Article

Water Multi-Parameter Sampling Design Method Based on Adaptive Sample Points Fusion in Weighted Space

Mingjian Zhai 1,2, Zui Tao 1,*, Xiang Zhou 1, Tingting Lv 1, Jin Wang 1 and Ruoxi Li 1,2

1 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; zhaimingjian20@mails.ucas.ac.cn (M.Z.); zhouxiang@radi.ac.cn (X.Z.); lvtt@aircas.ac.cn (T.L.); wangling@aircas.ac.cn (J.W.); liruoxi19@mails.ucas.ac.cn (R.L.)
2 University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: taozui@radi.ac.cn; Tel.: +86-150-1132-8757

Abstract: The spatial representativeness of the in-situ data is an important prerequisite for ensuring the reliability and accuracy of remote sensing product retrieval and verification. Limited by the collection cost and time window, it is essential to simultaneously collect multiple water parameter data in water tests. In the shipboard measurements, sampling design faces problems, such as heterogeneity of water quality multi-parameter spatial distribution and variability of sampling plan under multiple constraints. Aiming at these problems, a water multi-parameter sampling design method is proposed. This method constructs a regional multi-parameter weighted space based on the single-parameter sampling design and performs adaptive weighted fusion according to the spatial variation trend of each water parameter within it to obtain multi-parameter optimal sampling points. The in-situ datasets of three water parameters (chlorophyll a, total suspended matter, and Secchi-disk Depth) were used to test the spatial representativeness of the sampling method. The results showed that the sampling method could give the sampling points an excellent spatial representation in each water parameter. This method can provide a fast and efficient sampling design for in-situ data for water parameters, thereby reducing the uncertainty of inversion and the validation of water remote sensing products.

Keywords: sampling design; water; multi-parameter; remote sensing; validation

1. Introduction

Inland water worldwide has undergone tremendous changes from the pressure of climate change and human activities [1,2]. With the development of remote sensing technology, remote sensing data have been widely used in water monitoring, development, and water resources assessment [3–7]. More and more water remote sensing products are applied based on different sensors or spatial and temporal scales. For example, Tortini has used the multiple satellite data and the surface extent estimated data of the Moderate Resolution Imaging Spectrometer (MODIS) to produce the water remote sensing dataset of 347 lakes and reservoirs [8]; Wang released the global inland water remote sensing Forel-Ule index data set from 2000 to 2018 based on MODIS data [9]. These water remote sensing products have been widely used in water pollution, environmental monitoring, eutrophication, etc. [10–14]. However, many uncertainties exist in the quality of water remote sensing products retrieved by different inversion algorithms. Therefore, scientific evaluation of these water remote sensing products is the basis for applying water remote sensing products. The validation of water remote sensing products independently evaluates their uncertainty by comparing them with reference data (relative true values) [15,16]. The validation of water remote sensing products includes essential technical methods, such as spatial sampling design, scaling conversion, validation strategy, etc. [17]. Among them, as significant work in validation, the scientificity and reliability of spatial sampling design directly influence the quality of the in-situ data [15,18].
The spatial sampling design of shipboard measurements usually faces the difficulty of the limited sampling cost, the spatial distribution heterogeneity of water parameters, and the stability of the sampling design method. First, the number of points for water tests is limited by factors, such as the area of the lake, sampling time, and the speed of the water sampling vessel. In order to reduce the sampling cost and obtain more in-situ samples simultaneously, a unique sampling plan is usually designed in one test to collect multiple water parameters. Secondly, the on-site sampling points should have a high spatial representation to reflect the spatial distribution of the different water parameters and provide sufficient in-situ data for the inversion and the validation of remote sensing products. However, there are many water parameters (such as chlorophyll a (Chl-a), total suspended matter (TSM), and Secchi-disk Depth (SD), etc.) in the field sampling, and their spatial distribution is different. Finally, the optical properties of water parameters in different water bodies are quite different, and some sampling designs based on experience usually face the situation of insufficient prior knowledge. Therefore, the objectivity and robustness of the sampling design method can effectively improve the applicability of the experimental scheme and meet the data collection needs of different water bodies.

The research on water sampling design focuses on the spatial representation and stability of the water single-parameter sampling method. In contrast, the water multi-parameter synchronous sampling design research is insufficient [19–21]. The existing water sampling design methods are divided into two directions: traditional and intelligent optimal sampling methods. The traditional sampling methods include random, stratified, and systematic [22–24]. Random sampling is to use a random number generator to determine the sampling points, and the uncertainty of the spatial representation of the sampling points is relatively considerable. The systematic sampling uses a regular grid to select the sample locations at a fixed, periodic interval. Stratified sampling uses the prior information to divide the sampling area into different subclasses and randomly determines the sampling points from various subclasses according to the specified proportion. The intelligent optimal sampling method takes the geostatistical model as the objective function [25]. It uses optimization algorithms, such as spatial simulated annealing (SSA), genetic algorithm (GA), and particle swarm optimization (PSO), to solve the sampling points [26–30]. Compared with traditional sampling methods, intelligent optimal sampling methods utilize autocorrelation and spatial heterogeneity to improve the spatial representativeness of the sampling points [31–33]. These sampling methods are widely used in the water single-parameter sampling design. Still, few researchers consider the sampling design method of water multi-parameters [19,34,35]. The universal cokriging (UCK) sampling method is a water multi-parameter sampling optimization method proposed by Ge and Wangle et al. The sampling design results deployed long-term water parameter monitoring sensors [36]. However, there are still no effective multi-parameter sampling design methods for shipboard measurements. As mentioned above, the shipboard measurement generally adopts a unique sampling plan to obtain the in-situ data of multiple water parameters. Therefore, the focus of this study is that one sampling design can make multiple water parameters in all sampling sites have good spatial representation and stability.

