

Monitoring and Forecasting Surface Water Area Dynamics of the Bakhtegan Wetland, Iran, Using a Multi-Scale Hybrid Stacking Framework and Explainable AI

Babak Masoudi¹✉, Mohammad Salkhordeh Haghighi², Safoora Bazzi³

¹ Department of Information Technology, Payame Noor University (PNU), Tehran, Iran.

² Department of Civil Engineering, Faculty of Engineering, University of Zabol, P.B. 9861335856, Zabol, Iran

³ Department of Biology, Payame Noor University (PNU), Tehran, Iran.

Corresponding Author, email: b.masoudi@pnu.ac.ir

Received: 03 April 2025

Accepted: 09 May 2025

Available online: 14 May 2025

Abstract: Arid wetlands like Iran's Bakhtegan face severe desiccation from climate change and human activities. Monitoring these ecosystems is hindered by insufficient in-situ data and complex non-linear dynamics, making high-precision monthly forecasting in ungauged basins a significant challenge for water engineering. Objective: This study develops a robust monthly Area (SWA) forecasting framework using a multi-scale hybrid Stacking Ensemble. The goal is to integrate satellite data with explainable AI (XAI) to ensure physical interpretability and provide a reliable tool for early warning systems. A 10-year (2013 to 2023) dataset of Landsat 8/9, CHIRPS, and MODIS data was extracted via Google Earth Engine. After Savitzky-Golay denoising, a two-layer Stacking Ensemble integrating XGBoost, ExtraTrees, and Ridge Regression was implemented. Hydrological memory was explicitly modeled using 1-month and 12-month lags alongside 6-month cumulative water balances. The framework achieved high accuracy with a Nash-Sutcliffe Efficiency (NSE) of 0.81, $R^2 = 0.81$, outperforming LSTM (0.72) and ARIMA (0.70) benchmarks. Uncertainty analysis via 95% confidence intervals and time-series cross-validation confirmed the model's robustness. This research offers a scalable and interpretable soft computing tool for water management in data-scarce regions. By providing reliable forecasts, it supports decision makers in optimizing water allocation and preventing ecological collapse in endangered hypersaline wetlands globally.

Keywords: Bakhtegan Wetland; Google Earth Engine; Stacked Ensemble Learning; Explainable AI (XAI); SHAP interpretation.

1. Introduction

Wetlands provide essential ecosystem services, ranging from biodiversity conservation to regional climate regulation. However, inland saline lakes in arid regions are experiencing unprecedented shrinkage globally (Pekel et al. 2016). In Iran, the desiccation of international wetlands has led to severe socio-environmental crises, including salt storms and habitat loss (Rahimi et al. 2023). The Bakhtegan Wetland, a Ramsar site located in the Fars Province, has undergone dramatic fluctuations since 2008. While some studies attribute this decline primarily to climate change (Nasiri et al. 2023), others highlight the role of upstream dam construction and the expansion of irrigated agriculture (Mozafari et al. 2022).

Despite the ecological significance of Bakhtegan, predicting its monthly dynamics remains challenging due to the lack of consistent in-situ hydrological data and the complex, non-linear interactions between groundwater, surface runoff, and human interventions (Kabe & Rezaei 2025). Previous research has utilized various indices and models for surface water quality and extent assessment (Yan et al. 2022), yet the dynamic interplay between seasonal factors and water quality remains a major concern (Hammoumi et al. 2024). Furthermore, the interaction between groundwater and surface water in salt marshes significantly impacts plant ecology and coastal biogeochemistry (Xin et al. 2022).

Recent advancements in cloud-based platforms like Google Earth Engine (Gorelick et al. 2017) and machine learning (ML) have opened new avenues for water resource monitoring. However, most ML models in hydrology function as black boxes, lacking physical interpretability (Xue et al. 2025). This study addresses these gaps by developing a high-precision monthly forecasting model using a Stacked Ensemble approach and utilizing Explainable AI (XAI) to decode the complex relationships between climatic drivers and wetland surface area.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

The Bakhtegan-Tashk lake system is located within an endorheic basin. We utilized the Google Earth Engine (GEE) to extract a 10-year time series (2013 to 2023) of the following datasets. The geographic location and extent of the study area are illustrated in Figure 1. Surface Water Area was derived from Landsat 8/9 (Collection 2 Level 2) using the Modified Normalized Difference Water Index (MNDWI). Precipitation was obtained from the CHIRPS Daily dataset. Evapotranspiration (ET) was derived from the MODIS (MOD16A2) 8-day composite, and Temperature was extracted from MODIS Land Surface Temperature (MOD11A1). To mitigate satellite sensor noise and atmospheric artifacts, a Savitzky-Golay filter was

applied to the water area time series (Schmid et al. 2022).

