Supplement of Hydrol. Earth Syst. Sci., 27, 3581–3599, 2023
https://doi.org/10.5194/hess-27-3581-2023-supplement
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Remote quantification of the trophic status of Chinese lakes

Sijia Li et al.

Correspondence to: Zhidan Wen (wenzhidan@iga.ac.cn)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Methods

S1.1 Linear regression (MLR)

Multiple linear regression (MLR) is developed to formulate an approximation linear function between a set of independent variables and the dependent variable. The MLR can be described in a compact matrix form as below:

$$Y = X\beta + \varepsilon \tag{S1}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ and } X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ 1 & x_{2,1} & \cdots & x_{2,k} \\ 1 & \vdots & x_{m,i} & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{bmatrix}_{n\times(k+1)} \tag{S2}$$

where n is the number of samples, $x_{m,i}$ is the value of the i^{th} independent variable in the m^{th} sample, and ε_i is the i^{th} residual error in the m^{th} sample.

The standard least-square method can be used to calculate the coefficient vector β, which is described as follows:

$$\beta = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}Y \tag{S3}$$

Then the obtained MLR model can be adopted to predict the possible dependent variable related with the newly input vector.

S1.2 Support vector machine (SVM)

SVM aims to present the dataset into a high-dimensional features space via nonlinear mapping, and solve the prediction problem. Assuming that a given observation dataset, (X, Y), $x_i \in X \in R$, $f(x_i) \in Y \in R$ (x_i and $f(x_i)$ are the input vector and target value, respectively), are randomly and independently generated by function. The SVM regression function can be expressed as,

$$f(x) = w \times \varphi(x) + b \tag{S4}$$

where x is the input data, $\varphi(x)$ represents the nonlinear mapping function, w is the weight vector, and b is the bias term. In addition, the radial basis kernel function in SVM has the advantages of nonlinear forecasting performance and less numerical difficulties. Hence, the radial basis kernel function was used in this study.

S1.3 XGBoost (XGB)

XGBoost proposed by Chen et al., (2016) is an improved algorithm depended on the gradient-enhanced decision tree (GBDT). Compared the traditional GBDT, XGBoost performs the second-order Taylor expression on the loss function to approximate the objective function, providing efficiency of solving the optimal solution. XGB Integrates the tree model with addition method, assuming a total of K trees, and use F to represent the basic tree model as following,

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i), f_k \in F \quad \text{(S5)}$$

The objective function is:

$$L = \sum_i l(\widehat{y_i}, y_i) + \sum_k \Omega(f_k) \quad \text{(S6)}$$

Where l is the loss function, which represents the error between the predictive value and the true value. Ω is the function used for regularization to prevent overfitting.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \quad \text{(S7)}$$

where *T* represents the number of leaves per tree, and *w* represents the weight of the leaves of each tree.

S1.4 Random forest (RF)

Random forest (RF) is a tree-based ensemble-learning algorithm and was developed to address the shortcomings of traditional method. RF consists of combination model that has a large number of regression trees, with the average of *n* decision trees taken as the ensemble estimate (Breiman, 2001). In RF, bootstrap samples are chosen from original dataset to build a unpruned regression tree. Only a small and fixed number of randomly sampled *K* predictor are selected as split candidates. These two steps are then repeated until *C* such trees are grown, and new data is predicted by aggregating the prediction of the C trees. RF can be expressed as follows:

$$\hat{f}_{RF}^{C}(\mathbf{x}) = \frac{1}{C}\sum_{i=1}^{C} T_i(\mathbf{x}) \tag{S8}$$

Where x is the vectored input variable, C is the number of trees, and Ti ðx Þ is a single regression tree constructed based on a subset of input variables and the bootstrapped samples. A bagging is used to increase the diversity of the trees by growing them from different training data-sets, and this reduced the overall variance of the model (Rodriguez-Galiano et al., 2015). RF performs out-of-bag error estimation in the process of constructing the forest, which can compute an unbiased estimation of generalisation error without using an external text data subset (Breiman,2001). In order to improve the model's performance on high-dimensional datasets, RF also enables assessment of relative importance of input features (Ahmad et al.,2017).

Section 2 Accuracy assessment of model

Mean normalized error (MAE) and root mean square error (RMSE) were used to provide general descriptions of model performances between measured- and estimated- *TSI*. They are defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|E_i - M_i| \tag{S9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - M_i)^2} \tag{S10}$$

where N is the number of data pairs, and i denotes individual data; M and E are measured- and estimated- values, respectively.

Section 3 Five limnetic regions

EPLR, located in the middle and southeastern of China, has approximately 830 lakes with >1 km^2, accounting to 25.3% of the total lake area in China (Ma et al., 2011). There are rapid development of social economy and human economic activities in EPLR. NLR has a humid and sub-humid, continental monsoon climate and most of lakes are located in Songnen Plain and the remaining is in mountainous area. MXPLR has a high density are of saline lakes accounting for 22.1% of the total lake area in China, because of the arid and semi-arid climate characteristics (Song et al., 2018). YGPLR is a Karst topography region with high altitude and low latitude, and there has relatively small and deep lakes. TQPLR, known as a Asia's water tower, has thousands of closed lakes accounting to more than half of the total area of China. The averaged altitude is above 4 000 m and most lakes are fed by glacier melting and snow, can be more sensitive to climate change.

