



Article

Stability Analysis of Unmixing-Based Spatiotemporal Fusion Model: A Case of Land Surface Temperature Product Downscaling

Min Li ^{1,2,†}, Shanxin Guo ^{2,3,*}, Jinsong Chen ^{2,3}, Yuguang Chang ^{1,†}, Luyi Sun ^{2,3} , Longlong Zhao ^{2,3} , Xiaoli Li ^{2,3} and Hongming Yao ^{2,3}

¹ School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

² Center for Geo-Spatial Information, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

³ Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis and Application, Shenzhen 518055, China

* Correspondence: sx.guo@siat.ac.cn; Tel.: +86-755-8639-2331

† These authors contributed equally to this work.

Abstract: The unmixing-based spatiotemporal fusion model is one of the effective ways to solve limitations in temporal and spatial resolution tradeoffs in a single satellite sensor. By using fusion data from different satellite platforms, high resolution in both temporal and spatial domains can be produced. However, due to the ill-posed characteristic of the unmixing function, the model performance may vary due to the different model setups. The key factors affecting the model stability most and how to set up the unmixing strategy for data downscaling remain unknown. In this study, we use the multisource land surface temperature as the case and focus on the three major factors to analyze the stability of the unmixing-based fusion model: (1) the definition of the homogeneous change regions (HCRs), (2) the unmixing levels, and (3) the number of HCRs. The spatiotemporal data fusion model U-STFM was used as the baseline model. The results show: (1) The clustering-based algorithm is more suitable for detecting HCRs for unmixing. Compared with the multi-resolution segmentation algorithm and k-means algorithm, the ISODATA clustering algorithm can more accurately describe LST's temporal and spatial changes on HCRs. (2) For the U-STFM model, applying the unmixing processing at the change ratio level can significantly reduce the additive and multiplicative noise of the prediction. (3) There is a tradeoff effect between the number of HCRs and the solvability of the linear unmixing function. The larger the number of HCRs (less than the available MODIS pixels), the more stable the model is. (4) For the fusion of the daily 30 m scale LST product, compared with STARFM and ESTARFM, the modified U-STFM (iso_USTFM) achieved higher prediction accuracy and a lower error (R^2 : 0.87 and RMSE:1.09 k). With the findings of this study, daily fine-scale LST products can be predicted based on the unmixing-based spatial-temporal model with lower uncertainty and stable prediction.

Keywords: spatiotemporal fusion; unmixing; U-STFM; land surface temperature



Citation: Li, M.; Guo, S.; Chen, J.; Chang, Y.; Sun, L.; Zhao, L.; Li, X.; Yao, H. Stability Analysis of Unmixing-Based Spatiotemporal Fusion Model: A Case of Land Surface Temperature Product Downscaling. *Remote Sens.* **2023**, *15*, 901. <https://doi.org/10.3390/rs15040901>

Academic Editors: Jungho Im, Seonyoung Park and Cheolhee Yoo

Received: 26 December 2022

Revised: 26 January 2023

Accepted: 3 February 2023

Published: 6 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a key physical parameter of land surface, land surface temperature (LST) provides essential information on spatial-temporal changes in surface energy and water balance at local or global scales [1–3]. LST is now a key input of the physical model for evapotranspiration estimation, urban heat island analysis, vegetation monitoring, agricultural monitoring, etc. [4–7]. However, limited by a single satellite platform and sensor, it is difficult to obtain surface temperature with both high spatial and temporal resolution, limiting applications such as fine-scale wildfire monitoring and local-scale farm management [8–10].

To address this problem, in the last decade, many downscaling models have been introduced in both computer vision and remote sensing fields [11]. Generally, all these downscaling models can be classified as super-resolution models and spatiotemporal fusion models. Compared to the super-resolution models, which can be defined as the point spread function rebuilding problem, the spatiotemporal fusion methods mainly focus on how to take advantage of different remote sensing platforms and sensors by a fusing process [12,13]. Under the proper fusion hypothesis with suitable auxiliary data, the spatial-temporal fusion model can overcome the larger downscaling ratio and benefit from temporal continuity features in time series remote sensing observations [14,15]. From the perspective of key ideas, spatial-temporal fusion models can be commonly divided into three categories: weight function-based methods, learning-based methods, and unmixing-based methods.

The weight function-based methods assume that the temporal variation of surface reflectance is spatially consistent nearby, so the reflectance residual between the low- and high-resolution image can be formed with a linear weighted function, then selecting multiple nearby pixels under the measurement of spatial, spectral, and temporal similarity to solve this weighted function [16]. Under this idea, the spatial and temporal adaptive reflectance fusion mode (STARFM) [17] is the first weight function-based method to be developed and introduces the concepts of moving windows to search the candidate pixels to solve the weighted function. In this model, the search processing is limited by the window size. Sufficient pixels can easily be selected in the homogeneous region. However, for the regions with complex landscapes or changes, this method may result in large errors in the fusion results. Aiming at the shortcomings of the STARFM method, the spatial temporal adaptive algorithm for mapping reflectance change (STAARCH) uses tasseled cap transformations to detect reflectance changes, thereby improving the ability to detect land cover changes from low-resolution imagery [18]. To improve the temporal fusion capability, the enhanced STARFM (ESTARFM) [19] was formed by involving conversion coefficients, linear spectral unmixing, and spectral correlation coefficients in the temporal domain. With the extra-high-resolution images, ESTARFM shows higher prediction accuracy in heterogeneous landscapes with more spectral details. Similar results can be found with the robust adaptive spatial and temporal fusion model (RASTFM) [20], which consists of a nonlocal linear regression (NL-LR)-based weighted average module. Overall, the weight function-based methods usually have few parameters with high computational efficiency. However, for the parameters with rapid change, such as land surface temperature, these models still have limitations.

The second category is the learning-based model, which predicts high spatial and temporal resolution images using machine learning algorithms to solve the nonlinear problem between high- and low-resolution images [13]. In the context of this idea, the sparse representation-based spatiotemporal reflectance fusion model (SPSTFM) uses dictionary learning to establish the relationship between low and high spatial resolution images corresponding to reflectance changes and predicts target high-resolution images by time weighting [21]. For a stable dictionary, error-bound regularized semi-coupled dictionary learning (EBSCDL) [22] introduced a regularized error-bound method and constructed an optimized semi-coupled dictionary to capture the divergence between low spatial resolution images and high-resolution images. However, the sparse representation-based fusion method requires a properly designed dictionary. In algorithm implementation, dictionary learning, sparse coding, and image reconstruction are separated, which increases the instability and complexity of the algorithm. In this regard, the fusion method based on deep convolutional neural networks (STFDCNN) combines CNNs and nonlinear mapping models to design double CNNs, which realizes the automatic extraction of image features and improves prediction accuracy [23]. To reduce the impact of a single algorithm failure, a modified ensemble LST model was used, combining three regressors including a random forest, a ridge, and a support vector machine [24]. Generally, based on large training samples, learning-based methods can capture more surface spatial details and are suitable

for heterogeneous regions, etc. However, this generalization of the model across different locations and times can be a challenge.

Unmixing-based methods are based on linear spectral mixing theory, where endmembers are extracted from high spatial resolution images to decompose low spatial resolution images to obtain the abundances of each endmember [25]. The typical steps of an unmixing-based model include endmember selection, abundance calculation, unmixing processing, and fusion [26]. The multisensory multiresolution technique (MMT) is the first unmixing-based method to fuse the different resolution images, assuming a similar class contains a similar reflectance [27]. The unmixing processing is handled by a moving window and then assigns the result with high spatial resolution classes. However, the assumptions in the algorithm may not hold for locations where the landcover types vary with time. In this regard, Zurita-Milla et al. introduced constraints in the linear unmixing process to deal with the negative and outlier problems of the unmixing results [28]. The spatial-temporal data fusion approach (STDFA) is based on the assumption that the time-varying attributes of each land cover class are constant, using a modified linear mixed model that accounts for spatial variation [29]. Based on the STDFA algorithm, Zhang et al. combined the multiscale segmentation algorithm and ISODATA algorithm to generate classification maps, used the moving window method to unmix low spatial resolution images, and introduced temporal weights to fuse the temporal images [30]. To improve the accuracy of model prediction in land cover changes, the unmixing-based spatial-temporal reflectance fusion model (U-STFM) assumes that the reflectance change ratio remains the same across different sensors under the given location during the same period [31]. In the spatial domain, regions share the same ratio as the homogeneous change regions (HCRs), which can be obtained using the multiresolution segmentation algorithm. Then, the image on the target date can be predicted by using the high spatial resolution data before and after with the change ratio. The unmixing process is based on a linear unmixing function with low computational complexity [26]. Based on the multiple predictions with the time series data, U-STFM improves the fusion accuracy for predicting features with rapid change such as ocean surface chlorophyll *ab* concentration [32].