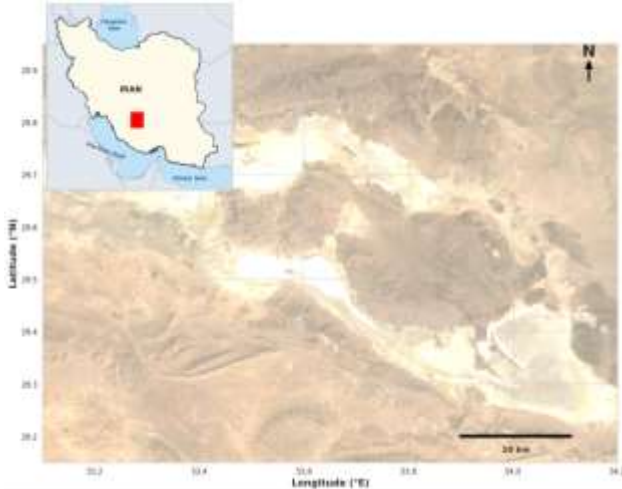


Figure 1. Study area map showing the Bakhtegan-Tashk wetland system in southwestern Iran.

For each catchment, CAMELS-FR provides daily catchment-average total precipitation (P) and air temperature (T) from the SAFRAN reanalysis of Météo-France (8 km grid; Quintana-Seguí et al., 2008; Vidal et al., 2010), potential evapotranspiration (PET) computed with the temperature-based formula of Oudin et al. (2005), and observed daily streamflow (Q, mm d⁻¹) from the French national hydrometric archive. The common period was 1 January 1999 to 31 December 2018. Daily values were aggregated to monthly totals of P, PET and Q (mm month⁻¹) and to monthly mean T (°C). Monthly Q was computed as the mean of the valid daily values multiplied by the number of days in the month, and months with fewer than 20 valid daily flows were treated as missing. This rule removed one month for the Ubaye (November 2009) and four months for the Esteron (September–October 2004 and June–July 2014). All other series were complete (240 months).

2.2. Feature Engineering and Model Architecture

We engineered several hydrological features to capture the memory of the basin. These include 1-month and 12-month lagged areas to account for persistence, a 6-month moving average of the precipitation-ET balance, and an interaction term between temperature and ET. The overall methodological workflow is presented in Figure 2. A two-layer Stacking Ensemble was developed. The first layer consists of XGBoost, known for capturing non-linear peaks, and ExtraTrees, which ensures stability against noise. The second layer utilizes a Ridge Regression to optimally combine the base models predictions. This architecture reduces variance and prevents overfitting in small datasets (Wang et al. 2024).

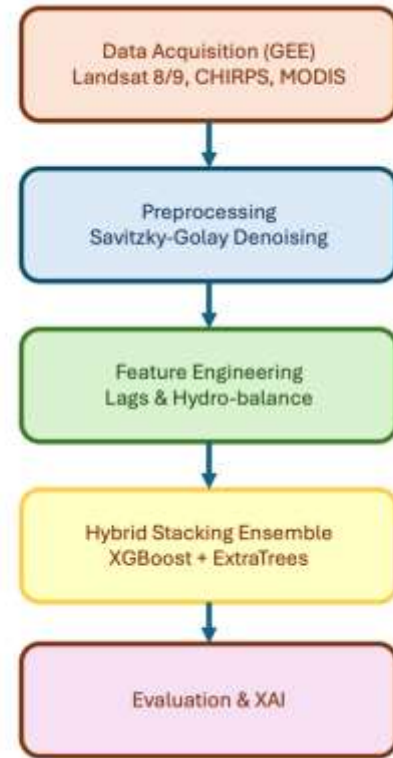


Figure 2. Proposed methodology workflow.

