



Article

Monitoring Optical Variability in Complex Inland Waters Using Satellite Remote Sensing Data

Yunxia Du ¹, Kaishan Song ^{2,*} and Ge Liu ²

¹ School of Geography and Environmental Sciences, Hainan Normal University, Haikou 571158, China; 920219@hainnu.edu.cn

² Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; liuge@iga.ac.cn

* Correspondence: songkaishan@iga.ac.cn

Abstract: Optical classification for water bodies was carried out based on satellite remote sensing data, which avoided the limitation of having a limited amount of in situ measured spectral data. Unsupervised cluster analysis was performed on 53,815 reflectance spectra extracted at 500-m intervals based on the same season or quasi-same season Landsat 8 SR data using the algorithm of fuzzy c-means. Lakes and reservoirs in the study area were comprehensively identified as three optical types representing different limnological features. The shape and amplitude characteristics of the reflectance spectra for the three optical water types indicated that one corresponds to the clearest water, one corresponds to turbid water, and the other is moderate clear water. The novelty detection technique was further used to label the match-ups of the in situ data set collected during 2006 to 2019 in 12 field surveys based on mathematical rules of the three optical water types. The results confirmed that each optical water type was associated with different bio-optical properties, and the total suspended matter of the clearest, moderate clear and turbid water types were 14.99 mg/L, 41.06 mg/L and 83.81 mg/L, respectively. Overall, the clearest, moderate clear and turbid waters in the study area accounted for 49.3%, 36.7% and 14.0%, respectively. The spatial distribution of optical water types in the study area was seamlessly mapped. Results showed that the bio-optical conditions of the water distributed across the southeast region were roughly homogeneous, but in most of other regions and within some water bodies, they showed a patchy distribution and heterogeneity. This study is useful for monitoring water quality and provides a useful foundation to develop or tuning algorithms to retrieve water quality parameters.

Keywords: Landsat imagery; spectra; optical water types; inland water



Citation: Du, Y.; Song, K.; Liu, G. Monitoring Optical Variability in Complex Inland Waters Using Satellite Remote Sensing Data. *Remote Sens.* **2022**, *14*, 1910.

<https://doi.org/10.3390/rs14081910>

Academic Editor: Teodosio Lacava

Received: 26 January 2022

Accepted: 11 April 2022

Published: 15 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Satellite remote sensing presents a significant opportunity for large spatio-temporal water environment monitoring programs. The retrieval of water quality parameters such as chlorophyll-a (chl-a), total suspended matter (TSM) and colored dissolved organic matter (CDOM) in inland waters using the observations from satellite sensors is an active field of research [1,2]. However, the estimation of key water quality parameters of inland and coastal water bodies has no operational method due to the difficulty of stripping the spectral signatures of different optically active constituents in water [3]. The wide range of concentrations of chl-a, TSM and CDOM and their independent variability make it difficult to develop a general algorithm for retrieval of the constituents in complex lake systems [4–6]. Therefore, previous studies reporting retrieval of inland or coastal water quality parameters across several lakes or estuaries were tailored to water bodies with similar optical properties [7–9]. Refining the optical classification of inland water to improve the performance of water quality retrieval algorithms is an effective way to deepen the research of water quality remote sensing [10–12]. Since

optical water types can represent diverse ranges of water quality parameters such as chl-a, TSM and CDOM content, they are extremely useful for detecting water eutrophic and calibrating or tuning retrieval algorithms [13]. The purpose of researching the optical characteristics is to optimize the band combination and obtain the best spectral feature information, improving the accuracy of the algorithm for water quality parameter estimation.

As early as the 1970s, Morel and Prieur (1977) distinguished the color of water according to their absorption [14]. Thereafter, they firstly qualitatively defined Case 1 and 2 waters [14], depending on the predominance of water constituents such as phytoplankton and other covarying compounds. For bio-optical modeling purposes on global water scales, the Case 1 and 2 scheme is commonly used to classify global water optical types. A large number of studies on the optical properties of open ocean and marine water have been performed [15]. Case 2 water bodies with very complex bio-optical properties were noticed by the water color community with the improvement and refinement of satellites or sensors [4,16]. In order to solve the problem of estimating the water quality parameters of complex waters, researchers have further introduced the idea of optical classification of water bodies into the remote sensing of water color of inland Case 2 waters [11,17]. The optical classification of water bodies is a comprehensive evaluation result of the optical characteristics of water bodies, mainly based on inherent optical quantity, remote sensing reflectance (R_{rs}) or water quality parameters [18,19]. The R_{rs} as an apparent optical property (AOP) displays different shapes and magnitudes [13], responding mostly to the backscattering and absorption properties of chl-a, TSM and CDOM [20]. Even though R_{rs} could be modified by the light conditions, it was widely used for identifying water optical types with distinct chl-a, TSM and CDOM content due to its easy accessibility [11,15,21]. Statistical methods for clustering such as k-means, hierarchical clustering and fuzzy clustering are widely used based on the spectral shape or spectral magnitude [10,11]. Additionally, researchers developed algorithms for identifying optical water types based on remote sensing data from different sensors, e.g., Uudeberg et al. (2019) characterized the optical water types of lakes from inland and coastal regions in Europe and developed specific methods for monitoring them based on Sentinel MSI and OLCI imagery [11]; optical water types were identified by Vantrepotte et al. (2012) for contrasted turbid coastal areas of the eastern English Channel, southern North Sea and French Guiana using a clustering approach performed on the spectrally normalized reflectance spectra [22].

Since the launch of Landsat, remote sensing data from multispectral sensors have been widely used in remote sensing monitoring of inland and coastal water quality. Wang et al. (2004) researched the relationship between Landsat TM and total organic carbon (TOC), chemical oxygen demand (COD) and biochemical oxygen demand (BOD) to monitor water quality change in Shenzhen Reservoir, China in 1988 and 1996 [23]. Gao et al. (2020) evaluated the Crosta method for the retrieval of water quality parameters from remote sensing data in the Pearl River estuary based on Landsat TM data. They argued that the Crosta method is good for retrieval of TSM from Landsat data [24]. Landsat satellite imagery was also used to investigate long-term trends in phytoplankton blooms [25] and to show change of lake water clarity [1]. The accuracy of the atmospheric correction method may be one of the factors influencing the accuracy of remote sensing retrieval of water quality parameters. Many previous studies showed that performance of the atmospheric corrections used to generate Landsat surface reflectance products is consistent with that of the water-specific atmospheric correction algorithms [26].

A growing number of studies have assessed optical water types for several regions, but none of them have comprehensively addressed inland waters bodies in Northeast China. Optical water types have been researched for some water bodies in China [27,28], but they lack research on suitable optical types of dissolved organic substances rich waters. The water bodies in Northeast China are affected by continental territory and diversified

environments resulting in aquatic systems with significant variability of TSM, chl-a and CDOM content. Most waters in the Songnen Plain of Northeast China are turbid with high TSM concentration [29]. Some big reservoirs and lakes across this region are oligotrophic, water transparency is high and TSM, chl-a, and CDOM concentrations are low [30,31]. The extensive range of environments in Northeast China has led to the occurrence of different optical water types. Therefore, a study that focuses on water bodies in Northeast China could improve the present characterization of optical water types in Northeast China. Due to the limited in situ spectral data, the optical classification of waters based on in situ spectral cannot be performed continuously.

The main aim of this study was (1) to identify optical water types using satellite remote sensing data and define them for the optically complex inland water in Northeast China; (2) to analyze the spectral characteristics and the bio-optical characteristics of the optical types identified; (3) to analyze the spatial pattern of the optical variation of the water in the study area. This research could be helpful to understand the variation characteristics of water optical properties and for water quality management. This research is also important for detecting eutrophic and improving the accuracy of water quality parameters retrieval.

2. Materials and Methods

2.1. Study Area

The study was carried out in water bodies across northeast of China (118°E–135°E, 38°N–55°N) (Figure 1). China is divided into five main lake regions by region, namely, Eastern Lake Region (ELR), Northeast Lake Region (NLR), Mengxin Lake Region (MXR), Qinghai-Tibet Plateau Lake Region (TQR), and Yunnan-Guizhou Lake Region (YGR) [32]. The selected study areas include the NLR region and the eastern part of the MXR region. The study area belongs to the temperate humid, semi-humid monsoon climate zone, surrounded by high mountains, and the Songnen Plain and Sanjiang Plain water systems are distributed in the middle region. Numerous lakes distribute across the northeast of China. The development and utilization of lakes in the study area are mainly irrigation and aquaculture, and some have the function of drinking water source. This study considers lakes and reservoirs, excluding rivers.

2.2. In Situ Data Collection and Processing

An in situ hyperspectral R_{rs} dataset from several widely-distributed lake and reservoirs across the study area were used for the analysis of satellite data. A dataset of the in situ TSM (mg/L) was used as the main bio-optical parameter in this research. Radiometric measurements ($n = 344$) and water samples ($n = 790$) were collected from 2006 to 2019 in 12 field surveys encompassing the waters with different optical and limnological conditions in the study area (Figure 1).