Table S1 Field surveys of lake names, abbreviation, location and sampling dates and the averaged concentration of water quality in China. Abb. represents the abbreviation of lakes, Time is sampling time, *N* denotes numbers of samples, EC is electrical conductivity (μs cm^{-1}), Turb represents turbidity (NTU), SDD denotes Secchi disk depth (m), Chl-a represents chlorophyll-a concentration (μg L^{-1}), TP is total phosphorus concentration (mg L^{-1}), DOC denotes dissolved organic carbon concentration (mg L^{-1}) and TSM represents total suspended matter concentration (mg L^{-1}), respectively.

Lakes	Abb.	Location(N/E)	Time[a]	*N*	pH	EC	Turb	SDD	Chl-a	TP	DOC	TSM
TaiPingChi	TPC(1)	44 °02',124 °57'	2018-06	3	8.2±0.8	875.7±72.3	41.3±2.7	0.4±0.02	75.3±27.1	0.1±0.005	10.9±0.9	37.6±3.4
XinMiaoPao	XMP(2)	45 °10',124 °27'	2018-10	6	8.4±0.2	348.3±5.9	43.0±10.6	0.3±0.03	27.4±6.1	0.2±0.2	8.7±2.2	32.3±7.3
ChaoHu	CH(3)	31 °41',117 °23'	2019-04	9	8.4±0.3	406.7±105.0	35.7±29.5	0.2±0.04	27.8±20.0	0.09±0.05	4.5±1.5	55.1±19.0
DaLi	DL(4)	43 °17',116 °35'	2019-07	19	8.8±0.2	9793.2±2375.1	22.3±3.4	0.5±0.1	10.5±3.7	2.1±0.09	64.3±13.4	20.5±2.6
Yueliang Pao	YLP(5)	45 °43', 123 °59'	2018-10	6	8.3±0.05	207.2±14.4	43.7±6.0	0.3±0.03	10.9±0.8	0.14±0.08	8.2±0.8	33.8±7.9
GaoYou	GY(6)	32 °48',119 °18'	2019-01	10	8.2±0.2	294.0±9.6	46.9±17.0	0.4±0.05	9.6±3.5	0.1±0.009	4.3±0.8	37.5±14.6
			2019-04	5	11.0±0.3	503.4±6.5	3.1±2.0	0.8±0.4	26.6±15.7	0.07±0.02	4.9±0.4	14.7±10.2
			2019-11	10	7.8±0.2	505.6±68.2	91.1±51.6	0.2±0.06	15.1±5.3	0.07±0.01	7.7±1.4	68.2±32.6
QingHe	QH(7)	42 °32',124 °12'	2018-10	4	7.7±0.2	262.8±5.2	19.8±10.2	0.6±0.1	14.6±3.8	0.1±0.03	2.8±0.9	11.8±5.5
ErlongLake	EL(8)	43 °14',124 °50'	2018-07	10	8.3±0.2	472.3±14.2	9.0±2.0	0.8±0.1	44.3±12.7	0.006±0.002	6.5±0.7	11.7±4.5
XingXingShao	XXS(9)	43 °37',126 °3''	2018-10	4	7.5±0.1	129.4±4.8	10.8±1.5	1.1±0.06	25.3±5.3	0.04±0.01	4.5±0.6	7.3±0.3
HongShan	HS(10)	42 °43',119 °41'	2017-04	4	8.5±0.1	1.1±0.02	45.9±26.8	0.6±0.2	6.1±3.4	0.05±0.03	9.7±1.2	46.7±28.8
			2019-07	5	8.2±0.2	1601.0±28.4	38.2±21.3	0.4±0.09	16.4±3.2	0.07±0.002	5.7±0.4	24.9±10.2
XiaoXingKaiHu	XXKH(11)	45 °22',132 °22'	2018-07	11	8.4±0.07	273.4±22.2	48.7±11.0	0.3±0.02	8.2±1.6	0.02±0.00	5.9±0.6	47.1±14.4
			2018-10	9	7.8±0.1	158.9±7.1	46.8±18.7	0.4±0.07	9.3±2.2	0.2±0.1	6.9±0.5	33.0±15.8
Qingnian	QN(12)	45 °40',131 °49'	2017-06	3	8.6±0.5	0.2±0.03		0.4±0.1	13.6±9.7	0.03±0.001	7.5±0.5	11.4±5.1