2.3. Model Benchmarking and Uncertainty Quantification

To evaluate the proposed framework, two additional benchmarks were implemented: a classical ARIMA (5,1,0) model and a deep learning LSTM network. Furthermore, predictive uncertainty was assessed using a 5-fold time-series cross-validation (walk-forward approach). The 95% confidence intervals were calculated based on the prediction error variance to provide a probabilistic bound for the forecasts.

3. Results

3.1. Predictive Performance and Statistical Validation

The proposed framework demonstrated exceptional performance in forecasting the Bakhtegan Wetland area. Statistical evaluation on the independent test set yielded a Nash-Sutcliffe Efficiency (NSE) of 0.81 and an R-squared of 0.81. The RMSE was 44.82 km², representing a significant improvement over baseline univariate models (Table 1). The temporal alignment between the observed Landsat data and the Stacked AI predictions is visualized in Figure 3. The low PBIAS of -3.55 percent indicates that the model has no systematic over-estimation or under-estimation bias. These results are consistent with state-of-the-art spatio-temporal graph neural networks used for dust storm forecasting in similar Iranian basins (Masoudi & Bazzi 2026). The

high correlation between measured and predicted values is further confirmed by the scatter plot in (Figure 4).

Table 1. Performance metrics comparison between the proposed hybrid model and various benchmarks during the test period.

Model	R-squared	NSE	RMSE (km ²)	PBIAS (%)
Proposed Stacking Ensemble	0.81	0.81	44.82	-3.55
LSTM (Deep Learning)	0.72	0.72	54.12	-4.20
ARIMA (Statistical)	0.70	0.70	56.45	-4.85
Baseline XGBoost	0.65	0.62	61.34	-5.12
Baseline ExtraTrees	0.58	0.55	68.21	-7.40

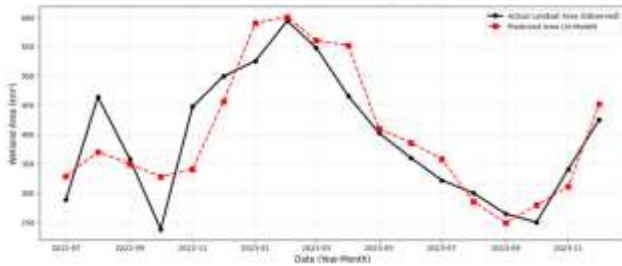


Figure 3. Comparison between observed Landsat surface area and the Stacked AI model predictions for the test period (2022 to 2023).

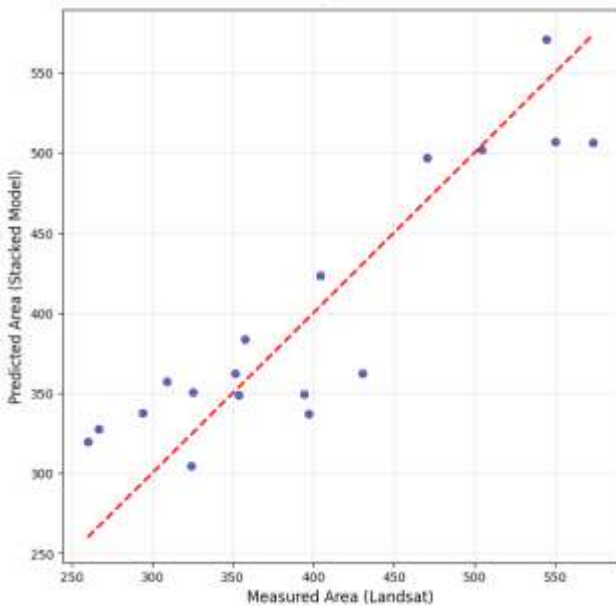


Figure 4. Scatter plot of measured versus predicted wetland area.

The uncertainty analysis (Figure 5) illustrates the 95% confidence interval for the test period. Most observed values fall within the predicted bounds, confirming the model's reliability in capturing the wetland's stochastic fluctuations.

