Detection and monitoring of HABs using remote sensing and machine learning in inland reservoirs: A case study in Tri An reservoir, Vietnam

Nam-Thang Ha

Faculty of Fisheries, Hue University (Vietnam)

Hao-Quang Nguyen

Institute of Port and Airport Research (Japan)

Thanh-Luu Pham

Vietnam Academy of Science and Technology (VAST), Institute of Tropical Biology (Vietnam)

CONTENT

- Tri An reservoir, Vietnam
- HABs in Tri An reservoir
- Semi-automatic mapping of HABs from water quality dataset
- Semi-automatic mapping of HABs from remote sensing dataset

Tri An reservoir (Vietnam)

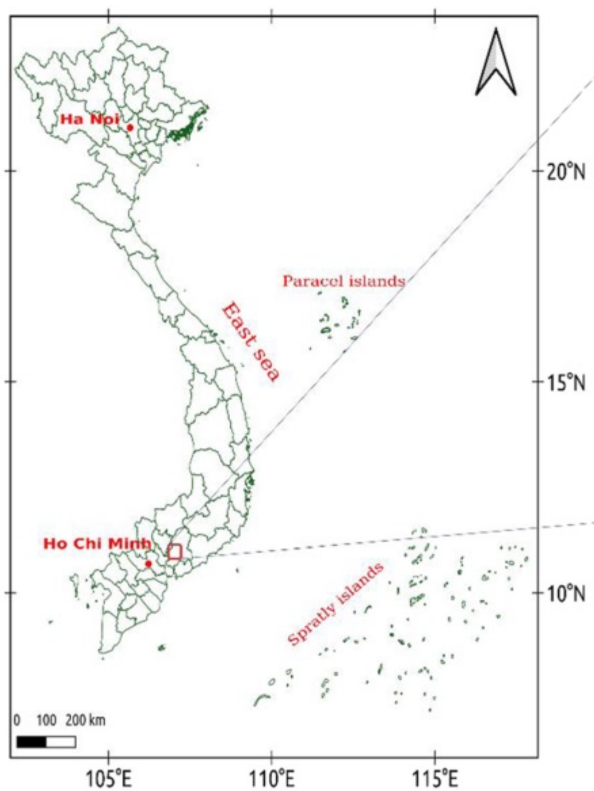
- Dong Nai river basin – second largest basin in Vietnam

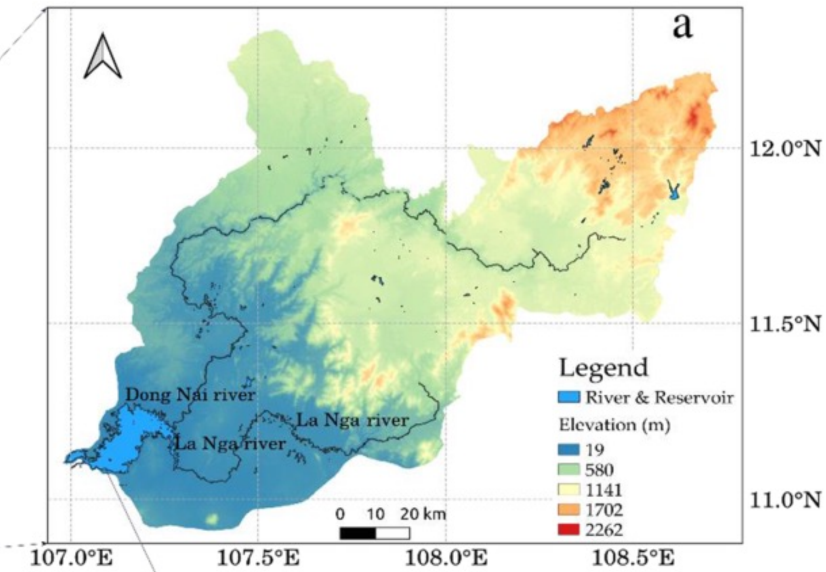
- Provide freshwater for agriculture, irrigation, fisheries, hydro-power

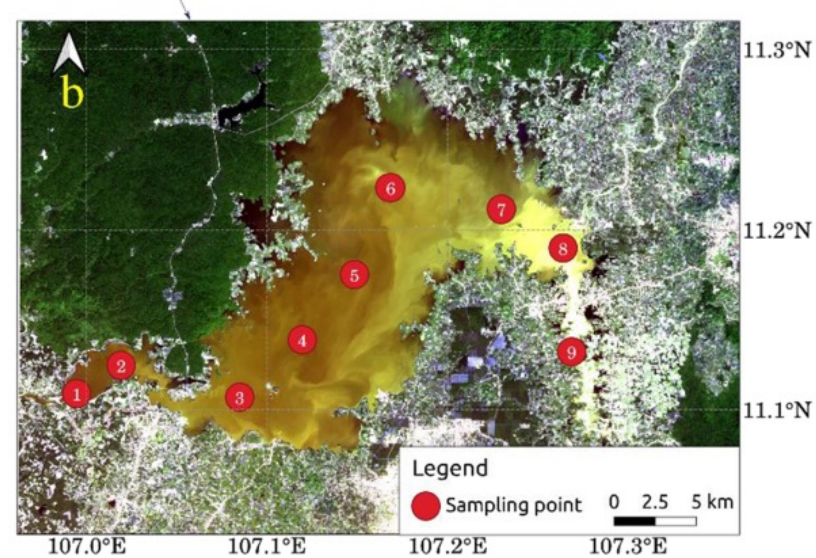
- Surface area: 320 km^2
- Maximum depth: 27 m
- Suffer intensive nutrient loading from surrounding anthropogenic activities

- TN (0.25–1.3 mg/L) and TP (0.05–0.14 mg/L) → eutrophic reservoir.

