

A multi-model ecosystem risk assessment methodology for the Massaciuccoli Lake basin wetland[☆]

Gian Luca Vannini^{a,b},^{*}, Gianpaolo Coro^a

^a Institute of Information Science and Technologies “A. Faedo” of the National Research Council of Italy (CNR-ISTI), Via Moruzzi 1, Pisa, 56124, Italy

^b University of Pisa, Lungarno Pacinotti, Pisa, 56126, Italy

ARTICLE INFO

Keywords:

Ecosystem risk assessment
Environmental stressors
Geospatial data
Cumulative impacts
Unsupervised learning
Anomaly detection

ABSTRACT

Wetlands are among the ecosystems most exposed to degradation due to the concurrent action of anthropogenic pressures and climate-related stressors, particularly in highly anthropised landscapes. This study presents a multi-model comparative methodology for assessing and mapping the risk of forest and grassland habitat degradation and associated biodiversity loss in wetland environments. It integrates three complementary, unsupervised models applied to a harmonised geospatial dataset: (i) a deterministic cumulative overlap index, (ii) a multivariate clustering model, and (iii) a predictive, deep learning model based on a Variational Autoencoder combined with cluster analysis. The methodology is applied to the Massaciuccoli Lake basin (Tuscany, Italy), a Ramsar-listed wetland characterised by intense land-use pressure and high ecological value. Using 12 stressor variables describing ecological fragility and harmful driving forces at approximately 50 m spatial resolution, the models produced spatially explicit risk hotspot distribution maps consistent with previous localised studies. Results also highlight complementary aspects among the models: deterministic mapping identified areas of strong stressor co-occurrence; clustering revealed structurally homogeneous ecological conditions; and deep learning captured non-linear patterns and anomalous configurations associated with emerging ecosystem risks. Areas consistently classified as medium or high risk across multiple models delineated robust priority hotspots, while divergent classifications identified transitional or uncertain conditions requiring further investigation. The proposed methodology is potentially re-applicable to other wetlands to enhance the robustness and interpretability of ecosystem risk assessments. It provides a reproducible, scalable tool to support conservation planning, environmental management, and spatial decision-making in wetlands affected by complex, interacting stressors.

1. Introduction

Wetlands are among the most complex and functionally important ecosystems, providing key services such as biodiversity support, carbon sequestration, climate regulation, water purification, and flood mitigation (Guo et al., 2017; Junk et al., 2013). Despite their ecological and socio-economic importance, they are also among the most vulnerable ecosystems, exposed to combined anthropogenic and climatic pressures, including land-use intensification, habitat fragmentation, pollution, and hydrological alterations (on Wetlands, 2025; Haddad et al., 2015). These pressures are further exacerbated by climate change, which alters temperature and precipitation regimes and increases the frequency of extreme events (IPCC, 2023; Trenberth et al., 2014). Within wetland landscapes, forest and grassland habitats play a critical role in maintaining ecological connectivity and

supporting biodiversity, yet they are particularly sensitive to these stressors (Loreau et al., 2001). Their degradation can trigger cascading effects on ecosystem functioning and resilience, especially in Mediterranean regions already characterised by water scarcity and strong anthropogenic pressure (Loreau et al., 2001; Cramer et al., 2018).

Assessing ecosystem risk in such contexts requires analytical frameworks that integrate multiple stressors across spatial and temporal scales. Traditional deterministic and expert-based approaches, while widely used, are limited by subjectivity, oversimplified assumptions, and a reduced ability to capture nonlinear relationships (Jørgensen, 2008; Schmolke et al., 2010). Recent advances in geospatial data availability and machine learning methods offer new opportunities to overcome these limitations by modelling complex interactions and heterogeneous data sources (Rolnick et al., 2022; Tuia et al., 2022).

[☆] This article is part of a Special issue entitled: ‘DTESD’ published in Ecological Informatics.

^{*} Corresponding author at: Institute of Information Science and Technologies “A. Faedo” of the National Research Council of Italy (CNR-ISTI), Via Moruzzi 1, Pisa, 56124, Italy.

E-mail address: gianluca.vannini@isti.cnr.it (G.L. Vannini).

<https://doi.org/10.1016/j.ecolinf.2026.103872>

Received 26 January 2026; Received in revised form 4 June 2026; Accepted 5 June 2026

Available online 11 June 2026

1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

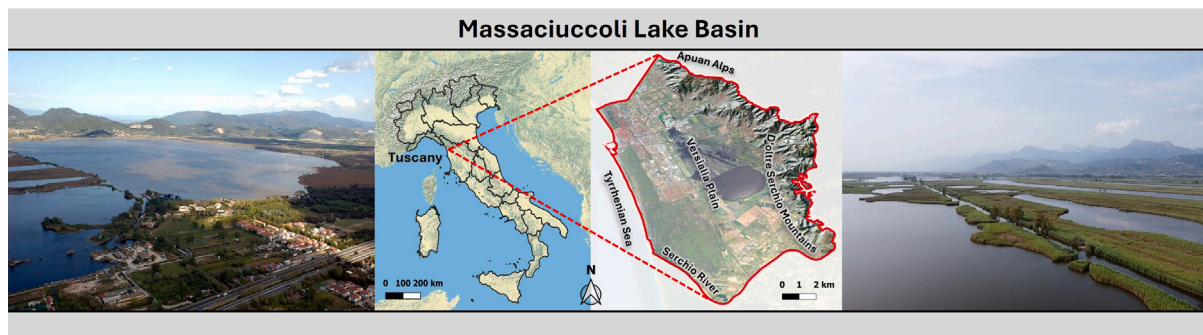


Fig. 1. General overview and geographic context of the Massaciuccoli Lake basin wetland, in Tuscany (Italy).

In this context, our study introduces a multi-model analytical framework to assess the risk of forest and grassland habitat degradation associated with biodiversity loss in wetlands. The novelty of the approach lies in the combination and explicit comparison of complementary modelling paradigms, applied jointly to the same set of harmonised geospatial data. Specifically, the framework integrates a deterministic cumulative risk index, a multivariate clustering technique, and an unsupervised deep learning model, enabling the identification of reliable risk hotspots supported by multiple models and the detection of transitional or uncertain conditions characterised by differing model responses. By systematically comparing risk models and using expert studies to assess reliability, the proposed methodology enhances the robustness of the assessments and improves their ecological interpretability, offering a reproducible, scalable tool for ecosystem risk analysis in highly anthropised wetland environments.

The methodology was first developed to assess risks of change in the ecosystem state of the Massaciuccoli Lake basin (Tuscany, Italy; Fig. 1) while remaining transferable to other wetlands. This basin is one of the most anthropogenically modified and ecologically sensitive wetlands in central Italy and is recognised as a Ramsar site of international importance (Baneschi et al., 2013; Coro et al., 2024a). It is both a biodiversity hotspot and a socio-economically relevant area, attracting tourism and scientific interest (Ciccolini et al., 2013; Nikologianni, 2019). The basin covers 114 km² and includes a shallow lake of about 13 km² (1–2.5 m depth; ~14 Mm³ volume) (Rossetto et al., 2010a). Its hydrological regime is predominantly pluvial and strongly influenced by surrounding agricultural landscapes. Human pressures are substantial, with agriculture accounting for approximately 40% of land use, alongside urban and infrastructural development (Silvestri et al., 2019). Additional impacts arise from wastewater discharges, affecting water quality and ecosystem dynamics. These conditions influence habitat suitability for both native species (e.g., black bullhead, eel, tench, largemouth bass, carp, heron, kingfisher, and others) and non-native species (e.g., sheatfish and red swamp crayfish) and highlight the ecological importance of the basin as a migratory stopover for birds (Ciccolini et al., 2013). As in many wetlands, avian communities are highly sensitive to hydrological and biogeochemical changes, making them particularly vulnerable to climate change (Bertacchi et al., 2015; Bove et al., 2026). Climate change poses significant risks to the Massaciuccoli ecosystem. Rising temperatures and increasing aridity may render habitats unsuitable for many avian species by 2100 (Coro et al., 2024a), while also altering species distributions and ecosystem services (Pignalosa et al., 2022a,b). In response, policy-makers are exploring mitigation and adaptation strategies, particularly to ensure long-term water availability (Toscana, 2018). However, these efforts require robust scientific support to anticipate ecosystem transitions and identify the most vulnerable species.

This paper makes a significant contribution to the existing body of knowledge on the ecosystem status of the Massaciuccoli Lake basin and aims to support local authorities in developing effective conservation

and mitigation strategies based on risk assessment. It is organised as follows: Section 1.1 defines ecosystem risk assessment, illustrates current approaches, presents the studies available for the Massaciuccoli Lake basin, and highlights the complementary aspects introduced by our methodology. Section 2 describes the data available for the analysed area, the methodology pipeline, and the evaluation strategy. Section 3 reports the results of qualitative and quantitative evaluations, and, finally, Section 4 summarises the main findings, limitations, and future work.

1.1. Overview

Ecosystem Risk Assessment (ERA) is a framework for identifying and spatially representing the likelihood of critical ecosystem transitions involving changes in structure, functioning, and ecosystem services due to natural and anthropogenic pressures (generally referred to as *stressors*) (De Lange et al., 2010; IPBES, 2019). Unlike impact assessments focused on single stressors, ERA addresses multiple interacting pressures across spatial and temporal scales and accounts for non-linear ecosystem responses (Halpern et al., 2008; Landis et al., 2013).

In wetlands, ERA is particularly relevant due to strong interconnections among hydrology, vegetation, biodiversity, climate, and land use (Mitsch and Gosselink, 2015; Davidson, 2018), which sustain key ecosystem functions and increase vulnerability to cumulative pressures such as land conversion, fragmentation, pollution, and hydro-climatic changes (Finlayson et al., 2019; Kingsford et al., 2016). Ecosystem risk, therefore, encompasses not only observed degradation but also conditions that increase the likelihood of future changes, including biodiversity loss and reduced resilience (Hooper et al., 2012; Cardinale et al., 2012), supporting preventive and adaptive management.

ERA methodologies vary in assumptions and objectives. Traditional approaches rely on deterministic, expert-based methods, such as multicriteria analysis and cumulative impact mapping (Malczewski, 2006; Halpern et al., 2008), valued for transparency but often limited by subjectivity and oversimplified assumptions (Linkov et al., 2006; Ascough II et al., 2008). Statistical and multivariate methods, including clustering and latent-variable analyses (Legendre and Legendre, 2012; Coro et al., 2024b), reduce subjectivity and identify recurring patterns but remain largely descriptive and limited in capturing complex non-linear relationships (Coro et al., 2025; Pavirani et al., 2025b). More recently, machine learning approaches, particularly unsupervised methods such as clustering and anomaly detection (Olden et al., 2008; Reichstein et al., 2019), and deep generative models (Doersch, 2016; An and Cho, 2015), have enabled the identification of latent structures and nonlinear dependencies in heterogeneous data. However, single-method approaches often face challenges in interpretability, comparability, and robustness (Pichler and Hartig, 2023; Ma et al., 2019).

Recent studies have emphasised the need for integrated frameworks combining cumulative, structural, and data-driven approaches to improve reliability and interpretability (Wang et al., 2021; Gong et al., 2025). In the Massaciuccoli Lake basin, ecosystem assessment studies

have focused on specific subregions to investigate (i) hydrological dynamics and water quality, including nutrient loads, hydrodynamic instability, subsidence of reclaimed peat soils, and peat oxidation processes (Pasquetti et al., 2024; Venturi et al., 2021); (ii) anthropogenic pressures related to intensive peri-lacustrine agriculture, urban and infrastructural expansion, and diffuse contamination, resulting in habitat fragmentation and loss of ecological connectivity (Prato et al., 2020; Baldaccini, 2018); (iii) biological components and habitats of community interest (Ministero dell'Ambiente e della Sicurezza Energetica, 2025), including coastal pine forests, meadow-woodland mosaics, and hygrophilous vegetation, as well as phytosanitary and microclimatic issues (Viciani et al., 2017; Bertacchi et al., 2021); and (iv) bioclimatic signals and potential impacts of climate change on species and habitats (Coro et al., 2024a; Joyce et al., 2016). However, these analyses remain sectoral and spatially limited, whereas no comprehensive, basin-scale ERA integrating multiple stressors has been conducted, partly due to the historical lack of harmonised geospatial data. Recently, this gap has been partially filled by the availability of an open-access geospatial Atlas of the Massaciuccoli basin, covering conditions from the 1950–1980 period to future climate scenarios between 2050 and 2100 (Vannini et al., 2025).

Based on this information, the present study represents the first attempt to apply a multi-model, comparative ERA methodology across the entire basin, with a specific focus on the degradation of forest and grassland habitats and the consequent loss of biodiversity.

The main novelty of the proposed methodology, compared to other ERA methods, lies in its explicit integration and systematic comparison of three analytical paradigms (deterministic, statistical, and deep learning). The deterministic *cumulative* index provides an inclusive and precautionary perspective, highlighting areas where multiple stressors co-occur and supporting exploratory and preventive planning. The clustering-based approach offers a *structural* interpretation of the risk-variables vector space, identifying homogeneous ecological conditions and clarifying how specific stressor combinations characterise different landscape portions. The deep learning component, based on a Variational Autoencoder integrated with cluster analysis, was newly designed for the present study. It introduces a *predictive* dimension, enabling the detection of non-linear patterns and anomalous configurations that may signal emerging ecosystem changes. These models were selected among baselines that demonstrated effectiveness in different contexts of anomaly detection (Kingma et al., 2014; Dilokthanakul et al., 2016), including ecosystem risk assessment and ecological modelling in marine and terrestrial areas (Coro, 2024; Lou et al., 2025). In particular, the work in Pavirani et al. (2025a) presented a comprehensive comparison of the pros and cons of the models mentioned and provides guidance on selecting the most suitable model for specific assessment purposes (general vs localised). Following the advice from this work, we reused the Multi K-means approach, extended the Variational Autoencoder model by combining it with cluster analysis to refine the assessments, and introduced a cumulative analysis representing a traditional approach to risk assessment (Section 2.2).

Rather than selecting a single optimal model, the presented methodology analyses their convergence and divergence to produce a more robust, hierarchical representation of ecosystem risk. Areas consistently classified as high risk across models can be interpreted as priority hotspots, while divergent classifications highlight transitional or uncertain conditions. This multimodal strategy enhances interpretability, reduces methodological bias, and provides a scalable and reproducible framework applicable to other wetland systems.

