



# Article WenSiM: A Relative Accuracy Assessment Method for Land Cover Products Based on Optimal Transportation Theory

Rui Zhu <sup>1</sup><sup>(b)</sup>, Yumin Tan <sup>1</sup><sup>(b)</sup>, Ziqing Luo <sup>1</sup><sup>(b)</sup>, Yanzhe Shi <sup>1</sup><sup>(b)</sup>, Jiale Wang <sup>1</sup>, Guifei Jing <sup>2</sup> and Xiaolu Wang <sup>1</sup>,\*<sup>(b)</sup>

- <sup>1</sup> School of Transportation Science and Engineering, Beihang University, Beijing 102206, China; buaa\_zhurui@buaa.edu.cn (R.Z.); tanym@buaa.edu.cn (Y.T.); zqluo@buaa.edu.cn (Z.L.); syz\_cannot@buaa.edu.cn (Y.S.); wangjiale13@buaa.edu.cn (J.W.)
- <sup>2</sup> Research Institute for Frontier Science, Beihang University, Beijing 100191, China; guifeijing@buaa.edu.cn
- \* Correspondence: zy2113207@buaa.edu.cn

Abstract: Land cover (LC) products play a crucial role in various fields such as change detection, resource management, and urban planning. The diversity in methods and principles used to create different products poses a challenge for researchers in choosing the most suitable one for research needs. Mainstream evaluation methods typically consider only a portion of the accuracy information from the product and require a significant effort in creating validation datasets. Here, we propose a relative accuracy assessment method for LC products based on optimal transport theory, which provides a comprehensive evaluation by utilizing a broader range of accuracy information within the product. The method can directly compute the similarity between the target product and the reference truth at a global scale, addressing the issue of quantitatively assessing product accuracy in the absence of a validation dataset. To validate the effectiveness of the method, we select Beijing as the study area to assess the accuracy of four LC products. The results suggest that the method allows for precise quantification of product accuracy, aligning closely with validation outcomes, which can provide valuable guidance to researchers in both product creation and selection.

Keywords: accuracy assessment; land cover products; optimal transportation; relative accuracy

## 1. Introduction

Land cover (LC) products, as common remote sensing products, depict the interactions between human activities and the natural environment, which are widely applied in environmental change monitoring, ecosystem assessment, and sustainable development planning [1–3]. With the rapid advancement of satellite remote sensing technology and the continuous improvement in data accessibility, LC products continue to evolve and be updated. However, due to differences in the classification algorithms and schemes of products, researchers find it challenging to directly assess the quality of these products [4]. The accuracy assessment and rational selection of LC products have become key issues in research [5,6]. It is a widely held view that the absolute ground truth does not exist, making it impossible to directly compare remote sensing products with it [7].

Previous research has found that a validation dataset can be created for the study area using a full-sampling method and make the dataset I reference truth to evaluate the overall accuracy of LC products. But the high cost of creating the dataset has made the full-sampling method rarely used directly [8]. To strike a balance between cost and accuracy, a variety of studies have proposed two major categories of accuracy assessment methods: direct assessment methods and indirect assessment methods. The direct assessment method involves first obtaining samples through random sampling and then creating a validation dataset through field surveys or visual interpretation based on high-resolution satellite images such as Google Earth [9]. Finally, a confusion matrix is computed to determine the overall accuracy of the product [10]. Although the method saves costs, the number and methods of sampling can introduce errors and lead to unstable results. Furthermore, most



Citation: Zhu, R.; Tan, Y.; Luo, Z.; Shi, Y.; Wang, J.; Jing, G.; Wang, X. WenSiM: A Relative Accuracy Assessment Method for Land Cover Products Based on Optimal Transportation Theory. *Remote Sens.* 2024, *16*, 257. https://doi.org/ 10.3390/rs16020257

Academic Editor: Dino Ienco

Received: 27 November 2023 Revised: 30 December 2023 Accepted: 3 January 2024 Published: 9 January 2024



**Copyright:** © 2024 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/). validation datasets are not interoperable, making it difficult for researchers to select the highest-quality one. Moreover, the rationale of relying solely on one validation dataset to assess the accuracy of all LC products need to be questioned [11]. To avoid the creation of validation datasets, indirect evaluation methods that compare the similarity between products have been reported. Unlike direct assessment, the indirect assessment method evaluates relative accuracy by analyzing differences between products, primarily through regional consistency and spatial consistency analysis [12-14]. Specifically, regional consistency involves statistically assessing the composition of different land cover types within the study area [15]. Spatial consistency compares whether different products have the same land cover type at the same location through pixel-by-pixel spatial overlay analysis [16]. The number of matching pixels indicates spatial consistency between products, implying the degree of similarity [17]. However, it is important to note that the method primarily provides qualitative assessments and may not offer a definitive ranking of product quality. The methods for assessing the accuracy of LC products described above can be summarized in Figure 1 based on their distinctive characteristics and developmental trajectories. As can be seen from the figure, the full-sampling method is rarely used due to its high cost. Although direct assessment methods and indirect assessment methods are cost-effective, they result in the loss of a significant amount of information accuracy.



Figure 1. The characteristics and development history of accuracy assessment methods for LC products.

This paper argues that the accuracy information determined by LC products can be divided into two types: spatial position information and global feature information. Most studies that assess LC product accuracy have only focused on spatial position information, which means the classification correctness of each pixel, disregarding the distribution of features formed by all pixels [18,19]. Every land cover type exhibits distinctive morphologic characteristics, which are expressed through the arrangement of pixels. Therefore, even with the absolute ground truth, the methods depicted in Figure 1 are still unable to fully exploit the accuracy information of products and the ground truth for a comprehensive assessment. This paper extends the research of Tan et al. [20], which primarily took account of the global features of pixels to quantitatively compare the similarity of remote sensing products. However, the similarity index proposed in Tan's study also only considers a

portion of the accuracy information in the products and may be challenging to explain in some application scenarios. In this paper, we attempt to utilize more accuracy information to assess LC products and ensure that the new similarity index maintains strict mathematical logic and interpretability, revealing both commonalities and differences among products. The new metric will assist researchers in making optimal selections quickly from LC products.

The primary innovations and contributions of this paper include:

- Proposal of a relative accuracy assessment method for LC products based on optimal transport theory. The method considers a broader range of accuracy information to compute the Wasserstein similarity (WenSiM) between the target product and the reference ground truth at a global level without the need for registration and validation datasets.
- Evaluation of the classification accuracy of four LC products in Beijing. Through a comparative analysis of WenSiM and the results obtained from mainstream approaches, the effectiveness of the relative accuracy assessment method is validated.

#### 2. Data

## 2.1. Data Preprocessing

This study made use of Dynamic World [21], ESA WorldCover v100 [22], FROM\_GLC10 [23], and SinoLC\_1 [24] in Beijing to conduct accuracy assessments. ESRI World Cover [25] was chosen as the reference ground truth because of the highest overall official validation accuracy. Table 1 shows the basic information for every LC product.

Table 1. Basic information of the five LC products.