High occurrence of harmful cyanobacteria blooms







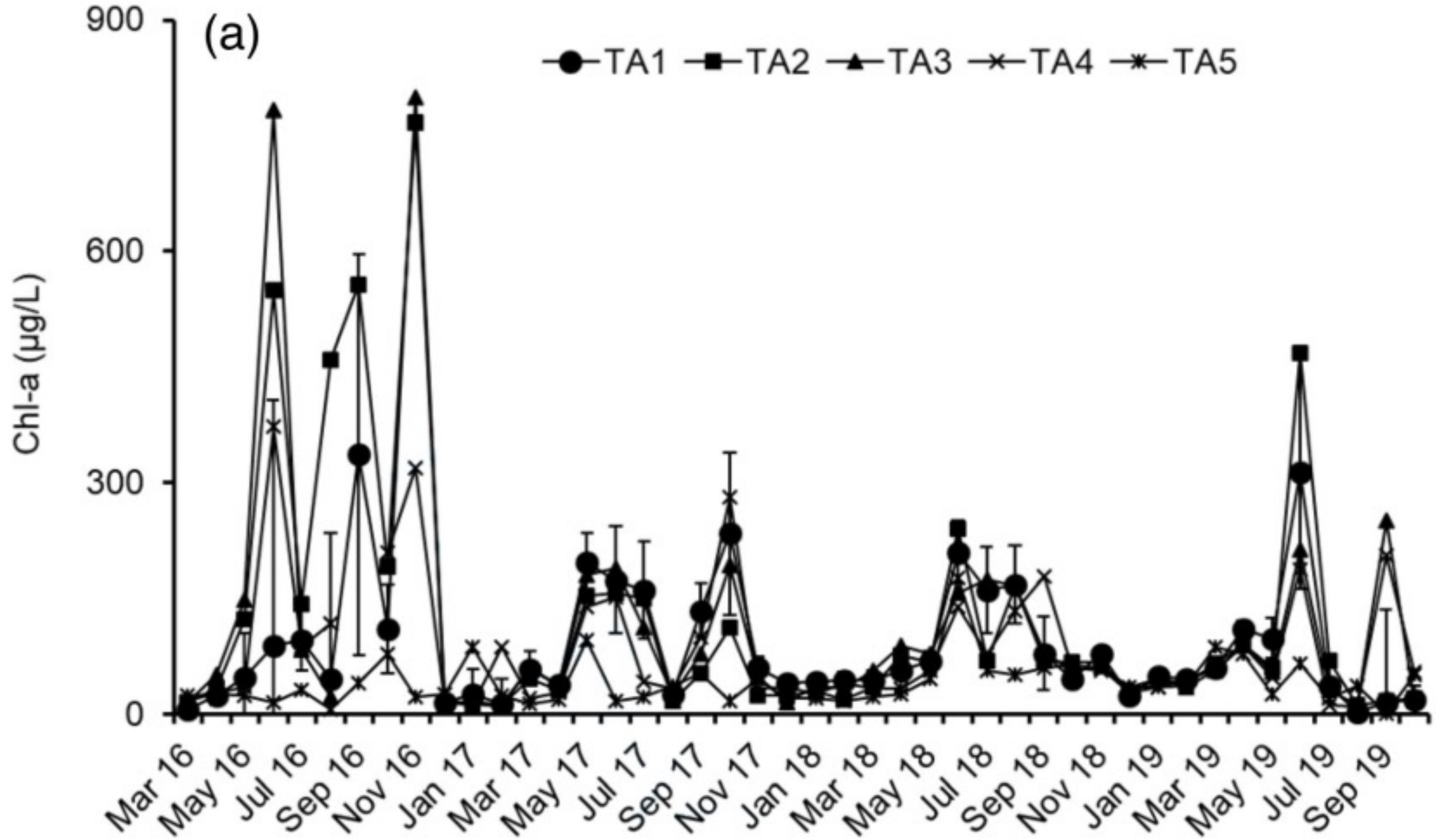
Harmful cyanobacteria blooms in Tri An reservoir

- *Mycrocytis* and *Anabaena* colonies dominant (Dao et al., 2016), producing toxins and hence, HABs

- When cyanobacteria dominate, Chl-a should be used for HCBs monitoring from remote sensing (Stumpf et al., 2016).

Heavy bloom of cyanobacteria in June (a), September (b), November, 2016 (c) and water without bloom (d)

Harmful cyanobacteria blooms in Tri An reservoir



Different blooms were observed in both rainy and dry months.

