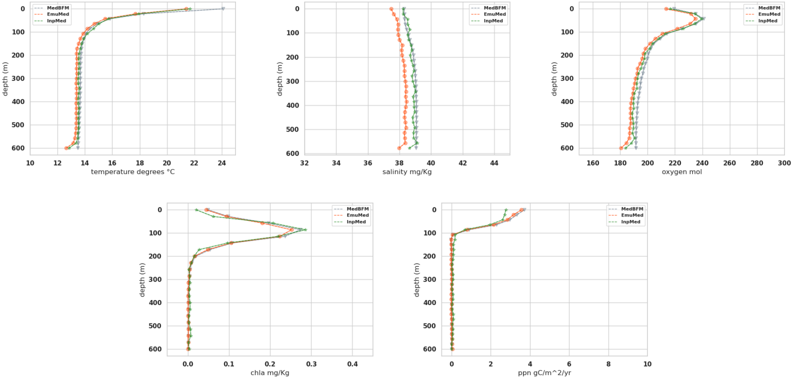
Outline

Experimental Result



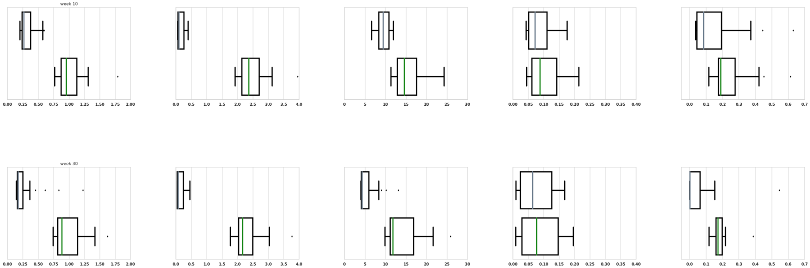
Vertical profile of, respectively, temperature, salinity, oxygen, chlorophyll, primary production during the **winter** season.

Experimental Result



Vertical profile of, respectively, temperature, salinity, oxygen, chlorophyll, primary production during the **summer** season.

Experimental Result



Box-plots of the distributions of the standard deviation of the marine variables computed from the spatial maps at different depths. From left to right are shown: temperature, salinity, oxygen, chlorophyll, primary production. From top to bottom are shown: week 10 and week 30.

Outline

Conclusion

- The models can **infer correctly information** from the deterministic model and observations.
- A deep learning model is **less computationally expensive**, once trained, with respect to a deterministic one.
- Possibility of **spread information** provided from the observed variables also to variables that are not observed.

Next steps

- Validate the model through the recognition of **specific events** (e.g. blooms).
- Extension of the ML model to a **larger number of channels**, so that it became able to reproduce all marine ecosystem variables (in particular a wider range of unobserved variables).
- Extension of the ML model to the **entire Mediterranean Sea**.

Thank you for your attention!