Supplementary materials including tables, charts, Mapas and analysis results

Table S1: Sentinel-2 scenes processed and selected for the six in situ sampling campaigns

In situ sampling campaign	Sampling period	Sentinel-2 scenes processed/screened	Selected scenes
Aug-16	Aug-16	5	2
Dec-16	Dec-16	7	3
Mar-17	Mar-17	8	3
Oct-21	Oct-21	8	3
Apr-22	Apr-22	7	3
Dec-25	Dec-25		4
Total	**Six campaigns**	**44**	**18**

Table S2: Sensitivity Analysis of RF based Model Performance Across Window Sizes for best window size determination for Lake Tana Water Quality study

Water Quality Parameters	3X3 window			5X5window			7X7window			Best window (Higher R^2)	Best window (Lowest RMSE)	Best window (Lowest MARE)
	R^2	RMSE	MARE	R^2	RMSE	MARE	R^2	RMSE	MARE			
Chlorophyll a(Ug/l)	0.50	12.20	0.65	0.80	7.80	0.41	0.72	9.20	0.58	5x5	5X5	5X5
Turbidity (NTU)	0.63	8.10	0.81	0.60	10.20	0.87	0.58	9.50	0.89	3x3	3X3	3X3
Total nitrogen(mg/l)	0.69	7.10	0.64	0.76	6.50	0.53	0.61	12.50	0.72	5x5	5X5	5X5
Total phosphorus(mg/l)	0.72	5.20	0.60	0.77	4.80	0.54	0.70	5.30	0.56	5x5	7X7	5X5
Average (all parameters)	**0.64**	**8.15**	**0.68**	**0.73**	**7.33**	**0.59**	**0.65**	**9.13**	**0.69**	**5X5**	**5X5**	**5X5**

Table S3: Initial(raw) candidates proposed for feature selection of the four water quality parameters

Water quality parameters	Candidate predictors
Chlorophyll a, Turbidity, TN and TP	NDTI,NDCI,MCI,FAI,NDGI,NDVI, NDWI, MNDWI,B2B3B4,B2B3B8, STI, B8B7, B8B6, B8B5, B8B4, B8B3, B8B2, B8B11,B7B8,B7B6, B7B5, B7B4, B7B3, B7B2, B7B11,B6B8, B6B7, B6B5, B6B4, B6B3, B6B2, B6B11,B5B8, B5B7, B5B4, B5B3, B5B2, B5B11,B4B8, B4B7, B4B6, B4B5, B4B2, B4B11, B3B8, B3B7, B3B6, B3B5, B3B4, B3B2, B3B11,B2B8, B2B7, B2B6, B2B5, B2B4, B2B3, B2B11, B11B8,B11B7,B11B6,B11B5,B11B4,B11B3,B11B2,B2,B3,B4,B5,B6,B7,B8, B11

Table S4: Getis-Ord GI* based clustering of parameter for both dry and weight seasons and their Jaccard hotspot index

Season	Pair	Kappa	Interpretation	N Pixels	Jaccard Hotspot	Jaccard Cold Spot	Subset Kappa excl double NotSig	Subset Kappa Interpretation	N Pixels Subset
Dry	TU-TN	0.99	Almost Perfect	1847910	0.99	0.99	0.98	Almost Perfect	1279396
Dry	TU-TP	0.99	Almost Perfect	1847910	1.00	0.99	0.99	Almost Perfect	1277679
Dry	TN-TP	0.99	Almost Perfect	1847910	0.99	0.98	0.97	Almost Perfect	1282578
Wet	TU-TN	0.95	Almost Perfect	1860020	0.96	0.96	0.92	Almost Perfect	1426129
Wet	TU-TP	0.95	Almost Perfect	1860020	0.96	0.95	0.92	Almost Perfect	1430763
Wet	TN-TP	0.97	Almost Perfect	1860020	0.97	0.97	0.94	Almost Perfect	1418412