LST data with strong volatility of temporal changes and low spectral features makes the parameters of the weight function class method more sensitive [33]. Because LST is the continuous variable, generally from 250 k to 350 k, it is hard to build sufficient training for the learning-based model to capture the relationship with continuous mapping. In contrast, the unmixing-based methods, such as the U-STFM model, are more suitable for this problem since the linear unmixing process is applied to the time-changing signal, which can capture the rapid changes in LST when these changes are recorded in the low-spatial resolution data. Moreover, it is temporally based, which means each location can have unique weights, which affords the model more flexibility for continuous mapping.

However, due to the complexity of temperature changes and the ill-condition problem of unmixing processing, the stability of the unmixing model can be affected by the following three aspects [34]: Firstly, the definition of the HCRs or endmembers. HCRs are mainly obtained from two types of models: clustering and segmentation. Segmentation processing focuses on the local similarity based on Tobler's first law, while clustering concentrates on the similarity in feature space, so the pixels nearby may be classified into different HCRs; secondly, the level of unmix processing. In U-STFM, the least squares algorithm is commonly used to solve linear mixing problems, which can be applied at the single date level, the differential level, and the change ratio level. Different levels require a different number of applications of the unmixing process, which affects the stability of the results; thirdly, the number of HCRs or endmembers. Generally, the higher the number of HCRs, the greater the detail in which the surface changes can be described. However, more HCRs may lead the unmixing process to the ill-posed problem then increase the uncertainty of the prediction.

Overall, for the unmixing-based fusion model, the aforementioned factors are essential for the model's stability. The quantitative analysis is needed to answer the question of how

these factors affect the prediction and how to find a suitable strategy to minimize the prediction error. In this study, we focus on these three factors for quantitative analysis of the effect of the different HCR definitions and unmixing levels for LST spatial temporal fusion. Part of the Guangdong–Hong Kong–Macao Greater Bay Area (GBA) was selected as the study area, and Landsat7 LST products and MODIS LST products were selected to build the high-spatial and -temporal products. The main unmixing model we used in this study was the U-STFM model. In Section 2, we describe our study area and datasets; Section 3 describes the experimental design and the theoretical basis of U-STFM. The performance of the model was compared with the commonly used STARFM and ESTARFM to verify the effectiveness and stability of the LST prediction. The results and discussion are delivered in Sections 4 and 5. The main conclusion is in Section 6.

2. Study Area and Dataset

2.1. Study Area

The study area was selected within the part of the Guangdong–Hong Kong–Macao Greater Bay Area (GBA) (113°49′13″E–114°16′10″E and 22°37′17″N–22°59′48″N) with an area of about 1843 km² (Figure 1). The topography of the region is complex, with various types of landforms. With rapid economic development, Dongguan and Shenzhen, as the core cities of GBA, have undergone significant changes in land use and urban expansion. A large area of forest and wasteland has been transformed into a city area. The spatial distribution of land surface temperatures has changed rapidly.

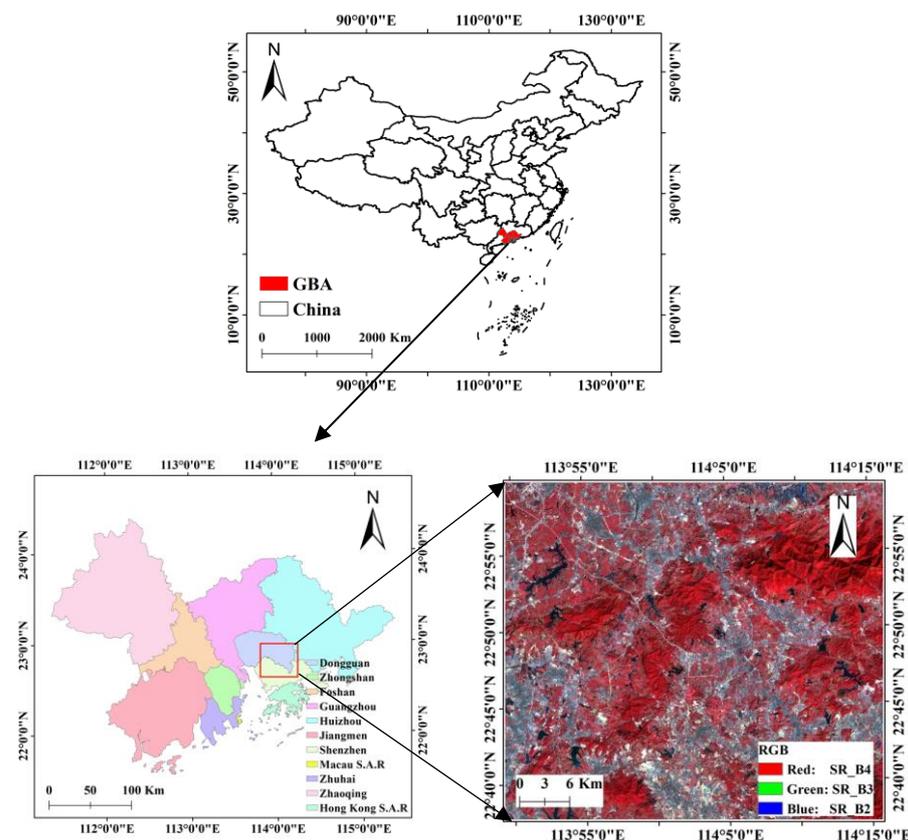


Figure 1. The study area is located in Shenzhen and Dongguan within the GBA of China.

2.2. Dataset

The Landsat7 ETM+ LST products and MODIS LST products (MOD11A1.006) were downloaded from USGS Earth Explorer (<https://earthexplorer.usgs.gov> (accessed on 12 April 2022)), and the “ST_B6” band and “LST_Day_1km” band were used, respectively. The spatial resolution of ETM + LST is 60 m, and the revisit frequency is 16 days; the spatial resolution of MODIS LST products is 1000 m, and the revisit frequency is 1 day. Landsat7

data were resampled (nearest neighbor method) to 30 m using arcgis10.8. Due to the failure of ETM + SLC after 31 May 2003, and the perennial cloudy and rainy conditions in the study area, this study selected data with a cloud cover threshold of less than 1% from September 2000 to May 2003 and collected eight pairs of valid Landsat7 LST and MODIS LST images pairs. Details are shown in Table 1.

Table 1. Data list of Landsat 7 LST and MODIS LST products used in the study area.

Date	Landsat 7 LST and MODIS LST Data Names	Spatial Resolution (m)
14 September 2000	LE71220442000258SGS00	30
	MOD11A1.A2000258.h28v06.061	1000
1 November 2000	LE71220442000306SGS00	30
	MOD11A1.A2000306.h28v06.061	1000
17 September 2001	LE71220442001260SGS00	30
	MOD11A1.A2001260.h28v06.061	1000
20 November 2001	LE71220442001324SGS00	30
	MOD11A1.A2001324.h28v06.061	1000
22 December 2001	LE71220442001356BKT00	30
	MOD11A1.A2001356.h28v06.061	1000
7 January 2002	LE71220442002007SGS00	30
	MOD11A1.A2002007.h28v06.061	1000
7 November 2002	LE71220442002311EDC00	30
	MOD11A1.A2002311.h28v06.061	1000
10 January 2003	LE71220442003010EDC00	30
	MOD11A1.A2003010.h28v06.061	1000

3. Methodology

3.1. Experimental Design

Three aspects of the stability of the unmixing model were evaluated: (1) the definition of the HCRs or endmembers, (2) the level of unmix processing, and (3) the number of HCRs or endmembers. The U-STFM method is used as the baseline model and then we compared the performance of the improved model with the STARFM and ESTARFM models.

Experiment I: For the definition of HCRs, the ISODATA [35], K-means [36], and multiresolution segmentation algorithms [37] were used to obtain the HCRs, respectively. The selected prediction date was 20 November 2001, and a total of 12 experimental date pairs were arranged and combined.

Experiment II: For analyzing the impact of the level of unmix processing, we unmixed the single date level (the unmixing process was applied to the image from each date, unmixing three times), two-date differential level (the unmixing process was applied to the differential images among three dates, unmixing twice), and three-date change ratio level (the unmixing process was applied to the change ratio of the three dates, unmixing once). For 6 available target dates (Table 1), 56 pairs of image groups in total were used to analyze how the different unmixing levels affect the stability of the model.

Experiment III: For testing the effect of the different number of HCRs, the number of HCRs was set from 10 to 259 by using the ISODATA algorithm; as in experiment two, 56 groups of image pairs were also used to evaluate the performance.

Experiment IV: The performance of the model with the best setup strategy from experiment I–IV was tested by comparing with the three classical models, STARFM, ESTARFM, and the original USTFM. The overall experimental design is shown in Figure 2.

