



OPEN

## Hybrid spatiotemporal modeling of nutrient cycling in wetland ecosystems using advanced mapping techniques and machine learning approaches

Eric Ariel L. Salas<sup>✉</sup>, Kelsie Schrack, Sakthi S. Kumaran & Robert Bennett

Accurate spatiotemporal monitoring of nutrient cycling in wetlands is critical for conservation. However, traditional field-based methods are often inadequate for capturing the overall dynamics of wetlands. To address this challenge, we developed and validated a two-stage hybrid Random Forest regression framework that seamlessly integrated in-situ water quality data with wetland features and satellite imagery from Sentinel-1 and Sentinel-2 within the Google Earth Engine platform. The framework first models baseline nutrient concentrations using discrete wetland characteristics (stage 1) and then models the resulting spatial residual using continuous satellite-derived predictors (stage 2). We applied the model to predict quarterly nitrogen and phosphorus concentrations over four years (2021–2024) in the Beaver Creek Wetlands Greenway (BWG), a mixed-use landscape. The two-stage model demonstrated exceptional predictive performance for both nitrogen (final  $r^2 = 0.90$ , RMSE = 0.129 mg/L) and phosphorus (final  $r^2 = 0.89$ , RMSE = 0.007 mg/L). The variable importance analysis revealed divergent predictive pathways: nitrogen concentration was driven by both landscape-level factors (e.g., land use classes, area and perimeter of wetlands, rainfall) and in-stream biophysical conditions captured by remote sensing (e.g., vegetation and SAR indices), whereas phosphorus was controlled by source-loading from developed land uses. We produced spatiotemporal maps to visualize these distinct patterns. The maps revealed that the BWG system is subject to seasonal nitrogen stress but has experienced significant recovery from prior phosphorus impairment. This study offers a diagnostic framework that could inform focused wetland management strategies and aid in monitoring the health and function of wetland ecosystems.

Wetlands have been known to purify water, sequester carbon, and provide habitat biodiversity<sup>1,2</sup>. However, more than 35% of the global wetlands have already disappeared since 1970<sup>3</sup>. This loss threatens one of their most important functions: the nitrogen (N) and phosphorus (P) nutrient cycle that maintains the water quality in connected aquatic systems<sup>4</sup>. As agricultural runoff and urban development increase the load of nutrients in surface waters<sup>5</sup>, processing of nutrients from wetland becomes more problematic<sup>6</sup>.

The cycling of nutrients in wetlands depends on the way morphology, water flow, and surrounding land use<sup>6,7</sup> interact. These interactions result in varying nutrient processing rates in wetland areas that field sampling cannot capture efficiently<sup>8</sup>. Although traditional point measurements provide detailed information, they also require much labor and time. In addition, they are ineffective in showing how nutrients move through entire wetland systems. When measuring how effective wetlands clean water or predicting how they respond to changing conditions, using traditional methods can be problematic<sup>9</sup>.

Researchers have started to combine field measurements with remote sensing (RS) and machine learning (ML) models to better understand nutrient patterns<sup>10</sup>. Early work used simple methods to estimate nutrient levels between sampling points. Recent studies have used multiple environmental factors to better explain where nutrients are found<sup>11,12</sup>. These studies show that the cycling of nutrients in wetlands is influenced by variables operating on different scales, from local water chemistry to land use throughout the landscape<sup>13</sup>.

Agricultural Research Development Program (ARDP), Central State University, Wilberforce, OH 45384, USA.  
✉email: eric\_asalas@yahoo.com

Many ML algorithms have improved the prediction of nutrients in wetlands by analyzing complex relationships between environmental factors and nutrient levels<sup>14</sup>. For instance, the Random Forest (RF) model works for wetland studies because of its ability to handle large amounts of data. Although it identifies non-linear patterns, RF does not make assumptions about the behavior of the data<sup>15</sup>. However, current studies are still struggling to find an efficient way to combine two different types of data<sup>16</sup>. One type describes the general system conditions<sup>17</sup> (e.g., discrete land use percentages, morphological measurements such as wetland area, perimeter, and shapes). The other type captures small-scale spatial differences and spatial environmental conditions within the wetlands<sup>18</sup> (e.g., continuous satellite images and derivatives). Most studies have examined the cycling of nutrients in wetlands in different regions<sup>8</sup> and wetland types<sup>19</sup>. However, we did not find studies that developed a method to properly combine discrete and continuous data for a time-series prediction of nutrients in wetlands.

Two-stage modeling approaches have proven useful in other fields for handling complex spatial data<sup>20,21</sup>. These approaches can handle the practical challenge of working with discrete and continuous datasets that operate on different temporal scales<sup>10</sup>. In this study, we developed a two-stage modeling approach to address the challenge of combining different types of data. The first stage of the model predicted nutrients based on general wetland characteristics. The second stage predicted the spatial errors from the first stage using satellite data. We tested this method on nitrogen (N) and phosphorus (P) data from water samples collected over four years in three Beavercreek Wetlands Greenway (BWG) wetlands in Ohio. Specifically, our objectives were to: (1) build models of current nutrient patterns in wetland systems; (2) make quarterly maps showing nutrient levels in high resolution; and (3) compare how environmental factors affect N versus P patterns differently.

## Results

### Efficiency of the RF model

Our hybrid RF approach showed strong predictive performance for both N and P concentrations throughout the two-stage modeling process (Table 1). For the prediction of N, the  $r^2$  values improved from 0.86 in stage 1 to 0.88 in stage 2 and achieved a final combined model  $r^2$  of 0.90. Consequently, RMSE decreased from 0.249 mg/L in stage 1 to 0.169 mg/L in stage 2 and reached the lowest value of 0.129 mg/L in the final combined model. MAE followed a similar pattern, decreasing from 0.222 mg/L to 0.155 mg/L to 0.116 mg/L in the final model. The values of MBE remained relatively low throughout, which indicated minimal systematic bias in the predictions.

For the prediction of P, the models also showed consistently high accuracy with  $r^2$  values of 0.86, 0.87, and 0.89 for stage 1, stage 2, and the final combined model, respectively. RMSE values were notably low in all stages (0.007 mg/L, 0.005 mg/L, and 0.007 mg/L), reflecting the precision of the models in predicting P concentrations. The MAE values were also low, decreasing from 0.005 mg/L in stage 1 to 0.004 mg/L in stage 2, and 0.002 mg/L in the final model. Similarly to N models, the MBE values remained consistently low at 0.003 mg/L, 0.002 mg/L, and 0.002 mg/L.

### Spatial mapping of nitrogen concentrations

The N maps for the four-year study period (2021–2024) revealed different temporal and spatial patterns within the wetland ecosystem (Fig. 1). In 2021, N concentrations exhibited a seasonal progression, and Q1 showed predominantly low concentrations ( $\leq 0.2$  mg/L to 0.6 mg/L) throughout most of the wetland area. Quarters Q2 and Q3 displayed elevated concentrations ranging from 1.2 mg/L to 1.4 mg/L, and from 1.8 mg/L to 2.0 mg/L in parts of the wetland. Q4 showed a transition back to moderate concentrations in the range of 1.2 mg/L to 1.6 mg/L.

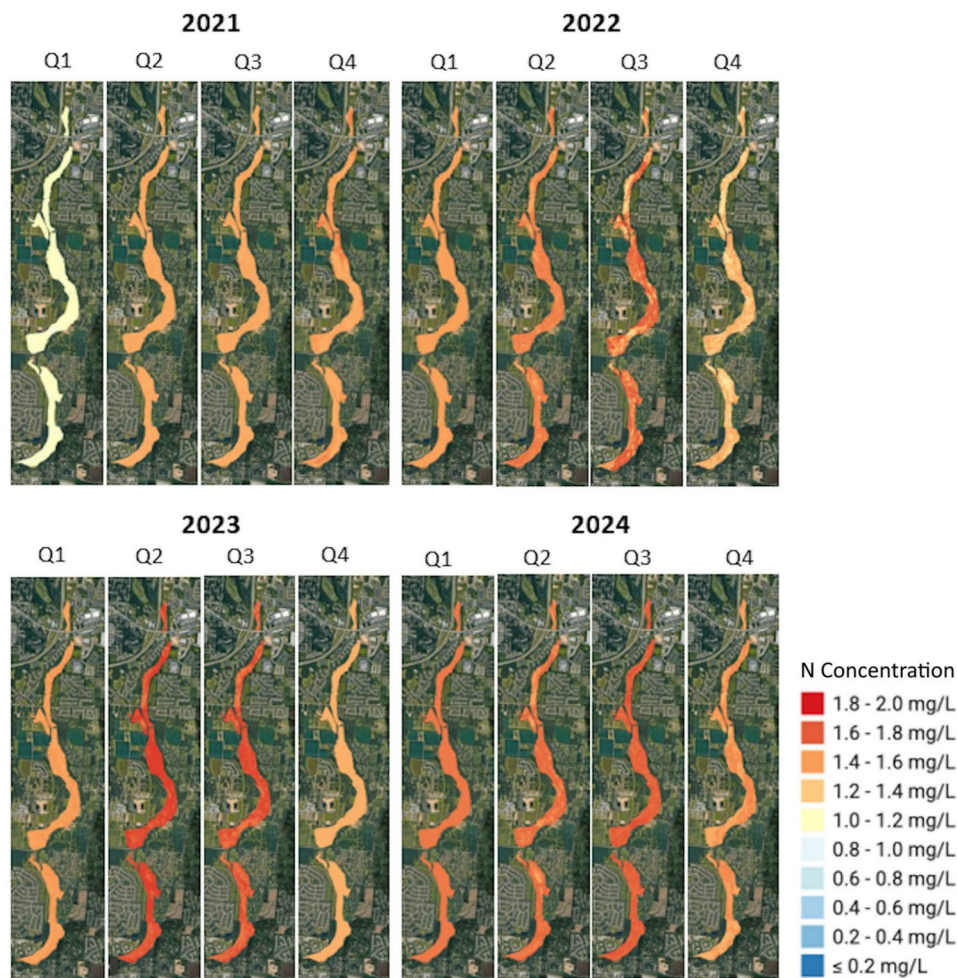
The N map results in 2022 demonstrated a similar seasonal pattern, but we observed higher baseline concentrations compared to 2021. Q1 maintained relatively low concentrations (0.6 mg/L to 1.2 mg/L), while the Q2 and Q3 exhibited the highest levels of N concentration ( $\sim 2.0$  mg/L) throughout large areas of the wetland. Q4 showed a decrease in moderate to high concentrations (1.4 mg/L to 1.8 mg/L).