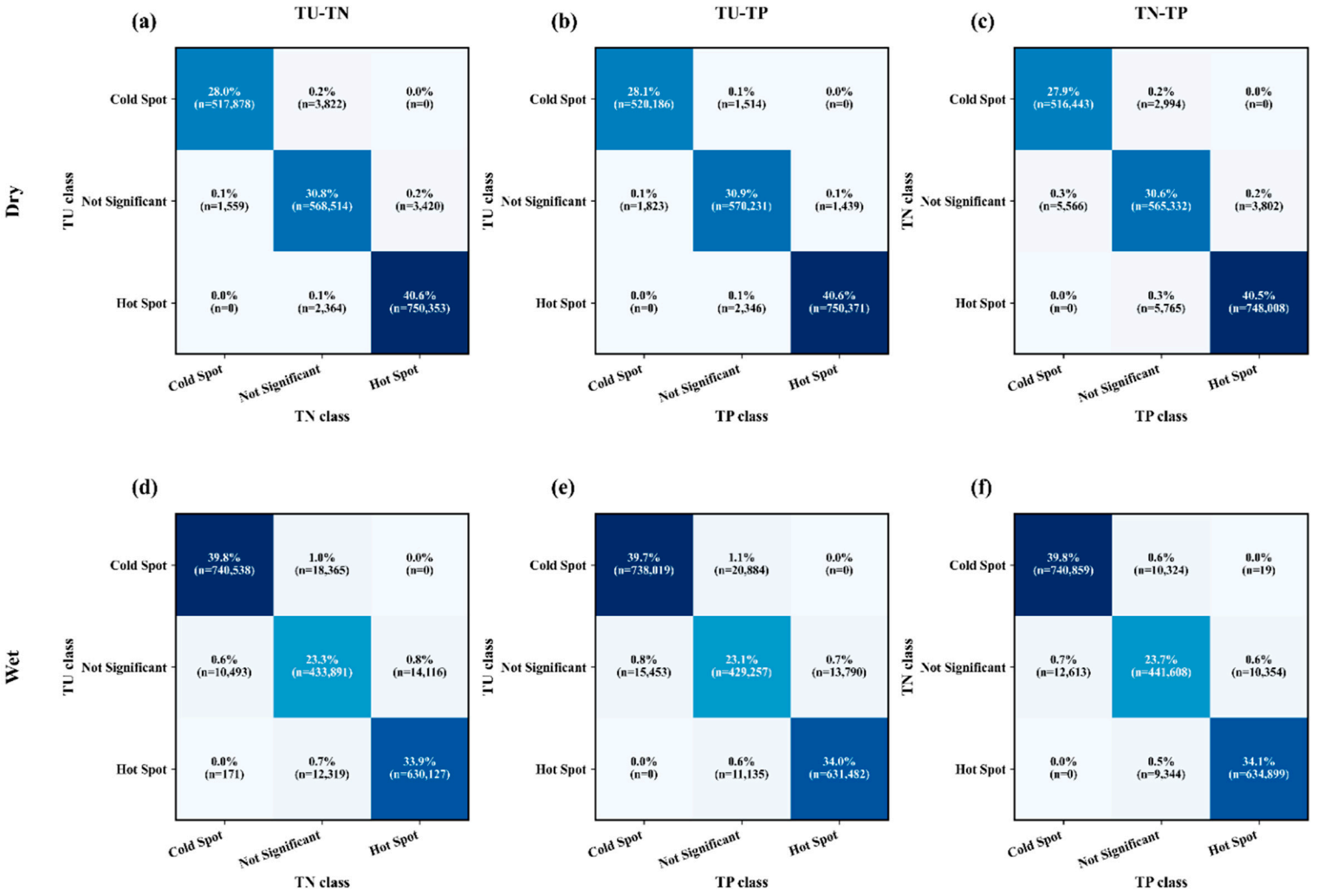


Figure S5: A 3x3 confusion matrix for the dry(a-c) and wet (d-f) seasons: TU-TN, TN-TP and TU-TP respectively for the Getis-Ord GI* hotspot clustering for Lake Tana

Table S6: Paired t-test comparing water quality parameter values between December 2016 and December 2025, Lake Tana.