	ESRI_GLC10	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_1
Institution	Esri	WRI	ESA	THU	WHU
Resolution (m)	10	10	10	10	1
Coverage	Global	Global	Global	Global	National (China)
Classification scheme	10 classes	9 classes	11 classes	10 classes	11 classes
Version and timeline	2020	2020	2020	2017	2020
Overall accuracy (%)	85.96	77.80	74.40	72.76	73.61 55.55 (Beijing)

Notes: ESRI = Environmental Systems Research Institute; WRI = World Resources Institute; ESA = European Space Agency; THU = Tsinghua University; WHU = Wuhan University.

In contrast to common data preprocessing procedures, this study only dealt with resampling and reclassifying the products without requiring georeferencing. According to Table 1, only the SinoLC\_1 product had a resolution of 1 m. To facilitate a more comprehensive comparison of each product's accuracy and ensure fairness in the process, we standardized the resolutions of the five products to 10 m. It is also important to emphasize that, although the FROM\_GLC10 was produced in 2017, existing research has demonstrated that, compared to classification errors in LC products, land cover changes in large regions within a five-year timeframe can be negligible [26,27]. Additionally, for the DW\_GLC10, as a continuously updated image collection, the study utilized the confusion matrix results obtained during the testing process as its overall accuracy. Finally, the land cover categories for products were reclassified into forestland, cropland, grassland, wetland, settlements, and other land according to the classification criteria and details provided in Tables A1 and A2.

### 2.2. Study Area

The study area was Beijing (Lat 39°56′N, Long 116°20′E), which covers an area of 1,641,054 hectares and is renowned as both an ancient capital and a modern international city. The city is surrounded by mountains on its western, northern, and northeastern



sides, whereas the southeastern part consists of a plain that slopes toward the Bohai Sea. Detailed geographic information about Beijing, as well as the visual comparative effects of LC products in Changping District, are depicted in Figure 2.

**Figure 2.** (**a**) Geographical location information of Beijing. (**b**) Depictions of the visual comparison of Changping District.

## 3. Methods

The section begins by highlighting the theoretical underpinnings of the relative accuracy assessment method. Following that, the section elaborates on the characteristics and computational process of the WenSiM index. Finally, the section introduces the creation of a validation dataset for the study area by using stratified sampling theory.

### 3.1. Wasserstein-p Distance

The Wasserstein distance (earth mover's distance, EMD) is a type of histogram similarity measure [28,29]. Compared to other similarity measures, like the Kullback–Leibler (KL) divergence and the Jensen–Shannon (JS) divergence, the Wasserstein distance has the advantage of being able to assess the similarity between two probability distributions that do not overlap at all. In addition, its value represents the minimum "cost" when two distributions are transformed into each other.

Let  $\phi$  be a metric space. For this study, we considered  $\phi$  to be compact d-dimensional Euclidean spaces, i.e.,  $\phi = [0, 1]^d$ . Let  $\phi_{\mu}$  and  $\phi_{\nu}$  denote the set of Borel probability measures defined on  $\phi$ . The Wasserstein-p distance for  $p \in [1, \infty)$  between two distributions  $P_{\mu} \in \phi_{\mu}$  and  $P_{\nu} \in \phi_{\nu}$  can be defined as Equation (1) with the cost functions  $c(x, y) = d^p(x, y)$ .

$$W_p(P_{\mu}, P_{\nu}) \triangleq \left\{ \inf_{\gamma \in \Gamma(P_{\mu}, P_{\nu})} \int_{\phi_{\mu} \times \phi_{\nu}} d^p(x, y) d\gamma(x, y) \right\}^{\frac{1}{p}}$$
(1)

Here,  $\Gamma(P_{\mu}, P_{\nu})$  is the set of all transportation plans  $\gamma(x, y)$  whose marginals are  $P_{\mu}$  and  $P_{\nu}$ , respectively,  $\gamma \in \Gamma(P_{\mu}, P_{\nu})$ . Additionally,  $d^{p}(x, y)$  represents the "cost" associated with transforming x in probability distribution  $P_{\mu}$  into y in  $P_{\nu}$ , and p is the dimension size of the probability distribution. Thus, the Wasserstein p-distance reflects the "cost" of the optimal transportation plan.

$$W_p(P_\mu, P_\nu) = \left\{ \inf_{f \in MP(P_\mu, P_\nu)} \int_{\phi_\mu} d^p(f(x), x) dP_\mu(x) \right\}^{\frac{1}{p}}$$
(2)

where  $MP(P_{\mu}, P_{\nu}) = \left\{ f : \phi_{\mu} \to \phi_{\nu} \middle| f_{\#P_{\mu}} = P_{\nu} \right\}$  and  $f_{\#P_{\mu}}$  is used to indicate the pushforward of measure  $P_{\mu}$ .

Since LC products consist of pixels, they belong to a special two-dimensional discrete probability distribution and can be calculated numerically using a matrix. Thus, we focused on how to calculate the Wasserstein distance of absolute discrete probability measures in various dimensions.

The process of calculating higher-dimensional Wasserstein distance is difficult and complicated. But the development of the sliced Wasserstein distance provides a closed-form solution to the Wasserstein-p distance calculation problem [31]. By reducing the probability distribution to one dimension through random projection, computing the Wasserstein distance becomes fast and straightforward [32,33].

The idea behind the sliced Wasserstein distance is to obtain the marginal distribution family (i.e., one-dimensional distribution) of high-dimensional probability distribution through linear random projection and then calculate the Wasserstein distance of two marginal distributions. The method transforms the challenging high-dimensional optimal transport problem into several one-dimensional optimal transport problems with closed-form solutions. Then, the sliced Wasserstein-p distance between two probability distributions  $P_{\mu}$  and  $P_{\nu}$  can be defined as in Equation (3).

$$\widetilde{W}_{p}(P_{\mu}, P_{\nu}) = \left[ \int_{\theta \in \Omega} W_{p}^{p} \left( P_{\mu}^{\theta}, P_{\nu}^{\theta} \right) d\theta \right]^{\frac{1}{p}}$$
(3)

where  $P_{\mu}^{\theta}$  and  $P_{\nu}^{\theta}$  denote all projections of marginal distributions  $P_{\mu}$  and  $P_{\nu}$  on the direction  $\theta$ , respectively, and  $\Omega$  is the set of all possible directions on the unit sphere. It has also been proven that  $\widetilde{W}_{p}$  satisfies sub-additivity and coincidence axioms, making it a genuine metric [34].

Due to the various direction selections, calculating the corresponding projections during the actual process is still very complicated. To address the challenge of projection complexity while keeping the sample complexity constant, Deshpande et al. [35] proposed a method to first find the most meaningful projection direction. The result obtained by Equation (4) in the projection direction  $\theta$  is known as the max sliced Wasserstein distance.

$$\widetilde{W}_{2-max}(P_{\mu}, P_{\nu}) = \left[\max_{\theta \in \Omega} W_2^2 \left(P_{\mu}^{\theta}, P_{\nu}^{\theta}\right)\right]^{\frac{1}{2}}$$
(4)

 $W_{2-max}$  is an effective measure that overcomes the limitations of projection complexity. It can directly compute the difference between two-dimensional distributions. Therefore, in this study, it was employed to calculate the Wasserstein distance between LC products and the reference ground truth within the study area.