Mapping of HCBs using linear approach (Sentinel -2)

No.	No. Samples for training	No. Samples for testing	Variables (x)	Model	R^2	RMSE (µg/L)	Bias	Mean Chl-a (µg/L)
1	22	47	B1/B3 vs $log_{10}Chl-a$	$y=-1.0668x+2.4018$	0.27	3.12	0.02	26
2			B6/B3 vs $log_{10}Chl-a$	$y=-1.5854x+2.3435$	0.69	4.22	0.14	
3			*B3/B6 vs $log_{10}Chl-a$*	*$y=0.3438x+0.7736$*	*0.72*	*5.95*	*0.24*	
4			B3/B6 vs $log_{10}Chl-a$	$y=-0.0538x^2+0.6149x+0.469$	0.73	4.53	0.16	
5			B3/B6 vs Chl-a	$y=36.363x-36.111$	0.74	187.03	−20.43	
6			B3/B7 vs Chl-a	$y=37.271x-37.804$	0.71	186.82	−19.45	
7			B3/B7 vs $log_{10}Chl$	$y=0.355x+0.7514$	0.70	6.42	0.25	
8			B7/B3 vs $log_{10}Chl$	$y=-1.5224x+2.314$	0.65	5.00	0.15	
9			B2/B6	$y=51.278x-39.882$	0.69	185.21	−29.56	
10			B2/B7 vs Chl-a	$y=52.037x-40.838$	0.65	185.10	−28.63	
11			B2/B6 vs $log_{10}Chl-a$	$y=0.4699x+0.7636$	0.63	5.00	0.15	
12			B2/B7 vs $log_{10}Chl-a$	$y=0.481x+0.7478$	0.61	4.35	0.16	
13			B6 vs Chl-a	$y=1.322x^{-1.101}$	0.49	202.24	−12.73	
14			B7 vs Chl-a	$y=131.65x^2-24.824x+2.4481$	0.49	4.74	0.20	
15			(B5 + B6)/B4 vs $log_{10}Chl-a$	$y=-0.4811x+2.4517$	0.20	4.62	−0.03	
16			B5/B4 vs $log_{10}Chl-a$	$y=1.5412x^{-0.023}$	0.0002	3.25	−0.03	
17			B5-(B4 + B6)/2 vs $log_{10}Chl-a$	$y=2.3932x^{0.1022}$	0.13	2.92	−0.03	
18			(B1-B2)/(B1 + B2) vs $log_{10}Chl-a$	$y=-2.4056x+1.5881$	0.22	3.08	−0.05	

RMSE value has been converted into unit of µg/L with the variables using log_{10} Chl-a. The italic line in the table indicates the best performance of the linear model.

Mapping of HCBs using linear approach (Sentinel -2)