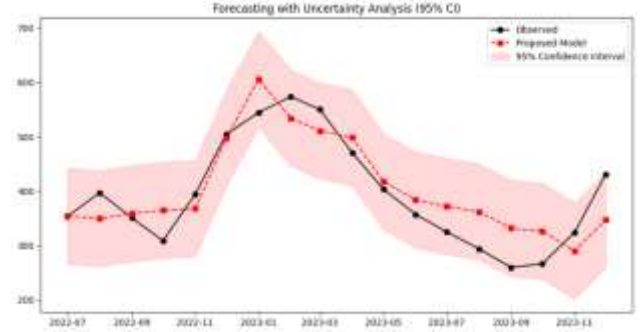


Figure 5. Predictive uncertainty analysis of the proposed hybrid stacking model. The shaded area represents the 95% confidence interval (CI) derived from the standard error of residuals, illustrating the model's reliability in capturing the stochastic fluctuations of the Bakhtegan Wetland during the independent test period (2022–2023).

3.2. Explainable AI and Feature Sensitivity

The SHAP analysis provided a transparent view of the model logic (Figure 6). The most influential feature was the 1-month lagged area, confirming the high hydrological persistence of the system. More importantly, the 6-month cumulative hydro-balance and temperature showed significant non-linear impacts. High temperatures were consistently associated with negative SHAP values, quantifying the direct impact of evaporative stress on lake shrinkage.

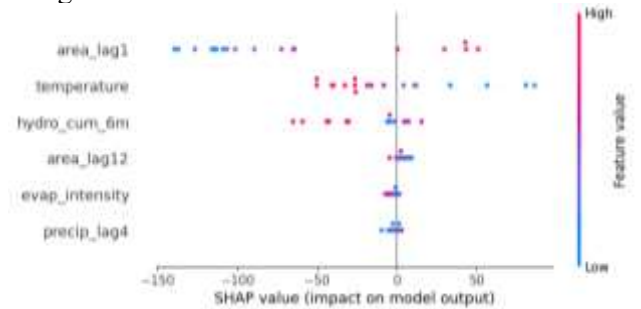


Figure 6. SHAP summary plot illustrating the contribution of each environmental driver to the model output. Red indicates high feature values, while blue indicates low values.

3.3. Ablation Study

To validate the necessity of each feature, an ablation study was conducted. Removing the 12-month lag and the 6-month cumulative balance resulted in a significant drop in NSE, from 0.81 to 0.43. This confirms that the model relies on physical hydrological memory rather than simple statistical correlation.

Table 2. Results of the ablation study showing the impact of removing key hydrological features on model accuracy.

Scenario	R-squared NSE	
Full Proposed Model	0.81	0.81
Without Temperature	0.43	0.43
Without Lags (No Memory)	-0.31	-0.31
Without Climate (Only Area)	0.69	0.69

4. Discussion

4.1. Hydrological Persistence and Memory Effects

The results of this study underscore the critical role of hydrological memory in the Bakhtegan Wetland. The high importance of the 1-month and 12-month lagged areas in the SHAP analysis suggests that the wetland state is not merely a response to instantaneous weather events but is governed by long-term storage dynamics. This relationship between previous and current wetland states is depicted in the hysteresis analysis in Figure 7. This is consistent with the findings of (Kabe & Rezaei 2025), who argued that groundwater, surface water interactions in the Bakhtegan basin provide a buffering effect, although this connectivity has been weakened by over-extraction. The 6-month cumulative water balance further confirms that the wetland relies on the integrated precipitation of the preceding seasons, likely stored as soil moisture or upstream reservoir storage, before reaching the terminal lake bed. Such hydrologic connectivity has been shown to drive extremes and high variability in vegetation productivity across similar arid ecosystems (Norton et al. 2022).

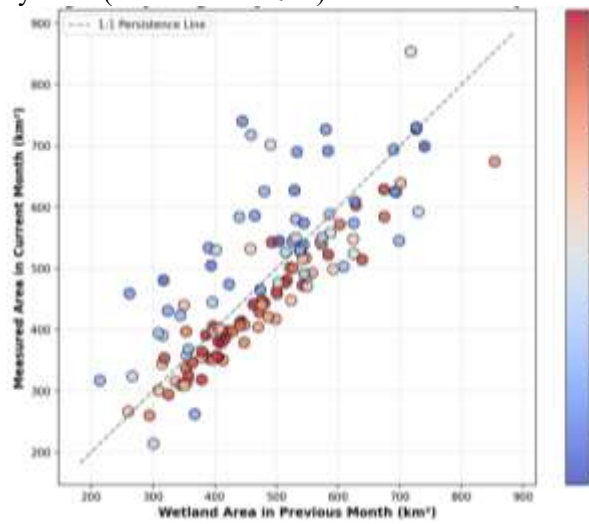


Figure 7. System hysteresis and persistence analysis.