#### 3.2. Wasserstein Similarity Index

This study divided the accuracy information carried by the LC products into two types: global feature information and spatial position information. Although the Wasserstein distance is effective in quantifying the similarity between products and the reference truth from a global feature perspective, its interpretability is affected because the total weight is evenly distributed among all pixels. Therefore, we normalized  $W_{2-max}$  to ensure that the global feature similarity  $\mathcal{F}$  had a strict mathematical significance and was independent of the size of the study area.

The existence of target product A and reference truth B was assumed, whose matrix spaces are denoted as  $A_{\mu}$  and  $B_{\nu}$ , respectively. In the study area, there was a total of *k* land cover types, and the transformed numerical matrices satisfied  $P_{\mu} \in A_{\mu}$  and  $P_{\nu} \in B_{\nu}$ . Then,  $\mathcal{F}$  can be defined as Equation (5).

$$F_{i} = 1 - \frac{\widetilde{W}_{2-max}^{(i)}}{D^{(i)}}$$
(5)

where  $\mathcal{F}_i$  represents the global feature similarity for the type  $i(i = 1, 2, \dots, k)$  in the study area, with  $0 < \mathcal{F}_i \leq 1$ .  $D^{(i)}$  denoting the maximum centroid distance for the pixels of the land cover type *i* in A and B, which can be defined with Equation (6).

$$D^{(i)} = max \left\{ d\left(n_A^{(i)}, n_B^{(i)}\right) \right\}$$
(6)

where  $n_A^{(i)}$  and  $n_B^{(i)}$  represent the pixels of the land cover type *i* in A and B, respectively. *d* denotes the centroid distance for the pixels of land cover type *i* in A and B. When the pixels in the two products are distributed in the diagonal corners of the area, the maximum centroid distance can be computed.

After considering the global feature information of the products comprehensively, there is also a need for an evaluation metric to assess spatial positional information. This study used the correlation coefficient to quantify the spatial position similarity between the target product and the reference truth. The correlation coefficient *K* can be defined with Equation (7), primarily reflecting the commonalities and differences in category composition between the product and the truth [36].

$$K = \frac{\sum_{i=1}^{k} \left( P_{\mu}^{(i)} - \overline{P}_{\mu} \right) \left( P_{\nu}^{(i)} - \overline{P}_{\nu} \right)}{\sqrt{\sum_{i=1}^{k} \left( P_{\mu}^{(i)} - \overline{P}_{\mu} \right)^{2} \sum_{i=1}^{k} \left( P_{\nu}^{(i)} - \overline{P}_{\nu} \right)^{2}}}$$
(7)

Here,  $P_{\mu}^{(i)}$  and  $P_{\nu}^{(i)}$  represent the areas of type  $i(i = 1, 2, \dots, k)$  for A and B, respectively.  $\overline{P}_{\mu}$  and  $\overline{P}_{\nu}$  are the average areas of each type in A and B, respectively within the study area.

With  $\mathcal{F}$  and K serving as measures for different types of accuracy information, this study further introduced the Wasserstein similarity (WenSiM) between the target product A and reference truth B. The index comprehensively extracts accuracy information to reflect the overall similarity, and it is defined as follows:

$$WenSiM = \frac{1}{k} \sum_{i=1}^{k} K \mathcal{F}_i$$
(8)

where the *WenSiM* satisfies the condition  $0 < WenSiM \le 1$  and WenSiM = 1 only when  $K = \mathcal{F}_i = 1$ , indicating that the target product and the reference truth are entirely identical in both spatial position and global features within the study area.

The WenSiM index fully takes use of more accuracy information to evaluate LC products. While maintaining properties such as symmetry, non-negativity, and identity, the WenSiM also possesses strict mathematical significance and interpretability. It is an excellent metric for accurately and rapidly assessing LC product accuracy. The detailed calculation process is summarized in Algorithm 1.

**Algorithm 1.**  $WenSiM(P_{\mu}, P_{\nu}, k, \theta, \Omega)$ 

1: Initialize  $\theta$ ,  $\Omega$ 2: Data preprocessing for  $(P_{\mu}, P_{\nu})$ 3: Repeat 4:  $(A^{(i)}, B^{(i)}, P_{\mu}^{(i)}, P_{\nu}^{(i)}, n_A^{(i)}, n_B^{(i)}) \leftarrow (P_{\mu}, P_{\nu})$ 5:  $D^{(i)}, K \leftarrow (A^{(i)}, B^{(i)}, n_A^{(i)}, n_B^{(i)})$ 6:  $\widetilde{W}_{2-max}^{(i)} \leftarrow (P_{\mu}^{(i)}, P_{\nu}^{(i)}, \theta, \Omega)$ 7:  $\mathcal{F}^{(i)} \leftarrow (\widetilde{W}_{2-max}^{(i)}, D)$ 8:  $WenSiM^{(i)} \leftarrow (K, \mathcal{F}^{(i)})$ 9:  $i \leftarrow i + 1$ 10: Until i > k11:  $WenSiM \leftarrow (WenSiM^{(i)}, k)$ 12: Output WenSiM

## 3.3. Production of Validation Dataset

The evaluation method proposed in this paper can directly assess the accuracy of LC products without registration and validation datasets. In order to demonstrate the effectiveness and rationality of the WenSiM index, it is necessary to create a validation dataset to analyze the consistency between validation accuracy and assessment results.

Sampling is a crucial step in creating a validation dataset, as the sample size and sample variance determine the accuracy of the dataset. Although increasing the sample size is beneficial for improving accuracy, it also comes with the added cost of production. The optimal sampling scheme aims to obtain the most reliable validation results with the lowest cost [37]. Therefore, once the sample size is determined, the key to improving the accuracy lies in reducing sample variance [38].

Stratified sampling (SS) is a great method to reduce variance and improve sampling accuracy when the sample size is constant [39,40]. To implement SS, it is crucial to confirm that every object in the sampling range shares similar properties. Subsequently, researchers can use a particular characteristic or rule to divide the population into L non-repeating sub-groups, each referred to as a layer. Finally, samples for each layer can be acquired through random sampling.

LC products possess both geographical and administrative attributes, allowing them to be divided into multiple levels using administrative boundaries. Therefore, researchers can divide the study area into *L* layers based on the characteristic. After getting sampling points through SS, an unbiased estimate of the variance  $V(y_{ss})$  of all samples can be calculated by using Equation (9).

$$(y_{ss}) = \sum_{l=1}^{L} W_l^2 \cdot V(y_l) = \sum_{l=1}^{L} W_l^2 \cdot \frac{1 - f_l}{m_l} S_l^2 = \sum_{l=1}^{L} \frac{W_l^2 S_l^2}{m_l} - \sum_{l=1}^{L} \frac{W_l S_l^2}{M_{ss}}$$
(9)

where  $y_{ss}$  represents all samples, and  $M_{ss}$  is the total population size.  $W_l = \frac{M_l}{M_{ss}}$  and  $f_l = \frac{m_l}{M_l}$  are the weight parameter and the sampling ratio of layer l (l = 1, 2, ..., L), respectively.  $m_l$  and  $M_l$  denote the number of samples and pixels in layer l, respectively;  $y_l$  is the sample mean of layer l; and the population variance of layer l can be written as  $S_l^2$ .