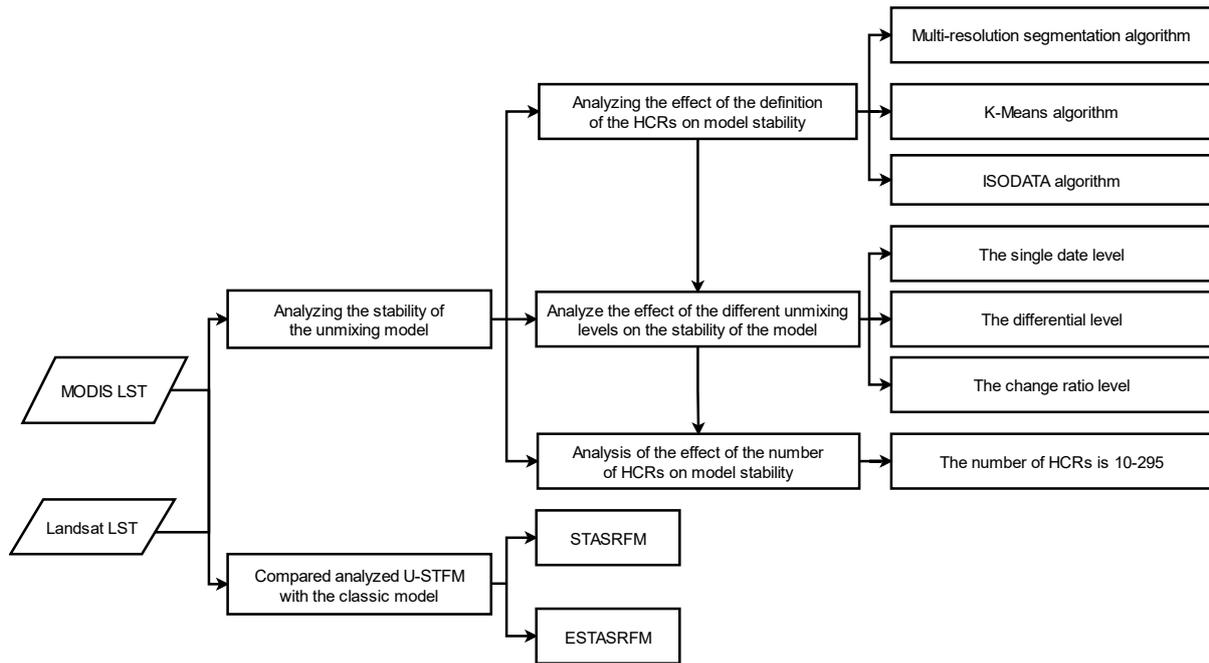


Figure 2. Experimental design of this study.

3.2. The Theoretical Basis of U-STFM

U-STFM was selected as the representative algorithm among the unmixing-based methods. The reason for choosing this model is that USTFM is a linear unmixing-based built based on the temporal change ratio, which suits the rapid land cover changes.

The U-STFM model needs to input the MODIS LST data of the target date (t_p), as well as the MODIS LST data and Landsat LST data before and after the target date (t_0 , t_e) to predict the Landsat7 LST data of the target date. The model is based on a linear spectral mixing model, which assumes that a low spatial resolution pixel can be represented by the linear combination of the spectral values of HCRs weighted by the fractional coverage within that pixel. The linear mixing processing is described in Equation (1):

$$M_t(i, j) = \sum_{x=1}^n f_t(x) \overline{M_t^{SR}(x)} \quad (1)$$

where $M_t(i, j)$ represents the reflectance of MODIS pixel (i, j) on date t ; $\overline{M_t^{SR}(x)}$ is the average reflectance on the x th HCRs, where SR stands for “super-resolution endmember”; $f_t(x)$ is the coverage of the x th HCRs on MODIS (i, j) pixels.

Assuming that the ratio of LST change in Landsat and MODIS sensors remains consistent over the same observation period (t_0 , t_p , t_e):

$$\alpha_p^L(i, j) = \alpha_p^{MSR}(i, j) \quad (2)$$

$$\alpha_p^L(i, j) = \frac{\Delta L_{pe}(i, j)}{\Delta L_{0p}(i, j)} = \frac{L_e(i, j) - L_p(i, j)}{L_p(i, j) - L_0(i, j)} \quad (3)$$

$$\alpha_p^{MSR}(i, j) = \frac{\Delta M_{pe}^{SR}(i, j)}{\Delta M_{0p}^{SR}(i, j)} = \frac{M_e^{SR}(i, j) - M_p^{SR}(i, j)}{M_p^{SR}(i, j) - M_0^{SR}(i, j)} \quad (4)$$

where $\alpha_p^L(i, j)$ is the time-series change trend ratio of a Landsat pixel (i, j); $\alpha_p^{MSR}(i, j)$ is the change trend ratio of HCRs in MODIS pixel (i, j); $\Delta L_{pe}(i, j)$ and $\Delta L_{0p}(i, j)$ represent the difference in Landsat pixel reflectance over the two periods [t_p , t_e] and [t_0 , t_p]; $\Delta M_{pe}^{SR}(i, j)$

and $\Delta M_{0p}^{SR}(i, j)$ represent the difference in HCR pixel reflectance over the two periods $[t_p, t_e]$ and $[t_0, t_p]$, respectively.

Therefore, combining Equations (2)–(4), we can speculate the Landsat pixel reflectance for the target date (t_p):

$$L_p(i, j) = \frac{L_e(i, j) - \alpha_p^{MSR}(i, j)L_0(i, j)}{1 + \alpha_p^{MSR}(i, j)} \quad (5)$$

Equation (5) shows that the Landsat image of the target date is related to the Landsat images of the two periods before and after and the rate of change of HCRs. The linear unmixing theory (Equation (1)) is applied to the MODIS time series to obtain α_p^{MSR} , and finally, the Landsat image at the target date t_p is predicted.

3.3. Evaluation

In this study, qualitative and quantitative evaluation metrics are used to evaluate the effect of the model in predicting the land surface temperature. For each prediction, the Landsat LST production with 30 m spatial resolution was used as the ground truth. There are 6 reliable dates in total across 8 dates. For each date, the different three-date combination groups are evaluated, for example, there are 12 available three-date groups for 20 November 2001. The qualitative evaluation of model fusion is evaluated by comparing and analyzing the visualization effects of predicted and real LST images. The peak signal-to-noise ratio (PSNR), correlation coefficient (CC), root mean square error (RMSE) and mean absolute error (MAE) are used for quantitative evaluation. PSNR is a full-reference image quality evaluation index. The effective range of CC value is between the interval $(-1, 1)$; closer to 1 indicates a better fusion result. Higher PSNR means better image fusion, while smaller RMSE and MAE values mean higher-quality fused images. All the above quantitative evaluation indicators are implemented by calling the evaluation indicator function corresponding to the scikit-learn module. The definitions of PSNR, CC, RMSE, and MAE are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (L(i, j) - P(i, j))^2}{M \times N}} \quad (6)$$

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (7)$$

$$CC = \frac{\sum_{i=1}^M \sum_{j=1}^N (L(i, j) - \mu_L)(P(i, j) - \mu_P)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (L(i, j) - \mu_L)^2 (P(i, j) - \mu_P)^2}} \quad (8)$$

$$RMSE = \sqrt{MSE} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^M \sum_{j=1}^N |L(i, j) - P(i, j)|}{M \times N} \quad (10)$$

where $L(i, j)$ and $P(i, j)$ represent the actual observed Landsat pixel (i, j) and the predicted image pixel (i, j), respectively; M and N represent the height and width of the image, respectively; MAX_I represents the maximum value of the image color; μ_L and μ_P represent the average value of the observed image and the predicted image, respectively.

4. Results and Discussion

4.1. Analyzing the Effect of the Definition of the HCRs on Model Stability

The HCRs or endmembers definition is the first step of the unmixing model, so accurate HCRs are crucial to improve the performance of the unmixing process. In this study, the multiresolution segmentation algorithm, K-means algorithm, and ISODATA algorithm are selected to obtain HCRs.

The multiresolution algorithm is a bottom-up region merging technique, which divides different regions spatially by identifying the differences in spectra, shapes, and other features of remote sensing images. The K-means algorithm is a clustering algorithm based on distance, which uses the distance between data objects as the similarity criterion and divides the data into different clusters by iterative update; the ISODATA algorithm is a dynamic cluster analysis algorithm based on the K-means algorithm with the “merge” and “split” operations, and the cluster centers and cluster numbers can be updated based on the characteristic of data distribution [38].

