Remote Quantification of the Trophic Status of Chinese Lakes

Sijia Li [a,b,c], Shiqi Xu[a], Kaishan Song[a], Tiit Kutser[c], Zhidan Wen[a*], Ge Liu[a], Yingxin

Shang[a], Lili Lyu[a], Hui Tao[a], Xiang Wang[a], Lele Zhang[a], Fangfang Chen[a]

[a] *Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun*

130102, China. P. R

[b] *Key Laboratory of Space Ocean Remote Sensing and Application, Ministry of Natural*

Resources, National Satellite Ocean Application Service, Beijing 100081, China

[c]*Estonian Marine Institute, University of Tartu, Mäealuse 14, 12618 Tallinn, Estonia*

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

Abstract: Assessing eutrophication in lakes is of key importance, as this parameter constitutes a major aquatic ecosystem integrity indicator. The trophic state index (*TSI*), which is widely used to quantify eutrophication, is a universal paradigm in scientific literature. In this study, a methodological framework is proposed for quantifying and mapping *TSI* using the Sentinel Multispectral Imager sensor and fieldwork samples. The first step of the methodology involves the implementation of stepwise multiple regression analysis of the available *TSI* dataset to find some band ratios, such as blue/red, green/red, and red/red, which are sensitive to lake *TSI*. Trained with in situ measured *TSI* and match-up Sentinel images, we established the XGBoost of machine learning approaches to estimate *TSI*, with good agreement (R^2=0.87, slope=0.85) and fewer errors (MAE= 3.15 and RMSE=4.11). Additionally, we discussed the transferability and applications of XGBoost in three lake classifications: water quality, absorption contribution, and reflectance spectra types. We selected the XGBoost to map *TSI* in 2019-2020 with good quality Sentinel-2 Level-1C images embedded in ESA to examine the spatiotemporal variations of the lake trophic state. In a large-scale observation, 10-m *TSI* products from investigated 555 lakes in China facing

27 eutrophication and unbalanced spatial patterns associated with lake basin characteristics,

28 climate, and anthropogenic activities. The methodological framework proposed herein

29 could serve as a useful resource toward a continuous, long-term, and large-scale

30 monitoring of lake aquatic ecosystems, supporting sustainable water resource

31 management.

32 **1 Introduction**

33 Lakes, as valid sentinels of global or regional responses, are sensitive to

34 anthropogenic activities and climate change (Mortsch et al., 1996; Quayle et al., 2002;

35 Tranvik et al., 2009). The commonly used paradigm for studying eco-environmental

36 monitoring and controlling of lakes is the status of eutrophication (Carlson, 1977). It is

37 a combination of light, heat, hydrodynamics, and nutrients, such as nitrogen and

38 phosphorus, which occurs through a series of biological, chemical, and physical

39 processes of lakes (Guo et al., 2020). As a result of eutrophication, nutrient loading and

40 productivity grow sharply, and even hypoxia and frequent outbreaks of harmful algal

41 blooms are likely to produce toxins (Paerl et al., 2008, 2011). These processes can cause

42 serious degradation of water quality and are detrimental to the ecosystem services

43 functionality of lakes and reliable supply of drinking water (OECO, 1982). Once the

44 eutrophication phenomenon becomes intense, ecological imbalances generally follow

45 (Smith et al., 2006). Hence, knowledge of eutrophication process can provide us with an

46 understanding of the structure and function of lake ecosystems that give rise to

47 environmental changes. We can then predict future trends and develop appropriate

48 mitigation strategies.

49 Several lakes experience eutrophication processes because of excessive nutrient

50 enrichment (Lund, 1967; Smith et al., 1999; Wetzel, 2001). At the global scale, 63.1%

51 of lakes larger than 25 km^2 are eutrophic and 54% of Asian lakes (Wang et al., 2018), as

well as 53% of European lakes (ILEC et al., 1994). Lake eutrophication has become a global water quality issue affecting most freshwater ecosystems (Matthews, 2014). Currently, many pollutions control measures and management strategies have been implemented that are specific to individual lakes or to lakes, in general (USEPA, 2002). However, there is still insufficient information to address lake eutrophication related to environmental disturbances or changes. Realization of lake eutrophication has been a serious situation for some lakes; therefore, we provided some reasons to suggest the need for large-scale research. First, different environmental factors control the trophic status of lakes at local and multiple scales (e.g., Wiley et al., 1997). Specifically, biotic factors may dominate the eutrophic state of individual lakes, and we can understand the mechanism processes by lake-specific sampling. In contrast, abiotic factors and their linkages are pivotal factors that determine lake biogeochemistry at multiple scales (Sass et al., 2007). It is often necessary to study a number of lakes with different characteristics and catchments to understand the mechanisms of spatio-temporal patterns. Therefore, an up-scaling study of trophic status is required to understand the evolution prospects of lakes in response to changes in global and regional environments. Second, multi-year environmental and climatic conditions require long-term field studies and observations to understand the temporal pattern in important trophic status processes. In addition, relatively large datasets are needed considering the spatial extent because environmental factors are integrated to determine the trophic status of lakes. It can promote data organization and enable us to address an emergency and establish scientific measures for water resource management (Cunha et al., 2013; Smith and Schindler, 2009). Thus, eutrophication should be rapidly assessed using easy-to-analyze indices and enforcement methods for large-scale and high-frequency applications.

Evaluating the trophic state of lakes has been an important topic for decades

77 (Carlson, 1977; Smith and Schindler 2009). The traditional method uses chlorophyll-a,

78 transparency, nutrients, and other variables as water quality indicators by field in situ

79 sampling and laboratory measurements (Rodhe, 1969). Subsequently, Carlson (1977)

80 introduced a numerical *TSI* that should have replaced descriptive values like

81 "oligotrophic," "mesotrophic," or "eutrophic". The replacement has not occurred, but

82 the *TSI* proposed by Carlson is a common method to determine the trophic state level of

83 aquatic environments (Aizaki et al., 1981). The traditional method for calculating *TSI* is

84 based on collected in situ data. The sampling itself and subsequent laboratory

85 measurements are labor-intensive and expensive, often also logistically difficult to

86 perform. This limits our capability to monitor hundreds or thousands of lakes for

87 eutrophication, not speaking about the majority of 117 million of lakes on Earth

88 (Verpoorter et al. 2014). Moreover, the TSI calculated for one or a few discrete samples

89 do not represent spatial distribution of TSI within (especially larger) lakes. This could

90 limit the large-scale assessment of eutrophication as well as the understanding of

91 biogeochemical cycles.

92 　　Satellite remote sensing is a useful tool for monitoring inland waters (Palmer et

93 al 2015). Ocean water-color sensors, such as Medium Resolution Imaging Spectrometer

94 (MERIS) or Ocean and Land Colour Instrument (OLCI) have too low spatial resolution

95 (300 m) for majority of lakes on Earth. Land remote sensing seosor like Landsat

96 Operational Land Imager (OLI), Sentinel-2 Multispectral Imager (MSI; 10-60 m) and

97 Satellite pour l'Observation de la Terre (SPOT) with high spatial resolution (5–30 m) are

98 not designed for water remote sensing (lack critical spectral bands, SNR is not sufficient

99 for water, etc.). Compared to OLI and SPOT sensors, MSI has a more adequate

100 radiometric resolution (12-bits) and 13 spectral bands, including four visible and SWIR

101 channels (Drusch et al., 2012). Inland water TSI has been produced for large lakes using

102 MODIS sensor (Wang et al 2018). However, this study is for more than 2000 large lakes

103 (due to the spatial resolution of the sensor). The Copernicus Land Monitoring Service

104 has started to produce TSI for lakes large enough to be mapped with 100 m pixel size

105 using Sentinel-2 MSI. However, this product is available only for Europe and some

106 parts of Africa.

107   Instead of individual parameters, several studies (e.g., Morel and Prieur, 1977;

108 Gurlin et al., 2011; Huang et al., 2014; Sass et al., 2007; Thiemann & Kaufmann, 2000;

109 Yin et al. 2018) have also provided empirical relationships expressed as band

110 combinations or baseline methods to acquire Chl-a, Secchi or nutrients related to

111 potential *TSI* calculations in regional lakes. However, the accuracy of these empirical

112 relationships for transferring knowledge from some representative lakes to large-scale

113 lake groups is limited by large uncertainties (i.e., in areas with different water quality

114 concentrations and atmospheric component influences, fewer lakes can be used with

115 more heterogeneous influences and uniform algorithms) (Oliver et al., 2017).

116 Considering the requirement of a uniform and universal relationship to quantify the

117 trophic status of lakes, an alternative method using high-frequency and spatial

118 resolution of the sensor is a significant challenge. Recently, technological developments,

119 such as machine learning algorithms, have allowed the usage of remotely sensed

120 imagery to successfully investigate water quality parameters using artificial intelligence

121 (Reichstein et al., 2019; Pahlevan et al., 2020; Cao et al., 2020). The potential

122 application and development of machine learning for remote quantification of water

123 quality is attributed to the following advantages: requirement of little prior knowledge,

124 rich features can be captured, and robust relationships can be obtained. These processes

125 avoid bias and uncertainty from the regional environmental background as well as

126 complications due to atmospheric components of traditional remote sensing-derived

127 relationships over large-scale, i.e. for multiple lakes. Given the novel application of

128 remote sensing and machine learning, this is a gap to fill for large-scale research of

129 monitoring trophic states.

130   Environmental issues fueled by rapid economic growth in China have significantly

131 increased in the last three decades. Lake eutrophication is a serious issue, with large

132 variability in terms of trophic status and optical properties. However, most studies (Jin,

133 2003, 2005; Fragoso et al., 2011; Huang et al., 2014) have addressed eutrophication

134 concerns in only a single lake or two lakes since the 1990s. It is acknowledged that a

135 rapidly growing economy and anthropogenic activities (e.g., elevated nutrient loading

136 and increasing air pollution) accelerate the aging process of lakes (Wu et al., 2011; Shi

137 et al., 2020). Therefore, it is critical to objectively assess the trophic status and pay

138 attention to protect the aquatic environment. We aim to provide a robust machine

139 learning algorithm and remote sensing flowchart from simultaneously retrieved *TSI*

140 over a wide range of bio-optical compositions in different lakes. The objectives of our

141 study were to: (1) examine biogeochemical parameters and assess trophic status, (2)

142 calibrate and validate the *TSI* model using different machining learning algorithms from

143 MSI-imagery derived remote sensing reflectance spectra *(Rrs)*, with different lake

144 classifications; and (3) quantify and map the trophic status of typical 555 lakes in five

145 Chinese limnetic regions.

146 **2 Materials and methods**

147 **2.1** *Study area and sampling process*

148   China is located in the east of Asia with a land area of 9,600,000 square kilometers

149 and a population of over 1.4 billion. The terrain of China descends from west to east in

150 three steps. Due to a vast territory span, this country has diverse climatic, geographical,

151 and geological conditions. There are 2,693 natural lakes (with area >1.0 km^2) that are

152   distributed in China (Ma et al., 2011). Protection and sustainable management of these

153   lakes have been priorities, considering the degradation of water quality over several

154   decades. In this study, a total of 45 lakes were visited and 431 samples were collected in

155   early April 2016 to late October 2019 (Table S1 and Fig. 1), which was the highest

156   productive season, as identified by Carlson's *TSI* model. These datasets were analyzed

157   and published in (Li et al., 2021; Song &Li et al., 2019; Song et al., 2020). Our lake

158   dataset was collected from various types of lakes across China, and efforts were made to

159   examine lake trophic status from a wide range of water quality parameters, lake sizes

160   (0.5 to 4, 256 km^2), lake elevation (10 to 4, 525 m), and climatic zones (Song &Li et al.,

161   2019). In the field, some small-size lakes were sampled in the middle, and signal sample

162   was used to represent the water qualities, while others were sampled at multiple

163   locations evenly distributed over the lake. The water samples were collected

164   approximately 0.5 m below the surface, and then stored in 1 L amber HDPE bottles and

165   kept in a portable refrigerator (4 ℃) before being transported to the laboratory. During

166   the sampling process, the Secchi disk depth (SDD, m) was measured using a

167   black-and-white Secchi disk. The pH and electrical conductivity (EC, μs cm^{-1}) were