The year 2023 showed a notable change in the N pattern, with Q2 and Q3 showing the most dramatic increase in concentration. Q2 and Q3 indicated the highest concentrations of N in the study area (1.8 mg/L to 2.0 mg/L). In contrast, the Q1 and Q4 quarters maintained moderate concentrations (1.2 mg/L to 1.6 mg/L). The results of 2024 showed a more moderate pattern compared to previous years, showing concentrations of 1.0 mg/L to 1.8 mg/L throughout most quarters. We found that the peak concentrations of Q2 and Q3 were less extreme than those in 2022 and 2023, although still elevated compared to the corresponding quarters in 2021.

The comparison of results in the four years showed a gradual increase in N concentrations in Q1 from 2021 to 2024. The 2021 Q1 had the lowest values ( $\leq 0.2$  mg/L to 0.6 mg/L), while 2024 Q1 had moderate concentration

Statistics	Nitrogen (N)			Phosphorus (P)		
	Stage 1	Stage 2	Final	Stage 1	Stage 2	Final
$r^2$	0.86	0.88	0.90	0.86	0.87	0.89
RMSE	0.249	0.169	0.129	0.007	0.005	0.007
MAE	0.222	0.155	0.116	0.005	0.004	0.002
MBE	-0.015	0.004	-0.052	0.003	0.002	0.002

**Table 1.** Summary of model performance metrics for the two-stage regression framework. The table shows validation statistics for both N and P predictions at each modeling phase: model for discrete variables (stage 1), model for continuous variables (stage 2), and the final combined model.



**Fig. 1.** Predicted spatial distribution of N concentrations in the wetlands, presented quarterly from 2021 to 2024. The time-series maps, produced by the two-stage model, revealed significant seasonal and inter-annual variability in N levels. This figure was created in Google Earth Engine (GEE) (<https://earthengine.google.com/>) and further processed in ArcGIS Pro v.3.4.0 software by Esri (source: <https://www.esri.com/>). The basemap was provided by ESRI World Imagery (source: ESRI, Maxar, Earthstar Geographics, and the GIS USER Community).

values (1.0 mg/L to 1.4 mg/L). We also observed consistently high N concentrations in Q2 and Q3; specifically in 2022 Q2 and 2023 Q2 to Q3 (1.8 mg/L to 2.0 mg/L). Except for slightly lower values in the last quarter of 2021, the concentration levels in Q4 remained stable across years (1.2 mg/L to 1.6 mg/L).

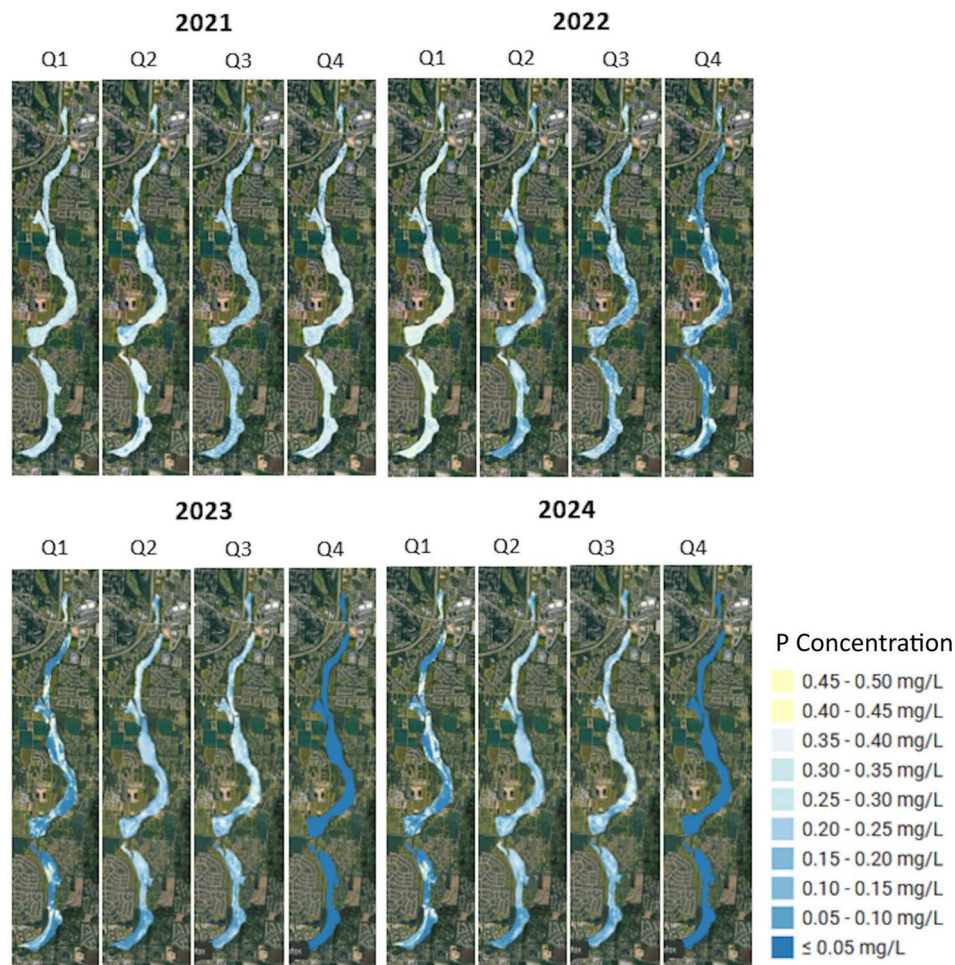
Lastly, we observed consistent seasonal patterns throughout the four years of study, with Q2 and Q3 showing higher N concentrations than Q1 and Q4. Periods 2022 and 2023 showed the most pronounced magnitude of seasonal variation. The difference between quarters with low concentrations (Q1, Q4) and quarters with high concentrations (Q2, Q3) was 1.2 mg/L to 1.4 mg/L (from 0.6 mg/L to 0.8 mg/L in Q1 to 1.8 mg/L to 2.0 mg/L in Q2-Q3). In contrast, 2021 showed the smallest seasonal variation (0.6 mg/L to 0.8 mg/L), with Q1 concentrations around 0.2 mg/L to 0.6 mg/L and Q2-Q3 around 1.2 mg/L to 1.4 mg/L. The year 2024 showed an intermediate seasonal concentration variation (0.8 mg/L to 1.0 mg/L), with Q1 around 1.0 mg/L to 1.2 mg/L and Q2-Q3 around 1.6 mg/L to 1.8 mg/L. The last year showed more gradual transitions between quarters.

### Spatial mapping of phosphorus concentrations

The P concentration maps from 2021 to 2024 (Fig. 2) showed a temporal progression in water quality throughout the study area. The 2021 Q1 map, for example, showed elevated P concentrations throughout the wetland system, ranging from 0.25 mg/L to 0.35 mg/L and higher. This pattern remained stable through Q2 and Q4 of 2021, with minimal variation in P levels.

We observed a significant improvement in water quality in 2022 compared to 2021. We observed a decrease in P concentrations in the 0.15 mg/L to 0.25 mg/L range. This improvement was consistent in all quarters, specifically in Q3 and Q4, with notable reductions in P levels throughout the wetlands.

The resulting P maps for 2023 demonstrated a continued improvement in water quality from 2022. The P concentrations appeared to have further decreased (0.10 mg/L to 0.20 mg/L) in most areas. The two quarters,



**Fig. 2.** Predicted spatial distribution of P concentrations in the wetlands, presented quarterly from 2021 to 2024. The time-series maps, produced by the two-stage model, revealed significant seasonal and inter-annual variability in P levels. This figure was created in Google Earth Engine (GEE) (<https://earthengine.google.com/>) and further processed in ArcGIS Pro v.3.4.0 software by Esri (source: <https://www.esri.com/>). The basemap was provided by ESRI World Imagery (source: ESRI, Maxar, Earthstar Geographics, and the GIS USER Community).