This paper proposed a water multi-parameter sampling design method based on sample points fusion in adaptive weighted space to solve the multi-parameter sampling design problem. The research structure of this paper is as follows: (1) by comparing the results of multiple spatial sampling design methods, the optimal method suitable for single-parameter sampling is selected. (2) By constructing a multi-parameter weighted evaluation space, an adaptive sample point fusion sampling method is established, and its accuracy is evaluated. (3) Evaluate the model using measured field data, verify its accuracy, and apply it to field experiments in different water bodies.
2. Data

2.1. Study Area

This study's two sampling areas are located in Nanyi Lake and Bosten lake in China, as shown in Figure 1. As the experimental area, Nanyi Lake and Bosten Lake were used for construction and sampling method verification.

![Sampling areas](image)

Figure 1. The geographical location of the two experimental areas in this study, the sampling area, is in the red box: Nanyi Lake (a) and Boston Lake (b).

Nanyi Lake is located in the southeastern part of Anhui Province, China. Its center coordinates are 118°56'E and 31°05'N. It is 20 km long and 17 km wide and has a water area of 189 km². Bosten Lake is located in the northwestern part of Xinjiang Province, China. It is the largest inland freshwater lake in China. Its location is between 86°40' ~ 87°25'E and 41°56' ~ 42°14'N. It is 55 km long and 25 km wide and has a water area of 1646 km².
2.2. Experimental Data

The experimental data included remote sensing image data and in-situ data. The former included Sentinel-2 and GaoFen (GF) high-resolution satellite remote sensing image data, and the latter was a water multi-parameter dataset measured by various instruments. The specific information is shown in Table 1.

Table 1. Study Data in Nanyi Lake and Bosten Lake.

| Lake       | Purpose       | Satellite Image Date | In-Situ Data                   |
|------------|---------------|----------------------|--------------------------------|
| Nanyi Lake | Sampling design | Time: 28 August 2021 | Time: 31 August 2021           |
|            | Validation    | Satellite: Sentinel-2| Water parameters: Chl-a, TSM, SD|
| Bosten Lake| Sampling design | Time: 10 September 2021 | Time: 14 September 2021         |
|            | Validation    | Satellite: GF-1      | Water parameters: Chl-a, TSM, SD|

2.3. Preprocessing of Remote Sensing Dataset and In-Situ Dataset

Remote sensing images were used from multiple satellites as prior knowledge in this paper. The sentinel-2 satellite remote sensing images were obtained through the European Union’s Earth observation program (https://scihub.copernicus.eu, accessed on 28 August 2021). The GaoFen satellite remote sensing data were obtained through the China Centre for Resources Satellite Data and Application (https://www.cresda.com, accessed on 10 September 2021). Firstly, radiometric and atmospheric corrections are carried out on remote sensing images. Then, the GF data are resampled to 10 m, and the effect of outliers is removed. The water extraction algorithm was used to separate land and water. Erosion Operation in Mathematical Morphology was used to remove the influence of the shore area. Third, the in-situ dataset is divided into two parts. The in-situ dataset of Nanyi Lake includes three water parameter data of Chl-a, TSM, and SD of 15 sampling points. The collected water parameters were Chl-a, TSM, and SD. The in-situ dataset of Bosten Lake includes three water quality parameter data of 32 sampling points, and 16 sampling points are arranged using the adaptive weight sampling method and the systematic sampling method, respectively. The spatial distribution of sampling points is shown in Figure 1. In field measured data. A water parameter spectrometer directly measured the Chl-a; the TSM was measured by conventional drying, baking, and weighing methods; the SD was measured by the Secchi disk [37]. The preliminary statistics of the measured data are shown in Table 2.

Table 2. In-situ data information statistics table.

| Lake       | Parameter | Maximum | Minimum | Average | Standard Deviation |
|------------|-----------|---------|---------|---------|--------------------|
| Nanyi Lake | Chl-a     | 2.23    | 0.71    | 1.39    | 0.46               |
|            | TSM       | 16      | 4       | 9.60    | 3.59               |
|            | SD        | 125     | 82      | 104.60  | 11.10              |
| Bosten Lake| Chl-a     | 1.12    | 0.23    | 0.63    | 0.26               |
|            | TSM       | 9       | 1       | 5.28    | 2.20               |
|            | SD        | 366     | 240     | 309.97  | 26.62              |

3. Method

The water multi-parameter sampling design method proposed is based on the existing water single-parameter sampling design method. Therefore, comparing the current spatial sampling design methods is necessary to evaluate their effectiveness and representativeness in water sampling design. This paper studies the spatial sampling design using high-
resolution satellite imagery data near the field sampling time as prior knowledge. The flow chart of the sampling method is shown in Figure 2.

![Flowchart of the method used in this paper.](image)

3.1. Spatial Distribution of Water Parameters

Understanding the spatial distribution of water parameters is essential for research sampling design. Because the spatial distribution of the parameters changes with time, in most cases, prior knowledge of various water parameters cannot be obtained in time. Remote sensing images have the characteristics of fast information acquisition and short cycles. Therefore, it is an effective method to get the spatial distribution of the parameters by using remote sensing images near the sampling time [38–41]. To ensure the broad applicability of the sampling design method, the sensitive band or the combination of sensitive bands for water parameters on remote sensing images was applied to represent the spatial distribution characteristics of the parameters [42]. In this paper, taking the experimental sample area of Nanyi Lake as an example, a sampling design study was carried out on the three parameters of Chl-a, TSM, and SD. Previous studies have shown that the near-infrared/red, red/green, and green bands can be used as remote sensing feature bands for Chl-a, TSM, and SD [43–47]. Therefore, this paper uses Band 8/Band 3, Band 4/Band 3, and Band 3 of Sentinel-2 satellites as the sensitive bands of Chl-a, TSM, and SD and generates the spatial distribution characteristics map of the water parameters.