Water samples for the dataset of limnological parameter were taken from the water surface (below 0.1 m) and stored in 1 L amber high-density polyethylene bottles and then transported back to the laboratory at cold conditions for processing. For TSM measurements, samples were filtered using the pre-ashed and pre-weighed Whatman GF/F filters, and the dried filters were gravimetrically measured before and after filtration [33]. These samples covered large variations of TSM concentrations, representing optically different types of waters and geographical background information. The in situ radiometric measurements were measured using a portable ASD FieldSpec spectrometer (ASD, Inc., Boulder, CO). The details of in situ spectral measurements and instrumentation could be found in [34]. The method of [31] was used for computing the R_{rs} . The R_{rs} of each wavelength (from 400 to 900 nm) was acquired and derived. Using the methods of [35], the R_{rs} shape was further emphasized by normalizing the R_{rs} using its integer.

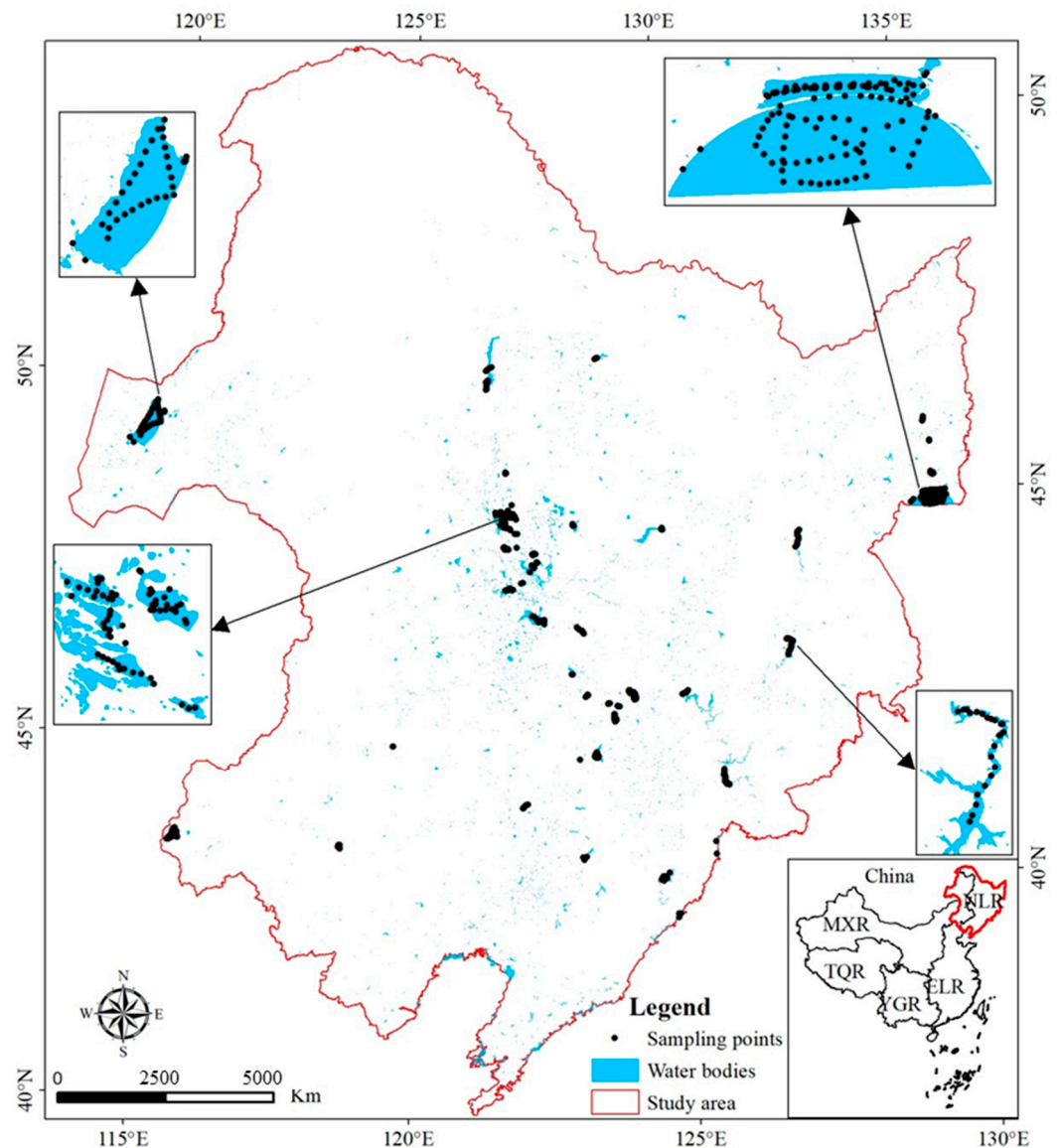


Figure 1. The study area and distribution of the sampling points.

2.3. Satellite Data and Processing

The surface reflectance (SR) data from the Landsat 8 OLI sensor was used in this study. The Landsat SR products generated by LEDAPS or LaSRC have been widely used in remote sensing monitoring of water quality [33,36–38]. It has been demonstrated that the performance of the atmospheric corrections methods used to generate these surface reflectance products is consistent with the water-specific atmospheric correction algorithms [39]. The images from August to October of 2019 covering the water bodies of the study area were acquired for optical classification of water bodies, and there is no cloud covering the water body in these images. In the absence of images, the time limit is extended from May to October of 2019 or 2018 or 2020. Then, water bodies were extracted by masking out the pixels with an average of NDWI and MNDWI less than 0.15 based on the normalized difference water index (NDWI) and modified NDWI (MNDWI).

Due to the large number and wide distribution of lakes and reservoirs in the study area, it is a great challenge to comprehensively analyze the spectra of water bodies through in situ hyperspectral measurement. Based on the continuous spectral sampling from Landsat OLI at equal intervals, this research comprehensively conducted an analysis for the spectra of lakes and reservoirs in the study area. The 500 m × 500 m grid

center points covering the lakes and reservoirs in the study area were generated using the ArcGIS 10.3 software and were used as the spectra sampling points set. A total of 54,635 points were generated (Figure 2). The R_{rs} spectra of these spectra sampling points were extracted based on the Landsat 8 SR data (B1–B5) obtained and used as spectra clustering input.

In addition, some Landsat imagery overpasses during 2006 to 2019 were selected to evaluate Landsat data and analyze limnological characteristics. For the analysis of match-ups, cloud flags and values of 1×1 pixels in OLI images were used to construct two datasets of match-ups (dataset of in situ R_{rs} -OLI and dataset of in situ TSM-OLI). In situ R_{rs} and Landsat satellite overpasses were acquired on the same day. In situ TSM and Landsat satellite overpasses were acquired on the same day or within 5 days. In total, 90 in situ R_{rs} -OLI match-ups and 693 in situ TSM-OLI match-ups were obtained, respectively.