HongZe	HZ(13)	33 °06',118 °43'	2019-01	10	8.2 ±0.1	367.9 ±51.1	74.8 ±34.5	0.3 ±0.2	4.7 ±0.8	0.2 ±0.02	3.8 ±1.6	60.3 ±24.8
			2019-04	2	9.5 ±0.007	568.5 ±4.9	6.2 ±2.8	0.2 ±0.01	6.7 ±3.4	0.03 ±0.008	3.7 ±0.4	14.0 ±3.6
			2019-08	3	10.6 ±0.1	327.3 ±12.0	52.0 ±3.2	0.2 ±0.02	3.3 ±0.80	0.07 ±0.02	2.1 ±0.07	38.4 ±9.6
			2019-11	8	8.1 ±0.4	381.5 ±20.4	56.5 ±26.2	0.3 ±0.1	5.1 ±1.8	0.08 ±0.008	5.8 ±0.5	59.9 ±25.1
DaXingKaiHu	DXKH(14)	45 °10',132 °25"	2018-10	19	7.3 ±0.2	109.4 ±17.3	49.6 ±9.4	0.3 ±0.04	4.2 ±0.8	0.2 ±0.02	3.8 ±1.2	49.8 ±6.6
LuoMa	LM(15)	34 °06',118 °13'	2019-04	7	9.7 ±0.4	658.0 ±43.7	8.2 ±5.8	0.6 ±0.2	8.3 ±2.9	0.04 ±0.007	3.9 ±0.5	12.4 ±5.6
GuTian	GT(16)	26 °36', 118 °48'	2017-06	4	9.1 ±0.06	59.4 ±0.9	6.6 ±0.9	0.9 ±0.04	12.5 ±1.7	0.02 ±0.005	2.3 ±0.2	3.3 ±0.2
DaGuangBa	DGB(17)	18 °57',109 °0'	2017-07	1	7.4	99.7	35.0	0.9	5.3	0.06	2.0	9.4
PoYang	PY(18)	29 °41',116 °10'	2019-04	9	7.7 ±0.04	145.8 ±31.2	11.5 ±11.0	0.7 ±0.1	3.0 ±1.1	0.05 ±0.01	1.9 ±0.4	13.6 ±5.0
			2019-07	9	10.9 ±0.06	104.3 ±2.5	9.7 ±2.7	0.8 ±0.1	10.3 ±2.2	0.05 ±0.005	1.6 ±0.2	7.0 ±2.5
YangXi	YX(19)	23 °55',116 °52'	2017-07	3	9.2 ±0.06	94.3 ±2.1	9.4 ±3.5	0.9 ±0.1	8.2 ±0.5	0.02 ±0.001	2.0 ±0.03	3.6 ±0.5
WeiShan	WS(20)	34 °38',117 °18'	2019-01	6	8.2 ±0.09	588.5 ±285.4	2.2 ±1.8	1.4 ±0.5	7.2 ±6.5	0.2 ±0.02	3.9 ±1.0	4.4 ±3.9
WuLiangSuHai	WLSH(21)	40 °54',108 °52'	2019-08	10	8.1 ±2.3	2343.0 ±534.2	6.3 ±5.6	1.0 ±0.6	5.1 ±2.6	0.06 ±0.03	6.7 ±1.8	5.7 ±5.4
FenHe	FH(22)	38 °04',111 °53'	2017-04	3	8.3 ±0.05	0.8 ±0.01	6.1 ±0.01	1.3 ±0.2	8.0 ±5.8	0.02 ±0.002	2.8 ±0.2	16.7 ±6.0
KeLuKe	KLK(23)	37 °15',96 °54'	2017-08	8	8.1 ±0.3	967.0 ±1.4	3.7 ±1.1	0.5 ±0.2	3.7 ±0.8	0.02 ±0.001	3.2 ±0.6	10.1 ±5.5
			2019-09	2	8.3 ±0.1	1004.5 ±46.5	3.4 ±1.5	1.4 ±0.6	4.3 ±2.5	0.07 ±0.01	5.6 ±0.6	2.8 ±1.4
FengShuBa	FSB(24)	24 °27',115 °23'	2019-04	2	8.5 ±0.2	129.6 ±1.0	0.5 ±0.4	3.4 ±0.03	5.0 ±0.9	0.02 ±0.003	2.1 ±0.5	1.0 ±0.001
			2019-11	5	12.0 ±0.5	88.9 ±1.1	2.8 ±1.2	1.0 ±0.06	4.0 ±2.0	0.05 ±0.004	1.9 ±0.1	1.8 ±0.4
HengLong	HL (25)	41 °19',125 °28'	2018-10	9	7.7 ±0.08	211.5 ±5.5	2.6 ±0.9	1.9 ±0.3	3.6 ±1.7	0.07 ±0.02	3.3 ±0.2	2.6 ±1.0
DaiHai	DH(26)	40 °33',112 °41'	2017-04	2	8.8 ±0.007	13.4 ±0.08	17.3 ±2.0	0.5 ±0.09	1.9 ±1.7	0.1 ±0.02	45.8 ±1.1	12.4 ±11.2
			2019-08	8	8.9 ±0.02	19350.0 ±83.1	6.6 ±3.4	1.3 ±0.3	1.6 ±0.3	0.1 ±0.006	46.9 ±4.1	6.0 ±1.4
ZhaRiNanMuCo	ZRNM(27)	31 °00',85 °31'	2017-08	4	9.2 ±0.01	10325.9 ±661.3	11.8 ±10.3	0.8 ±0.2	1.1 ±0.4	0.1 ±0.02	6.9 ±0.7	20.4 ±7.8
BaiShan	BS(28)	42 °37',127 °08'	2019-06	6	10.7 ±1.6	144.6 ±14.6	2.3 ±1.1	2.1 ±0.4	5.1 ±3.5	0.03 ±0.007	4.4 ±0.7	2.6 ±1.6
XinFengJiang	XFJ(29)	23 °50',114 °35'	2019-01	5	7.6 ±0.03	79.7 ±0.6	2.5 ±0.4	3.6 ±0.1	1.7 ±0.2	0.10 ±0.003	1.2 ±0.1	1.0 ±0.09
			2019-04	6	11.2 ±0.08	78.2 ±0.8	2.0 ±0.6	1.3 ±0.03	6.8 ±1.1	0.04 ±0.002	1.9 ±0.1	1.7 ±0.4