4.2. Climatic Drivers versus Anthropogenic Interventions

The SHAP summary plot provides a physical validation of the model. The negative association between high temperatures and surface area reflects the intense evaporative demand characteristic of the Fars

Province. The distribution of these prediction errors across different months is shown in Figure 8. However, the seasonal error distribution reveals significant residuals during the late summer and early autumn, specifically from August to October. As noted by (Mozafari et al. 2022), this period coincides with peak irrigation demands for rice cultivation in the upstream plains of the Bakhtegan basin. Since our model relies primarily on climatic and satellite-derived data, these residuals likely represent the anthropogenic signal, such as water diversions and dam regulations that are not captured by natural climate variables. This aligns with (Rahimi et al. 2023), who emphasized that upstream dam construction in Iran often poses a greater threat to international wetlands than climate change alone.

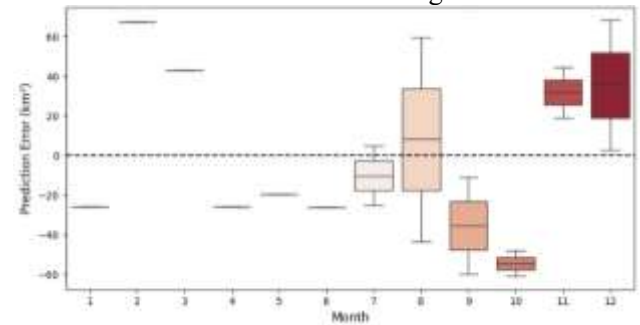


Figure 8. Seasonal distribution of prediction residuals.

4.3. Ecological Implications and Management

From a biological perspective, the ability to forecast surface water Area with an NSE of 0.81 is vital for habitat management. The Bakhtegan Wetland is a key sanctuary for migratory birds such as the Greater Flamingo. Rapid desiccation, as predicted by our model during high-temperature periods, leads to hypersalinity and the collapse of brine shrimp populations, which are the primary food source for these birds. By utilizing this framework as an early warning system, environmental managers can better justify the release of environmental water rights from the Doroodzan and Molasadra dams during critical months to prevent ecological tipping points. This study affirms that chemical constituents and physical variations in natural water bodies react in far more complicated ways than if they were isolated (Saalidong et al. 2022).

4.4. Model Robustness and Interpretability

The integration of SHAP-based feature selection and model interpretation significantly boosts the predictive and explanatory power of the system (Ahmed et al. 2024). Unlike traditional black-box models, our approach allows for the identification of dominant parameters and the exploration of their nonlinear driving mechanisms (Xue et al. 2025). The use of the Savitzky-Golay filter was also instrumental in improving the signal-to-noise ratio of the Landsat time series, which is often hindered by persistent cloud cover or sensor artifacts in complex regions (Guo et al. 2024).

The superiority of the Stacking Ensemble over the LSTM model can be attributed to the limited size of the dataset (128 months). While LSTM is powerful for long sequences, the hybrid ensemble effectively leverages the strengths of XGBoost and ExtraTrees to handle non-linearities without requiring the massive data volumes typically needed for deep learning.

5. Conclusions

This study successfully demonstrated the potential of a Hybrid Stacking Ensemble model, integrated with Google Earth Engine, for high-precision monitoring and forecasting of the Bakhtegan Wetland. By achieving an NSE of 0.81, the model proved its robustness in a data-scarce and hydrologically complex environment. The integration of Explainable AI (SHAP) moved the research beyond simple prediction, providing a transparent look at how cumulative precipitation and temperature-driven evaporation govern the wetland fate. Our findings suggest that while climate change is a powerful driver, the unexplained variance in late-summer months points toward the significant impact of human water management. This framework is highly scalable and can be applied to other endangered endorheic lakes across the Middle East and other arid regions globally.