To describe the heterogeneity of the spatial distribution characteristics of each water parameter, this paper uses K-means to perform spatial clustering on the spatial distribution characteristic map of each water parameter [48,49]. The optimal number of clusters for K-means clustering was determined using the sum of squared errors (SSE). SSE is the sum of the squared errors of each observation and its cluster center, which measures the closeness within a category [50]. As shown in Formula (1):

\[
SSE = \sum_{i=1}^{k} \sum_{x \in C_i} d^2(x, m_i)
\]  

(1)
where $C_i$ represents all points in the $i$th category, $k$ is the number of categories, and $d^2(x, m_i)$ represents the Euclidean distance between the points in $C_i$ and the cluster center $m_i$.

### 3.2. Comparison of Water Single-Parameter Sampling Methods

High-quality single-parameter sampling design results are the premise of multi-parameter spatial sampling design. To verify the effectiveness of different single water parameter spatial sampling methods, this paper selects and evaluates six water parameter spatial sampling methods, including three conventional methods (random sampling, systematic sampling, stratified sampling) and three intelligent optimization methods based on objective function (SSA, GA, PSO). The specific information of each sampling design method is shown in Table 3. Among them, the objective function of spatial sampling design based on an intelligent optimization algorithm is the minimization of mean square error ($MSE$). $MSE$ is the mean square error between the kriging interpolation surface of the sampling points and the spatial distribution characteristic map of the water parameters [51,52]. As shown in Formula (2):

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (V_i - \hat{V}_i)^2$$  \hspace{1cm} (2)

where $V_i$ represents the pixel value of the spatial distribution map of the parameters, $\hat{V}_i$ represents the estimated value of sampling area obtained by kriging interpolation surface, and $N$ represents the number of pixels of the sampling area.

### Table 3. Water single-parameter sampling design method.

| Sampling Method | Objective Function | Theory |
|-----------------|--------------------|--------|
| random sampling  | Null               | Select the sample points by random number [53] |
| stratified sampling | Divide the study area into sub-areas and use random sampling within each sub-area [24,54,55] |
| systematic sampling | Select the sample points by using a regular grid [24,56] |
| GA sampling     | $MSE$              | The objective function is minimized by genetic operations, such as selection, crossover, and variation of different initial sampling points [57,58] |
| SSA sampling    |                    | The objective function is minimized by allocating sampling locations randomly [36,59,60] |
| PSO sampling    |                    | The objective function is to minimize collaboration and information sharing among individuals in the group [61] |

### 3.3. Water Multi-Parameter Sampling Method

The water multi-parameter sampling design needs to comprehensively consider the spatial distribution of various parameters. Therefore, to balance the local spatial representation of multiple parameter sampling points, we propose an adaptive sampling points fusion method in multi-parameter weighted space in this paper. The basic idea is to use the sampling points of multiple groups of parameters to construct a local weighted space. The multi-parameter sampling points in the weighted space are obtained by weighted fusion of the single-parameter sampling points. The weighted space refers to the non-overlapping polygonal areas (or lines) constructed adaptively with multiple parameter sampling points as vertices in the local sampling area. An example of the weighted space is shown in Figure 3. When there are two water parameters, the weighted space is a line segment $(a, b)$. When there are three or more water parameters, the weighted space is a polygonal area $(c, d)$. The basic framework of multi-parameter sampling design:

$$Q(x, y) = w_1 \times P_1(x_1, y_1) + w_2 \times P_2(x_2, x_2) \cdots w_n \times P_n(x_n, y_n)$$  \hspace{1cm} (3)

where $Q(x, y)$ represents adaptive sampling points, $w_1, w_2 \cdots w_n$ represents weight factor, $P_1(x_1, y_1), P_2(x_2, y_2) \cdots P_n(x_n, y_n)$ represents each parameter sampling points.
Among them, the size of each weight factor determines the position of the adaptive sampling points in the weighted space and the ability to balance the sampling points of each water parameter. In conventional methods, equal weights are usually used to construct weighting factors, but equal weights are more suitable for the case where the spatial variation of each parameter is consistent. Since the spatial variation characteristics of various water parameters in the weighted space are different and in order to determine the optimal adaptive sampling point in the weighted space, this paper uses the spatial variation characteristics of each water parameter in the weighted space to construct the weights. The faster the characteristics change, the stronger the spatial heterogeneity and it should be given a higher weight. In this paper, directional derivatives are used to describe the rate of spatial variation of the water parameters at the sampling point [62]. The solution process is as follows:

The weight coefficient $w_i$ depends on the directional derivatives of $P_i(x, y)$ and $Q(x, y)$, and the calculation method is as follows:

$$w_i = f_i(\varphi_i)$$  \hspace{1cm} (7)

The multi-parameter sample point locations are solved by the ordinary least square method:

$$\alpha / f_i(\varphi_i) = \sqrt{(x-x_i)^2 + (y-y_i)^2}, \ i = 1, 2 \cdots n$$  \hspace{1cm} (8)

where $\alpha$ is the scaling factor.

A simple numerical example is shown to explain the solution process of the adaptive weight sampling design. First, as shown in Figure 3d, $P_1$, $P_2$, $P_3$ are Chl-a, TSM, and SD sampling points, respectively; Q is the adaptive weight sampling point. Their coordinates and values are assumed for:

$$P_1 = (1, 1), \ P_2 = (1, 2), \ P_3 = (3, 1), \ Q = (x, y)$$  \hspace{1cm} (9)

$$V_{P_1} = 0.8, \ V_{P_2} = 0.7, \ V_{P_3} = 0.6$$  \hspace{1cm} (10)
Secondly, since the water parameters show a very obvious continuous change in the water, the rate of change of Chl-a, TSM and SD values near the sampling points is assumed to be linear with the distance. Therefore, the change functions of the three water parameters are:

\[
\begin{align*}
F_{\text{Chl-a}}(r) &= V_P - 0.28r \\
F_{\text{TSM}}(r) &= V_P - 0.17r \\
F_{\text{SD}}(r) &= V_P - 0.15r
\end{align*}
\]  

(11)

where \( r \) is the Euclidean distance between the other sampling points and the sampling points of the three water parameter.