The quantitative evaluation results are shown in Figure 3. It can be seen that the ISODATA algorithm (the median PNSR: 46.93, CC: 0.94, RMSE: 1.80 K, MAE: 1.08 K) is more accurate in dividing HCRs and more suitable for capturing the LST variation in time series than the K-means (the median PNSR: 46.73, CC: 0.94, RMSE: 1.84 K, MAE: 1.19 K) and the segmentation algorithm (the median PNSR: 45.76, CC: 0.95, RMSE: 2.07 K, MAE: 1.18 K) with high PNSR and low RMSE and MAE. The lower CCs of ISODATA are caused by the outliers of the prediction. Overall, the segmentation-based methods consider the local similarity and provide the most accurate spatial division of homogeneous change regions. However, it easily over-segmented the region and lead to an unstable solution of the linear unmixing system (Equation (1)). The clustering-based algorithm can reduce this problem by providing more accurate division and reducing the number of HCRs at the same time, which is more suitable for the linear unmixing system.

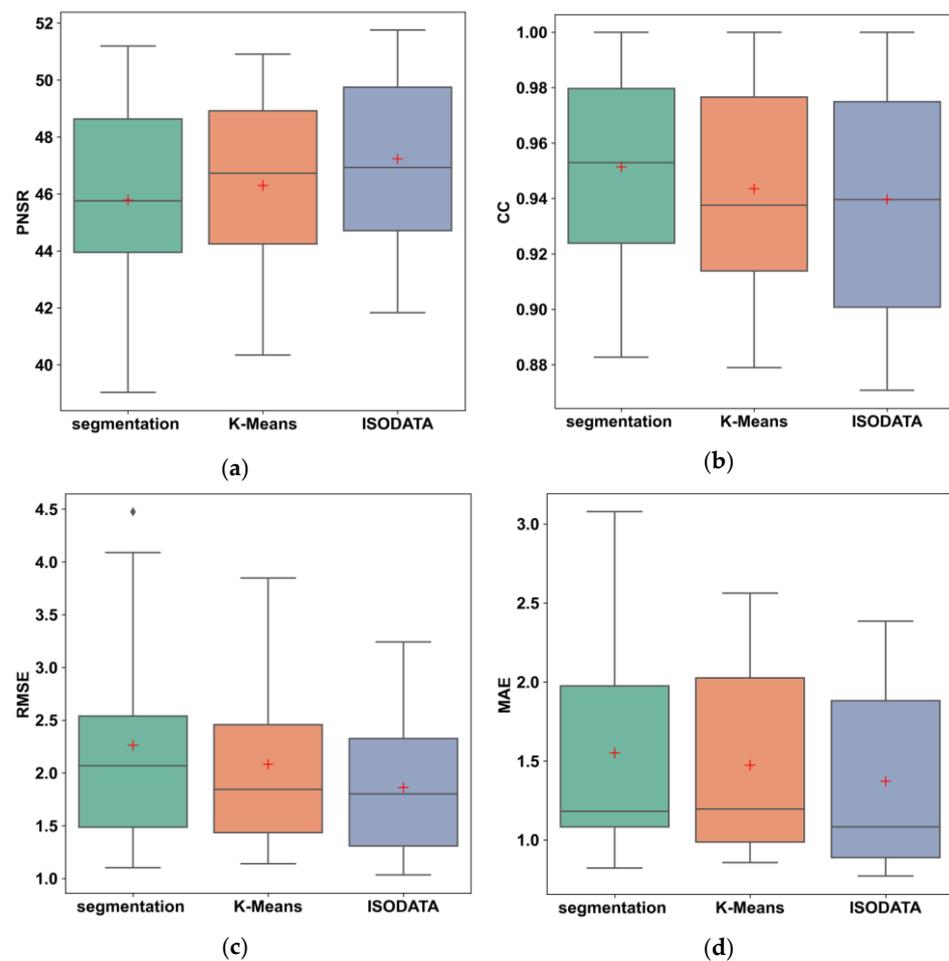


Figure 3. Comparison of model fusion results based on multiresolution segmentation algorithm, k-means algorithm, and ISODATA algorithm. (a) PNSR; (b) CC; (c) RMSE; (d) MAE.

Figure 4 shows the distribution of the absolute error of predictions on 20 November 2001. For each HCR definition, most of their absolute errors are distributed within 5 K.

The ISODATA-based model (Figure 4c) is generally less than 2 K, showing better results compared to the K-means and multiresolution segmentation. Overall, compared to the segmentation method (the original chosen in U-STFM), the clustering algorithms have better performance with lower RMSE and MAE variation. The reason is that the clustering-based HCRs is more related to the different land surface type, which is more likely to share the same LST change ratio. Compared to the K-means, the ISODATA algorithm can more accurately obtain LST with lower variation, so it is selected for experiment IV.

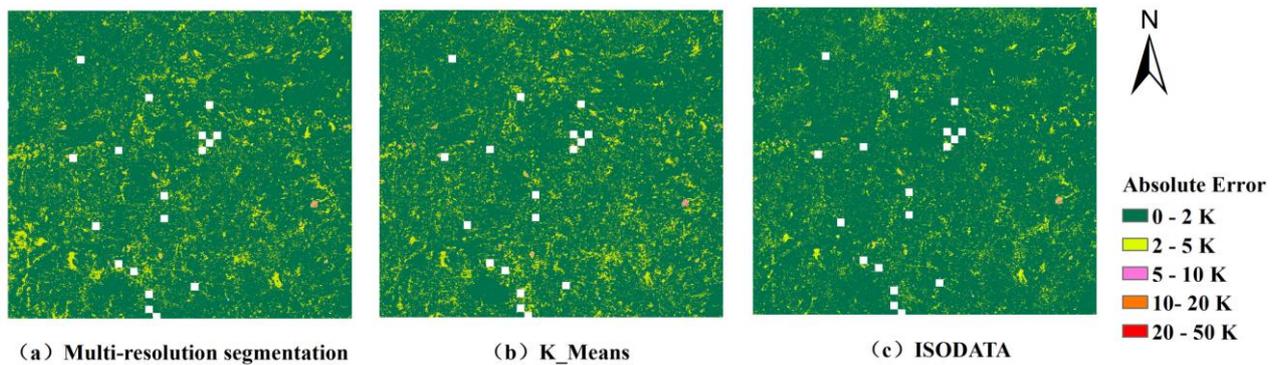


Figure 4. Comparison of absolute error distribution of model fusion results based on multi-scale segmentation (a), k-means (b) and ISODATA (c) on 20 November 2001. The missing pixels are influenced by cloud cover.

4.2. Analyze the Effect of the Different Unmixing Levels on the Stability of the Model

The linear unmixing equation is the core component of the unmixing-based models. The unmixing matrix composed of HCR values, abundance, and MODIS pixels can easily become an ill-conditioned matrix when the number of HCRs is larger than the number of MODIS pixels. At the same time, the least squares optimization is more sensitive to outliers. As mentioned before, there are three different levels for applying the unmixing processing (Figure 5): (1) the single-date level: unmixing each datum on t_0 , t_p , t_e , then calculating the change ratios; (2) the differential level: unmixing the difference between $t_e - t_p$ and $t_p - t_0$ then calculating the change ratios; and (3) the change ratio level: calculating the change ratio first then solving the unmixing problem directly on the change ratio. In this section, we will explore the impact of the different levels on the model performance.

The results are shown in Figure 6. As we can see, the fusion accuracy at the single-date level is significantly lower than the other two levels, and the performance of the model in the differential level and the change ratio level is much more similar. Based on the RMSE and MAE, the model on the change ratio level has lower variation (the median PNSR: 42.00; CC: 0.94, RMSE: 3.18 K, and MAE: 2.18 K). The reason behind this is that most prediction error comes from the unmixing process. So, the less unmixing, the lower the error. On the other hand, the differential image between different dates can reduce the additive systematic noise of sensors, so the performance of the model at the differential level is much better than that at the single date level. The change ratio level can further reduce the multiplicative noise of the system and then obtain better results.