Parameter	Unit	Mean ± SD (Dec 2016)	Mean ± SD (Dec 2025)	Mean Difference (2025 − 2016)	t-statistic	p-value
Chl-a	ug/L	16.00 ± 8.40	27.80 ± 8.40	11.80	16.15	< 0.001***
TN	mg/L	2.00 ± 0.60	2.60 ± 0.80	0.60	20.10	< 0.001***
TP	mg/L	0.70 ± 0.10	0.80 ± 0.10	0.10	14.25	< 0.001***

*Note: Values are pixel-paired means ± standard deviation for each survey date. Mean Difference is Dec 2025 minus Dec 2016. All comparisons were statistically significant. Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.*

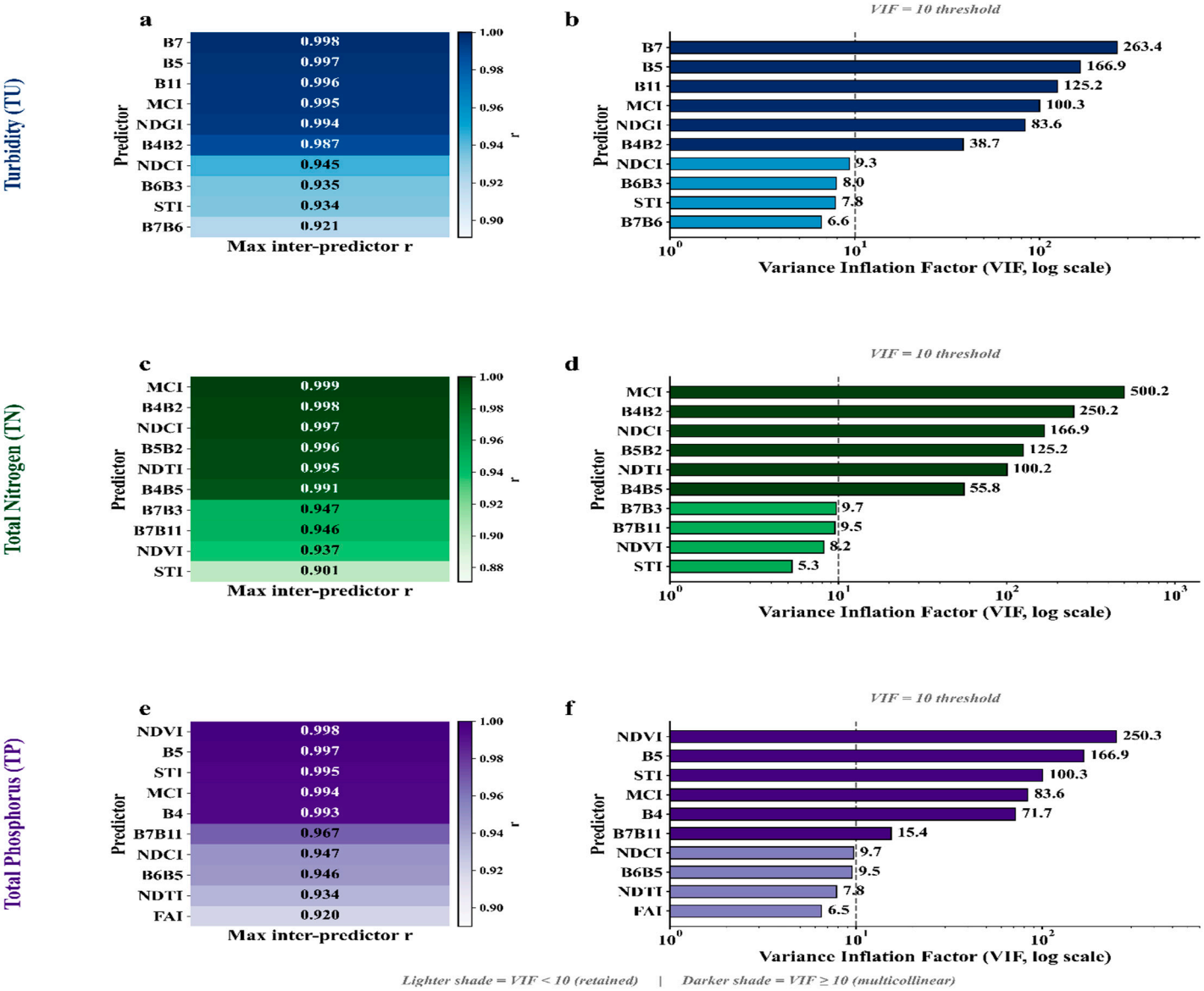


Figure S7. SHAP- and VIF-based diagnostics of the final predictors for turbidity (TU), total nitrogen (TN), and total phosphorus (TP). Columns show VIF values (a, e, i), SHAP effects (b, f, j), SHAP interaction values (c, g, k), and relative feature importance (d, h, l).

