

Multi-risk assessment of river water quality under climate and land use changes using machine learning and explainable AI

Diep Ngoc Nguyen^{(b),(a)}, Jacopo Furlanetto^{(a),(b),(c)}, Majid Niazkar^{(b),(a)}, Silvia Torresan^{(a),(b),(c)}, Andrea Critto^{(b),(a)}

(a) CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy
 (b) Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy
 (c) National Biodiversity Future Center (NBFC), Palermo, Italy

CONTACT US!
 ngocdiep.nguyen@cmcc.it
 jacopo.furlanetto@cmcc.it

SUMMARY

Freshwater is influenced by compound climate multi-hazards and anthropogenic pressures, posing serious threats to its ecological status. We present two complementary methods tested for the Veneto river network:

(1) **Explainable AI for multi-risk assessment:** Random Forest - SHAP provides driver attribution, interaction/threshold detection, and basin-level evidence for priority setting across hazards and land use, with vulnerability/exposure context.

(2) **Hybrid Graph Neural Network accounting for river connectivity:** a regime-aware graph model that exploits river topology, lower-uncertainty LIMeco predictions, mitigating sparse/uneven observations.

Together, these tools make risk assessment explainable and actionable for decision-making support.

CONCEPTUAL FRAMEWORK

Case study: Veneto Region, Italy

CLIMATE EXTREMES

- Extreme precipitation
- Droughts
- Heatwaves

LAND USE

- Artificial surfaces
- Agricultural lands
- Fertilizer used

RISKS

Water Framework Directive (2000/60/EC)

River Ecological Status

High, Good, Moderate, Bad, Poor

No impact, Impact

Adapted from EEA, 2018

VULNERABILITIES AND EXPOSURES

- Water body typologies
- Basin characteristics
- Riparian buffer zones
- Wetlands

(1) EXPLAINABLE AI TO UNDERSTAND RISK DRIVERS FOR RIVER ECOLOGICAL STATUS IN MULTI-HAZARD CONTEXTS

(Nguyen et al., under review)

METHODOLOGICAL WORKFLOW

Temporal data: Extreme hot, Extreme dry, Extreme wet, Land use

Static data: Water body typologies, Topography, Geomorphology

Random Forest Classifier → SHAP

ANNUAL ELEMENTAL BASIN
Impacts on LIMeco Index (2010-2022)

1887, 1647

SHAP dependence reveals thresholds and synergies among drivers.

- Capturing non-linear interactions between extreme climate events, vulnerabilities, and exposures.
- Yielding insight into which factors most shift predicted impacts, clarifies regime boundaries for intervention design and targeted adaptation planning.

HOT-DRY CONDITION

EXTENSIVE AGRICULTURE UNDER HOT CONDITION

[Each point is an observation with its feature value on the x-axis and SHAP contribution to predicted impact on the y-axis; color encodes an interacting driver. Curves and dashed guides are added post-hoc to aid threshold and interaction identification.]

- Annual layers of hazards and vulnerability/exposure are computed for each basins.
- A Random Forest classifier predicts ecological status (nutrients, specific pollutants, biological) and splits outcomes into **No-impact** vs. **Impact** classes
- SHAP provides local/global attributions to identify the most influential drivers and interaction dynamics.

(2) HYBRID GRAPH NEURAL NETWORK (GNN) FOR RIVER ECOLOGICAL STATUS PREDICTION UNDER CLIMATE AND LAND USE CHANGE SCENARIOS

(Nguyen et al., in preparation)

From sparse observations to graph learning: A heterogeneous river system with data gaps is modeled as a direct, clustered graph. This topology lets the model propagate signals along hydrological connections and learn status patterns despite missing data.

The hybrid GNN outperforms other models: Across calibration and residual plots, predictions lie closest to the 1:1 line with the narrowest 95% bands, indicating the best calibration and precision (RMSE=0.028, MAE=0.02, R²=0.98). Residuals stay centered near zero over most of the range, unlike GNN, XGBoost, and MLP, which show range compression (overpredict low, underpredict high) and larger uncertainty.

The hybrid graph neural network integrates unsupervised (SOM + DBSCAN) and supervised (GNN) AI that learns complex spatio-temporal patterns in river water quality dynamics under different climate and land-based regimes.

GRU for temporal encoding of climate and water quality dynamics

Unsupervised ML to incorporate latent environmental heterogeneity

DBSCAN → BMU prototypes [N×D] → SOM mapping → codebook grid (25×25)

GraphSAGE to capture upstream-downstream connectivity

Aggregate Concat [self // neighbors] Linear W, ReLU, Dropout

Climate change

How do projected changes in temperature, precipitation patterns, and extreme weather events influence ecological status?

Land use change

What synergies or trade-offs exist between land use practices and water quality under future climate scenarios?

Adaptation measures

Where climate adaptation measures are most effective in mitigating risks to water quality?

REFERENCES:

- EEA. European waters - assessment of status and pressures. (2018).