			2019-11	7	7.3±0.05	75.4±1.0	1.3±0.5	3.5±0.8	2.8±1.2	0.03±0.002	2.2±0.3	0.8±0.3
GuanYinGe	GYG(30)	41 °22',124 °18'	2018-10	5	7.8±0.09	225.1±0.8	1.7±0.7	4.1±1.3	4.8±1.3	0.04±0.02	2.3±0.3	2.1±1.1
LongYangXia	LYX(31)	36 °09',100 °48'	2019-09	7	8.0±0.2	351.7±0.7	0.8±0.9	2.3±0.1	3.0±0.9	0.03±0.02	1.6±1.2	2.8±1.1
BaiPenZhu	BPZ(32)	23 °06',115 °08'	2017-07	1	7.7	33.1	14.0	1.0	0.7	0.01	1.0	5.1
			2019-04	5	7.6±0.2	34.2±2.3	5.2±1.4	1.1±0.1	2.1±0.4	0.01±0.004	2.9±1.7	6.6±3.8
			2019-07	3	11.3±0.5	31.4±0.3	4.9±0.5	1.2±0.2	0.8±0.06	0.07±0.05	1.1±0.2	5.5±2.6
			2019-11	6	7.7±0.2	33.3±0.7	2.0±0.6	1.9±0.2	2.9±0.5	0.04±0.002	2.0±0.2	2.4±0.9
XiaoLangDi	XLD(33)	34 °56',112 °18'	2019-09	19	7.7±0.2	739.5±8.5		1.5±0.3	2.1±1.0	0.01±0.002	1.9±0.8	4.9±4.1
YunFeng	YF(34)	41 °33',126 °35'	2018-10	1	7.4	118.7	1.5	3.4	1.3	0.1	3.5	1.5
TuoSu	TS(35)	37 °10',96 °54'	2017-08	10	8.6±0.03	32674.9±741.6	0.9±0.8	3.8±0.4	1.4±0.8	0.2±0.01	41.2±8.9	2.8±1.2
QingHaiHu	QHH(36)	36 °57',100 °21'	2019-09	32	8.9±0.05	14653.0±537.6	2.0±1.2	2.9±0.8	1.6±0.8	0.03±0.05	35.4±9.1	3.4±2.5
LiJiaXia	LJX(37)	36 °07',101 °47'	2017-09	4	8.3±0.02	335.6±4.2	1.2±0.5	3.4±0.3	0.5±0.1	0.05±0.003	2.3±0.68	1.2±0.4
			2019-09	13	8.2±0.1	339.8±3.8	0.7±0.6	2.2±0.3	1.7±1.0	0.04±0.07	2.2±0.5	2.7±0.8
ZheLin	ZL(38)	29 °14',115 °28'	2019-07	7	10.7±0.3	109.7±4.6	0.3±0.2	4.6±0.9	1.1±0.5	0.04±0.002	1.5±0.3	0.6±0.2
			2019-11	4	7.2±0.07	101.9±2.6	0.7±0.3	3.9±0.5	1.6±0.2	0.1±0.2	1.8±0.08	0.8±0.2
LiuJiaXia	LJX(39)	35 °51',103 °15'	2017-09	8	8.3±0.02	392.8±11.5	1.4±0.6	4.3±0.8	1.3±0.3	0.05±0.002	1.8±0.3	1.0±0.7
			2019-09	11	7.7±0.1	376.6±2.9	1.1±0.8	2.6±0.4	1.4±1.0	0.01±0.006	2.0±0.2	3.1±3.0
SeLinCo	SLC(40)	31 °49',89 °11'	2017-08	4	8.9±0.01	11256.5±102.8	1.3±1.6	2.8±0.2	0.3±0.1	0.1±0.01	20.3±10.9	0.4±0.07
TaRuoCo	TRC(41)	31 °07',84 °17'	2017-08	4	8.5±0.01	759.8±4.0	0.3±0.3	2.7±0.5	0.2±0.04	0.07±0.003	3.5±0.2	2.0±2.9
CuoNa	CN(42)	32 °03',91 °27'	2017-08	5	8.3±0.01	440.6±5.0	0.3±0.2	4.8±2.2	0.3±0.1	0.02±0.005	4.7±0.5	0.8±0.4
BaMuCuo	BMC(43)	31 °32', 90 °34'	2017-08	2	9.3±0.01	10449.6±15.0	1.4±0.3	6.1±1.6	0.4±0.07	0.01±0.001	15.5±8.1	1.6±0.5
HaLa	HL(44)	38 °12',97 °35'	2019-09	1	8.95	16515.0	1.9	6.6	0.4	0.02	5.7	2.4
NamoCo	NMC(45)	30 °47',90 °51'	2017-08	2	8.9±0.01	1373.5±5.0	9.2±12.8	9.2±0.3	0.3±0.2	0.05±0.001	4.2±0.1	0.6±0.02

