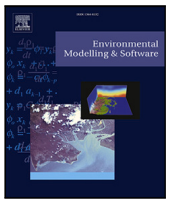
Environmental Modelling and Software 194 (2025) 106678

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Leveraging artificial intelligence methods to map seagrass ecosystems in Italian Seas: Tackling human impact and climate change☆

Angelica Bianconi [a,b,c], Sebastiano Vascon [b,d,*], Elisa Furlan [c], Andrea Critto [b,c]

[a] University School for Advanced Studies Pavia, Piazza della Vittoria 15, Pavia, 27100, Italy
[b] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, I-30170, Venice, Italy
[c] CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy
[d] European Center for Living Technology, Ca' Foscari University of Venice, Venice, Italy

ARTICLE INFO

Dataset link: https://github.com/Angebii/GNN4Seagrass

Keywords:
Italian coastal waters
Random Forest
Multi-Layer Perceptron
Support Vector Machine
Graph neural networks
Seagrass ecosystem

ABSTRACT

Marine coastal ecosystems (MCEs) are crucial for human health, playing a key role in climate change adaptation. However, MCEs are globally threatened by environmental and human pressures. This study applies Graph Neural Networks (GNNs) to model seagrass distribution in the Italian Seas using a dataset of 2244 spatial units with environmental, climatic, and anthropogenic factors harmonised at 4 km resolution. GNN models, including Graph Convolutional and Attention Networks, were benchmarked against traditional machine learning methods: Random Forest, Support Vector Machine, and Multi-Layer Perceptron. GNNs achieved comparable overall accuracy (91%) but delivered more spatially consistent predictions and higher F1-scores (0.89) for the minority class (seagrass presence). Sensitivity analysis identified climatic and human variables as key drivers of seagrass distribution. These insights support the implementation of blue Nature-based Solutions (NbS) to protect and restore seagrass habitats, aiding biodiversity conservation and climate change mitigation while guiding effective policymaking.

1. Introduction

Marine and coastal ecosystems (MCEs) are of key importance for human well-being, playing a critical role in climate regulation mechanisms on Earth and producing life-support functions and services such as carbon sequestration, coastal protection against sea level rise and flood control (Albert et al., 2021). However, they face serious threats, Watson et al. (2019) including one deriving from the interaction between multiple human pressures (e.g. land and sea-based pollution, over-exploitation of fish, aquaculture) and pressures more closely related to climate change (e.g. rising sea temperatures sometimes resulting in marine heat waves, increased occurrence of climate and weather extremes, ocean acidification, etc.). The cumulative and interacting nature of these pressures poses a significant challenge to the health and resilience of MCEs (Salomidi et al., 2012; Berrouet et al., 2018; Watson et al., 2019). To respond to this urgent global challenge, recent studies have highlighted the critical role of Nature-based Solutions (NbS), including coastal vegetation, wetlands, and particularly seagrasses, in mitigating the impacts of climate change and preserving biodiversity (do Amaral Camara Lima et al., 2023; O'Leary et al., 2023; Forrester et al., 2024; Ozkiper et al., 2024; Casal et al., 2025). These nature-driven interventions harness the intrinsic functions of healthy ecosystems to deliver both climate resilience and ecological benefits, offering cost-effective and sustainable alternatives or complements to engineered solutions. Among MCEs, seagrass meadows, salt marshes, and mangroves are especially important due to their capacity to act as blue carbon sinks, stabilise coastlines, and support high levels of biodiversity (Duarte et al., 2010; Zhu and Yan, 2022). However, the effectiveness of NbS is inherently dependent on the health and integrity of the ecosystems involved. Degraded or fragmented habitats significantly reduce their potential to buffer climate extremes, regulate nutrient cycles, or provide refuge for marine species. Protecting the condition of MCEs is therefore essential not only for maintaining biodiversity but also to ensure that NbS delivers their full potential in achieving climate adaptation and mitigation goals. Monitoring, modelling, and protecting these systems, especially under growing anthropogenic pressures, has become a priority in international conservation agendas, including the EU Biodiversity Strategy and the UN Decade on Ecosystem Restoration. However, the inherent

☆ This article is part of a Special issue entitled: 'Sys-mod in Big Sci' published in Environmental Modelling and Software.

* Correspondence to: Ca' Foscari University of Venice, Via Torino 155, Office 514A, Alfa Building (5th Floor), 30170 Venice Mestre, Italy.
E-mail addresses: angelica.bianconi@iusspavia.it (A. Bianconi), sebastiano.vascon@unive.it (S. Vascon), elisa.furlan@cmcc.it (E. Furlan), andrea.critto@unive.it (A. Critto).

https://doi.org/10.1016/j.envsoft.2025.106678
Received 29 October 2024; Received in revised form 8 August 2025; Accepted 3 September 2025
Available online 11 September 2025

complexity and multidimensionality of MCEs often limit the effectiveness of traditional ecological modelling approaches. Mathematical models have been widely used to investigate the impacts of climate change on marine ecosystems (Mandal et al., 2021a,b, 2022a,b), providing valuable insights into species interactions, trophic dynamics, and long-term ecological scenarios. These models, however, often rely on predefined equations or assumptions and are less flexible when dealing with heterogeneous, large-scale environmental datasets. In contrast, machine learning (ML) techniques offer data-driven alternatives that can efficiently model non-linear relationships and extract patterns from high-dimensional environmental data (Simeoni et al., 2022).

Notable examples include the study by Teichert et al. (2016) that investigated the influence of multiple stressors on fish ecological quality in European estuaries by using a RF. This algorithm is employed to model the response of the 'ecological quality ratio' (EQR), the indicator representing ecological status, to the effects of different stressors. Another study concerning the prediction of ecological indicators using ML methods is performed by Stock et al. (2018), in which ten statistical and ML algorithms (e.g., Classification and Regression Trees, RF and Boosted regression trees) were used and compared in order to understand the cumulative effects of multiple stressors through the forecast of three ecological indicators of MCEs condition (i.e., kelp biodiversity, fish biomass, and rocky intertidal biodiversity) within the California coast. Although current ML methods are useful for the assessment of environmental stressors on MCEs, they often over-look the spatial dependence of pressure effects. This limitation could be addressed through the adoption of alternative models, particularly those leveraging Graph Neural Networks (GNNs).

In recent years, GNNs have gained some popularity in research related to ML (Wu et al., 2020), but in the field of environmental science are still barely used. Few applications use them for predicting extreme events (Yang et al., 2022; Cachay et al., 2021), monitoring air quality (Wang et al., 2020; Ouyang et al., 2021; Dun et al., 2022), forecasting weather (Yu et al., 2021; Ni et al., 2022; Lira et al., 2022; Ma et al., 2022; Lin et al., 2022) and managing water resources (Zanfei et al., 2022; Sun et al., 2021). These studies demonstrate that modelling data as graphs provides a powerful way to capture the spatial structure of environmental processes, enabling the model to learn from both local and non-local interactions. In parallel, Gaussian Processes (GPs) (Jiao and Tao, 2025) have demonstrated substantial value in modelling environmental data, particularly in contexts where quantifying predictive uncertainty is critical. By leveraging kernel-based formulations, GPs can naturally model spatial and spatiotemporal correlations, making them especially suitable for interpolation tasks such as kriging in geostatistics or for characterising complex physical processes in data-scarce regimes (Feng et al., 2025; Camps-Valls et al., 2016; Rasmussen, 2003). Their non-parametric nature allows flexible modelling without strong assumptions on functional forms. Recent applications include: air pollution mapping (Xu et al., 2021; Patel et al., 2022), groundwater salinity prediction (Lal and Datta, 2018; Cui et al., 2021; Shadrin et al., 2021), weather and climate forecasting (Wang et al., 2022; Donnelly et al., 2024) and carbon flux estimation from satellite data (Campos-Taberner et al., 2024). While traditional GPs are computationally expensive with large datasets due to their cubic scaling with data size, recent innovations such as sparse approximations, inducing point methods, and variational inference have significantly enhanced their scalability and practical use in environmental applications. More recently, Physics-Informed Neural Networks (PINNs) have emerged as a compelling alternative for problems governed by known physical laws. PINNs integrate partial differential equations (PDEs) directly into the training objective of neural networks, ensuring that learned models respect physical constraints such as conservation of mass, momentum, or energy (Raissi et al., 2017). This approach is particularly advantageous for scenarios where observational data is limited but physical equations are well established. In environmental sciences, PINNs have been successfully used as surrogate models for computationally intensive simulators, including hydrodynamic models, climate systems, and atmospheric transport. Applications include simulating coastal flooding and storm surges (Fu et al., 2024; Zhu et al., 2025), contaminant transport (Bertels and Willems, 2023; Ke et al., 2025), and groundwater systems (Cuomo et al., 2023; Ali et al., 2024), where PINNs can capture complex dynamics with relatively little training data. By embedding domain knowledge directly into the learning process, PINNs offer data efficiency, interpretability, and physical consistency, which are essential for trustworthy modelling in environmental contexts. Within this diverse and evolving modelling landscape, this study explores the potential of Graph Neural Networks (GNNs) for large-scale ecological prediction. Particularly, this is the first application of GNNs to predict the spatial distribution of seagrass (i.e., *Posidonia Oceanica* and *Cymodocea Nodosa*) across the entire Italian coastal zone. While previous AI-based studies on seagrasses have primarily focused on local mapping tasks using remote sensing with Sentinel-2 data or image classification (Traganos et al., 2018; Carpenter et al., 2022; Li et al., 2023), this approach investigates whether spatially structured deep learning models can improve predictive accuracy under the combined influence of anthropogenic and climate-related pressures. The main contributions of the study are as follow:

- A comprehensive, spatially explicit dataset of 2244 grid cells (4 km resolution) was compiled, integrating environmental drivers and human-induced pressures to model seagrass presence/absence.
- Two GNN architectures, Graph Convolutional Network and Graph Attention Network, were developed and benchmarked against widely used ML models, including Random Forest, Support Vector Machine, and Multi-layer Perceptron.
- A detailed performance evaluation and sensitivity analysis was conducted to assess model accuracy and identify the most relevant predictors.
- By bridging spatial ecology, environmental data science, and deep learning, the study introduces a novel modelling framework that can be scaled or adapted to other marine ecosystems worldwide.

Unlike the few environmental applications that employ GNNs and the many traffic forecasting applications where they are widely used (Jiang et al., 2022; Lan et al., 2022; Li et al., 2022; Wang et al., 2020), the temporal component of the data is not modelled in this study due to the significant lack of available data on seagrasses at the Mediterranean and Italian scale (McKenzie et al., 2020). However, the spatial component is included in the graph construction. Accordingly, this work aims to evaluate how GNNs can be a key tool in representing the environmental processes that characterise MCEs, especially the Seagrasses ecosystem. Moreover, as they create biodiversity hotspots, this work aims to evaluate how deep learning can be a strategic tool to support the implementation of EU Marine Strategy Framework Directive, contributing to the monitoring of biodiversity as an element in achieving good environmental status.