168   recorded using a portable multi-parameter water quality analyzer (YSI 6600, 170 U.S).
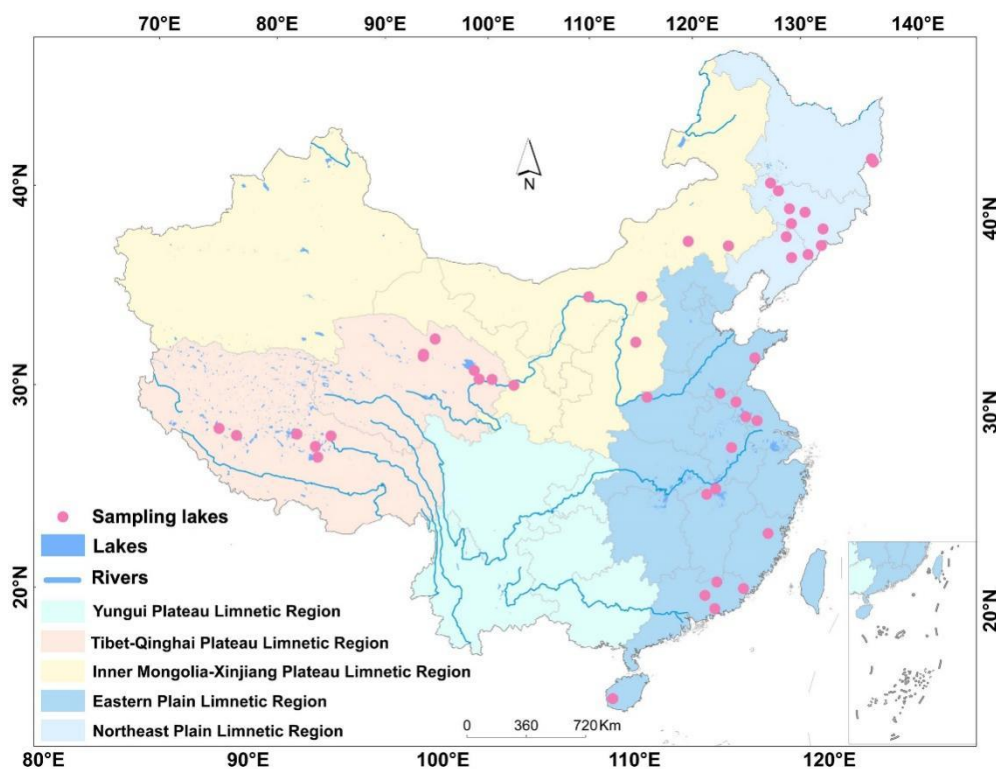
Figure 1：**Location of lake sites.**

171 **2.2 Laboratory analysis**

172   A transferred portion of each bulk water sample was immediately filtered with

173 0.45-μm pore size Whatman cellulose acetate membrane filters in the laboratory. It is to

174 be noted that some remote Tibet and Qinghai lake samples had to be filtered during

175 fieldwork. Chlorophyll-a (Chl-a) was extracted from the filters using a 90 % buffered

176 acetone solution at 4° C under 24 h dark conditions. According to the SCOR-UNESCO

177 equations (Jeffrey and Humphrey, 1975), the concentration of Chl-a ($\mu g\ L^{-1}$) was

178 determined using a UV-2600PC spectrophotometer at 750 nm, 663 nm, 645 nm, and

179 630 nm. Dissolved organic carbon ($mg\ L^{-1}$) concentrations were determined using a

180 total organic carbon analyzer. Total nitrogen (TN) and total phosphorus (TP)

181 concentrations ($mg\ L^{-1}$) were measured using a continuous flow analyzer (SKALAR,

182 San Plus System, the Netherlands) using a standard procedure (APHA/AWWA/WEF,

183 1998). In addition, total suspended matter (TSM, $mg\ L^{-1}$) concentrations were obtained

184 gravimetrically using pre-combusted 0.7-μm pore size Whatman GF/F filters. All

8

185 preprocesses (e.g., filtration and concentration quantification) of all water samples were

186 undertaken within two days in the laboratory. The procedures are provided in detail in

187 Li et al. (2021).

188　　The bulk samples were again filtered through a 0.7-µm pore size glass fiber

189 membrane (Whatman, GF/F 1825-047) to retain particulate matter. The water from

190 particulate matter measurements was then filtered through a 0.22-µm pore size

191 polycarbonate membrane (Whatman, 110606) in order to measure chromophoric

192 dissolved organic matter (CDOM) absorption of each sample. According to the

193 quantitative membrane filter technique (Cleveland and Weidemann, 1993), the light

194 absorption of total particulate matter $a_p(\lambda)$ can be separated into phytoplankton pigment

195 absorption $a_{ph}(\lambda)$, non-algal particles $a_d(\lambda)$, and CDOM absorption $a_{CDOM}(\lambda)$. The

196 optical density (OD) of the particulate matter retained in the filters was measured using

197 a UV-2600PC spectrophotometer at 380–800 nm, with a blank membrane as a reference

198 at 380–800 nm. The filters were then bleached using a sodium hypochlorite solution to

199 remove phytoplankton pigment and measured again using a spectrophotometer. Finally,

200 the phytoplankton pigment absorption $a_{ph}(\lambda)$ was calculated by subtracting $a_d(\lambda)$ from

201 the total particulate matter $a_p(\lambda)$. The absorption coefficients of the optical active

202 substance (OACs) were calculated according to Song et al. (2013).

203 **2.3 Trophic status assessment of lakes**

204　　Several studies have proposed different indices of the lake trophic state (Aizaki et al.,

205 1981; Carlson, 1977). Carlson's trophic state index used five variables, such as Chl-a,

206 TP, TN, SDD, and chemical oxygen demand (COD), to characterize the trophic state.

207 However, there are no optical characteristics for TN, TP and COD to manifest in

208 changes of remote sensing reflectance, which may bring more uncertainties or errors.

209 Thus, Chl-a, TP, and SDD were selected to assess the trophic status according to the

210 modified Carlson's trophic state index (*TSI*). The *TSI* can be calculated using individual

211 TSI_M(Chl-a), TSI_M(SDD), and TSI_M(TP) using the following equations:

212
$$TSI_M(Chl\text{-}a) = 10 \times \left(2.46 + \frac{\ln Chl - a}{\ln 2.5}\right) \tag{1}$$

213
$$TSI_M(SDD) = 10 \times \left(2.46 + \frac{3.69 - 1.52 \times \ln SDD}{\ln 2.5}\right) \tag{2}$$

214
$$TSI_M(TP) = 10 \times \left(2.46 + \frac{6.71 + 1.15 \times \ln(TP)}{\ln 2.5}\right) \tag{3}$$

215
$$TSI = 0.54 \times TSI_M(Chl\text{-}a) + 0.297 \times TSI_M(SDD) + 0.163 \times TSI_M(TP) \tag{4}$$

216 Where, the *TSI* below 30 correspond to oligotrophic waters, above 50 are eutrophic and

217 *TSI* between 30 and 50 in mesotrophic (Carlson, 1977).

218 **2.4 Muti-Spectral Instrument imagery and atmospheric correction**

219   Sentinel-2A/B MSI imagery was acquired from the Copernicus Open Access Hub

220 of the European Space Agency. Altogether, 210 scenes of cloud-free Level-1C images

221 covering the lakes were downloaded with a time window of ±7 days from in situ

222 measurements. The Case 2 Regional Coast Color processor (C2RCC) was used to

223 remove atmospheric effects. An average of 3×3-pixels centered at each in situ sampling

224 station was used in the further analysis. All the processes were performed using the

225 Sentinel Application Platform (SNAP) version 7.0.0. A flowchart of the process is

226 shown in Fig. 2.

227
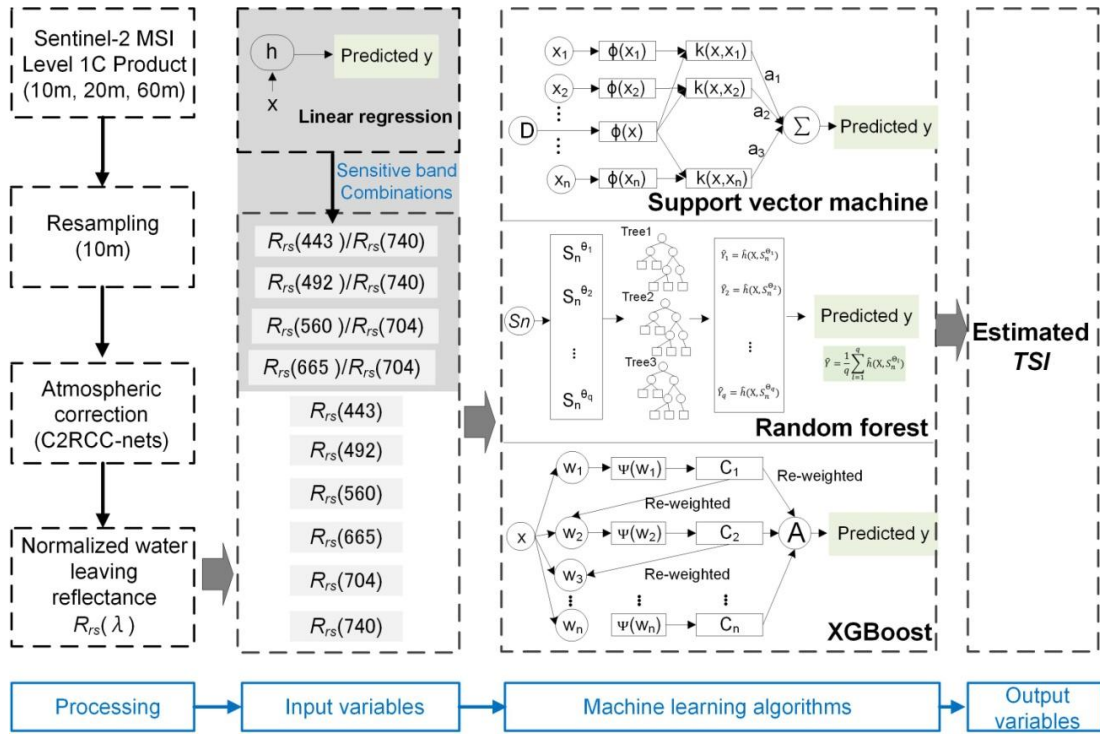
Figure 2：Workflow of the Sentinel-2 MSI data and machine learning algorithms for estimating *TSI*

230

2.5 Machine learning algorithms

As a branch of artificial intelligence, the application of machine learning is growing in the field. Machine learning can automatically analyze huge chunks of data, develop optimal models, generalize algorithms, and make predictions. These approaches have been applied in a variety of eco-environmental and remote sensing fields (Mountrakis et al., 2011; Pahlevan et al., 2019). Hence, we employed four representative machine learning algorithms, namely linear regression (LR), support vector machine (SVM), XGBoost (XGB), and random forest (RF) (Supplementary data, methods), to establish a *TSI* model. To strengthen the robustness, band combinations sensitive to *TSI* were determined by LR (Fig. 2), and were added to the procedure of machine learning algorithms as input variables. Subsequently, the output variable was the predicted *TSI*. The in situ measured samples were then randomly divided into a calibration dataset (70%, 287 lake samples) and validation dataset (30%, 144 lake

244 samples) using MATLAB software. The *TSI* modeling procedure considering machine

245 learning and Multiple Linear Regression (MLR) was processed using the R software.

246 **2.6 Classifications of lakes**

247 In order to provide further feasibility for the application and availability of the *TSI*

248 model, the in situ measured samples were classified in three ways (Fig. 3):

249 a) based on water quality: Salinity classification referred to the threshold value of

250 electrical conductivity (named EC, EC=1000 μS cm^{-1}) (Duarte et al., 2008), following

251 which the lakes were divided into brackish lakes (*N*=100 samples) and fresh water lakes

252 (*N*=331 samples). Dissolved organic carbon (DOC) in global lake water classification

253 referred to the volume weighted averaged DOC level of global lakes (3.88 mg L^{-1})

254 according to Toming et al., (2020), following which lakes were divided into high DOC

255 lake (*N*=224 samples) and low DOC lake (*N*=207 samples).

256 b) based on optical absorption contribution: Optical absorption classification

257 referred to Prieur and Sathyendranath (1981), where the total light absorption of water

258 can be separated from phytoplankton pigment absorption, non-algal particles, and

259 CDOM absorption, respectively. The relative percentage of absorption contribution of

260 OACs can be divided into phytoplankton-type (Phy-type) lakes (*N*=54 samples),

261 non-algal particles-type (NAP-type) lakes (*N*=109 samples), CDOM-type lakes (*N*=177

262 samples), and mix-type lakes (*N*=91 samples).