[a] The time is Year-month; [+] means brackish lake

Table S2 Bandwidths, central wavelengths, and corresponding pixel resolution of Sentinel-2A MSI Sensor

Band	1	2	3	4	5	6	7	8	8a	9	10	11	12
λ[a]	443	492	560	665	704	740	783	833	865	945	1374	1614	2202
B[b]	21	66	36	31	15	15	20	106	21	20	31	91	175
Res.[c]	60	10	10	10	20	20	20	10	20	60	60	20	20

[a] denotes central wavelength, nm; [b] denotes bandwidth, nm; [c] denotes pixel resolution, m.

The data comes from European Space Agency (https://earth.esa.int/web/sentinel/)

Table S3 ANOVA analysis of reflectance spectra *Rrs*(λ) according to the k-means clustering (Avg. ± SD.)

Wavelength (nm)	cluster_1(*N*=39)	cluster_2(*N*=73)	cluster_3(*N*=161)	*F*	*p*
443	0.03 ± 0.01	0.01 ± 0.01	0.03 ± 0.01	474.587	0.000**
492	0.04 ± 0.01	0.02 ± 0.01	0.05 ± 0.01	608.241	0.000**
566	0.06 ± 0.02	0.02 ± 0.01	0.10 ± 0.02	831.889	0.000**
665	0.02 ± 0.01	0.01 ± 0.01	0.07 ± 0.01	1537.509	0.000**
704	0.01 ± 0.01	0.01 ± 0.01	0.07 ± 0.01	1644.586	0.000**
740	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	1546.425	0.000**

*$p<0.05$, **$p<0.01$

TableS4 Descriptive statistics of water quality and bio-optical properties

Parameters	*Units*	*N*	Avg.	SD.	Min.	Max
pH	-	431	8.51	1.04	6.86	13.05
EC	µS cm^{-1}	431	3252.3	6739.31	0.17	33453.10
Turb	UTN	431	18.62	26.46	0	183.35
SDD	m	431	1.6	1.5	0.17	9.47
Chl-a	µg L-1	431	7.56	11.28	0.13	100.22
TP	mg L^{-1}	431	0.16	0.42	0.003	2.17
DOC	mg L^{-1}	431	11.01	16.83	0.09	85.36
TSM	mg L^{-1}	431	16.94	21.9	0.24	147.50
$a_{ph}(440)$	m^{-1}	431	0.48	0.72	0	5.33
$a_{d}(440)$	m^{-1}	431	0.93	1.43	0	6.96
$a_{CDOM}(440)$	m^{-1}	431	0.54	0.43	0	1.89

TableS4 The linear regression coefficients (2-tiailed) between band combinations and *TSI*

Band combinations	Regression Coefficients(R^2)	Band combinations	Regression Coefficients(R^2)
Band 4/ Band 5	0.71**	Band 2+Band 3	0.54**
Band 3/ Band 5	0.68**	Band 2+Band 4	0.58**
Band 2/ Band 6	0.60**	Band 2+Band 6	0.50**
Band 1/ Band 6	0.59**	Band 3-Band 4	0.40**
Band 1/ Band4	0.47**	Band 3-Band 5	0.40**
Band 3/ Band 2	0.53**	Band 4-Band 5	0.50**
Band 6/ Band 5	0.58**	Band 4-Band 6	0.49**
Band 1-Band 3	0.48**	Band 5-Band 6	0.49**

** presents the significance level <0.01

TableS5 Comparison of *TSI* calculating from our XGBoost model and historical records from earlier national investigation by Wang and Dou (1998)

Lake	Our XGBoost	Earlier national investigation by Wang and Dou (1998)			
	TSI [a]	SDD (m)	Chl-a (μg L^{-1})	TP(mg L^{-1})	*TSI* [b]
Dongting Lake	57.57	0.5-1.2	0.548-3.31	0.02-0.04	40.36 – 48.07
Poyang Lake	57.29	1.36	0.5-3.47	0.144-4.736	38.93 -57.50
Chaohu Lake	53.98	0.15-0.25	11.81	0.121	68.07 -65.56
Taihu Lake	53.17	0.425-1	7.76	0.082	59.67 -55.45
Jingpo Lake	56.24	2	8.32	0.46	55.98 -55.98

[a] Note that the *TSI* calculating from our XGBoost model were averaged among the datasets in spring, summer and autumn, respectively; [b] The TSIs were calculated by Eq(1-4).