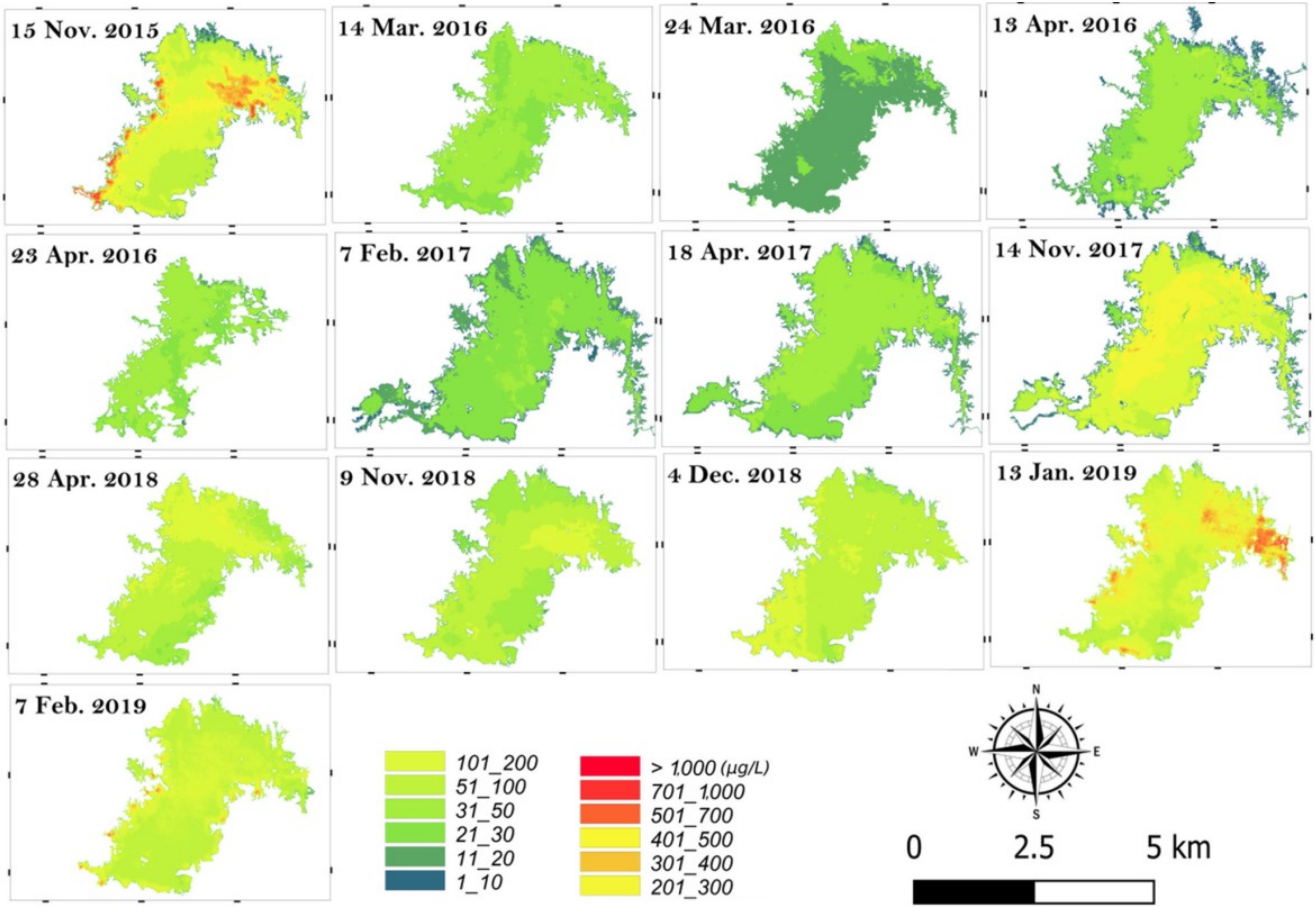


Fig. 12 Spatiotemporal distribution of HCBs in the Tri An Reservoir in dry season

Mapping of HCBs using machine learning: water quality dataset

Pearson correlation analysis between water quality parameters in the tri An reservoir (TAR) (significant level <0.05)

	Chl-a	pH	DO	Temp	Trans	TSS	$N\text{-}NO_3^-$	$P\text{-}PO_4^{3-}$	TN	TP
Chl-a	1									
pH	0.3	1								
DO	−0.08	0.37	1							
Temp	0.22	0.24	−0.08	1						
Trans	−0.19	0.37	0.61	−0.27	1					
TSS	0.49	0.12	−0.1	0.09	−0.29	1				
$N\text{-}NO_3^-$	−0.05	−0.02	−0.03	−0.1	−0.08	0.01	1			
$P\text{-}PO_4^{3-}$	0.25	0.1	−0.14	0.08	−0.18	0.19	−0.04	1		
TN	0.75	0.13	−0.15	0.18	−0.31	0.46	−0.01	0.15	1	
TP	0.44	0.06	−0.26	0.11	−0.26	0.31	−0.07	0.5	0.43	1

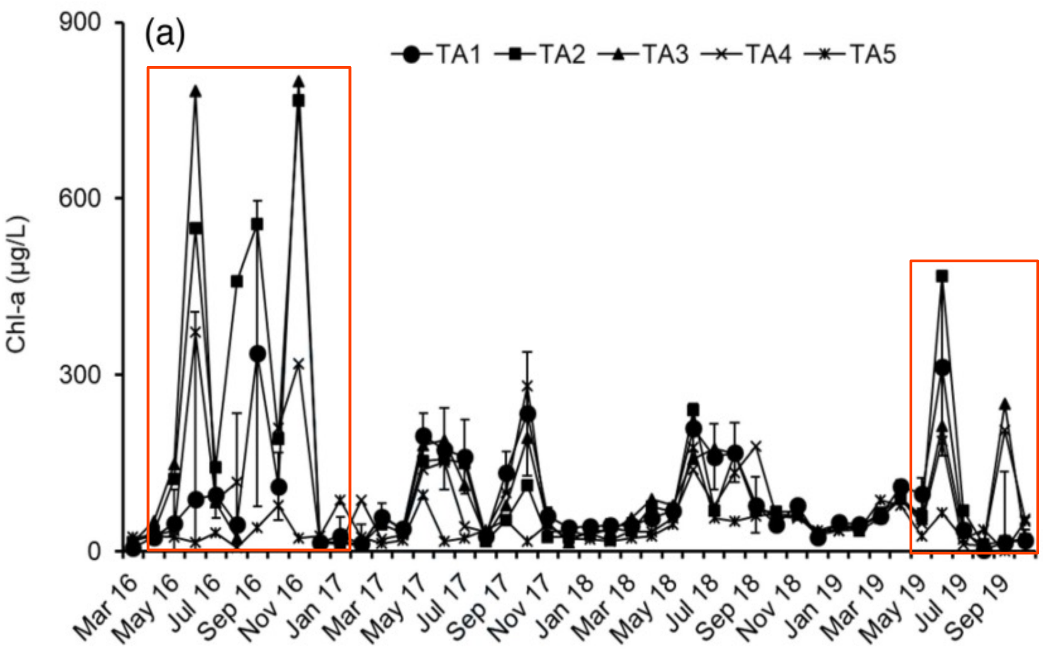
Abbreviations: Chlorophyll-a, Chl-a; DO, dissolved oxygen; TN, total nitrogen; TP, total phosphorous; Trans, transparency; TSS, total suspended solids.