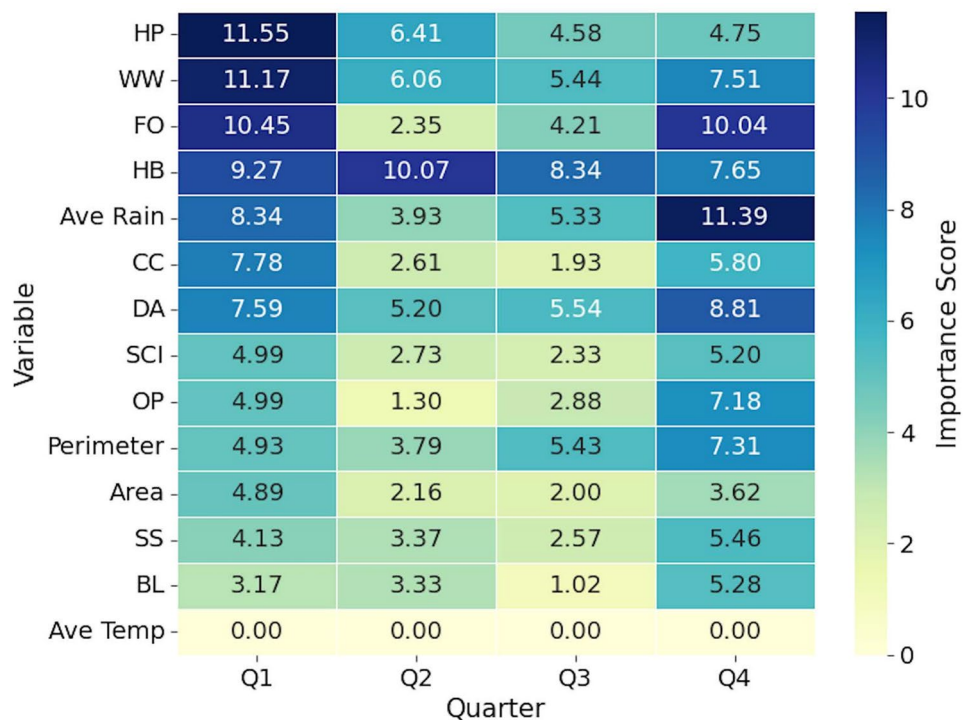
Q1 and Q4, exhibited better seasonal water quality compared to Q2 and Q3. We noticed that the 2023 trend remained favorable compared to 2021 and 2022. Our analysis indicated that the 2024 results had similar trends of P in 2023. Q1 and Q4 showed slightly better water quality than the Q2 and Q3 quarters.

The seasonal pattern of P that we observed in 2024 followed a consistent trend throughout the study period. Winter quarters (Q1 and Q4) generally exhibited lower P concentrations than summer quarters (Q2 and Q3). The spatial distribution showed that while the wetland areas have experienced a general improvement from 2021 to 2024, the greatest reductions in P concentrations were observed during the winter months of 2023 and 2024, particularly in Q1 and Q4 ( $\leq 0.05$  mg/L range).

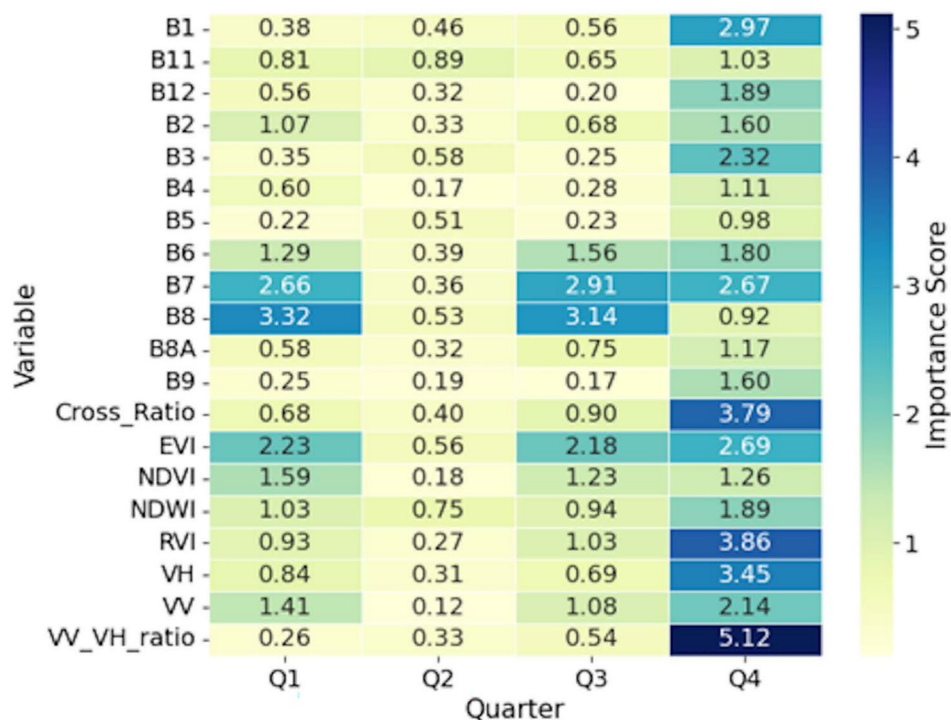
Lastly, the quarterly maps revealed that the P distribution is not uniform. We observed a persistent P "hotspot" in the wider middle section of the wetland. For example, during the summer peaks of Q3 in 2023 and 2024, this section consistently displayed the highest concentrations of P, ranging from 0.35 mg/L to more than 0.40 mg/L. This spatial pattern suggested that the middle section could be a key area for internal P loading, which may be due to low-flow conditions of late summer.

### Important variables for mapping nutrients

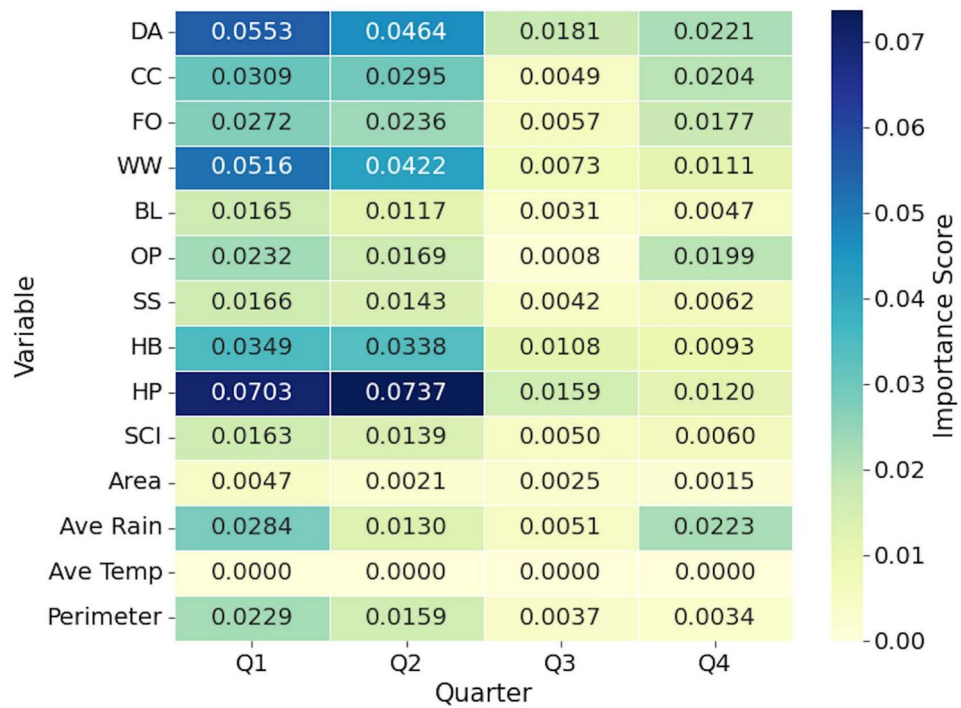
Analysis of the importance of RF variables revealed different predictive patterns between the concentrations of N (Figs. 3 and 4) and P (Figs. 5 and 6) in BWG, with landscape-level variables dominating both first-stage models but showing different magnitudes and seasonal dynamics. For N prediction, considering all quarters, herbaceous (HB) landuse class demonstrated the highest importance score, followed by woody wetland (WW), average rain, and hay/pasture (HP). At the same time, P models showed substantially lower peak importance values, with HP at the top of the list. The herbaceous class (HB) proved to be important for both nutrients. The developed class (DA) showed contrasting patterns, ranking lower for N (ranks 7, 4, 2, and 3 for Q1, Q2, Q3, and Q4, respectively)



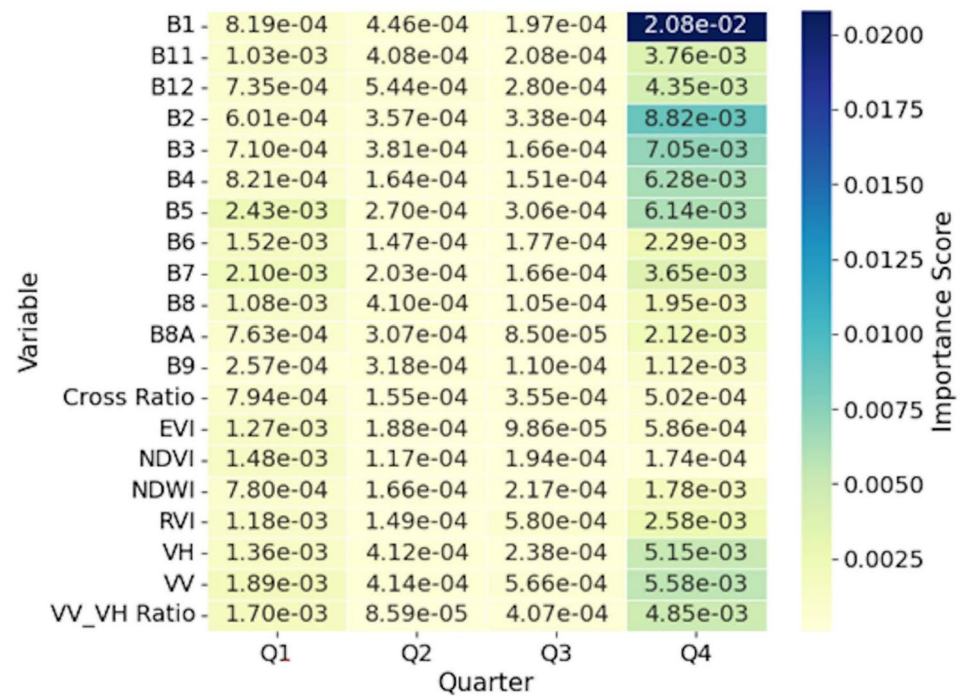
**Fig. 3.** Heatmap showing the 4-year averaged importance scores by quarter in stage 1 for variables used in the estimation of N concentration. *DA* developed area, *CC* cultivated crop, *FO* forest, *WW* woody wetland, *BL* barren land, *OP* open water, *SS* shrub scrub, *HB* herbaceous, *HP* hay/pasture, *SCI* Shape Complexity Index.