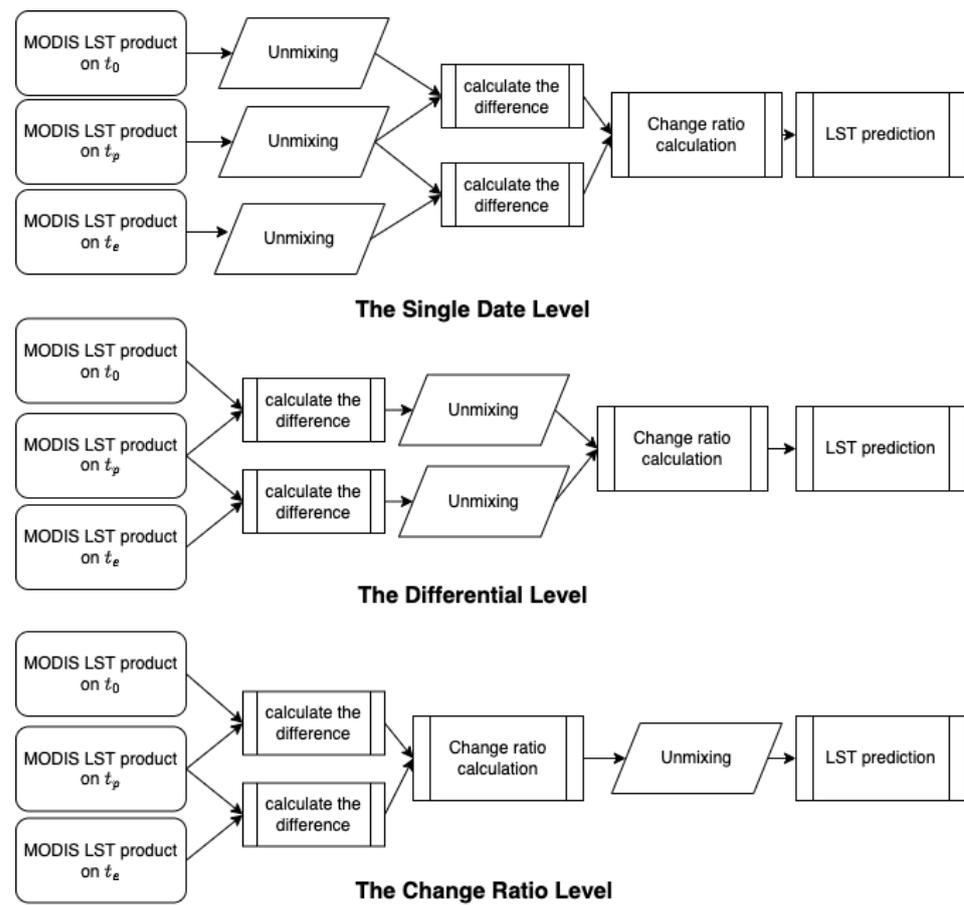


Figure 5. The different levels of the unmixing.

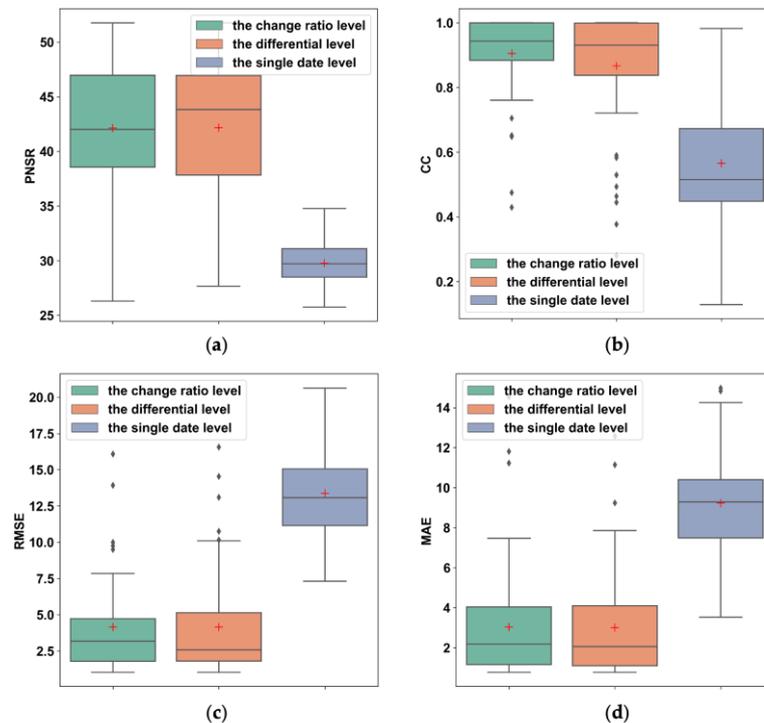


Figure 6. Comparison of model fusion results based on unmixing on the single date level, the differential level, and the change ratio level. (a) PNSR; (b) CC; (c) RMSE; (d) MAE.

Figure 7 shows the spatial distribution of the absolute error of the final prediction on 20 November 2001. We can see from Figure 7 that the prediction errors of the model in the change ratio level and the differential level are mostly distributed below 5 K. The errors of the model in the differential level have more regions within 2–5 K. However, the error of the model at the single-date level is relatively large, most of which is distributed between 5 K and 20 K. based on this result, the model in the change ratio level is recommended for experiment IV.

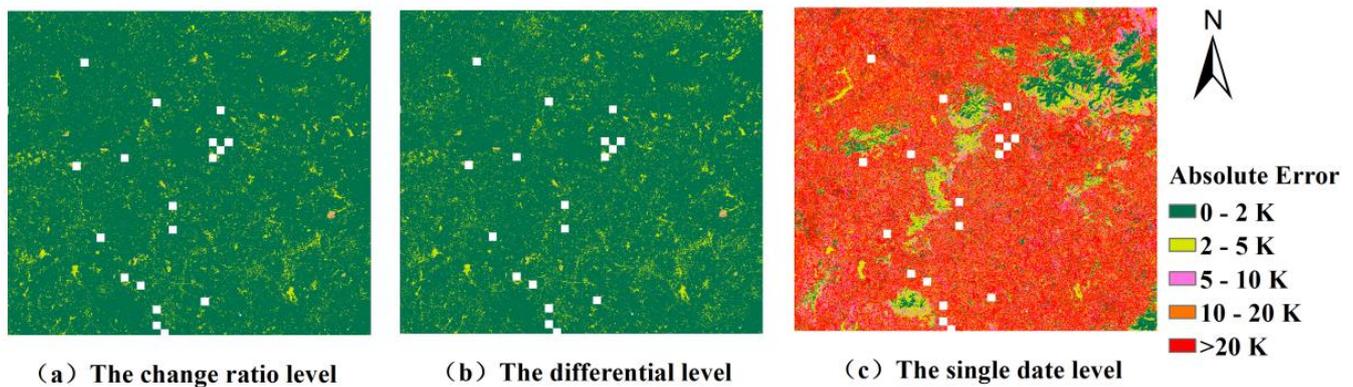


Figure 7. Comparison of the absolute error distribution of model fusion results based on unmixing on the single date level (a), the differential level (b), and the change ratio level (c) on 20 November 2001.

4.3. Analysis of the Effect of the Number of HCRs on Model Stability

Theoretically, the finer the HCRs, the more details of LST are obtained. However, when the number of HCRs is larger than the available number of MODIS pixels, it leads to instability in solving the linear unmixing equations. The number of HCRs also affects the accuracy of solving equations by the least squares optimization. Therefore, how to set up the appropriate number of HCRs is crucial to the performance of the model. In this section, we use the ISODATA algorithm to divide HCRs and apply the model at the change ratio level. The number of HCRs was set from 10 to 295 to evaluate the model performance.

As shown in Figure 8, we can see that there is a strong correlation between MAE and the number of HCRs. The MAE shows a decreasing trend as the number of HCRs increases. The PSNR and RMSE showed an increasing and decreasing trend when the number of HCRs categories was below 104, and then both gradually stabilized. The CC values were above 0.92, and the values were relatively high when the number of HCRs categories was within 100, and then had a decreasing trend. Figure 9 is a visualization of the absolute error of the LST results for selecting models based on 10, 104, and 207 HCRs. Their absolute errors are generally distributed within 5 K, with the largest distribution of absolute errors in the range of 2–5 K for the models based on HCR category 10. The overall distribution of the model based on the HCR category of 104 and the model of the HCR category of 207 is relatively similar, and the former is more concentrated. When the number of HCR categories in the model is 104, it has a median PSNR of 43.11 K, a median CC of 0.93, a median RMSE of 2.81 K, and a median MAE of 2.17 K. Overall, the model fusion effect is better when the HCRs category is 104.