To minimize the objective function  $V(y_{ss})$ , it is essential to choose an appropriate sample size and allocation method. This study adopted the Neyman optimal allocation method, which maximizes sampling accuracy by achieving a proportional distribution between  $m_l$  and  $W_lS_l$ .  $V_{min}(y_{ss})$  can be computed by Equation (10), where  $m_{ss}$  is used to indicate the total number of samples.

$$V_{min}(y_{ss}) = \frac{1}{m_{ss}} \left( \sum_{l=1}^{L} W_l S_l \right)^2 - \frac{1}{M_{ss}} \sum_{l=1}^{L} W_l S_l^2$$
(10)

The production method is based on SS theory and considers factors such as economic cost and sample quality, which involves scientifically selecting sample points to ensure that the accuracy meets research requirements. However, it is important to emphasize that the sample points selected through SS only reflect the spatial position information of the pixels and do not effectively reveal the global features formed by the pixels' interactions.

## 4. Results

This section is divided into two parts: the computation results of the Wasserstein similarity and the validation results obtained through the validation dataset within the study area. Ultimately, the section computes the confusion matrix to obtain the validation accuracy of each product and compares the results with the WenSiM and official validation accuracy.

## 4.1. Wasserstein Similarity Results

On completing data preprocessing, the product's classification scheme was modified to the IPCC format, categorizing the types into forestland, cropland, grassland, wetlands, settlements, and other land. Since the absolute ground truth did not exist, the paper selected ESRI\_GLC10, with the highest overall official validation accuracy from Table 1, as the reference truth. Accuracy assessment of DW\_GLC10, ESA\_GLC10, FROM\_GLC10, and SinoLC\_10 was performed by computing the WenSiM between the remaining products and the reference truth. The results for global feature similarity and correlation coefficients are shown in Tables 2 and 3, respectively.

Table 2. The global	feature similarity betw	een four LC products	and ESRI_GLC10 in Beijing
			, , ,

Types	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_10
FL (%)	98.05	97.57	97.17	93.00
CL (%)	96.37	98.66	96.51	96.15
GL (%)	85.86	85.06	93.09	75.55
WL (%)	98.86	95.45	99.10	88.26
SL (%)	98.45	98.72	96.40	97.95
OL (%)	96.02	93.52	84.06	77.89

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land. The bold results represent the maximum values of all products.

Table 3. The correlation coefficient between four LC products and ESRI\_GLC10 in Beijing.

	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_10			
ESRI_GLC10	0.9985	0.9596	0.9611	0.9925			
The held results represent the maximum values							

The bold results represent the maximum values

What stands out in the tables is that DW\_GLC10 exhibited the strongest correlation with the reference truth, with the highest global feature similarity in forestland and other land of 98.05% and 96.02%, respectively. It also can be seen from the data in Table 2 that ESA\_GLC10 had the most similar global features with the reference truth in cropland and settlements, with a value of 98.66%, and in wetlands, at 98.72%. The max similarity results for grassland of 93.09% and wetlands of 99.10% were from FROM\_GLC10.

Table 4 displays the WenSiM results obtained from the preliminary analysis of global feature similarity and correlation coefficients. Following the results, researchers can evalu-

ate the product accuracy at different levels, such as for each land cover type or the overall product, to meet the various requirements. From the data in Table 4, it is apparent that FROM\_GLC10 exhibited the highest WenSiM to the reference ground truth in grassland, with a value of 89.47%. The single most striking observation to emerge from the results is that DW\_GLC10 was the most similar to ESA\_GLC10 in all aspects except grassland. In summary, for the overall WenSiM results in Table 4, the products can be ranked in order of the overall WenSiM results as follows: DW\_GLC10 > ESA\_GLC10 > FROM\_GLC10 > SinoLC\_10.

Types	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_10
FL (%)	97.90	93.63	93.40	92.31
CL (%)	96.23	94.68	92.76	95.43
GL (%)	85.73	81.63	89.47	74.99
WL (%)	98.71	91.60	95.25	87.60
SL (%)	98.30	94.73	92.65	97.22
OL (%)	95.88	89.74	80.80	77.31
OS (%)	95.46	91.00	90.72	87.48

Table 4. The WenSiM results between four LC products and ESRI\_GLC10 in Beijing.

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land; OS = overall WenSiM. The bold results represent the maximum values of all products.

## 4.2. Validation Results

Following the SS theory outlined in Section 3.3, the study initially divided Beijing into 16 layers, corresponding to each administrative district. Subsequently, a total of 2001 samples were randomly selected within the study area. Table A3 shows the sampling results and computational parameters for each district. Table A4 displays the sample numbers for each land cover type in each layer. Figure 3 illustrates the visual distribution of districts in Beijing, as well as the distribution of sample points in Changping and Chaoyang districts.



**Figure 3.** (a) Visual distribution of districts in Beijing. (b) Distribution of the samples in Changping District. (c) Distribution of the samples in Chaoyang District.

In the follow-up phase of the study, multiple experts were invited to determine the land cover types of the sample points based on Google Earth and Sentinel-2 high-resolution remote sensing images. For samples with disputed categorizations, experts engaged in collective discussions and voting to assign the type with the highest number of votes, ensuring that the accuracy of the validation dataset met the task requirements. Figure 4 below illustrates the number of samples for each land cover type in the validation dataset.



Figure 4. Sample numbers for each land cover type in the validation dataset.

After the validation dataset was created, the confusion matrices were computed and the quantitative metrics for evaluating the product performance were provided, including user accuracy (U.A.), producer accuracy (P.A.), overall accuracy (O.A.), and the kappa coefficient. The results of the confusion matrices for each product can be found in Tables A5–A9. For clarity, this paper refers to the accuracy computed from the validation dataset as the validation accuracy (V.A.), whereas the accuracy reported by the production agencies is referred to as the official validation accuracy (O.V.A.). Table 5 displays the V.A. and O.V.A. for every product in the study area.

Accuracy (%)	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_10
V.A. (FL)	77.96	69.09	76.39	71.91
V.A. (CL)	27.78	43.33	34.44	28.89
V.A. (GL)	0.40	26.59	4.37	1.19
V.A. (WL)	77.97	61.01	54.23	15.25
V.A. (SL)	95.18	76.53	69.77	85.21
V.A. (OL)	6.10	14.63	1.22	0.85
V.A. (O.A.)	65.67	61.23	60.67	58.52
O.V.A. (O.A.)	77.80	75.00	72.76	55.55

Table 5. The V.A. and O.V.A. of four LC products in Beijing.

Note: V.A. = validation accuracy; FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land; O.A. = overall accuracy; V.O.A. = official validation accuracy. The bold results represent the maximum values of all products.

The most surprising aspect of Table 5 is that DW\_GLC10 and ESA\_GLC10 achieved the highest validation accuracy across all land cover types. As shown in the table, the peak validation accuracy for forestland of 77.96%, wetlands of 77.97%, and settlements of 95.18% came from DW\_GLC10. ESA\_GLC10 had the highest validation accuracy for cropland of 43.33%, grassland of 26.59%, and other land of 14.63%. Additionally, DW\_GLC10 had the maximum overall V.A. of 65.67% and O.V.A. of 77.80%. To further demonstrate the



superiority of the WenSiM metric, Figures 5 and 6 present a comparison of the WenSiM, V.A., and O.V.A. results.