2. Methods

The assessment of the Massaciuccoli Lake basin's ecosystem risks related to forest and grassland habitat degradation and biodiversity loss was conducted using a multi-phase methodological framework (Fig. 2) designed to ensure analytical robustness, transparency, and

reproducibility. The approach integrates deterministic, clustering, statistical, and deep learning strategies, supported by an interpretative procedure for detected clusters. Section 2.1 reports the details of the data collection and preparation phase, whereas Section 2.2 explains the modelling methodology. The code implementing this workflow was developed in R and is openly available, together with the data, on Zenodo and GitHub for reproducibility, repeatability, and reuse (Software and data availability).

2.1. Data

Data collection and preparation was composed by the following phases, detailed in the next sections: (i) spatial and temporal characterisation of the study area, through environmental, climatic, biological, and anthropogenic geospatial data; (ii) definition and identification of ecological stressor variables, to distinguish ecosystem fragility conditions from harmful driving forces; and (iii) data preparation and standardisation, which includes variable reprocessing, spatial alignment, and transformation into a unified dataset suitable for multivariate analysis.

2.1.1. Spatio-temporal description of the study area

To obtain a comprehensive and coherent environmental characterisation of the study area, a publicly available cartographic dataset on Zenodo was used (Vannini and Coro, 2025b; Vannini et al., 2025). Its content is described in a geographic Atlas of the Massaciuccoli Lake Basin, which provides a comprehensive description of the environmental variables and active driving forces acting in the ecosystem (Vannini and Coro, 2025a). The methodology that led to the realisation of this repository was based on FAIR (Findable, Accessible, Interoperable, Reusable) principles and a semi-automated workflow for data collection and harmonisation.

The data collection constitutes a spatially explicit environmental knowledge base, organised to support both exploratory and analytical workflows. It includes raster and vector datasets accompanied by standardised metadata, a harmonised coordinate reference system (EPSG 4326 - WGS84), and a consistent spatial resolution. Moreover, it provides an integrated spatial and temporal representation of the Massaciuccoli Lake basin, describing its climatic, environmental, pedological, anthropogenic, and biological features, with particular emphasis on biodiversity. Overall, this data collection includes 75 geo-variables at a spatial resolution of $0.0005^\circ \times 0.0005^\circ$ (approximately 50 m $\times$ 50 m), distributed across 148 datasets and organised into five temporal snapshots: remote past (1950–1980), recent past (1981–2015), present (2016–today), near future (today–2050), and far future (2051–2100). Future projections are provided under the Representative Concentration Pathways (RCP) 2.6, 4.5, and 8.5.

When aligned on a common spatial grid, the database comprises a total of 77,346 georeferenced spatial cells covering the study area. This spatio-temporal structure enables multivariate and multi-temporal analyses and supports the investigation of long-term environmental processes, ecosystem dynamics, and their responses to climatic and anthropogenic drivers within the Massaciuccoli basin.

2.1.2. Definition and identification of ecological stressors

The definition of ecological stressors was based on the specific characteristics of the Massaciuccoli ecosystem. Stressors are driving forces within the ecosystem that can influence changes in the ecosystem status, causing a critical transition (Coro et al., 2023; Pavirani et al., 2025b). Numerically, a stressor is a spatio-temporal variable whose *high* values influence the ecosystem status (Coro et al., 2025; Pavirani et al., 2025a). A stressor variable is defined on a collection of data, like that of the Massaciuccoli basin Atlas (Section 2.1.1). In our case, a stressor variable could directly correspond to a variable represented in one of the geospatial datasets (e.g., an increase in land imperviousness over the past decade) or to the inverse of a variable (e.g., reduced precipitation as the inverse of the increase in precipitation). Overall, the variables could be divided into two main categories:

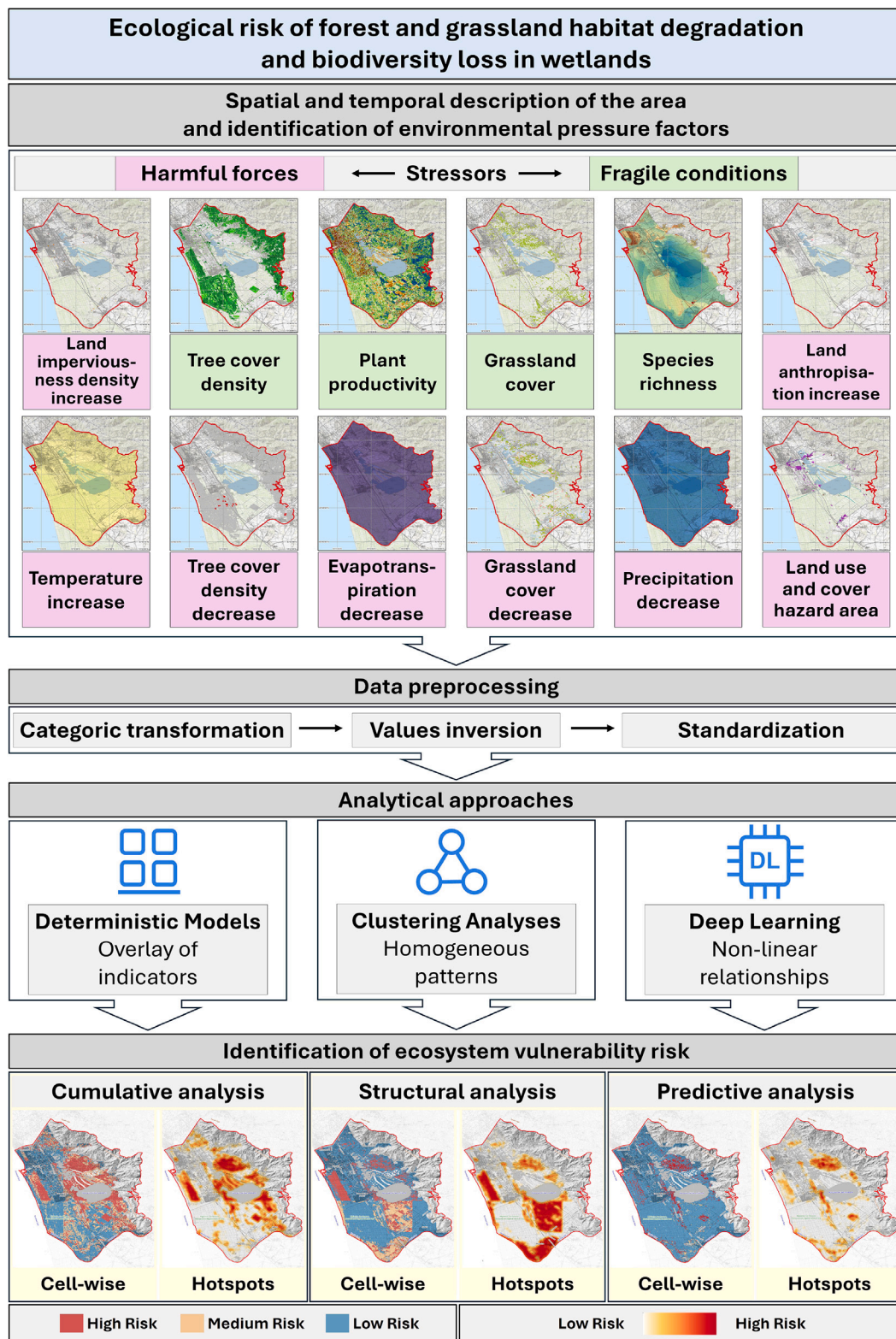


Fig. 2. Graphical representation of the developed methodology.

1. *Fragile conditions*: variables whose high values were associated with the sensitivity and resilience of the ecosystem (Fennessy et al., 2007; Scheffer, 2009);
2. *Harmful forces*: variables representing direct drivers with negative impact on habitats, potentially causing structural and functional changes (Foley et al., 2005; Malekmohammadi and Blouchi, 2014).

Although the original datasets were characterised by heterogeneous temporal references (single-year snapshots, short-term change periods, or multi-year reference windows), the stress variables used for this study were selected from levels explicitly marked as “present” (2016–today) in the Atlas. This time frame referred to recent and current ecosystem conditions, rather than distinct historical phases. Variables aggregated over short time periods in the original dataset

(e.g., 2015–2018 or 2020–2025) synthesised recent changes or multi-year baseline conditions that have led to the current state of habitats and biodiversity. The time frame of the analysis can therefore be interpreted as representative of the prevailing ecosystem conditions during 2016–2025, ensuring temporal comparability among stressor variables.

The restriction to this time frame was motivated by the need to ensure maximum consistency and alignment across the selected stressor variables, as this snapshot provides the most complete and harmonised representation of the ecosystem conditions within the Atlas. Focusing on a single time frame also allowed us to isolate and evaluate the performance of the proposed multi-model risk assessment methodology without introducing additional variability related to temporal dynamics. While the Atlas includes multi-temporal data, extending the analysis across multiple time periods would require dedicated methodological adaptations and evaluation strategies beyond the scope of the present study. The main objective of the present study was thus to identify from this dataset the areas most exposed to forest and grassland habitat degradation and related biodiversity loss; therefore, a target set of 12 related variables (Table 1) was defined based on the Massaciuccoli Atlas, to characterise ecological fragility and degradation processes.

2.1.3. Data preparation and standardisation

The variables defined in the previous step were extracted from the Atlas, reprocessed in the case of inverted variables, and aligned on the same ~50 m grid. Table 1 summarises the reprocessing operations applied to each stressor variable. A gridded dataset was prepared as a table file in comma-separated values (CSV) format to enable multi-variate analyses. Each row of this dataset represented a georeferenced spatial cell (or raster pixel, measuring $0.0005^\circ \times 0.0005^\circ$) within the study area. The columns corresponded to the different stressor variables. To ensure data consistency and reliability, rows were retained only if all variables had non-zero values. This aggregation and cleaning operation yielded 49,382 cells spanning the Massaciuccoli basin.

To enhance comparability among stressor variables and increase the robustness of subsequent analyses, the dataset was further reprocessed through categorical and numerical transformations. Categorical variables were converted into numerical form using either: (i) a normalised intensity scale ranging from 0 to 1 for variables expressing proportional information (“Grassland cover”, “Land use and cover Natura 2000 hazard areas”, and “Land use and cover anthropisation increase”), or (ii) a Boolean encoding (1 = presence, 0 = absence) for binary variables (“Tree cover density decrease” and “Grassland cover decrease”). Finally, unavailable data for numerical variables were reinterpreted when explicit values were unavailable (e.g., as reported for “Land imperviousness density increase”).

All variables were subsequently z-standardised to have a mean of 0 and a standard deviation of 1. This operation allowed direct comparison of high and low values across the variables. The purpose was to place the stressor variables on a common numerical scale, while preventing variables with larger measurement ranges from dominating the analyses. This procedure did not alter the ecological importance of the stressor variables, as the relative ecological relevance of each variable was determined during the stressor-definition phase, where variables were selected, transformed, and, when necessary, inverted to represent meaningful ecological pressures or fragile conditions. Although the methodology allows variable-specific weighting when supported by expert knowledge or ecological evidence, all stressors were intentionally assigned equal contribution to the risk assessment because no prior ecological evidence was available to justify differential weighting. This choice also helped maintain a data-driven, assumption-minimising assessment.

The result of this phase was a *stressor-variable table file*, ready for analysis by the risk assessment models.

2.2. Risk assessment

Our ecosystem risk assessment methodology to identify the areas most exposed to forest and grassland habitat degradation and related biodiversity loss, consisted of three main processing steps:

1. Step 1: *Risk assessment* through three complementary models;
2. Step 2: *Risk assessment refinement* to better identify conditions associated with ecosystem-change risks;
3. Step 3: *Ecosystem-risk hotspot identification* to highlight spatialised risk hotspots.

This methodology was designed to be independent of the specific study area, as it operated on the statistical structure of georeferenced stressor variables rather than on site-specific parameters. Consequently, it could be applied to other ecosystems, provided that a harmonised dataset of stressor variables describing ecological fragility and anthropogenic or climatic pressures were available. This design choice aligned with the objective of developing a reusable and scalable risk assessment framework.

The next sections provide detailed descriptions of these steps. These sections will explicitly distinguish three related but sequential concepts that should not be interpreted as equivalent and will be used throughout the methodology, i.e., (i) *attention levels* (lower, medium, closer), derived from clustering analysis to provide a statistical characterisation of stressor configurations within the data; (ii) *anomalies*, identified through unsupervised deep learning as observations with lower reconstruction probability, representing deviations from the dominant data distributions — used to filter and focus subsequent attention level classification; and (iii) *ecosystem risk levels* (low, medium, high), representing the final ecological interpretations, obtained through a refinement procedure that integrates biodiversity-related variables and attention scores.

2.2.1. Step 1 – Multi-model risk assessment

Three unsupervised, complementary models for ecosystem risk assessment were used to analyse the prepared data, as detailed below.

Cumulative analysis. The first model corresponded to a cumulative overlap index for the stressor variables. This overlap index represented a basic and common approach to risk-area identification. It simulated an intuitive analytical approach by an expert who would overlay the spatial distributions of the stressor variables to identify areas where the highest overlap among the highest values of all variables occurs.

The overlap index measured the degree of overlap between high values in all variables. It was obtained by first calculating the sum of all stressor variables (after standardisation) and then studying the statistical distribution of this sum. The quartiles of the distribution indicated regions of high, medium, or low stressor concurrency, based on the following rules:

- First and second quartiles (<50th percentile): *low risk*
- Third quartile (50th–75th percentile): *medium risk*
- Fourth quartile (>75th percentile): *high risk*

This model classified overall 12,346 cells as high risk, 12,345 as medium risk and 24,691 as low risk. The cells associated with a high risk included high values for most stressor variables, i.e., a high overlap between strong harmful forces and peculiarly fragile conditions. Consequently, they could be associated with a high risk of potential ecosystem change due to this cumulation. Therefore, the model's output represented a *cumulative analysis* of ecosystem risk.

As output, the model enriched the *stressor-variable table* with an additional column reporting the overlap index classification for each row (i.e., each cell-wise stressor-variable vector).

Although powerful and interpretable, this approach had two main limitations: the exclusion of interdependencies among variables and the

Table 1

List of the stressor variables used in our methodology. For each variable, the table reports the numeric identifier in the Massaciuccoli Atlas, the stressor name, a general description, the data type (categorical or numeric), our interpretation as a fragile condition or harmful force, whether the variable in the Atlas was used directly or inversely to define the stressor, the unit of measure, and the transformations used for mapping values onto numerical features. All variables represent recent ecosystem conditions (2016–today), derived from single-year datasets, short-term change periods, or aggregated multi-year windows, as described in the Atlas.