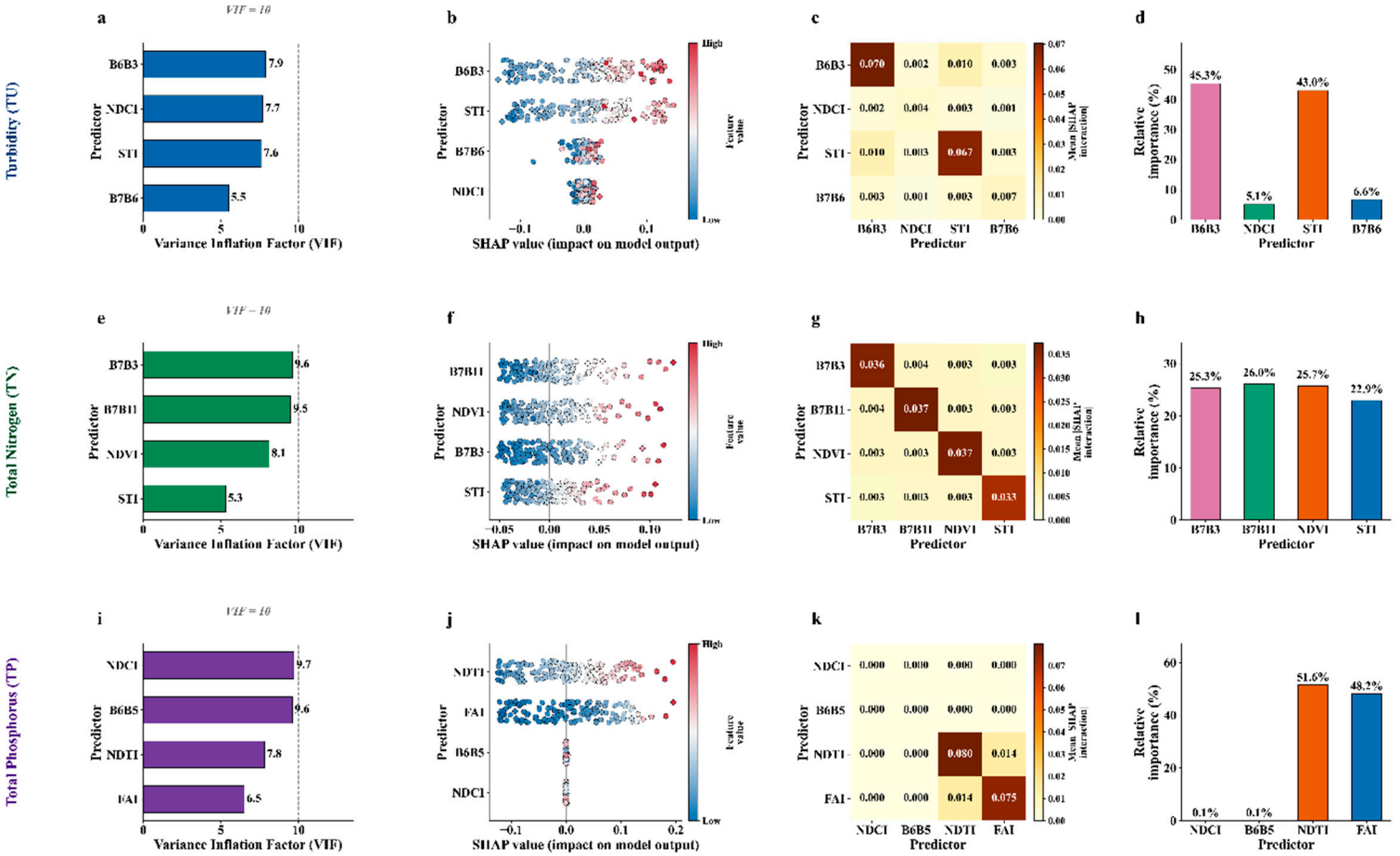


Figure S8: VIF- and correlation-based multicollinearity screening for TU, TN, and TP candidate predictors. (a,c,e) Maximum inter-predictor correlation. (b,d,f) VIF (log scale), VIF = 10 threshold shown. Lighter = retained; darker = excluded.

Table S9: Model performance outputs for training and testing

Parameters	ML Model	Model Training Metrics			Model Testing Metrics		
		R^2	RMSE	MARE	R^2	RMSE	MARE
Chl-a	ANN	0.75	5.62	0.11	0.55	7.85	0.22
	RF	0.96	1.98	0.04	0.94	2.11	0.05
	SVR	0.78	4.43	0.10	0.71	4.73	0.13
	XGB	0.93	2.52	0.05	0.9	2.77	0.07
Turbidity	ANN	0.91	5.20	0.08	0.87	5.75	0.10
	RF	0.94	4.87	0.07	0.92	5.25	0.07
	SVR	0.89	5.34	0.07	0.86	5.74	0.09
	XGB	0.94	4.90	0.07	0.93	5.17	0.07
TN	ANN	0.76	0.35	0.10	0.71	0.41	0.19
	RF	0.95	0.15	0.09	0.93	0.2	0.09
	SVR	0.9	0.21	0.10	0.86	0.3	0.12
	XGB	0.95	0.14	0.10	0.94	0.18	0.10
TP	ANN	0.93	0.23	0.09	0.88	0.3	0.10
	RF	0.93	0.24	0.09	0.91	0.26	0.09
	SVR	0.89	0.28	0.10	0.86	0.36	0.11
	XGB	0.93	0.24	0.09	0.9	0.28	0.09