**Fig. 4.** Heatmap showing the 4-year averaged importance scores by quarter in stage 2 for variables used in the estimation of N concentration. *B1* to *B12* Sentinel-2 bands, *NDVI* Normalized Difference Vegetation Index, *NDWI* Normalized Difference Water Index, *EVI* Enhanced Vegetation Index, *RVI* Simplified Radar Vegetation Index, *VH* Vertical transmit/Horizontal receive, *VV* Vertical transmit/Vertical receive.



**Fig. 5.** Heatmap showing the 4-year averaged importance scores by quarter in stage 1 for variables used in the estimation of P concentration. *DA* developed area, *CC* cultivated crop, *FO* forest, *WW* woody wetland, *BL* barren land, *OP* open water, *SS* shrub scrub, *HB* herbaceous, *HP* hay/pasture, *SCI* Shape Complexity Index.



**Fig. 6.** Heatmap showing the 4-year averaged importance scores by quarter in stage 2 for variables used in the estimation of P concentration. *B1* to *B12* Sentinel-2 bands, *NDVI* Normalized Difference Vegetation Index, *NDWI* Normalized Difference Water Index, *EVI* Enhanced Vegetation Index, *RVI* Simplified Radar Vegetation Index, *VH* Vertical transmit/Horizontal receive, *VV* Vertical transmit/Vertical receive.

but achieving higher importance for P (rank 2, 2, 1, and 2 for Q1, Q2, Q3, and Q4, respectively). Agricultural variables exhibited nutrient-specific responses, with the cultivated crop (CC) showing moderate importance for the N but lower importance for the P model.

We observed pronounced seasonal variations for the hydrological and morphological variables, which differed between the two nutrients. The average rainfall showed strong temporal dynamics in the prediction of P, dropping from 0.028 in Q1 to 0.005 in Q3 before recovering to 0.022 in Q4, while maintaining a more consistent moderate importance for N (8.34). The class of open water (OP) exhibited contrasting seasonal patterns, showing minimal importance for P but substantial variation for N. The perimeter measurements showed stability across quarters for both nutrients, although with different magnitudes. The area measurements remained consistently low for both the N and P predictions. The Shape Complexity Index (SCI) maintained moderate importance for both nutrients, but showed greater stability in the P model compared to that of N.

The second stage analysis revealed different profiles of variable importance between nutrients. Spectral and water quality indices showed reduced predictive power for P compared to N. The importance of spectral bands varied substantially, with band 1 (B1) reaching its peak importance in Q4 for N and lower values for P. Other spectral bands (B2 to B12) generally exhibited importance scores for N but remained low for P across all quarters. The B8A band demonstrated consistent patterns for both nutrients; maintaining moderate and minimal importance for N and P, respectively.

## Discussion

We developed an RF framework that combined discrete site-specific variables with continuous RS data for wetland nutrient mapping. The success of our hybrid framework stemmed from its ability to (1) utilize the strengths of both discrete and continuous variable inputs, and (2) separate the predictive problem into two distinct and more manageable tasks without requiring extensive data transformation. Our final combined model showed higher predictions for N ( $r^2 = 0.90$ , RMSE = 0.129 mg/L) and P ( $r^2 = 0.89$ , RMSE = 0.007 mg/L) than the two individual stages. The improvement from stage 1 to the final combined model (N  $r^2$  increased from 0.86 to 0.90; P  $r^2$  increased from 0.86 to 0.89) validated our hypothesis that hybrid modeling could better capture the multiscale dynamics and processes of nutrients in wetlands. Guo et al.<sup>20</sup> also observed an improved prediction accuracy when using a two-stage scheme to estimate heavy metal concentrations.

Xue et al.<sup>22</sup> found comparable  $r^2$  values of 0.85 to 0.87 for N and P prediction using RF regression with high-frequency sensor data. However, compared to our approach, we achieved higher accuracy by incorporating the spatial correction component. Similar to Giri and Qiu<sup>23</sup> who emphasized the importance of landuse factors in predicting water quality, stage 1 results captured broad-scale patterns from the landscape characteristics and morphological features. Stage 2 then spatially corrected the baseline predictions and captured localized environmental variations not explained by stage 1 alone. Our approach addressed a critical limitation identified in previous wetland mapping and monitoring studies, where single-factor analyses often proved insufficient to capture complex nutrient dynamics<sup>24,25</sup>. Widely-used single-model approaches usually train a single ML algorithm by combining a set of discrete and continuous predictors. Although powerful, this method could pose a challenge: the strong predictive signal from dominant, coarse-scale variables could overshadow the subtle, but critical, information contained within the high-resolution satellite imagery. Therefore, our two-stage hybrid framework was designed to overcome this limitation. By applying a two-stage approach, the interpretability of the model improved, such as a clearer understanding of the nutrient drivers.

Another advantage of our framework is its high performance when used on both nutrients on various time scales. The two-stage hybrid model accurately predicted nutrients with different biogeochemical behaviors ( $r^2 = 0.90$  for N and  $r^2 = 0.89$  for P). Over four years of study, the model maintained its accuracy despite seasonal variations. This result is important because it addresses the problem identified by Zhang et al.<sup>2</sup>, which is to develop a robust model appropriate for long-term monitoring. In addition, a model with consistent performance under different environmental conditions is also needed for effective management decisions<sup>26</sup>.

The variable importance analysis validated the conceptual basis of our two-stage approach. For N, the results showed a dual dependency on both stages of the model. The stage 1 model identified a competitive interaction between landuse considered nutrient sources (e.g. HP and CC)<sup>27,28</sup> and landuse that served as buffers (e.g. WW and FO)<sup>19,29,30</sup>. This interaction highlighted the baseline N concentration as a function of landscape composition. However, the high importance of the stage 2 variables, namely vegetation indices (EVI, RVI) and SAR metrics (VV/VH ratio), showed that the final N concentration was affected by localized biophysical conditions in the wetland<sup>31</sup>. This two-part dependency showed consistency with the high solubility and biological reactivity of N.

In contrast to N, the prediction of P was dominated by the stage 1 model. This is in agreement with the results from Ding et al.<sup>13</sup> where the distribution of P concentrations corresponded to the shore landuse. Although the most important predictor was the HP, the DA land use class was the most consistent in all quarters. We also saw this consistency of the DA for N. This result was expected since the DA had the highest percentage of landuse class contribution surrounding the sampling points (Supplementary Appendix 1). We found our results to be in agreement with studies linking P levels directly with anthropogenic landscapes<sup>32,33</sup>. The negligible predictive power of the stage 2 variables was an important finding. It suggested that once the land-use source signal has been explained in stage 1, very little remaining variance could be explained by the optical or radar properties of the water. This observation was aligned with the chemical nature of P that is less mobile and more particle-bound than N<sup>34</sup>. The weak signal detected in stage 2 was likely due to the turbidity of the water or suspended solids to which the P was adsorbed<sup>35</sup>.

The final nutrients maps provided a visual record of the ecological health of the wetlands. The N maps (Fig. 4) showed seasonal stress in the wetlands from diffuse agricultural sources (about 19% of the land use surrounding the wetlands is CC). The health of the wetlands depends on the buffering capacity provided by the contributing landscape (about 11% of the land use surrounding the wetlands is FO). The P maps (Fig. 5) showed

a sustained decrease in concentrations from 2022 to 2024, particularly during the winter months of 2023 and 2024. We believe that the reduction in P levels was caused by the source loading rather than a change in in-stream processing. The variable importance analysis supported this interpretation.

The distinct maps that we generated for N and P confirmed the utility of our framework beyond a mere prediction. The framework diagnosed two distinct issues that impacted the health of the BWG wetlands. First, a diffuse N problem that is tied to landscape-scale processes and, second, a source-driven P problem where a significant reduction in loading has yielded a clear ecological improvement.

## Methods

This study focuses on three small wetlands (total area is about 2.0 km<sup>2</sup>) managed by Beaver Creek Wetlands Greenway (BWG) Community Land Trust. Located in Beavercreek, Ohio, USA, the wetlands are protected by the Ohio Environmental Protection Agency (OEPA) Clean Water Act. Due to the variety of flora and fauna species, school groups, birdwatchers, and wildflower enthusiasts frequent the wetlands. The Beaver Creek stream runs through the wetlands from north to south. The BWG boundary and the perimeter of our study area are depicted in Fig. 7. According to NLCD land use, most urban areas (26.2%) of the BWG are in the east, while cultivated crops (38.2%) and forest areas (14.7%) predominate in the west.