Figure S1 Relationships between in situ and derived *TSI* for both model training and testing samples using reflectance of B1-B6 as input variables by support vector machine (a), XGBoost (b) and random forest (c), as well as their errors (d).

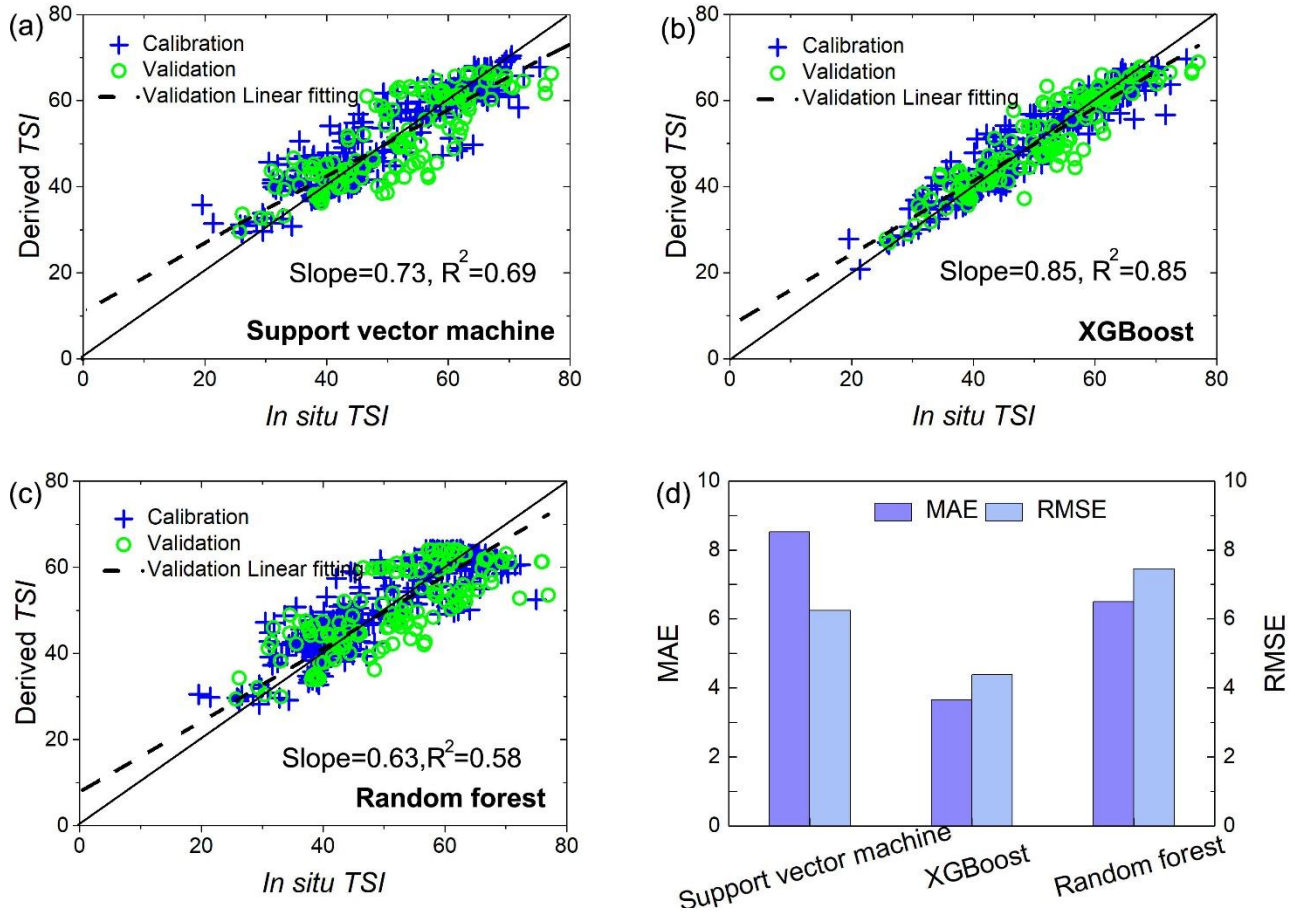

Figure S2 The annual Gross Domestic Product (GDP) and population of China.