Id (Atlas ref.)	Stressor name	Description	(C)ategorical (N)umeric	(F)ragile condition (H)armful Force	(I)nverse (D)irect	Unit of measure	Value mapping
3	Land imperviousness density increase	Change in degree of land imperviousness, from -100% to +100%.	N	H	D	%	Non-sealed in both years (stable non-built up)→100% Unclassifiable (no satellite image available or clouds or shadows or snow)→NA Outside area→NA Others→as is
9	Tree cover density	Level of tree cover density, from 0% to 100%.	N	F	D	%	As is
10	Tree cover density decrease	Changes in tree cover density across thematic classes (unchanged areas with no tree cover / new tree cover / loss of tree cover / unchanged areas with tree cover).	C	H	I	classes	Unchanged areas with no tree cover →0 New tree cover →-1 Loss of tree cover →-1 Unchanged areas with tree cover →0 Others→0
11	Grassland cover	Classification of grassland cover with two thematic classes (grassland / non-grassland).	C	F	D	classes	Grassland→1 0→0
12	Grassland cover decrease	Changes in grassland vegetation cover across thematic classes (all non-grassland areas, grassland gain and grassland loss, unchanged grassland in both years and unverified grassland gain and loss areas).	C	H	I	classes	Unchanged non-grassland in both years→0 Grassland gain→-1 Grassland loss→-1 Unchanged grassland in both years→0 Unverified grassland gain→-1 Unverified grassland loss→1 Others→0
17	Land use and cover (Natura 2000) hazard area	Land cover and land use in Natura 2000 sites. Only classes with the greatest potential impact on natural ecosystems were selected (i.e., industrial, commercial and military units, mining activities, landfills and construction sites, heavily modified watercourses and canals).	C	H	D	classes	Industrial→1 Commercial and military units→1 Mineral extraction→1 Dump and construction sites→1 Highly modified water courses and canals→1 Others→0
40	Plant productivity	Total productivity as the sum of the daily values of the vegetation phenological index during the growing season. It provides an annual estimate of accumulated plant biomass and overall ecosystem productivity.	N	F	D	range	As is
57	Temperature increase	Absolute change in air temperature values (approximately 2 m above the Earth's surface) relative to the average across 1971–2010.	N	H	D	°C	As is
58	Potential evapotranspiration decrease	Relative change in potential evapotranspiration values compared to the average across 1971–2010.	N	H	I	%	As is
59	Precipitation decrease	Relative change in precipitation values per unit area compared to the average across 1971–2010.	N	H	I	%	As is
65	Species richness	Biodiversity index based on cumulative species richness, including birds, fish and crustaceans, insects and clams, mammals, reptiles and amphibians, and other groups.	N	F	D	number of species	As is
147	Land use and cover anthropisation increase	Changes in land use and vegetation cover. Only classes with the greatest potential impact on natural ecosystems were selected (i.e., artificially shaped territories, utilised agricultural areas, Water bodies).	C	H	D	classes	Artificially shaped territories→1 Utilised agricultural areas→0.5 Water bodies→-1 Others→0

likely misestimation of higher-risk areas. In fact, the high-risk conditions identified by this model required the simultaneous presence of high values for most variables, whereas actual risk was most likely due to combinations and interdependencies between variable subsets (Pavirani et al., 2025a). These limitations justified our search for alternative models that could substitute or complement the cumulative analysis.

Structural analysis. To account for interdependencies among the stressor variables in identifying ecosystem-risk conditions, our methodology used a Multi K-means model. This model, in a variant introduced in Coro et al. (2025), is an extended version of the K-Means algorithm that automatically estimates the optimal number of clusters (K^*) by analysing different partitions of the stressor-variable space. The centroids, i.e., the vectors representing the averages of the variables within each cluster, were randomly initialised and updated iteratively until convergence, using the Euclidean distance as the similarity metric to estimate variable co-occurrence. The quality of clustering was assessed through the $UNIF(K)$ function, which compared the observed distribution of vectors with a theoretical uniform distribution by a χ^2 (Chi-square) test. The optimal solution was selected as the one with the best fit, indicative of more homogeneous and statistically robust clusters.

As an additional processing, our model conducted a statistical analysis on the model's output. For each identified cluster, the centroid vector values were compared with the statistical distributions of the variables across the entire dataset. In particular, each stressor variable value of a centroid was classified as follows:

- H (closer attention), if the value was higher than the 75th percentile
- M (medium attention), if the value fell between the 50th and 75th percentile
- L (lower attention), if the value was lower than the 50th percentile

An overall *attention level* was subsequently assigned to each cluster centroid, and consequently to the entire cluster, by counting the variables labelled as H, M, and L:

- Closer attention level (dominance of H variables)
- Medium attention level (dominance of M variables)
- Lower attention level (dominance of L variables)

Multi K-means identified overall 21 clusters ($K^* = 21$) in the Masaciucoli dataset, of which 15 (including 38,632 vectors, overall) were associated with closer attention, 5 (including 8340 vectors, overall) with medium attention, and 1 (including 2410 vectors, overall) with lower attention.

Notably, the assignment of attention levels was not intended to model ecological interactions among stressors. These interactions were approximated during the clustering phase by analysing multivariate similarities among complete stressor-variable vectors through Euclidean distance. The centroid interpretation procedure served instead as a transparent and reproducible mechanism for summarising the dominant characteristics of each cluster and facilitating ecological interpretation. In this sense, the *attention levels* did not replace the multivariate information used to generate the clusters, but provided a synthetic description of the stressor configurations represented by their centroids. For example, one cluster classified as deserving *closer* attention was characterised by high values of land imperviousness density increase, grassland cover, grassland cover loss, plant productivity, temperature increase, and land-use anthropisation, indicating the coexistence of productive ecosystems with strong anthropogenic and climatic pressures. Conversely, a cluster classified as deserving *lower* attention exhibited low values of grassland loss and plant productivity, low species richness, and only minor changes in temperature and potential evapotranspiration, representing comparatively stable conditions with

limited evidence of active degradation pressures. These examples illustrate how the centroid-based interpretation complements, rather than substitutes for, the multivariate structure identified by the clustering algorithm.

Overall, this model extrapolated a structure for the vector space of the stressor variables; therefore, it performed a *structural* analysis of the dataset, a term that will hereafter refer to this model. As the output, the model generated an updated *risk assessment* table containing an additional column that reported, for each stressor-variable vector, the attention level of the belonging cluster.

Predictive analysis. The third ecosystem risk assessment model addressed the modelling of non-linear relationships between the stressor variables. The Euclidean distance used by the Multi K-means model could detect vector similarities but could not capture the complex relations among the variables that led to the observed data. One way to model these relations was to infer the latent variables that generated the observed stressor-variable combinations (Coro et al., 2026).

A Variational Autoencoder (VAE) (Kingma and Welling, 2022) is a deep, unsupervised probabilistic neural network that learns a latent representation $\bar{z}$ underlying the observed data $\bar{x}$. It models the data distribution $p(\bar{x})$ through a lower-dimensional latent space, capturing the generative structure of the input vectors. The VAE consists of two neural components: (1) an *Encoder*, which maps each input vector to the parameters (mean and standard deviation) of a multivariate Gaussian distribution over the latent variables, thereby approximating the posterior distribution $q(\bar{z}|\bar{x})$; and (2) a *Decoder*, which reconstructs the input vector $\bar{x}$ from a sample drawn from this latent distribution ($\bar{z} \sim q(\bar{z}|\bar{x})$). The Decoder also estimates the reconstruction likelihood $p(\bar{x}|\bar{z})$, which can be used to identify anomalous observations as those with low reconstruction quality (Zhang and Peng, 2018; Liu et al., 2021). During training, the model parameters are optimised to maximise reconstruction likelihood by minimising a loss function composed of two terms: (i) the *reconstruction loss*, which measures how accurately the Decoder reproduces the input (e.g., via mean squared error or cross-entropy), and (ii) the *Kullback–Leibler divergence*, which regularises the learned posterior $q(\bar{z}|\bar{x})$ towards a standard multivariate Gaussian prior with $\bar{\mu} = (0, \dots, 0)$ and $\bar{\sigma} = (1, \dots, 1)$.

In our study, the optimal architecture was automatically determined during training using a “growing” strategy, in which nodes were incrementally added to the hidden layers, and the configuration achieving the highest total reconstruction probability of the input data was selected. For simplicity and consistency with previous studies (Pavirani et al., 2025a; Bove et al., 2026), the Encoder and Decoder architectures were designed to be symmetrical. The dimension of $\bar{z}$ was varied from 1 to $\dim(\bar{x})$ (=12) to identify possible compressed and less noisy representations of the input data. More specifically, for each candidate latent-space dimension, the hidden architecture was progressively expanded from a minimal configuration of one hidden layer with one neuron. Neurons were incrementally added until either no further substantial increase in total reconstruction probability was observed (increase <1%) or the reconstruction probability decreased. Once the optimal number of neurons for a layer was identified, an additional hidden layer was introduced and evaluated using the same procedure. The architecture search stopped when adding further neurons or hidden layers did not produce a substantial improvement in reconstruction probability. Each candidate architecture was trained 10 times independently with different random initialisations, and the architecture yielding the highest average reconstruction probability was retained. The selected architecture was found to be stable across repetitions, with only minor variations in reconstruction probability and negligible differences in the resulting risk classifications. For the Masaciucoli dataset, the optimal Encoder architecture consisted of 10 neurons in the first hidden layer and 5 neurons in the second, while $\bar{z}$ had the same dimensionality as $\bar{x}$ (12). The Decoder mirrored this structure, with hidden layers of 5 and 10 neurons and a final reconstruction layer of 12 neurons.

A threshold at the 50th percentile of the reconstruction probability distribution across the data was adopted to identify the subset of vectors with lower reconstruction probability. These vectors were referred to as “anomalous” in a relative sense, i.e., as observations that deviated more from the dominant data distribution learned by the VAE, rather than as rare or extreme outliers in the classical sense of anomaly detection. The choice of the 50th percentile was motivated by the absence of prior knowledge on the proportion of anomalous conditions in the ecosystem. Instead of assuming a fixed or small anomaly rate, the VAE was used as a preliminary filtering mechanism to rank the data by reconstruction probability and to reduce the vector space size for subsequent analysis. This threshold represented a balanced trade-off between excluding potentially relevant anomalous configurations (if too restrictive) and including an excessive number of well-reconstructed, typical observations (if too permissive). Previous studies have highlighted a strong correspondence between this type of anomalous stressor variable vectors and higher ecosystem risk (Pavirani et al., 2025a; Coro et al., 2026). In this framework, the VAE did not provide a final anomaly classification but rather served as a complexity-aware filter, leveraging its ability to model non-linear relationships among stressor variables to isolate less typical ecosystem configurations. The actual discrimination of ecosystem risk was subsequently refined through structural analysis applied only to this subset.

In our dataset, the VAE filter selected 24,674 vectors. The additional structural analysis was applied to this subset to characterise different types of anomalies. In fact, the VAE distinguished only between relatively well- and poorly-reconstructed observations, the clustering step enabled a more nuanced interpretation of anomaly patterns and the identification of those deserving closer ecological attention. Attention levels from the structural analysis were thus extracted only for the subset of the original dataset vectors that the VAE identified as anomalous. From the 24,674 anomalous vectors, the structural analysis identified 27 clusters ($K^* = 27$): 7 clusters (including 7855 vectors, overall) were associated with closer attention, 12 (including 9612 vectors, overall) with medium attention, and 8 (including 7207 vectors, overall) with lower attention.

The output of this modelling phase consisted of an updated *risk assessment* table with an additional column reporting the attention levels classified by this model.

One important difference between this model and the mere structural analysis is that the latent variables learned by the VAE can be used to generate new data beyond the training dataset. This feature could be used to hypothesise future anomalies and, consequently, project new scenarios of ecosystem change. In this view, this system can be interpreted (and referred to) as a *predictive analysis*.

Notably, the 50th percentile threshold over the reconstruction probability distribution was intended to provide the later structural analysis with suitable and stable information for assessing attention levels. Lower thresholds (e.g., the 25th percentile) yielded a smaller subset of anomalous vectors and, consequently, reduced the abundance of the clusters identified in the subsequent structural analysis. However, the clusters deserving closer ecological attention remained broadly consistent, with similar centroid profiles and spatial distributions. The main effect of the lower threshold was a reduction in cluster size, which increased variability across repeated executions and reduced the robustness of the subsequent statistical refinement procedure. Conversely, higher thresholds increased the number of retained vectors and progressively diluted the distinction between anomalous and typical observations. Under these conditions, the predictive analysis became increasingly similar to the structural analysis applied to the full dataset, reducing the added value of the VAE-based filtering stage. Overall, the 50th-percentile threshold provided the best compromise between retaining sufficiently representative anomalous configurations for stable clustering and preserving a meaningful distinction from the structural analysis.