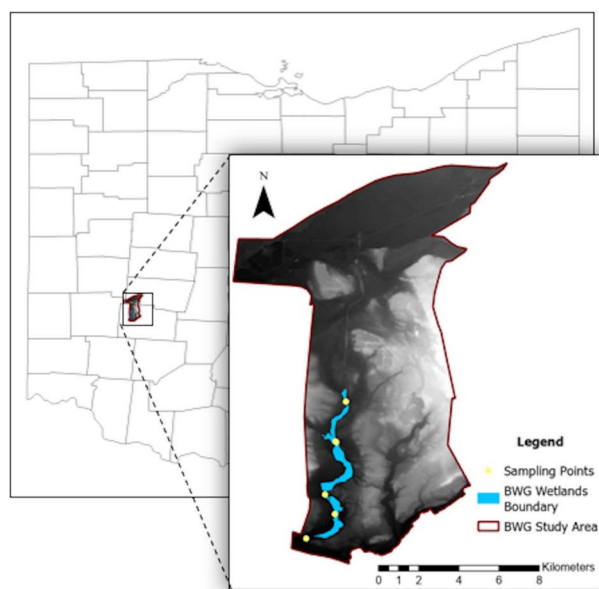
Figure 8 shows the two-stage hybrid modeling approach we used in the prediction and mapping of the BWG nutrients, N and P. Stage 1 models the baseline nutrient prediction using non-spatial wetland attributes. Next, stage 2 models the spatial error (residual) from the baseline model in stage 1 using satellite-derived environmental variables. This 2-step modeling strategy provides fine-scale spatial refinement to the final nutrient prediction and mapping. The main components of each stage include data collection, image processing, random forest modeling, and analysis/output. All major steps are performed within the Google Earth Engine (GEE) platform.

## Field sampling and laboratory analysis

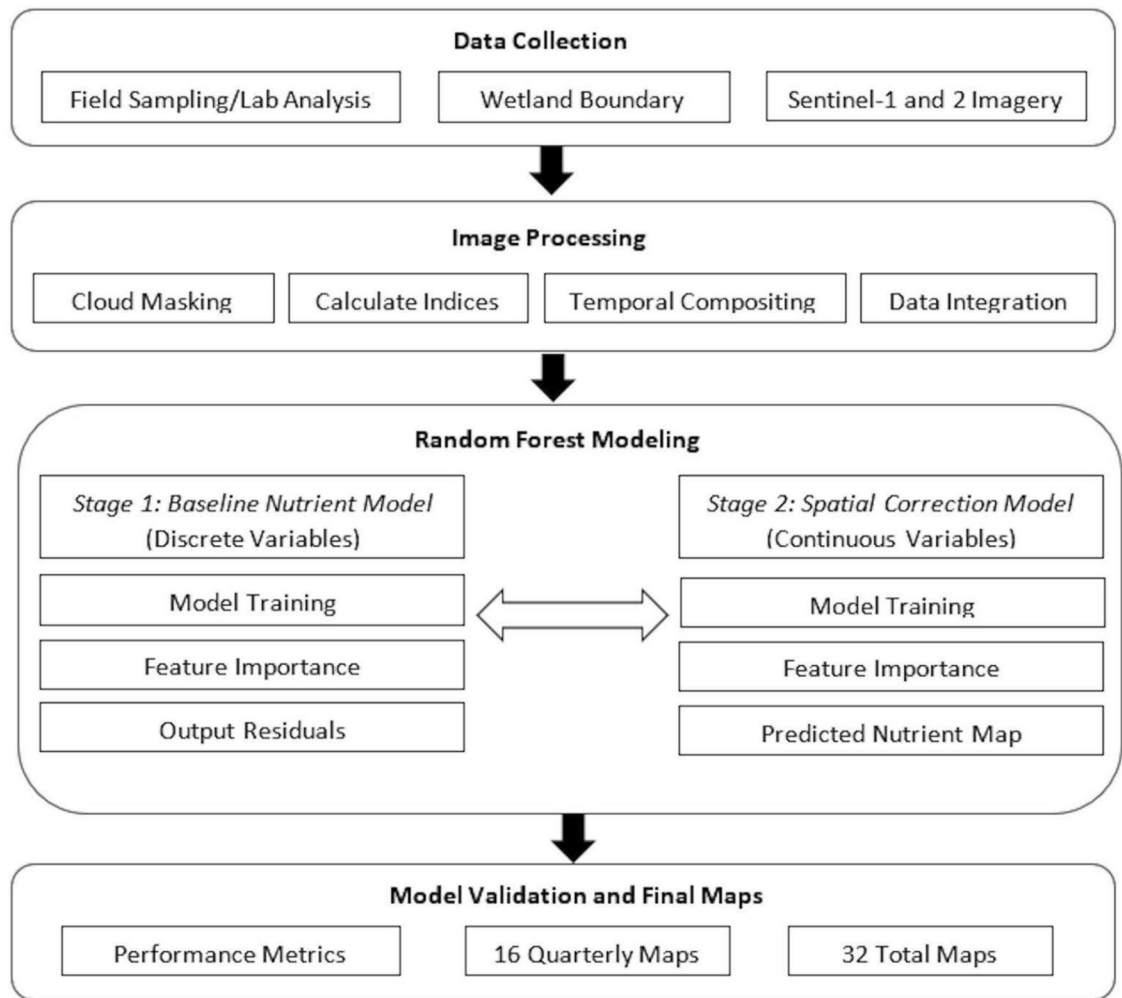
We selected wetland sampling locations based on initial reconnaissance surveys. We considered factors such as the manageable size of the three wetlands (< 1.0 km<sup>2</sup> each), their spatial distribution throughout the study area, ecological importance, and ease of accessibility for repeated field sampling. We then collected monthly water samples from representative locations within each wetland – inlet, middle, and outlet – to capture spatial variability in water quality parameters. We continuously collected monthly samples from 2021 to 2024 to capture seasonal nutrient fluctuations. During the 4-year study period, we obtained a total of 324 field water samples. In the laboratory, we filtered the samples and tested them for concentrations of nitrate (NO<sub>3</sub><sup>-</sup>) and orthophosphate (PO<sub>4</sub><sup>3-</sup>). Finally, we used the nutrient concentration values as input to the machine learning models to predict and spatially map the N and P distributions. The average concentrations of N and P are in Supplementary Appendices 2 and 3.

## Discrete variables and dataset

For the first stage of the modeling approach, we used several discrete characteristics of the wetlands as variable inputs to represent the physical, hydrological, and ecological properties of the wetland ecosystem (Table 1). We added these physical variables, namely, area and perimeter, to capture the basic morphometric characteristics that influence the cycle and retention capacity of nutrients<sup>36</sup>. We also calculated the Shape Complexity Index



**Fig. 7.** The BWG study area is in Beavercreek, Ohio, USA, showing the small-sized wetlands. This figure was created using ArcGIS Pro v.3.4.0 software by Esri (source: <https://www.esri.com/>).



**Fig. 8.** The four main components of the methodology include data collection, image processing, two-stage modeling using random forest, and evaluation/output. These components were performed within the Google Earth Engine (GEE) platform.

(SCI) and Perimeter-to-Area Ratio (PPS) to quantify the geometry of the wetland, as the configuration of the wetlands could impact on their ability to process nutrients<sup>37,38</sup>. We included climatic variables such as average temperature (AveTemp) and average rainfall (AveRain) since hydrological dynamics affects nutrient concentrations and these variables are considered high priorities in nutrient fluctuations in wetlands<sup>39</sup>.

In addition to morphological variables, we included landuse classes such as deciduous forest (DF), evergreen forest (EF), mixed forest (MF), shrub/scrub (SS), woody wetlands (WW), emergent herbaceous wetlands (EHW), herbaceous (HB), and hay/pasture (HP) to represent vegetation variables, different plant functional types, and specific nutrient uptake capacities<sup>40</sup>. To capture anthropogenic influences, we added cultivated crops (CC) and developed areas (DA) as both have a strong positive correlation with N and P nutrients and are considered sources of water pollutants<sup>41</sup>.

### Sentinel-1 and 2 satellite images, continuous variables and dataset

We used Sentinel-1 C-band Synthetic Aperture Radar (SAR) Ground Range Detected (GRD) data for our study period, available from the Sentinel Scientific Data Hub of the European Space Agency (ESA) (<https://scihub.copernicus.eu/>) to detect surface water quality<sup>42</sup>. We filtered the data to include only images captured in the Interferometric Wide Swath (IW) mode containing dual-polarization backscatter coefficients for VV (vertical transmit, vertical receive) and VH (vertical transmit, horizontal receive). To ensure radiometric consistency and stable viewing geometry, we selected a single orbital direction (either ascending or descending) based on the pass that offered more images. We converted the backscatter coefficients to a linear power scale for physical calculations.

From the selected SAR data, we derived the ratio of VV to VH, the simplified Radar Vegetation Index (SRVI), and the cross-polarization ratio (VH/VV) for each month of the study period (Table 2). SRVI is a robust measure of vegetation density that leverages the depolarization properties of volume scattering within a plant canopy<sup>43</sup>. We also produced a temporally averaged composite image by calculating the mean of all processed images. This