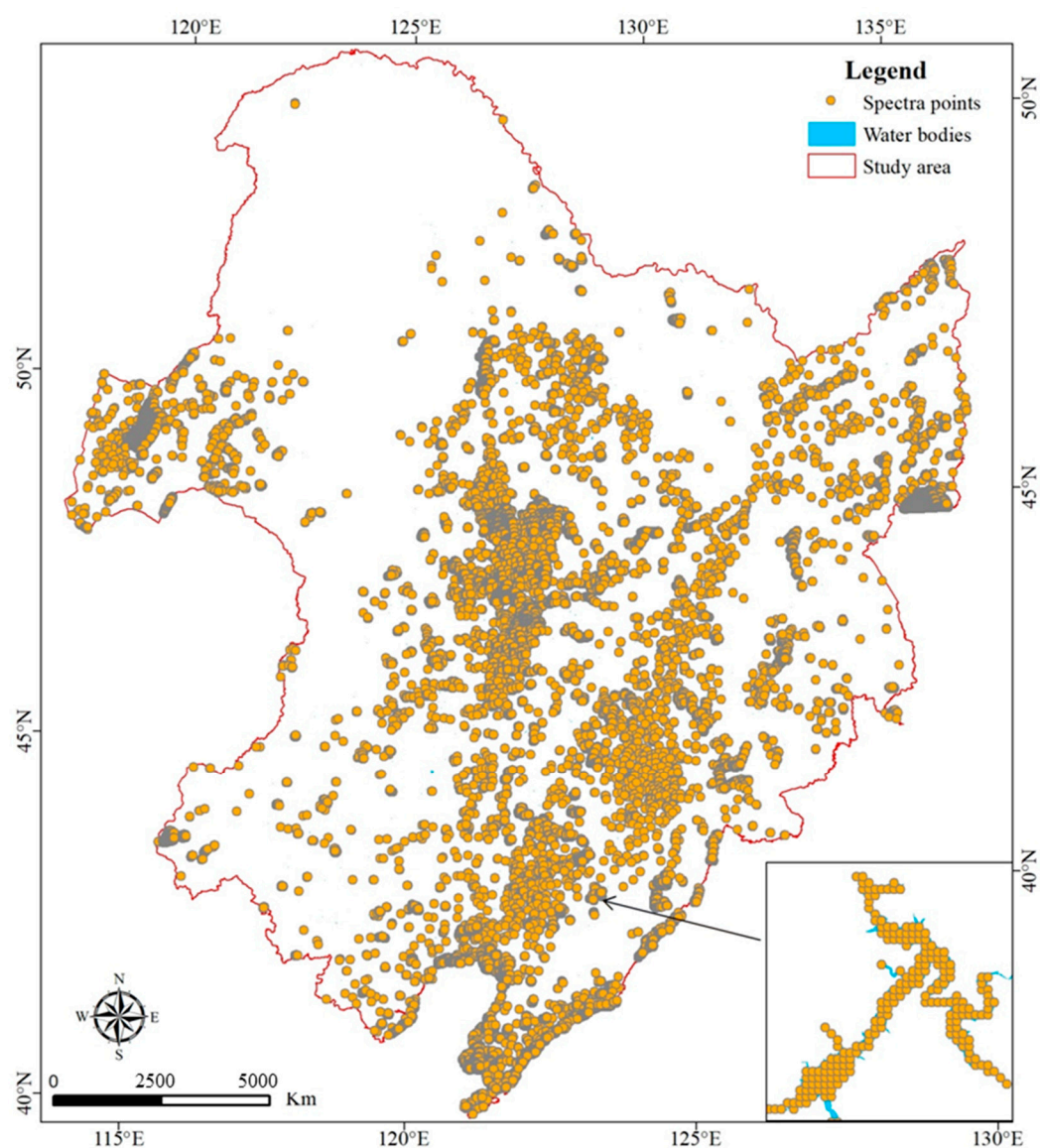


Figure 2. Distribution of spectra sampling points in the study area generated by $500 \text{ m} \times 500 \text{ m}$ grids.

2.4. Assessment of Satellite Bands Using in Situ R_{rs}

To validate the implementation capacity of the atmospherically corrected Landsat 8 OLI reflectance in the optical classification of water bodies, this study assessed the Landsat 8 OLI bands based on the 90 in situ R_{rs} matchups. The hyperspectral R_{rs} spectra were resampled to 5 bands of Landsat 8 OLI based on the spectral response function which excluded SWIR, Pan, Cirrus and TIRS bands. The bands were simulated from in situ R_{rs} and spectral response function of the Landsat 8 OLI sensor as

$$S_i(\lambda_k) = \frac{\int_{\lambda_i}^{\lambda_j} S(\lambda) R_{rs}(\lambda)}{\int_{\lambda_i}^{\lambda_j} S(\lambda)} \quad (1)$$

where S_i is the remote sensing reflectance convoluted from Landsat 8 OLI spectral bands; λ_j and λ_i are the upper and lower limit of the band λ_k , respectively; and S_λ is the spectral response function of the i th band of OLI.

The error and differences of the shape and magnitude between simulated spectra and OLI spectra were used to assess the Landsat data. Considering the distributions of in situ simulated (S_i) and satellite (R_i) bands values, respectively, the mean absolute *bias* and mean *bias*, root mean square error (RMSE) between them were adopted and could be respectively expressed as

$$|bias|_i = \frac{1}{N} \sum_{j=1}^N |S_i - R_i| \quad (2)$$

$$bias_i = \frac{1}{N} \sum_{j=1}^N (S_i - R_i) \quad (3)$$

$$RMSE_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (S_i - R_i)^2} \quad (4)$$

where the i is the i th band; the j is the j th sample; N is the number of samples.

2.5. Clustering Algorithms and Establishment of Optical Water Types

Unsupervised classification technique was performed based on the whole spectra data set from Landsat 8 OLI spectra, in order to split the spectra data into optically homogeneous groups. The fuzzy c-means (FCM) is one of the most popular techniques among the unsupervised classification techniques [10,40]. The FCM algorithm was selected in this study because it can solve the overlap problem of an object might belong to two or more clusters. For the FCM algorithm, the membership degrees of one object to different clusters usually are quantitated between 0 and 1, and the sum of them is equal to 1 is as the main constraint [41]. The membership matrixes and cluster centroids can be obtained for a specific data set after the minimization of the objective function, which is defined as follows:

$$J_\varnothing(U, V) = \sum_{j=1}^N \sum_{i=1}^K u_{ij}^\varnothing \|x_j - v_i\|_A^2 \quad (5)$$

with

$$\sum_{i=1}^k u_{ij} = 1 \text{ and } 0 \leq \sum_{j=1}^N u_{ij} \leq N \quad (6)$$

where N represents the total number of objects in the spectra data set; K is the cluster number; U is the membership matrix of the j th object x_j to each cluster i ; $\left\| |x_j - V_i| \right\|_A^2$ is the square of the feature distance between x_j and cluster centroids v_i ; the Euclidean distance was used as the matrix norm A , which was the rule of similarity measurement; the weighting exponent ϕ ($1 < \phi < \infty$) is the fuzzifier used to control the fuzziness of FCM together with the matrix norm A .

This clustering framework needs to define the cluster number (K), which depends on the specific data set. Since the FCM algorithm is an iterative update process, it is important to wisely choose the starting centers at each iteration. In our research, the starting centers of FCM were selected randomly in the clustering process, assuming the cluster K was set from 2 to 10 and clustering was repeated 500 times by randomly bootstrapping over 49,100 subsamples (about 90% of the whole data) to avoid inappropriate weight to individual spectra. According to the research of Bi et al. (2019), the weighting exponent ϕ was set to 1.3. The convergence tolerance of FCM was set to 0.001. The algorithm was implemented based on the fuzzy-c-means software FuzME3.5 in this study.

The optimal number of cluster K was determined by calculating and comparing the metrics for assessing clustering performance including fuzzy performance index (FPI), improved classification entropy (MPE) and cluster independence index (S). For each unsupervised clustering process of FCM, all these metrics values were computed. The optimal K value was determined based on the lower values of all the metrics.

The cluster definition obtained from spectra clustering was preserved as optical water type for labeling the satellite R_{rs} . For the definition of optical water types, the maximum wavelength, slopes, and amplitude of R_{rs} were used as key features, and a mathematical rule set was created in order to associate the Satellite R_{rs} spectra with the optical water types.

2.6. Satellite R_{rs} Labeling

For analysis of limnological features, each of Landsat OLI reflectance spectra from the 693 match-ups was labeled with an optical type. The novelty detection technique [42] was used to associate each Landsat reflectance spectrum with different optical types defined. The threshold of Mahalanobis distance was determined following the calculation result of the Mahalanobis distance for cluster training set. For a given pixel of match-ups, the class-matching procedure test was alternatively performed for each of the types. When the Mahalanobis distance value was lower than the threshold, the spectrum statistically belongs to corresponding type, otherwise it was considered to be very different from the R_{rs} distribution of the defined type.

3. Results

3.1. Analysis of in Situ R_{rs}

A wide variability was found in the in situ R_{rs} spectra collected during the 12 field surveys from 2006 to 2019 both in magnitude and shape (Figure 3a,b). All the in situ R_{rs} spectra monotonely increased from 400 nm to approximately 560 nm where some started decreasing down to 700 nm and others started increasing up to 700 nm, with however a valley near 680 nm corresponding to the chlorophyll strong absorption (Figure 3a,b). All the in situ R_{rs} spectra showed a decreased tendency in the wavelengths longer than 700 nm and drop rapidly with increasing pure sea water absorption $a_w(\lambda)$ [43] approximately between 700 nm and 750 nm. In approximately between the 400 and 560 nm wavelengths domain, the variability of spectra was relatively smaller, and it was minimal between the 700 nm and 900 nm wavelengths domain (Figure 3a,b). The greatest spectral variation both in the shape and amplitude was observed between approximately 560 nm and 700 nm (Figure 3a,b), where the standard deviation spectrum severely exceeds the mean spectrum (Figure 3b). The remote sensing reflectance was lower at the wavelength range from 400 nm to 500 nm (Figure 3a,b), which was caused by the strong absorption of chlorophyll, carotenoids and yellow substances [22]. The reflection peak observed near 560 nm suggested an

important role of backscattering by particles [35]. The impact of the maxima chlorophyll absorptions and phycocyanin absorptions might also slightly be noticed around both 440 and 620 nm of some R_{rs} spectra (Figure 3a,b). Near the wavelength 700 nm, a secondary peak corresponding to the chlorophyll a fluorescence was showed. The peak of the in situ spectra was shown near 810 nm in the near-infrared region, but the peak of some spectra was very obvious and of some others was weak (Figure 3a,b). These in situ spectra were gathered from different sampling sites and collected during different years (Section 2.2), indicating that the samples were representative.