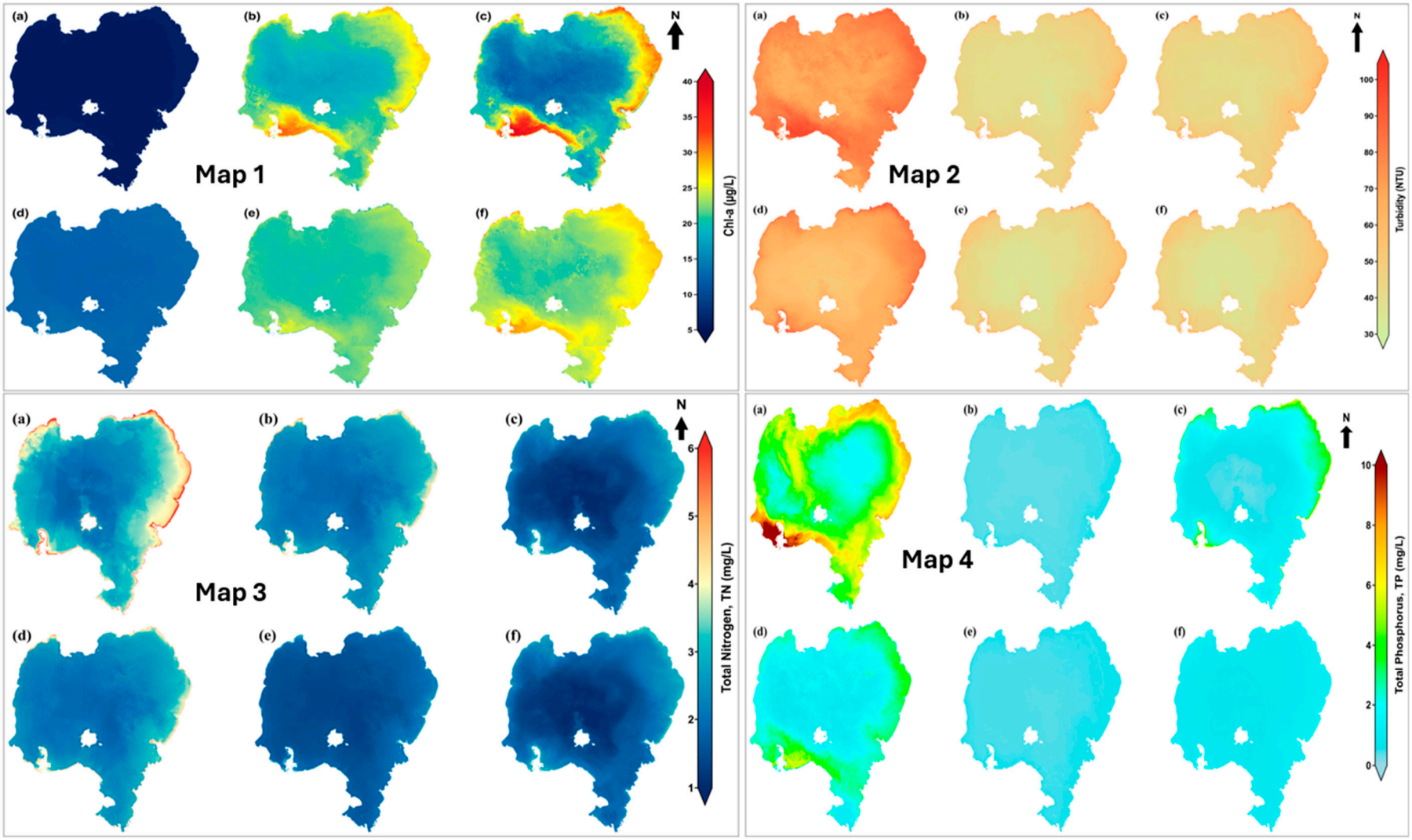


Figure S10: Yearly spatial distribution of Chlorophyll a chl-a, and Total Phosphorus, TP predicted using the Random Forest RF, (Map1 and Map 4) Turbidity and Total Nitrogen, TN (Map 2 and Map 4) using, XGB for the seasons of August 2016 December 2016, March 2017, October 2021, April 2022 and December 2025 (a-f) for each map respectively.

Table S11: Wet- and dry-season descriptive statistics and Mann–Whitney U test comparing water quality parameter values between seasons, Lake Tana, 2024.

Parameter	Wet Mean	Dry Mean	Δ Mean (%)	Wet Median	Dry Median	Δ Median (%)	Mann–Whitney U	p-value	Effect Size (r)
Chl-a	11.65	27.62	+137.1%	9.93	27.84	+180.4%	3,848,990,720	< 0.001***	0.91
TN	2.48	2.38	-3.9%	2.25	2.51	+11.3%	34,376,888,320	< 0.001***	0.24
TP	1.07	0.84	-21.7%	1.03	0.83	-19.0%	66,779,910,144	< 0.001***	-0.48
Turbidity	70.19	64.94	-7.5%	70.15	66.28	+5.5%	57,030,144,000	< 0.001***	-0.27

*Note: ΔMean and Δ Median express the percent change from the wet- to dry-season value. Sample sizes: wet season n = 13,299,365–13,309,050 pixels; dry season n = 13,274,571–13,283,401 pixels. Significance: *** p < 0.001, ** p < 0.01, * p < 0.05.*