2.2.2. Step 2 – Risk assessment refinement

In marine ecosystem–risk assessments, stressor variables are typically dominated by fishing pressure and biodiversity indicators (Coro et al., 2025; Pavirani et al., 2025b). These variables are generally well-mapped, spatially dense, and tightly coupled, enabling a relatively direct integration of human-induced pressures and ecological responses (Halpern et al., 2015; Korpinen and Andersen, 2016). Therefore, the spatial overlap between these pressures and biodiversity patterns facilitates the identification of risks and the attribution of ecosystem impacts (Halpern et al., 2008). In terrestrial systems, however, the configuration of stressors can be more heterogeneous. Datasets, such as those from the Massaciucoli Atlas, include diverse anthropogenic and climatic drivers — such as urbanisation, increasing impervious surface area, land-use change, soil degradation, or rising aridity — whose spatial patterns may not align neatly with ecosystem processes. In many cases, these stressors intersect primarily through human-driven dynamics rather than through direct evidence of ecological change (Foley et al., 2005; Nelson et al., 2006). Consequently, ecosystem-risk assessments should identify areas with a complex interplay among anthropogenic pressures, climatic variability, and biological components. These areas should encompass intricate interactions among plant and animal communities, habitat structures, and ecological functions, which cannot be inferred solely from physical or human-induced stressors. Our ecosystem risk evaluation, therefore, had to identify areas where biodiversity and ecological indicators adequately captured potential impacts from human activities and climate-related drivers on the ecosystem. To this aim, a refinement procedure was developed to optimise the selection of the closer-attention clusters generated by the structural and predictive analyses. This procedure aimed to identify the clusters that exhibited stronger association with risks of ecosystem change, based on statistical properties of the clusters. The procedure consisted of the following steps:

1. The clusters classified as deserving *closer attention* were extracted along with their associated vectors (*selected clusters*).
2. For each stressor variable, the average across all *selected clusters* ($\bar{X}$) and the related standard deviation $sd_{\bar{X}}$ were calculated.
3. Each element in a *selected-cluster* centroid vector ($x_{i|c_j}$) was compared to $\bar{X}$ to produce a *positive deviation score* (D) through the following rules:
 - If $x_{i|c_j} > \bar{X} + sd_{\bar{X}} \rightarrow D = 1$ (highly significant deviation)
 - Else, if $\bar{X} + (sd_{\bar{X}}/2) \leq x_{i|c_j} \leq \bar{X} + sd_{\bar{X}} \rightarrow D = 0.5$ (significant deviation)
 - Else - all other cases - if $x_{i|c_j} < \bar{X} + (sd_{\bar{X}}/2) \rightarrow D = 0$ (non significant deviation)
4. Among the *selected clusters*, an additional selection was conducted to identify the clusters with at least a significant deviation ($D > 0$) for fragile-condition related to biodiversity (*species richness*), forest presence (*tree cover density*), and grassland presence (*grassland cover*).
5. For each biodiversity-related *selected cluster*, an overall ecosystem risk index was calculated by summing and classifying the D-scores of all centroid variables according to the following rules:
 - If $\sum D > \text{number of features}/2 \rightarrow \text{high ecosystem risk}$
 - Else, if $\sum D > 0 \rightarrow \text{medium ecosystem risk}$
 - Else $\rightarrow \text{low ecosystem risk}$

Notably, this classification assigned the highest risk to clusters that involved very high values ($D = 1$) for most variables or at least significant values ($D = 0.5$) for all variables.

Overall, the presented risk assessment refinement process identified clusters indicating strong concurrence between anthropogenic and/or

climatic stressors on biodiversity. As a result, the process added a new column to the *risk assessment* table that, for each row vector, contained the ecosystem risk re-classification of the corresponding cluster, categorised as *low*, *medium*, or *high*. This table served as the basis of the cell-wise analysis presented in the evaluation (Section 2.3).

2.2.3. Step 3 - Ecosystem-risk hotspots

To identify hotspots of ecosystem risks, the cell-wise assessments from the previous step were reprocessed using Kernel Density Estimation (KDE) (Rosenblatt, 1956). The coordinates of the cells associated with the vectors in the *medium* and *high* clusters were processed through a quartic Kernel. The interpolation radius was set to the estimated spatial autocorrelation of the at-risk cells (0.003° , ~ 300 m). As the output, the KDE produced a density map (*heatmap*) of the ecosystem-risk cells, with higher-density areas indicating greater risk hotspots. This output served as the basis for the hotspot analysis presented in the evaluation (Section 2.3), which used the ecosystem risk hotspots after excluding values below the 25th percentile to improve heatmap visualisation and at-risk area identification. In particular, values above the 25th percentile were used to assess similarities and discrepancies between the maps produced by the three models.

2.3. Evaluation methodology

The following sections describe the evaluation methodology for the three models. It was a two-step approach: the first step was a *qualitative* assessment of the results that displayed and characterised the principal hotspots of ecosystem risks. It offered ecological interpretations of the major hotspots identified by the models. This analysis aimed to identify the likely reasons for ecosystem risk arising from the interactions among the stressor variables. Additionally, this evaluation compared the identified spatial risk hotspots and their associated explanations with the risks reported by localised agronomic and ecological studies pertinent to those regions.

Following this, a *quantitative* evaluation aimed to highlight the differences and similarities among the maps produced by the models. It was conducted using a *pairwise* analysis to measure agreement between the risk assessment maps generated by the models. This analysis was conducted separately for cell-wise maps and hotspot maps. The agreement was always calculated on a ~ 50 m grid, the spatial resolution of the maps. Calculation was based on binarised maps. For cell-wise comparisons, a value of 1 was assigned to cells corresponding to medium- and high-risk clusters (*at-risk cells*), and 0 to those belonging to low-risk clusters (*non-at-risk cells*). Similarly, for hotspot comparison, a value of 1 was assigned to grid points with KDE values exceeding the 20th percentile of the distribution, and 0 to those with lower KDE values. For each map pair, we assumed one risk distribution (A) as the *reference* distribution and the other (B) as the evaluated *prediction*. We defined similarity as the proportion of cells (or grid points, equivalently) categorised as at-risk or non-at-risk that were jointly assessed by the models. Consequently, we calculated the following metrics:

- True positives (TP) as the ratio of jointly assessed risk cells: $TP = \frac{|A=1 \text{ and } B=1|}{|A|}$, with $|A| = |B| = \text{total number of cells}$
- True negatives (TN) as the ratio of jointly assessed non-at-risk cells: $TN = \frac{|A=0 \text{ and } B=0|}{|A|}$
- False negatives (FN) as fraction of cells where A indicated risk, whereas B indicated non-risk: $FN = \frac{|A=1 \text{ and } B=0|}{|A|}$
- False positives (FP) as fraction of cells where A indicated non-risk, whereas B indicated risk: $FP = \frac{|A=0 \text{ and } B=1|}{|A|}$
- Accuracy as $A = \frac{TP+TN}{(TP+TN+FP+FN)}$
- Agreement through Cohen's Kappa (Cohen, 1960), which compared accuracy with chance agreement (P_e): $\kappa = \frac{\text{Accuracy} - P_e}{1 - P_e}$ with $P_e = \frac{(TP+FP)(TP+FN)+(TN+FN)(TN+FP)}{(TP+TN+FP+FN)^2}$

- Precision as the proportion of cells jointly classified as at-risk compared to the total number of cells classified by B as at-risk: $P = \frac{TP}{TP+FP}$
- Recall as the proportion of cells jointly classified as at-risk with respect to those at risk for A: $R = \frac{TP}{TP+FN}$
- F1-score as the harmonic average between Precision and Recall, which summarised the overall assessment performance of B with respect to A: $F_1 = \frac{2 \cdot P \cdot R}{P+R}$

As a further quantitative evaluation, a *simultaneous* comparison was conducted, in which the three maps were overlapped to highlight patterns of *medium* or *high* ecosystem risk that were collectively identified, treating the three models as a unified assessment. In this case, we calculated the following metrics:

- *All medium/high risk*: percentage of cells classified as medium or high risk by all three models compared.
- *One non-medium/high risk*: percentage of cells classified as low risk by one model but as medium or high risk by the other two.
- *All non-medium/high risk*: percentage of cells classified as low risk by all three models compared.
- *One medium/high risk*: percentage of cells classified as medium or high risk by one model but as low risk by the other two.

The *pairwise* and *simultaneous* analyses therefore provided an overview of the similarities, differences, and complementary aspects among the models, as well as the advantages of combining them.

Notably, the evaluation framework adopted in this study does not constitute a formal validation against independent ground-truth data. In ecosystem risk assessment, risk refers to the likelihood of potential ecosystem transitions rather than directly observable degradation states, and comprehensive, spatially explicit reference datasets are typically unavailable. For the Massaciuccoli basin, existing studies are largely sectoral, focusing on specific components such as hydrology, biodiversity, agriculture, or pollution, and therefore do not provide a unified benchmark for ecosystem-wide risk. Accordingly, the proposed evaluation strategy combines a *qualitative* comparison with independent domain-specific studies to verify the plausibility of the identified risk patterns, and a *quantitative* analysis to assess the consistency and complementarity of the models. This approach, therefore, evaluated the *external* alignment of the methodology with established ecological knowledge and its *internal* coherence. The results should thus be interpreted as a plausibility-based assessment rather than a strict empirical validation.

3. Results

This section reports the results of the qualitative and quantitative evaluations of the models described in the previous section.

3.1. Qualitative results

A characterisation of the areas at risk as indicated by each model is provided below. Subsequently, a comparison with other scientific studies is presented to evaluate the consistency of the findings.

Cumulative analysis. A total of 12,346 cells out of 49,382 were classified as high-risk cells (Fig. 3-a). The hotspot map (Fig. 4-a, magnified in the Supplementary Material along with the other hotspot maps) provides a more concentrated overview of the areas where at-risk cells accumulate. These areas were characterised by the following features (Table 2):

- Marshy areas surrounding the lake, particularly to the west near Torre del Lago, to the north-west between the canals, and to the east on the border with the town of Massaciuccoli. (H1.O);

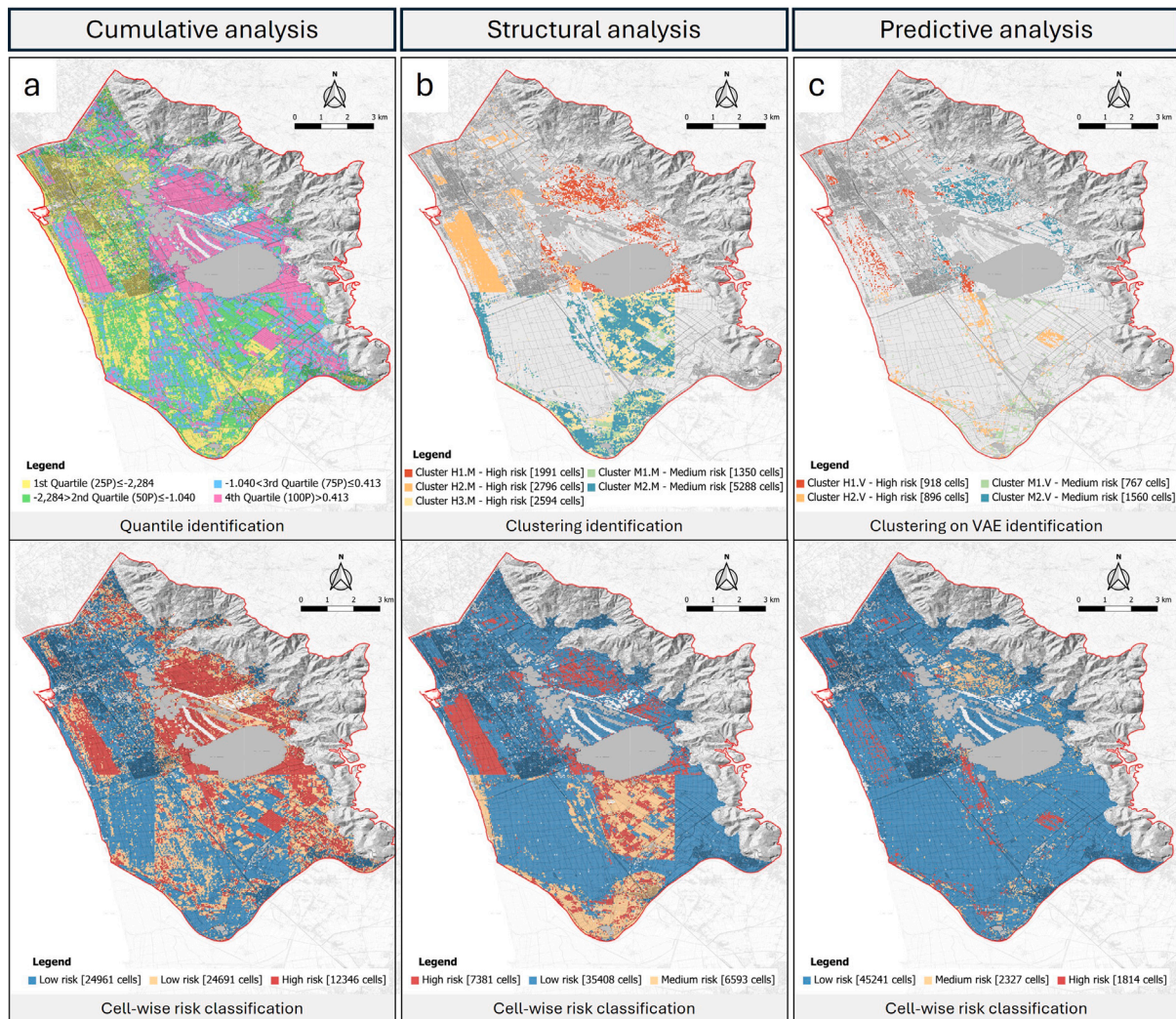


Fig. 3. Distribution of the classification results (first row) estimated by the three models included in our methodology (cumulative, structural, and predictive analyses), along with respective cell-wise risk classifications as low/medium/high (second row). The spatial resolution of the cells is approximately 50 m.

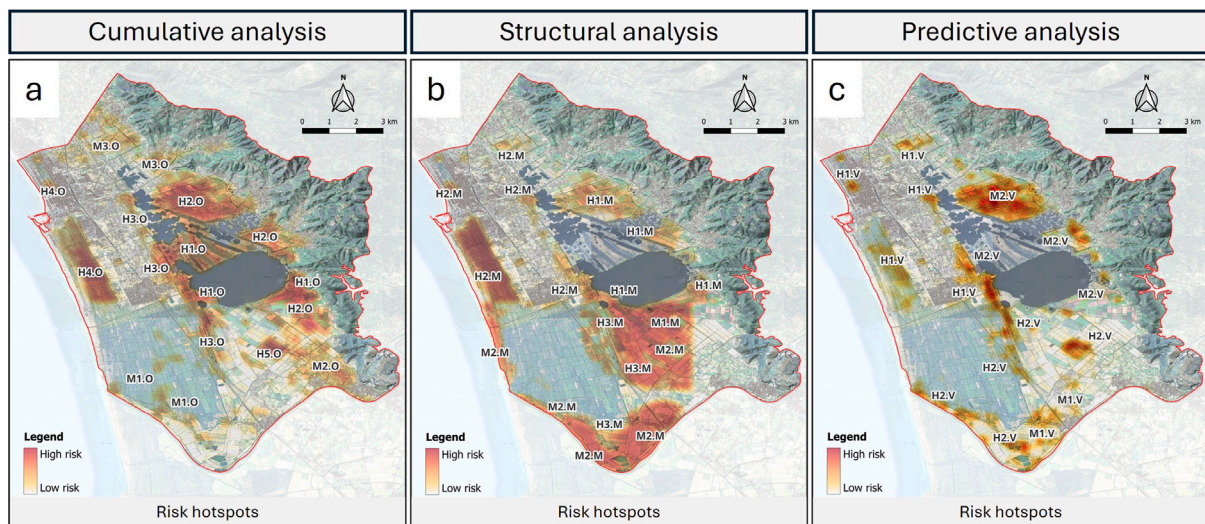


Fig. 4. Distribution of the hotspots of ecosystem risk estimated by the three models included in our methodology (cumulative, structural, and predictive analyses). The spatial resolution is approximately 50 m. Labels are intended as references for the characterisations reported in the article. Prefix *H* indicates hotspots with a prevalence of high-risk cells. Prefix *M* indicates hotspots with a prevalence of medium-risk cells.