263 c) based on reflectance spectra: In order to discern the different optical

264 characteristics of lakes, the derived MSI reflectance was clustered using the k-means

265 clustering approach with a gap statistic (Neil et al., 2018). We identified 431 MSI

266 reflectance *Rrs*(λ) spectra for three branches (Table S3), and the *Rrs*(λ) spectra are
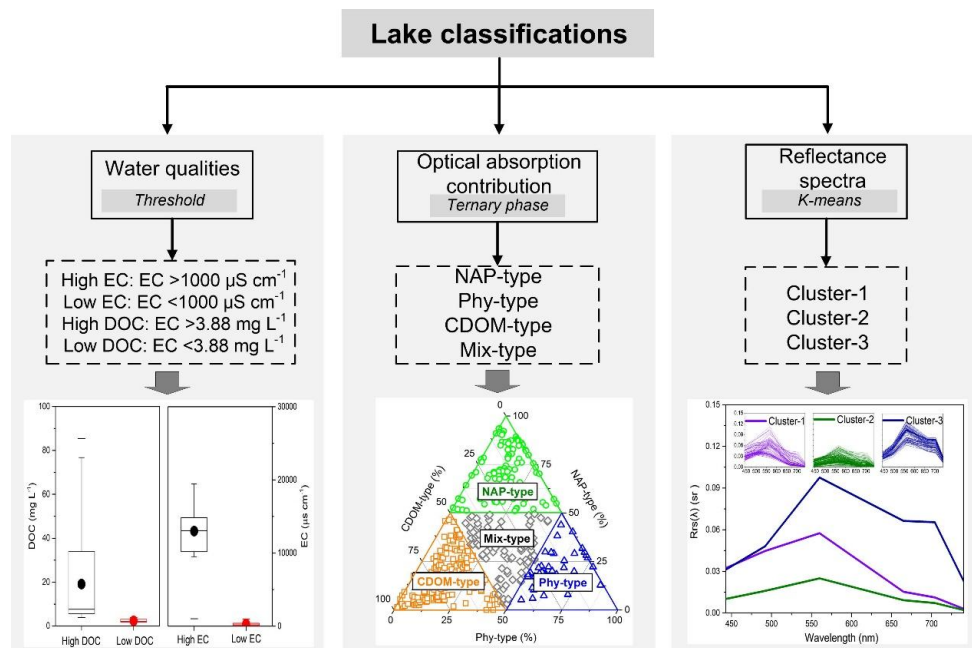
267 shown in Fig.3.

Figure 3：Lake classifications considering three ways, i.e., water quality, optical absorption contribution and reflectance spectra. ANOVA analysis was conducted in different classifications (*p*<0.001) (Table S3).

2.7 Statistical analyses and accuracy assessment

Statistical analysis, including descriptive statistics, correlation (*r*), regression (R^2), and ANOVA analyses, were implemented with Statistical Program for Social Science software (version 16.0; SPSS, Chicago, IL, USA). Correlation and regression analyses were used to examine the relationships between the water quality parameters and absorption coefficients of OACs as well as the *TSI* model calibration and validation. The differences in trophic status, EC classification, DOC classification, absorption coefficients of OAC classification, and MSI reflectance spectra classification for *TSI* model validation were assessed using one-way ANOVA. The significance level was set at $p<0.05^*$. The mean normalized error (MAE) and root mean square error (RMSE) were used to assess the performance of the *TSI* model (Supplementary data, accuracy assessment).

3 Results

3.1 Aquatic environmental scenery

287   The water qualities and bio-optical properties of our samples covered a wide range,

288   revealing different geographical environmental scenery (Tables S1 and S2-4). The EC

289   and DOC concentration showed high variability, ranging for example, from 3345.31 μs

290   cm^{-1} (TuoSu, TS20) in Tibet-Qinghai region to 0.17 μs cm^{-1} (Qingnian, QN2) in

291   Northeast region. For the water quality parameters to characterize *TSI*, the Chl-a

292   concentration ranged from 0.12 to 100.22 μg L^{-1}, with the highest value recorded in

293   TaiPingChi (TPC5) and the lowest value in NamoCo (NMC36). The range of TP was

294   from 0.003 mg L^{-1} (Erlong, EL8) to 2.17 mg L^{-1} (Dali, DL7), and SDD ranged from

295   0.17 m (Chalhu, CH32) to 9.47 m (NMC36) for surveyed lakes, respectively. Overall,

296   the maximum values of EC, DOC, turbidity, Chl-a, TSM, and SDD were 196782.35,

297   948.4, 723.3, 770.92, 614.58, and 55.71 fold greater than the minimum values,

298   respectively, indicating that our dataset was representative of diverse water qualities.

299   Lake samples were grouped into different classifications based on water quality

300   (e.g., EC and DOC), optical absorption contribution, and reflectance spectra (Table 1

301   and Fig. 3). The results indicated that all water qualities showed significant differences

302   ($p<0.05$) under different lake classifications. For example, brackish lakes showed higher

303   average values of SDD, TP, DOC, and optical attributions of OAC values than those of

304   fresh water lakes, but the turbidity, Chl-a, and TSM concentrations were lower. Lakes

305   equipped with low DOC levels had a low average value of SDD than that of lakes with

306   high DOC levels. NAP-type lakes exhibited the highest average Chl-a and DOC values,

307   whereas Phy-type lakes had the highest average turbidity and TSM values, and the

308   highest average SDD and TP values were recorded in CDOM-type and Mix-type lakes,

309   respectively. For reflectance spectra classifications (Fig. 3), the highest average EC,

310   SDD, and DOC were recorded in cluster-1 lakes, the highest average turbidity and TP

311   was shown in cluster-3 lakes and the highest average TSM was found in cluster-2 lakes.

Table1 (a) Averaged values (Avg.) of water quality and bio-optical properties considering lake classifications and (b) ANOVA analysis (F value) among them

:ations	N	EC	Turbidity	SDD	Chl-a	TP	DOC	TSM	$a_{ph}(440)$
Brackish	100	12986.28	8.83	2.21	4.18	0.45	33.31	8.42	0.23
Fresh	331	302.39	21.75	1.43	8.58	0.07	4.28	19.52	0.56
High DOC	224	5988.93	23.90	1.39	10.42	0.25	19.07	21.50	0.68
Low DOC	207	276.19	12.45	1.85	4.46	0.06	2.29	11.98	0.27
NAP-type	54	5156.02	11.28	1.58	14.26	0.09	18.75	15.99	1.29
Phy-type	109	825.48	43.28	0.65	6.85	0.10	4.75	37.18	0.46
CDOM-type	177	4081.96	4.44	2.43	3.64	0.13	9.70	4.99	0.13
Mix-type	91	3424.07	19.40	1.17	12.05	0.34	16.48	16.22	0.70
Cluster-1	87	6948.28	4.46	2.38	2.64	0.08	17.92	5.76	0.26
Cluster-2	215	2728.71	6.18	2.05	8.57	0.07	7.18	5.81	0.35
Cluster-3	129	1626.05	46.68	0.36	9.19	0.37	12.73	42.59	0.84
Brackish	100	-	18.7^{**}	21.8^{**}	12.0^{**}	68.9^{**}	486.5^{**}	20.4^{**}	16.6^{**}
Fresh	331								
High DOC	224	93.8^{**}	19.8^{**}	10.0^{**}	32.2^{**}	23.3^{**}	-	21.0^{**}	38.0^{**}
Low DOC	207								
NAP-type	54	7.4^{**}	71.6^{**}	46.0^{**}	21.0^{**}	7.1^{**}	13.5^{**}	73.0^{**}	-
Phy-type	109								
CDOM-type	177								
Mix-type	91								
Cluster-1	87	220.9^{**}	17.9^{**}	25.2^{**}	312.7^{**}	11.0^{**}	18.5^{**}	18.9^{**}	26.1^{**}
Cluster-2	215								
Cluster-3	129								

The unit of TN, TP, DOC and TSM is mg L^{-1}; EC is μs cm^{-1}; Chl-a is μg L^{-1}; turbidity is NTU (nephelometric turbidity unit). Significance levels are reported as significant (noted with *, $0.05 > p > 0.01$) or highly significant (noted with **, $p < 0.01$).

313 **3.2 Trophic status assessment**

314   The trophic status of 45 lakes across China, from where in situ samples were

315 collected, was evaluated (Fig. 4a). Our results showed that there were 13 oligotrophic

316 (3.02 %), 199 mesotrophic (46.17 %), and 219 eutrophic (50.81 %) samples. Because

317 our samples were collected in different seasons and eutrophication is time-dependent,

318 the *TSI* values of samples within a lake were averaged. It can be shown that only five

319 lakes accounting for 11.1% of investigated lakes were characterized with an

320 oligotrophic status, 17 lakes accounting for 37.8 % were mesotrophic, and 23 lakes

321 accounting for 51.1 % were characterized with eutrophic status. These eutrophic lakes

322 were distributed in the eastern region of China (Fig. 4b), and were associated with a

323 highly concentrated human population and economic development. Moreover, the

324 ANOVA results showed that the *TSI* of lake samples were significantly different

325 considering lake classifications (Fig. 4c, and d).
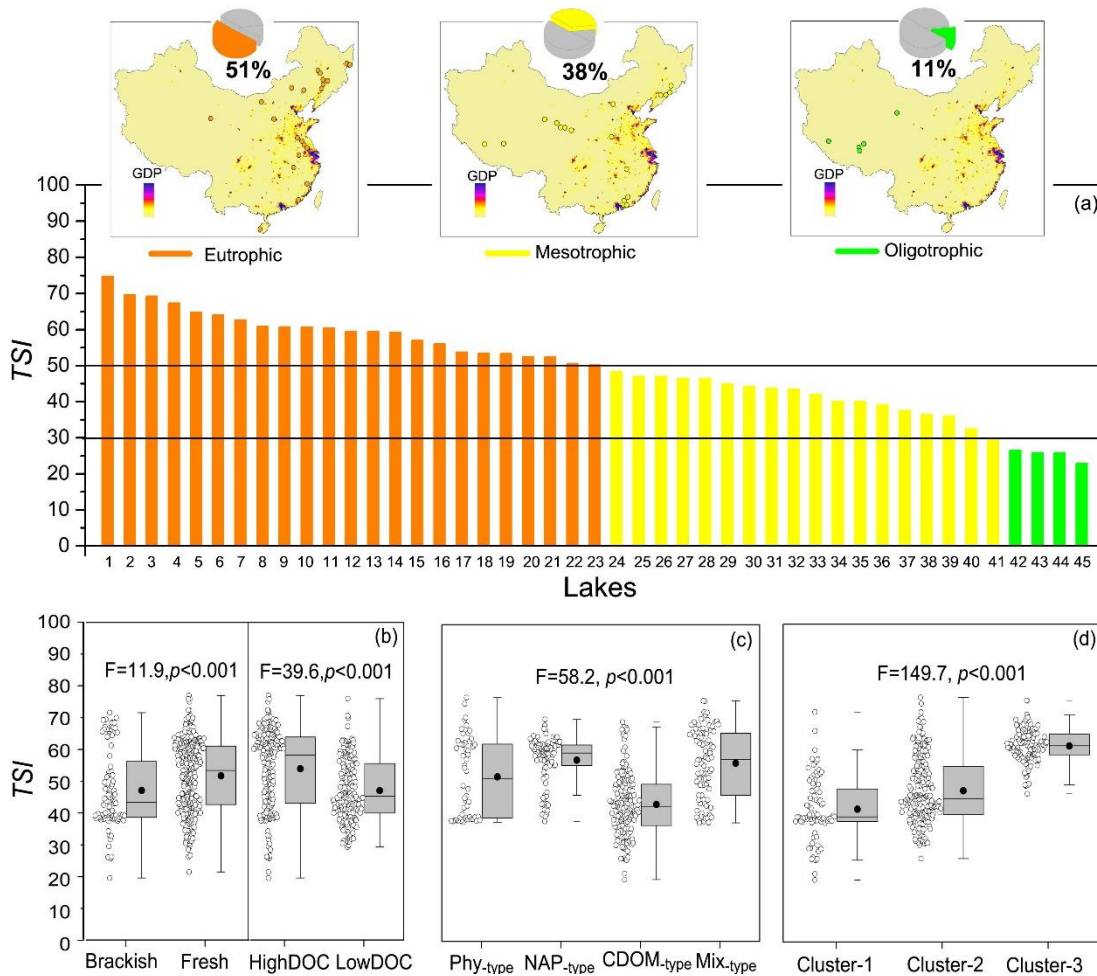
Figure 4： **(a) is the averaged *TSI* in collected samples from lakes across China and their spatial distribution. The number of lakes can be found in TableS1. The box plots of *TSI* at different classifications of water quality (b), optical absorption contribution types (c) and reflectance spectra (d). The balls beside the boxes are the lake samples, and the black balls in the boxes represent the mean values. The horizontal edges of the boxes denote the 25th and 75th percentiles; the whiskers denote the 10th and 90th percentiles.**