2. Case study

The Italian Seas (Fig. 1) are part of the Mediterranean Sea. Their maximum depth exceeds 5000 m (in the Ionian Sea), and their average depth is about 1500 m. The Italian peninsula exhibit diverse oceanographic features influenced by the entire Mediterranean dynamics, encompassing ecological conditions from both Western and Eastern subbasins. Understanding their oceanographic features requires considering the Mediterranean Sea as a whole (Danovaro and Boero, 2019). Having narrow continental shelves and a large surface of open sea, much of the Mediterranean basin can be classified as a deep sea and includes some unusual features such as the variation of temperatures from 12.8 °C–13.5 °C in the western basin to 1–3.5 °C–15.5 °C in the eastern basin and the high salinity of 37.5–39.5 psu (Coll et al.,

2010; Piroddi et al., 2015; Guiot et al., 2020). In terms of climate, the Mediterranean region is characterised by windy, mild, wet winters, hot and dry summers (UNEP, 2014), while in terms of biodiversity, it is one of the world's 25 hotspots for biodiversity because of the presence of a high number of different ecosystems (Coll et al., 2010). Many of them are subjected to the rapid alteration under the combined pressure of climate change and human impact (Coll et al., 2010). Among all the species in the Mediterranean, the seagrass ecosystem deserves special attention as one of the most productive and valuable ecosystems in the entire Mediterranean basin (Chefaoui et al., 2018). Indeed, it plays an important role in providing critical ES such as carbon sequestration and storage, nursery habitat and food, water quality improvement and coastal erosion prevention (UNEP, 2020). Seagrasses are flowering plants which produce seeds and grow through the substrate by extension of their underground rhizomes and production of new leaves, creating complex, rich and highly productive habitats (Hemminga and Duarte, 2000). Although in the Mediterranean basin the photic zone reaches a maximum depth of 150 m (Maes et al., 2020), seagrasses are mainly located in shallow water (up to 40–50 m depth) (Weatherdon et al., 2017; UNEP, 2020), where there is the right amount of light to allow their photosynthetic pigments to perform photosynthesis. The collective term "seagrasses" includes more than 70 species around the world. Nevertheless, in the Mediterranean Sea, species that can be recognised are Posidonia oceanica (representing the 23% of all shallow bottoms) (Castejón-Silvo and Terrados, 2012), Cymodocea nodosa, Zostera marina, Zostera noltii, Ruppia cirrhosa, Ruppia maritima, and Halophila stipulacea (Ruiz et al., 2015). Focusing on the seas of the Italian peninsula, however, the most widespread species is P. oceanica. Initially, the meadows were characterised by a rather continuous distribution along the continental and insular coasts of the Tyrrhenian, Ionian, and southwestern Adriatic, with the exception of the main river mouths. Subsequently, regression has been documented along the continental coasts of Liguria, the Ionian Sea, and the southwestern Adriatic, recording a total regressed area of 34,472 hectares nationwide over the past 20–30 years (Telesca et al., 2015). In addition, a decrease on the habitat has also been observed in the Venice lagoon where the sub-habitat of Cymodocea nodosa beds had greatly declined since the beginning of the 20th century. Small patches are also present in Friuli (northeast) (Telesca et al., 2015). In the map in Fig. 1, the distribution of seagrasses in the Italian Seas is reported, based on Emodnet[1] and UNEP[2] data for the years 2016, 2017 and 2018. Key requirements for seagrasses growth can be summarised into three main classes: (i) habitat suitability: depth, sediment substrate, temperature and water movement; (ii) water quality: adequate light for photosynthesis (high minimum light requirements, 4.4%–29% of surface irradiance), salinity, absence of toxicants; (iii) grazing and recruitment processes: suitable assemblages of grazing animals, water movement to transport seeds and vegetation fragments (O'Brien et al., 2018). As mentioned above, Seagrass habitat plays a key ecological role in the marine environment, providing a long list of valuable ES such as food security, climate change mitigation, ocean acidification buffer (covering only 0.1% of the ocean floor, they efficiently store up to 18% of the world's ocean carbon) (UNEP, 2020). Moreover, they also contribute to fishery by supporting food webs, enrichment of biodiversity by providing valuable nursery habitat, nutrients cycling, absorption of pollutants by filtrating water, diseases control, protection against coastal erosion and tourism (Campagne et al., 2015). Unfortunately, global climate change is an emerging threat to terrestrial and aquatic ecosystems worldwide. Mediterranean and Italian seagrasses are facing a rapid decline in growth and demography due to the impact of pressures related to climate change (Jordà et al., 2012; Grizzetti et al., 2016; Díaz et al., 2019; Guiot et al., 2020) and human activities (Roca et al., 2017; Tičina et al., 2020; Smith and Rodríguez-Labajos, 2021) such as the discharge of excess nutrients and hazardous substances, marine litter and overfishing (Chefaoui et al., 2018). The impact of these pressures threatens the functioning and resilience of the ecosystem to further disruption, resulting in a reduction in the flow of ES and the ability to ensure human well-being.

3. Dataset

3.1. Data collection

The dataset used to conduct the experiments was obtained from the collection and pre-processing of a huge amount of heterogeneous data able to represent spatial distribution and intensity of both endogenic and exogenic pressures (Elliott et al., 2015) related to climate change and human impact, as well as detailed information on seagrasses health and biodiversity. To this objective, several open-source web-data platforms were screened (e.g., Copernicus Services, EU-Atlas of the Sea, Worldclim, UNEP and EMODnet data), paying particular attention to the availability of high spatio-temporal resolution data.

First of all, bathymetric data,[3] useful for framing the boundaries of the case study area, were retrieved from the EMODnet database[1]. Then, focusing on the most relevant stressors affecting seagrasses meadows in the Italian region, data on both endogenic (e.g. variables regarding nutrients, dissolved oxygen, water transparency, etc.) and exogenic (e.g. sea surface temperature, pH, etc.) pressures were recovered from the Copernicus Marine Environment Monitoring Service (CMEMS).[4] This platform provides free and open scientifically-assessed ocean data across the global ocean to enable marine policy implementation and scientific innovation. In addition to these stressors, spatial layer on the "kinetic energy at the seabed due to currents" was retrieved from EMODnet Platform[1]. In particular, this indicator was calculated by the EMODnet Seabed Habitats project consortium exploiting CMEMS products. As regards shipping traffic, map on the vessel traffic density (number of routes per square km per month), was collected from the EMODnet Human Activities database web portal.[5]

Additionally, to evaluate the influence of human coastal activities and urban areas on seagrasses health and distribution, different indicators and metrics related to the distance to the human settlements (i.e. ports, main cities and river mouths) have been collected and pre-processed. In particular, an open-source layer representing the distance from ports along the Mediterranean coastline was gathered from Global Fishing Watch.[6] Then, the minimum distance of each seagrass polygon to closest main Italian cities and river mouths have been calculated by applying the Haversine[7] formula (based on two different datasets containing information on major rivers and cities located near the Italian coasts, respectively).

Finally, for data on ecosystems' health, several seagrass distribution maps across the Italian coast were collected. Specifically, data from broad range UNEP-WCMC global biodiversity standardised database for the year 2017 were combined with the seagrass coverage layer produced by EMODnet Seabed Habitats (for the years 2016, 2017, 2018) in order to obtain the most complete representation of seagrasses distribution in the investigated marine region.

[1] https://www.emodnet-seabedhabitats.eu.
[2] https://www.unep.org/data-resources.
[3] The EMODnet Digital Terrain Model (DTM) has been generated for European sea regions (36 W,15N; 43E,90N) from selected bathymetric survey data sets, composite DTMs, Satellite Derive Bathymetry (SDB) data products, while gaps with no data coverage were completed by integrating the GEBCO Digital Bathymetry.
[4] https://marine.copernicus.eu/.
[5] www.emodnet-humanactivities.eu.
[6] https://globalfishingwatch.org/data-download/.
[7] Angular distance between two points on the surface of a sphere.

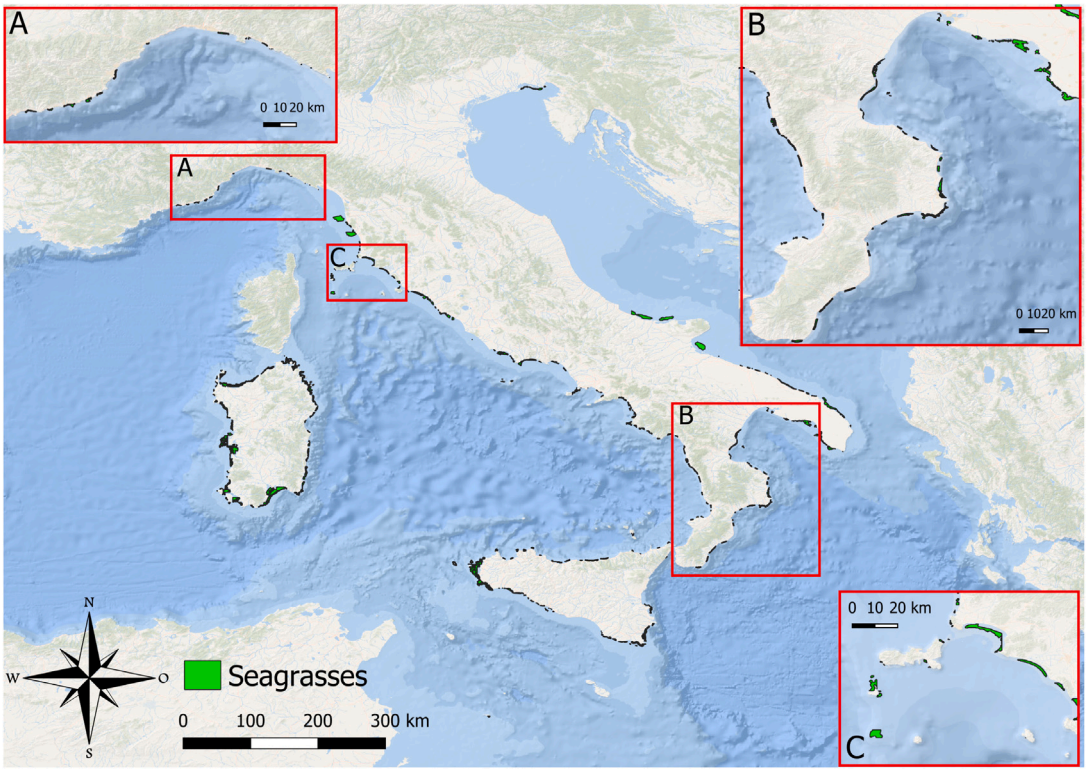


Fig. 1. The distribution of Seagrasses in the case study area for 2017; Focus A: Ligurian coast; Focus B: Calabria, Basilicata and part of Puglia coasts; Focus C: Elba Island and part of the Tuscan coast.

Table 1
List of indicators and metrics selected for the Italian Seas to represent pressures.

	Indicator	Annual metrics (2017 as reference)	Source
Endogenic pressures	Distance from ports	Derived from the source	Global Fishing Watch[6]
	Distance from rivers	Calculated through Haversine distance	Global Fishing Watch[6]
	Distance from cities	Calculated through Haversine distance	Global Fishing Watch[6]
	Nutrients (NH4, NO3, PO4)	Minimum, 5th percentile, 95th percentile	CMEMS[4]
	Oxygen	Minimum, 5th percentile	CMEMS[4]
	Chlorophill-a	Maximum, 90th percentile	CMEMS[4]
	Secchi depth	Minimum, 5th percentile	CMEMS[4]
	Light attenuation	Minimum, 5th percentile	CMEMS[4]
	Shipping traffic (density)	Mean	EMODnet[5]
Exogenic pressures	Sea surface temperature	95th percentile, standard deviation, number and intensity of marine heatwaves	CMEMS[4]
	pH	Minimum, mean, maximum	CMEMS[4]
	Salinity	Minimum, standard deviation, 5th percentile	CMEMS[4]
	Max significant wave height	Maximum	CMEMS[4]
	Eastward and northward sea water velocity	Mean, maximum	CMEMS[4]
	Kinetic energy at the seabed due to currents	90th percentile	EMODnet[1]
	Sea surface height	Mean	CMEMS[4]

3.2. Data preprocessing and analysis

With the aim of obtaining a homogeneous and high-quality dataset, the collected data were pre-processed and analysed. It was firstly necessary to frame the study area, taking, for all the pressures, as a reference the year 2017, due to the more complete seagrass ecosystem coverage in the whole Italian Seas. As emerged in literature, seagrasses are mainly located within 40–50 m of depth (UNEP, 2020), therefore a bathymetry layer up to 50 m depth was defined (Fig. 2). Then, all data collected from different open-source data platforms were preprocessed to homogenise their different spatial resolutions into a 4 km based raster grid. For each of the selected environmental indicators, representing human and climate change pressures, a set of yearly-based metrics (e.g. minimum, maximum, standard deviation, etc.) were calculated and mapped (Table 1).

