EXPLORING GRAPH NEURAL NETWORKS TO PREDICT THE SEAGRASSES ECOSYSTEM STATE IN THE ITALIAN SEAS

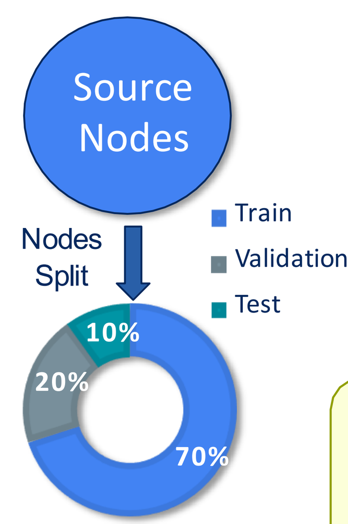
Angelica Bianconi[1,2,3], Sebastiano Vascon[1,4], Elisa Furlan[1,2] and Andrea Critto[1,2]
[1] Ca' Foscary University of Venice, Department of Environmental Sciences, Informatics and Statistics, Via Torino 155, 30172, Venice, Italy
[2] Centro Euro-Mediterraneo sui Cambiamenti Climatici, Risk Assessment and Adaptation Strategies Division, via Marco Biagi 5, 73100 Lecce, Italy
[3] University School for Advanced Studies IUSS Pavia, Pavia, Italy
[4] Ca' Foscari University of Venice, European Center for Living Technology, Dorsoduro 3246, 30123 Venice, Italy
angelica.bianconi@iusspavia.it

1. Motivations & objective

- **Marine and Coastal Ecosystems (MCEs)** and their ability to flow ecosystem services are **threatened** by the complex interplay between human-made and climate change-related pressures
- Although existing **ML** methods can **simplify the impact evaluation** of multiple environmental stressors on **MCEs**, they **do not consider the potential spatial dependence of the effect of pressures**
- Recent advancements in **Graph Neural Networks (GNNs)** offer potential solutions and are gaining popularity in environmental sciences

- **Explore** the application of GNNs to **assess** the impact of anthropogenic and climate change-derived pressures on **Seagrass ecosystem** in Italian seas
- **Compare** GNNs with commonly used models in this context (i.e., Random Forest, Support Vector Machine, Multi-Layer Perceptron)

4. Experiments



Architectures were tested with **different layer configurations**, including a **single convolutional layer** (or **attention layer**) followed by a linear layer and **three convolutional layers** (or **attention layers**) followed by a linear layer, with generally better performance observed with the latter.

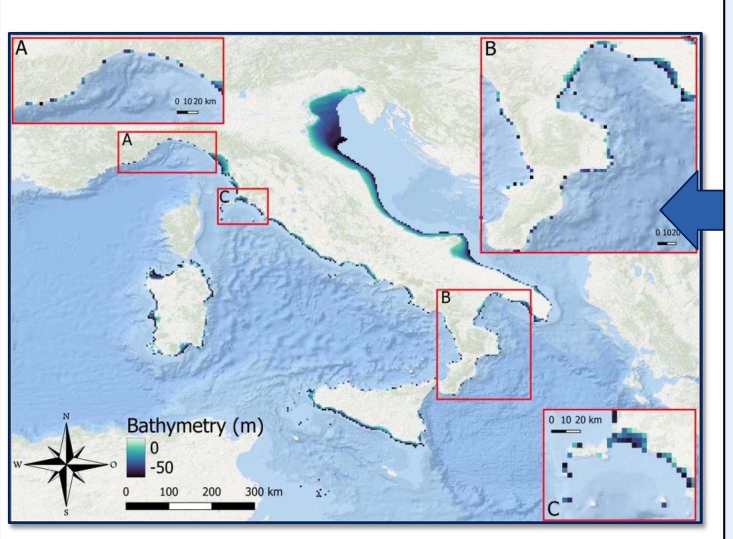
Architecture configurations were explored on graphs with both **4-node** and **8-node** connectivity.

Models applied to **8-node connected** graphs **perform better** and are **used for comparison**

Training:

1. On the **entire graph**
2. On **subgraphs** obtained through a **weighted random sampling**[1] based on class distribution.