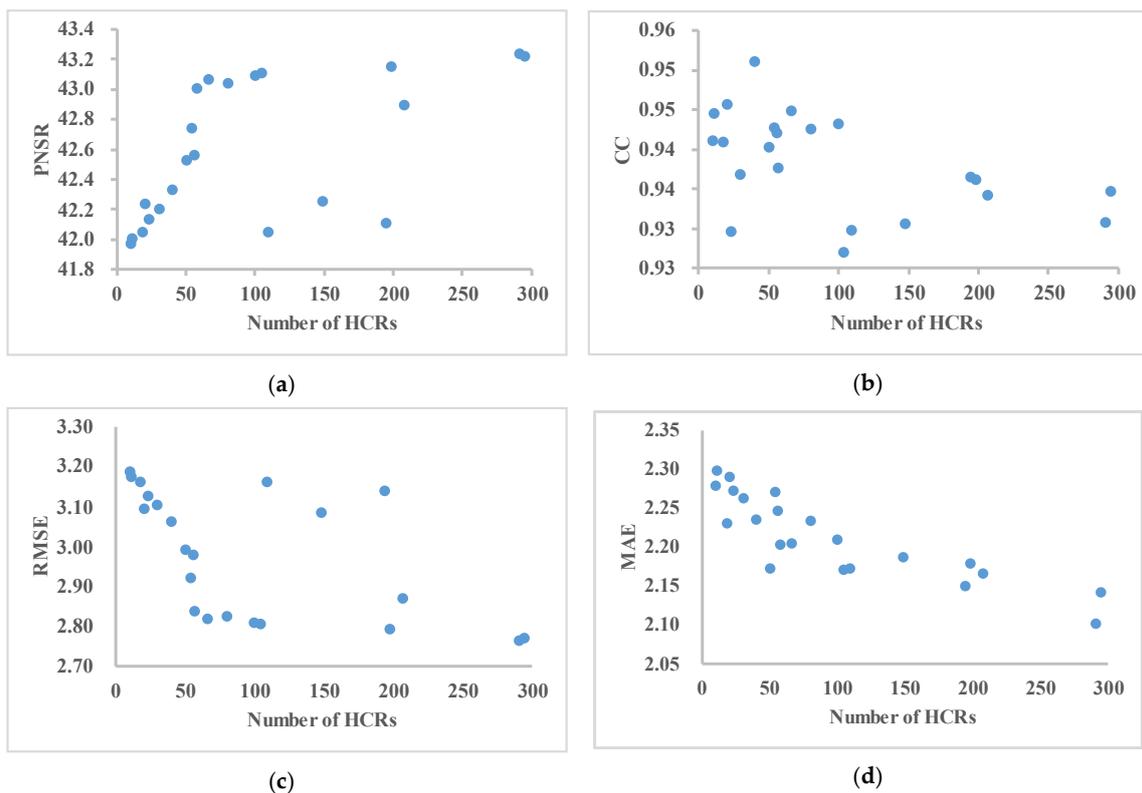


Figure 8. Comparison of model fusion results based on different numbers of HCRs. (a) PNSR; (b) CC; (c) RMSE; (d) MAE. Each point is the median value of 56 predictions (about 2027671 pixels for each prediction) across 6 predicted dates (from 1 November 2000 to 7 November 2002).

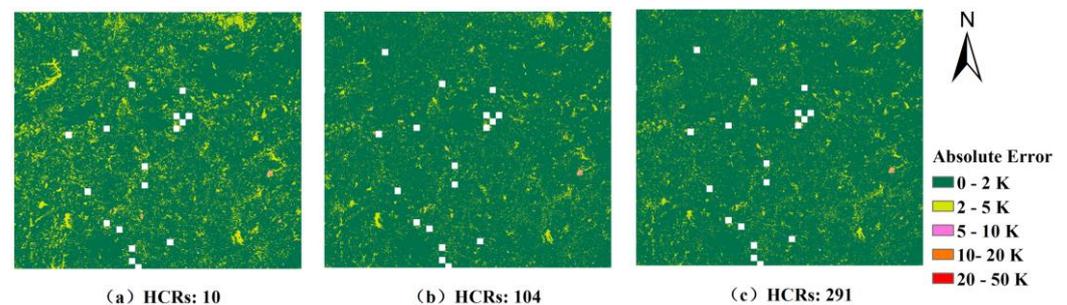


Figure 9. Comparison of the absolute error distribution of model fusion results based on different numbers of HCRs on 20 November 2001.

4.4. Compare Analyzed USTFM with STARFM and ESTARFM

Based on the above experimental results, the best setup for the USTFM model uses the ISODATA algorithm for HCR detection, the number of HCRs was set above 100 (the final number is 104), and the unmixing processing was set at the change ratio level. To distinguish from the original model, we denoted these setups as iso_USTFM and compared this model with the three classical models, STARFM, ESTARFM, and USTFM (the best segment size: 59).

Figure 10 is a visualization of the absolute error of the LST results for selecting models based on STARFM (Figure 10a), ESTARFM (Figure 10b), USTFM (Figure 10c), and iso_USTFM (Figure 10d). We can see that the prediction error of the model in iso_USTFM is mostly distributed below 1.5 K. The error of the model in ESTARFM and USTFM are mostly distributed below 2.5 K. However, the error of the model in STARFM is relatively large, most of which is distributed below 5 k. STARFM has the largest error in predicting LST, and the LST distribution shows many low- and high-temperature extremes that do

not match reality. A comparison of the absolute error distribution of model fusion results for land cover categories are shown in Figure 11. We can see that the performance of the iso_USTFM is better than STARFM, ESTARFM, and USTFM on barren, cropland, forest, grassland, and impermeable surfaces, with lower absolute error, but has higher errors for water and shrubs (Figure 11b).

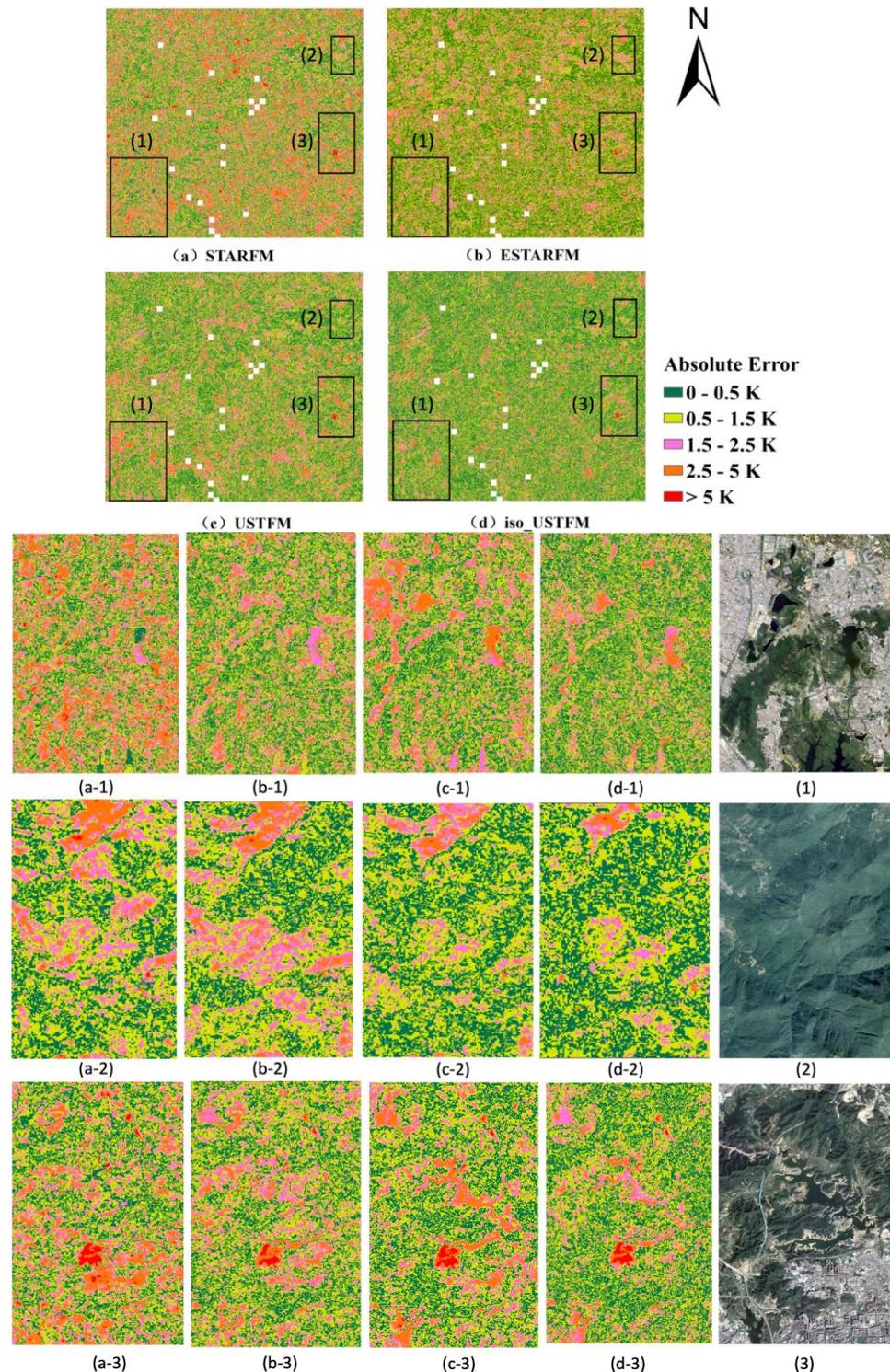


Figure 10. Comparison of the absolute error distribution of model fusion results on 20 November 2001. (a) STARFM, (b) ESTARFM, (c) USTFM, (d) iso_USTFM in three sub-regions: (1) city, (2) forest, (3) lakes. The true-color composite Landsat 7 image from Google earth on 31 December 2001.