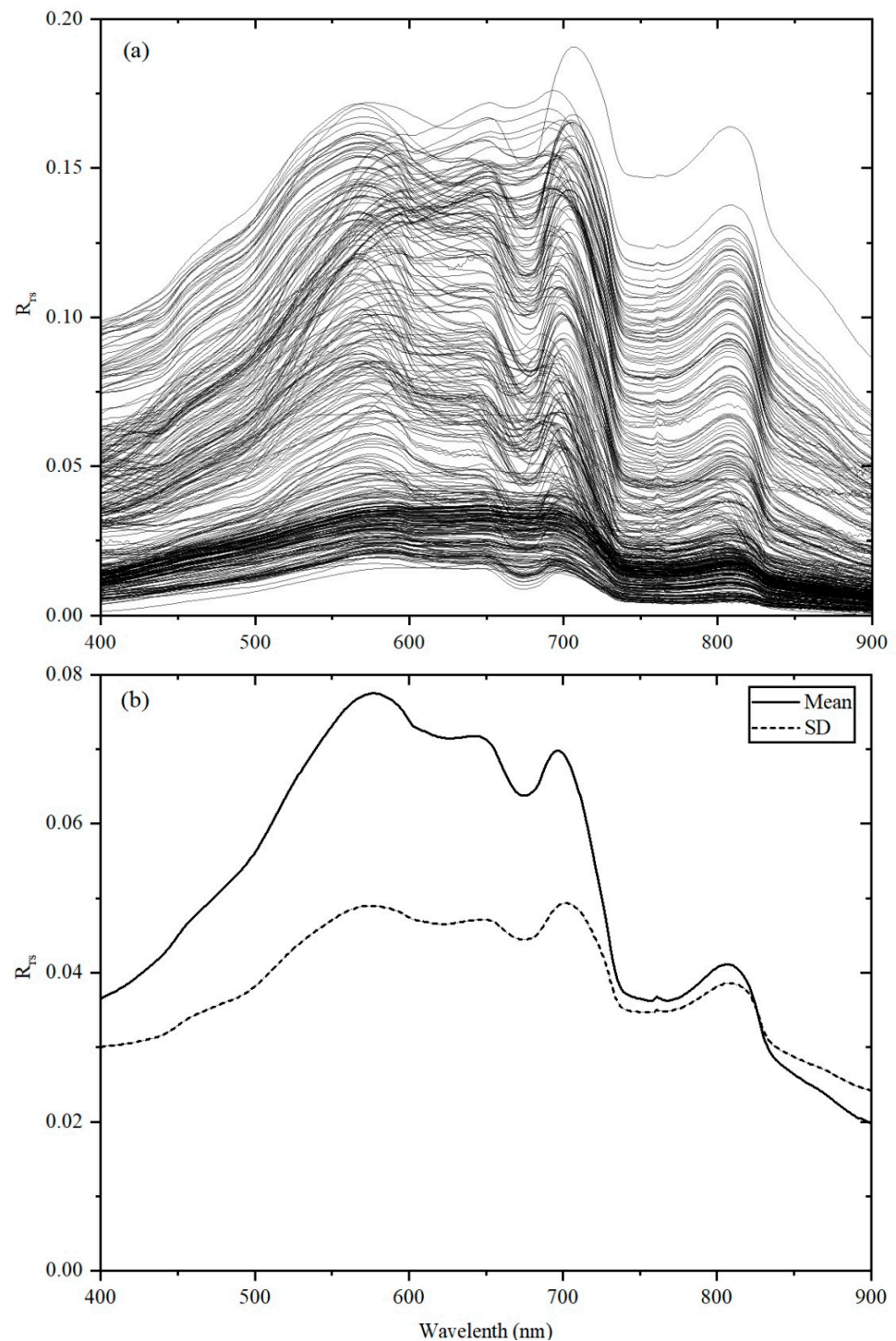


Figure 3. (a) In situ spectra. (b) Mean and standard deviation (SD) of the in situ spectra.

3.2. Comparing Landsat 8 OLI and In Situ R_{rs}

Landsat 8 OLI bands were simulated from 90 in situ R_{rs} match-ups and spectral response function, and the results were shown in Figure 4a. The comparison of simulated spectra derived by in situ spectra with the Landsat 8 OLI spectra was illustrated by Figure 4a,b, and statistics were given in Table 1 for the OLI bands Coastal, Blue, Green, Red and NIR. Most of the simulated spectra curves (Figure 4a) were very similar in shape and size to the Landsat 8 OLI surface reflectance curves (Figure 4b). In general, the reflectance values at the Coastal band (B1) and Blue band (B2) were lower, while at the Green band (B3) and Red band (B4) were higher and at NIR (B5) dropped sharply (Figure 4a,b).

The mean *bias* at bands ranged from -0.001 to -0.005 $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$. The mean absolute *bias* was equal to 0.005, 0.006, 0.006, 0.006 and 0.006 $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ at bands Coastal, Blue, Green, Red and NIR, respectively (Table 1). The mean absolute *bias* and mean *bias* at each band were both lower than its reflectance value by an order of magnitude. The mean *bias* at the Green band (-0.001 $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) was lower than at other bands, and the mean absolute *bias* at the Coastal band (0.005 $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) was lower than that at others (Table 1). The RMSE decreased with wavelength from 0.009 to 0.007 $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ (Table 1).

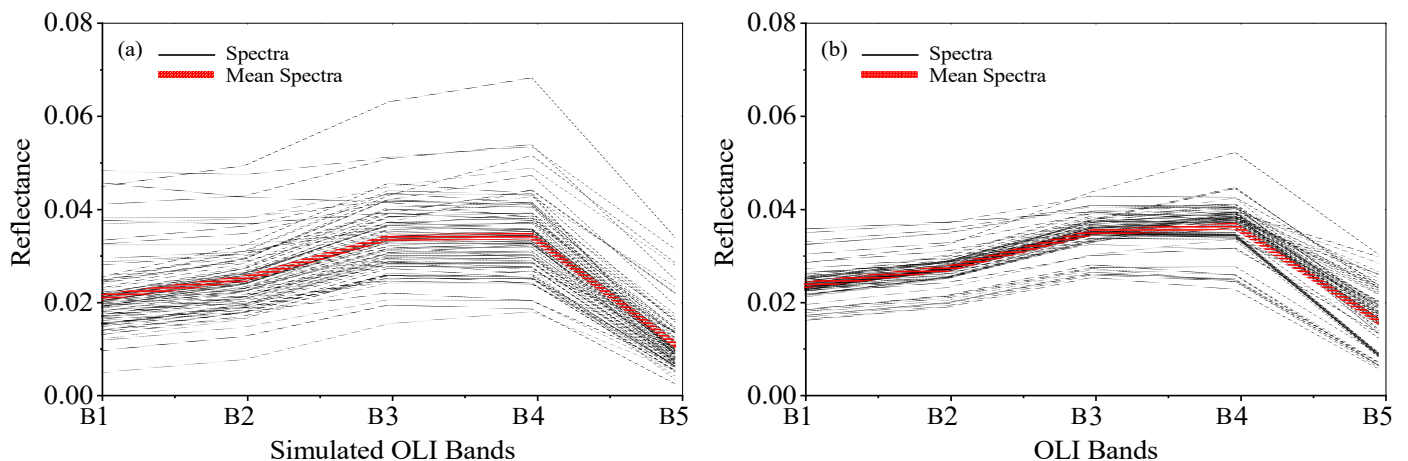


Figure 4. Comparison of surface reflectance curves between Landsat and simulated Landsat spectra: (a) for simulated Landsat bands, (b) for Landsat bands.

Table 1. Statistics of validation of the Landsat OLI R_{rs} .

Bands	<i>bias</i>	<i>bias</i>	RMSE
B1	-0.003	0.005	0.009
B2	-0.002	0.006	0.008
B3	-0.001	0.006	0.008
B4	-0.002	0.006	0.008
B5	-0.005	0.006	0.007

3.3. Clusters and Assessment for Clustering

All spectral data from satellites were subjected to unsupervised clustering using the FCM algorithm with $m = 1.3$. The values of metrics FPI, MPE and S were computed in each unsupervised clustering process of FCM (the cluster $K = 2, 3, 4, \dots, 10$), respectively (Figure 5). When $K = 2, 5, 6, 7, 8, 9$ or 10 , FPI and MPE and S were higher than $K = 3$ or 4 , which was inconsistent with the evaluation criteria of the metric (the lower the value, the better the clustering effect [44]). When $K = 3$, the values of FPI were lowest (0.0685). The values of MPE and S were lowest (0.0708 and 0.1907, respectively) when $K = 4$. Overall, their average was the lowest when $K = 4$ (Figure 5), so the FCM could obtain appropriate membership distributions in which

the majority of the spectra are strongly associated with one given cluster. Four optimized clusters were obtained by utilizing FCM with the FPI, MPE and S was 0.0751, 0.0708 and 0.1907, respectively, showing a well-performed clustering structure. Among the 54635 spectra participating in the clustering, 48.5%, 36.1%, 1.5% and 13.8% of them belong to clusters 'a', 'b', 'c' and 'd' (Figure 6). The centroids of cluster 'c' were obviously far away from the other three clusters in the distribution on the first two principal component (PC1 and PC2) (Figure 6). The variation in the distribution of cluster centroids reflected the difference in the shape and amplitude of the spectra among these clusters.