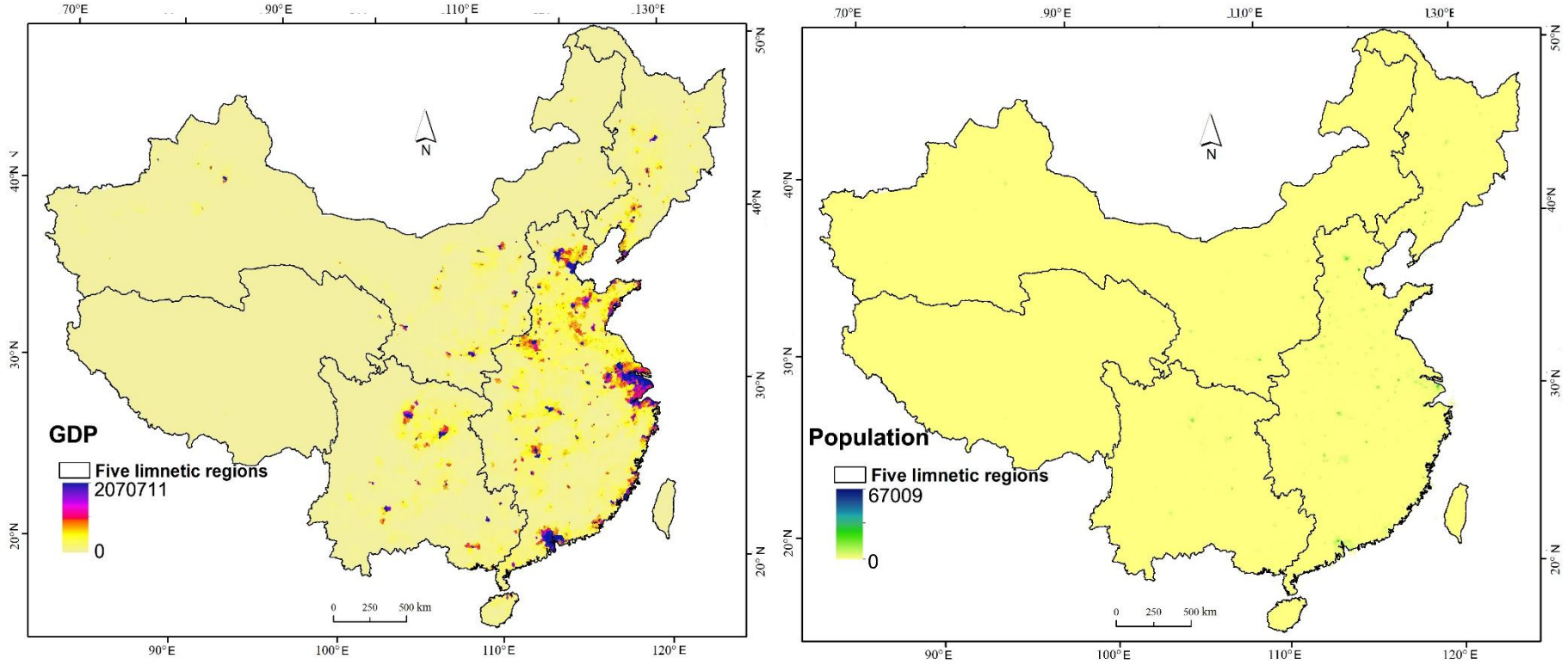

Figure S3 (a) The annual total solar radiation intensity of China; (b) The annual sunshine hours of China (Figures are from the Physical Geography of China).