Table 2

Interpretation and classification of medium and high ecosystem risk by the cumulative analysis. For each area, the table reports a unique identifier, its spatial location, the main fragile conditions present (along with their identifiers in the Massaciuccoli Atlas), and the main harmful forces at work.

Identifier	Localization	Fragile conditions	Harmful forces
High Risk			
H1.O	Marshy areas distributed along the perimeter of the lake, with a higher concentration in the western sector near Torre del Lago Puccini, in the north-western sector along the canal network, and in the eastern sector bordering the town of Massaciuccoli, characterised by hygrophilous vegetation.	<ul style="list-style-type: none"> • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in grassland cover (id 12) • High increase in absolute average temperature (id 57) • High decrease in potential evapotranspiration (id 58)
H2.O	Agricultural areas immediately adjacent to the lake and marshland, with a higher concentration in the western sectors near the towns of Pian del Quercione and Massarosa, west of Bozzano and Quiesa, and in the southern sector south of the town of Massaciuccoli.	<ul style="list-style-type: none"> • High presence of grassland cover (id 11) • High plant growth (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • High reduction in grassland cover (id 12) • High increase in absolute average temperature (id 57) • High decrease in potential evapotranspiration (id 58)
H3.O	Mixed agricultural and wooded areas located between the towns of Viareggio and Torre del Lago Puccini, near the marshland area, extending south-southeast along the A12 motorway and the Aurelia state road.	<ul style="list-style-type: none"> • Moderate tree cover density (id 9) • Moderate presence of grassland cover (id 11) • High plant growth (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in tree cover density (id 10) • High presence of hazard areas (id 17) • Moderate increase in absolute average temperature (id 57) • Moderate decrease in potential evapotranspiration (id 58)
H4.O	Agricultural area cultivated with olive groves, located in the south-eastern sector of Massaciuccoli Lake.	<ul style="list-style-type: none"> • Moderate tree cover density (id 9) • High presence of grassland cover (id 11) • Moderate plant growth (id 40) • Moderate species richness (id 65) 	<ul style="list-style-type: none"> • High reduction in grassland cover (id 12) • High reduction in precipitation (id 59)
H5.O	Wooded area, known as Pineta di Levante, located between the coast and the urbanised areas of Torre del Lago Puccini and Viareggio.	<ul style="list-style-type: none"> • High tree cover density (id 9) • Moderate plant growth (id 40) • Moderate species richness (id 65) 	<ul style="list-style-type: none"> • Moderate decrease in potential evapotranspiration (id 58)
Medium Risk			
M1.O	Wooded areas within the Migliarino, San Rossore and Massaciuccoli Natural Park, including the area around the mouth of the Serchio River.	<ul style="list-style-type: none"> • High tree cover density (id 9) • Moderate plant growth (id 40) 	<ul style="list-style-type: none"> • High reduction in precipitation (id 59)
M2.O	Agricultural areas near and north-east of the towns of Vecchiano and Nodica.	<ul style="list-style-type: none"> • Moderate species richness (id 65) 	<ul style="list-style-type: none"> • High reduction in precipitation (id 59)
M3.O	Agricultural areas near the urban centres of Piano di Mommio and Piano di Conca	<ul style="list-style-type: none"> • Moderate tree cover density (id 9) • Moderate presence of grassland cover (id 11) • Moderate plant growth (id 40) 	<ul style="list-style-type: none"> • Moderate reduction in grassland cover (id 12) • High increase in absolute average temperature (id 57) • High decrease in potential evapotranspiration (id 58)

- Agricultural zones located immediately surrounding the lake and marshy areas, particularly to the west of the towns of Pian del Quercione and Massarosa, west of the towns of Bozzano and Quiesa, and south of the town of Massaciuccoli (H2.O);
- Mixed agricultural-wooded regions situated between the towns of Viareggio and Torre del Lago Puccini, extending south-southeast along the routes of the A12 highway and the Aurelia state road (H3.O);
- Fields with olive groves located to the southeast of the lake (H4.O);
- The wooded area of Pineta di Levante, located west of the town of Torre del Lago Puccini (H5.O).

A more heterogeneous distribution of high and medium-risk areas was observed in the wooded regions of the Natural Park of Migliarino, San Rossore, and Massaciuccoli, as well as at the mouth of the Serchio River in the southwest (M1.O). Additional mixed areas were found in the agricultural fields adjacent to the towns of Vecchiano and Nodica (M2.O), and in the northern part of the basin, near the towns of Piano di Mommio and Piano di Conca (M3.O).

All identified areas were characterised by elevated values for most fragile conditions and harmful forces, as the model did not account for the interconnection between stressor variables.

Structural analysis. The high-risk clusters identified by this model corresponded to 7381 cells within the basin (Fig. 3-b). Table 3 provides a detailed characterisation of these areas, while the hotspot map (Fig. 4-b) illustrates the accumulation areas of the at-risk cells. The characterisation can be summarised as follows:

- Agricultural areas surrounding the lake and located between the towns (H1.M). In these areas, the presence of grasslands and plant productivity is threatened by a gradual artificial reduction in grassland coverage and an increase in aridity observed in recent years;
- Wooded areas within the Pineta di Levante, near the Tyrrhenian coast (H2.M). These areas, rich in biodiversity with high tree cover density and plant productivity, face increasing aridity (particularly a decrease in potential evapotranspiration) as the primary threat;
- Agricultural fields south of the lake and within the bends of the Serchio River (H3.M). These areas are characterised by high biodiversity and substantial plant production, yet they are also subject to an increasing trend of aridity (notably a decrease in precipitation) and ongoing artificial reductions in grassland coverage in recent years.

Medium-risk areas (6593 cells) were predominantly located in mixed agricultural-wooded zones south of the lake (M1.M and M2.M),

Table 3

Interpretation and classification of medium and high ecosystem risk by the structural analysis. For each area, the table reports a unique identifier, its spatial location, the main fragile conditions present (along with their identifiers in the Massaciuccoli Atlas), and the main harmful forces at work.

Cluster	Localization	Fragile conditions	Harmful forces
High Risk			
H1.M	Agricultural areas surrounding the lake and marshes, located between neighbouring towns (Montramito, Pian del Quercione, Massarosa, Bozzano, Quiesa and Massaciuccoli). To a lesser degree, it also involves the lacustrine areas, particularly those located to the south-east and south-west of the lake, near Torre del Lago Puccini and Massaciuccoli, characterised by the presence of hygrophilous vegetation.	<ul style="list-style-type: none"> • High presence of grassland cover (id 11) • High plant productivity (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • High reduction in grassland cover (id 12) • High increase in absolute average temperature (id 57) • High decrease in potential evapotranspiration (id 58)
H2.M	Wooded area, known as Pineta di Levante, located between the coast and the urbanised areas of Torre del Lago Puccini and Viareggio. Smaller wooded areas are located along the highway sections between Torre del Lago Puccini and the lake and between Viareggio and the marshland area; within the town of Viareggio (Pineta di Ponente); south of the town of Piano di Mommio.	<ul style="list-style-type: none"> • High tree cover density (id 9) • High plant growth (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • High decrease in potential evapotranspiration (id 58)
H3.M	Agricultural areas located south of the lake, near the town of Migliarino and within the bend of the Serchio River. To a lesser extent, it also includes tree crops distributed along the highway. Cluster complementary to M1.M and M2.M.	<ul style="list-style-type: none"> • High plant growth (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • High reduction in grassland cover (id 12) • High reduction in precipitation (id 59)
Medium Risk			
M1.M	Mixed agricultural areas with tree crops located south of the lake, with a higher concentration around the town of Migliarino and along the bend of the Serchio River, almost reaching its mouth. Cluster complementary to H3.M and M2.M.	<ul style="list-style-type: none"> • High presence of grassland cover (id 11) • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in tree cover density (id 10) • High reduction in precipitation (id 59)
M2.M	Agricultural areas located south of the lake, with a higher concentration near the town of Migliarino and along the bend of the Serchio River. It also includes part of the wooded areas near the highway (on the west side) and part of the coastal area (Marina di Vecchiano), both of which fall within the Migliarino, San Rossore and Massaciuccoli Natural Park. Cluster complementary to H3.M and M1.M.	<ul style="list-style-type: none"> • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in tree cover (id 10) • Moderate reduction in grassland cover (id 12) • High reduction in precipitation (id 59)

complementary to the high-risk agricultural areas classified as H3.M. They were characterised by a good level of biodiversity and occasional grassland coverage, but were increasingly affected by reduced precipitation and moderate declines in tree cover density and grassland extent.

Predictive analysis. The high-risk clusters identified by this model corresponded to 1814 cells within the basin (Fig. 3-c). Similar to the structural analysis, Table 4 reports the detailed area characterisation, and the hotspot map (Fig. 4-c) highlights the accumulation areas of the at-risk cells. The characterisation can be succinctly summarised as follows:

- Wooded areas and grasslands within the Pineta di Levante (H1.V). These areas exhibit high tree cover density and plant productivity, alongside moderate biodiversity. However, they are also affected by the artificial reduction of green areas coupled with increasing aridity in recent years;
- Areas adjacent to the highway and state road, south of the lake (H2.V). These regions are characterised by grasslands and plant productivity, featuring high biodiversity but also reduced precipitation and a decline in green areas in recent years;
- Localised biodiversity-rich areas within the bends of the Serchio River (H2.V). These areas primarily consist of grasslands characterised by rich biodiversity, yet are increasingly subject to aridity due to rising temperatures, declining precipitation, and evapotranspiration.

Areas adjacent to the highway and state road, south of the lake and within the bends of the Serchio River (H2.V). These zones mainly corresponded to forest margins interspersed with grassland patches, characterised by ecosystems with substantial tree cover, high plant productivity, and high species richness. However, these areas have been increasingly exposed to climatic stressors, particularly reduced precipitation, accompanied by a moderate decline in grassland extent and a progressive increase in aridity.

Similar to the structural analysis, medium-risk areas (2327 cells) were mainly situated in mixed agricultural-wooded zones south of the lake and in localised regions along the bends of the Serchio River (M1.V and M2.V), characterised by rich biodiversity but a moderate reduction in tree cover and grassland.

3.1.1. Consistency with other studies

The predictive analysis emerged as the most focused and localised model, because it produced at-risk areas that were also classified by the other models. Consequently, the risk areas delineated by this model demonstrated greater reliability in identifying ecosystem risk. Our qualitative evaluation, therefore, compared the hotspot areas from this model (Table 4 and Fig. 4-c) with the findings from other localised scientific studies. In particular:

Agricultural areas north and east of the lake (M2.V). Numerous studies have identified the peri-urban agricultural belt that separates agricultural settlements from the town of Massaciuccoli as one of the most vulnerable areas within the basin (Lastrucci et al., 2017; Viciani et al.,

Table 4

Interpretation and classification of medium and high ecosystem risk by the predictive analysis. For each area, the table reports a unique identifier, its spatial location, the main fragile conditions present (along with their identifiers in the Massaciuccoli Atlas), and the main harmful forces at work.

Cluster	Localization	Fragile conditions	Harmful forces
High Risk			
H1.V	Extensive coverage of the wooded area known as Pineta di Levante, located between the coast and the urbanised areas of Torre del Lago Puccini and Viareggio. Additional mixed forest areas with grass cover are located along sections of the A12 highway between Torre del Lago Puccini and the lake and between Viareggio and the marshland area; within the town of Viareggio (Pineta di Ponente); and south of the town of Piano di Mommio.	<ul style="list-style-type: none"> • High tree cover density (id 9) • High plant productivity (id 40) • Consistent species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in grassland cover (id 12) • High increase in absolute average temperature (id 57) • High increase in potential evapo-transpiration (id 58)
H2.V	Areas covered by trees located: south-east of Torre del Lago along the A12 motorway, the Aurelia state road and the railway; south of the lake (agricultural area planted with olive groves), along the bend of the Serchio river to its mouth.	<ul style="list-style-type: none"> • High tree cover density (id 9) • High plant productivity (id 40) • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in grassland cover (id 12) • High reduction in precipitation (id 59)
Medium Risk			
M1.V	Heterogeneous coverage of agricultural areas located south of the lake. Greater concentration in the vicinity of the town of Migliarino and within the bend of the Serchio River, extending towards the mouth.	<ul style="list-style-type: none"> • High presence of grassland cover (id 11) • High species richness (id 65) 	<ul style="list-style-type: none"> • Moderate reduction in grassland cover (id 12) • High reduction in precipitation (id 59)
M2.V	Agricultural areas surrounding the lake and located between neighbouring urban centres (Torre del Lago Puccini, Montramito, Pian del Quercione, Massarosa, Bozzano, Quiesa and Massaciuccoli).	<ul style="list-style-type: none"> • High presence of grassland cover (id 11) • Consistent species richness (id 65) 	<ul style="list-style-type: none"> • High increase in absolute average temperature (id 57) • High decrease in potential evapo-transpiration (id 58)

2017). These reclaimed and intensively cultivated fields (Conti et al., 2021; SNPA and Consumo di suolo, dinamiche territoriali e servizi ecosistemici, 2025) place considerable pressure on natural habitats, primarily by introducing nutrients into the lake system and indirectly affecting adjacent grassland and forest ecosystems (Pasquetti et al., 2024). Hydrological analyses have substantiated that nitrogen- and phosphorus-rich agricultural runoff exacerbates trophic imbalance and facilitates siltation processes (Rossetto et al., 2010a; Baneschi et al., 2013). Concurrently, the historical reclamation and oxidation of peat soils have induced subsidence and hydrological instability (Pasquetti et al., 2024), thereby jeopardising the water regime essential for the conservation of grassland and forest habitats within the marsh system. In summary, this hotspot aligned with one of the basin's most critical areas in terms of hydrological and ecological vulnerability.