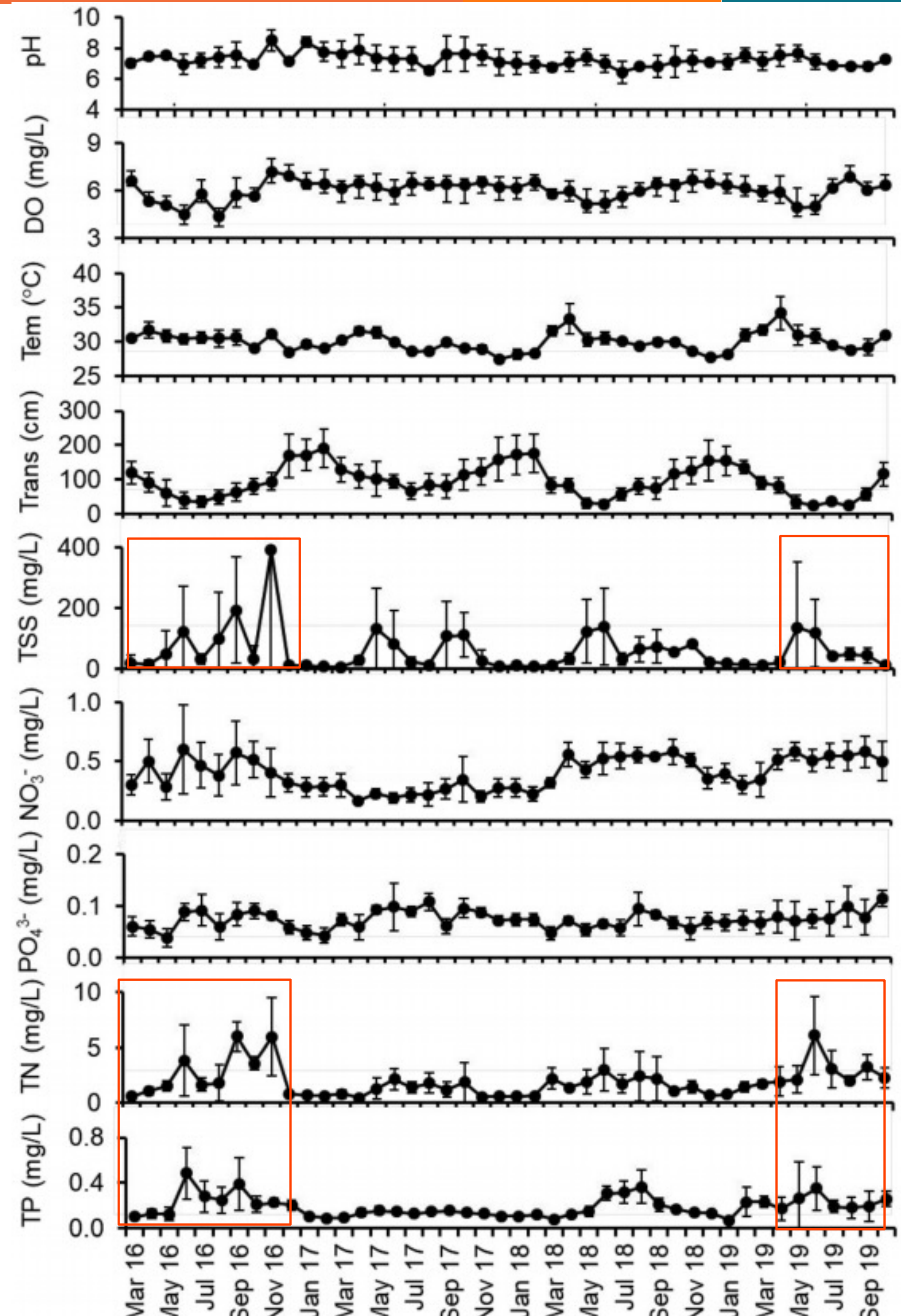
RMSE and MAE values were converted into unit of μg/L

Mapping of HCBs using machine learning and water quality dataset

- Mean values and standard deviation of water quality variables from March 2016 to October 2019 in the Tri An Reservoir (TAR)

229 time-series observations:
– 70% for training: ~ 159 obs
– 30% for testing: ~ 70 obs