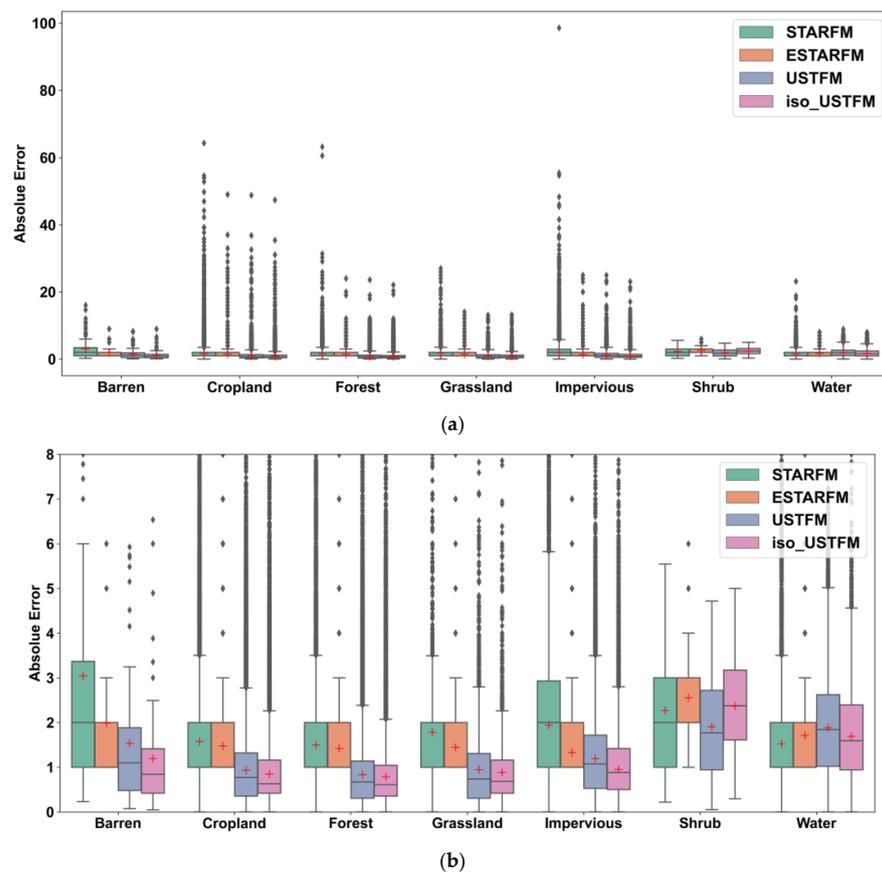


Figure 11. (a) Comparison of absolute error distribution of model fusion results for land cover categories in 2001. (b) 0–8 k range of (a).

Figure 12 shows the scatterplots of the LST predicted by the four models. According to the data distribution, all groups have data points close to the 1:1 diagonal, STARFM has a large number of outliers, and iso_USTFM has the highest correlation. Based on the metrics, the determination coefficient R^2 of the iso_USTFM model is 0.875 with the RMSE being 1.089 K, which is better than STARFM (R^2 : 0.764; RMSE: 1.729), ESTARFM (R^2 : 0.821; RMSE: 1.334) and USTFM (R^2 : 0.841; RMSE: 1.275).

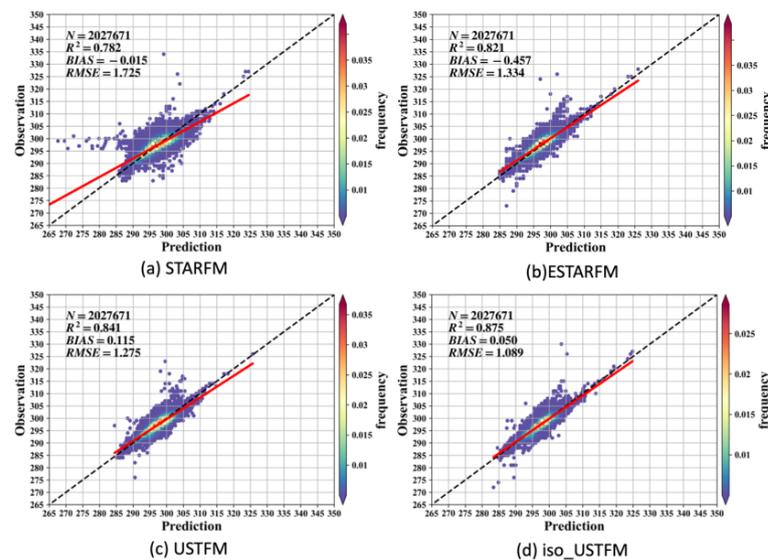


Figure 12. Scatter plot of predicted LST using these three models with Landsat7 LST on 20 November 2001.

5. Discussion

5.1. The Error Source of the Unmixing-Based Model

Generally, the unmixing-based model has two key components, one is the unmixing strategy and the other is the weighting function.

For the unmixing strategy, when the actuarial mixing process is unknown, the linear unmixing strategy is commonly used. However, the linear system may not fit reality. The thermal infrared signal mixed from the fine to coarse pixels may be dominant by the hottest object inside this area. Simply mixing them by the linear function may cost systematic error. The other linear system problem is the instability in solving the linear unmixing equations. When the number of unknown variables (the HCRs units) is larger than the number of equations (the available MODIS pixels), many unknown variables covered only by one or two equations will lead to no reliable and stable solution. This can explain the finding in Figure 3, where the segmentation-based method tends to over-segment the area.

The weighting function used by the U-STFM model is the area coverage ratio of each HCRs in each MODIS pixel. This strategy may suit the reflectance data fusion but may cost the problem in thermal infrared signals in the area containing hot or cold spots. In these areas, the hot or cold spots dominate the thermal infrared signal, so the weighting function fails in these areas. This can explain the finding in Figure 10(c3) where the hot spot region shows the biggest error.

The outliers (Figure 11) are mainly shown in cropland, forest, and impervious area. One of the reasons behind this is the land cover changes rapidly in these areas compared to the stable ones such as barren, shrub, and water. The USTFM model was originally designed for land cover change situations, as we can see the USTFM performed better compared to the STARFM and ESTARFM. However, for the stable categories, such as water and shrub, USTFM lost this benefit.

5.2. The Difficulty of LST Data Fusion

One of the big differences between the LST and the surface reflectance data is that the thermal infrared signal in LST data has the radiation effect of the hot spots. These hot spots will dominate the signal in the coarse pixel during the mixing process. So, the weighting function based on area coverage ratio (USTFM), the nearby stable pixels (STARFM), or tempol information (ESTARFM) may not be suitable for these areas. More accurate weighting functions and unmixing strategies need to be developed.

The other key factor for LST fusion is the tempol consistency. Most unmixing-based spatial-temporal fusion models require consistency between fine and coarse sensors. This consistency can be held under careful calibration for the surface reflectance data. But for the LST data, the model error involved by the different models may break this consistency, leading to unreliable prediction.

5.3. The Model Generalization

The most unmixing-based fusion model is the unsupervised model. This means that the unmixing process and the corresponding solution for HCRs only can be applied in the fixed region and dataset. In the GBA area, the land cover and land use units are piecemeal with rapid change, So the finding in this study may not be general in other places with different land cover conditions. The learning-based supervised nonlinear model can be developed to add generalizability to the model across different locations.

6. Conclusions

In this study, we identified three key factors for the unmixing-based spatial-temporal fusion model: (1) the definition of the HCRs, (2) the unmixing levels, and (3) the number of HCRs. Based on the step-by-step control experiment, we evaluated these three key factors quantitatively to answer the question of how these factors affect the stability of the model. The U-STFM model was used as the baseline model. Landsat7 and MODIS LST products were used to generate daily 30 m high spatiotemporal resolution LST products.

The best setups of U-STFM for LST prediction were compared with the classical STARFM and ESTARFM model. Based on the findings in our study area, the main conclusions are as follows:

1. The clustering-based algorithm is more suitable for detecting the HCRs and end-members for unmixing. Compared with the multi-scale segmentation algorithm and K-means algorithm, the ISODATA clustering algorithm can more accurately describe LST's temporal and spatial changes to HCRs.
2. The fewer times the unmixing processing is used, the more stable the model predictions are. For the U-STFM model, applying the unmixing processing at the change ratio level can significantly reduce the additive and multiplicative noise of the prediction.
3. The larger the number of HCRs (less than the available MODIS pixels), the more stable the model is. There is a tradeoff effect between the number of HCRs and the solvability of the linear unmixing function. The larger number of HCRs means that more details of the LST change can be captured, but the prediction of the change ratios in small HCRs that are only covered by a few MODIS pixels will be unstable.
4. For the fusion of the daily 30 m scale LST product, the suitable setups for U-STFM use ISODATA as the detector for HCRs, controlling the HCRs above 100 and applying the unmixing model at the change ratio level. Compared with STARFM and ESTARFM, modified U-STFM (iso_USTFM) achieved higher prediction accuracy and lower error.