Tree transition areas along infrastructure and urban margins (H1.V and H2.V). Previous research has indicated that the urban and infrastructural margins of the Massaciuccoli basin are characterised by significant fragmentation, resulting in a diminished ecological connectivity (Pasquetti et al., 2024; Lastrucci et al., 2017). These regions, situated along road corridors and at the periphery of urban settlements, predominantly feature residual vegetation cover that is particularly vulnerable to microclimatic fluctuations and anthropogenic disturbances. This vegetation acts as an ecological buffer between urbanised zones and wetlands. The dynamics of fragmentation and anthropogenic pressures, as previously noted in national environmental reports (SNPA and Consumo di suolo, dinamiche territoriali e servizi ecosistemici, 2025), may contribute to a decline in ecosystem resilience. Hotspot H1.V indeed indicated these transitional zones as ecologically sensitive areas under significant environmental stress.

Olive groves in the southern part of the lake (H2.V). The olive groves located in the southern agricultural fields play a crucial role in promoting landscape diversity and facilitating biodiversity conservation. However, the implementation of intensive management practices, including the use of fertilisers, pesticides, and irrigation, has been linked to pervasive contamination and habitat degradation (Silvestri et al., 2012; Pistocchi

et al., 2012). The conversion of permanent grasslands and agroecological mosaics into tree crops has reduced the landscape's structural complexity, thereby increasing its vulnerability to climatic and hydrological stresses. The hotspot identified by our model was located within the area indicated by other studies, and its characterisation in terms of stressor variables reflected the conditions described in these studies.

Wooded areas of the levante and ponente pine forests (H1.V). The coastal pine forests in this region are increasingly vulnerable to phytosanitary degradation, soil compaction, and salinisation, alongside disturbances stemming from wildfires and tourist activities (Lastrucci et al., 2017; Pasquetti et al., 2024). These factors have led to a progressive decline in tree cover, alterations in floristic composition, and reduced ecological functionality. Our model successfully identified these areas as hotspots, highlighting the concurrence of anthropogenic and climatic pressures that impact these vulnerable forest habitats.

Mixed woodland and grassland areas along the bend of the River Serchio to its mouth (H2.V and M1.V). The areas delineated by the model in the lower Serchio valley, situated between the town of Migliarino and the mouth of the River Serchio, represent a complex mosaic wherein agricultural land intersects with woodland and grassland formations. This coastal strip features a highly dynamic hydrogeomorphological system, susceptible to fluctuations in water regimes, flooding, salinisation, and marine intrusion (Pasquetti et al., 2024; Rossetto et al., 2010b). The riparian habitats, along with the grassland and woodland formations present within the delta plain, are crucial for maintaining biodiversity and regulating the ecological integrity of the river system. However, they are particularly susceptible to climatic stresses such as drought and alterations in water levels, as well as anthropogenic pressures related to the proximity of cultivated areas and the settlements of Migliarino. The presence of agricultural land disrupts ecological continuity, heightening the risk of habitat degradation and diminishing its resilience. The localised hotspots identified by our model in this region, along with their characterisation regarding the concurrency of stressor variables, aligned with findings from the other studies cited.

Overall, the regions identified by the model were known to basin experts for their ecological significance, persistent anthropogenic pressures, and ongoing degradation. Furthermore, the highlighted hotspots

Table 5

Pairwise comparison metrics for the ecosystem risk assessments produced by the three models compared (cumulative, structural, and predictive analyses). The table reports Accuracy, Precision, Recall, F1 score, and Cohen's Kappa for cell-wise and heatmap comparisons, with one model serving as the reference and the other as the prediction. Symmetrical values for Accuracy, F1, and Kappa are not reported to avoid information redundancy.

		Cell-wise			Hotspots		
		Accuracy			Accuracy		
Reference		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
	Structural	0.65	—	—	0.63	—	—
Reference		Precision			Precision		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	0.33	0.76	—	0.51	0.81
Reference		Recall			Recall		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	0.37	0.25	—	0.49	0.35
Reference		F1 score			F1 score		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		Cohen's Kappa			Cohen's Kappa		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		Accuracy			Accuracy		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		Precision			Precision		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		Recall			Recall		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		F1 score			F1 score		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—
Reference		Cohen's Kappa			Cohen's Kappa		
		Prediction			Prediction		
		Cumulative	Structural	Predictive	Cumulative	Structural	Predictive
	Cumulative	—	—	—	—	—	—

were known to have experienced degradation and undergone ecological transformations characterised by renaturalisation dynamics, including recolonisation by grassland and woodland.

3.2. Quantitative evaluation results

Pairwise analysis. For each model pair, the results of the cell-wise and hotspot comparisons were calculated separately. Fig. 5 visually represents these comparisons, emphasising the differences and similarities in the spatial patterns of the cell-wise (left-hand side) and hotspot (right-hand side) maps. Additionally, Table 5 reports the confusion matrices. It is noteworthy that the at-risk areas identified by the predictive analysis were predominantly encompassed within those identified by the other models, whereas the cumulative analysis exhibited complementary patterns to the structural analysis.

Cumulative vs structural analyses: The cumulative analysis achieved a moderate accuracy (0.65) in replicating the assessments of the structural analysis, along with only a slight agreement ($\kappa = 0.12$) as per Fleiss' interpretation of Cohen's Kappa (Fleiss and Cohen, 1973) (Fig. 5-a). When the two models were alternated as references, the Precision (0.37–0.33), Recall (0.33–0.37), and F1-score (0.35) remained modest, indicating limited overlap between the at-risk areas identified jointly by the two models, suggesting substantial complementarity.

This limited disagreement was corroborated by the comparison between the hotspots (Fig. 5-b), which revealed an Accuracy of 0.63 and a slight agreement ($\kappa = 0.21$). The Precision (0.49–0.51), Recall (0.51–0.49), and F1-score (0.50) confirmed that the two models were largely complementary.

Cumulative vs predictive analyses: These two models achieved the highest level of agreement among the comparisons, with an Accuracy of 0.79 and a fair Kappa (0.29) (Fig. 5-c). Precision and Recall had dissimilar values (0.76 vs 0.25) and swapped when predictive analysis was used as the reference, with the previous false negatives becoming false positives and vice versa. This dissimilarity was reflected in a modest F1 Score (0.38), indicating some complementarity between the two models. However, these results also indicate that the false positives — at-risk areas for the predictive analysis not classified by the cumulative analysis — were few. Therefore, the at-risk areas were almost entirely included in the cumulative analysis. On the other hand, the low Recall indicated that the cumulative analysis produced more at-risk areas. The hotspot comparison confirmed this scenario, with a fair Accuracy (0.73) and Kappa (0.35), and an F1-score of 0.49 (Fig. 5-d).

Structural vs predictive analyses: The comparison between these two models yielded a moderate Accuracy (0.73) but only a slight agreement ($\kappa = 0.16$) (Fig. 5-e). When structural analysis was utilised as the reference, the Precision of predictive analysis was 0.17, while the

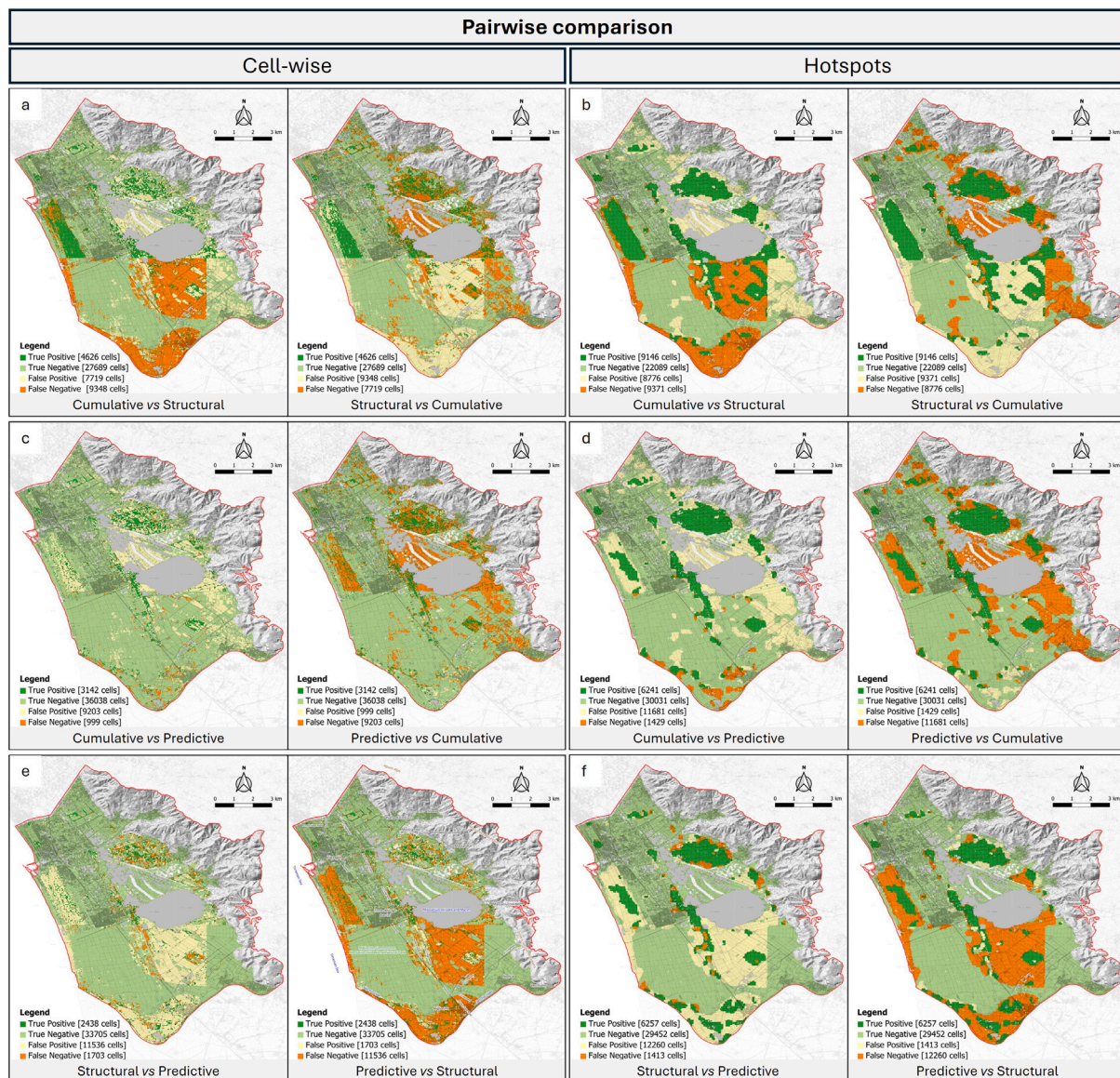


Fig. 5. Pairwise differences between the three models included in our methodology (cumulative, structural, and predictive analyses), reported at the levels of the individual cells classified by the models and the hotspot maps. For each map, the first-mentioned model is used as the reference, and the second as the prediction. The spatial resolution of each map is approximately 50 m.

Recall was 0.59, resulting in an F1-score of 0.27. This observation suggests that the at-risk areas delineated by the predictive analysis were nearly fully included within those identified by the structural analysis, indicating that the structural analysis produced a broader classification. This assertion was further corroborated by the hotspot comparison (Accuracy = 0.72 and F1-score = 0.48), despite a slightly higher agreement between the maps ($\kappa = 0.33$) (Fig. 5-f).

Simultaneous analysis. The simultaneous analysis, applied to both cell-wise and hotspot maps, evaluated the spatial consistency among the three models, thereby distinguishing areas characterised by joint agreement from those exhibiting divergent classifications. This approach offered an integrated perspective on the models, as reflected in their complementary outputs. Fig. 6 illustrates the spatial distributions of the agreement categories, while Table 6 presents the evaluation metrics. The category “all medium/high risk” encompassed 3.40% of the cells and 9.92% of the hotspots, indicating a low level of joint assessment in areas classified as medium or high risk. Conversely, the “one non-medium/high risk” category accounted for 10.40% of the cells and 14.07% of the hotspots, suggesting that high concordance

among the models occurred when at least two models indicated a cell as being at medium or high risk. The “all non-medium/high risk” category accounted for 55.64% of the cells and 44.59% of the hotspots, reflecting substantial agreement in classifying numerous areas as non-medium/high risk. Finally, the “one medium/high risk” category accounted for 30.56% of the cells and 31.42% of the hotspots, as several locations were classified as medium or high risk by only one model due to the complementary nature of their assessments.

4. Discussion

This paper has presented a multi-step methodology for conducting ecosystem risk assessments for the Massaciuccoli Lake basin in Tuscany (Italy). The presented results provide valuable, previously unavailable information for conservation planning by local authorities. The approach incorporates three complementary models: a *cumulative* analysis based on an overlap index, a *structural* analysis based on Multi K-means, and a *predictive* analysis based on applying structural analysis to selected outputs from a Variational Autoencoder. These models

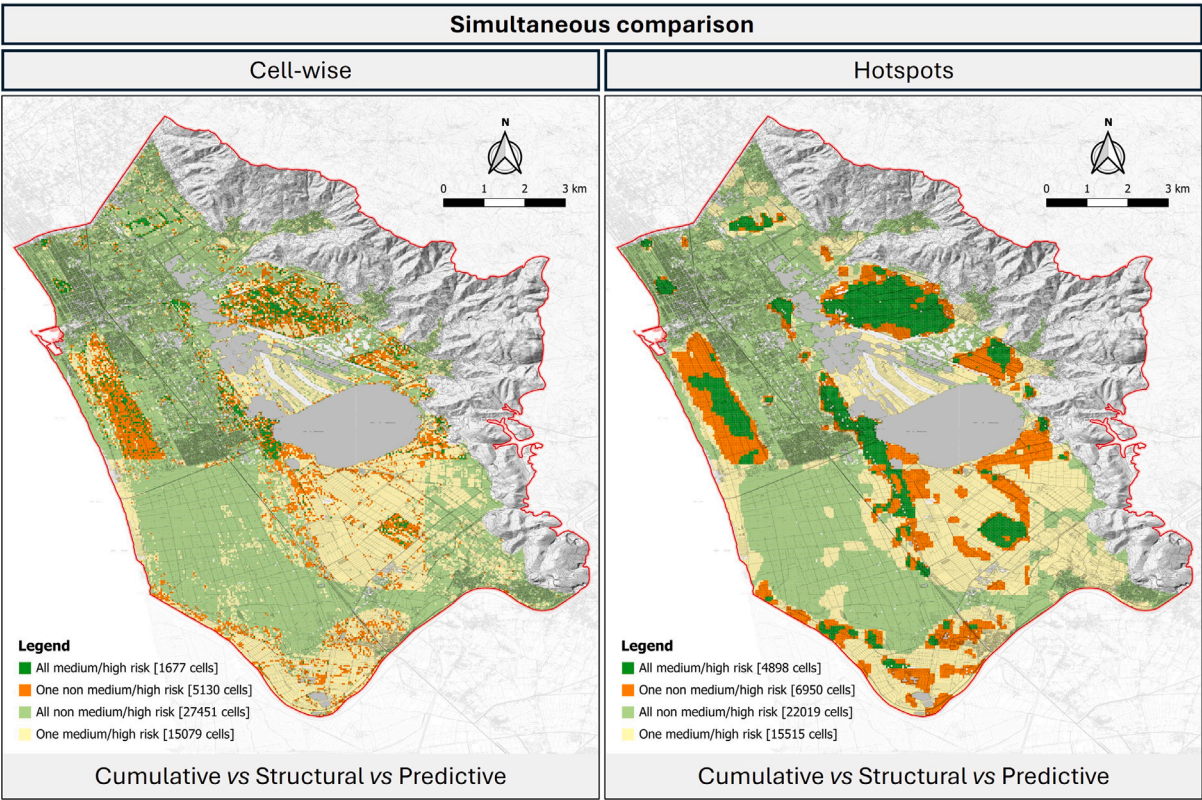


Fig. 6. Simultaneous comparison of the three models included in our methodology (cumulative, structural, and predictive analyses), reported at the levels of the individual cells classified by the models and the hotspot maps. Different colours/tones are used to distinguish different agreement conditions among the models. The spatial resolution of the maps is approximately 50 m.