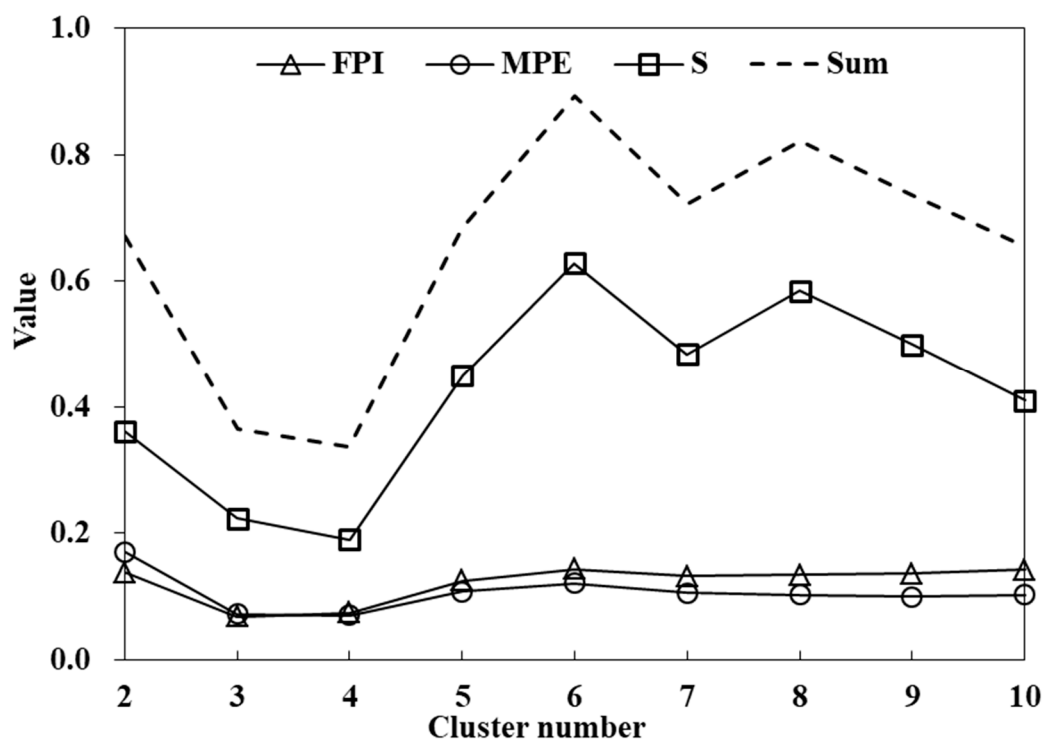


Figure 5. The metrics of spectra clustering derived from Landsat 8 OLI data.

3.4. Optical Water Types and Limnological Features

The centroid spectra curves of the above four clusters generated by clustering were shown in Figure 7. Among them, the cluster 'c' was obviously not the signals of the water body (Figure 7), and it was generated by the mixed pixels and land pixels at the water edge. Except for this non-aqueous cluster, the water spectra of the study area were clustered into three clusters representing three optical water types (clusters 'a', 'b' and 'd' in Figures 6 and 7). For analysis, these three water clusters were re-arranged into Type 1, Type 2 and Type 3, respectively (Figure 8). Compared to the other types, the mean reflectance of Type 2 spectrum displayed the least values at all the bands as well as the reflectance at green band was higher than at red band. The reflectance at all bands of Type 3 was the highest. In view of this, the Type 1 was defined as: the reflectivity at the green band (B3) is higher than at the red band (B4), and the reflectance of the green band is greater than 0.02; the Type 2 was defined as: the reflectance at the green band (B3) is higher than at the red band (B4), and the reflectance of the green band is less than 0.02; and the Type 3 was defined as: the reflectance at the red band (B4) is greater than at the green band (B3).

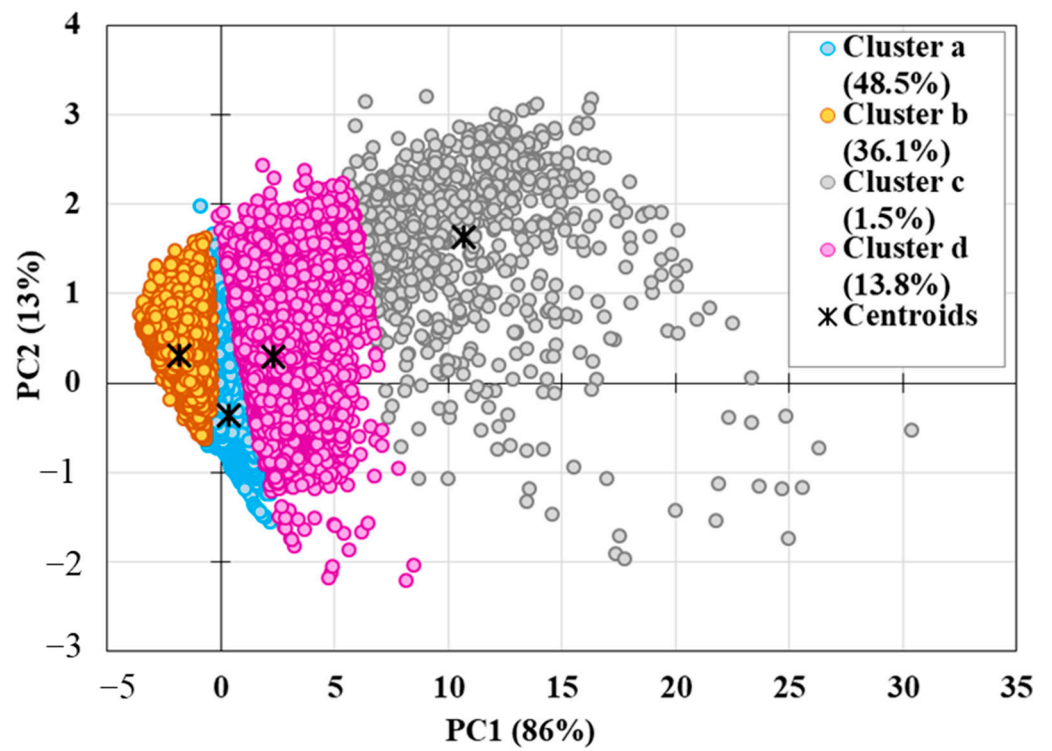


Figure 6. The clustering scatters of the spectra and the values of PC1 and PC2.

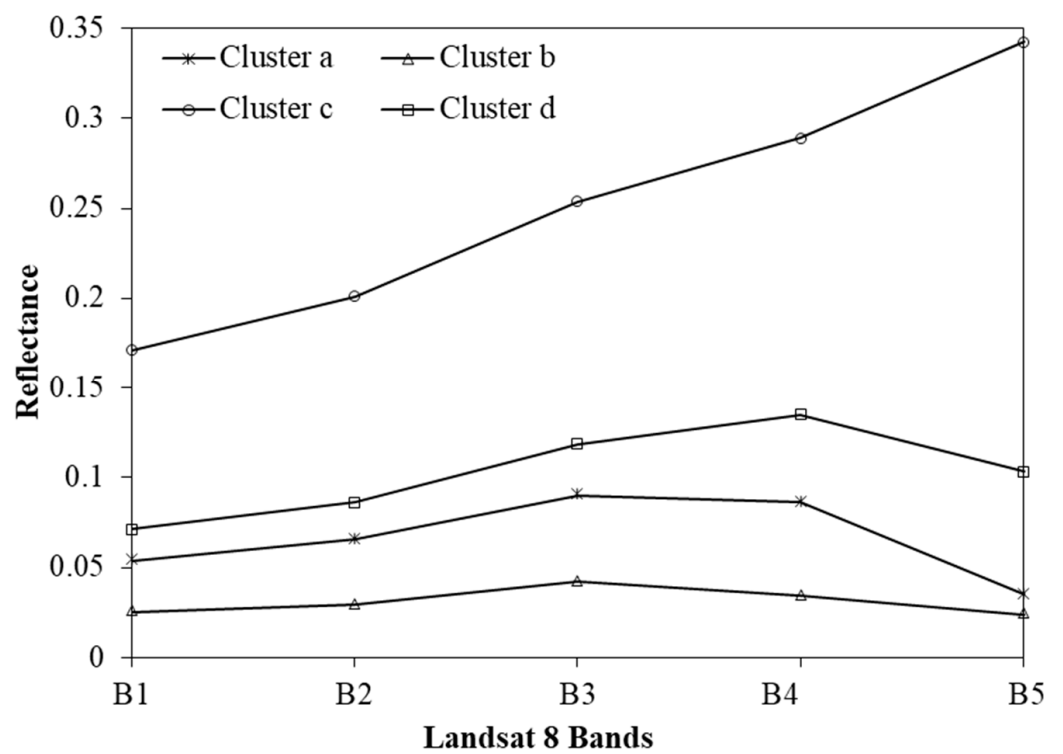


Figure 7. Centroid spectra curves of the clusters.