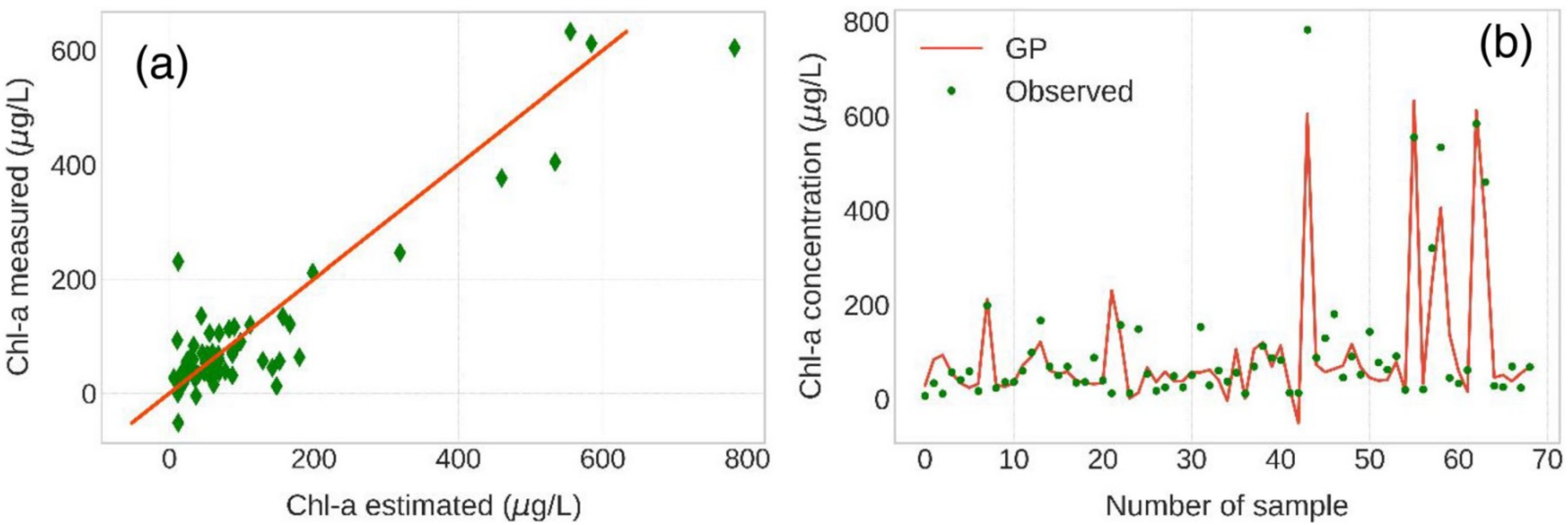




Mapping of HCBs using machine learning: water quality dataset

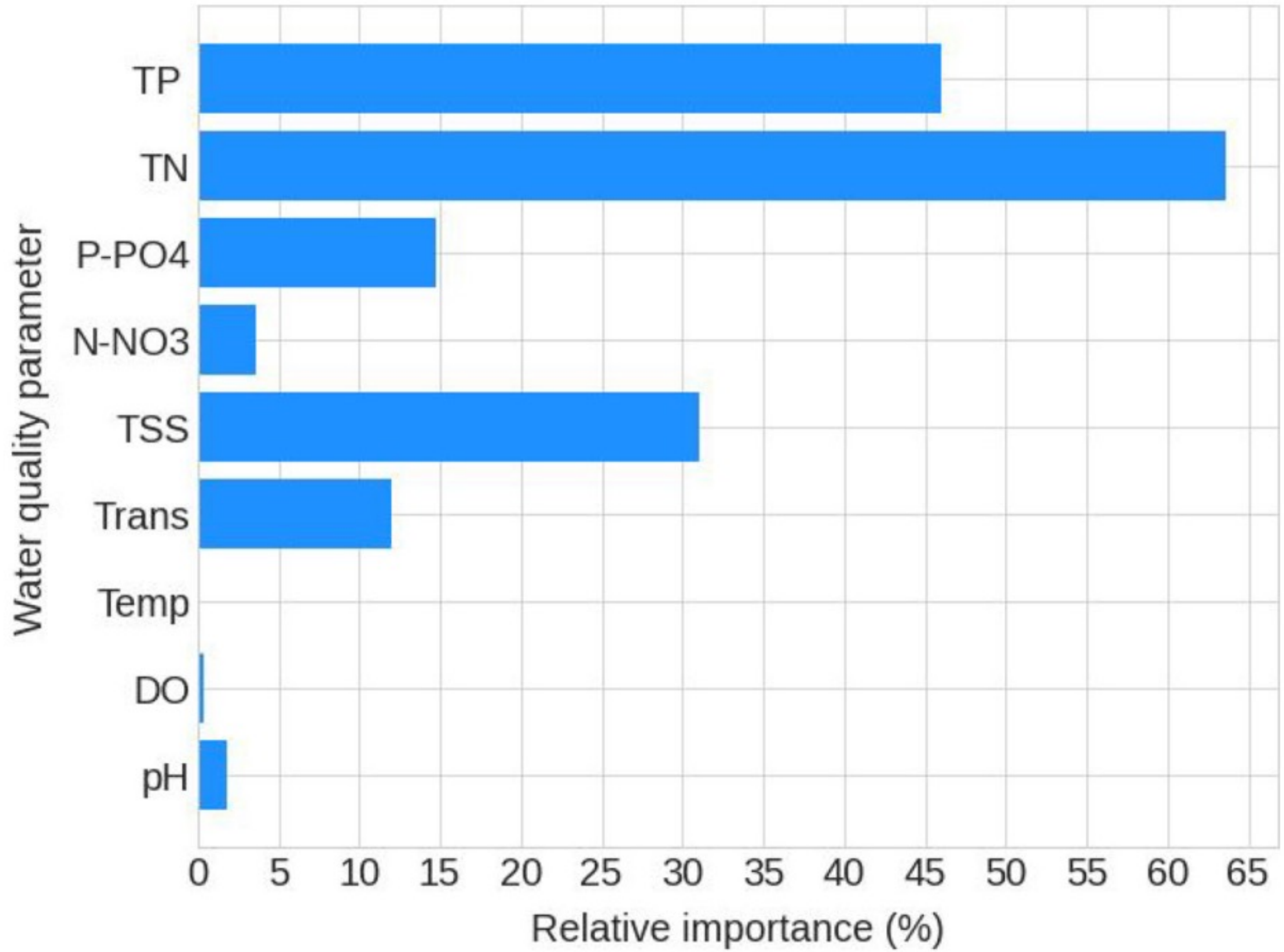
Model	R^2 (training)	R^2 (testing)	AIC	BIC	RMSE (μg/L)	Mean Chl-a (μg/L)
RF	0.85	0.57	641.34	661.45	96.51	105.61
SVM	0.44	−0.05	710.58	730.69	151.21	
GP	**0.97**	**0.85**	**575.10**	**595.24**	**56.65**	
XGB	0.99	0.66	631.83	651.94	85.45	
CB	0.99	0.65	635.68	655.79	86.23	