**Figure 5.** Comparison between the WenSiM and V.A. results for every land cover type. (**a**) Forestland. (**b**) Cropland. (**c**) Grassland. (**d**) Wetlands. (**e**) Settlements. (**f**) Other land.



Figure 6. Comparison of the overall WenSiM, V.A., and O.V.A. results for LC products.

Looking at the figures, it is apparent that the WenSiM consistently had higher values compared to the V.A. and O.V.A. Figure 5 shows that the WenSiM exhibited similar trends to the V.A. for forestland, wetlands, and settlements, whereas it differed significantly for cropland, grassland, and other land categories. What is interesting about the trend in Figure 6 is that the overall WenSiM, V.A. and O.V.A. displayed the same accuracy assessment results: DW\_GLC10 > ESA\_GLC10 > FROM\_GLC10 > SinoLC\_10.

#### 5. Discussion

This section initially delves into latent information within the results of the WenSiM and validation. Through a comparison with mainstream methods, the section further analyzes the strengths and weaknesses of the relative accuracy assessment method based on optimal transport theory. Finally, the section provides a reasonable accuracy assessment and ranking of LC products within the study area.

#### 5.1. Analysis and Comparison of Assessment Results

Due to the fact that both the relative accuracy assessment method and the direct assessment method require the reference ground truth for accuracy evaluation, it is imperative to engage in a discussion about the scientific basis for the selection standards. Firstly, a number of recent studies have already demonstrated the rationality of using a validation dataset as the reference ground truth [41–43]. Although the V.A. and O.V.A. do not fully exploit all accuracy information, they are still considered the primary indicators for effectively assessing LC product accuracy [44]. Therefore, it is scientific to choose O.V.A. as the standard for selecting the reference ground truth. Moreover, the results of the V.A. and O.V.A. and O.V.A. can also serve as important references for evaluating the WenSiM index.

As shown in Figure 5, nearly all quantitative results of the validation accuracy were lower than those of the WenSiM, which is because the WenSiM results encompass more accuracy information about the products. Figure 6 demonstrates that the overall WenSiM, V.A., and O.V.A. results had consistent trends, further confirming the effectiveness of the WenSiM index. The reason for the low V.A. of grassland and other land was attributed to the limitations of direct evaluation method.

The 10 m resolution LC product had hundreds of millions of pixels in Beijing. Considering cost constraints, it was not feasible to validate the true land cover type for each pixel. Despite the fact that the sample proportion chosen in this study surpassed that of other validation datasets, it remained insufficient for a region of this magnitude. The situation had the potential to create imbalanced sampling data ratios among different classes, ultimately leading to an unreasonable accuracy assessment [45].

On the one hand, when there is an imbalance in land cover composition within a study area, the direct evaluation method, with random sampling, will inevitably result in insufficient or even zero samples for certain land cover categories. This phenomenon may lead to unstable validation accuracy for those land cover types. On the other hand, the direct evaluation method only considers the spatial position information of sample points to quantify product accuracy, which can unavoidably yield lower results.

The indirect evaluation method achieves cost-effective utilization of spatial position information for all pixels by comparing the consistency between products [46]. The study categorized spatial consistency into five levels from high to low. Taking forestland as an example, level 5 indicates that the classification results of all LC products at that pixel were forest land, whereas level 1 means that only one LC product classified the pixel as forestland. Figures 7 and 8 present the regional consistency analysis results and the spatial consistency analysis results, respectively, for the five LC products.



Figure 7. Regional consistency analysis of LC products.



Figure 8. Spatial consistency analysis of LC products.

Looking at Figure 7, it is apparent that there was a significant difference for each product in grassland and other land, whereas the differences are relatively small in the forestland and wetlands. Interestingly, a noticeable grouping effect was observed in the regional consistency in cropland and settlements. Except for FROM\_GLC10, the products exhibited stronger consistency in cropland. The consistency results of settlements in ESRI\_GLC10, ESA\_GLC10, and SinoLC\_10 were similar, whereas DW\_GLC10 was more consistent with FROM\_GLC10.

Figure 8 shows that the spatial consistency analysis not only revealed the distribution of land cover types in the study area but also indirectly assessed the classification accuracy of the products. For instance, the analysis results indicate that forestland was mainly distributed in the northern and southwestern parts of Beijing. Through visual interpretation,

SinoLC\_10 exhibited the highest spatial consistency in forestland, which suggests that SinoLC\_10 had a higher classification accuracy in forestland.

The indirect evaluation method can effectively mitigate issues arising from low samples, preventing insufficient utilization of spatial location information in products. However, the method lacks on-site validation and the ability to quantify product accuracy, providing only qualitative suggestions for product selection [47].

## 5.2. Advantages and Limitations of Wasserstein Similarity

Land cover types often exhibit anomalous morphologies, manifesting as anomalous pixels. Anomalous pixels carry accuracy information that is markedly different from other pixels of the same type, representing unique features. The causes of anomalous pixels can be attributed to three main factors.

- 1. The random distribution of the actual class;
- 2. Errors introduced during development;
- 3. Misclassification resulting from insufficient accuracy in the classification algorithms.

Therefore, anomalous pixels primarily reflect the commonalities and differences between products. The direct evaluation method, due to the randomness of sampling, cannot effectively utilize the accuracy information of anomalous pixels. Although the indirect evaluation method considers the spatial position information of anomalous pixels, it falls short in quantifying the accuracy information. This paper provides the first method to take into account various types of accuracy information, with a specific emphasis on anomalous pixels, to quantify product accuracy globally.

The advantages of Wasserstein similarity are primarily evident in two application scenarios. Firstly, when the absolute ground truth is available, serving as a reference for LC products, the method reasonably evaluates the products by thoroughly exploring their accuracy information. This results in more comprehensive findings compared to other approaches. Secondly, when the absolute ground truth is unattainable, WenSiM enables a cost-effective evaluation of new LC products. Unlike direct evaluation methods, the method selects the product with the highest overall accuracy as the reference truth for computing WenSiM, eliminating the need for a validation dataset. In contrast to indirect evaluation methods, it allows for a quantitative assessment of product accuracy and provides a clear ranking of accuracy for each product at different levels.

However, the method also has some limitations. Firstly, the presence of anomalous pixels can make the evaluation results overly stringent. Secondly, when performing assessments on an extensive scale, such as at the global or super-regional level, the large number of pixels can lead to slower computation speeds. Furthermore, in the absence of the absolute ground truth, the evaluation results are influenced by the choice of reference truth. If the criteria are chosen arbitrarily, the reliability of the results may significantly decrease. Therefore, the criteria for selecting the reference truth should be flexibly determined in applications. In the future, the method could be expanded for the accuracy assessment of a broader range of remote sensing products, such as land surface temperature and digital elevation models, thereby aiding researchers in rapidly selecting the most suitable products based on task requirements [48,49].

After elucidating the characteristics of the accuracy evaluation method based on optimal transportation theory, the final evaluation ranking was obtained by combining the WenSiM with the validation results: DW\_GLC10 > ESA\_GLC10 > FROM\_GLC10 > SinoLC\_10.