Variable	Description	Equation/Notes
<b>Continuous Variables</b>		
NDVI	Normalized Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$
NDWI	Normalized Difference Water Index	$\frac{Green - NIR}{Green + NIR}$
EVI	Enhanced Vegetation Index	$2.5 \times \frac{NIR - Red}{NIR + 6 \times Red - 7.5 \times Blue + 1}$
B2	Sentinel-2 Blue Band (490 nm)	B2
B3	Sentinel-2 Green band (560 nm)	B3
B4	Sentinel-2 Red band (665 nm)	B4
B5	Sentinel-2 Red Edge 1 band (705 nm)	B5
B6	Sentinel-2 Red Edge 2 band (740 nm)	B6
B7	Sentinel-2 Red Edge 3 band (783 nm)	B7
B8	Sentinel-2 Near-Infrared (NIR) band (842 nm)	B8
B8A	Sentinel-2 Narrow NIR band (865 nm)	B8A
B11	Sentinel-2 Shortwave Infrared 1 (SWIR 1) band (1610 nm)	B11
B12	Sentinel-2 Shortwave Infrared 2 (SWIR 2) band (2190 nm)	B12
VV	Sentinel-1 SAR Co-polarization (Vertical transmit, Vertical receive)	VV
VH	Sentinel-1 SAR Cross-polarization (Vertical transmit, Horizontal receive)	VH
VV/VH Ratio	Ratio of VV to VH backscatter	$\frac{VV}{VH}$
SRVI	Simplified Radar Vegetation Index	$\frac{4(aVV)}{aVV + aVH}$
Cross Ratio	Cross-polarization ratio (VH / VV)	$\frac{VH}{VV}$
<b>Discrete Variables</b>		
SCI	Shape Complexity Index	$\frac{Perimeter}{Area}$
Area	Area for each wetland	-
Perimeter	Perimeter for each wetland	-
PPS	Polsby-Popper Score	$\frac{4\pi \times Area}{Perimeter^2}$
Ave Rain	Average 7-day precipitation for each wetland in mm	-
Ave Temp	Average 7-day temperature for each wetland in °F	-
OP	Open Water	%OP
DA	Developed Open Space	%DOS
DA	Developed Low Intensity	%DLI
DA	Developed Medium Intensity	%DMI
DA	Developed High Intensity	%DHI
BL	Barren Land	%BL
FO	Deciduous Forest	%DF
FO	Evergreen Forest	%EF
FO	Mixed Forest	%MF
SS	Shrub/Scrub	%SS
HB	Herbaceous	%HB
HP	Hay/Pasture	%HP
CC	Cultivated Crops	%CC
WW	Woody Wetlands	%WW
EHW	Emergent Herbaceous Wetlands	%EHW

**Table 2.** List of the discrete and continuous variables used in the RF model, which consists of spectral indices, morphological properties, environmental factors, landuse classes, and the individual Sentinel-2 bands. The aVH represents the backscatter coefficient from the cross-polarization channel (vertical transmit, horizontal receive) in a linear power scale. The aVV represents the backscatter coefficient from the co-polarization channel (vertical transmit, vertical receive) in a linear power scale.

new average image serves to reduce speckle noise and provides a representative measure of surface conditions during the study period. We utilized the final SAR-derived bands and indices<sup>44</sup> as spatial predictors in stage 2 of our model (Fig. 8).

For Sentinel-2, we used the Level-2A (L2A) surface reflectance collection. Sentinel-2 has 13 spectral bands: four bands at 10 m resolution, six bands at 20 m resolution and three bands at 60 m spatial resolution. We resampled all bands that were 20 m resolution to 10 m to maintain consistency with the four native bands (B2

in blue, B3 in green, B4 in red, and B8 in NIR). We then computed the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and the Normalized Difference Water Index (NDWI). We based the selection of the spectral indices from previous studies that have applied ML to remote sensing data to predict nutrients for surface waters<sup>45,46</sup>. We also produced a cloud-reduced composite image by computing the per-pixel average of the entire time series. This composite, including the original spectral bands – band 2 (B2, 490 nm), band 3 (B3, 560 nm), band 4 (B4, 665 nm), band 5 (B5, 705 nm), band 6 (B6, 740 nm), band 7 (B7, 783 nm), band 8 (B8, 842 nm), band 8A (B8A, 865 nm), band 9 (B9, 945 nm), band 11 (B11, 1610 nm), and band 12 (B12, 2190 nm) – and the derived indices, represented the continuous variables from Sentinel-2.

When images (Sentinel-1 and Sentinel-2) were combined, GEE resampled the bands to a common projection and scale. This resampling process forced the Sentinel-1 data to align with the data grid of Sentinel-2 using the 10 m resolution. All of this image integration occurred at the feature level before the modeling stage. A complete list of continuous variables is found in Table 2. We used these variables as spatial predictors in the stage 2 model, in addition to the SAR-derived variables (Fig. 8).

### Random forest modeling

We utilized the Random Forest (RF) as the ML algorithm to predict and map nutrient concentrations<sup>22,46,47</sup>. RF is a popular model for predicting wetlands and other surface waters, including water quality parameters, using remote sensing data<sup>38,48</sup>. RF provides a way to select important covariates based on changes in prediction accuracy when variables are added or deleted from models. It is a non-parametric supervised classifier that uses the Classification and Regression Tree (CART) through bagging, where it randomly picks a set of features and creates a classifier with a bootstrapped sample of the training data to grow a tree<sup>49</sup>. With RF training data selection, it is possible that the same sample could be picked several times, whereas others may not be picked at all. Apart from RF being quite robust with highly collinear variables, the random selection process at each tree node causes low correlation among trees and avoids overfitting<sup>50</sup>.

We guaranteed a precise alignment of the variables before applying the RF model. For temporal alignment, first, we grouped the in-situ nutrient samples for that particular quarter (e.g., Q1-2021: January 1 through March 31). Next, we used all available Sentinel-1 and Sentinel-2 images that were collected in the same quarter to generate one mean composite image. This part of our methodology was crucial because (1) it provides a strong representation of the average environmental conditions during the period when the samples were taken, and (2) it minimizes the impact of atmospheric conditions or day-to-day variability, thus, our process makes a more stable set of predictors. We ensured that every point sample was temporally matched to the mean environmental state of its corresponding quarter. In terms of spatial alignment, our model extracted the corresponding pixel values from the quarterly mean composite image for each in-situ sample point through its geographic coordinates. This part of our model ensured that each sample point was precisely linked to a single vector of predictor values that represents the pixel area where the sample was collected.

#### *Two-stage hybrid RF modeling framework*

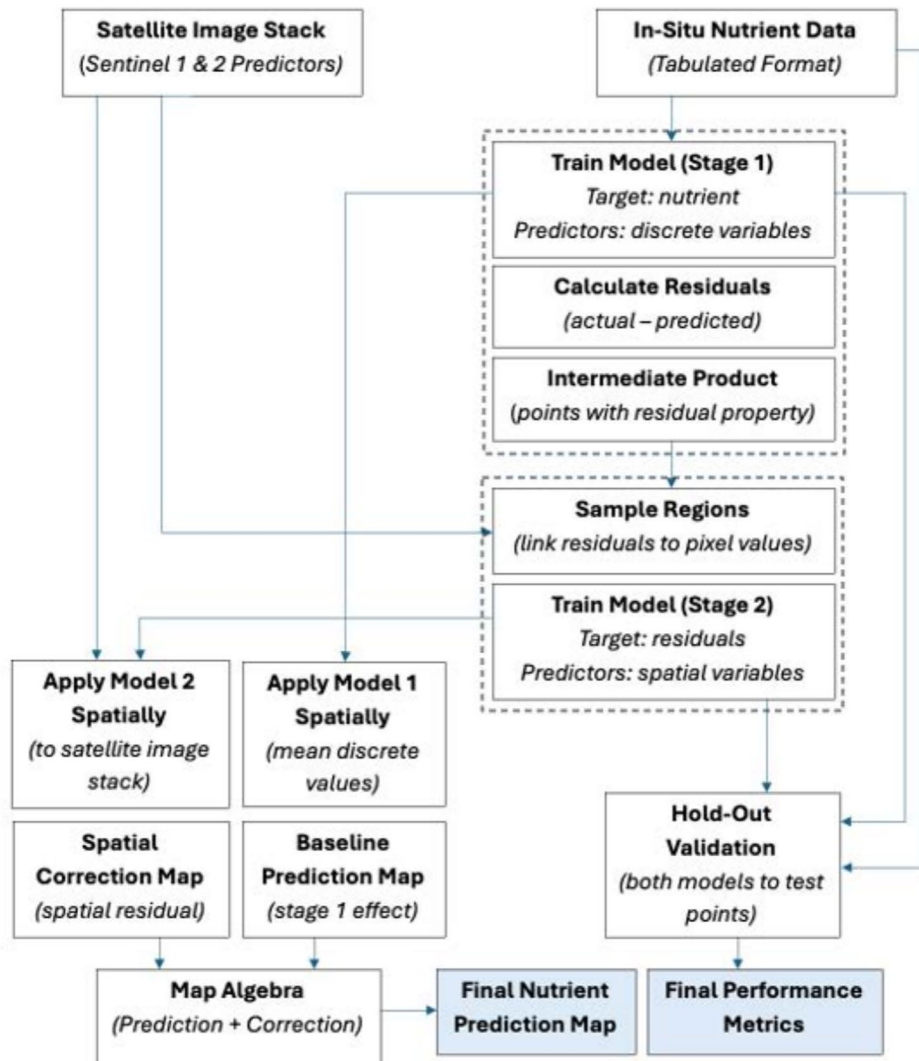
In this study, we employed a two-stage hybrid RF regression framework designed to combine discrete, site-specific variables (stage 1) with continuous and spatially explicit RS data (stage 2). The first stage used an RF model to establish a baseline nutrient prediction as a function of the discrete wetland characteristics (e.g., area, shape, climate). The model was trained using 70% of the in-situ data points. The output of the stage 1 model are the residuals or the difference between observed and predicted values. These values captured the variance not explained by the site-level variables. Next, we used these residuals as explicit target variables for the next stage. In stage 2, we employed another independent RF model to explain the residual variance using the continuous predictors derived from Sentinel-1 SAR and Sentinel-2 imagery. For mapping the spatial distribution of the nutrients, the stage 1 model generated a general baseline prediction map. In contrast, the stage 2 model produced a detailed correction map on a per-pixel basis. Finally, we generated a nutrient prediction map by summing the baseline prediction and the spatial correction maps using map algebra. This last step of adding maps created a product that uses both the general trends from physical site data and the localized environmental variations captured by RS. Figure 9 shows a more detailed diagram of the two-stage modeling framework.