3.3 Calibration and validation of *TSI* model

In this section, multiple linear regression was used to identify significantly sensitive spectral variables related to *TSI* (Table 2 and Fig. 2). Of the band combinations validated in the study (*N*=144), the blue/red [*Rrs*(443)/*Rrs*(740), *Rrs*(492)/*Rrs*(740)], and green/red [*Rrs*(560)/*Rrs*(704), *Rrs*(665)/*Rrs*(704)] band ratios showed a good regression coefficient ($R^2 > 0.59$) with *TSI* (Table S5). These band combinations provided certain sensitive spectral variables that responded to the lake eutrophic status. Hence, to strengthen the robustness of the three machine learning models, the blue/red

343 and green/red combinations above were considered as the input variables as well as six

344 spectral variables ($Rrs(\lambda)$ at 443, 492, 560, 665, 709, and 740 nm). Likewise, the output

345 variables were estimated using *TSI* to examine the performances (Fig. 5). The results

346 showed that when XGBoost was applied to the validation data (*N*=144), the

347 performance of the model was excellent (R^2=0.87, slope=0.85) with low errors (MAE=

348 3.15, RMSE=4.11). The support vector machine (R^2=0.71, slope=0.77, MAE=4.67,

349 RMSE=6.11) and random forest (R^2=0.85, slope=0.84, MAE=3.31, RMSE=4.34)

350 models also showed significant performance. These results demonstrate the potential of

351 using XGBoost by considering band combinations to derive *TSI* from Sentinel products.
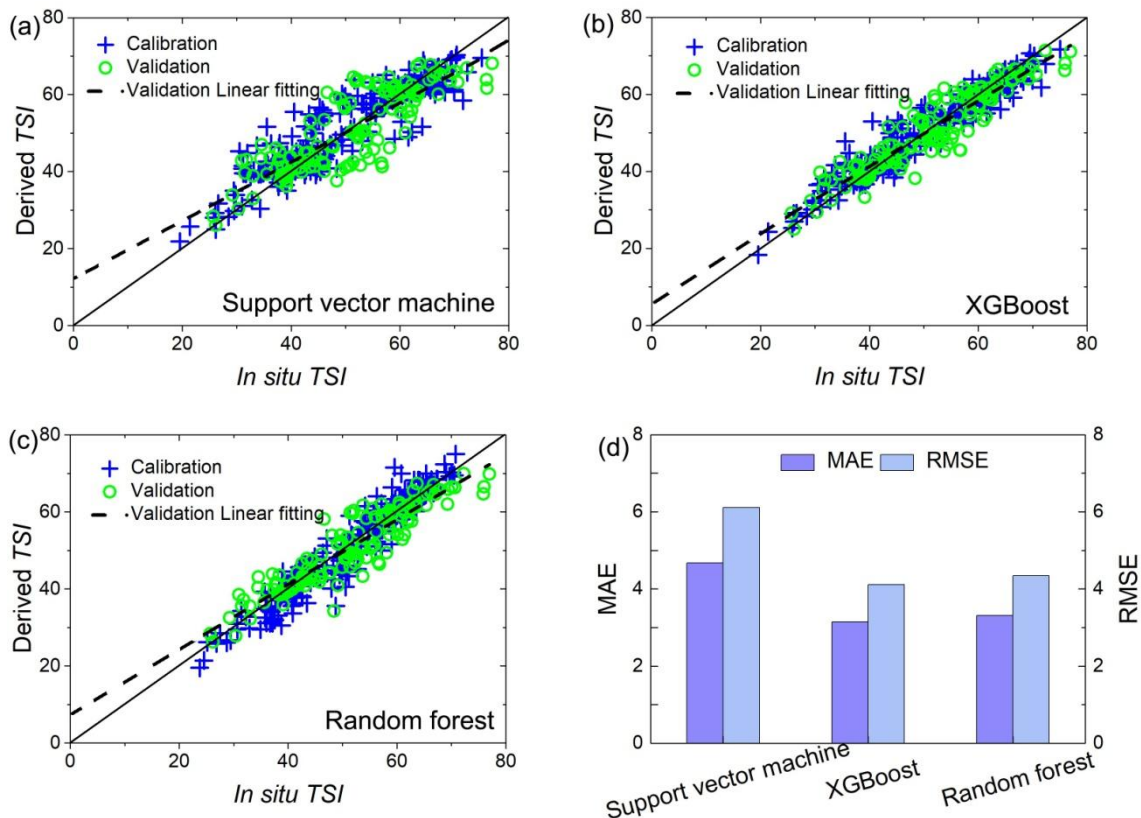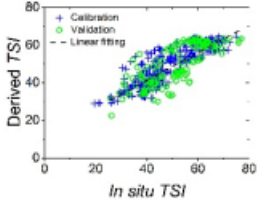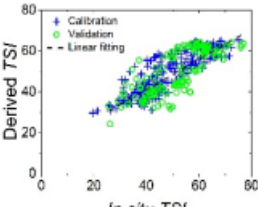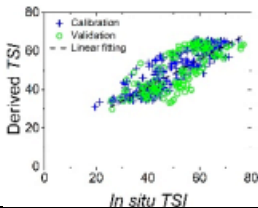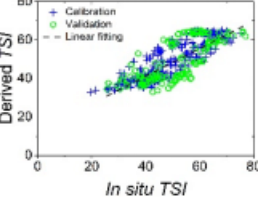

352
353 **Figure 5： Relationships between in situ and derived *TSI* for both model training**
354 **and testing samples by support vector machine (a), XGBoost (b) and random**
355 **forest (c), as well as their errors (d).**

Table 2 Multiple linear regression between measured- and estimated- *TSI* from the MSI spectral bands after using C2RCC processor

Band combinations	Datasets	N	Fitting equation	R^2	Errors	Plots figures
Band 1/ Band 6 （Blue/Red）	Calibration	287	$TSI = -8.51\ln[Rrs(B1)/Rrs(B6)] + 63.47$	0.76	MAE = 6.45 RMSE = 5.85	
	Validation	144	$TSI_{derived} = 0.73 \times TSI_{in\ situ} + 11.868$	0.61	MAE = 6.26 RMSE = 7.48	
Band 2/ Band 6 （Blue/Red）	Calibration	287	$TSI = -8.87\ln[Rrs(B2)/Rrs(B6)] + 67.91$	0.77	MAE = 4.57 RMSE = 5.74	
	Validation	144	$TSI_{derived} = 0.74 \times TSI_{in\ situ} + 11.751$	0.60	MAE = 6.32 RMSE = 7.57	
Band 3/ Band 5 （Green/Red）	Calibration	287	$TSI = -13.63\ln[Rrs(B3)/Rrs(B5)] + 67.26$	0.77	MAE = 4.55 RMSE = 5.70	
	Validation	144	$TSI_{derived} = 0.72 \times TSI_{in\ situ} + 12.44$	0.59	MAE = 6.39 RMSE = 7.66	
Band 4, Band 5 （Red/Red）	Calibration	287	$TSI = -44.15 \times [Rrs(B4)/Rrs(B5)] + 108$	0.80	MAE = 4.39 RMSE = 5.43	
	Validation	144	$TSI_{derived} = 0.72 \times TSI_{in\ situ} + 12.32$	0.59	MAE = 6.85 RMSE = 7.94	

3.4 *TSI* model application to lake classifications

The *TSI* model calculated by XGBoost was assessed by comparing derived and in situ *TSI* considering different lake classifications (Fig. 6). We aimed to provide a universal *TSI* model and evaluate its feasibility in different aquatic environments. Significant agreement (slope>0.91, R^2>0.91) between derived and in situ *TSI* was observed in lakes with high DOC levels (DOC>3.88 mg L^{-1}) and EC values (EC>1000 μS cm^{-1}), with low errors. For lakes classified by different absorption contributions, the NAP-type (slope=0.98, R^2=0.88) and Phy-type (slope=0.82, R^2=0.92) samples generally showed a positive derived performance than those of Phy-type, CDOM-type, and Mix-type, respectively. In addition, a significant relationship between derived and in situ *TSI* can be described for lakes with cluster-1 reflectance spectra, with slope=0.91, R^2=0.87, RMSE=2.87, and MAE=2.29.
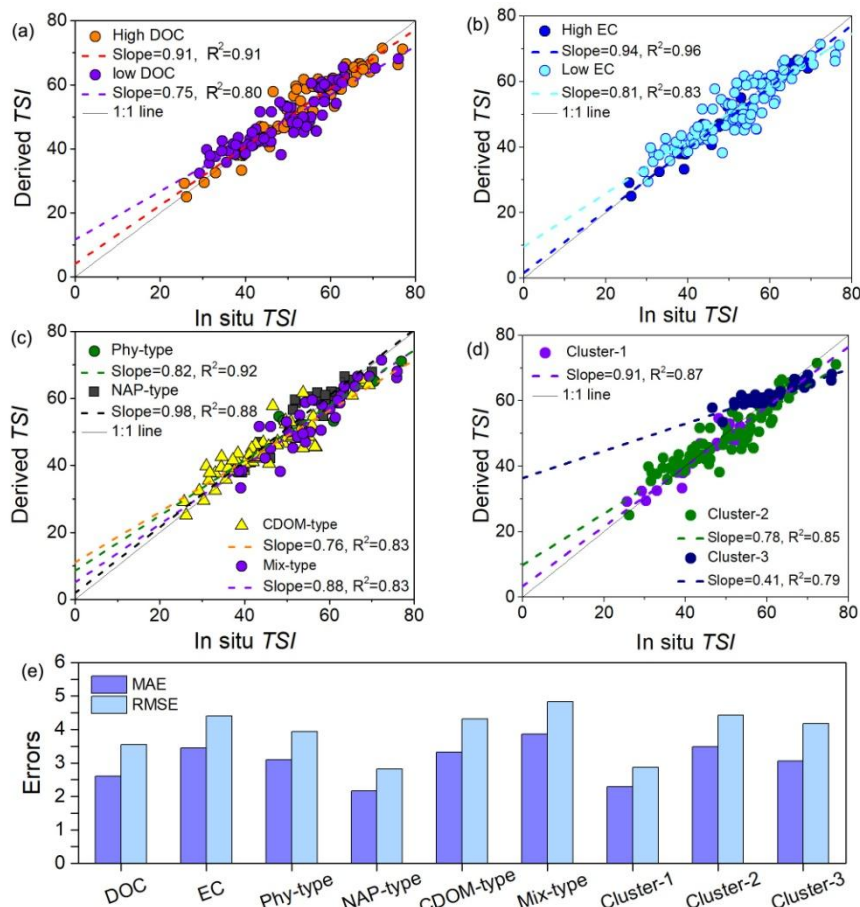
Figure 6：Scatter plots of derived- and in situ- *TSI* by XGBoost for validation samples (*N*=144) according to lake classifications, such as water quality (DOC and

EC) (a-b), absorption contribution (c), reflectance spectra(d) with the 1:1 line (red solid) and errors (e).

3.5 Spatial and Seasonal patterns of trophic states: Five lake limnetic regions

Previous studies have demonstrated that some lakes disappeared or increased numbers recently according to statistics from Ma et al. (2011). Thus, we selected some representative and stable lakes (N=555) to qualify spatial trophic states using the XGBoost algorithm. The preprocessing of MSI data were referred to the Fig.2, and a total of 139 cloud-free images in spring (Apr. and May.), summer (Jul. and Aug.) and autumn (Sep. and Oct.) covered investigated lakes were acquired. According to the different geographic and limnological types in China, lakes were divided into five limnetic regions (Wang and Dou 1998, Early National Investigation): Eastern Plain Limnetic Region (EPLR, N=123), Northeast Plain Limnetic Region (NPLR, N=37), Inner Mongolia-Xinjiang Plateau Limnetic Region (IMXPLR, N=56), Yungui Plateau Limnetic Region (YGPLR, N=15), and Tibet-Qinghai Plateau Limnetic Region (TQPLR, N=324) (Fig. 1 and Supplementary data).