#### 6. Conclusions

The paper set out to find a relative accuracy assessment method based on optimal transport theory for LC products and demonstrates its effectiveness through a research case. Three key conclusions are drawn from the paper. First, the method enables a rapid quantification of LC product accuracy without the need for registration and validation datasets. Second, the WenSiM index derived from the method measures the similarity between the target product and the reference truth at a global scale, utilizing a broader

range of accuracy information and resulting in more accurate assessment. Third, this study evaluated the accuracy of four LC products in Beijing, with results closely aligning with validation outcomes. Although the method is limited by the small number of cases, it possesses universality and can be extended to the accuracy assessment of various remote sensing products in global regions. This would be a fruitful area for further work, with the goal of providing researchers with a fast and reliable accuracy assessment tool for the selection and production of remote sensing products.

**Author Contributions:** Conceptualization, R.Z. and Y.T.; methodology, R.Z., Y.T., Y.S. and G.J.; software, R.Z. and X.W.; validation, R.Z., Z.L. and J.W.; formal analysis, R.Z.; investigation, R.Z.; resources, R.Z.; data curation, R.Z.; writing—original draft preparation, R.Z.; writing—review and editing, and Y.T.; visualization, R.Z.; supervision, Y.T. and G.J.; project administration, X.W.; funding acquisition, Y.T. and G.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was co-funded by Special Fund for the follow-up work of the Three Gorges project (grant number 102126222020270019081) and the National Key R&D Program (grant number 2019YFE0126400).

Data Availability Statement: ESRI\_GLC10 comes from the website https://livingatlas.arcgis.com/landcover/ (accessed on 25 November 2023). DW\_GLC10 comes from the website https://dynamicworld. app/ (accessed on 25 November 2023). ESA\_GLC10 comes from the website https://esa-worldcover. org/en (accessed on 25 November 2023). FROM\_GLC10 comes from the website https://data-starcloud.pcl.ac.cn/resource/1 (accessed on 25 November 2023). SinoLC\_1 comes from the website https://ata-starcloud.pcl.ac.cn/resource/1 (accessed on 25 November 2023). Other data will be made available on request.

Acknowledgments: The authors thank all survey participants and reviewers of the paper. The data used for this paper were provided by ESRI\_GLC10 from ESRI, DW\_GLC10 from Google, ESA\_GLC10 from ESA, FROM\_GLC10 from Tsinghua University, and SinoLC\_1 from Wuhan University. We greatly appreciate the constructive comments from the anonymous reviewers and the *Remote Sensing* editorial team that helped us improve our paper.

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

## Appendix A

Table A1. The detailed definitions of IPCC land categories.

La	nd Categories	Detailed Definitions
	Forestland	<ul><li>All land with woody vegetation consistent with thresholds;</li><li>Systems with vegetation that are expected to exceed the threshold of the forestland.</li></ul>
	Cropland	<ul> <li>Arable and tillage land;</li> <li>Agro-forestry systems where vegetation falls below the thresholds.</li> </ul>
•	Grassland	<ul> <li>Rangelands and pastureland that is not considered cropland;</li> <li>Systems with vegetation that fall below the threshold and are not expected to exceed the threshold without human intervention;</li> <li>All grassland from wildlands to recreational areas, as well as agricultural and silvi-pastural systems.</li> </ul>
	Wetlands	<ul> <li>Land that is covered by or saturated with water for all or part of the year;</li> <li>Reservoirs, natural rivers and lakes.</li> </ul>
	Settlements	• All developed land, including transportation infrastructure and human settlements.
	Other land	• Bare soil, rock, ice, and all unmanaged land areas that do not fall into any of the other five categories.

	IPCC	ESRI_GLC10	DW_GLC10	ESA_GLC10	FROM_GLC10	SinoLC_1
		Trees		Tree cover	Forest	Tree cover
	Forestland		Trees	Shrubland		
		Scrub and shrub		Mangroves	Shrubland	Shrubland
	Cropland	Crops	Crops	Cropland	Cropland	Cropland
	Grassland	Grass	Grass	Grassland Grassland		Grassland
_	Water	Water	Permanent water bodies	Waterbodies	Water	
	Wetlands	Flooded vegetation Fl	Flooded vegetation	Herbaceous wetland	Wetlands	Wetlands
_	01			D 11	T	Roads
	Settlements	Built area	Built	Built up	Impervious	Built up
	Other land	Bare ground Bare		Bare and sparse vegetation	Barren land	Barren and spare vegetation
	Outer fallu	Snow and ice	Concernent inc	Snow and ice	Snow and ice	Snow and ice
		Clouds	Show and ice	Moss and Lichen	Tundra	Moss and lichen

Table A2. Conversion relations between five LC products and the IPCC land cover classification.

Table A3. Stratified sampling calculation parameters.

-

District	$M_l$	$\overline{y}_l$	$S_l^2$	W <sub>l</sub>	$W_l S_l$	$m_l$
Changping	23,067,539	2.408	1.323	0.082	0.09430	197
Chaoyang	7,925,627	4.401	0.363	0.028	0.01697	35
Daxing	17,461,322	3.221	0.720	0.062	0.05264	110
Dongcheng	713,476	4.879	0.092	0.003	0.00077	2
Fangshan	33,808,138	2.095	1.284	0.120	0.13612	285
Fengtai	5,195,318	4.264	0.478	0.018	0.01276	27
Haidian	7,322,717	3.972	0.679	0.026	0.02145	45
Huairou	36,841,342	1.417	0.896	0.131	0.12396	259
Mentougou	24,734,083	1.269	0.764	0.088	0.07681	161
Miyun	38,531,881	1.742	1.065	0.137	0.14133	295
Pinggu	16,263,416	1.998	1.140	0.058	0.06171	129
Shijingshan	1,436,001	3.957cc	0.763	0.005	0.00446	9
Shunyi	17,292,399	3.288	0.758	0.061	0.05349	112
Tongzhou	15,341,172	3.359	0.695	0.055	0.04544	95
Xicheng	856,827	4.965	0.012	0.003	0.00034	1
Yanqing	34,618,782	1.576	0.861	0.123	0.11415	239

Notes:  $M_l$  = the number of pixels in layer l;  $\overline{y}_l$  = the sample mean of layer l;  $S_l^2$  = the population variance of layer l;  $W_l$  = the weight parameter of layer l;  $m_l$  = the number of samples of layer l.

Table A4. Sample numbers for each land cover type in each layer.

District	FL	CL	GL	WL	SL	OL
Changping	125	5	20	3	37	7
Chaoyang	6	1	3	2	22	1
Daxing	38	5	30	0	36	1
Dongcheng	0	0	0	0	2	0
Fangshan	185	4	33	3	54	6
Fengtai	5	2	0	0	17	3
Haidian	15	3	1	1	22	3
Huairou	210	12	15	1	14	7
Mentougou	139	10	4	0	2	6
Miyun	186	17	27	24	20	21

District	FL	CL	GL	WL	SL	OL
Pinggu	73	1	33	9	11	2
Shijingshan	5	0	0	1	3	0
Shunyi	28	6	35	6	32	5
Tongzhou	28	7	24	5	27	4
Xicheng	0	0	0	0	1	0
Yanqing	164	17	27	4	11	16
Total	1207	90	252	59	311	82

Table A4. Cont.