#### *Model validation*

To quantitatively assess the predictive performance of our model, we performed a validation on an independent hold-out dataset. We formed the test set by randomly partitioning 30% of the original data points, and the rest was used to train or adjust the model. This approach guaranteed that the performance of our model is tested on a set of data that was not used during the training phase. We evaluated the performance of the combined two-stage model using a list of statistical metrics<sup>51</sup>: (1) Coefficient of Determination ( $r^2$ ) to quantify the proportion of the variance in the observed nutrient concentrations that is explained by the model; (2) Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) to measure the magnitude of prediction error; and (3) Mean Bias Error (MBE) to diagnose any systematic tendency of the model to over-predict (positive bias) or under-predict (negative bias). This set of metrics ensures a general understanding of the performance of our model and how the model captured the goodness-of-fit, magnitude of error, and directional bias<sup>52</sup>.

### Google Earth engine

We ran the hybrid RF framework through the GEE platform (earthengine.google.com). GEE is a cloud-based platform that provides researchers with a vast collection of satellite imagery and geospatial datasets<sup>53</sup>. It offers a comprehensive and diverse range of data sources, including Sentinel-1 SAR and Sentinel-2 Level-2A (L2A) images. The extensive GEE data archive, coupled with its efficient data storage and processing capabilities, allowed this research seamless access to the necessary image data for analysis. Using the power of GEE, we retrieved



**Fig. 9.** The detailed two-stage RF hybrid modeling framework, illustrating the integration of discrete data (stage 1) and satellite imagery (stage 2) to generate a final nutrient prediction map.

and preprocessed the images for all the months and years required for this study. The use of GEE ensured a streamlined and efficient workflow, enabling us to take advantage of its vast data resources and advanced analysis capabilities for accurate nutrient mapping<sup>54</sup>.

### Implications

The findings of this study have implications for wetland management and nutrient cycling studies. We demonstrated that our model could predict N and P nutrients using a dynamic, robust, and spatially comprehensive two-stage approach. The quarterly nutrient mapping provides valuable information for the rapid identification of hotspots and assessment of management interventions.

The ability of our model to distinguish between N and P drivers could help target management strategies in wetlands. For instance, to improve P concentrations, specific point sources have to be addressed, urban runoff policies have to be implemented, or buffer zones have to be established around DA to reduce direct runoff inputs. For N management, landscape-scale implementation of agricultural best management practices is required. One practice that can be implemented is cover cropping and nutrient management plans in surrounding farm fields.

From a scientific perspective, this study provides a framework for integrating diverse data sources to model complex environmental processes. The two-stage hybrid framework enhances ecological interpretability through the quantitative separation of localized biophysical processes and broad source patterns. Our modeling approach demonstrates high transferability and could be adapted to model additional water quality parameters, such as total suspended solids. Lastly, our framework does not only apply for current wetland monitoring. It also has the potential to assess the impacts of future development on wetland water quality.

## Caveats and future research directions

We recognize that our two-stage hybrid framework has inherent limitations. Below, we list these limitations so that we can set directions for further investigation.

- (1) The BWG is small and has unique hydrological and biogeochemical conditions, which were used to calibrate the model. To apply the model to other wetland systems with larger areas, different climates, geologies, or percentage of land-use contributions, local training data and recalibration are needed to ensure precision.
- (2) Our model is useful for capturing seasonal trends and is based on the quarterly temporal composites of the satellite image. However, this could mitigate the effects of high intensity short-term events, such as large storms, which could affect nutrient transport. To better capture these event-based dynamics, future research could utilize high-temporal monthly data.
- (3) Our model is only sensitive to surface conditions. RS-based models do not have the capability to measure processes that occur deeper in the water column. A future study is necessary to understand these subsurface processes through the integration of our current framework with process-based hydrological models.
- (4) The level or depth of the water is a factor in changes in nutrient concentrations in wetlands through dilution and concentration mechanisms. For future studies, we recommend adding in situ water level data as one of the input variables.
- (5) Lastly, although our model successfully highlighted the factors that influence nutrient change, the increase in nutrient levels in some seasons needs a focused examination to determine the root cause. In order to transform our predictive framework into a comprehensive and globally applicable tool for wetland monitoring, it will be important to address these limitations.

## Conclusion

This study successfully created and validated a two-stage hybrid modeling framework for high-resolution spatiotemporal mapping of nutrient concentrations in wetland ecosystems. The predictive accuracy of the model for both nutrients (N and P) highlighted its ability to integrate continuous satellite imagery with discrete site-level datasets. Being multi-seasonal, our framework allowed us to move beyond a single static model. Our results showed which seasons are the most suitable for remote sensing-based nutrient monitoring. Also, our analysis identified divergent biogeochemical pathways for each nutrient. Source loading from DA landuse class dominated P concentrations, while seasonal agricultural pressures and in-stream processing affected N dynamics. Finally, this study offers a strong diagnostic framework that could inform focused wetland management strategies and monitor the health and function of wetland ecosystems.

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 3 August 2025; Accepted: 13 February 2026

Published online: 19 February 2026

## References

1. Zhang, X. et al. Global annual wetland dataset at 30m with a fine classification system from 2000 to 2022. *Sci. Data* **11**, 310. <https://doi.org/10.1038/s41597-024-03143-0> (2024).
2. Zhang, S. et al. Investigating the dynamic change and driving force of isolated marsh wetland in Sanjiang plain, northeast China. *Land* **13**, 1969. <https://doi.org/10.3390/land13111969> (2024).
3. Guelmami, A. Large-scale mapping of existing and lost wetlands: Earth observation data and tools to support restoration in the sebou and medjerda river basins. *Euro-Mediterr. J. Environ. Integr.* **9**, 169–182. <https://doi.org/10.1007/s41207-023-00443-6> (2024).
4. Howard-Williams, C. Cycling and retention of nitrogen and phosphorus in wetlands: a theoretical and applied perspective. *Freshw. Biol.* **15**, 391–431. <https://doi.org/10.1111/j.1365-2427.1985.tb00212.x> (1985).
5. Lacher, I. L. et al. Scale-dependent impacts of urban and agricultural land use on nutrients, sediment, and runoff. *Sci. Total Environ.* **652**, 611–622. <https://doi.org/10.1016/j.scitotenv.2018.09.370> (2019).
6. Fisher, J. & Acreman, M. C. Wetland nutrient removal: a review of the evidence. *Hydrol. Earth Syst. Sci.* **8**, 673–685. <https://doi.org/10.5194/hess-8-673-2004> (2004).
7. Jackson, C. R., Thompson, J. A. & Kolka, R. K. Wetland soils, hydrology, and geomorphology. In *Ecology of Freshwater and Estuarine Wetlands* (eds Batzer, D. P. & Sharitz, R. R.) 23–60 (University of California Press, Oakland, CA, USA, 2014).
8. Little, A. Sampling and analyzing wetland vegetation. In *Wetland Techniques* 273–324 (Springer, Dordrecht, 2013). [https://doi.org/10.1007/978-94-007-6860-4\\_5](https://doi.org/10.1007/978-94-007-6860-4_5).
9. Salari, A., Zakaria, M., Nielsen, C. C. & Boyce, M. S. Quantifying tropical wetlands using field surveys, spatial statistics and remote sensing. *Wetlands* **34**, 565–574. <https://doi.org/10.1007/s13157-014-0524-3> (2014).
10. Gallant, A. L. The challenges of remote monitoring of wetlands. *Remote Sens.* **74**, 10938–10950. <https://doi.org/10.3390/rs70810938> (2015).
11. Corcoran, J. M., Knight, J. F. & Gallant, A. L. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern minnesotas. *Remote Sens.* **5**, 3212–3238. <https://doi.org/10.3390/rs5073212> (2013).
12. Drisiya, J. & Shibu, K. Assessment of nutrient dynamics in Sasthamkotta wetland using gis remote sensing and artificial neural network. *Trans. Indian Natl. Acad. Eng.* **10**, 453–470. <https://doi.org/10.1007/s41403-025-00528-4> (2025).
13. Ding, C. et al. Combining artificial neural networks with causal inference for total phosphorus concentration estimation and sensitive spectral bands exploration using modis. *Water* **12**, 2372. <https://doi.org/10.3390/w12092372> (2020).
14. Gaballah, M. S. & Lammers, R. W. Assessing the factors influencing constructed wetland performance for mitigating agricultural nutrient runoff in the U.S.. *J. Water Process Eng.* **71**, 107293. <https://doi.org/10.1016/j.jwpe.2025.107293> (2025).
15. Singh, S. et al. Machine learning application for nutrient removal rate coefficient analyses in horizontal flow constructed wetlands. *ACS ES T Water* **4**, 2619–2631. <https://doi.org/10.1021/acsestwater.4c00121> (2024).