2. Case study & Dataset

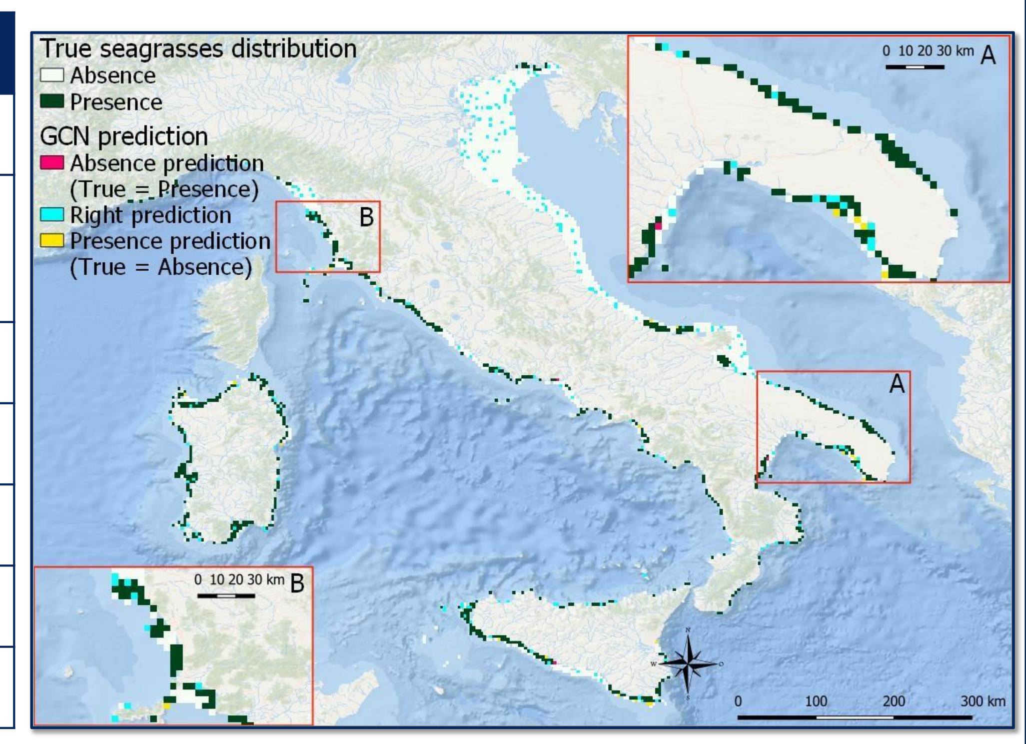


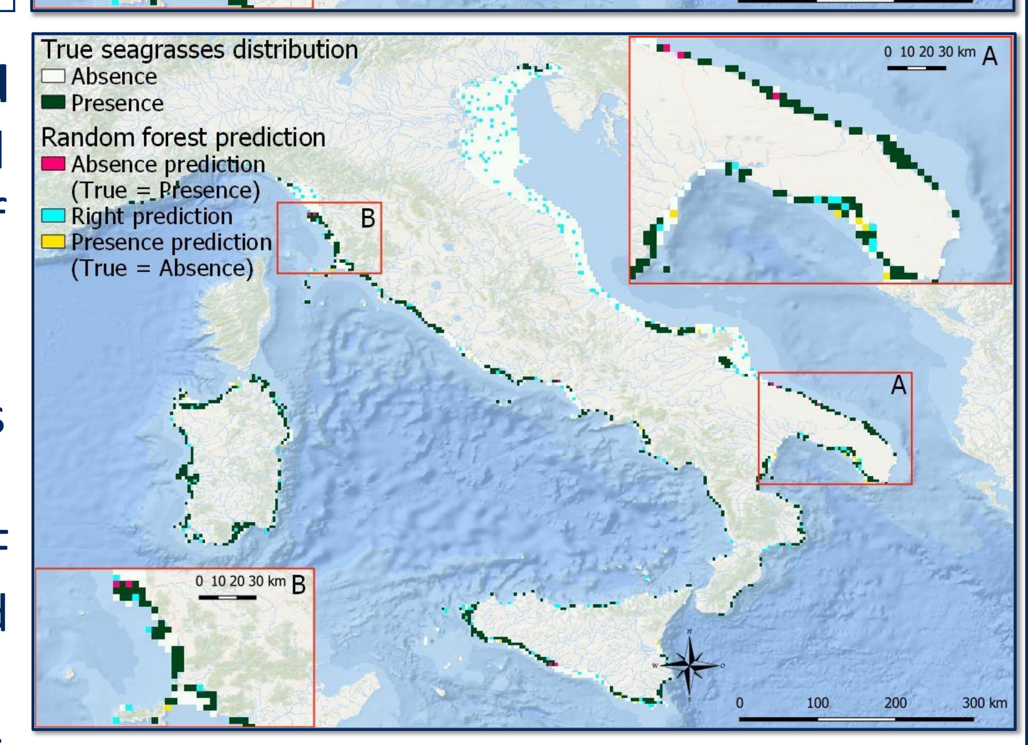
- The dataset was compiled **involving experts** and utilizing data from **open-source platforms** (e.g., Copernicus Services, UNEP, EMODnet)
- Collected data were pre-processed, and environmental indicators were aggregated into **4km-spatial resolution** raster grids
- Annual metrics were calculated for input variables representing pressures, considering **2017 as reference year**
- Seagrass spatial distribution was categorized as **presence or absence** pixels.