Notes: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

Table A5. Confusion matrix for ESRI\_GLC10 according to the validation dataset.

Classification	FL	CL	GL	WL	SL	OL	Total	P.A. (%)
FL	983	127	0	2	94	1	1207	81.44
CL	51	25	0	1	13	0	90	27.78
GL	25	157	15	0	47	8	252	5.95
WL	7	4	1	42	5	0	59	71.19
SL	5	13	0	0	291	2	311	93.57
OL	35	14	0	4	27	2	82	2.44
Total	1106	340	16	49	477	13	2001	
U.A. (%)	88.88	7.35	93.75	85.71	61.01	15.38		
O.A. (%) Kappa	67.87 0.4816							

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

 Table A6. Confusion matrix for DW\_GLC10 according to the validation dataset.

Classification	FL	CL	GL	WL	SL	OL	Total	P.A. (%)
FL	941	131	42	2	89	2	1207	77.96
CL	40	25	5	1	16	3	90	27.78
GL	19	170	1	3	55	4	252	0.40
WL	3	3	0	46	6	1	59	77.97
SL	5	5	0	0	296	5	311	95.18
OL	29	11	0	4	33	5	82	6.10
Total	1037	345	48	56	495	20	2001	
U.A. (%)	90.74	7.25	2.08	82.14	59.80	25.00		
O.A. (%)	65.67							
Kappa	0.4610							

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

 Table A7. Confusion matrix for ESA\_GLC10 according to the validation dataset.

Classification	FL	CL	GL	WL	SL	OL	Total	P.A. (%)
FL	834	112	222	0	18	21	1207	69.10
CL	20	39	20	0	5	6	90	43.33
GL	52	107	67	1	9	16	252	26.59
WL	5	6	3	36	0	9	59	61.02
SL	30	20	1	0	238	22	311	76.53
OL	19	15	13	3	20	12	82	14.63
Total	960	299	326	40	290	86	2001	
U.A. (%)	86.88	13.04	20.55	90.00	82.07	13.95		
O.A. (%)	61.27							
Kappa	0.4118							

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

Classification	FL	CL	GL	WL	SL	OL	Total	P.A. (%)
FL	922	176	77	1	25	6	1207	76.39
CL	41	31	9	0	9	0	90	34.44
GL	34	182	11	1	24	0	252	4.36
WL	3	15	5	32	4	0	59	54.24
SL	10	55	25	0	217	4	311	69.77
OL	27	25	8	2	19	1	82	1.22
Total	1037	484	135	36	298	11	2001	
U.A. (%)	88.91	6.40	8.15	88.89	72.82	9.09		
O.A. (%) Kappa	60.67 0.3894							

**Table A8.** Confusion matrix for FROM\_GLC10 according to the validation dataset.

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

Table A9. Confusion matrix for SinoLC	_10 according to	the validation dataset
---------------------------------------	------------------	------------------------

Classification	FL	CL	GL	WL	SL	OL	Total	P.A. (%)
FL	868	130	123	3	83	0	1207	71.91
CL	47	26	6	0	11	0	90	28.89
GL	45	154	3	1	49	0	252	1.19
WL	30	13	0	9	7	0	59	15.25
SL	16	28	2	0	265	0	311	85.21
OL	32	19	3	0	28	0	82	0
Total	1038	370	137	13	443	0	2001	
U.A. (%)	83.62	7.03	2.19	69.23	59.82	0		
O.A. (%) Kappa	58.52 0.3474							

Note: FL = forestland; CL = cropland; GL = grassland; WL = wetlands; SL = settlements; OL = other land.

## References

- Song, X.P.; Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Tyukavina, A.; Vermote, E.F.; Townshend, J.R. Global land change from 1982 to 2016. *Nature* 2018, 560, 639–643. [CrossRef] [PubMed]
- 2. Stanimirova, R.; Graesser, J.; Olofsson, P.; Friedl, M.A. Widespread changes in 21st century vegetation cover in Argentina, Paraguay, and Uruguay. *Remote Sens. Environ.* 2022, 282, 113277. [CrossRef]
- Li, L.; Zhan, W.; Ju, W.; Peñuelas, J.; Zhu, Z.; Peng, S.; Zhu, X.; Liu, Z.; Zhou, Y.; Li, J.; et al. Competition between biogeochemical drivers and land-cover changes determines urban greening or browning. *Remote Sens. Environ.* 2023, 287, 113481. [CrossRef]
- Wang, J.; Yang, X.; Wang, Z.; Cheng, H.; Kang, J.; Tang, H.; Li, Y.; Bian, Z.; Bai, Z. Consistency Analysis and Accuracy Assessment of Three Global Ten-Meter Land Cover Products in Rocky Desertification Region—A Case Study of Southwest China. *ISPRS Int. J. Geo-Inf.* 2022, 11, 202. [CrossRef]
- 5. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131. [CrossRef]
- 6. Zhang, W.; Wang, J.; Lin, H.; Cong, M.; Wan, Y.; Zhang, J. Fusing Multiple Land Cover Products Based on Locally Estimated Map-Reference Cover Type Transition Probabilities. *Remote Sens.* **2023**, *15*, 481. [CrossRef]
- Comber, A.; Fisher, P.; Brunsdon, C.; Khmag, A. Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* 2012, 127, 237–246. [CrossRef]
- Chen, J.; Chen, L.; Chen, F.; Ban, Y.; Li, S.; Han, G.; Tong, X.; Liu, C.; Stamenova, V.; Stamenov, S. Collaborative validation of GlobeLand30: Methodology and practices. *Geo-Spat. Inf. Sci.* 2021, 24, 134–144. [CrossRef]
- 9. Wang, Y.; Zhang, J.; Liu, D.; Yang, W.; Zhang, W. Accuracy Assessment of GlobeLand30 2010 Land Cover over China Based on Geographically and Categorically Stratified Validation Sample Data. *Remote Sens.* **2018**, *10*, 1213. [CrossRef]
- Venter, Z.S.; Barton, D.N.; Chakraborty, T.; Simensen, T.; Singh, G. Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sens.* 2022, 14, 4101. [CrossRef]
- 11. Foody, G.M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sens. Environ.* **2020**, *239*, 111630. [CrossRef]
- 12. Bie, Q.; Shi, Y.; Li, X.; Wang, Y. Contrastive Analysis and Accuracy Assessment of Three Global 30 m Land Cover Maps Circa 2020 in Arid Land. *Sustainability* 2022, *15*, 741. [CrossRef]