Abbreviations: AIC, Akaike's information criterion; BIC, Bayesian information criterion; Chl-a, chlorophyll-a; CB, CatBoost; GP, Gaussian process; RF, random forest; RMSE, root-mean-square error; SVM, support vector machine; XGB, extreme gradient boost.



Mapping of HCBs using machine learning and water quality dataset

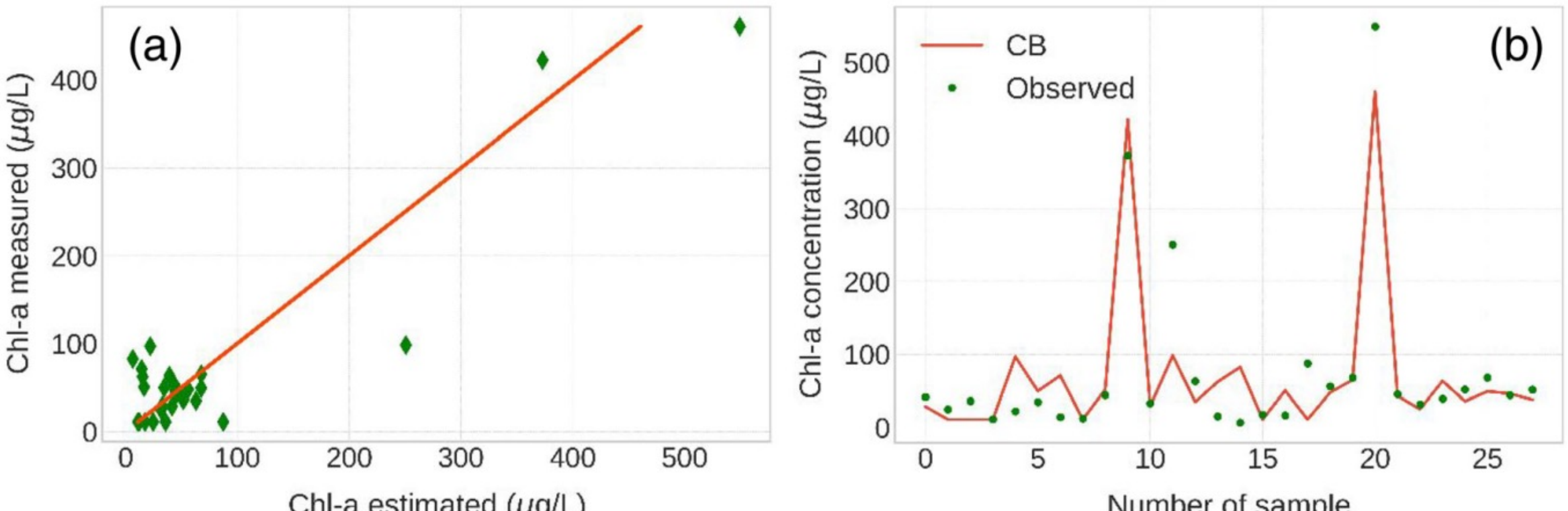
Feature importance of water quality parameters in Gaussian process (GP) model