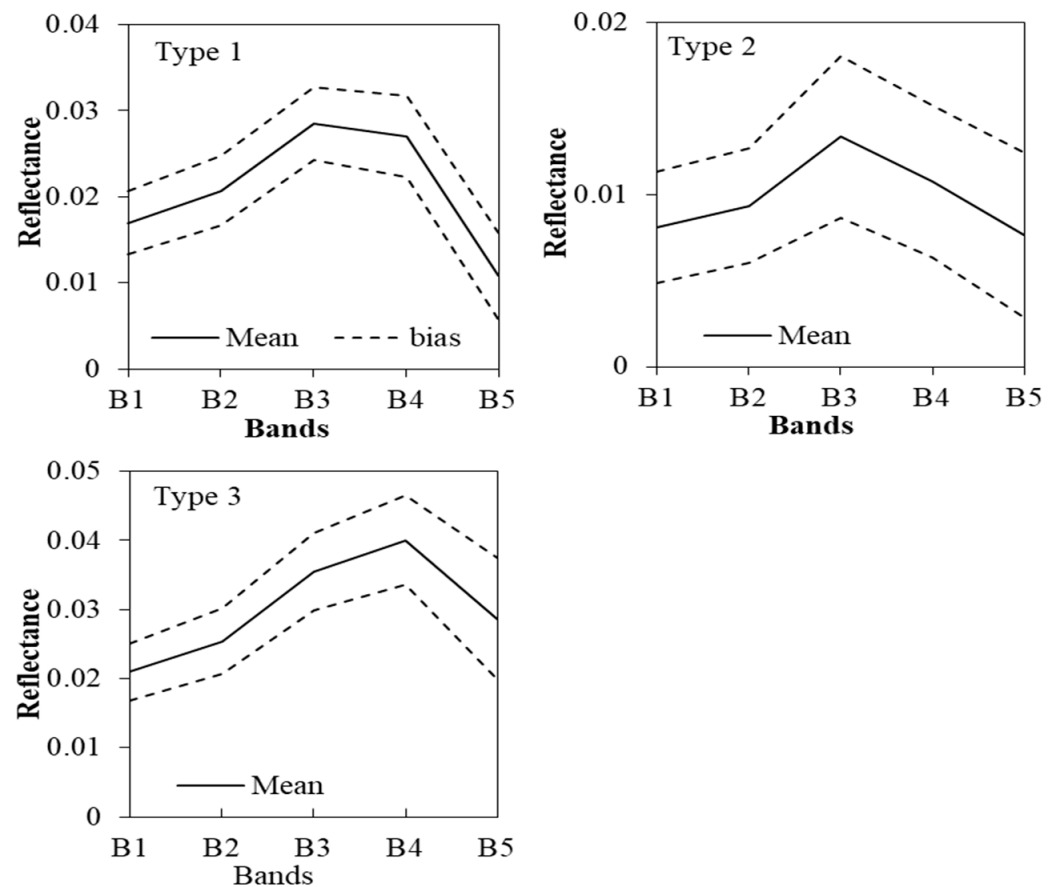


Figure 8. Average spectral curves of three optical water types generated by clustering for the study area (bands were for Landsat 8).

These mathematical rules of the definition of optical water types were applied to label 693 reflectance spectra from OLI matched with the in situ TSM measurements by the novelty detection technique. As a result, all these spectra were detected as one of these three types, and this underlined the significant representativeness of our data set generated in the entire study area, although a year only. Of these reflectance spectra, 279 were labeled as type 1; 212 were labeled as type 2, and 202 were labeled as type 3 (Table 2). The distributions of each type for TSM were also analyzed (Table 2). The variation and aggregation of each type performed obviously for TSM. The type 2 corresponded to the most transparent water with the lowest average TSM of 14.99 mg/L; the turbid type 3 showed the highest TSM value of 83.81 mg/L, and the average TSM of the moderate clear type 1 was 41.06 mg/L (Table 2).

Table 2. Statistics of TSM concentration for different optical water types.

Classification	N	TSM (mg/L)		
		Min	Mean	Max
Type 1	279	1.00	41.06	293.99
Type 2	212	0.83	14.99	126.03
Type 3	202	4.50	83.81	472.19

3.5. Spatial Pattern of Optical WATER Types

The type 1, type 2 and type 3 were assigned to the satellite R_{rs} of 49.3%, 36.7% and 14.0% of the water grids from the study area, respectively, meaning that a small part of the water in the lakes and reservoirs of the study area was turbid. Figure 9 shows the spatial distribution pattern of 3 types of water in the study area. The waters in the southeastern region of the study area corresponded mostly to the optical Type 2 that represented clear water (Figure 9). The central and northeast regions of the study area are the Songnen Plain and the Sanjiang Plain, respectively, and most of the water in these regions was more turbid Type 1 and Type 3 (Figure 9). Most of the water in the northwestern region was also classified as Type 1 or Type 3 (Figure 9). In other regions of the study area, the spatial distribution of the optical water types presented more likely a patchy distribution. For the perspective of a water body, the variation of the optical property within some water bodies was relatively weak, e.g., Lake Lake Hulun (Figure 9), where the water was mostly Type 1. In contrast, there were also some water bodies where the water exhibited different optical water types, e.g., Lake Songhua and Lake Xingkai.

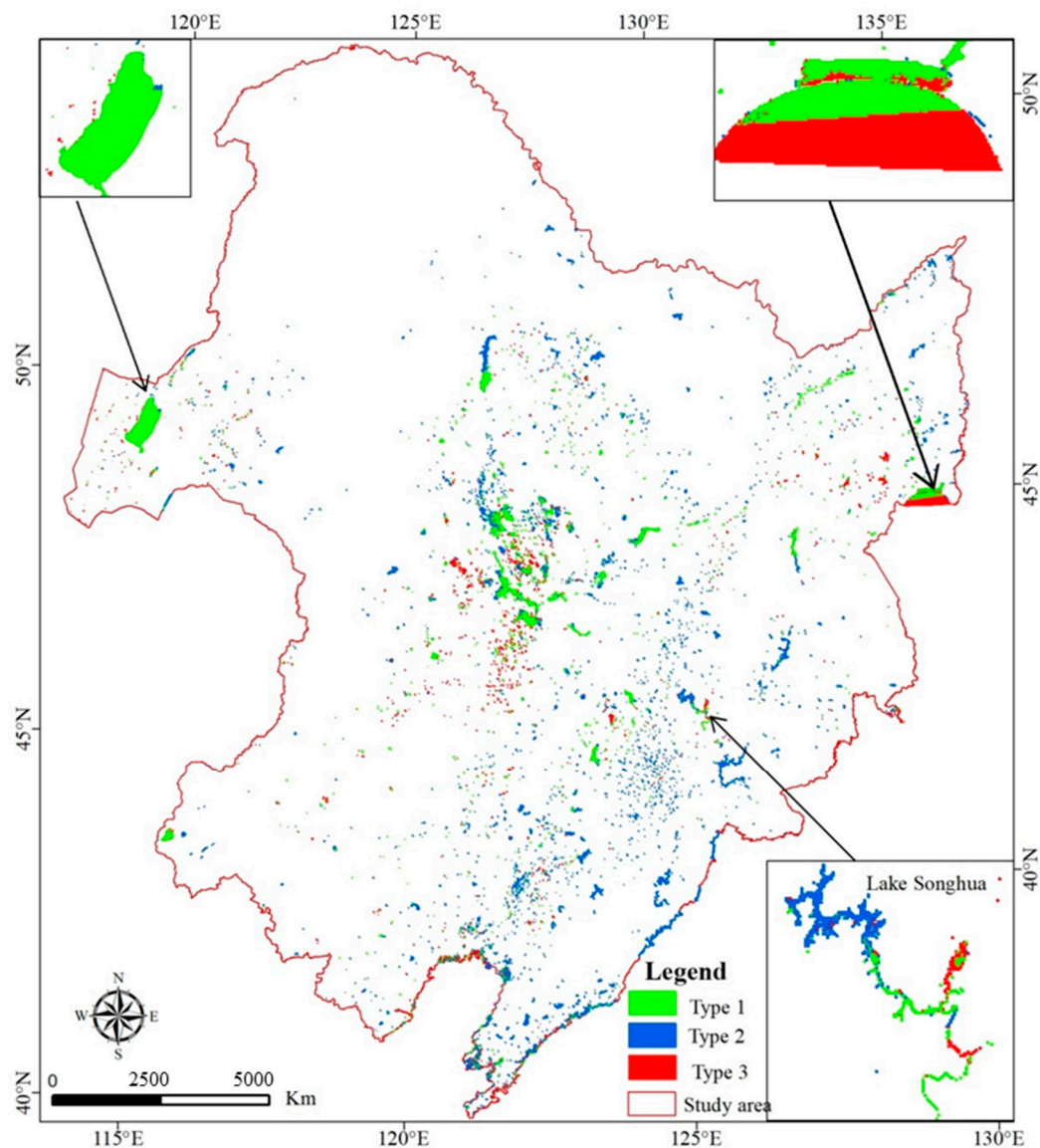


Figure 9. Spatial pattern of optical water types in the study area.