As regards input variables (i.e., the exogenic and endogenic pressures), metrics were calculated using python codes. In particular, each of Netcdf files, as collected from different open-source portals, were processed through the *xarray*[8] library allowing to manipulate the data and calculate aggregated metrics.

As far as pressures related to coastal developments are concerned, the distance of seagrasses meadows to the closest major river mouths and urban areas was calculated by applying the Haversine distance formula and implemented in Python by means of the following libraries: *geopandas*,[9] *geocube*,[10] *rasterio*.[11] and *xarray*[8]

[8] https://xarray.pydata.org/en/stable/index.html.
[9] https://geopandas.org/en/stable/.
[10] https://corteva.github.io/geocube/stable/.
[11] https://rasterio.readthedocs.io/en/latest/.

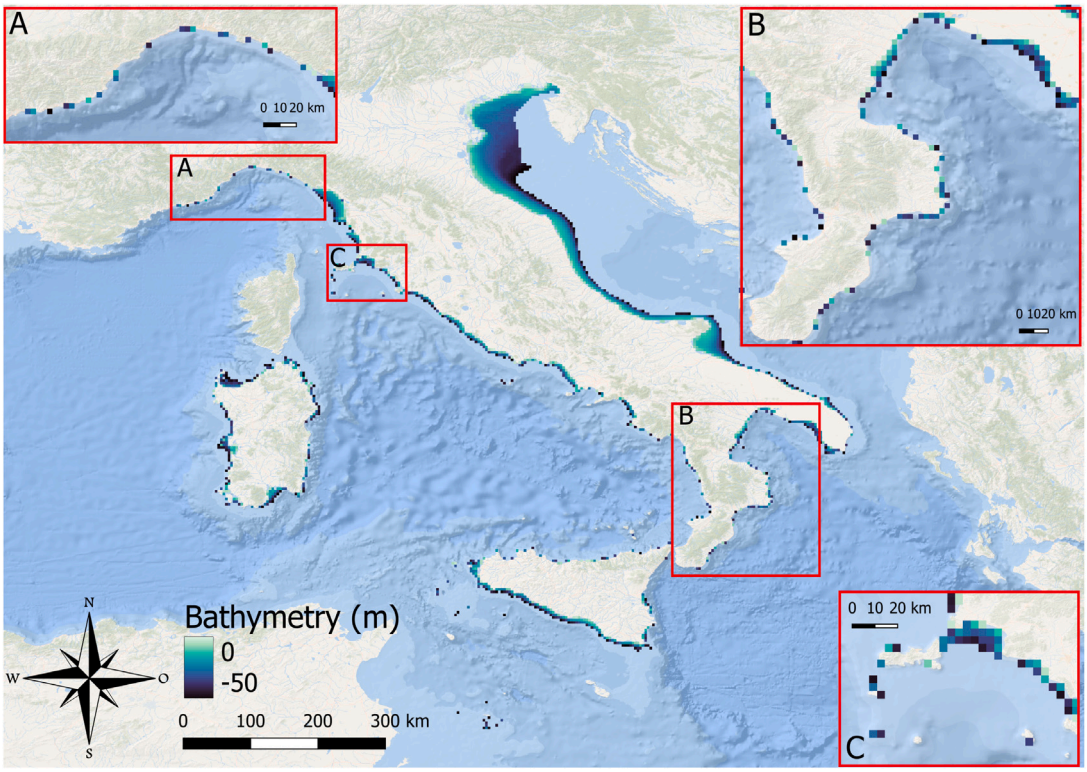


Fig. 2. Case study area corresponding to a bathymetry layer up to 50 m depth; Focus A: Ligurian coast; Focus B: Calabria, Basilicata and part of Puglia coasts; Focus C: Elba Island and part of the Tuscan coast.

A correlation analysis using a Pearson correlation coefficient was also performed to measure the strength of the linear relationship between the analysed variables and attempt to identify certain relationships, patterns, significant connections and variables that show a high correlation rate. In this study, the threshold (i.e. the degree of dependence) selected to remove highly correlated features is equal or greater than 0.85 (less than −0.85 respectively). As a result of this filtering process, the following variables were excluded from the dataset: the maximum value of pH, chlorophyll-a concentration and northward sea water velocity; the minimum values of salinity and nitrate; the 5th percentile of phosphate and light attenuation coefficient; the 95th percentile of nitrate; all metrics related to ammonium due to their high correlation with other nutrient parameters; and the mean value of eastward sea water velocity. After this step, the final set of input features used for model training and evaluation consists of 26 variables. Once the overall process of predictive variables' calculation and definition was completed, the descriptive statistics, including minimum, maximum, mean, median, standard deviation and interquartile range were computed and reported in Table 2, providing an overview of their variability and distribution. The data reveal a high degree of variability across multiple indicators, reflecting the heterogeneity of marine conditions within the Italian Seas. Water clarity, as represented by Secchi depth, showed substantial variation, with minimum values ranging from 1.32 m to 14.2 m and a mean of 4.23 m for the minimum metric. The 5th percentile values indicate deeper light penetration in some areas (up to 27.9 m), suggesting the presence of clearer waters, likely associated with offshore or low-nutrient environments. Currents, described via eastward and northward sea water velocities, showed low mean values (e.g., −0.02 m/s northward), with maximums not exceeding 0.88 m/s. This suggests moderate to weak circulation regimes, consistent with enclosed or semi-enclosed basins such as the Adriatic and Tyrrhenian Seas. The light attenuation coefficient also remained relatively low (mean = 0.04 m^{-1}), indicating generally good light penetration, an important condition for seagrass survival. In terms of anthropogenic pressures, variables like distance from ports, cities, and rivers revealed considerable spatial heterogeneity. For example, the distance from ports ranged from 0 to nearly 62 km (mean = 13.42 km), while distance from rivers extended up to 372 km. These distances are crucial to assess land-based runoff and potential pollutant exposure gradients. Vessel density showed a highly skewed distribution (mean = 9.57, median = 0.5), with extreme values reaching up to 4615 routes/km^2/month, indicating concentrated maritime traffic in specific high-pressure zones such as major shipping lanes and port vicinities. Regarding oceanographic energy conditions, kinetic energy presented a wide range (0–80.14 N m^{-2}) with a skewed distribution (median = 0.73), highlighting the presence of energetic coastal or shelf areas. The biogeochemical pressures showed patterns aligned with expected nutrient and oxygen dynamics. Nitrate concentrations (5th percentile) ranged broadly (0.31–34.79 mmol/m^3), with a mean of 4.51 mmol/m^3, while phosphate (95th percentile) reached up to 0.76 mmol/m^3. These elevated nutrient values suggest potential eutrophic zones. Dissolved oxygen minimum values ranged from 197 to 233 mmol/m^3, with a mean of 211, which remains above hypoxic thresholds, although some areas may approach low oxygen conditions. Sea surface temperature (SST) and marine heat waves represent climate-related pressures. SST 95th percentiles ranged from 296.8 to 300.9 K, while standard deviation values were moderate (mean = 5.35 K), indicating seasonal variability. Marine heat waves showed an average intensity of 1.87 and occurred roughly 2 times per year on average, with relatively low standard deviations, suggesting widespread warming patterns across the area. Chlorophyll-a, as a proxy for primary productivity, exhibited a wide range in its 90th percentile values (0.11–3.7 mg/m^3), reflecting variability in trophic state and possible coastal eutrophication. Finally, salinity showed expected ranges, with 5th percentile values from 31.58 to 38.93 PSU (mean = 37.26), capturing both freshwater-influenced coastal areas and more saline offshore waters. Standard deviation values remained low, indicating stable salinity regimes over time. The dataset thus captures a comprehensive and diverse range of pressures,

Table 2
Descriptive statistics of input variables based on 2244 observations.

Pressure	Unit	Min	Max	Mean	Median	Std	IQR
Secchi depth - minimum	m	1.32	14.2	4.23	3.36	2.82	3.86
Secchi depth - 5th percentile	m	2.35	27.92	13.89	13	6.15	9.47
Eastward Sea Water Velocity - maximum	m s^{-1}	0	0.88	0.19	0.17	0.1	0.13
Northward Sea Water Velocity - mean	m s^{-1}	−0.25	0.07	−0.02	−0.01	0.03	0.03
Light attenuation coefficient - minimum	m^{-1}	0.02	0.08	0.04	0.04	0.01	0.02
Distance from port	km	0	61.93	13.42	10.25	11.15	13.54
Vessel density	n.routes/km^2/month	0	4615	9.57	0.5	147.69	2.89
Distance from cities	km	0	161.32	44.37	43.86	29.37	41.41
Distance from rivers	km	0	372.32	72.43	45.76	79.22	75.32
Kinetic energy	N m^{-2}	0	80.14	3.74	0.73	7.13	3.49
Nitrate - 5th percentile	mmol m^{-3}	0.31	34.79	4.51	2.1	5.23	6.36
Ocean acidification - mean	pH	8.05	8.14	2.1	8.11	0.02	0.04
Ocean acidification - minimum	pH	7.91	7.99	7.96	7.96	0.01	0.02
Phosphate- 95th percentile	mmol m^{-3}	0.01	0.76	0.08	0.05	0.08	0.05
Phosphate - minimum	mmol m^{-3}	0	0.2	0.01	0.01	0.02	0
Max significant wave height	m	0.71	6.95	3.19	3.22	0.79	0.72
Number of marine heat waves - mean	–	1.20	2.75	2.04	2	0.23	0.3
Intensity of marine heat waves - mean	–	1.25	2.55	1.87	1.84	0.25	0.34
Sea Surface Temperature - std	Kelvin	3.45	6.91	5.35	5.68	0.98	1.86
Sea Surface Temperature - 95th percentile	Kelvin	296.83	300.9	299.14	299.3	0.6	0.72
Chlorophyll-a - 90th percentile	mg m^{-3}	0.11	3.7	0.65	0.5	0.59	0.64
Dissolved Oxygen - 5th percentile	mmol m^{-3}	226.33	270.65	243.82	241.56	9.33	13.31
Dissolved Oxygen - minimum	mmol m^{-3}	197.56	233.01	211.47	210.83	4.89	6.84
Salinity - 5th percentile	PSU	31.58	38.93	37.26	37.73	1.21	1.47
Salinity - std	PSU	0.04	1.41	0.44	0.33	0.32	0.55
Sea Surface Height - mean	m	−0.42	−0.28	−0.39	−0.39	0.01	0.02

encompassing both natural gradients and human-induced stressors. These variations are essential for modelling seagrass distribution and understanding their sensitivity to multiple interacting environmental conditions.

Similar procedures were carried out for the output variable depicted by seagrasses distribution. Specifically, the collected data were mapped in a 4-km resolution grid, covering the same extent as delineated in the case study. Each pixel of the obtained map was associated with a value of 1 in the case of the presence of meadows, 0 otherwise.

Based on the spatio-temporal resolution of the available data covering the selected case study area, the final dataset includes 2244 observations (i.e. pixels) within the Italian Seas. Concerning the response variable, it was divided into two classes (i.e. class 0 and class 1), class 0 specifies the absence of meadows in the considered pixel, while class 1 specifies their presence.

4. Method

With the main aim of comparing different artificial intelligence approaches to assess the Seagrasses ecosystem distribution in response to multiple pressures, firstly the potential of the GNNs was exploited, then the most widely used models in the frame of environmental applications (i.e. RF, SVM, MLP) were implemented and evaluated. Fig. 3 illustrates the overall workflow: starting from data collection and preprocessing, already detailed in Section 3, then continuing through graph construction, model development, training, and evaluation. While only the data collection-related steps have been previously addressed, the subsequent stages depicted in the workflow anticipate the structure of the following paragraphs. More precisely, in this section, GNNs are described, specifying the main features and peculiarities. Starting with the explanation of graph construction, the implementation description of the two GNNs models is provided. In addition, details of the basic models that were compared with GNNs are further defined.