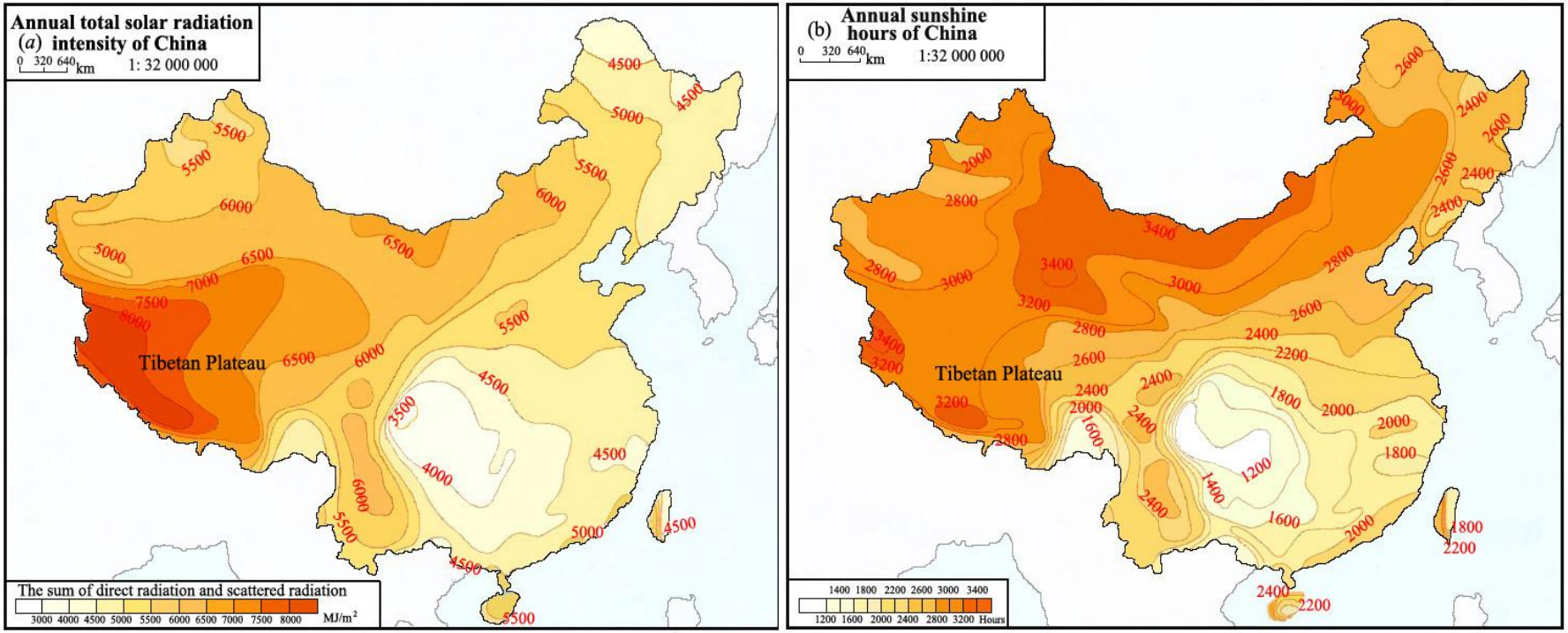

Figure S4 (a) The annual precipitation of China; (b) The annual temperature of China

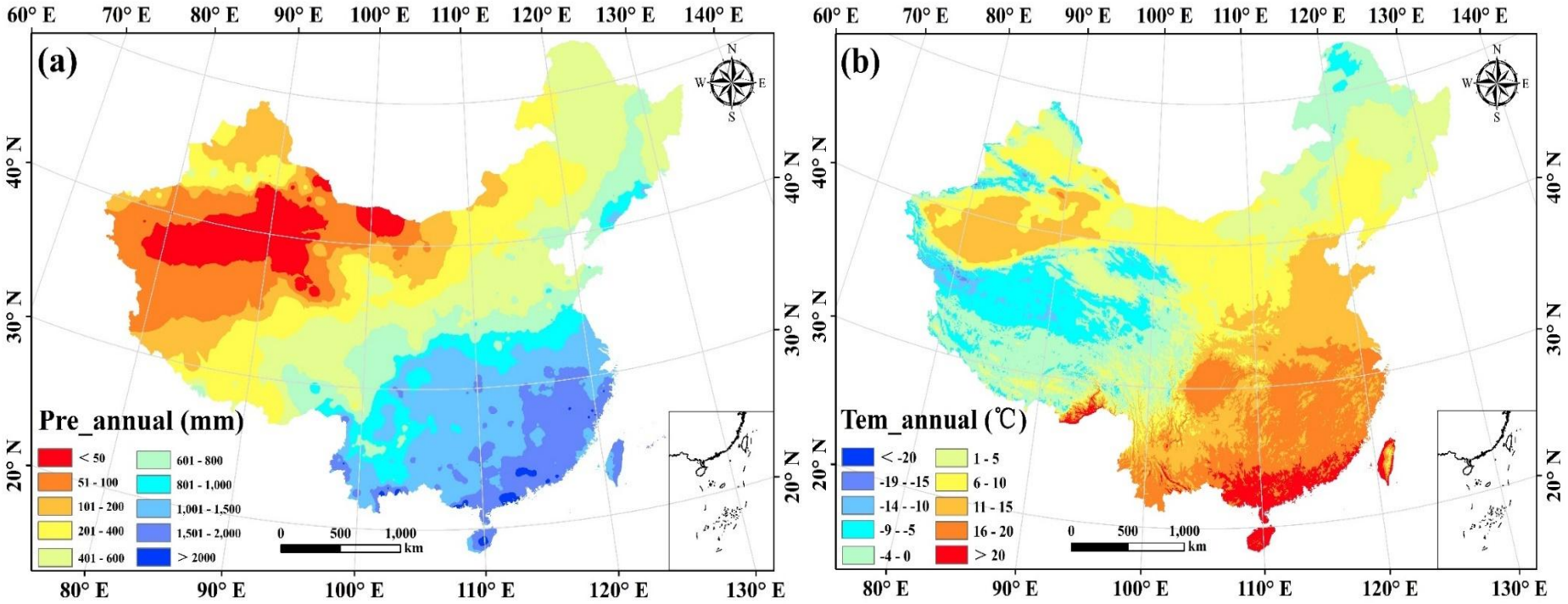

Figure S5 Spatial variations of *TSI* calculated by our XGBoost model in typical lakes in five limnetic regions, i.e., Taihu Lake, Hulun Lake, Fuxian Lake, Qinghai Lake and Chagan Lake.