- Pflugmacher, D.; Krankina, O.N.; Cohen, W.B.; Friedl, M.A.; Sulla-Menashe, D.; Kennedy, R.E.; Nelson, P.; Loboda, T.V.; Kuemmerle, T.; Dyukarev, E.; et al. Comparison and assessment of coarse resolution land cover maps for Northern Eurasia. *Remote Sens. Environ.* 2011, 115, 3539–3553. [CrossRef]
- 14. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy assessment of seven global land cover datasets over China. *ISPRS J. Photogramm. Remote Sens.* **2017**, 125, 156–173. [CrossRef]
- 15. Shi, W.; Zhao, X.; Zhao, J.; Zhao, S.; Guo, Y.; Liu, N.; Sun, N.; Du, X.; Sun, M. Reliability and consistency assessment of land cover products at macro and local scales in typical cities. *Int. J. Digit. Earth* **2023**, *16*, 486–508. [CrossRef]
- 16. Liu, P.; Pei, J.; Guo, H.; Tian, H.; Fang, H.; Wang, L. Evaluating the Accuracy and Spatial Agreement of Five Global Land Cover Datasets in the Ecologically Vulnerable South China Karst. *Remote Sens.* **2022**, *14*, 3090. [CrossRef]
- 17. Gao, Y.; Liu, L.; Zhang, X.; Chen, X.; Mi, J.; Xie, S. Consistency Analysis and Accuracy Assessment of Three Global 30-m Land-Cover Products over the European Union using the LUCAS Dataset. *Remote Sens.* **2020**, *12*, 3479. [CrossRef]
- 18. Ye, S.; Pontius, R.G.; Rakshit, R. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 137–147. [CrossRef]
- Kang, J.; Sui, L.; Yang, X.; Wang, Z.; Huang, C.; Wang, J. Spatial Pattern Consistency among Different Remote-Sensing Land Cover Datasets: A Case Study in Northern Laos. *ISPRS Int. J. Geo-Inf.* 2019, *8*, 201. [CrossRef]
- 20. Tan, Y.; Shi, Y.; Xu, L.; Zhou, K.; Jing, G.; Wang, X.; Bai, B. An Optimal Transport Based Global Similarity Index for Remote Sensing Products Comparison. *Remote Sens.* **2022**, *14*, 2546. [CrossRef]
- 21. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
- 22. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. ESA WorldCover 10 m 2020 v100 (Version v100) [dataset]. *Zenodo* 2021. [CrossRef]
- Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover. *Sci. Bull.* 2017, 64, 370–373. [CrossRef] [PubMed]
- 24. Li, Z.; He, W.; Cheng, M.; Hu, J.; Yang, G.; Zhang, H. SinoLC-1: The first 1-meter resolution national-scale land-cover map of China created with the deep learning framework and open-access data. *Earth Syst. Sci. Data.* **2023**, *15*, 4749–4780. [CrossRef]
- Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global land use/land cover with Sentinel 2 and deep learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4704–4707.
- 26. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]
- McCallum, I.; Obersteiner, M.; Nilsson, S.; Shvidenko, A. A spatial comparison of four satellite derived 1km global land cover datasets. *Int. J. Appl. Earth Obs. Geoinf.* 2006, *8*, 246–255. [CrossRef]
- Rubner, Y.; Tomasi, C.; Guibas, L.J. The Earth Mover's Distance as a Metric for Image Retrieval. Int. J. Comput. Vis. 2000, 40, 99–121. [CrossRef]
- 29. Villani, C. Optimal Transport: Old and New; Springer: Berlin/Heidelberg, Germany, 2009.
- 30. Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **1991**, *44*, 375–417. [CrossRef]
- Kolouri, S.; Rohde, G.K.; Hoffmann, H. Sliced Wasserstein Distance for Learning Gaussian Mixture Models. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3427–3436.
- 32. Deshpande, I.; Zhang, Z.; Schwing, A. Generative Modeling Using the Sliced Wasserstein Distance. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3483–3491.
- Chen, Y.; Li, C.; Lu, Z. Computing Wasserstein-p Distance between Images with Linear Cost. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 509–518.
- Kolouri, S.; Zou, Y.; Rohde, G.K. Sliced Wasserstein Kernels for Probability Distributions. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5258–5267.
- Deshpande, I.; Hu, Y.T.; Sun, R.; Pyrros, A.; Siddiqui, N.; Koyejo, S.; Zhao, Z.; Forsyth, D.; Schwing, A.G. Max-Sliced Wasserstein Distance and Its Use for GANs. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 10640–10648.
- 36. Wang, H.; Yan, H.; Hu, Y.; Xi, Y.; Yang, Y. Consistency and Accuracy of Four High-Resolution LULC Datasets—Indochina Peninsula Case Study. *Land* **2022**, *11*, 758. [CrossRef]
- 37. Stehman, S.V. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 2009, 30, 5243–5272. [CrossRef]
- 38. Liu, Y.; Shi, W.; Zhang, H.; Zhang, M. A multilevel stratified spatial sampling approach based on terrain knowledge for the quality assessment of OpenStreetMap dataset in Hong Kong. *Trans. GIS* **2023**, *27*, 290–318. [CrossRef]
- 39. Stehman, S.V.; Foody, G.M. Key issues in rigorous accuracy assessment of land cover products. *Remote Sens. Environ.* 2019, 231, 111199. [CrossRef]
- 40. Tsendbazar, N.E.; de Bruin, S.; Mora, B.; Schouten, L.; Herold, M. Comparative assessment of thematic accuracy of GLC maps for specific applications using existing reference data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *44*, 124–135. [CrossRef]

- 41. Wagner, J.E.; Stehman, S.V. Optimizing sample size allocation to strata for estimating area and map accuracy. *Remote Sens. Environ.* **2015**, *168*, 126–133. [CrossRef]
- 42. Wickham, J.D.; Stehman, S.V.; Gass, L.; Dewitz, J.; Fry, J.A.; Wade, T.G. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* **2013**, *130*, 294–304. [CrossRef]
- Gong, Y.; Xie, H.; Liao, S.; Lu, Y.; Jin, Y.; Wei, C.; Tong, X. Assessing the Accuracy of Multi-Temporal GlobeLand30 Products in China Using a Spatiotemporal Stratified Sampling Method. *Remote Sens.* 2023, 15, 4593. [CrossRef]
- Li, Z.; Chen, X.; Qi, J.; Xu, C.; An, J.; Chen, J. Accuracy assessment of land cover products in China from 2000 to 2020. *Sci. Rep.* 2023, 13, 12936. [CrossRef]
- 45. Morales-Barquero, L.; Lyons, M.B.; Phinn, S.R.; Roelfsema, C.M. Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources. *Remote Sens.* **2019**, *11*, 2305. [CrossRef]
- Zhao, T.; Zhang, X.; Gao, Y.; Mi, J.; Liu, W.; Wang, J.; Jiang, M.; Liu, L. Assessing the Accuracy and Consistency of Six Fine-Resolution Global Land Cover Products Using a Novel Stratified Random Sampling Validation Dataset. *Remote Sens.* 2023, 15, 2285. [CrossRef]
- Liu, B.; Yang, X.; Wang, Z.; Ding, Y.; Zhang, J.; Meng, D. A Comparison of Six Forest Mapping Products in Southeast Asia, Aided by Field Validation Data. *Remote Sens.* 2023, 15, 4584. [CrossRef]
- Polidori, L.; El Hage, M. Digital Elevation Model Quality Assessment Methods: A Critical Review. *Remote Sens.* 2020, 12, 3522.
   [CrossRef]
- 49. Liu, Y.; Yu, Y.; Yu, P.; Göttsche, F.M.; Trigo, I.F. Quality Assessment of S-NPP VIIRS Land Surface Temperature Product. *Remote Sens.* 2015, 7, 12215–12241. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.