4.1. Graph neural network

Graph neural networks (GNNs), introduced by Gori et al. (2005), are a family of deep neural network capable of manipulating graph-structured data in which information propagation occurs across nodes, so that the aggregated information can capture both feature and topological information. In the propagation module, the convolution operator is the one most commonly used for GNN models. The main idea of convolution operators is to generalise convolutions from other domains to the graph domain. Advances in this direction are often classified as spectral and spatial approaches. Spectral approaches include the Graph Convolutional Network (GCN), whose goal is to learn a function of features on a graph which takes as input a feature description for each node, summarised in a feature matrix, and a representative description of the structure of the graph in the form of adjacency matrix, and produces a node-level output. This type of network has undergone several evolutions (Hammond et al., 2011; Bruna et al., 2013; Henaff et al., 2015; Defferrard et al., 2016; Kipf and Welling, 2016) and the most recent (Kipf and Welling, 2016) was used in this study. As far as spatial approaches are concerned, Graph Attention Networks (GATs) are composed of an attention-based architecture to perform node classification of graph-structured data. The idea is to compute the hidden representations of each node in the graph, by attending over its neighbour, following a self-attention strategy, which allows the model to weigh the importance of neighbouring nodes (Veličković et al., 2017).

4.1.1. Graph notations

A graph is represented as $G(V, E)$, where $V = \{v_i\}_{i=1}^{N}$ is a set of N nodes and $E = \{e_{ij}\}$ is a set of edges connecting pairs of nodes (v_i, v_j), where $v_i \in V$ and $v_j \in V$. The neighbourhood of a node v is defined by a set of nodes connected to it, $\mathcal{N}(v) = \{u \in V | (u, v) \in E\}$. Node connections are described by the adjacency matrix $A \in \{0, 1\}^{N \times N}$, where a_{ij} is equal to 1 when there exists an edge from node v_i to v_j, and 0 otherwise. The feature vector of node v is denoted as $x_v \in \mathbb{R}^F$, where F is the total number of features.

4.1.2. Graph convolutional neural networks

The goal of GCNs is to learn a signal/feature function on a graph $G(V, E)$ that takes as input a feature description x_v for each node v and a representative description of the graph structure in matrix form (i.e., the adjacency matrix A) and produces a node-level output Z (a matrix of $N \times M$ features, where M is the number of output features per node). To provide the mathematical formulation of GCNs, some further notations must be introduced. Given a graph $G(V, E)$, D is defined as

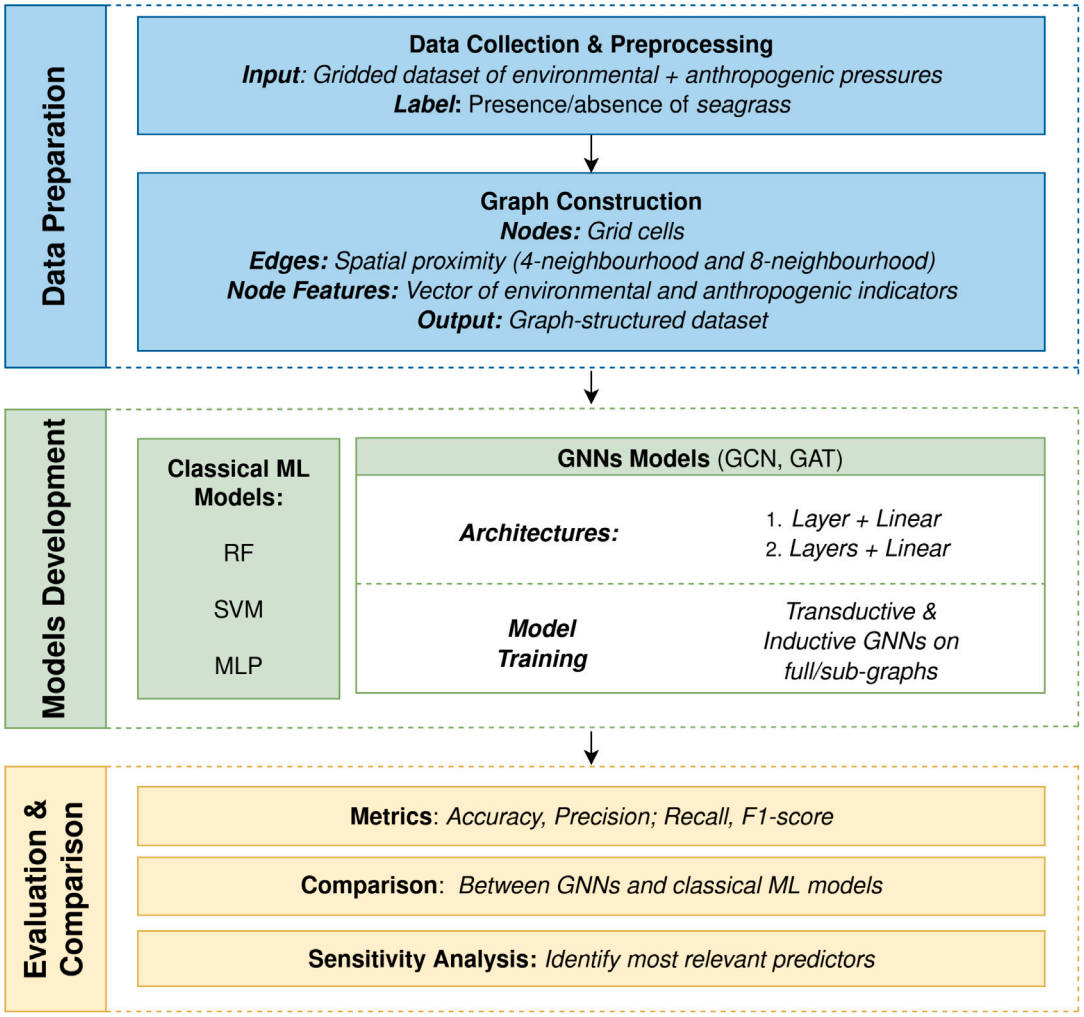


Fig. 3. Methodological workflow for seagrass distribution modelling. The workflow consists of three main phases: (1) Data Preparation, where environmental and anthropogenic variables are processed into a graph-structured dataset with grid cells as nodes and spatial proximity as edges; (2) Model Development, involving both classical ML models (RF, SVM, MLP) and Graph Neural Networks (GCN, GAT) trained in transductive and inductive modes; and (3) Evaluation and Comparison, where models are assessed using classification metrics and sensitivity analysis is used to identify key predictors.

the degree matrix, in which $D_{i,j} := \begin{cases} \deg(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$, where the degree $\deg(v_i)$ of the vertex v_i is the number of edges connecting it.

A layer-wise propagation rule of a multi-layer GCN is defined as:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

Here, $\widetilde{A} = A + I$ is the adjacency matrix of the undirected graph G with added self-connections corresponding to the identity matrix I, $\widetilde{D}$ is the diagonal degree matrix $\widetilde{D_{ii}} = \sum_j \widetilde{A_{ij}}$ and $W^{(l)}$ is a layer-specific trainable weight matrix. $\sigma(\cdot)$ denotes a non-linear activation function and $H^{(l)} \in R^{N\times D}$ is the matrix of activations in the lth layer; $H^{(0)} = X$.

This propagation rule is strictly connected with spectral graph convolution, fully explained in Kipf and Welling (2016).

4.1.3. Graph attention neural networks

GATs, introduced by Veličković et al. (2017), offer a more nuanced approach to neighbourhood information aggregation than traditional GNNs by exploiting attention mechanisms for feature learning on graphs. In fact, unlike standard GNNs, such as GCNs, they assign an attention coefficient to each neighbour, indicating the importance of that neighbour's features for updating node features.

The coefficients computed by the attention mechanism are expressed as:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T[W\vec{h}_i \parallel W\vec{j}]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\vec{a}^T[W\vec{h}_i \parallel W\vec{j}]\right)\right)} \tag{2}$$

where a^T represents transposition and $\parallel$ is the concatenation operation. $\vec{a} \in \mathbb{R}^{2F'}$ is the weight vector, the LeakyReLU is a nonlinear activation function and $W \in \mathbb{R}^{F'\times F}$ is the weight matrix.

The final output features for every node is computed as the weighted sum by a of the neighbouring new node representation passed through a non linearity:

$$\vec{h}'_i = \sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}W\vec{h}_j\right) \tag{3}$$

To stabilise the learning process of self-attention a *multi-head attention* mechanism is employed. Particularly, multiple k independent attention score estimators are computed:

$$\vec{h}'_i = \parallel_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k W^k\vec{h}_j\right) \tag{4}$$

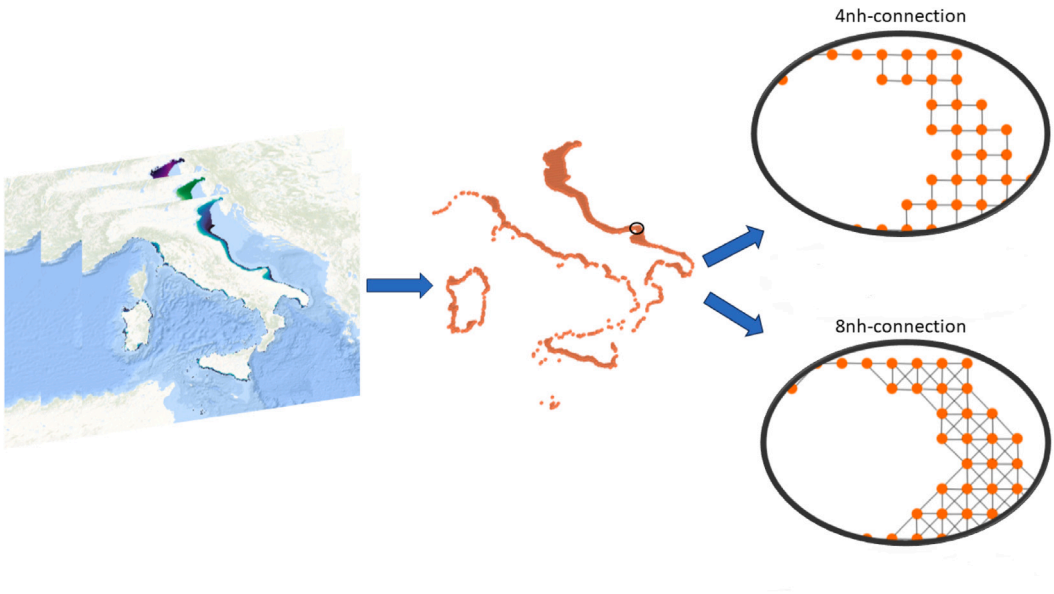


Fig. 4. Graph construction scheme. On the left, the layers corresponding to the pressure variables are shown. They are mapped as a feature vector associated with each node (in orange) across the study area. To illustrate how spatial connections are defined, two zoomed-in views are displayed on the right for a randomly selected subregion. These insets visually demonstrate how the entire study area is encoded as a graph, where each node is connected to its four immediate neighbours (4-neighbourhood, top inset) or to all surrounding pixels (8-neighbourhood, bottom inset), depending on the connectivity strategy adopted.

where $\|$ represents concatenation, α_{ij}^k are normalised attention coefficients computed by the kth attention mechanism, and W^k is the weight matrix of the corresponding input linear transformation. In the final layer, multi-head attention is computed by the average:

$$\vec{h}_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k W^k \vec{h}_j\right) \tag{5}$$

4.2. Graph structure and GNNs implementation

The main component of GNNs is definitely the graph construction (Zhou et al., 2020). To map the described dataset, given the fact that it is gridded, grid cells (i.e. geographic locations) can be naturally mapped into nodes of a graph. As can be seen in Fig. 4, each pixel belonging to the case study and identified by a latitude and longitude coordinate pair, is mapped to a node v of the graph. In addition, layers containing metrics related to the indicators selected as input variables are vectorised into an x_v feature vector associated to each node. The edges of the graph, which model the flow of information between nodes, were selected based on the distance between nodes. Specifically, to construct the adjacency matrix A, each node is connected to the four or eight nearest nodes (i.e., pixels) as can be seen in the enlargements "4nh-connection" and "8nh-connection" in Fig. 3.