4. Discussion

4.1. Applicability of OLI SR Product in Water Optical Classification

Most of the simulated in situ spectra curves (Figure 4a) were very similar in shape and size to the Landsat 8 OLI surface reflectance curves (Figure 4b). The mean absolute *bias* and mean *bias* at each band were both lower than its reflectance value by an order of magnitude. These statistics at bands Blue, Green and Red agreed with the analysis published by Song et al. (2020). The uncertainty of the Coastal and Blue bands was due to the low values of reflectance at these spectral regions that are prone to produce large relative differences [45]. These differences at all bands might be explained by the optical properties of the measure stations that were not properly represented by the aerosol models available in the process of atmospheric correction [45]. Even so, Landsat 8 OLI SR data were acceptable for optical classification of the inland water in the study area, because a higher mean absolute *bias* value was only due to the translation of the larger negative *bias* or positive *bias* from a few samples' spectra. It was considered that the results of the classification appear to be fairly independent of the documented R_{rs} uncertainties, even if the reduction of the *bias* affecting bands reflectance would further tune the classification [35]. Furthermore, a growing number of studies have demonstrated that the Landsat surface reflectance products can be used to accurately estimate inland or coastal water quality parameters [37–39], which proves the ability of Landsat surface reflectance to reflect the properties of water bodies.

4.2. Optical Water Types from Unsupervised Clustering

Unsupervised clustering methods such as the FCM algorithm aim to identify the most characteristic optical water types by clustering all samples into different types and retain all data in the analysis [27]. All available water spectral samples were clustered into three types in this research (Figure 8). Whether in terms of shape or amplitude, the average spectra of these three optical water types all showed significant distinction (Figure 8). Type 1 represented moderate clear water (average TSM was 41.06 mg/L) with the reflectance relatively higher than Type 2 and lower than Type 3. The Type 2 pattern was directly related to the fact that this class corresponded to water with low optical substance concentrations and the highest water transparency (average TSM was 14.99 mg/L). Type 3 was turbid water with high TSM concentration (average TSM was 83.81 mg/L) due to its reflectance maximum at all bands and its reflectance at the red band being higher than at the green band. The analysis of in situ R_{rs} confirmed that there were spatio-temporal differences in the water compositions and their concentration in the study area, indicating the spatio-temporal variability of their bio-optical feature. Thus, it was proved that each optical water type was associated with a specific bio-optical condition, implying the reliability of optical classification. Furthermore, this optical variation directly indicated the wide range of bio-optical and biogeochemical conditions covered by the current data set [22]. The spatial distribution pattern of optical water types (Figure 9) indicates that unsupervised clustering could effectively monitor the optical variability of water bodies. The water in the southeastern region of the study area is mostly the clearest Type 2 (Figure 9) due to the high vegetation coverage and weak impact from human activities in this region. Most water in the central and northeast regions of the study area is moderately clear and turbid Type 1 and Type 3 (Figure 9), which is related to the intense agricultural activity. The northwestern region might be related to the poor vegetation coverage, so the water in this region is also classified as Type 1 or Type 3 (Figure 9). Other regions of the study area are affected by local natural and human factors, resulting in the patchy distribution of the optical water types (Figure 9). The regional heterogeneity of optical water types could be explained by the differences of regional natural factors and human activities, and the variability within a water body could be attributed to some local factors.

5. Conclusions

A total of 53,815 (excluding 820 non-water spectra) reflectance spectra of water were extracted at 500 m intervals from the same season or quasi-same season Landsat SR data and used for spectral cluster analysis to generate optical water types. The clustering results were assessed to generate the optical water types representing the bio-optical conditions of water bodies in the study area. The optical water types were further defined according to shape and magnitude, and their bio-optical conditions were analyzed by in-situ TSM data collected from different dates and sites. In addition, the spatial distribution of optical water types in the study area was seamlessly mapped. The key findings of this study are summarized below:

- (1) The water environmental conditions of lakes and reservoirs across the study area had variability among regions and within some water bodies, which was manifested in three optical types. Most lakes and reservoirs were moderately clear or very clear, while only a few were turbid.
- (2) The novelty detection technique did not find a new type (different from the defined three types) in the labeling process of the reflectance spectra of the 693 match-ups collected from different dates. Therefore, this additionally showed that these three optical water types provided by contemporaneous or quasi-contemporaneous Landsat 8 data represent not only spatial variability but also temporal variability.
- (3) The total suspended solids content in different optical water types showed obvious differences. Thus, the difference of the total suspended matter content in water could reflect the variation of water optical properties.
- (4) The applicability of Landsat OLI in such optically-based unsupervised clustering approaches for different scale applications was demonstrated by the analysis from the clustering and the bio-optical conditions of the optical water types.

Author Contributions: Conceptualization, Y.D. and K.S.; Data curation, G.L.; Funding acquisition, G.L. and K.S.; Investigation, Y.D.; Methodology, Y.D. and G.L.; Software, G.L.; Supervision, K.S.; Writing—original draft, Y.D.; Writing—review & editing, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly funded by the National Key Research and Development Program of China (Grant Number 2021YFB3901101), the National Natural Science Foundation of China (Grant Number 42171374, 42171385, 41730104) and the Natural Science Foundation of Hainan Province hosted by Yunxia Du.

Acknowledgments: Thanks to the reviewers for their important comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chawla, I.; Karthikeyan, L.; Mishra, A.K. A review of remote sensing applications for water security: Quantity, quality, and extremes. *J. Hydrol.* **2020**, *585*, 124826. [[CrossRef](#)]
2. Palmer, S.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2014**, *157*, 1–8. [[CrossRef](#)]
3. Lehmann, M.K.; Nguyen, U.; Allan, M.; Van der Woerd, H.J. Colour classification of 1486 lakes across a wide range of optical water types. *Remote Sens.* **2018**, *10*, 1273. [[CrossRef](#)]
4. Mouw, C.B.; Greb, S.; Aurin, D.; Digiacomio, P.M.; Lee, Z.; Twardowski, M.; Binding, C.; Hu, C.; Ma, R.; Moore, T. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **2015**, *160*, 15–30. [[CrossRef](#)]
5. Sun, D.; Li, Y.; Qiao, W.; Le, C.; Huang, C.; Shi, K. Development of optical criteria to discriminate various types of highly turbid lake waters. *Hydrobiologia* **2011**, *669*, 83–104. [[CrossRef](#)]
6. Tyler, A.N.; Hunter, P.D.; Spyarakos, E.; Groom, S.; Constantinescu, A.M.; Kitchen, J. Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. *Sci. Total Environ.* **2016**, *572*, 1307–1321. [[CrossRef](#)]
7. Hou, X.; Feng, L.; Duan, H.; Chen, X.; Sun, D.; Shi, K. Fifteen-year monitoring of the turbidity dynamics in large lakes and reservoirs in the middle and lower basin of the Yangtze River, China. *Remote Sens. Environ.* **2017**, *190*, 107–121. [[CrossRef](#)]