The advantage of the above assumption is that the rate of change of the water parameters in any direction near the sampling point is constant. In other words, its directional derivative is the first derivative of the change function:

\[
\begin{align*}
\omega_{\text{Chl-a}} &= f_{\text{Chl-a}}(\phi) = \frac{\partial F_{\text{Chl-a}}}{\partial r} = 0.28 \\
\omega_{\text{TSM}} &= f_{\text{TSM}}(\phi) = \frac{\partial F_{\text{TSM}}}{\partial r} = 0.17 \\
\omega_{\text{SD}} &= f_{\text{SD}}(\phi) = \frac{\partial F_{\text{SD}}}{\partial r} = 0.15
\end{align*}
\]  

(12)

Finally, the adaptive sampling point is solved by its Euclidean distance relationship with the three water parameter sampling points:

\[
\begin{align*}
\alpha_{0.28} &= \sqrt{(x - 1)^2 + (y - 1)^2} \\
\alpha_{0.17} &= \sqrt{(x - 1)^2 + (y - 2)^2} \\
\alpha_{0.15} &= \sqrt{(x - 3)^2 + (y - 1)^2}
\end{align*}
\]  

(13)

with \( \alpha = 0.313, \; x = 1, \; y = 0.5, \; Q = (1, 0.5) \)

3.4. Evaluation Method

Relative precision (\( RP_{\text{mean}} \)) is used to evaluate the representativeness of sampling, and it is defined as the ratio of the mean of the sample data to the mean of all the data in the sampling [23]. The formula is as follows:

\[
RP_{\text{mean}} = \frac{\text{mean}(x_i)}{\text{mean}(x_b)}
\]  

(14)

where \( \text{mean}(x_i) \) represents the sample mean of the \( i \)th sampling method, \( \text{mean}(x_b) \) represents the average of all data in the sampling area.

The root means square error (\( RMSE \)) was utilized to assess the effectiveness of the sampling methods. \( RMSE \) is the root mean square error between the kriging interpolation surface of the sampling points and the spatial distribution characteristics map of the parameters [63]. The formula for \( RMSE \) is as follows:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (V_i - \hat{V}_i)^2}{N}}
\]  

(15)

where \( V_i \) represents the pixel value of the spatial distribution characteristics map of the water parameters, \( \hat{V}_i \) represents the estimated value of the sampling area obtained by kriging interpolation, and \( N \) represents number of pixels of the sampling area.

The Spearman correlation coefficient (\( R \)) was applied to evaluate the correlation between the kriging interpolation surface of the in-situ data and remote sensing image [64]. The formula is as follows:

\[
R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}
\]  

(16)
where \( n \) is the number of sample points, and \( d_i \) is the grade difference between the value on the remote sensing image and the kriging interpolation surface of the in-situ data value. In this paper, 1000 sampling points are uniformly selected in the sampling area.

In order to evaluate the representativeness of the measured data over the range. The average relative error (MRE) is used to evaluate the representativeness of the in-situ data in the range. Ideally, the water parameter data collected should be equally spaced across the range. It can make the in-situ data more representative in the experimental area. Therefore, we assume a set of uniformly distributed simulated sampled data as the standard value, and the MRE assess the error of the in-situ data and the standard value [65]. The lower the MRE, the better the distribution of the in-situ data in the range. The formula is shown in Formula (12):

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{(V_i - S_i)}{S_i}
\]

where \( V_i \) represents the in-situ data, \( S_i \) represents the simulated data uniformly distributed in the range, and \( n \) is the number of the in-situ data.

4. Results and Analysis
4.1. Spatial Distribution of Various Water Parameters

Figure 4 reveals the trend of SSE and the spatial distribution characteristics of the three parameters. The SSE decreases with the increase in clusters K. When K > 4, the SSE of change tends to be flat, and the SSE does not improve much as the number of clusters K increases. Therefore, we determined that the number of clusters of each parameter is four. From Figure 4b–d, each class of Chl-a was relatively discrete. The spatial distribution of each category of the TSM and SD was relatively concentrated, and there is heterogeneity in the spatial distribution characteristics of each water parameter.

![Figure 4](image)

**Figure 4.** Variations of SSE with increasing K (a) and the spatial distribution of Chl-a (b), TSM (c), and SD (d) in Nanyi Lake.

4.2. Effectiveness Evaluation of the Various Sampling Methods

4.2.1. Spatial Representative Comparison of Various Sampling Design Methods

Sampling points can reflect the overall distribution characteristics of water parameters in the sampling area. The representativeness of the sampling points can be represented by the difference between the statistical data of the sampling point and the statistical data of the population \( (RP_{mean}) \). The \( RP_{mean} \) is used to analyze the representativeness of sampling points in statistical theory in this paper. This paper used six sampling methods to conduct 100 simulated samplings on the spatial distribution characteristic maps of the three water parameters. The number of sampling points was set to 15. The \( RP_{mean} \) value of each simulated sample was extracted as shown in Figure 5. It can be concluded that
the interquartile ranges (IQR) of SSA in the three water parameters are 0.0051, 0.0076 and 0.0070, and the results are all smaller than other sampling methods. Its range is also smaller than other sampling methods, which indicates the representativeness of the SSA sampling points are less volatile. Then, although the median value of stratified sampling in Chl-a is slightly higher than that of the SSA, in TSM and SD, the median of the SSA is closer to one than other methods. Random and GA methods have too high or too low outliers, and the representativeness of their sampling results is relatively unstable. Therefore, the $R_{\text{mean}}$ results show that the sampling points obtained by the SSA method are well represented in the characteristic distribution maps of the three water parameters.

Figure 5. Boxplots of $R_{\text{mean}}$ for 100 simulations for each sampling method, Chl-a (a), TTS (b), SD (c).