As for the experiments conducted on GNNs, they were performed on the dataset introduced in Section 3. The dataset was initially used to construct the graph as described above. Secondly a mask was applied to each node of the graph specifying whether the node is a training, validation or test node. Particularly, 20% of the nodes in the graph (i.e. 20% of the observations in the dataset) were specified as validation set nodes, 10% as test set nodes and the remaining 70% as training set nodes. Regarding GNNs, several architecture-related configurations were tried out for GCNs and GATs for both the graph with 4-node connectivity and the graph with 8-node connectivity. First, an architecture with a convolutional layer (attention layer, respectively) followed by a linear layer (reported in Tables 3 and 4 as 1ly) was tested, then with three convolutional layers (attention layers, respectively) and a linear layer (reported in Tables 3 and 4 as 3ly) was also implemented. In addition, for each architecture, two distinct training approaches were employed: transductive, in which the model was applied to the entire graph, and inductive, in which the graph was divided into subgraphs, and training was performed on each of them. To obtain the subgraphs, a sampling technique was adopted, whereby a weighted random sampler selects the elements based on the class distribution. As such, it will remove samples from the majority class (undersampling) or add more samples from the minority class (oversampling).

4.3. Models configurations and parameters

4.3.1. Compared models

In order to validate the effectiveness of the proposed GNN-based models, they were compared with several models usually used in the context of ecosystem state assessment:

- RF: combines multiple deep decision trees, using bootstrap samples and a bagging strategy to reduce variance. RF also employs random feature subspacing (Ho, 1998), where each tree is built using a random subset of features. This reduces correlation between trees and increases model independence. For predictions, each tree votes on the output, and these votes are aggregated to determine the final result.
- SVM: is a non-parametric supervised learning technique important for handling both linear and non-linear data. It aims to find an optimal hyperplane by maximising the margin between the classes' closest points, known as support vectors. For non-linear data, SVM uses a kernel function to map the training data into a higher-dimensional feature space, where it computes the separating hyperplanes that maximise class separation (Cervantes et al., 2020).
- MLP: is a supervised artificial neural network with input, hidden, and output layers of interconnected neurons. Neurons connect to all neurons in the next layer without feedback. Initial weights are set randomly. During training, input/output pairs are used to calculate outputs and compare them to actual values, producing an error term. Weights are adjusted to reduce this error. This process repeats until the error is minimised or a maximum number of iterations is reached (Hastie et al., 2009).

Table 3
Comparison results between different GCNs.

GCNs comparison results							
Model	Accuracy	Precision		Recall		F1 - score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
GCN-inductive-4nh-1ly	0.88	0.96	0.73	0.85	0.92	0.91	0.81
GCN-inductive-4nh-3ly	0.88	0.95	0.74	0.87	0.89	0.91	0.81
GCN-inductive-8nh-1ly	0.88	0.94	0.80	0.86	0.90	0.90	0.85
GCN-inductive-8nh-3ly	**0.91**	**0.98**	0.82	0.87	**0.98**	**0.92**	**0.89**
GCN-transductive-4nh-1ly	0.88	0.92	0.79	**0.91**	0.80	0.91	0.80
GCN-transductive-4nh-3ly	0.88	0.95	0.74	0.87	0.89	0.91	0.81
GCN-transductive-8nh-1ly	0.89	0.93	**0.83**	0.89	0.89	0.91	0.86
GCN-transductive-8nh-3ly	0.89	0.95	0.82	0.88	0.92	0.91	0.87

Table 4
Comparison results between different GATs.

GATs comparison results							
Model	Accuracy	Precision		Recall		F1 - score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
GAT-inductive-4nh-1ly	0.88	0.96	0.73	0.86	0.92	0.91	0.82
GAT-inductive-4nh-3ly	0.88	0.96	0.75	0.87	0.91	0.91	0.82
GAT-inductive-8nh-1ly	0.89	**0.98**	0.79	0.85	**0.96**	0.91	0.87
GAT-inductive-8nh-3ly	0.89	0.96	0.81	0.86	0.94	0.91	0.87
GAT-transductive-4nh-1ly	0.87	0.92	0.76	**0.89**	0.82	0.91	0.79
GAT-transductive-4nh-3ly	0.88	0.97	0.72	0.85	0.94	0.91	0.82
GAT-transductive-8nh-1ly	0.88	0.93	**0.82**	**0.89**	0.88	0.91	0.85
GAT-transductive-8nh-3ly	**0.91**	**0.98**	**0.82**	0.87	**0.96**	**0.92**	**0.89**

4.3.2. Experimental settings

In this study, open-source Python libraries were used to build the models. Specifically, the RF and SVM models were developed using the *sklearn*[12] library, typically used in basic ML projects, while the MLP and GNNs were implemented using *PyTorch*,[13] a fully featured framework for building deep learning models, and *PyTorch Geometric*,[14] designed to facilitate deep learning on irregularly structured data like graphs. In order to make the results fully comparable, for models that are not applied directly on the graph structure, the observations corresponding to the training nodes were used to construct the training set, applying the same procedure to the validation and test data. To implement the hyperparameter tuning, random search for RF and SVM was used using *sklearn's RandomizedSearchCV* method. Specifically, in this approach, a search space can be defined as a bounded domain of hyperparameter values and points in this domain can be randomly sampled. As regard the MLP and GNNs, using the *optuna*[15] and *PyTorch* libraries, the Tree-structured Parzen Estimator (TPE) algorithm (Bergstra et al., 2011) was used for sampling, i.e. a Bayesian optimisation based on kernel fitting that, after sampling different areas of the search space, focuses its attention on the point where it obtained the best results and continues to search it. Moreover, for each combination, models were trained by performing cross validation and, at the end, the combination that obtains the highest accuracy on the validation test was selected.

In particular, for the RF a list of possible values have been set for: (i) *n_esimator*: the number of trees in the forest; (ii) *min_samples_split*: the minimum number of observations required to split a node; (iii) *min_samples_leaf*: the minimum number of samples required to be at a leaf node, (iv) *max_depth*: the maximum depth of the trees (i.e. the dimension of the longest path between the root node and the leaf). Focusing on the SVM model, the parameters searched were (i) *C*: the regularisation parameter, (ii) *kernel*: the type of kernel to be used in the algorithm, and (iii) *gamma*: the kernel coefficient. As far as MLP is concerned, two hyper-parameters were optimised. The *learning_rate*, responsible for minimising the loss function, and the *n_hidden_unit*, the number of neurons in the hidden layer. Regarding the activation function, the ReLU was adopted, as it is widely used. For the GNNs-based models hyperparameters, again the learning rate was optimised and the ReLU activation function was chosen, while list of values around the input units (i.e. 28) were selected for the number of units of each convolutional and attention layers. For both MLP and GNNs the final linear layer has only one output unit, corresponding to the response variable.

4.4. Evaluation of performances

In order to compare the different models, a consistent way to evaluate the model performance is required, so, in addition to accuracy, also precision, recall and F1 score will be examined, as the response variable have an unbalanced distribution of classes. Model accuracy returns the number of classifications the model correctly predicts divided by the total number of predictions made. Mathematically, model **accuracy** is expressed as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{6}$$

where TP stands for "True Positives", FP for "False Positives", TN for "True Negatives" and FN for "False negatives".

Recall (also known as sensitivity) highlights the number of members of a class that the classifier identified correctly, divided by the total number of members in that specific class. Mathematically, model **recall** is defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

On the other hand, model **precision** (also called positive predictive value) is the ratio between the True Positives and all the Positives. Mathematically, it is defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

Finally, the F1 score will be also calculated. This score is the weighted average of Precision and Recall, providing a way to express both concerns with a single score. As a consequence, this evaluation metric takes both false positives and false negatives into account. F1 is usually

[12] https://scikit-learn.org/stable/.
[13] https://pytorch.org.
[14] https://pytorch-geometric.readthedocs.io/en/latest/.
[15] https://optuna.org.

more useful than accuracy, especially if you have an uneven class distribution. Mathematically, the **F1 score** it is defined as follows:

$$\text{F1\_Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{9}$$

4.5. Sensitivity analysis

Given the importance of accurately predicting the distribution of seagrasses, it is crucial to understand which variables most significantly influence the outcome of this problem. To this end, a sensitivity analysis (SA) was conducted to identify all relevant input parameters from a set of potential variables. This analysis compared the results of ML and graph-based models to determine the key variables impacting the predictions. Additionally, the analysis aimed to examine the contributions of various climate change and human impact variables to the final predictions under different or randomised scenarios. Specifically, the Morris method was employed for this SA to address the complexity and interaction effects of the input variables within the models. The Morris method, introduced by Morris (1991), is a global SA technique based on a one-at-a-time (OAT) approach, allowing for the evaluation of the importance and interactions between variables. Unlike the traditional OAT method used in local SA, where variables are individually altered from a single baseline, the Morris method perturbs variables from multiple starting points in a randomised manner. This approach calculates the Elementary Effect (EE) of input variables on the output, utilising randomised sampling matrices. The EE is determined using the following formula:

$$EE_i = \frac{y(x_1, \dots, x_i + \Delta \dots x_j) - y(x_1, \dots, x_i, \dots, x_j)}{\Delta} \tag{10}$$

where x represents the model parameters as a j-dimensional vector $(x_1, \dots x_i, \dots x_j)$; $y(x)$ is the model output from the Morris simulation; Δ is the step size, defined as a multiple of $1/(p-1)$ with p being the number of levels.

Two key metrics (i.e., the mean μ_i and the standard deviation σ_i) are used to assess the significance and interactions of variables based on the set of EEs. After computing numerous independent EEs across several trajectories (r) (i.e., distinct/elementary paths for sampling in the multi-dimensional space of parameters) for each input variable, the mean value, representing the overall influence of the variable, is calculated as:

$$\mu_i = \frac{\sum_{n=1}^{r} EE_n}{r} \tag{11}$$

A high μ_i indicates a strong influence of the variable on the output. Additionally, σ_i, which reflects the distribution of EE values, represents higher-order effects due to the non-linear influence of the variable or its interactions with other variables:

$$\sigma_i = \sqrt{\frac{1}{r}\sum_{n=1}^{r}(EE_n - \mu_i)^2} \tag{12}$$

Furthermore, Campolongo et al. (2007) introduced a third measure, μ^*, representing the mean of the absolute values of the EEs to prevent the cancellation of positive and negative EE values and to provide an overall sensitivity measure:

$$\mu_i^* = \frac{\sum_{n=1}^{r} |EE_n|}{r} \tag{13}$$

5. Results

This section includes a comparison of different configurations of GCNs and GATs, highlighting performance trends across various graph structures and sampling methods. Subsequently, the top-performing model from each GNN family was compared with the baseline models. Additionally, the sensitivity analysis results comparing the best GNN model and the best baseline model are presented.

Table 5
Comparison results between benchmark and GNN-based models.