In general, there were significant seasonal variations in eutrophic state for lakes from the EPLR (F=39.56, p<0.001) and TQPLR (F=5.0, p<0.05) (Fig. 7). The averaged *TSI* in EPLR were 56.37 (Spring), 57.73(summer) and 54.26 (autumn) indicating serious eutrophication of investigated lakes, consistent with the results from Li et al., (2022). Recognizing that over 94% of the Chinese population lives in eastern watersheds with great demands of water use, this may be due to different water qualities management in provincial scales. Likewise, we found there was spatial heterogeneity of *TSI* results in TQPLR and some of which were the widespread saline lakes in Qinghai-Tibet Plateau with high reflectance in satellite images. On the contrary, there were no seasonal differences of *TSI* for lakes from IMXPLR, NPLR and YPLR, respectively. The eutrophic lakes dominated the proportions of the investigated lakes in

401 the EPLR (93.5 %), followed by the NPLR (89.2 %), YGPLR (86.7 %), IMXPLR

402 (69.6%) and TQPLR (3.7%) (Fig.8). It can be also found that mesotrophic lakes were

403 found in the decreased order of TQPLR (45.7 %), IMXPLR (30.4%), YGPLR (13.3 %),

404 NPLR (10.8 %) and EPLR (6.5 %), respectively. In comparison, most oligotrophic lakes

405 (50.6%) were distributed in the TQPLR.

406
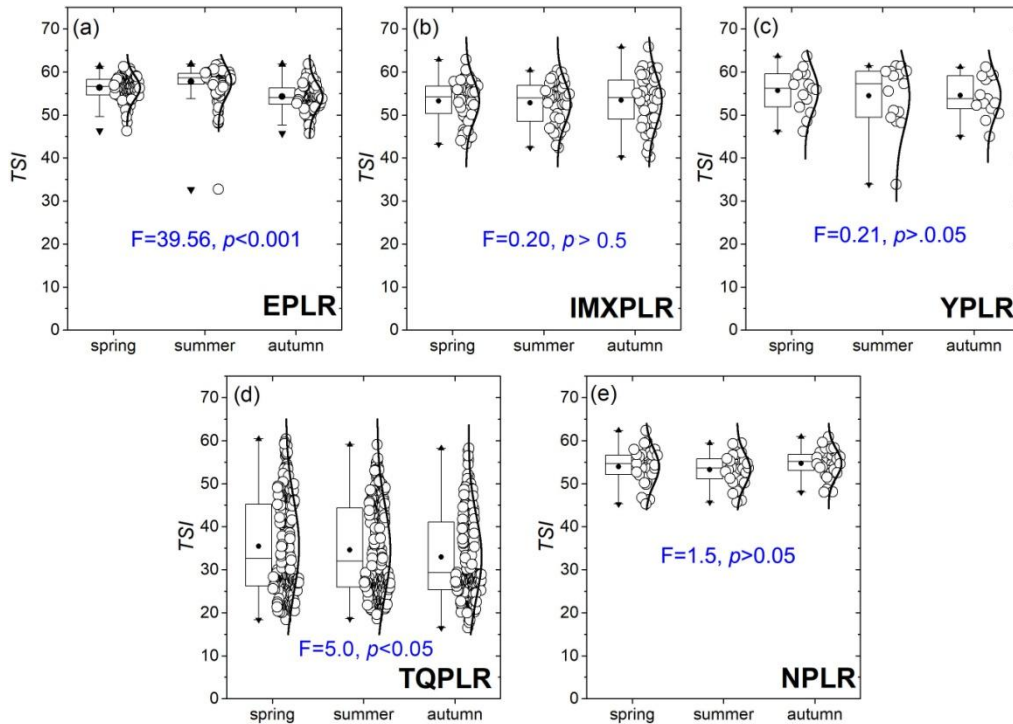
407
408 **Figure 7：Box plots of *TSI* derived from XGBoost model in investigated lakes**
409 **from the five limnetic regions (Wang & Dou 1998), i.e., (a) EPLR, (b) IMXPLR, (c)**
410 **YPLR, (d) TQPLR and (e) NPLR. The black line and balls in the boxes represent**
411 **the median and mean values, respectively. The horizontal edges of the boxes denote**
412 **the 25th and 75th percentiles; the whiskers denote the 10th and 90th percentiles.**
413
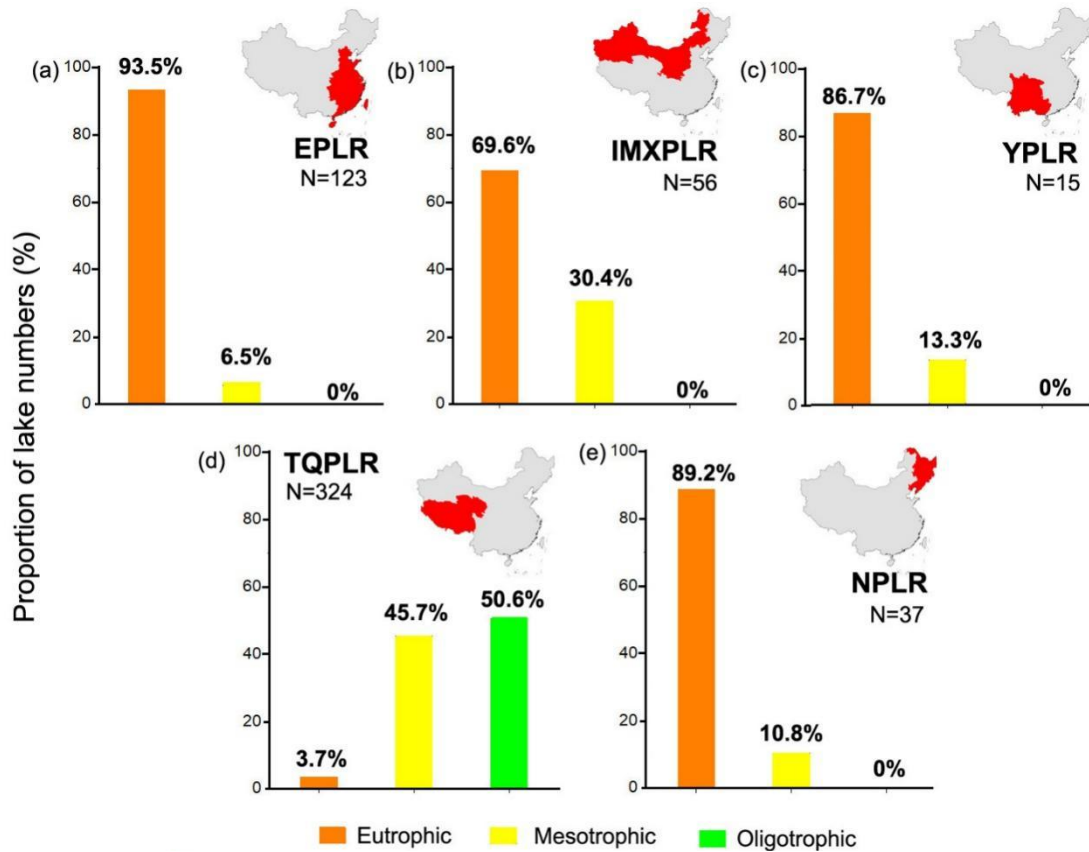
Figure 8：The proportions of lake numbers (%) for different trophic state in the five limnetic regions (Wang & Dou 1998), i.e., (a) EPLR, (b) IMXPLR, (c) YPLR, (d) TQPLR and (e) NPLR. N represents the lake numbers.

4 Discussion

4.1 Remote-sensed and machine-learning-based *TSI* model

Traditional approaches to quantitatively characterize the trophic status rely on field measurements of trophic parameters, for example, Chl-a, nutrients, and SDD, to calculate the *TSI* (Carlson, 1977). It is difficult and costly to make field measurements in lakes in remote locations. The *TSI* calculation does not need all of these trophic parameters but just one, for example, Chl-a (Thiemann and Kaufmann, 2000), SDD (Olmanson et al., 2008; Song et al., 2020), TP (Kutser et al., 1995) and total absorption coefficients (Lee et al., 1999; Shi et al., 2019), etc. There have been many lake studies (Chl-a and SDD, Sheela et al., 2011; Chl-a, SDD and TP, Song et al., 2012) where two or three water quality parameters were mapped, which would allow to subsequently

430 gather them to calculate comprehensive *TSI*. Although these studies provided the

431 potential to evaluate the trophic status of lakes, *TSI* is a synthetic indicator that is

432 affected by biological, physical, and chemical factors that co-vary in most instances.

433 Huang et al. (2014) also tried to derive *TSI* using remote sensing spectrum reflectance,

434 but the accuracy was not completely usable. It shows that variability in remote sensing

435 estimates of the *TSI* are not bad.

436 With advances in artificial intelligence technology and the increasing use of

437 computer applications in recent years, machine learning has become a useful tool for

438 monitoring aquatic environments by remote sensing (Mountrakis et al., 2011). It allows

439 us to develop and evaluate a machine-learning-based *TSI* model that addresses quality

440 and accuracy problems more effectively (Li et al., 2021). Hence, we propose a new

441 approach to directly characterize the trophic status and accurately reflect spatial

442 variations in this study, but should also be conveniently available for the different lake

443 classifications (Figs. 5, 6). Using machine learning algorithms, in order to improve the

444 robustness and applicability of the *TSI* model, a sufficient database of trophic state

445 parameters (*N*=431) was collected from lakes with different biogeochemical

446 characteristics, such as water quality, absorption contributions of different optically

447 active substances, and reflectance spectra (Table1). We first used B1-B6 reflectance as

448 input variables of machine learning algorithms, and XGBoost showed a significant

449 performance with R^2 and a slope of 0.85 (Fig. S1). The SVM performed worse than

450 XGBoost and random forest, and did not produce the sufficient performance. This is

451 because the latter model are integrated algorithms with trees are unpruned and diverse,

452 signifying the high resolution in the feature space and smoother decision boundary.

453 There was no optical response bands or appropriate band ratios for *TSI*. We thus used a

454 multiple linear regression to find some suitable sensitive band combinations responding

455   to the *TSI*, which made it possible to develop a robust machine-learning-based *TSI*

456   model. It is important to note that the blue/red [$Rrs(443)/Rrs(740)$, $Rrs(492)/Rrs(740)$],

457   and green/red [$Rrs(560)/Rrs(704)$, $Rrs(665)/Rrs(704)$] band ratios were significantly

458   correlated with *TSI* (Table 2). This result indicated that the blue/red and green/red band

459   ratios were more sensitive to the *TSI*, although the nutrients and SDD had no optical

460   response. It was known for decades that the blue part of spectrum is useless when water

461   itself is not blue (i.e. outside of ocean or very oligotrophic mountain lakes), owing to

462   the noneffective atmospheric correction and complex reflectance signals. However, our

463   dataset to train *TSI* models contain the samples from blue and oligotrophic Tibetan lakes,

464   which are like the oceanic environments (Liu et al., 2021). The blue bands responding

465   to *TSI* were thus used in this study. Most empirical Chl-a estimation studies adopted

466   red/near infrared (NIR) band ratios to calibrate models using reflectance signatures

467   (Gitelson et al., 1992). Similarly, empirical SDD retrieval models provided by previous

468   studies used empirical algorithms or models to figure out what bands should work the

469   best considered the following ratios: blue/green, red/blue plus red/green, and red/blue

470   plus blue (Bindling et al., 2007), and Red/Blue ratio plus Blue (Kloiber et al., 2002).

471   Kutser et al. (1995) also built a TP retrieval model using the red and NIR ratios, which

472   is consistent with Chl-a empirical models. Overall, it is not surprising for our *TSI* model

473   to have strong correlations with the blue/red and green/red band ratios because the *TSI*

474   incorporates the optical properties.

475   For this reason, we used MSI bands in the visible band ratios at six bands,

476   considering the comprehensive spectrum information about the trophic status of lakes as

477   input variables (Fig. 2). The three representative machine learning *TSI* models improved

478   the accuracy of the traditional linear regression (Table 2 and Fig. 5), and the results

479   were better than those obtained with B1-B6 reflectances as input variables (Fig. S1). As

480   a type of supervised machine learning algorithm, linear regression can be used to obtain

481   certain learning criteria as expressions ($y=w_0+w_1{\times}x_1+{\ldots}+w_p{\times}x_p$) about the optimal w_i

482   solution. However, for complex targeted tasks, the fitting ability of linear regression is

483   limited, and it cannot represent the real situation well. For example, a support vector

484   machine can map data to another space, which can use a linear regression to distinguish

485   the categories well. In complex environments (real world in machine learning), such as

486   our large-scale database collected from different lakes (Fig. 1), there are various

487   environmental factors as well as different seasons within a lake, that have an impact on

488   the trophic parameters and optical characteristics of lakes (Wen et al., 2016). Likewise,

489   we found that the enhanced input variables, like the band ratios, if appropriately

490   corrected for the *TSI*, resulted in a better performance (Fig. S1). This is consistent with

491   some applications of machine learning algorithms (Cao et al., 2020), in which the

492   performance of machine learning was reduced when covariances of input features were

493   incorporated. This allows us to find more interesting *TSI*-correlated band ratios for MSI

494   imagery in machine learning.