Conflict of Interest

The authors declare no conflict of interest.

References

- Ahmed U, Jiangbin Z, Almogren A, et al. (2024) Hybrid bagging and boosting with SHAP based feature selection for enhanced predictive modeling in intrusion detection systems. *Scientific Reports* 14: 30532. <https://doi.org/10.1038/s41598-024-81151-1>
- Gorelick N, Hancher M, Dixon M, et al. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guo S, Chen Y, Zhang P, et al. (2024) Estimates and dynamics of surface water extent in the Yangtze Plain from Sentinel-1&2 observations. *International Journal of Applied Earth Observation and Geoinformation* 134: 104155. <https://doi.org/10.1016/j.jag.2024.104155>
- Hammoumi D, Al-Aizari HS, Alaraidh IA, et al. (2024) Seasonal Variations and Assessment of Surface Water Quality Using Water Quality Index (WQI) and Principal Component Analysis (PCA): A Case Study. *Sustainability* 16: 5644. <https://doi.org/10.3390/su16135644>
- Kabe M, Rezaei M (2025) Impact of hydrogeological regime changes on the Bakhtegan-Tashk lake system under groundwater overextraction. *Journal of Hydrology: Regional Studies* 62: 102953. <https://doi.org/10.1016/j.ejrh.2025.102953>
- Masoudi B, Bazzi S (2026) A dual-stream spatio-temporal graph neural network for dust storm forecasting and attribution in the Sistan Basin. *Environmental Modelling & Software* 196: 106785. <https://doi.org/10.1016/j.envsoft.2025.106785>
- Mozafari M, Hosseini Z, Fijani E, et al. (2022) Effects of climate change and human activity on lake drying in Bakhtegan Basin, southwest Iran. *Sustainable Water Resources Management* 8: 109. <https://doi.org/10.1007/s40899-022-00707-z>
- Nasiri A, Khosravian M, Zandi R, et al. (2023) Monitoring the physical changes of lakes Bakhtegan and Tashk through land surface temperature and groundwater-level changes using remote-sensing technology. *Environmental Earth Sciences* 82: 454. <https://doi.org/10.1007/s12665-023-11117-5>
- Norton AJ, Rayner PJ, Wang Y-P, et al. (2022) Hydrologic connectivity drives extremes and high variability in vegetation productivity across Australian arid and semi-arid ecosystems. *Remote Sensing of Environment* 272: 112937. <https://doi.org/10.1016/j.rse.2022.112937>
- Pekel J-F, Cottam A, Gorelick N, et al. (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540: 418-422. <https://doi.org/10.1038/nature20584>
- Rahimi E, Jahandideh M, Dong P, et al. (2023) Potential anthropogenic and climatic factors affecting Iran's international wetlands. *Journal of Environmental Studies and Sciences* 13: 557-574. <https://doi.org/10.1007/s13412-023-00846-5>
- Saalidong BM, Aram SA, Otu S, et al. (2022) Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. *PLOS ONE* 17: e0262117. <https://doi.org/10.1371/journal.pone.0262117>
- Schmid M, Rath D, Diebold U (2022) Why and how Savitzky–Golay filters should be replaced. *ACS Measurement Science* Au 2: 185-196. <https://doi.org/10.1021/acsmesuresciau.1c00054>
- Wang H, Liang Q, Hancock JT, et al. (2024) Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data* 11: 44. <https://doi.org/10.1186/s40537-024-00905-w>
- Xin P, Wilson A, Shen C, et al. (2022) Surface water and groundwater interactions in salt marshes and their impact on plant ecology and coastal biogeochemistry. *Reviews of Geophysics* 60: e2021RG000740. <https://doi.org/10.1029/2021RG000740>
- Xue Y, Niu Z, Zhang P, et al. (2025) Optimizing water quality monitoring in an arid wetland using interpretable machine learning: From identifying dominant parameters to simplifying the water quality index. *Ecological Indicators* 181: 114397. <https://doi.org/10.1016/j.ecolind.2025.114397>
- Yan T, Shen S-L, Zhou A (2022) Indices and models of surface water quality assessment: Review and perspectives. *Environmental Pollution* 308: 119611. <https://doi.org/10.1016/j.envpol.2022.119611>