4.2.2. Water Single-Parameter Sampling Points

Figure 7 presents the decrease in RMSE during SSA (a, b, c) and Chl-a TSM and SD sampling points (d, e, f). After this initial phase, the RMSE decreased steadily and gradually stabled after 10,000 iterations. No further reduction in the RMSE was achieved after about 14,000 iterations, indicating that the sampling design reached the optimal solution. The RMSE for the three water parameters was reduced from 0.0537, 0.119, and 0.01 to

In order to further evaluate the spatial representativeness of various sampling methods, the RMSE between the kriging interpolation of sampling points and the spatial distribution characteristics maps of the three water parameters was used as the evaluation index. The RMSE value of each simulated sampling is recorded, as shown in Figure 6. The RMSE box plot of a sample with good representability should have the following characteristics: (1) the median and IQR should be small; (2) no outliers [23]. From Figure 6, in the characteristic distribution map of the three water parameters, the median of the RMSE of the SSA sampling method was 0.0135, 0.0216, and 0.0055, respectively. Compared with other sampling methods, the median of the SSA sampling method was the smallest. In addition, the IQR of the SSA method was significantly better than the other sampling methods. It showed that the SSA sampling method has good spatial representation and stability. The random, stratified, and GA all had local outliers, indicating that their sampling results were unstable. The result of the systematic sampling method after a fixed interval was more dependent on the heterogeneity of the sampling area. If the sampling area has firm heterogeneity, the representativeness of its sampling points is relatively weak. The median and the IQR of the PSO sampling method are second only to the SSA and outperform other sampling methods, and the sampling results also have good stability. Therefore, it can be concluded that the SSA method is more suitable for the water single-parameter optimal sampling method.
4.2.2. Water Single-Parameter Sampling Points

Figure 7 presents the decrease in RMSE during SSA (a, b, c) and Chl-a TSM and SD sampling points (d, e, f). After this initial phase, the RMSE decreased steadily and gradually stabled after 10,000 iterations. No further reduction in the RMSE was achieved after about 14,000 iterations, indicating that the sampling design reached the optimal solution. The RMSE for the three water parameters was reduced from 0.0537, 0.119, and 0.01 to 0.0128, 0.0196, and 0.006. It showed that the prediction accuracy of the three water parameters was improved by 76%, 83%, and 50%. At the same time, we found that the sampling points of the three water parameters were evenly distributed in each category of the sampling area. This showed that the SSA improves the representativeness of the sampling points of the three water parameters.

4.3. Accuracy Evaluation of Water Multi-Parameter Sampling Method

Figure 8 shows the multi-parameter weighted space and sampling points of the adaptive weight sampling design method. Table 4 lists the RMSE of Chl-a sampling points, TSM sampling points, SD sampling points, adaptive weight sampling points, the centroid of the triangle, and the incentre of the triangle in the distribution characteristic map of the three water parameters. The optimization results show that sampling points of each parameter...
only had the smallest RMSE on its spatial distribution characteristic map. In contrast, the RMSE on the spatial distribution characteristic map of other water parameters was higher, indicating that the single-parameter sampling points were less representative of the other water parameters’ spatial distribution characteristic map. The RMSE of adaptive weight sampling points in the spatial distribution characteristic map of the three water parameters were 0.0156, 0.0216, and 0.0065. Compared with the optimal sampling points for single-parameter, the spatial representativeness of the adaptive weight sampling points was slightly reduced. Still, its advantage was that it could maintain relatively high spatial representativeness simultaneously among the three water parameters and demonstrated the ability to balance the spatial distribution of multiple water parameters.

![Image of Adaptive weight sampling points]

**Figure 8.** Example of Adaptive weight sampling points.

### 4.4. Accuracy Evaluation of In-Situ Dataset of Water Parameters

To assess the spatial representativeness of the adaptive weight sampling design method, we used the in-situ dataset to verify the sampling method. In the water experiment of Nanyi Lake, this sampling method was applied to design 15 sampling points to assess the effectiveness of the sampling method. In the water experiment of Bosten Lake, 16 sampling points were designed using systematic sampling and adaptive weight sampling design method to prove the spatial representativeness of the sampling method.

Figure 9 displays the 15 sampling points and the MRE of the in-situ and simulated data in Nanyi Lake. The simulated data are a set of optimal values obtained by assuming that the data of the sampling points are uniformly distributed in the range. The range of the simulated values is the maximum and minimum values of the in-situ data. The simulated values of the other sampling points were evenly distributed in the value. The MRE of the in-situ and simulated data for the three water parameters was 6.76%, 5.72%, and 2.67%. The errors were within an acceptable range (10%), indicating that the in-situ data were uniformly distributed and had a good representation in the range.

| RMSE                  | Chl-a       | TSM         | SD          |
|-----------------------|-------------|-------------|-------------|
| Chl-a sampling points | 0.0128      | 0.0227      | 0.0081      |
| TSM sampling points   | 0.0241      | 0.0196      | 0.0067      |
| SD sampling points    | 0.0236      | 0.0267      | 0.0060      |
| Adaptive weight points| 0.0156      | 0.0216      | 0.0065      |

**Table 4.** Spatial representative evaluation of various sampling points.

![Image of Multi-parameter sampling points in Nanyi lake](a–c), and the MRE of the in-situ and simulated data (d–f), and the black dots are the simulated data, and the red dots are the in-situ data.
On the other hand, taking SD as an example, the range of systematic samplings was 17.15%, 29.54%, and 6.54%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 11.76%, and 4.68%, while the MRE of the systematic sampling method was 6.24%, 9.63%, and 4.68%.

Figure 9 displays the 15 sampling points and the MRE of the in-situ data and simulated data in Bosten Lake. On the one hand, the MRE of the adaptive weight sampling method was 6.24%, 9.63%, and 4.68%, while the MRE of the systematic sampling method was 17.15%, 29.54%, and 6.54%. The MRE of Chl-a, TSM, and SD improved by 10.91%, 19.91%, and 1.86%. On the other hand, taking SD as an example, the range of systematic sampling points was [340–330], while the range of the multi-parameter sampling points was [240, 366], and the breadth of the range of the latter was improved by 27%. The results showed that the adaptive weight sampling points were more uniformly distributed and had more capacity in the range than the systematic sampling points.