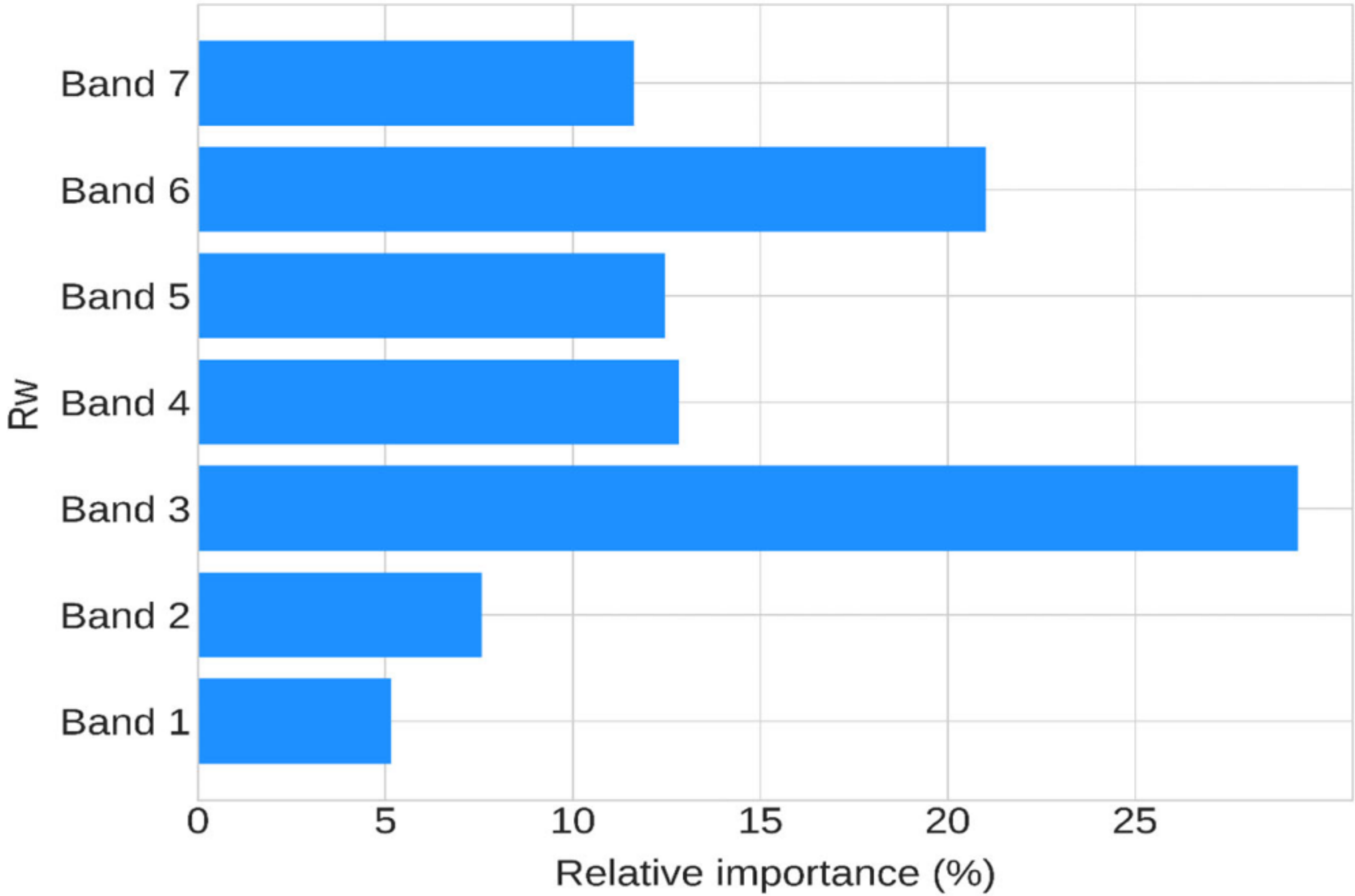
Mapping of HCBs using machine learning and remote sensing dataset (Sentinel - 2)

Model	R^2 (training)	R^2 (testing)	AIC	BIC	RMSE (µg/L)	Mean Chl-a (µg/L)
RF	0.44	0.10	274.25	283.57	111.47	105.61
SVM	0.42	0.03	280.27	289.60	116.14	
GP	0.99	0.55	258.38	267.70	78.56	
XGB	0.84	0.81	233.82	243.14	50.66	
CB	**0.99**	**0.84**	**229.18**	**238.50**	**46.28**	

Abbreviations: AIC, Akaike's information criterion; BIC, Bayesian information criterion; Chl-a, chlorophyll-a; CB, CatBoost; GP, Gaussian process; RF, random forest; RMSE, root-mean-square error; SVM, support vector machine; XGB, extreme gradient boost.



Mapping of HCBs using machine learning and remote sensing dataset



FROM SEMI-AUTOMATIC TO AUTOMATIC MAPPING OF HCBs

- Develop transferring and general ML model that deals with high variation of Chl-a in different waters
- Develop python-based program with ML and DL models to automatic processing remote sensing data for HCBs mapping
- Develop fast and accurate methods for Chl-a measurement to increase dataset for ML and DL model training.

CHALLENGES

– High cloud coverage —> less remote sensing image fit the sampling date

– Insufficient time-series dataset of to train the ML models at different levels of bloom

– Coarse spatial resolution of free satellite image whilst very costly in using very high spatial resolution image

– Hyper-spectral image would be prefer, but not availability

– Variation and very high concentration of Chl-a in inland waters

REFERENCES

1. Comparing the performance of machine learning algorithms for remote and in situ estimations of chlorophyll-a content: A case study in the Tri An Reservoir, Vietnam.
DOI: 10.1002/wer.1643

2. Estimation of nitrogen and phosphorus concentrations from water quality surrogates using machine learning in the Tri An Reservoir, Vietnam
DOI: 10.1007/s10661-020-08731-2

3. Inland harmful cyanobacterial bloom prediction in the eutrophic Tri An Reservoir using satellite band ratio and machine learning approaches
DOI: 10.1007/s11356-019-07519-3

THANK YOU
FOR YOUR ATTENTION!