495   Several machine learning algorithms generally have different advantages and

496   applicability owing to their different main principles (Cao et al., 2020; Li et al., 2021).

497   This can be found in our results of the validation exercise, which showed that XGBoost

498   provided stable *TSI* estimates, with a slope close to 1 and a good fitting coefficient of

499   the measured and derived values (R^2=0.87, slope=0.85, MAE= 3.15, RMSE=4.11) (Fig.

500   4). Similarly, we can also find an excellent performance (R^2=0.85, slope=0.84,

501   MAE=3.31, RMSE=4.34) for estimating *TSI* values by the random forest algorithm.

502   This was likely because it is a summation of all weak learners, weighted by the native

503   log odds of error. In the case of boosting, we make decision trees into weak learners by

504   allowing every tree to make only one decision before prediction (Chen et al., 2016). In

some cases, XGBoost outperformed random forest. In addition, the support vector machine performed worse than XGBoost and random forest (Fig. 4). Li et al. (2021) used a support vector machine to estimate Chl-a concentrations with a relatively small dataset of 32 samples and 273 samples, respectively. This is consistent with the recent process in the development of support vector machines and has many advantages for remote sensing applications with a small number of training datasets. Overall, the remote sensing and machine learning-based *TSI* model aims to reduce the dependence of traditional field measurements, while also providing a cost-effective approach to rapidly quantify the trophic state.

4.2 *TSI* model for lake classifications

We validated the XGBoost *TSI* model considering different scenarios of lake classification, for example, water quality, optical absorption contributions, and reflectance spectra (Figs. 2 and 6). The results indicate three application scenarios for our model with low errors. The first one is of the XGBoost *TSI* model, which in particular, performed well (slope>0.91, R^2>0.91) in high DOC (>3.88 mg L^{-1}) and EC (>1000 μS cm^{-1}) lakes (Fig. 6). We found that lakes with high EC level correspondingly showed a high DOC level (Table 1), for example, high average EC value of 5156.02 μS cm^{-1} and high average DOC value of 18.75 mg L^{-1} for NAP-type lakes. These brackish or saline lakes were distributed in the Tibet-Qinghai Plateau Region (e.g., KLK20, TS21, QHH22, SLC32, BMC34, ZRNMC36, NMC37) and Inner Mongolia-Xinjiang Plateau Limnetic Region (e.g., DL8, HSH10, DH17, HL18, WLSH16) (Table S1). Our results are in agreement with those of previous studies that DOC and EC of inland waters located in semi-arid region can be attributed to the evapo-concentration and accumulation processes (Curtis and Adams, 1995) as well as anthropogenic activities. Further, it can be observed that oligotrophic lakes accounting for 11.1% were also

530 distributed in the Tibet-Qinghai (Fig. 4).

531 Secondly, we found that our XGBoost *TSI* model performed well if the trophic

532 parameters that correlated to the TSI_M(Chl-a) or TSI_M(SDD) dominated the lake

533 classifications. Specifically, the high Chl-a (averaged 14.26 μg L^{-1}) and a_{ph}(440)

534 (averaged 0.26 m^{-1}) levels in NAP-type lakes showed the best performance (slope=0.98,

535 R^2=0.88) than those of other optical absorption contribution classifications (Fig. 6). In

536 fact, there was a negligible difference in the performance for application in Phy-type

537 and NAP-type lakes. For the third scenario, for the reflectance spectra classification,

538 cluster-1 lakes with low TSM (averaged 5.76 mg L^{-1}), turbidity (averaged 4.46 NTU),

539 and a_d(440) (averaged 0.26 m^{-1}) level, and high SDD level (average 2.38 m) also

540 showed good performance (slope=0.91, R^2=0.87) (Fig. 6). In general, *TSI*, as a

541 comprehensive index incorporating the optical properties of itself, was calculated using

542 trophic state parameters [(TSI_M(Chl-a), (TSI_M(SDD), and TSI_M(TP) in Eq. 7]. Our

543 XGBoost *TSI* model performed best in the present study, which confirmed that the

544 performance was mostly determined by biogeochemical environments in larger-scale

545 regions. We cannot explain the dependence of the *TSI* model on the physico-optical

546 properties. From another point of view, it can be inferred that the XGBoost *TSI* model

547 applications mostly correlated to the Chl-a and SDD because of their high weight

548 allocation in *TSI* equation.

549 Although we conducted a large-scale *TSI* observation across Chinese lakes, and if

550 the XGBoost could also perform well for signal lake is required to evaluate. Hence, the

551 *in situ* measured samples were classified in three scenarios, with XGBoost *TSI* model

552 was analyzed. Overall, in future work, for lakes mainly located in high elevation and

553 arid region with high DOC/EC levels, the input band combinations responding to color

554 dissolved organic matter (CDOM, Green/Red) could be added in XGBoost *TSI* model.

This is because that CDOM and DOC generally showed positive correlations for investigated lakes (Song et al., 2013), and CDOM is one of optical active substance. It also confirmed that non-algal particles could cover the reflectance signals and impact on the model performance in second and third scenarios. More classifications based on reflectance spectra (Spyrakos et al., 2018) and water color index (Wang et al., 2018) should be first used and then developed corresponding models for high turbid lakes.

4.3 Trophic status in five limnetic regions

According to this study more than 50% of lakes were eutrophic, indicating a long-standing status of eutrophication (Fig. 4), as seen by the mapping of 555 lakes by our XGBoost *TSI* model (Fig. 7). Some lake investigations undertaken earlier in China during 1978–1980 concluded that 41.2% lakes of eutrophication in China (Jin, 2003), during 1988-1992 demonstrated that 51.2% lakes (Wang & Dou, 1998), during 2001-2005 indicated that 84.5% lakes, during 2011-2019 showed that 50% lakes (Wen et al., 2019) were eutrophic or undergoing eutrophication. In our study, some historical records of Chl-a, SDD and TP from in comparison to earlier national investigation by Wang and Dou (1998) were collected in typical lakes, e.g., Dongting Lake, Poyang Lake, Chaohu Lake, Taihu Lake and Jingpo Lake, respectively (Table S6). Evidently, Chinese lakes have deteriorated considerably in terms of water quality at an alarming rate for typical lakes, e.g., Jingpo Lake, Dongting Lake and Poyang Lake, during past ~22 years (Table S6). Lake eutrophication is influenced by both natural (hydrological processes, topography, lake depth, and buffer capacity) factors as well as anthropogenic factors (land-use changes, urbanization construction, and domestic and industrial pollution) (Müller et al., 1998). A large-scale overview of lake eutrophication indicated there was a significant difference (ANOVA, F=255.2, $p<0.001$) in the five limnetic

580　regions (Wang & Dou 1998). Owing to the imbalanced development of economic

581　(Fig.S2, GDP and population), geological topography (Fig.S3, solar radiation intensity

582　and sunshine hours) and climate (Fig.S4, annual temperature and precipitation), it was

583　not surprising that the eutrophic lakes were generally distributed in the Eastern Plain

584　Limnetic Region and Northeast Plain Limnetic Region, as well as that the oligotrophic

585　lakes were found in the Tibet-Qinghai Plateau Limnetic Region (Fig.4 and Fig.7).

586　　　Considering the natural factors for the distributions of Chinese lake eutrophication,

587　we could suppose some possibility that lake depth and lake hydrological processes

588　cause the eutrophication of lakes in China. Previous studies (Wang & Dou 1998; Huang

589　et al., 2014) have demonstrated that lakes with mean depths > 5 m in China are mainly

590　located in the Yungui Plateau Limnetic Region, Inner Mongolia-Xinjiang Plateau

591　Limnetic Region, and Tibet-Qinghai Plateau Limnetic Region, whereas almost all lakes

592　located in the Eastern Plain Limnetic Region are shallow. Both these lakes in the

593　Eastern Plain Limnetic Region are hydraulically connected with the Yangtze River with

594　a temporary residence time of approximately 30 days (Fig. S7). In shallow lakes, due to

595　wind waves or disturbance by fishes, the phosphorus/nitrogen nutrients stored in the

596　sediment can be easily resuspended and released into the overlying water (Niemistö et

597　al., 2008). Consequently, an increased frequency of algal blooms can be found in

598　Eastern Plain Limnetic Region, in lakes, such as Taihu, Chaohu, and Hongze (Qin et al.,

599　2019; Yao et al., 2016). Instead, deeper lakes, such as the ones in YGPLR and TQPLR,

600　possess relatively good buffer capacity for waste-water runoff (Huang et al., 2014).

601　Carvalho et al. (2009) found that Chl-a levels decreased with lake water depth and

602　geographic location. Qin et al., (2020) and Tong et al., (2006) demonstrated that

603　phosphorus reduction can mitigate eutrophication in deep lakes, and more efforts to

604　reduce both N and P need to be undertaken in shallow lakes. This can be demonstrated

605   in our case of Fuxian Lake with changeable eutrophication levels, with an average depth

606   of 87 m, which was the deepest lake in southwest China (Fig. S7). In addition, the

607   annual precipitation and air temperatures were relatively high in the EPLR (Fig. S4).

608   Hydrological and meteorological processes can scour land surfaces and bring nutrients

609   into lakes via rivers. Therefore, lake ecosystems were strongly related to the lake basin

610   morphology and its hydrologic characteristics, which were higher in shallow lakes than

611   in deep ones (Köiv et al., 2011).

612   On the other hand, human-induced eutrophication, for example, agricultural

613   fertilization (Carpenter, 2008; Huang et al., 2017), aquaculture (Guo & Li, 2003) and

614   sewage discharge (Paerl et al., 2011), are increasing terrestrial nutrient phosphorus but

615   not nitrogen concentration inputs (Schindler et al., 2008). We suspected that two

616   interactive factors, such as land-use and nutrient variations cause lake eutrophication,

617   because this can be found in our investigation of distributed lakes in the EPLR in

618   comparison to earlier national investigation by Wang and Dou (1998). Many lakes in the

619   EPLR that were naturally connected with rivers have been modified to paddy fields, and

620   some small lakes have become isolated for lake aquaculture. For instance, Lake

621   Dongting was artificially shifted from being river-fed to dammed/isolated. Logically it

622   should a dam can settle down the suspended matter and nutrients via river inputs. But

623   the shallow characteristic and wind mixing influence process significantly increased the

624   probability of eutrophication (Liu et al., 2019). In EPLR and NPLR, 94% of China's

625   population lives in 43% of its eastern region, which visually demonstrates the

626   distribution of GDP with a densely populated east (Fig. S2). Owing to the requirements

627   of water source utilization, the EPLR has lost one-third of its original lake areas to

628   cropland since 1949 (Yin and Li, 2001). Lake aquaculture is highly active in these areas.

629   These processes could lead to terrestrial nutrient loading into lakes, from either

630   agriculture or aquaculture, and thereby alter the trophic state levels of a lake ecosystem.

631   In 2019, the total fish catch in Hubei was 4,695 tons; in Jiangxi was 432, 25 tons; in

632   Anhui was 588,135 tons; 2,314,603 and 4,841,159 tons in Anhui and Jiangsu in the east,

633   respectively (China rural statistical yearbook).

634   Although we have not systematically analyzed the effects of environmental factors

635   on trophic status, some of the sparse existing comparative literature supported certain

636   spatiotemporal patterns. It should be emphasized that China has been facing serious lake

637   eutrophication and unbalanced distributions. Almost invariably, lake ecosystem health

638   would still be impacted by stresses integrating anthropogenic and overexploitation of

639   catchment resources. Consequently, addressing the issue of worsening eutrophication

640   requires a better understanding of the environmental interactive mechanisms in the

641   future.

642

643   **4.4 Limitations, uncertainties and future**

644   Toward the United Nation's Sustainable Development Goal (SDG) 6.3.2, satellite

645   imagery and machine learning still provides great potential for evaluating water qualities

646   state from global observations, particularly in developing countries. Machine learning

647   algorithms could serve as good alternatives for empirical and semi-analytical algorithms

648   to quantify on large-scale spatial applications, which could avoid or minimize the

649   errors.Our results further demonstrated machine learning algorithms could improve the

650   accuracy of water quality models (e.g., *TSI*) when the linear regression was used to find

651   sensitive band combinations with red/red edge bands. Previous studies (Li et al., 2021,

652   2022) found red and red edge band could help us to quantify the spatial and temporal

653   changes of Chl-a concentration or a synthetic parameter-such as *TSI* with high Chl-a

654   weight ratio-from regional lakes. It is enable us to use sentinel-2 or similar sensors

655   equipped with these bands to capture records of *TSI* dynamics.