To further evaluate the spatial representativeness of the adaptive weight sampling method, we extracted the sensitive bands of the three water parameters in the synchronously observed remote sensing images during on-site sampling and calculated the spatial distribution characteristic maps of the three water parameters. The interpolation surface of the sampling area is obtained by performing kriging interpolation on the in-situ data.
of the sampling points. Finally, in this paper, 1,000 sampling points are simultaneously extracted from the spatial distribution characteristics map of water parameters and the kriging interpolation surface for correlation analysis; the results are shown in Figure 11. The correlation coefficients of the systematic sampling method were 0.45, 0.54, and 0.61, while the correlation coefficients of the adaptive weight sampling method were 0.60, 0.65, and 0.60. The correlation coefficients of Chl-a and TSM are relatively improved by 0.13 and 0.11, while the correlation coefficients of SD are equal. The results show that the correlation of adaptive weight sampling points in the three water parameters is better than the systematic sampling method. Meanwhile, we find that the systematic sampling points have better spatial representation in SD, but the spatial representation of Chl-a and TSM is insufficient. The adaptive weight sampling design method has a similar correlation among the three water parameters, indicating that it can balance the spatial distribution of three water parameters. Thus, it is reasonable to conclude that the adaptive weight sampling design method is significantly better than the systematic sampling method.

![Figure 11](image.png)

Figure 11. The Spearman correlation coefficient of multi-parameter sampling design method and systematic sampling method, Chl-a (a,d), TSM (b,e), SD (c,f).

5. Discussion

The spatial sampling design is an essential prerequisite for ensuring the validity and rationality of in-situ data for the inversion and validation of water remote sensing products. More and more researchers pay attention to the spatial representation of sampling points research. The field measurement of water parameters mainly includes fixed station and shipboard measurements. Some scholars have researched water multi-parameter sampling design, but they are primarily used to select long-term fixed measurement points [36]. Different from measurement at fixed stations, shipboard measurement faces more challenges. On the one hand, the sampling design will make necessary changes to factors, such as weather, ships, and waterways. This requires sampling designs to be fast, efficient, and stable. On the other hand, the number of sampling points is limited due to the limitations of economic cost and the time window of satellite-ground synchronization observation. It is usually necessary to collect multiple water quality parameters at one sampling point. However, there is a specific heterogeneity in the spatial distribution of different water quality parameters. Designing sampling points to improve their spatial representation in different water parameters under a limited number of sampling points to meet the needs of the inversion and verification of remote sensing satellite products is also a key problem to be solved in this paper. Therefore, this paper focuses on the sampling design problem in the shipboard water parameter measurement tests. We propose a water multi-parameter sampling design method based on adaptive sample points fusion in multi-parameter weighted space. The results show that it effectively solved the problem of multi-parameter spatial representation under the condition of heterogeneous spatial
distribution characteristics. According to our calculations, taking the Nanyi Lake sampling area as an example, the algorithm’s time complexity is one-fourth to one-sixth of the UCK sampling method. The fast and efficient sampling design capability enables the sampling plan to be redesigned quickly in variable environmental conditions. This makes the method more suitable for sudden needs, such as algal blooms, post-disaster assessment, and other applications. This dramatically improves the ability of the shipboard measurements scheme to cope with changes in various conditions.

Although the multi-parameter sampling method has achieved good analysis results, the practical application capabilities need further evaluation. The multi-parameter sampling method needs some constraints to better apply to the actual sampling design. The most crucial issue is the validity of prior knowledge. This paper uses high-resolution remote sensing satellites as prior knowledge to reflect the spatial representation of different water parameters. Due to weather and satellite transit time, we may not be able to obtain the nearest high-resolution satellite remote sensing imagery before field sampling. Therefore, satellite images more days away from the experiment time and even contemporaneous satellite images from other years in history are used for prior knowledge acquisition. For example, we use remote sensing images four days before on-site measurement as the previous knowledge of water sampling in Bosten Lake. Due to the influence of various environmental conditions, the spatial distribution of water parameters will change with time, and this change cannot be effectively simulated and predicted. Therefore, it is necessary to analyze the time scale effect of the spatial distribution of water parameters in combination with the influencing factors of water environment changes to ensure the reliability of the spatial representation of remote sensing images. Moreover, the remote sensing image data may come from different satellites. For example, the remote sensing image data of GF and sentinel-2 satellites are used in this paper. Different satellites have different sensitive bands for water parameters. Different remote sensing satellite data may not be able to use the same band or band combination when describing the spatial representation of water parameters. The wrong band selection may lead to insufficient spatial representation, which will significantly affect the reliability of the sampling design method. The literature analysis of water quality parameter modeling is an excellent way to obtain the sensitive bands of different satellites to water quality parameters [47,66]. Therefore, it is necessary to study the sensitive bands of different satellites to different water parameters to determine the optimal combination of sensitive bands.

Spatial–temporal constraints in sampling design are also factors to consider. This paper pays more attention to evaluating the spatial representation of sampling, ignoring the time constraints. In fact, in the sampling design, the effect of time cannot be overlooked; it directly affects whether the experiment can be completed. For example, to ensure the reliability of the in-situ data, we need to be collected within ±1.5 h of the satellite transit. Usually, we determine the sampling range of the lake surface based on the experience of previous experiments. In fact, under the common constraints of the sampling ship’s sailing speed and the dock’s position, the area of a single sampling is within a limited range and cannot be estimated empirically. Especially for large inland waters (such as Lake Bosten), it is not easy to perform on-site sampling of the entire water surface within the specified time (±1.5). In this case, it is necessary to consider reducing the sampling points or the sampling area to ensure the quality of in-situ data. Therefore, we should also consider the time and space cost (for example, geographic accessibility, satellite transit time, water area, ship speed, dock location, etc.) of on-site sampling in the sampling design to consider the sampling design methods comprehensively.