16. Isles, P. D. F. A random forest approach to improve estimates of tributary nutrient loading. *Water Res.* **248**, 120876. <https://doi.org/10.1016/j.watres.2023.120876> (2024).
17. Ayele, G. T., Yu, B., Bruere, A. & Hamilton, D. P. Response of streamflow and nutrient loads in a small temperate catchment subject to land use change. *Environ. Monit. Assess.* **195**, 1418. <https://doi.org/10.1007/s10661-023-11828-z> (2023).
18. Zhu, B. et al. Satellite remote sensing of water quality variation in a semi-enclosed bay (Yueqing bay) under strong anthropogenic impact. *Remote Sens.* **14**, 550. <https://doi.org/10.3390/rs14030550> (2022).
19. Euliss, N. H. J. et al. Placing prairie pothole wetlands along spatial and temporal continua to improve integration of wetland function in ecological investigations. *J. Hydrol.* **513**, 490–503. <https://doi.org/10.1016/j.jhydrol.2014.04.006> (2014).
20. Guo, B. et al. Using a two-stage scheme to map toxic metal distributions based on gf-5 satellite hyperspectral images at a northern chinese opencast coal mine. *Remote Sens.* **14**, 5804. <https://doi.org/10.3390/rs14225804> (2022).
21. Tarantino, C. et al. Intra-annual sentinel-2 time-series supporting grassland habitat discrimination. *Remote Sens.* **13**, 277. <https://doi.org/10.3390/rs13020277> (2021).
22. Xue, J., Yuan, C., Ji, X. & Zhang, M. Predictive modeling of nitrogen and phosphorus concentrations in rivers using a machine learning framework: a case study in an urban-rural transitional area in wenzhou china. *Sci. Total Environ.* **910**, 168521. <https://doi.org/10.1016/j.scitotenv.2023.168521> (2024).
23. Giri, S. & Qui, Z. Understanding the relationship of land uses and water quality in twenty first century: A review. *J. Environ. Manag.* **173**, 41–48. <https://doi.org/10.1016/j.jenvman.2016.02.029> (2016).
24. Wang, Y., Sun, J. & Wu, Y. Spatiotemporal variation of small and micro wetlands and their multiple responses to driving factors in the high-latitude region. *Wetlands* **44**, 126. <https://doi.org/10.1007/s13157-024-01882-9> (2024).
25. Nijjati, N., Rusuli, Y., Maimaitiaili, K. & Kuluwan, Y. Spatiotemporal variations and driving forces analysis of ecosystem health in the Bosten lake wetland in China. *Land Degrad. Dev.* <https://doi.org/10.1002/ldr.5686> (2025).
26. Gaballah, M. S., Yousefyani, H., Karami, M. & Lammers, R. W. Free water surface constructed wetlands: review of pollutant removal performance and modeling approaches. *Environ. Sci. Pollut. Res.* **31**, 44649–44668. <https://doi.org/10.1007/s11356-024-34151-7> (2024).
27. Voli, M. T. et al. Fingerprinting the sources of suspended sediment delivery to a large municipal drinking water reservoir: Falls lake, neuse river, north carolina, usa. *J. Soils Sediments* **13**, 1692–1707 (2013).
28. Arhonditsis, G. B., Javed, A., Saber, A., Neumann, A. & Arnillas, C. A. Characterization of biogeochemical cycles in agricultural watersheds: Integrating regional assessment with downstream water quality. *SSRN* **13**, 52. <https://doi.org/10.2139/ssrn.5249940> (2025).
29. Carvalho, W. S., Filho, F. J. C. M., Rodrigues, L. R. & Calheiros, C. S. C. Influence of land use and land cover on the quality of surface waters and natural wetlands in the Miranda river watershed, Brazilian Pantanal. *Appl. Sci.* **14**, 5666. <https://doi.org/10.3390/app14135666> (2024).
30. Kumwimba, M. N. et al. Nutrient and sediment retention by riparian vegetated buffer strips: impacts of buffer length, vegetation type, and season. *Agric. Ecosyst. Environ.* **369**, 109050. <https://doi.org/10.1016/j.agee.2024.109050> (2024).
31. Sendrowski, A., Castañeda-Moya, E., Twilley, R. & Passalacqua, P. Biogeochemical and hydrological variables synergistically influence nitrate variability in coastal deltaic wetlands. *J. Geophys. Res.: Biogeosci.* **126**, e2020JG005737. <https://doi.org/10.1029/2020JG005737> (2021).
32. Hogan, D. & Walbridge, M. Urbanization and nutrient retention in freshwater riparian wetlands. *Ecol. Appl.* **17**, 1142–1155. <https://doi.org/10.1890/06-0185> (2007).
33. Hobbie, S. et al. Contrasting nitrogen and phosphorus budgets in urban watersheds and implications for managing urban water pollution. *Proc. Natl. Acad. Sci.* **114**, 4177–4182. <https://doi.org/10.1073/pnas.1618536114> (2017).
34. Allan, J. D., Castillo, M. M. & Capps, K. A. Nutrient Dynamics. In *Stream Ecology* (Springer, Cham, 2021).
35. Jin, X. D., He, Y. L., Kirumba, G., Hassan, Y. & Li, J. B. Phosphorus fractions and phosphate sorption-release characteristics of the sediment in the Yangtze river estuary reservoir. *Ecol. Eng.* **55**, 62–66. <https://doi.org/10.1016/j.ecoleng.2013.02.001> (2013).
36. Hou, Y., Zhao, G., Chen, X. & Yu, X. Improving satellite retrieval of coastal aquaculture pond by adding water quality parameters. *Remote Sens.* **14**, 3306. <https://doi.org/10.3390/rs14143306> (2022).
37. Kadlec, R. H. The effects of wetland vegetation and morphology on nitrogen processing. *Ecol. Eng.* **33**, 126–141. <https://doi.org/10.1016/j.ecoleng.2008.02.012> (2008).
38. Zhang, J., Liu, X., Qin, Y., Fan, Y. & Cheng, S. Wetlands mapping and monitoring with long-term time series satellite data based on google earth engine, random forest, and feature optimization: A case study in gansu province, china. *Land* **13**, 1527. <https://doi.org/10.3390/land13091527> (2024).
39. Kull, A., Kull, A., Jaagus, J., Kuusemets, V. & Mander, U. The effects of fluctuating climatic conditions and weather events on nutrient dynamics in a narrow mosaic riparian peatland. *Boreal Environ. Res.* **13**, 243–263 (2008).
40. Venkateswarlu, T. & Anmala, J. Importance of land use factors in the prediction of water quality of the upper green river watershed, Kentucky, USA, using random forest. *Environ. Dev. Sustain.* **26**, 23961–23984. <https://doi.org/10.1007/s10668-023-03630-1> (2024).
41. Weng, M. et al. Exploring the impact of land use scales on water quality based on the random forest model: A case study of the Shaying river basin, China. *Water* **16**, 420. <https://doi.org/10.3390/w16030420> (2024).
42. Kaplan, G. & Avdan, U. Sentinel-1 and sentinel-2 data fusion for wetlands mapping: Balıkdami, turkey. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3*, 729–734, <https://doi.org/10.5194/isprs-archives-XLII-3-729-2018> (2018).
43. Gonenc, A., Ozerdem, S. & Acar, E. Comparison of ndvi and rvi vegetation indices using satellite images. In *International Conference on Agro-Geoinformatics*, 1–4. <https://doi.org/10.1109/Agro-Geoinformatics.2019.8820225> (Istanbul, Turkey, 2019).
44. Jafarzadeh, H., Mahdianpari, M., Gill, E. W. & Mohammadimanesh, F. Enhancing wetland mapping: Integrating sentinel-1/2, gedi data, and google earth engine. *Sensors* **24**, 1651. <https://doi.org/10.3390/s24051651> (2024).
45. Behrouz, M. S., Yazdi, M. N. & Sample, D. J. Using random forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff. *J. Environ. Manag.* **317**, 115412. <https://doi.org/10.1016/j.jenvman.2022.115412> (2020).
46. Harrison, J. W., Lucius, M. A., Farrell, J. L., Eichler, L. W. & Relyea, R. A. Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using random forests regression. *Sci. Total Environ.* **763**, 143005. <https://doi.org/10.1016/j.scitotenv.2020.143005> (2021).
47. Bhattarai, A., Dhakal, S., Gautam, Y. & Bhattarai, R. Prediction of nitrate and phosphorus concentrations using machine learning algorithms in watersheds with different landuse. *Water* **13**, 3096. <https://doi.org/10.3390/w13213096> (2021).
48. Wang, X., Jiang, W., Peng, K., Li, Z. & Rao, P. A framework for fine classification of urban wetlands based on random forest and knowledge rules: taking the wetland cities of haikou and yinchuan as examples. *GIScience Remote Sens.* **59**, 2144–2163. <https://doi.org/10.1080/15481603.2022.2152926> (2022).
49. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
50. Salas, E. A. L., Subburayalu, S. K., Partee, E. B., Willis, L. P. & Mitchell, K. Potential of mapping dissolved oxygen in the little Miami river using sentinel-2 images and machine learning algorithms. *Remote Sens. Appl. Soc. Environ.* **26**, 100759. <https://doi.org/10.1016/j.rsase.2022.100759> (2022).
51. Ohijeagbon, O., Waheed, A. & Ajayi, O. Machine learning-based solar radiation forecasting models for enhancing renewable energy integration in Nigeria. Available at SSRN: <https://ssrn.com/abstract=5200160>, <https://doi.org/10.2139/ssrn.5200160> (2025).

52. Santos, V. O., Rocha, P. A. C., Thé, J. V. G. & Gharabaghi, B. Evaluation of machine learning methods for forecasting turbidity in river networks using sentinel-2 remote sensing data. *Ecol. Inform.* **90**, 103313. <https://doi.org/10.1016/j.ecoinf.2025.103313> (2025).
53. Tamiminia, H. et al. Google earth engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.* **164**, 152–170. <https://doi.org/10.1016/j.isprsjprs.2020.04.001> (2020).
54. Merchant, M. et al. Leveraging google earth engine cloud computing for large-scale arctic wetland mapping. *Int. J. Appl. Earth Obs. Geoinf.* **125**, 103589. <https://doi.org/10.1016/j.jag.2023.103589> (2023).

### Author contributions

E.A.L.S. conceptualized the study. E.A.L.S. and K.S. conducted the experiments. E.A.L.S. and S.S.K. provided inputs for methodology, software, and resources. S.S.K. project administration. E.A.L.S., K.S., and R.B. conducted the field survey and experiments. E.A.L.S. and K.S. analyzed the results. R.B. and K.S. conducted the laboratory experiments. All authors read and approved the final manuscript.

### Funding

This research was supported by the Evans-Allen funds of the U.S. Department of Agriculture, National Institute of Food and Agriculture [NI241445XXXXG004].

### Declarations

### Competing interests

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.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-40585-5>.

**Correspondence** and requests for materials should be addressed to E.A.L.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026