656   As a medium-resolution (10~60 m) satellite, Sentinel-2 MSI offers the potential to

657   monitor small-size lakes and produce reliable *TSI* estimates. However, there are

658   significant obstacles in generating a Sentinel-2 (~10m) lake *TSI* distribution, including

659   the acquisition of high quality atmospheric corrected Rrs(λ) and massive computational

660   overhead by C2RCC processor (Li et al., 2023). C2RCC processor designed for waters

661   based on neural networks is data-driven approach and uses huge datasets collected from

662   *in situ* and simulation measurements. *In situ* reflectance measurements were not

663   conducted in these investigated Chinese lakes when sampling. Our recently study

664   reported that C2RCC (SNAP 8.0) and Polymer (v4.13) processors both performed best

665   with *in situ* field radiometry in typical lakes across China (Li et al., 2023), but the latter

666   could work better when all bands are pooled together in derived algorithms.

667   Considering the growing requirements of *TSI* products, more *in situ* measurements

668   would be required to be added the already-implemented processors in following work.

669   In addition, there is a need for a robust model developed from different locations

670   and optical water types that accounts for the interplay of different water quality

671   parameters. Machine learning *TSI* model required a highly calibrated dataset, including

672   high nutrients (e.g., TP >2.50 mg L^{-1} in this study) and Chl-a concentrations (>100 μg

673   L^{-1} in this study). Likewise, for our developed universal *TSI* model, the feasibility

674   application performances were different considering lake classifications. Hence, the

675   extensive field–lab materials with complex source variations would be required first and

676   water optical typologies further is a good compromise to develop groups of optimized

677   algorithms in future. Nevertheless, we aim to provide technical operation approach,

678   which could prompt more analysis responding to warming climate and anthropogenic

679   activities. The strong linkages between reflectance and several trophic state defining

680 indexes further underscore the potential of remote sensing for resources-limited

681 countries meet their SDG goals.

682

683 **5 Conclusions**

684 Our study presents a novel remote sensing- and machine-learning-based algorithm

685 applied in that allow to retrieve the lake *TSI* from Sentinel-2 MSI imagery. We used a

686 match-up database (*N*=431) over a diverse range of bio-optical regimes to train machine

687 learning algorithms and validated it against the in situ data. The trophic states of 555

688 lakes were then evaluated. These results provide a better understanding how remote

689 sensing and machine learning-based models allow to estimate eutrophication over a

690 large scale of different lakes. Our main findings can be summarized as follows:

691 1) Linear regression enabled us to find certain band combinations sensitive to *TSI*

692 (R^2>0.59), for example, the blue/red [*Rrs*(443)/*Rrs*(740), *Rrs*(492)/*Rrs*(740)] and

693 green/red [*Rrs*(560)/*Rrs*(704), *Rrs*(665)/*Rrs*(704)] band ratios.

694 2) XGBoost algorithm resulted in optimum performance with R^2=0.87 and

695 slope=0.85, considering the low errors (MAE=3.15, RMSE=4.11), compared to the

696 support vector machine and random forest algorithms.

697 3) If there is some preliminary data available from the study area one can improve

698 the performance of the machine learning by dividing the lakes based on high DOC/EC,

699 NAP-type and Phy-type, and cluster-1 reflectance spectra.

700 4) The trophic states of 555 lakes were evaluated in five limnetic regions;

701 eutrophic lakes dominated in Eastern Plain Limnetic Region and Northeast Plain

702 Limnetic Region, and most lakes in Tibet-Qinghai Plateau Limnetic Region were

703 mesotrophic or oligotrophic.

704 In our subsequent research and management, qualification and mapping of *TSI* will

be implemented as a remote sensing and machine learning model in a large-scale study, allowing for an improved performance. In the future, Sentinel-2 MSI data could be used to reveal spatiotemporal variations in lake trophic states in long-term time-series responding to climate and anthropogenic activities.

737

738 References

739 APHA, A. W., Standard Methods for the examination of water and wastewater 20th

740   edition.,https://doi.org/10.2105/AJPH.51.6.940-a,1961.

741 American Public Health Association/American Water Works Association/Water Environment

742   Federation, Washington DC, USA,https://doi.org/10.1080/23267224.1919.10651076,1998.

743 Aizaki, M., Otsuki, A., Fukushima, T., et al., Application of Carlson's trophic state index to Japanese

744   lakes and relationships between the index and other parameters: With 2 figures and 4 tables in

745   the text. Internationale Vereinigung für theoretische und angewandte Limnologie:

746   Verhandlungen, 21(1), 675-681.Res.rep.natl.inst.environ.stud.jpn, 1981.

747 Carpenter, S. R., Brock, W. A., Cole, J. J., et al., Leading indicators of trophic cascades. Ecol. Lett.

748   11(2), 128-138,https://doi.org/10.1111/j.1461-0248.2007.01131.x,2008.

749 Carlson, R. E., A trophic state index for lakes 1. Limnol. Oceanogr. 22(2),

750   361-369.https://doi.org/10.2307/2834910,1977.

751 Chen, T., Guestrin, C., Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm*

752   *sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

753 Cleveland, J. S., Weidemann, A. D., Quantifying absorption by aquatic particles: A multiple

754 scattering correction for glass-fiber filters. Limnol. Oceanogr. 38(6), 1321-1327,

755 https://doi.org/10.4319/lo.1993.38.6.1321,1993.

756 Cunha, D. G. F., do Carmo Calijuri, M., Lamparelli, M. C., A trophic state index for

757 tropical/subtropical reservoirs (TSItsr). Ecol. Eng. 60, 126-134.

758 https://doi.org/10.1016/j.ecoleng.2013.07.058,2013.

759 Curtis, P. J., Adams, H. E., Dissolved organic matter quantity and quality from freshwater and

760 saltwater lakes in east-central Alberta. Biogeochemistry 30(1), 59-76.

761 https://doi.org/10.1007/bf02181040,1995.

762 Doerffer, R., Schiller, H., The MERIS Case 2 water algorithm. Int. J Remote sens. 28(3-4), 517-535.

763 https://doi.org/10.1080/01431160600821127,2007.

764 Duarte, C. M., Prairie, Y. T., Montes, C., et al., CO2 emissions from saline lakes: A global estimate

765 of a surprisingly large flux. J. Geophys. Res.-Biogeosciences 113(G4).

766 https://doi.org/10.1029/2007jg000637,2008.

767 Drusch, M., Del Bello, U., Carlier, S., et al., Sentinel-2: ESA's optical high-resolution mission for

768 GMES operational services. Remote Sens. Environ.120, 25-36,

769 https://doi.org/10.1109/igarss.2007.4423394,2012.

770 Fragoso Jr, C. R., Marques, D. M. M., Ferreira, T. F., et al., 2011. Potential effects of climate change

771   and eutrophication on a large subtropical shallow lake. Environ. Modell. Softw. 26(11),

772   1337-1348, https://doi.org/10.1016/j.envsoft.2011.05.004,2011.

773   Gitelson, A.The peak near 700 nm on radiance spectra of algae and water: relationships of its

774   magnitude and position with chlorophyll concentration. International Journal of Remote

775   Sensing, 13(17), 3367-3373, https://doi.org/10.1080/01431169208904125,1992.

776   Gitelson, A. A., Dall'Olmo, G., Moses, W., et al., 2008. A simple semi-analytical model for remote

777   estimation of chlorophyll-a in turbid waters: Validation. Remote Sens. Environ. 112(9),

778   3582-3593. https://doi.org/10.1080/01431169208904125,1992.

779   Gurlin, D., Gitelson, A. A., Moses, W. J., Remote estimation of chl-a concentration in turbid

780   productive waters—Return to a simple two-band NIR-red model?. Remote Sens. Environ.

781   115(12), 3479-3490. https://doi.org/10.1016/j.rse.2011.08.011,2011.

782   Guo, L., Li, Z., Effects of nitrogen and phosphorus from fish cage-culture on the communities of a

783   shallow lake in middle Yangtze River basin of China. Aquaculture 226(1-4), 201-212.

784   https://doi.org/10.1016/S0044-8486(03)00478-2,2003.

785   Guo, M., Li, X., Song, C., et al., Photo-induced phosphate release during sediment resuspension in

786   shallow lakes: A potential positive feedback mechanism of eutrophication. Environ Pollut. 258,

787   https://doi.org/113679. 10.1016/j.envpol.2019.113679

788   Huang, C., Wang, X., Yang, H., et al., Satellite data regarding the eutrophication response to human

789 activities in the plateau lake Dianchi in China from 1974 to 2009. Sci. Total Environ. 485, 1-11.

790 https://doi.org/10.1016/j.scitotenv.2014.03.031,2014.

791 Huang, J., Xu, C. C., Ridoutt, B. G., et al., Nitrogen and phosphorus losses and eutrophication

792 potential associated with fertilizer application to cropland in China. J. Clean. Prod. 159,

793 171-179. https://doi.org/10.1016/j.jclepro.2017.05.008,2017.

794 Hu, M., Zhang, Y., Ma, R., et al., Optimized remote sensing estimation of the lake algal biomass by

795 considering the vertically heterogeneous chlorophyll distribution: Study case in Lake Chaohu of

796 China. Sci. Total Environ. 771, 144811. https://doi.org/10.1016/j.scitotenv.2020.144811,2021.

797 ILEC/Lake Biwa Research Institute 1994 1988-1993 survey of the state of the world's lakes Volumes

798 I-IV (International Lake Environment Committee, Otsu and United Nations Environment

799 Programme: Nairobi, Kenya)

800 Jeffrey, S. T., Humphrey, G. F., New spectrophotometric equations for determining chlorophylls a, b,

801 c1 and c2 in higher plants, algae and natural phytoplankton. Biochemie und physiologie der

802 pflanzen 167(2), 191-194. https://doi.org/10.1016/0022-2860(75)85046-0,1975.

803 Jin, X., Hu, X., A comprehensive plan for treating the major polluted regions of Lake Taihu, China.

804 Lakes & Reservoirs: Research & Management 8(3-4), 217-230.

805 https://doi.org/10.1111/j.1440-1770.2003.00220.x,2003.

806 Jin, X., Xu, Q., Huang, C., Current status and future tendency of lake eutrophication in China.

807   Science   in   China   Series   C: Life   Sciences   48(2),

808   948-954.https://doi.org/10.1007/BF03187133,2005.

809   Kloiber, S. M., Brezonik, P. L., Olmanson, L. G., et al., A procedure for regional lake water clarity

810   assessment using Landsat multispectral data. Remote Sens. Environ. 82(1), 38-47.

811   https://doi.org/10.1016/S0034-4257(02)00022-6,2002.

812   Koiv, T., Nõges, T., Laas, A., Phosphorus retention as a function of external loading, hydraulic

813   turnover time, area and relative depth in 54 lakes and reservoirs. Hydrobiologia 660(1),

814   105-115. https://doi.org/10.1007/s10750-010-0411-8,2011.

815   Li, S., Song, K., Wang, S., et al., Quantification of chlorophyll-a in typical lakes across China using

816   Sentinel-2 MSI imagery with machine learning algorithm. Sci. Total Environ. 778, 146271.

817   https://doi.org/10.1016/j.scitotenv.2021.146271,2021.

818   Liu, D., Duan, H., Yu, S., et al., 2019. Human-induced eutrophication dominates the bio-optical

819   compositions of suspended particles in shallow lakes: Implications for remote sensing. Sci.

820   Total Environ. 667, 112-123. https://doi.org/10.1016/j.scitotenv.2019.02.366

821   Lund, J. W., Eutrophication. Nature 214(5088), 557-558.https://doi.org/10.1038/214557a0,1967.

822   Ma, Y. L., Zhang, W. B., Yu, B., et al., Prevalence of allergic bronchopulmonary aspergillosis in

823   Chinese patients with bronchial asthma. Zhonghua jie he he hu xi za zhi= Zhonghua jiehe he

824   huxi zazhi= Chinese journal of tuberculosis and respiratory diseases, 34(12), 909-913.

825   https://doi.org/10.1038/cmi.2011.4,2011.

826   Matthews, M. W., Eutrophication and cyanobacterial blooms in South African inland waters: 10

827   years of MERIS observations. Remote Sens. Environ. 155, 161-177.

828   https://doi.org/10.1016/j.rse.2010.04.013,2014.