Comparison results							
Model	Accuracy	Precision		Recall		F1 - score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
RF	0.90	0.95	0.84	0.89	0.92	**0.92**	0.87
SVM	0.89	0.94	0.82	0.88	0.90	0.91	0.86
MLP	0.89	0.91	**0.87**	**0.92**	0.85	0.91	0.86
GCN	**0.91**	**0.98**	0.82	0.87	**0.98**	**0.92**	**0.89**
GAT	**0.91**	**0.98**	0.82	0.87	0.96	**0.92**	**0.89**

5.1. Model performances

After the training and validation phases, the models were investigated against the testing dataset. First the results of different architectures of GCNs (Table 3) and GATs (Table 4) were compared, then the GNNs-based models were analysed in comparison with the basic ML models, and Table 5 shows the results. As can be seen from Table 3, the architecture with 3 convolutional layers applied to the 8-node connected graph, trained with a sampling technique, exhibited superior performance. Similar behaviour is observed for GATs (Table 4), where 3 attentional layers applied to the 8-connection graph obtained the best performances. Models applied to the graph with 8-connected nodes consistently outperformed those with fewer connections. The inclusion of more connected nodes allows these models to better capture the spatial relationships and interactions within the data, leading to more accurate predictions. Despite the minimal differences in overall accuracy between the GNN-based models and the baseline models (Table 5), the GNNs demonstrated a distinct advantage in handling the presence of seagrass (class 1). This advantage is reflected in their higher F1 scores and recall rates for class 1, indicating that GNNs are better at recognising the presence of seagrass under varying environmental pressures. The spatial information captured by the 8-node connected graph allows these models to understand complex interactions that baseline models cannot effectively learn. For the absence class (class 0), the performance remained stable across all models, showing that while baseline models can handle the more straightforward task of identifying absence, they struggle with the nuances of predicting presence as effectively as GNNs. The precision and recall balance achieved by GNNs in predicting class 1 further demonstrates their robustness in making accurate predictions without sacrificing precision. Moreover, Fig. 5 presents the error maps for the RF and GCN models, respectively, when applied to the test set. In particular, correctly classified pixels are highlighted in light blue, while pixels misclassified as absence are marked in red. In contrast, instances classified as presence are shown in yellow. In Focus A of Fig. 5, it is evident that the RF model, compared with the GCN model, erroneously predicts seagrass absence in the middle of connected presence pixels. Additionally, in the same focus area, the RF model incorrectly classifies an isolated pixel as presence among absence pixels, thus demonstrating that treating pixels in isolation could not contribute to accurate predictions. Conversely, the errors associated with the GCN model are primarily located at the edges of the seagrass meadows. This observation, consistent in Focus B as well, further confirms that GNNs' incorporation of spatial context enhances prediction accuracy, particularly in complex ecological settings.

5.2. Sensitivity analysis

In this study, to assess the significance of pressures for the two most proficient models in ML and graph-based contexts (i.e., RF and GCN), the Morris method (outlined in Section 4.5) was applied utilising the open-source *SALib* library in Python (Herman and Usher, 2017; Iwanaga et al., 2022). The results were depicted through bar diagrams showcasing the values of μ^*, thereby highlighting the relative importance of each variable. Variables with high μ^* values are deemed

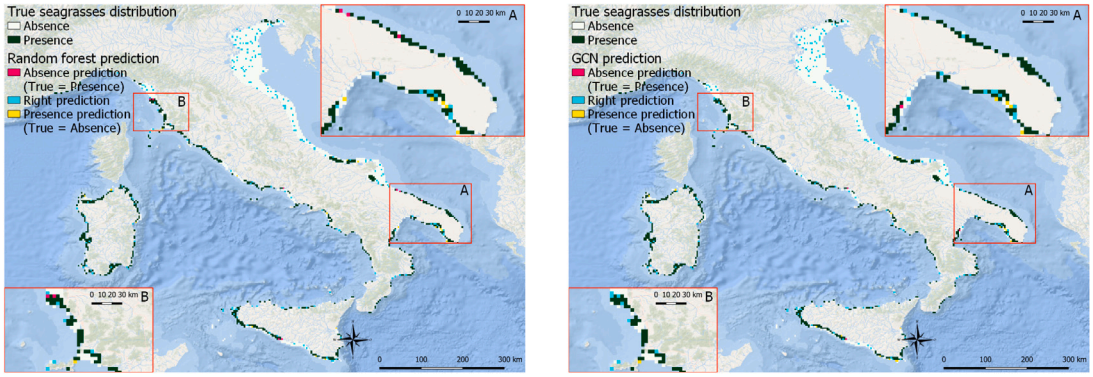


Fig. 5. Test set error maps related to RF (left) and GCN (right). Focus A: part of the Apulia region. Focus B: part of the Tuscany region.

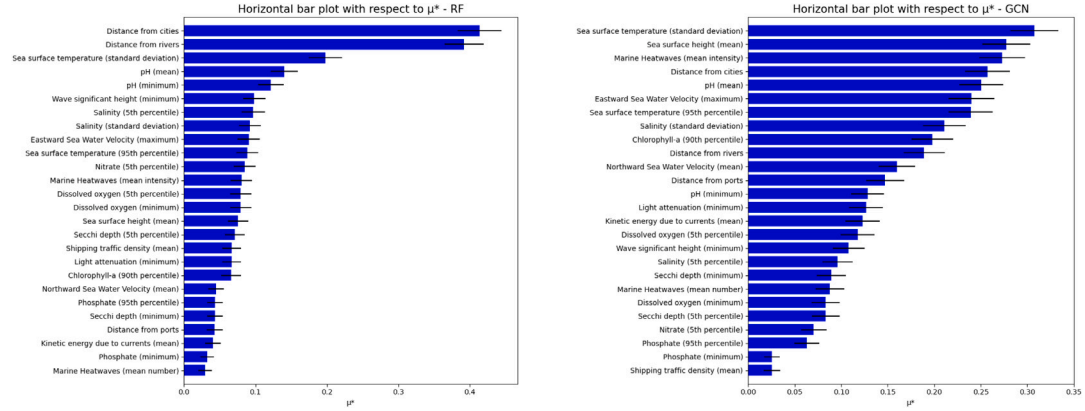


Fig. 6. Sensitivity analysis conducted using the Morris method. Left: results for RF. Right: results for GCN.

highly influential, whereas lower μ^* values indicate lesser significance. Confidence intervals of μ^* are also presented alongside the bars. Particularly, in Fig. 6, it is immediately evident that pivotal variables for RF include distances from ports and cities, with a μ^* value of approximately 0.4, followed by temperature and pH with μ^* values of approximately 0.2 and 0.1, respectively. The remaining variables have μ^* values equal to or below 0.1, rendering them relatively insignificant. Conversely, the results for GCN indicate no distinct disparity between the most and least significant variables. However, temperature, sea surface height, marine heatwaves, pH, distances from cities, currents and salinity are, in order, the most significant variables, exhibiting μ^* coefficients ranging from 0.2 to 0.35.

6. Discussion

6.1. Unveiling the power of spatial context

This study developed and evaluated GNN-based models to investigate the spatial dynamics between environmental stressors and biological responses, with a particular focus on seagrass distribution. A comparative modelling framework was employed to benchmark the performance of GNNs against conventional ML algorithms, such as RF, which typically treat observations as independent and do not incorporate spatial dependencies. While traditional models achieved high levels of predictive accuracy, GNN-based approaches, and specifically GCNs, demonstrated marginally superior results in terms of overall performance, and more notably, exhibited a distinct advantage in generating spatially coherent predictions. In particular, GCNs performed slightly better than RF models in predicting seagrass presence, with improvements in key metrics like the F1 score. In addition, visual analysis of the predicted maps showed that GCNs were able to capture more ecologically realistic spatial patterns, while RF models often produced fragmented and spatially inconsistent results. This enhanced spatial expressiveness, resulting from the graph-based representation of observations, constitutes a central contribution of the study. It enables a more ecosystemic and realistic depiction of spatial processes, which is especially valuable in the context of marine spatial planning where stressors and environmental influences exhibit clear spatial structure (e.g., sediment transport, currents). These findings build upon and extend prior research applying GNNs in environmental domains. For instance, Wang et al. (2020) utilised GNNs to predict PM2.5 concentrations in urban environments. Their study demonstrated that GNNs could effectively capture the spatial correlations between different monitoring stations, resulting in more accurate predictions compared to traditional models like SVM and RF. The ability of GNNs to model spatial dependencies among air pollution sources and monitoring points provided a clearer understanding of pollution dispersion patterns, which is critical for urban air quality management. This is also related to the domain-knowledge integrated in the construction of the graph. The latter is based on the physical factors influencing pollutant dispersion such as wind direction, speed, and geographical barriers resulted in more accurate models that could capture both horizontal and vertical pollutant transport. In climate science, Cachay et al. (2021) employed GNNs to forecast El Niño events by modelling the complex interactions between sea surface temperatures, atmospheric pressure patterns, and ocean currents. The GNN model outperformed conventional time-series models by effectively capturing the non-linear spatial dependencies driving these climatic events. Similar applications underscore the effectiveness of GNNs in environmental modelling, particularly in scenarios where spatial dynamics are crucial. The application of GNNs in seagrass monitoring, as explored in this study, extends this growing body of research by demonstrating how these models can be used to understand and predict ecological changes in marine environments. By leveraging the spatial relationships between stressors, GNNs offer a more

holistic approach to modelling that can lead to more informed and effective environmental management strategies. In contrast to existing studies on seagrass mapping that primarily rely on high-resolution satellite imagery at localised scales, the proposed approach integrates spatially distributed environmental variables to provide a broader, system-level perspective. Although remote sensing remains essential for high-resolution habitat mapping, it often lacks the ability to represent underlying ecological processes and interactions. The GCN-based approach presented here complements remote sensing by embedding spatial context and connectivity into the learning process, thereby enabling a more integrated and interpretable ecological analysis. Future work could enhance this framework by incorporating remote sensing-derived features within the graph structure, further improving spatial resolution and predictive accuracy. Despite these advantages, the modelling framework was limited by the resolution and completeness of available spatial datasets. Crucial physical processes known to affect seagrass dynamics such as ocean currents, wave exposure, and sediment transport could not be directly incorporated due to data constraints. Addressing this limitation represents a key frontier in environmental machine learning. In this regard, PINNs offer a promising pathway. By integrating physical laws directly into the learning architecture, PINNs enable the synthesis of mechanistic understanding with data-driven inference, improving both accuracy and interpretability. The integration of PINNs with GNNs remains a largely unexplored but potentially transformative research direction, capable of advancing the modelling of dynamic and spatially complex ecological systems.

6.2. Identify key stressors for effective management

In addition to methodological contributions, the study provides insights of direct relevance to environmental management, particularly within the Italian coastal and marine context. The sensitivity analysis conducted on the models provides critical insights into the primary factors influencing seagrass meadow ecosystems. According to the RF analysis, proximity to urban centres and river mouths emerges as the principal determinant shaping seagrass distribution. These findings are consistent with previous research (e.g., Catucci and Scardi, 2020) which demonstrated how nutrient enrichment and sedimentation from agricultural and industrial sources compromise seagrass health. In Italy, major river systems such as the Po and Arno transport substantial pollutant loads into coastal zones, particularly in highly urbanised and industrialised regions such as the northern Adriatic and Tyrrhenian coasts. These insights underscore the need for targeted mitigation strategies, including the enhancement of wastewater treatment infrastructure, the implementation of vegetated buffer strips, and stricter regulatory frameworks governing land-based pollution and runoff. Conversely, the GCN analysis emphasises the importance of sea surface temperature and other climatic variables, such as sea level rise and ocean acidification. These factors are critical because they directly influence seagrass growth rates, reproductive cycles, and overall ecosystem dynamics. The Mediterranean is particularly susceptible to these climatic changes, with seagrasses facing an accelerated rate of warming. The vulnerability of *Posidonia Oceanica* underlines the severity of the threat, as this species plays a crucial role in maintaining the health and stability of the Mediterranean marine ecosystem. Similar conclusions were drawn by Trois et al. (2024), who demonstrated that rising temperatures and shifts in salinity regimes could significantly alter the spatial extent and health of seagrass meadows. The results stress the importance of incorporating climate variables into conservation strategies, as they play a significant role in determining the resilience and long-term survival of seagrass meadows in the Mediterranean. In the Italian context, these findings call for the revision of national and regional Marine Strategy Framework Directive (MSFD) programmes of measures to incorporate climatic stressors explicitly and to develop early warning systems for thermally induced degradation of seagrass habitats. The integration of these results into management frameworks also supports the broader deployment of blue Nature-Based Solutions (NbS), particularly within the scope of Italy's marine and coastal restoration policies. Seagrass meadows, particularly *Posidonia Oceanica*, provide multiple ecosystem services, including carbon sequestration, sediment stabilisation, and biodiversity support, all of which are crucial for enhancing coastal resilience. Protecting and restoring these meadows represents a strategic intervention to address interconnected challenges such as climate change, coastal erosion, and biodiversity loss. In line with this, the recent adoption of the EU Nature Restoration Law, which mandates the restoration of at least 20% of the EU's terrestrial and marine ecosystems by 2030, provides a legislative framework that Italy can leverage to prioritise seagrass restoration. Model outputs from this research provide spatially explicit guidance to inform the allocation of resources and restoration priorities under the EU law. Specifically, areas near major urban agglomerations (e.g., Naples, Palermo, Venice) and large river estuaries can be targeted for restoration based on identified stressor intensity. Moreover, the integration of these predictive tools into existing marine spatial planning platforms can facilitate cross-sectoral coordination among national environmental agencies, regional authorities, fisheries, and tourism stakeholders. Such a systemic and evidence-based approach is essential to operationalise restoration efforts in line with the EU Green Deal, the national strategy for biodiversity conservation, and Italy's commitments under international frameworks for marine protection.