Table 6
Quantitative evaluation of the spatial agreement categories from the simultaneous comparison of the three methods included in our methodology (cumulative, structural, and predictive analyses). The table first summarises the agreement category definitions. Then, for cell-wise and hotspot comparisons separately, it reports the number of cells per category and the relative percentage. This table is meant to support the interpretation of the colour-coded spatial patterns reported in the related figure.

Agreement category	• Medium/high risk ◦ Non medium/high risk			Cell-wise		Hotspots	
	Cumulative	Structural	Predictive	Number of cells	%	Number of cells	%
All medium/high risk	•	•	•	1677	3.40%	4898	9.92%
	◦	•	•	716		1359	
One non medium/high risk		◦	•	1465	10.40%	1343	14.07%
	•	•	◦	2949		4248	
All non medium/high risk	◦	◦	◦	27451	55.64%	22019	44.59%
	•	◦	◦	6254		7433	
One medium/high risk	◦	•	◦	8587	30.56%	8012	31.42%
	◦	◦	•	238		70	
	Total			49337	100.00%	49382	100.00%

processed data focusing on ecosystem stressors within the analysed area, encompassing current fragile conditions and harmful forces likely to converge, thereby pushing the status of the Massaciuccoli ecosystem towards degradation.

A *simultaneous* comparative analysis provided an overview of the three models combined, highlighting their complementary and concordant assessments. This analysis, complemented by pairwise comparisons, demonstrated that the models yielded risk assessment maps with low to moderate agreement. This outcome also underscores complementary conceptual and operational foundations underlying the methodologies employed. The *cumulative* analysis showed an *inclusive* approach that typically identified larger at-risk areas. Generally, this model tended to overestimate ecosystem risk compared to the other models. However, by design, it identified regions of high stress within

the ecosystem, even though these regions were not necessarily critical. Conversely, the *structural* analysis identified clusters of vector similarities, thereby revealing homogeneous conditions rather than merely co-occurrences of stressors. This model produced risk maps interpretable as *precautionary*, as they tended to encompass extensive areas surrounding the identified risk-accumulation zones. Lastly, the *predictive* analysis inferred latent variables that captured the generative process underlying the data, thereby accounting for more complex interrelations among the stressor variables. By nonlinearly reducing the stressor variables, this model captured latent anomalies associated with ecosystem risks, aligning with findings from other studies (Pavirani et al., 2025b; Coro et al., 2025). The results indicated that it prioritised the reliability of risk assessments over their expansiveness; its high-risk areas were primarily included in those identified by the other models.

However, this emphasis on reliability may lead to an underestimation of risk, rendering the model's results more *conservative*.

These structural differences justify the ensemble overview presented in the simultaneous analysis. Hierarchical spatial structures also emerged: areas of maximum agreement (depicted in dark green in Fig. 6) represented critical ecosystem risk zones, often encircled by a zonal buffer (orange), within which the two models concurred on the presence of risk. This observable gradient suggests that ecosystem risk is not merely a binary construct but rather has a progressive intensity, consistent with the ecological complexity characteristic of the Massaciuccoli basin. From this perspective, the simultaneous maps in Fig. 6 were more effective at capturing the nuances of risk and identifying management priority areas for conservation or mitigation efforts. Therefore, the discrepancies observed across the three models should not be interpreted as inconsistencies but rather as complementary perspectives on a complex ecosystem. Future research will aim to integrate these three models into a hierarchical framework to bolster decision-making processes through spatial, statistical, and predictive evidence.

The qualitative analysis demonstrated the reliability of the jointly identified ecosystem risk hotspots, as they showed correspondence with other studies focused on specific subregions of the basin. Overall, the principal risks emerging were driven by the pressure of climate change (with increasing aridity) on biodiversity-rich or productive areas, as well as by the reduction of grassland and trees near urbanised areas due to urban or industrial expansion.

From an ecosystem management perspective, using either one risk assessment model or the simultaneous maps as a support for guiding interventions should depend on the specific ecological objectives, available data, and management context, as these factors influence prioritisation of conservation and intervention efforts. The *cumulative* analysis map (Fig. 4-a, magnified in the Supplementary Material along with the other hotspot maps) was more useful for providing a comprehensive overview of risk areas and general problems. It consistently identified marshy areas, agricultural zones around the lake, areas near rural towns, mixed agricultural-wooded regions, and areas surrounding the highway and the principal state road. From a management perspective, it provided a macroscopic view of the principal risk assets. The *structural* analysis map (Fig. 4-b) was more specific in detecting risk areas and might be more useful for spatial planning. However, its precautionary output would not be suitable for urgent management interventions that should prioritise targeted actions in specific areas. The *predictive* analysis map (Fig. 4-c), on the other hand, would support this type of planning at the expense of a comprehensive overview of the problems. Ongoing research (CNR News, 2025) is testing how the produced maps can practically support spatial planning by presenting them to local authorities to gather feedback and support their decision-making processes.

The main limitation of our approach is that risk depends on the selection of the stressor variables. A more exhaustive analysis would require examining new variables related to ecosystem services, such as tourist pressure, flood-reduction capacity, water-purification rates, aesthetic value, tourism revenue, soil health indicators, and nutrient cycling rates. Future research, in collaboration with local authorities, will investigate how the ecosystem-risk maps evolve after incorporating these variables and whether the methodology can be extended to assess risks associated with changes in ecosystem services.

Another limitation of the present study concerns the absence of independent, spatially explicit ground-truth data for validating the ecosystem risk maps. Ecosystem risk is inherently a forward-looking concept, representing the likelihood of future degradation rather than an immediately observable condition. Additionally, the available studies for the Massaciuccoli basin are fragmented and focus on specific environmental components, preventing the construction of a comprehensive validation benchmark. For this reason, validation was conducted through a combination of model comparison and qualitative

agreement with independent expert studies, which should be interpreted as an assessment of plausibility rather than strict accuracy. Future work will aim to strengthen validation by integrating additional data sources, such as updated ecological monitoring data, changes in species distributions, and ecosystem service indicators, to provide more direct empirical support for the identified risk patterns.

Future research will also analyse the multi-temporal dimensions of the Massaciuccoli Atlas data for long-term ecosystem monitoring and management. While the present study focuses on a single time frame to evaluate the methodology, the availability of historical and projected datasets offers the potential to analyse the temporal evolution of ecosystem risk. Extending the methodology to multiple time steps would enable the identification of risk trajectories, the detection of emerging or declining hotspots, and the assessment of ecosystem responses to climatic and anthropogenic changes. However, such an extension will require dedicated methodological developments, including the definition of temporal consistency metrics and comparative frameworks across time frames.

Finally, although the methodology was designed to be reproducible and reusable, in line with the Open Science paradigm (Coro et al., 2024c), its generalisability has not been empirically demonstrated across multiple study areas. However, the approach relies on general-purpose models that operate on the statistical properties of stressor variables and does not require site-specific parameterisation, thereby supporting its applicability across different ecological contexts. However, successful transfer depends on the availability of harmonised geospatial datasets representing comparable stressor variables. Future research will focus on applying and validating the methodology across different wetland and terrestrial ecosystems to assess its robustness under varying environmental conditions and data configurations. Such cross-site analyses will be essential to fully establish the generalisability and operational reliability of the proposed framework.

5. Conclusions

In conclusion, from the perspective of emerging technologies in ecology and sustainable development, the proposed methodology represents a disruptive approach in terms of integration, scalability, and applicability. By combining deterministic, statistical, and deep learning models within a unified framework, it enables the analysis of a complex ecological system without requiring labelled data or strong prior assumptions. This is particularly relevant in contexts where data are heterogeneous or incomplete, as is often the case in rapidly changing environments. Overall, the proposed approach helps bridge advanced computational methods and ecological applications, in line with the development of scalable, context-aware solutions for sustainable ecosystem management.

Several challenges emerged during the development of the methodology. The first concerned integrating heterogeneous geospatial information from different sources, temporal references, spatial representations, and measurement scales. This challenge was addressed through a dedicated harmonisation workflow that transformed the original datasets into a common stressor-variable representation suitable for joint analysis. A second challenge involved defining ecologically meaningful stressors from a much larger set of environmental variables. This challenge required identifying variables representing either harmful forces or fragile ecosystem conditions, along with appropriate transformations and inversions to ensure a consistent ecological interpretation. Finally, the absence of spatially explicit ground-truth data for ecosystem risk necessitated the adoption of a consistency-based evaluation framework that combined model agreement analysis with comparisons with independent ecological studies. These challenges motivated several design choices of the proposed methodology, including the use of complementary analytical models, stressor standardisation, and a transparent refinement procedure to improve the robustness and interpretability of the resulting risk assessments.

CRediT authorship contribution statement

Gian Luca Vannini: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Gianpaolo Coro:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Software and data availability

All input and output data are available on the Zenodo repository at <https://doi.org/10.5281/zenodo.17160216>.

The source code of all workflows and all trained models is available on GitHub at <https://github.com/cybprojects65/TerrestrialEcosystemRiskAssessment>

Ethics declaration

Our research did not include invasive research on human or animal subjects.

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

The authors wish to thank the Lega Italiana Protezione Uccelli (LIPU) for providing valuable feedback on the results of this research.

Appendix A. Supplementary material

A supplementary material document is attached to the present publication. It contains magnified versions of the figures to enhance inspection, as well as cartographic maps that complement those presented in the manuscript.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2026.103872>.

Data availability

All data are open access. Links are reported in the “software and data availability” section within the paper.

References

- An, J., Cho, S., 2015. Variational autoencoder based anomaly detection using reconstruction probability. *Spec. Lect. IE* 2 (1), 1–18.
- Ascough II, J., Maier, H., Ravalico, J., Strudley, M., 2008. Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecol. Model.* 219 (3–4), 383–399.
- Baldaccini, G.N., 2018. Zone umide: Dal degrado al recupero ecologico. Il caso del lago di Massaciuccoli (Toscana nord-occidentale). *Biol. Ambient.* 32, 85–98.
- Baneschi, I., Basile, P., Bonari, E., Ercoli, L., Gorrieri, L., Guidi, M., Paglialonga, S., Pistocchi, C., Risaliti, R., Rossetto, R., Sabbatini, T., Silvestri, N., 2013. Agricoltura e tutela delle acque nel bacino del lago di Massaciuccoli, Pacini.
- Bertacchi, A., Lombardi, T., Giannini, V., Silvestri, N., Bonari, E., et al., 2015. A case of ecological renaturation in a drained Mediterranean peatland: the case study of the Massaciuccoli lake basin (Tuscany, IT). In: 110 °Congresso della Società Botanica Italiana onlus Pavia, 14-17 September 2015: Abstracts, Keynote lectures, Communications, Posters. Società Botanica Italiana, 16–16.
- Bertacchi, A., Lombardi, T., Saggese, A., Lazzeri, V., et al., 2021. The vegetation of a relict salt marsh area in the pisan coast in the context of brackish wetlands of tuscany. *Plant Sociol.* 58 (1), 41–53.
- Bove, P., Bertini, A., Coro, G., 2026. Estimating species commonness and prevalence through unsupervised methods. *Sci. Rep.*

- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., et al., 2012. Biodiversity loss and its impact on humanity. *Nature* 486 (7401), 59–67.
- Ciccolini, V., Giannini, V., Bosco, S., Pellegrino, E., Pistocchi, C., Sabbatini, T., Rossetto, R., Silvestri, N., Bonari, E., 2013. Restoration of a Mediterranean drained peatland: the case study of the Massaciuccoli lake basin (Tuscany, IT). In: International Workshop AWARE Approaches in Wetland Restoration. pp. 21–23.
- CNR News, 2025. Il contributo del Cnr alla pianificazione ecosistemica del Lago di Massaciuccoli. available at <https://www.cnr.it/it/news/13903/il-contributo-del-cnr-alla-pianificazione-ecosistemica-del-lago-di-massaciuccoli>.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20 (1), 37–46.
- Conti, P., Carmignani, L., Massa, G., Meccheri, M., Fantozzi, P., Masetti, G., Rossetto, R., 2021. Note illustrative della carta geologica d'Italia alla scala 1:50.000, foglio 260 viareggio. Tech. Rep., ISPRA - Servizio Geologico d'Italia, progetto CARG, Università degli Studi di Siena, Dipartimento di Scienze della Terra, https://www.isprambiente.gov.it/Media/carg/note_illustrative/260_Viareggio.pdf.
- Coro, G., 2024. An open science oriented bayesian interpolation model for marine parameter observations. *Environ. Model. Softw.* 172, 105901.
- Coro, G., Bove, P., Baneschi, I., Bertini, A., Calvisi, L., Provenza, A., 2024a. Climate change effects on animal presence in the Massaciuccoli Lake basin. *Ecol. Informatics* 81, 102644.
- Coro, G., Bove, P., Ellenbroek, A., 2026. Detecting fishing areas from navigation radar detector data. *Int. J. Digit. Earth* 19 (1), 2617004.
- Coro, G., Bove, P., Kesner-Reyes, K., 2023. Global-scale parameters for ecological models. *Sci. Data* 10 (1), 7.
- Coro, G., Pavirani, L., Ellenbroek, A., 2024b. Extracting mediterranean hidden fishing hotspots through big data mining. *IEEE Access* 12, 85465–85483.
- Coro, G., Pavirani, L., Ellenbroek, A., 2025. Computing ecosystem risk hotspots: A mediterranean case study. *Ecol. Informatics* 85, 102918.
- Coro, G., Sana, L., Bove, P., 2024c. An open science automatic workflow for multi-model species distribution estimation. *Int. J. Data Sci. Anal.* 1, 1–20.
- Cramer, W., Guiot, J., Fader, M., Garrahou, J., Gattuso, J.-P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., et al., 2018. Climate change and interconnected risks to sustainable development in the mediterranean. *Nat. Clim. Chang.* 8 (11), 972–980.
- Davidson, N.C., 2018. Wetland losses and the status of wetland-dependent species. In: *The Wetland Book*. Springer, pp. 369–381.
- De Lange, H.J., Sala, S., Vighi, M., Faber, J.H., 2010. Ecological vulnerability in risk assessment—a review and perspectives. *Sci. Total Environ.* 408 (18), 3871–3879.
- Dilokthanakul, N., Mediano, P.A., Garnelo, M., Lee, M.C., Salimbeni, H., Arulkumar, K., Shanahan, M., 2016. Deep unsupervised clustering with gaussian mixture variational autoencoders. *arXiv preprint arXiv:1611.02648*.
- Doersch, C., 2016. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.
- Fennessy, M.S., Jacobs, A.D., Kentula, M.E., 2007. An evaluation of rapid methods for assessing the ecological condition of wetlands. *Wetlands* 27 (3), 543–560.
- Finlayson, C., Davies, G.T., Moomaw, W.R., Chmura, G., Natali, S.M., Perry, J., Roulet, N., Sutton-Grier, A.E., 2019. The second warning to humanity—providing a context for wetland management and policy. *Wetlands* 39 (1), 1–5.
- Fleiss, J.L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* 33 (3), 613–619.
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., et al., 2005. Global consequences of land use. *Science* 309 (5734), 570–574.
- Gong, D., Huang, M., Ge, Y., Zhu, D., Chen, J., Chen, Y., Zhang, L., Hu, B., Lai, S., Lin, H., 2025. Revolutionizing ecological security pattern with multi-source data and deep learning: An adaptive generation approach. *Ecol. Indic.* 173, 113315.
- Guo, M., Li, J., Sheng, C., Xu, J., Wu, L., 2017. A review of wetland remote sensing. *Sensors* 17 (4), 777.
- Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., et al., 2015. Habitat fragmentation and its lasting impact on earth's ecosystems. *Sci. Adv.* 1 (2), e1500052.
- Halpern, B.S., Longo, C., Lowndes, J.S.S., Best, B.D., Frazier, M., Katona, S.K., Kleisner, K.M., Rosenberg, A.A., Scarborough, C., Selig, E.R., 2015. Patterns and emerging trends in global ocean health. *PLoS One* 10 (3), e0117863.
- Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., d'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., et al., 2008. A global map of human impact on marine ecosystems. *Science* 319 (5865), 948–952.
- Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E., Hungate, B.A., Matulich, K.L., Gonzalez, A., Duffy, J.E., Gamfeldt, L., O'Connor, M.I., 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486 (7401), 105–108.
- IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services.
- IPCC, 2023. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. Rep., Intergovernmental Panel on Climate Change.

- Jørgensen, S.E., 2008. Overview of the model types available for development of ecological models. *Ecol. Model.* 215 (1–3), 3–9.
- Joyce, C.B., Simpson, M., Casanova, M., 2016. Future wet grasslands: ecological implications of climate change. *Ecosyst. Health Sustain.* 2 (9), e01240.
- Junk, W.J., An, S., Finlayson, C., Gopal, B., Květ, J., Mitchell, S.A., Mitsch, W.J., Robarts, R.D., 2013. Current state of knowledge regarding the world's wetlands and their future under global climate change: a synthesis. *Aquat. Sci.* 75 (1), 151–167.
- Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M., 2014. Semi-supervised learning with deep generative models. *Adv. Neural Inf. Process. Syst.* 27.
- Kingma, D.P., Welling, M., 2022. Auto-encoding variational bayes. *arXiv:1312.6114*, <https://arxiv.org/abs/1312.6114>.
- Kingsford, R.T., Basset, A., Jackson, L., 2016. Wetlands: conservation's poor cousins. *Aquat. Conserv.: Mar. Freshw. Ecosyst.* 26 (5), 892–916.
- Korpinen, S., Andersen, J.H., 2016. A global review of cumulative pressure and impact assessments in marine environments. *Front. Mar. Sci.* 3, 153.
- Landis, W.G., Durda, J.L., Brooks, M.L., Chapman, P.M., Menzie, C.A., Stahl Jr., R.G., Stauber, J.L., 2013. Ecological risk assessment in the context of global climate change. *Environ. Toxicol. Chem.* 32 (1), 79–92.
- Lastrucci, L., Dell'Omo, L., Foggi, B., Massi, L., Nuccio, C., Vicenti, C., Viciani, D., et al., 2017. Contribution to the knowledge of the vegetation of the Lake Massaciuccoli (northern Tuscany, Italy). *Plant Sociol.* 54, 67–87.
- Legendre, P., Legendre, L., 2012. *Numerical Ecology*, vol. 24, Elsevier.
- Linkov, I., Satterstrom, F.K., Kiker, G., Batchelor, C., Bridges, T., Ferguson, E., 2006. From comparative risk assessment to multi-criteria decision analysis and adaptive management: Recent developments and applications. *Environ. Int.* 32 (8), 1072–1093.
- Liu, K., Shuai, R., Ma, L., et al., 2021. Cells image generation method based on vae-rgan. *Procedia Comput. Sci.* 183, 589–595.
- Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., Hooper, D.U., Huston, M.A., Raffaelli, D., Schmid, B., et al., 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294 (5543), 804–808.
- Lou, J., Liu, B., Xiong, Y., Zhang, X., Yuan, X., 2025. Variational autoencoder framework for hyperspectral retrievals (hyper-vae) of phytoplankton absorption and chlorophyll a in coastal waters for nasa's emit and pace missions. *IEEE Trans. Geosci. Remote Sens.*
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 152, 166–177.
- Malczewski, J., 2006. Gis-based multicriteria decision analysis: a survey of the literature. *Int. J. Geogr. Inf. Sci.* 20 (7), 703–726.
- Malekmohammadi, B., Blouchi, L.R., 2014. Ecological risk assessment of wetland ecosystems using multi criteria decision making and geographic information system. *Ecol. Indic.* 41, 133–144.
- Ministero dell'Ambiente e della Sicurezza Energetica, 2025. Rete natura 2000. accessible online at <https://www.mase.gov.it/portale/rete-natura-2000>.
- Mitsch, W.J., Gosselink, J.G., 2015. *Wetlands*. John Wiley & sons.
- Nelson, G.C., Bennett, E., Berhe, A.A., Cassman, K., DeFries, R., Dietz, T., Dobermann, A., Dobson, A., Janetos, A., Levy, M., et al., 2006. Anthropogenic drivers of ecosystem change: an overview. *Ecol. Soc.* 11 (2).
- Nikologianni, A., 2019. Exploring the future: Landscape architects and emerging professionals. *Common Ground* 1, 47–47.
- Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biology* 83 (2), 171–193.
- on Wetlands, C., 2025. Global wetland outlook 2025: Valuing, conserving, restoring and financing wetlands. <http://dx.doi.org/10.69556/GWO-2025-eng>.
- Pasquetti, F., Natali, S., Luppichini, M., Bini, M., Seppia, N. Del, Delgado-Huertas, A., Gianecchini, R., 2024. Hydrodynamics and water quality of a highly anthropized wetland: the case study of the Massaciuccoli basin (Tuscany, Italy). *Environ. Sci. Pollut. Res.* 31 (30), 43117–43137.
- Pavirani, L., Bove, P., Coro, G., 2025a. Assessing marine ecosystem risks through unsupervised methods. *Ecol. Informatics* 103334.
- Pavirani, L., Bove, P., Coro, G., 2025b. Ecosystem risk assessment through stressor concurrency identification: A comparative analysis. In: *OCEANS 2025 Brest*. IEEE, pp. 1–6.
- Pichler, M., Hartig, F., 2023. Machine learning and deep learning—a review for ecologists. *Methods Ecol. Evol.* 14 (4), 994–1016.
- Pignalosa, A., Silvestri, N., Pugliese, F., Corniello, A., Gerundo, C., Del Seppia, N., Lucchesi, M., Coscini, N., De Paola, F., Giugni, M., 2022a. Long-term simulations of Nature-Based Solutions effects on runoff and soil losses in a flat agricultural area within the catchment of Lake Massaciuccoli (Central Italy). *Agric. Water. Manag.* 273, 107870.
- Pignalosa, A., Silvestri, N., Pugliese, F., Gerundo, C., Corniello, A., Del Seppia, N., Lucchesi, M., Coscini, N., De Paola, F., 2022b. Modelling the effects of NBS adoption in mitigating soil losses of a land reclamation area in the Massaciuccoli lake catchment (Central Italy). In: *EGU General Assembly Conference Abstracts*. pp. EGU22–7965.
- Pistocchi, C., Silvestri, N., Rossetto, R., Sabbatini, T., Guidi, M., Baneschi, I., Bonari, E., Trevisan, D., 2012. A simple model to assess nitrogen and phosphorus contamination in ungauged surface drainage networks: application to the Massaciuccoli Lake Catchment, Italy. *J. Environ. Qual.* 41 (2), 544–553.
- Prato, S. Da, Baneschi, I., et al., 2020. The coastal wetland systems of northern Tuscany: Massaciuccoli Lake and ex Porta Lake. State of knowledge and new opportunities for multidisciplinary approach. *Acque Sotterranee-Italian J. Groundw.* 9 (3).
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven earth system science. *Nature* 566 (7743), 195–204.
- Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al., 2022. Tackling climate change with machine learning. *ACM Comput. Surv.* 55 (2), 1–96.
- Rosenblatt, M., 1956. Estimation of a probability density-function and mode. *Ann. Math. Stat.* 27, 832–837.
- Rossetto, R., Baneschi, I., Basile, P., Guidi, M., Pistocchi, C., Sabbatini, T., Silvestri, N., Bonari, E., 2010a. Water management sustainability in reclaimed coastal areas. The case of the Massaciuccoli lake basin (Tuscany, Italy). In: *EGU General Assembly Conference Abstracts*. p. 12158.
- Rossetto, R., Basile, P., Cannavò, S., Pistocchi, C., Sabbatini, T., Silvestri, N., Bonari, E., et al., 2010b. Surface water and groundwater monitoring and numerical modeling of the southern sector of the Massaciuccoli Lake basin. In: *Rendiconti Online Società Geologica Italiana*, vol. 11, Michele Marroni, Mauro Rosi, pp. 189–190.
- Scheffer, M., 2009. *Critical Transitions in Nature and Society*. Princeton University Press.
- Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. *Trends Ecol. Evolut.* 25 (8), 479–486.
- Silvestri, N., Giannini, V., et al., 2019. 6.1. study case 1: restoration of an agricultural drained peatland: the case study of the Massaciuccoli Lake Basin in Tuscany (Italy). In: *Monografia fias do IBADER*. pp. 95–101.
- Silvestri, N., Pistocchi, C., Sabbatini, T., Rossetto, R., Bonari, E., 2012. Diachronic analysis of farmers' strategies within a protected area of central Italy. *Ital. J. Agron.* 7 (2), e20.
- SNPA and Consumo di suolo, dinamiche territoriali e servizi ecosistemici, 2025. Edizione 2025, report ambientali snpa, sistema nazionale per la protezione dell'ambiente. <https://www.snpaambiente.it/wp-content/uploads/2025/11/rapporto-consumo-suolo-2025.pdf>.
- Toscana, R., 2018. Contratto di Lago per il Massaciuccoli. <https://partecipa.toscana.it/web/contratto-di-lago-per-il-massaciuccoli>. (Accessed 18 December 2025).
- Trenberth, K.E., Dai, A., Van Der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. *Nat. Clim. Chang.* 4 (1), 17–22.
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., Langevelde, F. Van, Burghardt, T., et al., 2022. Perspectives in machine learning for wildlife conservation. *Nat. Commun.* 13 (1), 792.
- Vannini, G.L., Bove, P., Coro, G., 2025. A FAIR and open geographic data collection for the Massaciuccoli Lake basin wetland in Italy. *Data Brief* 59, 111303.
- Vannini, G.L., Coro, G., 2025a. Massaciuccoli lake basin - harmonized geographic atlas for ecosystem services assessment. <http://dx.doi.org/10.5281/zenodo.10953582>.
- Vannini, G.L., Coro, G., 2025b. Massaciuccoli Lake basin in Tuscany, Italy. Datasets of 75 environmental, geomorphologic, and socio-economic variables associated from remote past to remote future. <http://dx.doi.org/10.5281/zenodo.10953582>.
- Venturi, S., Tassi, F., Cabassi, J., Randazzo, A., Lazzaroni, M., Capeccchiacci, F., Vietina, B., Vaselli, O., 2021. Exploring methane emission drivers in wetlands: the cases of Massaciuccoli and Porta Lakes (Northern Tuscany, Italy). *Appl. Sci.* 11 (24), 12156.
- Viciani, D., Dell'Omo, L., Vicenti, C., Lastrucci, L., 2017. Natura 2000 protected habitats, Massaciuccoli Lake (northern Tuscany, Italy). *J. Maps* 13 (2), 219–226.
- Wang, Z., Zhang, L., Li, X., Li, Y., Fu, B., 2021. Integrating ecosystem service supply and demand into ecological risk assessment: a comprehensive framework and case study. *Landsc. Ecol.* 36 (10), 2977–2995.
- Zhang, C., Peng, Y., 2018. Stacking vae and gan for context-aware text-to-image generation. In: *2018 Fourth International Conference on Multimedia Big Data. BigMM, IEEE*, Xi'an, China, pp. 1–5.