	Indicator	Annual metrics
Endogenic Pressures	Distance from port	Derived from the source
	Distance from river	Calculated by Haversine distance
	Distance from cities	Calculated by Haversine distance
	Nutrients (NH4, mNO3, PO4)	Minimum, 5°percentile, 95°percentile
	Chlorophill-a	Minimum, 90°percentile
	Secchi depth	Minimum, 5°percentile
	Light attenuation	Minimum, 5°percentile
	Shipping traffic (density)	Mean
Exogenic pressures	Sea Surface Temperature	95°percentile, standard deviation, marine heat waves
	pH	Minimum, mean, maximum
	Salinity	5°percentile, minimum, standard deviation
	Max significant wave height	Maximum
	Eastward and Northward Sea Water Velocity	Mean, maximum
	Kinetic energy at the seabed due to currents	90°percentile
	Sea surface height	mean

5. Results

Comparison results			
Model	Accuracy	F1-score	
		Absence	Presence
RF	0.90	**0.92**	0.87
SVM	0.89	0.91	0.86
MLP	0.89	0.91	0.86
GCN	**0.91**	**0.92**	**0.89**
GAT	**0.91**	**0.92**	**0.89**

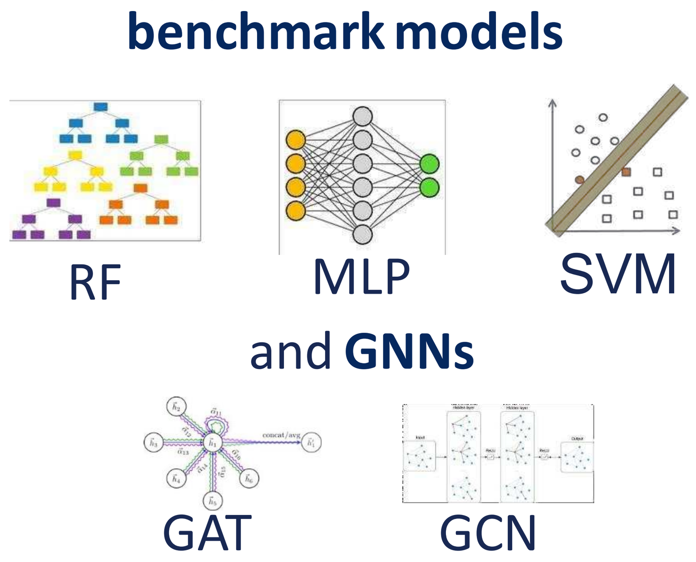




- The **RF** and **GNN-based models** show good performance in terms of accuracy (**0.91** vs **0.90**) and the latter have the highest **F1 score** for the seagrass presence class (**0.89**).
- Error maps show that the RF model incorrectly predicted seagrass absence among connected presence pixels (Focus A and B), demonstrating the limitation of treating pixels in isolation for accurate predictions.

6. Method

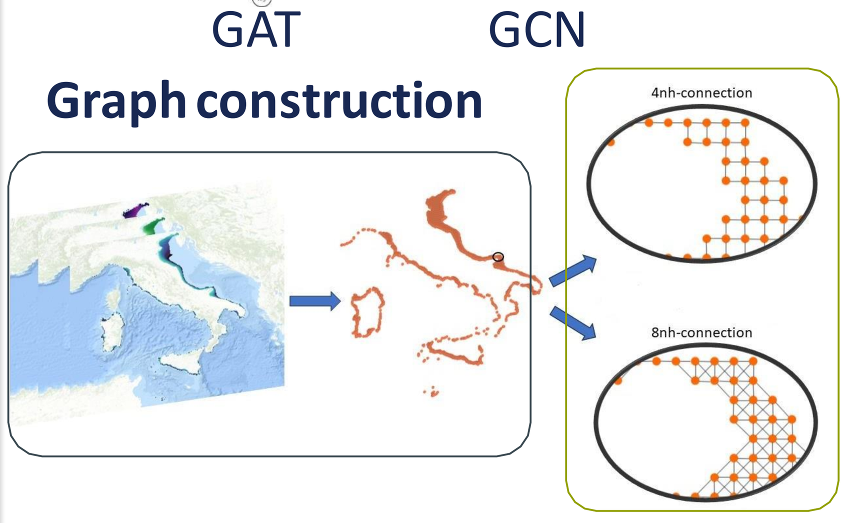
Comparison between **benchmark models**



Each pixel from the dataset, identified by latitude and longitude coordinates, was **mapped to a node** in the graph

Layers containing **metrics** related to selected **input variables** were vectorized into **feature vectors** associated with each node

Graph construction



The **edges** of the graph were determined based on the **distance** between nodes, particularly each node was connected to its **four** or **eight nearest** neighbouring **nodes** (pixels)

6. Conclusions

- **Unveiling Spatial Dynamics**: GNN-based models outperformed traditional models by incorporating spatial context, especially in predicting Seagrass presence
- **Impact on Graph Construction**: Limited availability of high-resolution data influenced graph construction. Integration was mainly limited to geographic information, neglecting temporal and ecosystem interaction dynamics.
- **Future Directions**: Integrate all data components, including temporal and "physical" aspects, can lead to more accurate predictions and support implementation of environmental directives for marine conservation.

References

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. arXiv preprint arXiv:1710.10903, 2017.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.

[1] torch_geometric.loader.imbalanced_sampler pytorch_geometric documentation (pytorch-geometric.readthedocs.io)