7. Conclusion

The main aim of this work was to explore the potential of GNN-based models to discover and unravel the complex relationships between human and climate change pressures and their impacts on the seagrass ecosystem in the Italian Seas. These models effectively integrate the spatial component of the data, extracting information and propagating it across the entire network. Therefore, the novelty of this work lies in the implementation of a spatially explicit GNN-based framework and its comparative evaluation against standard ML approaches. In contrast to previous studies that mainly rely on high-resolution satellite imagery and focus on small-scale patterns, this research addresses seagrass distribution at a broader spatial scale. It incorporates a comprehensive set of environmental drivers, including both climatic and anthropogenic variables, enabling a more systemic assessment of pressures across the entire study area. This large-scale perspective is particularly relevant for informing conservation strategies, spatial planning, and policy development. All models demonstrated the capability to predict the state of seagrass meadows under various pressures, however, the slightly superior performance of GNNs confirms the added value of integrating spatial dependencies. These results reinforce the growing recognition of spatially-aware artificial intelligence techniques in ecological modelling and demonstrate how graph-based representations can enhance model performance and interpretability compared to conventional approaches that treat input features independently. Despite these promising outcomes, several limitations must be acknowledged. The dataset used in this study suffers from limited spatial and temporal resolution, and the graph construction was constrained to geographical proximity, without incorporating temporal dynamics or oceanographic connectivity (e.g., currents, wave energy). Additionally, the lack of long-term monitoring data restricted the capacity to model temporal changes in seagrass distribution and resilience. These constraints limit the generalisability of the findings and highlight the need for more comprehensive, high-resolution, and multidimensional datasets. Future research should aim to address these limitations by integrating temporal components and physical processes into the graph structure. The use of satellite imagery could further enhance spatial granularity and help capture fine-scale ecological dynamics. Moreover, the integration of domain knowledge through PINNs represents a promising direction for improving model

robustness and generalisability. The methodology proposed here, particularly the graph-based modelling approach, offers a transferable and flexible tool for environmental monitoring and prediction. It can be adapted to other ecological contexts where spatial relationships are essential, such as coral reef degradation, wetland health, or forest ecosystem dynamics. Ultimately, these AI-powered tools can support informed decision-making, enhance biodiversity conservation strategies, and guide the optimisation of NbS, contributing to climate resilience and the protection of critical coastal habitats.

CRediT authorship contribution statement

Angelica Bianconi: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Sebastiano Vascon:** Writing – review & editing, Supervision, Software, Methodology, Conceptualization. **Elisa Furlan:** Validation, Funding acquisition, Conceptualization. **Andrea Critto:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This article was produced while attending the PhD program in Sustainable Development And Climate Change at the University School for Advanced Studies IUSS Pavia, Cycle XXXVIII, with the support of a scholarship financed by the Ministerial Decree no. 351 of 9th April 2022, based on the NRRP - funded by the European Union - NextGenerationEU - Mission 4 "Education and Research", Component 1 "Enhancement of the offer of educational services: from nurseries to universities" - Investment 3.4 "Advanced teaching and university skills" OR Investment 4.1 "Extension of the number of research doctorates and innovative doctorates for public administration and cultural heritage". This work has been funded by the European Union's Horizon Europe programme under grant agreement No. 101112972 as part of the DesirMED project (https://www.desirmed.eu) and the European Union's Horizon 2020 research and innovation programme under Grant Agreement N. 869710 as part of MaCoBioS project (https://macobios.eu/). The research leading to these results was carried out as part of the activities of the National Biodiversity Future Center (NBFC) Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n. 3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union - NextGenerationEU, project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP F87G22000290001, and was supported by the DoE 2023–2027 (MUR, AIS.DIP.ECCELLENZA2023_27.FF project). The authors would like to express their gratitude to Hung Vuong Pham and Christian Simeoni for their assistance in searching and downloading the data for this research.

Data availability

The data and open-source code is available on https://github.com/Angebii/GNN4Seagrass.

References

Albert, J.S., Destouni, G., Duke-Sylvester, S.M., Magurran, A.E., Oberdorff, T., Reis, R.E., Winemiller, K.O., Ripple, W.J., 2021. Scientists' warning to humanity on the freshwater biodiversity crisis. Ambio 50 (1), 85–94.

Ali, A.S.A., Jazaei, F., Clement, T.P., Waldron, B., 2024. Physics-informed neural networks in groundwater flow modeling: Advantages and future directions. Groundw. Sustain. Dev. 25, 101172.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. Adv. Neural Inf. Process. Syst. 24.

Berrouet, L.M., Machado, J., Villegas-Palacio, C., 2018. Vulnerability of socio—ecological systems: A conceptual framework. Ecol. Indic. 84, 632–647.

Bertels, D., Willems, P., 2023. Physics-informed machine learning method for modelling transport of a conservative pollutant in surface water systems. J. Hydrol. 619, 129354.

Bruna, J., Zaremba, W., Szlam, A., LeCun, Y., 2013. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203.

Cachay, S.R., Erickson, E., Bucker, A.F.C., Pokropek, E., Potosnak, W., Bire, S., Osei, S., Lütjens, B., 2021. The world as a graph: Improving El Ni\˜ no forecasts with graph neural networks. arXiv preprint arXiv:2104.05089.

Campagne, C.S., Salles, J.-M., Boissery, P., Deter, J., 2015. The seagrass Posidonia oceanica: ecosystem services identification and economic evaluation of goods and benefits. Marine Poll. Bull. 97 (1–2), 391–400.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Softw. 22 (10), 1509–1518.

Campos-Taberner, M., Gilabert, M.A., Sánchez-Ruiz, S., Martínez, B., Jiménez-Guisado, A., García-Haro, F.J., Guanter, L., 2024. Global carbon fluxes using multioutput Gaussian processes regression and MODIS products. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 11310–11321.

Camps-Valls, G., Verrelst, J., Munoz-Mari, J., Laparra, V., Mateo-Jimenez, F., Gomez-Dans, J., 2016. A survey on Gaussian processes for earth-observation data analysis: A comprehensive investigation. IEEE Geosci. Remote. Sens. Mag. 4 (2), 58–78.

Carpenter, S., Byfield, V., Felgate, S.L., Price, D.M., Andrade, V., Cobb, E., Strong, J., Lichtschlag, A., Brittain, H., Barry, C., et al., 2022. Using unoccupied aerial vehicles (UAVs) to map seagrass cover from sentinel-2 imagery. Remote. Sens. 14 (3), 477.

Casal, G., Fonseca, C., Allegri, E., Bianconi, A., Boyd, E., Cornet, C.C., de Juan, S., Espinoza Córdova, F., Furlan, E., Gil, A., et al., 2025. Informing implementation of Nature-based Solutions in marine and coastal environments: the MaCoBioS Blue NBS Toolbox. Pensoft Publishers.

Castejón-Silvo, I., Terrados, J., 2012. Patterns of spatial variation of nutrient content, epiphyte load and shoot size of Posidonia oceanica seagrass meadows (Mediterranean Sea). Mar. Ecol. 33 (2), 165–175.

Catucci, E., Scardi, M., 2020. A machine learning approach to the assessment of the vulnerability of Posidonia oceanica meadows. Ecol. Indic. 108, 105744.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408, 189–215. http://dx.doi.org/10.1016/j.neucom.2019.10.118, URL: https://www.sciencedirect.com/science/article/pii/S0925231220307153.

Chefaoui, R.M., Duarte, C.M., Serrão, E.A., 2018. Dramatic loss of seagrass habitat under projected climate change in the Mediterranean Sea. Global Change Biol. 24 (10), 4919–4928.

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C.N., Corbera, J., Dailianis, T., et al., 2010. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS One 5 (8), e11842.

Cui, T., Pagendam, D., Gilfedder, M., 2021. Gaussian process machine learning and kriging for groundwater salinity interpolation. Environ. Model. Softw. 144, 105170.

Cuomo, S., De Rosa, M., Giampaolo, F., Izzo, S., Di Cola, V.S., 2023. Solving groundwater flow equation using physics-informed neural networks. Comput. Math. Appl. 145, 106–123.

Danovaro, R., Boero, F., 2019. Chapter 11 - Italian seas. In: Sheppard, C. (Ed.), World Seas: An Environmental Evaluation (Second Edition), second ed. Academic Press, pp. 283–306. http://dx.doi.org/10.1016/B978-0-12-805068-2.00060-7, URL: https://www.sciencedirect.com/science/article/pii/B9780128050682000607.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. Adv. Neural Inf. Process. Syst. 29.

Díaz, S.M., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., et al., 2019. The global assessment report on biodiversity and ecosystem services: Summary for policy makers.

do Amaral Camara Lima, M., Bergamo, T.F., Ward, R.D., Joyce, C.B., 2023. A review of seagrass ecosystem services: providing nature-based solutions for a changing world. Hydrobiologia 850 (12), 2655–2670.

Donnelly, J., Daneshkhah, A., Abolfathi, S., 2024. Forecasting global climate drivers using Gaussian processes and convolutional autoencoders. Eng. Appl. Artif. Intell. 128, 107536.

Duarte, C.M., Marbà, N., Gacia, E., Fourqurean, J.W., Beggins, J., Barrón, C., Apostolaki, E.T., 2010. Seagrass community metabolism: Assessing the carbon sink capacity of seagrass meadows. Glob. Biogeochem. Cycles 24 (4).

Dun, A., Yang, Y., Lei, F., 2022. Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction. Ecol. Inform. 70, 101736.

Elliott, M., Borja, A., McQuatters-Gollop, A., Mazik, K., Birchenough, S., Andersen, J.H., Painting, S., Peck, M., 2015. Force majeure: Will climate change affect our ability to attain good environmental status for marine biodiversity? Marine Poll. Bull. 95 (1), 7–27.

Feng, D., Tan, Z., Lin, Z., Xu, D., Yu, C.-W., He, Q., 2025. Integrating physics-informed and data-driven neural networks into earth system models: A comparative study for compound flood simulation at river-ocean interfaces. Authorea Preprints.

Forrester, J., Leonardi, N., Cooper, J.R., Kumar, P., 2024. Seagrass as a nature-based solution for coastal protection. Ecol. Eng. 206, 107316.

Fu, C., Xiong, J., Yu, F., 2024. Storm surge forecasting based on physics-informed neural networks in the Bohai Sea. In: Journal of Physics: Conference Series. Vol. 2718, IOP Publishing, 012057, 1.

Gori, M., Monfardini, G., Scarselli, F., 2005. A new model for learning in graph domains. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. Vol. 2, IEEE, pp. 729–734.

Grizzetti, B., Lanzanova, D., Liquete, C., Reynaud, A., Cardoso, A., 2016. Assessing water ecosystem services for water resource management. Environ. Sci. Policy 61, 194–203.

Guiot, J., Cramer, W., Marini, K., et al., 2020. Climate and environmental change in the Mediterranean Basin–current situation and risks for the future. First mediterranean assessment report.

Hammond, D.K., Vandergheynst, P., Gribonval, R., 2011. Wavelets on graphs via spectral graph theory. Appl. Comput. Harmon. Anal. 30 (2), 129–150. http://dx.doi.org/10.1016/j.acha.2010.04.005, URL: https://www.sciencedirect.com/science/article/pii/S1063520310000552.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. vol. 2, Springer.