8. Li, S.; Song, K.; Wang, S.; Liu, G.; Mu, G. Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. *Sci. Total Environ.* **2021**, *778*, 146271. [[CrossRef](#)]
9. Odermatt, D.; Gitelson, A.; Brando, V.E.; Schaepman, M. Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sens. Environ.* **2012**, *118*, 116–126. [[CrossRef](#)]
10. Bi, S.; Li, Y.; Xu, J.; Liu, G.; Xu, J. Optical classification of inland waters based on an improved Fuzzy C-Means method. *Opt. Express* **2019**, *27*, 34838–34856. [[CrossRef](#)]
11. Uudeberg, K.; Ansko, I.; Põru, G.; Ansper, A.; Reinart, A. Using optical water types to monitor changes in optically complex inland and coastal waters. *Remote Sens.* **2019**, *11*, 2297. [[CrossRef](#)]
12. Xue, K.; Ma, R.; Wang, D.; Shen, M. Optical classification of the remote sensing reflectance and its application in deriving the specific phytoplankton absorption in optically complex lakes. *Remote Sens.* **2019**, *11*, 184. [[CrossRef](#)]
13. da Silva, E.F.F.; Novo, E.M.L.d.; Lobo, F.d.; Barbosa, C.C.F.; Noernberg, M.A.; Rotta, L.H.d.; Cairo, C.T.; Maciel, D.A.; Júnior, F.R. Optical water types found in Brazilian waters. *Limnology* **2021**, *22*, 57–68. [[CrossRef](#)]
14. Morel, A.; Prieur, L. Analysis of variations in ocean color. *Limnol. Oceanogr.* **1977**, *22*, 709–722. [[CrossRef](#)]
15. Spyarakos, E.; O'Donnell, R.; Hunter, P.D.; Miller, C.; Scott, M.; Simis, S.G.H.; Neil, C.; Barbosa, C.C.F.; Binding, C.E.; Bradt, S.; et al. Optical types of inland and coastal waters. *Limnol. Oceanogr.* **2018**, *63*, 846–870. [[CrossRef](#)]
16. Seegers, B.N.; Stumpf, R.P.; Schaeffer, B.A.; Loftin, K.A.; Werdell, P.J. Performance metrics for the assessment of satellite data products: An ocean color case study. *Opt. Express* **2018**, *26*, 7404–7422. [[CrossRef](#)]
17. Sayers, M.J.; Bosse, K.R.; Shuchman, R.A.; Ruberg, S.A.; Fahnenstiel, G.L.; Leshkevich, G.A.; Stuart, D.G.; Johengen, T.H.; Burtner, A.M.; Palladino, D. Spatial and temporal variability of inherent and apparent optical properties in western Lake Erie: Implications for water quality remote sensing. *J. Great Lakes Res.* **2019**, *45*, 490–507. [[CrossRef](#)]
18. Reinart, A.; Herlevi, A.; Arst, H.; Sipelgas, L. Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. *J. Sea Res.* **2003**, *49*, 357–366. [[CrossRef](#)]
19. Wernand, M.R.; Hommersom, A.; Woerd, H.J.V.D. MERIS-based ocean colour classification with the discrete Forel–Ule scale. *Ocean Sci.* **2013**, *9*, 477–487. [[CrossRef](#)]
20. Kirk, J. *Light and Photosynthesis in Aquatic Ecosystems*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2011.
21. Minu, P.; Lotliker, A.A.; Shaju, S.S.; Ashraf, P.M.; Kumar, T.S.; Meenakumari, B. Performance of operational satellite bio-optical algorithms in different water types in the southeastern Arabian Sea. *Oceanologia* **2016**, *58*, 317–326. [[CrossRef](#)]
22. Vantrepotte, V.; Loisel, H.; Dessailly, D.; Mériaux, X. Optical classification of contrasted coastal waters. *Remote Sens. Environ.* **2012**, *123*, 306–323. [[CrossRef](#)]
23. Wang, Y.; Xia, H.; Fu, J.; Sheng, G. Water quality change in reservoirs of Shenzhen, China: Detection using LANDSAT AT/TM data. *Sci. Total Environ.* **2004**, *328*, 195–206. [[CrossRef](#)] [[PubMed](#)]
24. Gao, F.; Wang, Y.; Zhang, Y. Evaluation of the Crosta method for the retrieval of water quality parameters from remote sensing data in the Pearl River estuary. *Water Qual. Res. J.* **2020**, *55*, 209–220. [[CrossRef](#)]
25. Ho, J.C.; Michalak, A.M.; Pahlevan, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* **2019**, *574*, 667–670. [[CrossRef](#)] [[PubMed](#)]
26. Olmanson, L.G.; Page, B.P.; Finlay, J.C.; Brezonik, P.L.; Bauer, M.E.; Griffin, C.G.; Hozalski, R.M. Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable minnesota lakes using landsat 8 imagery. *Sci. Total Environ.* **2020**, *724*, 138141. [[CrossRef](#)] [[PubMed](#)]
27. Shen, Q.; Li, J.; Zhang, F.; Sun, X.; Li, J.; Li, W.; Zhang, B. Classification of several optically complex waters in China using in situ remote sensing reflectance. *Remote Sens.* **2015**, *7*, 14731–14756. [[CrossRef](#)]
28. Shi, K.; Li, Y.; Lin, L.; Lu, H.; Song, K.; Liu, Z.; Xu, Y.; Li, Z. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Sci. Total Environ.* **2013**, *444*, 1–15. [[CrossRef](#)]
29. Du, Y.; Song, K.; Liu, G.; Wen, Z.; Fang, C.; Shang, Y.; Zhao, F.; Wang, Q.; Du, J.; Zhang, B. Quantifying total suspended matter (TSM) in waters using Landsat images during 1984–2018 across the Songnen Plain, Northeast China. *J. Environ. Manag.* **2020**, *262*, 110334. [[CrossRef](#)]
30. Song, K.; Li, L.; Wang, Z.; Liu, D.; Zhang, B.; Xu, J.; Du, J.; Li, L.; Li, S.; Wang, Y. Retrieval of total suspended matter (TSM) and chlorophyll-a (Chl-a) concentration from remote-sensing data for drinking water resources. *Environ. Monit. Assess.* **2012**, *184*, 1449–1470. [[CrossRef](#)]
31. Song, K.S.; Liu, G.; Wang, Q.; Wen, Z.D.; Lyu, L.L.; Du, Y.X.; Sha, L.W.; Fang, C. Quantification of lake clarity in China using Landsat OLI imagery data. *Remote Sens. Environ.* **2020**, *243*, 111800. [[CrossRef](#)]
32. Wang, S.; Dou, H. *Chinese Lake Catalogue*; Science Press: Beijing, China, 1998.
33. Du, Y.; Song, K.; Wang, Q.; Li, S.; Wen, Z.; Liu, G.; Tao, H.; Shang, Y.; Hou, J.; Lyu, L.; et al. Total suspended solids characterization and management implications for lakes in East China. *Sci. Total Environ.* **2021**, *806*, 151374. [[CrossRef](#)] [[PubMed](#)]
34. Song, K.; Lin, L.; Tedesco, L.P.; Shuai, L.; Clercin, N.A.; Hall, B.E.; Li, Z.; Shi, K. Hyperspectral determination of eutrophication for a water supply source via genetic algorithm–partial least squares (GA–PLS) modeling. *Sci. Total Environ.* **2012**, *426*, 220–232. [[CrossRef](#)] [[PubMed](#)]
35. Lubac, B.; Loisel, H. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sens. Environ.* **2007**, *110*, 45–58. [[CrossRef](#)]

36. Gardner, J.R.; Yang, X.; Topp, S.N.; Ross, M.R.V.; Altenau, E.H.; Pavelsky, T.M. The color of rivers. *Geophys. Res. Lett.* **2021**, *48*, e2020GL088946. [[CrossRef](#)]
37. Kuhn, C.; Valerio, A.D.; Ward, N.; Loken, L.; Sawakuchi, H.O.; Karnpel, M.; Richey, J.; Stadler, P.; Crawford, J.; Striegl, R.; et al. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sens. Environ.* **2019**, *224*, 104–118. [[CrossRef](#)]
38. Zhang, Y.; Zhang, Y.; Shi, K.; Zhou, Y.; Li, N. Remote sensing estimation of water clarity for various lakes in China. *Water Res.* **2021**, *192*, 116844. [[CrossRef](#)]
39. Topp, S.N.; Pavelsky, T.M.; Stanley, E.H.; Yang, X.; Griffin, C.G.; Ross, M.R.V. Multi-decadal 1 improvement in U.S. Lake water clarity. *Environ. Res. Lett.* **2021**, *16*, 055025. [[CrossRef](#)]
40. Feng, H.; Campbell, J.W.; Dowell, M.D.; Moore, T.S. Modeling spectral reflectance of optically complex waters using bio-optical measurements from Tokyo Bay. *Remote Sens. Environ.* **2005**, *99*, 232–243. [[CrossRef](#)]
41. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum Press: New York, NY, USA, 1981.
42. Bishop, C.M. Novelty detection and neural network validation. *Image Signal Process. IEE Proc. Vis.* **1994**, *141*, 217–222. [[CrossRef](#)]
43. Pope, R.M.; Fry, E.S. Absorption spectrum (380–700 nm) of pure water, 2, integrating cavity measurements. *Appl. Opt.* **1997**, *36*, 8710–8723. [[CrossRef](#)]
44. Guo, Y.; Shi, Z.; Li, H.Y.; Triantafilis, J. Application of digital soil mapping methods for identifying salinity management classes based on a study on coastal central China. *Soil Use Manag.* **2013**, *29*, 445–456. [[CrossRef](#)]
45. Mélin, F.; Vantrepotte, V.; Clerici, M.; D’Alimonte, D.; Zibordi, G.; Berthon, J.-F.; Canuti, E. Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. *Prog. Oceanogr.* **2011**, *91*, 229–244. [[CrossRef](#)]