To sum up, although the sampling design method proposed in this paper can quantitatively and objectively carry out the experimental design of water multi-parameters, it still faces some problems that need in-depth research. In future research, it is necessary to explore and design more effective and representative multi-parameter sampling design methods based on an in-depth analysis of the reliability of auxiliary data under more constraints to meet the needs of water remote sensing product inversion and verification.
6. Conclusions

In order to solve the problem of multi-parameter spatial representation in water sampling, this paper proposed a water multi-parameter sampling design method based on adaptive sample points fusion in multi-parameter weighted space, which has obvious advantages compared with the traditional sampling method. Using high-resolution satellite remote sensing images before the experiment as prior knowledge, the spatial distribution characteristics of water parameters can be reflected in time, which provides an effective reference for sampling design methods. On this basis, a multi-parameter sampling design method is constructed through techniques, such as regional weighted space construction and adaptive weight fusion. This method can consider various water parameters’ spatial distribution and spatial variation characteristics. The research results show that the sampling results have good spatial representation across multiple water parameters, which can provide effective data support for satellite-ground synchronization observations. Therefore, the water multi-parameter sampling design method can provide an efficient and reliable sampling design method for the inversion and the verification of water remote sensing products. The continuous improvement of sampling methods will be widely used in water remote sensing products, monitoring, and evaluation.

Author Contributions: Conceptualization, M.Z., Z.T. and X.Z.; Formal analysis, M.Z.; Investigation, Z.T. and X.Z.; Methodology, M.Z.; Project administration, J.W.; Software, M.Z.; Supervision, X.Z.; Validation, Z.T. and X.Z.; Visualization, R.L. Writing—original draft, M.Z.; Writing—review & editing, M.Z., Z.T., T.L. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2018YFE0124200).

Data Availability Statement: The satellite data used in this study are in the public domain, available from ESA (https://scihub.copernicus.eu/dhus/#/home, accessed on 13 December 2021) and CRESDA (http://www.cresda.com/CN/, accessed on 13 December 2021). Other data that support the findings of this study are available from the author upon reasonable request.

Acknowledgments: We would like to thank Python for helping us to design the sampling strategy and perform data analysis and plotting.

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

References
1. Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic–spectral classification algorithm. Int. J. Digit. Earth 2016, 9, 113–133. [CrossRef]
2. Woolway, R.I.; Kraemer, B.M.; Lenters, J.D.; Merchant, C.J.; O’Reilly, C.M.; Sharma, S. Global lake responses to climate change. Nat. Rev. Earth Environ. 2020, 1, 388–403. [CrossRef]
3. Dörnhöfer, K.; Oppelt, N. Remote sensing for lake research and monitoring—Recent advances. Ecol. Indic. 2016, 64, 105–122. [CrossRef]
4. Bonansea, M.; Ledesma, M.; Rodriguez, C.; Pinotti, L. Using new remote sensing satellites for assessing water quality in a reservoir. Hydrol. Sci. J. 2019, 64, 34–44. [CrossRef]
5. Aires, F.; Venot, J.-P.; Massuel, S.; Gratiot, N.; Pham-Duc, B.; Prigent, C. Surface water evolution (2001–2017) at the Cambodiva/Vietnam border in the upper mekong delta using satellite MODIS observations. Remote Sens. 2020, 12, 800. [CrossRef]
6. Rai, P.K.; Chandel, R.S.; Mishra, V.N.; Singh, P. Hydrological inferences through morphometric analysis of lower Kosi river basin of India for water resource management based on remote sensing data. Appl. Water Sci. 2018, 8, 15. [CrossRef]
7. Wang, X.; Xie, H. A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. Water 2018, 10, 608. [CrossRef]
8. Tortini, R.; Noujdina, N.; Yeo, S.; Ricko, M.; Birkett, C.M.; Khandelwal, A.; Kumar, V.; Marlier, M.E.; Lettenmaier, D.P. Satellite-based remote sensing data set of global surface water storage change from 1992 to 2018. Earth Syst. Sci. Data 2020, 12, 1141–1151. [CrossRef]
9. Wang, S.; Li, J.; Zhang, W.; Cao, C.; Zhang, F.; Shen, Q.; Zhang, X.; Zhang, B. A dataset of remote-sensed Forel-Ule Index for global inland waters during 2000–2018. Sci. Data 2021, 8, 26. [CrossRef]
10. Wang, X.; Yang, W. Water quality monitoring and evaluation using remote sensing techniques in China: A systematic review. Ecosyst. Health Sustain. 2019, 5, 47–56. [CrossRef]
11. Xu, J.; Lei, S.; Bi, S.; Li, Y.; Lyu, H.; Xu, J.; Xu, X.; Mu, M.; Miao, S.; Zeng, S. Tracking spatio-temporal dynamics of POC sources in eutrophic lakes by remote sensing. Water Res. 2020, 168, 115–162. [CrossRef] [PubMed]
12. Zang, W.; Lin, J.; Wang, Y.; Tao, H. Investigating small-scale water pollution with UAV remote sensing technology. In Proceedings of the World Automation Congress, Puerto Vallarta, Mexico, 24–28 June 2012; pp. 1–4.

13. Feng, L.; Hou, X.; Zheng, Y. Monitoring and understanding the water transparency changes of fifty large lakes on the Yangtze Plain based on long-term MODIS observations. Remote Sens. Environ. 2019, 221, 675–686. [CrossRef]

14. Palmer, S.C.; Kutser, T.; Hunter, P.D. Remote Sensing of Inland Waters: Challenges, Progress and Future Directions; Elsevier: Amsterdam, The Netherlands, 2015; pp. 1–8.

15. Rui, J.; Youhua, R.; Qinhuo, L. Key methods and experiment verification for the validation of quantitative remote sensing products. Adv. Earth Sci. 2017, 6, 630–642.