Hemminga, M.A., Duarte, C.M., 2000. Seagrass ecology. Cambridge University Press.

Henaff, M., Bruna, J., LeCun, Y., 2015. Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163.

Herman, J., Usher, W., 2017. SALib: An open-source python library for sensitivity analysis. J. Open Source Softw. 2 (9), http://dx.doi.org/10.21105/joss.00097.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844.

Iwanaga, T., Usher, W., Herman, J., 2022. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. Socio-Environ. Syst. Model. 4, 18155. http://dx.doi.org/10.18174/sesmo.18155, URL: https://sesmo.org/article/view/18155.

Jiang, W., Xiao, Y., Liu, Y., Liu, Q., Li, Z., 2022. Bi-GRCN: A spatio-temporal traffic flow prediction model based on graph neural network. J. Adv. Transp. 2022.

Jiao, Z., Tao, R., 2025. Geographical Gaussian process regression: A spatial machine-learning model based on spatial similarity. Geogr. Anal..

Jordà, G., Marbà, N., Duarte, C.M., 2012. Mediterranean seagrass vulnerable to regional climate warming. Nat. Clim. Chang. 2 (11), 821–824.

Ke, Z.-W., Wei, S.-J., Yao, S.-Y., Chen, S., Chen, Y.-M., Li, Y.-C., 2025. Pre-trained physics-informed neural networks for analysis of contaminant transport in soils. Comput. Geotech. 180, 107055.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Lal, A., Datta, B., 2018. Genetic programming and gaussian process regression models for groundwater salinity prediction: Machine learning for sustainable water resources management. In: 2018 IEEE Conference on Technologies for Sustainability. SusTech, IEEE, pp. 1–7.

Lan, S., Ma, Y., Huang, W., Wang, W., Yang, H., Li, P., 2022. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In: International Conference on Machine Learning. PMLR, pp. 11906–11917.

Li, Y., Bai, J., Chen, S., Chen, B., Zhang, L., 2023. Mapping seagrasses on the basis of sentinel-2 images under tidal change. Mar. Environ. Res. 185, 105880.

Li, Y., Zhao, W., Fan, H., 2022. A spatio-temporal graph neural network approach for traffic flow prediction. Mathematics 10 (10), 1754.

Lin, H., Gao, Z., Xu, Y., Wu, L., Li, L., Li, S.Z., 2022. Conditional local convolution for spatio-temporal meteorological forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, pp. 7470–7478, 7.

Lira, H., Martí, L., Sanchez-Pi, N., 2022. A graph neural network with spatio-temporal attention for multi-sources time series data: An application to frost forecast. Sensors 22 (4), 1486.

Ma, M., Xie, P., Teng, F., Li, T., Wang, B., Ji, S., Zhang, J., 2022. HiSTGNN: Hierarchical spatio-temporal graph neural networks for weather forecasting. arXiv preprint arXiv:2201.09101.

Maes, J., Czúcz, B., Keith, H., Jackson, B., Nicholson, E., Dasoo, M., et al., 2020. A review of ecosystem condition accounts: Lessons learned and options for further development.

Mandal, S., Islam, M.S., Biswas, M.H.A., 2021a. Modeling the potential impact of climate change on living beings near coastal areas. Model. Earth Syst. Environ. 7 (3), 1783–1796.

Mandal, S., Islam, M.S., Biswas, M.H.A., Akter, S., 2021b. Modeling the optimal mitigation of potential impact of climate change on coastal ecosystems. Heliyon 7 (7).

Mandal, S., Islam, M.S., Biswas, M.H.A., Akter, S., 2022a. A mathematical model applied to investigate the potential impact of global warming on marine ecosystems. Appl. Math. Model. 101, 19–37.

Mandal, S., Islam, M.S., Biswas, M.H.A., Islam, S., 2022b. Modeling with strategies to control the adverse effects of global warming on marine ecosystems. Model. Earth Syst. Environ. 8 (3), 3073–3088.

McKenzie, L.J., Nordlund, L.M., Jones, B.L., Cullen-Unsworth, L.C., Roelfsema, C., Unsworth, R.K., 2020. The global distribution of seagrass meadows. Environ. Res. Lett. 15 (7), 074041.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Ni, Q., Wang, Y., Fang, Y., 2022. GE-STDGN: A novel spatio-temporal weather prediction model based on graph evolution. Appl. Intell. 52 (7), 7638–7652.

O'Brien, K.R., Waycott, M., Maxwell, P., Kendrick, G.A., Udy, J.W., Ferguson, A.J., Kilminster, K., Scanes, P., McKenzie, L.J., McMahon, K., et al., 2018. Seagrass ecosystem trajectory depends on the relative timescales of resistance, recovery and disturbance. Marine Poll. Bull. 134, 166–176.

O'Leary, B.C., Fonseca, C., Cornet, C.C., de Vries, M.B., Degia, A.K., Failler, P., Furlan, E., Garrabou, J., Gil, A., Hawkins, J.P., et al., 2023. Embracing nature-based solutions to promote resilient marine and coastal ecosystems. Nat.-Based Solut. 3, 100044.

Ouyang, X., Yang, Y., Zhang, Y., Zhou, W., 2021. Spatial-temporal dynamic graph convolution neural network for air quality prediction. In: 2021 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Ozkiper, O., Allegri, E., Bianconi, A., Pham, H.V., Furlan, E., Simide, R., van der Geest, M., Critto, A., 2024. A GIS-MCDA approach to map environmental suitability of Posidonia oceanica meadows as blue nature-based solutions in the Mediterranean eco-region. Sci. Total Environ. 955, 176803.

Patel, Z.B., Purohit, P., Patel, H.M., Sahni, S., Batra, N., 2022. Accurate and scalable gaussian processes for fine-grained air quality inference. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, pp. 12080–12088, 11.

Piroddi, C., Coll, M., Steenbeek, J., Moy, D.M., Christensen, V., 2015. Modelling the Mediterranean marine ecosystem as a whole: Addressing the challenge of complexity. Mar. Ecol. Prog. Ser. 533, 47–65.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561.

Rasmussen, C.E., 2003. Gaussian processes in machine learning. In: Summer School on Machine Learning. Springer, pp. 63–71.

Roca, G., Romero, J., Farina, S., Martínez-Crego, B., Alcoverro, T., 2017. Using seagrasses to identify local and large-scale trends of metals in the Mediterranean sea. Marine Poll. Bull. 123 (1–2), 83–91.

Ruiz, J.M., Guillén, J., Ramos Segura, A., Otero, M., 2015. Atlas De Las Praderas Marinas De España. Instituto Español de Oceanografía Madrid.

Salomidi, M., Katsanevakis, S., Borja, A., Braeckman, U., Galparsoro, I., Mifsud, R., Mirto, S., Pascual, M., Pipitone, C., Rabaut, M., et al., 2012. Assessment of goods and services, vulnerability, and conservation status of European seabed biotopes: A stepping stone towards ecosystem-based marine spatial management.

Shadrin, D., Nikitin, A., Tregubova, P., Terekhova, V., Jana, R., Matveev, S., Pukalchik, M., 2021. An automated approach to groundwater quality monitoring—geospatial mapping based on combined application of Gaussian process regression and Bayesian information criterion. Water 13 (4), 400.

Simeoni, C., Furlan, E., Pham, H.V., Critto, A., de Juan, S., Trégarot, E., Cornet, C.C., Meesters, E., Fonseca, C., Botelho, A.Z., et al., 2022. Evaluating the combined effect of climate and anthropogenic stressors on marine coastal ecosystems: Insights from a systematic review of cumulative impact assessment approaches. Sci. Total Environ. 160687.

Smith, D.J., Rodríguez-Labajos, B., 2021. Turning the wheel away from biophysical indicators in coastal zone management: Towards a stakeholder-based systemic framework. Ecol. Indic. 125, 107527.

Stock, A., Haupt, A.J., Mach, M.E., Micheli, F., 2018. Mapping ecological indicators of human impact with statistical and machine learning methods: Tests on the California coast. Ecol. Inform. 48, 37–47.

Sun, A.Y., Jiang, P., Mudunuru, M.K., Chen, X., 2021. Explore spatio-temporal learning of large sample hydrology using graph neural networks. Water Resour. Res. 57 (12), e2021WR030394.

Teichert, N., Borja, A., Chust, G., Uriarte, A., Lepage, M., 2016. Restoring fish ecological quality in estuaries: Implication of interactive and cumulative effects among anthropogenic stressors. Sci. Total Environ. 542, 383–393.

Telesca, L., Belluscio, A., Criscoli, A., Ardizzone, G., Apostolaki, E.T., Fraschetti, S., Gristina, M., Knittweis, L., Martin, C.S., Pergent, G., et al., 2015. Seagrass meadows (Posidonia oceanica) distribution and trajectories of change. Sci. Rep. 5 (1), 12505.

Tičina, V., Katavić, I., Grubišić, L., 2020. Marine aquaculture impacts on marine biota in oligotrophic environments of the Mediterranean sea–a review. Front. Mar. Sci. 7, 217.

Traganos, D., Aggarwal, B., Poursanidis, D., Topouzelis, K., Chrysoulakis, N., Reinartz, P., 2018. Towards global-scale seagrass mapping and monitoring using sentinel-2 on google earth engine: The case study of the Aegean and Ionian Seas. Remote. Sens. 10 (8), 1227.

Trois, C., Del Fabro, L.D., Baulin, V.A., 2024. Machine learning reveals large-scale impact of Posidonia oceanica on Mediterranean sea water. arXiv preprint arXiv:2402.14459.

UNEP, 2014. Horizon 2020 mediterranean report. 6.

UNEP, 2020. Out of the blue: The value of seagrasses to the environment and to people. p. 95.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.

Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., Gao, F., 2020. Pm2. 5-gnn: A domain knowledge enhanced graph neural network for pm2. 5 forecasting. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems. pp. 163–166.

Wang, H., Zhang, Y.-M., Mao, J.-X., 2022. Sparse Gaussian process regression for multi-step ahead forecasting of wind gusts combining numerical weather predictions and on-site measurements. J. Wind Eng. Ind. Aerodyn. 220, 104873.

Watson, R., Baste, I., Larigauderie, A., Leadley, P., Pascual, U., Baptiste, B., Demissew, S., Dziba, L., Erpul, G., Fazel, A., et al., 2019. Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat: Bonn, Germany, pp. 22–47.

Weatherdon, L., Friedrich, L., Martin, C., King, S., 2017. Experimental Seagrass Ecosystem Accounts : A Pilot Study for One Component of Marine Ecosystem Accounts. European Union, 2018. URL: www.unep-wcmc.org.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y., 2020. A comprehensive survey on graph neural networks. IEEE Trans. Neural Netw. Learn. Syst. 32 (1), 4–24.

Xu, R., Deng, X., Wan, H., Cai, Y., Pan, X., 2021. A deep learning method to repair atmospheric environmental quality data based on Gaussian diffusion. J. Clean. Prod. 308, 127446.

Yang, X., Zhang, F., Sun, P., Li, X., Du, Z., Liu, R., 2022. A spatio-temporal graph-guided convolutional LSTM for tropical cyclones precipitation nowcasting. Appl. Soft Comput. 109003.

Yu, X., Shi, S., Xu, L., 2021. A spatial–temporal graph attention network approach for air temperature forecasting. Appl. Soft Comput. 113, 107888.

Zanfei, A., Brentan, B.M., Menapace, A., Righetti, M., Herrera, M., 2022. Graph convolutional recurrent neural networks for water demand forecasting. Water Resour. Res. 58 (7), e2022WR032299.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. AI Open 1, 57–81.

Zhu, Z., Wang, Z., Dong, C., Yu, M., Xie, H., Cao, X., Han, L., Qi, J., 2025. Physics informed neural network modelling for storm surge forecasting—A case study in the Bohai Sea, China. Coast. Eng. 197, 104686.

Zhu, J.-J., Yan, B., 2022. Blue carbon sink function and carbon neutrality potential of mangroves.