829   Matthews, M. W., Odermatt, D., Improved algorithm for routine monitoring of cyanobacteria and

830   eutrophication in inland and near-coastal waters. Remote Sens. Environ. 156, 374-382.

831   https://doi.org/10.1016/j.rse.2014.10.010,2015.

832   Mortsch, L. D., Quinn, F. H.,Climate change scenarios for Great Lakes Basin ecosystem studies.

833   Limnol. Oceanogr. 41(5), 903-911.https://doi.org/10.4319/lo.1996.41.5.0903,1996.

834   Morel, A., Prieur, L., 1977. Analysis of variations in ocean color 1. Limnol. Oceanogr. 22(4),

835   709-722.

836   M üller, B., Lotter, A. F., Sturm, M., et al., Influence of catchment quality and altitude on the water

837   and sediment composition of 68 small lakes in Central Europe. Aquat. Sci. 60(4), 316-337.

838   https://doi.org/10.1007/s000270050044,1998.

839   Mountrakis, G., Im, J., Ogole, C.,Support vector machines in remote sensing: A review. ISPRS-J.

840   Photogramm.   Remote   Sens. 66(3),

841   247-259.https://doi.org/10.1016/j.isprsjprs.2010.11.001,2011.

842   Neil, C., Spyrakos, E., Hunter, P. D., et al., A global approach for chlorophyll-a retrieval across

843   optically complex inland waters based on optical water types. Remote Sens. Environ. 229,

844   159-178. https://doi.org/10.1016/j.rse.2019.04.027,2019.

845   Niemistö, J., Holmroos, H., Pekcan-Hekim, Z., et al., Interactions between sediment resuspension

846   and sediment quality decrease the TN: TP ratio in a shallow lake. Limnol. Oceanogr. 53(6),

847   2407-2415. https://doi.org/10.2307/40058331,2008.

848   OECD (Organization for Economic Cooperation and Development). 1982. Eutrophication of waters:

849   monitoring, assessment and control. Organisation for Economic and Cooperative Development,

850   Paris, France.

851   Oliver, S. K., Collins, S. M., Soranno, P. A., et al.,Unexpected stasis in a changing world: Lake

852   nutrient and chlorophyll trends since 1990. Glob. Change Biol. 23(12), 5455-5467 .

853   https://doi.org/10.1111/gcb.13810,2017.

854   Olmanson, L. G., Bauer, M. E., Brezonik, P. L., A 20-year Landsat water clarity census of

855   Minnesota's 10,000 lakes. Remote Sens. Environ. 112(11), 4086-4097.

856   https://doi.org/10.1111/jawr.12138,2008.

857   Pan, Q., Dias, D., An efficient reliability method combining adaptive support vector machine and

858   Monte Carlo simulation. Struct. Saf. 67,

859   85-95.https://doi.org/10.1016/j.strusafe.2017.04.006,2017.

860   Paerl, H.,Nutrient and other environmental controls of harmful cyanobacterial blooms along the

861   freshwater–marine continuum. In Cyanobacterial harmful algal blooms: State of the science and

862   research needs (pp. 217-237). Springer, New York, NY.

863   https://doi.org/10.1007/978-0-387-75865-7_10,2008.

864   Paerl, H. W., Xu, H., McCarthy, M. J., et al., Controlling harmful cyanobacterial blooms in a

865   hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management

866   strategy. Water Res. 45(5), 1973-1983. https://doi.org/10.1016/j.waters.2010.09.018,2011.

867   Pahlevan, N., Chittimalli, S. K., Balasubramanian, S. V., et al.,Sentinel-2/Landsat-8 product

868   consistency and implications for monitoring aquatic systems. Remote Sens. Environ. 220,

869   19-29.　https://doi.org/10.1016/j.rse.2018.10.027,2019.

870   Pahlevan, N., Smith, B., Schalles, J., et al.,Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI)

871   and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sens.

872   Environ. 240, 111604. https://doi.org10.1016/j.rse.2019.111604,2020.

873   Prieur, L., Sathyendranath, S.,An optical classification of coastal and oceanic waters based on the

874   specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and

875   other particulate materials 1. Limnol. Oceanogr. 26(4), 671-689.

876   https://doi.org/10.4319/lo.1981.26.4.0671,1981.

877   Qin, B., Gao, G., Zhu, G., et al., Lake eutrophication and its ecosystem response. Chinese Sci. Bull.

878   58(9), 961-970. https://doi.org/10.1007/s11434-012-5560-x,2013.

879   Qin, B., Zhou, J., Elser, J. J., et al.,Water depth underpins the relative roles and fates of nitrogen and

880   phosphorus in lakes. Environ. Sci. Technol. 54(6), 3191-3198.

881   https://doi.org/10.1021/acs.est.9b05858,2020.

882   Quayle, W. C., Peck, L. S., Peat, H., et al.,Extreme responses to climate change in Antarctic lakes.

883   (Climate Change). Science 295(5555), 645-646.https://doi.org/10.1126/science.1064074,2002.

884   Reichstein, M., Camps-Valls, G., Stevens, B., et al.,Deep learning and process understanding for

885   data-driven Earth system science. Nature 566(7743), 195-204.

886   https://doi.org/10.1038/s41586-019-0912-1,2019.

887   Rodhe, W., 1969. Crystallization of eutrophication concepts in northern Europe.

888   Sass, G. Z., Creed, I. F., Bayley, S. E.et al.,Understanding variation in trophic status of lakes on the

889   Boreal Plain: A 20 year retrospective using Landsat TM imagery. Remote Sens. Environ.109(2),

890   127-141. https://doi.org/10.1016/j.rse.2006.12.010,2007.

891   Schindler, D. W., Hecky, R. E., Findlay, D. L., et al.,Eutrophication of lakes cannot be controlled by

892   reducing nitrogen input: results of a 37-year whole-ecosystem experiment. Proceedings of the

893   National Academy of Sciences, 105(32), 11254-11258.

894   https://doi.org/10.1109/ICASSP.2002.5745032,2008.

895   Sheela, A. M., Letha, J., Joseph, S., et al.,Trophic state index of a lake system using IRS (P6-LISS

896   III) satellite imagery. Environ. Monit. Assess. 177(1), 575-592.

897   https://doi.org/10.1007/s10661-010-1658-2,2011.

898   Shi, K., Zhang, Y., Song, K., et al., 2019. A semi-analytical approach for remote sensing of trophic

899   state in inland waters: Bio-optical mechanism and application. Remote Sens. Environ. 232,

900   111349.

901   Smith, V. H., Tilman, G. D., Nekola, J. C.,Eutrophication: impacts of excess nutrient inputs on

902   freshwater, marine, and terrestrial ecosystems. Environ. Pollut. 100(1-3), 179-196.

903   https://doi.org/10.1016/j.rse.2019.111349,1999.

904   Smith, V. H., Joye, S. B., Howarth, R. W., 2006. Eutrophication of freshwater and marine

905   ecosystems.   Limnol.   Oceanogr.   51(1part2),   351-355.

906   https://doi.org/10.4319/lo.2006.51.1_part_2.0351,2006.

907   Smith, V. H., Schindler, D. W.,Eutrophication science: where do we go from here?. Trends Ecol.

908   Evol. NLM. 24(4), 201-207. https://doi.org/10.1016/j.tree.2008.11.009,2009.

909   Song, K., Li, L., Tedesco, L. P., et al.,Hyperspectral determination of eutrophication for a water

910   supply source via genetic algorithm–partial least squares (GA–PLS) modeling. Sci. Total

911   Environ. 426, 220-232. https://doi.org/10.1016/j.scitotenv.2012.03.058,2012.

912   Song, K. S., Zang, S. Y., Zhao, Y., et al.,Spatiotemporal characterization of dissolved carbon for

913   inland waters in semi-humid/semi-arid region, China. Hydrol. Earth Syst. Sci. 17(10).

914   https://doi.org/10.5194/hessd-10-6559-2013,2013a.

915   Song, K., Li, L., Tedesco, L. P., Li, S., 2013b. Remote estimation of chlorophyll-a in turbid inland

916   waters: Three-band model versus GA-PLS model. Remote Sens. Environ. 136, 342-357.

917   https://doi.org/10.1016/j.rse.2013.05.017,2013b.

918   Song, K., Liu, G., Wang, Q., et al.,Quantification of lake clarity in China using Landsat OLI imagery

919   data. Remote Sens. Environ. 243, 111800. https://doi.org/10.1016/j.rse.2020.111800,2020.

920   Thiemann, S., Kaufmann, H., 2000. Determination of chlorophyll content and trophic state of lakes

921   using field spectrometer and IRS-1C satellite data in the Mecklenburg Lake District, Germany.

922   Remote Sens. Environ. 73(2), 227-235. https://doi.org/10.1016/S0034-4257(00)00097-3,2000.

923   Toming, K., Kotta, J., Uuemaa, E., Sobek, S., Kutser, T., & Tranvik, L. J. Predicting lake dissolved

924   organic   carbon   at   a   global   scale. Scientific   reports   10(1),   1-8.

925   https://doi.org/10.1038/s41598-020-65010-3,2020.

926   Tong, Y., Zhang, W., Wang, X., et al.,Decline in Chinese lake phosphorus concentration

927   accompanied by shift in sources since 2006. *Nature Geoscience*, *10*(7), 507-511.

928   https://doi.org/10.1038/ngeo2967,2017.

929   Tranvik, L. J., Downing, J. A., Cotner, J. B., et al.,Lakes and reservoirs as regulators of carbon

930   cycling   and   climate.   Limnol.   Oceanogr.   54(6part2),   2298-2314.

931   https://doi.org/10.4319/lo.2009.54.6_part_2.2298,2009.

932   USEPA, A., 2002. National water quality inventory 2000 report. EPA-841-R-02-001.

933   Wetzel, R. G.,Limnology: lake and river ecosystems. gulf professional publishing.

934   https://doi.org/10.1086/380040,2001.

935   Wang, S., Li, J., Zhang, B., et al.,Trophic state assessment of global inland waters using a

936   MODIS-derived Forel-Ule index. Remote Sens. Environ. 217, 444-460.

937   https://doi.org/10.1016/j.rse.2018.08.026,2018.

938   Wang, S. and Hongsheng, D., 1998. Chinese Lake Records, Beijing, Science Publishing (in

939   Chinese).

940   Warren, M. A., Simis, S. G., Martinez-Vicente, et al.,Assessment of atmospheric correction

941   algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote

942   Sens. Environ. 225, 267-289. https://doi.org/10.1016/j.rse.2019.03.018,2019.

943   Wen, Z., Song, K., Liu, G., et al., Quantifying the trophic status of lakes using total light absorption

944   of optically active components. Environ. Pollut. 245, 684-693.

945   https://doi.org/10.1016/j.envpol.2018.11.058,2019.

946   Wen, Z. D., Song, K. S., Zhao, Y., et al., Influence of environmental factors on spectral

947   characteristics of chromophoric dissolved organic matter (CDOM) in Inner Mongolia Plateau,

948   China. Hydrol Earth Syst sc, 20(2), 787-801. https://doi.org/10.5194/hess-20-787-2016

949   Wiley, C.,What motivates employees according to over 40 years of motivation surveys. International

950   journal of manpower. https://doi.org/10.1108/01437729710169373,1997.

951   Wu, G., Xu, Z., Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake.

952   Ecol Modell. 222(6), 1245-1252. https://doi.org/10.1016/j.ecolmodel.2010.12.021,2011.

953   Yao, Y., Wang, P., Wang, C., et al., Assessment of mobilization of labile phosphorus and iron across

954   sediment-water interface in a shallow lake (Hongze) based on in situ high-resolution

955   measurement.   Environ.   Pollut.   219,   873-882.

956   https://doi.org/10.1016/j.envpol.2016.08.054,2016.

957   Yin, H., Li, C., Human impact on floods and flood disasters on the Yangtze River. Geomorphology

958   41(2-3), 105-109. https://doi.org/10.1016/S0169-555X(01)00108-8,2001.

959   Yin, H., Douglas, G. B., Cai, Y., et al., Remediation of internal phosphorus loads with modified clays,

960   influence of fluvial suspended particulate matter and response of the benthic macroinvertebrate

961   community. Sci. Total Environ. 610, 101-110. https://doi.org/10.1016/j.scitotenv.2017.07.243,

962   2018.

963   Zhang, Y., Zhou, Y., Shi, K., et al., Optical properties and composition changes in chromophoric

964   dissolved organic matter along trophic gradients: Implications for monitoring and assessing

965   lake   eutrophication.   Water   Res.   131,   255-263.

966   https://doi.org/10.1016/j.watres.2017.12.051,2018.