16. Justice, C.; Belward, A.; Morissette, J.; Lewis, P.; Privette, J.; Baret, F. Developments in the ‘validation’ of satellite sensor products for the study of the land surface. Int. J. Remote Sens. 2000, 21, 3383–3390. [CrossRef]

17. Wu, X.; Xiao, Q.; Wen, J.; You, D.; Hueni, A. Advances in quantitative remote sensing product validation: Overview and current status. Earth Sci. Rev. 2019, 196, 102875. [CrossRef]

18. Aliilou, H.; Nia, A.M.; Keshkar, H.; Han, D.; Bray, M. A cost-effective and efficient framework to determine water quality monitoring network locations. Sci. Total Environ. 2018, 624, 283–293. [CrossRef]

19. Kiefer, I.; Odermatt, D.; Anneville, O.; Wüest, A.; Bouffard, D. Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake. Sci. Total Environ. 2015, 527, 493–506. [CrossRef]

20. Hansen, C.H.; Williams, G.P.; Adjei, Z.; Barlow, A.; Nelson, E.J.; Miller, A.W. Reservoir water quality monitoring using remote sensing with seasonal models: Case study of five central-Utah reservoirs. Lake Reser. Manag. 2015, 31, 225–240. [CrossRef]

21. Noges, P.; Poiskane, S.; Koiv, T.; Noges, T. Effect of chlorophyll sampling design on water quality assessment in thermally stratified lakes. Hydrobiologia 2010, 649, 157–170. [CrossRef]

22. Sun, P.; Zhang, J.; Congalton, R.G.; Pan, Y.; Zhu, X. A quantitative performance comparison of paddy rice acreage estimation using stratified sampling strategies with different stratification indicators. Int. J. Digit. Earth 2018, 11, 1001–1019. [CrossRef]

23. Lv, T.; Zhou, X.; Tao, Z.; Sun, X.; Wang, J.; Li, R.; Xie, F. Remote Sensing-Guided Spatial Sampling Strategy over Heterogeneous Surface Ground for Validation of Vegetation Indices Products with Medium and High Spatial Resolution. Remote Sens. 2021, 13, 2674. [CrossRef]

24. Brus, D.; Knotters, M. Sampling design for compliance monitoring of surface water quality: A case study in a Polder area. Water Resour. Res. 2008, 44, W11410. [CrossRef]

25. Ling, C.; Lu, Z. Adaptive Kriging coupled with importance sampling strategies for time-variant hybrid reliability analysis. Appl. Math. Model. 2020, 77, 1820–1841. [CrossRef]

26. Vaš, R.; Heuvelink, G.; Borůvka, L. Sampling design optimization for multivariate soil mapping. Geoderma 2010, 155, 147–153. [CrossRef]

27. Chen, B.; Pan, Y.; Wang, J.; Fu, Z.; Zeng, Z.; Zhou, Y.; Zhang, Y. Even sampling designs generation by efficient spatial simulated annealing. Math. Comput. Model. 2013, 58, 670–676. [CrossRef]

28. Wadoux, A.M.J.C.; Brus, D.J.; Heuvelink, G.B.M. Sampling design optimization for soil mapping with random forest. Geoderma 2019, 355, 113913. [CrossRef]

29. Puri, D.; Borel, K.; Vance, C.; Karthikeyan, R. Optimization of a water quality monitoring network using a spatially referenced water quality model and a genetic algorithm. Water 2017, 9, 704. [CrossRef]

30. Cai, X.; Qiu, H.; Gao, L.; Yang, P.; Shao, X. A multi-point sampling method based on kriging for global optimization. Struct. Multidiscip. Optim. 2017, 56, 71–88. [CrossRef]

31. Miralha, L.; Kim, D. Accounting for and predicting the influence of spatial autocorrelation in water quality modeling. ISPRS Int. J. Geo Inf. 2018, 7, 64. [CrossRef]

32. Yang, X.; Jin, W. GIS-based spatial regression and prediction of water quality in river networks: A case study in Iowa. J. Environ. Manag. 2010, 91, 1943–1951. [CrossRef]

33. Guedes, L.P.C.; Ribeiro, P.J., Jr.; De Stefano Piedade, S.; Uribe-Opazo, M.A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. Chil. J. Stat. 2011, 2, 39–50.

34. Li, J.; Tian, L.; Wang, Y.; Jin, S.; Li, T.; Hou, X. Optimal sampling strategy of water quality monitoring at high dynamic lakes: A remote sensing and spatial simulated annealing integrated approach. Sci. Total Environ. 2021, 777, 146113. [CrossRef]

35. Jiang, J.; Tang, S.; Han, D.; Fu, G.; Solomatine, D.; Zheng, Y. A comprehensive review on the design and optimization of surface water quality monitoring networks. Environ. Model. Softw. 2020, 132, 104792. [CrossRef]

36. Ge, Y.; Wang, J.; Heuvelink, G.B.; Jin, R.; Li, X.; Wang, J. Sampling design optimization of a wireless sensor network for monitoring ecophysiological processes in the Babao River basin, China. Int. J. Geogr. Inf. Sci. 2015, 29, 92–110. [CrossRef]

37. Rose, K.C.; Greb, S.R.; Diebel, M.; Turner, M.G. Annual precipitation regulates spatial and temporal drivers of lake water clarity. Ecol. Appl. 2017, 27, 632–643. [CrossRef] [PubMed]

38. He, Y.; Gong, Z.; Zheng, Y.; Zhang, Y. Inland Reservoir Water Quality Inversion and Eutrophication Evaluation Using BP Neural Network and Remote Sensing Imagery: A Case Study of Dashae Reservoir. Water 2021, 13, 2844. [CrossRef]

39. Cheng, M.; Jiang, M.; Huang, Z.; Lei, H.; Yan, D.; Zhu, F. Research on Baiyangdian Lake Water Body Changes and Water Quality Parameters Inversion Based on Landsat Dense Time Series Data. IOP Conf. Ser. Earth Environ. Sci. 2021, 783, 012134. [CrossRef]