The findings in this study can effectively help current unmixing-based spatiotemporal fusion models avoid the wrong setups and have the potential to achieve lower uncertainty and stable prediction. However, as shown in this study, the linear unmixing model based on the least squares algorithm is influenced by the outliers, which may not suit reality. Therefore, the nonlinear-based unmixing methods may further improve the model stability and robustness for daily fine-scale LST prediction.

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

Funding: This research was supported by the National Key Research and Development Program of China (Project No.2021YFF0703900), the Natural Science Foundation of China project (41601212, 42206035, 42171323, 42001286), and the Fundamental Research Foundation of Shenzhen Technology and Innovation Council (Project No. KCXFZ2020020111006298, KCXFZ20201221173613035), the Fundamental Research Foundation of Shenzhen Technology and Innovation Council (General Program) (Project No. JCYJ20190806170814498).

Acknowledgments: We thank all the GIS group members at the SIAT, Chinese Academy of Sciences, for their encouragement and discussion of the work presented here. We also thank to the anonymous reviewers for their valuable suggestions on the earlier drafts of this study.

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

References

1. Li, Z.-L.; Tang, B.-H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-Derived Land Surface Temperature: Current Status and Perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [[CrossRef](#)]
2. Ebrahimi, H.; Azadbakht, M. Downscaling MODIS Land Surface Temperature over a Heterogeneous Area: An Investigation of Machine Learning Techniques, Feature Selection, and Impacts of Mixed Pixels. *Comput. Geosci.* **2019**, *124*, 93–102. [[CrossRef](#)]
3. Duan, S.B.; Li, Z.L.; Tang, B.H.; Wu, H.; Tang, R. Direct Estimation of Land-Surface Diurnal Temperature Cycle Model Parameters from MSG–SEVIRI Brightness Temperatures under Clear Sky Conditions. *Remote Sens. Environ.* **2014**, *150*, 34–43. [[CrossRef](#)]
4. Liu, S.; Sun, Z.; Li, X.; Liu, C. A Comparative Study on Models for Estimating Evapotranspiration. *J. Nat. Resour.* **2003**, *18*, 161–167.
5. Alhawiti, R.H.; Mitsova, D. Using Landsat-8 Data to Explore the Correlation between Urban Heat Island and Urban Land Uses. *IJRET Int. J. Res. Eng. Technol.* **2016**, *5*, 457–466.

6. Semmens, K.A.; Anderson, M.C.; Kustas, W.P.; Gao, F.; Alfieri, J.G.; McKee, L.; Prueger, J.H.; Hain, C.R.; Cammalleri, C.; Yang, Y.; et al. Monitoring Daily Evapotranspiration over Two California Vineyards Using Landsat 8 in a Multi-Sensor Data Fusion Approach. *Remote Sens. Environ.* **2016**, *185*, 155–170. [[CrossRef](#)]
7. Weng, Q.; Fu, P.; Gao, F. Generating Daily Land Surface Temperature at Landsat Resolution by Fusing Landsat and MODIS Data. *Remote Sens. Environ.* **2014**, *145*, 55–67. [[CrossRef](#)]
8. Bo, H.; Yongquan, Z. Research Status and Prospect of Spatiotemporal Fusion of Multi-Source Satellite Remote Sensing Imagery. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 1492.
9. DONG, W.; MENG, J. Review of Spatiotemporal Fusion Model of Remote Sensing Data. *Remote Sens. Land Resour.* **2018**, *30*, 1–11.
10. Pu, R. Assessing Scaling Effect in Downscaling Land Surface Temperature in a Heterogenous Urban Environment. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *96*, 102256. [[CrossRef](#)]
11. Wang, S.; Luo, Y.; Li, X.; Yang, K.; Liu, Q.; Luo, X.; Li, X. Downscaling Land Surface Temperature Based on Non-Linear Geographically Weighted Regressive Model over Urban Areas. *Remote Sens.* **2021**, *13*, 1580. [[CrossRef](#)]
12. Li, J.; Li, Y.; He, L.; Chen, J.; Plaza, A. Spatio-Temporal Fusion for Remote Sensing Data: An Overview and New Benchmark. *Sci. China Inf. Sci.* **2020**, *63*, 140301. [[CrossRef](#)]
13. Zhu, X.; Cai, F.; Tian, J.; Williams, T. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sens.* **2018**, *10*, 527. [[CrossRef](#)]
14. Zhan, W.; Chen, Y.; Zhou, J.; Wang, J.; Liu, W.; Voogt, J.; Zhu, X.; Quan, J.; Li, J. Disaggregation of Remotely Sensed Land Surface Temperature: Literature Survey, Taxonomy, Issues, and Caveats. *Remote Sens. Environ.* **2013**, *131*, 119–139. [[CrossRef](#)]
15. Mao, Q.; Peng, J.; Wang, Y. Resolution Enhancement of Remotely Sensed Land Surface Temperature: Current Status and Perspectives. *Remote Sens.* **2021**, *13*, 1306. [[CrossRef](#)]
16. Shen, H.; Meng, X.; Zhang, L. An Integrated Framework for the Spatio-Temporal-Spectral Fusion of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7135–7148. [[CrossRef](#)]
17. Feng, G.; Masek, J.G.; Schwaller, M.R.; Hall, F.F. On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218. [[CrossRef](#)]
18. Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A New Data Fusion Model for High Spatial-and Temporal-Resolution Mapping of Forest Disturbance Based on Landsat and MODIS. *Remote Sens. Environ.* **2009**, *113*, 1613–1627. [[CrossRef](#)]
19. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model for Complex Heterogeneous Regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [[CrossRef](#)]
20. Zhao, Y.; Huang, B.; Song, H. A Robust Adaptive Spatial and Temporal Image Fusion Model for Complex Land Surface Changes. *Remote Sens. Environ. Interdiscip. J.* **2018**, *208*, 42–62. [[CrossRef](#)]
21. Huang, B.; Song, H. Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [[CrossRef](#)]
22. Wu, B.; Huang, B.; Zhang, L. An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6791–6803. [[CrossRef](#)]
23. Song, H.; Liu, Q.; Wang, G.; Hang, R.; Huang, B. Spatiotemporal Satellite Image Fusion Using Deep Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 821–829. [[CrossRef](#)]
24. Ouyang, X.; Dou, Y.; Yang, J.; Chen, X.; Wen, J. High Spatiotemporal Rugged Land Surface Temperature Downscaling over Saihanba Forest Park, China. *Remote Sens.* **2022**, *14*, 2617. [[CrossRef](#)]
25. Xie, D.; Zhang, J.; Sun, P.; Pan, Y.; Yun, Y.; Yuan, Z. Remote Sensing Data Fusion by Combining STARFM and Downscaling Mixed Pixel Algorithm. *J. Remote Sens.* **2016**, *20*, 62–72.
26. Liu, W.; Zeng, Y.; Li, S.; Huang, W. Spectral Unmixing Based Spatiotemporal Downscaling Fusion Approach. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102054. [[CrossRef](#)]
27. Zhukov, B.; Oertel, D.; Lanzl, F.; Reinhackel, G. Unmixing-Based Multisensor Multiresolution Image Fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [[CrossRef](#)]
28. Zurita-Milla, R.; Clevers, J.; Schaepman, M.E. Unmixing-Based Landsat TM and MERIS FR Data Fusion. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 453–457. [[CrossRef](#)]
29. Wu, M.; Niu, Z.; Wang, C.; Wu, C.; Wang, L. Use of MODIS and Landsat Time Series Data to Generate High-Resolution Temporal Synthetic Landsat Data Using a Spatial and Temporal Reflectance Fusion Model. *J. Appl. Remote Sens.* **2012**, *6*, 063507.
30. Zhang, W.; Li, A.; Jin, H.; Bian, J.; Zhang, Z.; Lei, G.; Qin, Z.; Huang, C. An Enhanced Spatial and Temporal Data Fusion Model for Fusing Landsat and MODIS Surface Reflectance to Generate High Temporal Landsat-like Data. *Remote Sens.* **2013**, *5*, 5346–5368. [[CrossRef](#)]
31. Huang, B.; Zhang, H. Spatio-Temporal Reflectance Fusion via Unmixing: Accounting for Both Phenological and Land-Cover Changes. *Int. J. Remote Sens.* **2014**, *35*, 6213–6233. [[CrossRef](#)]
32. Guo, S.; Sun, B.; Zhang, H.K.; Liu, J.; Chen, J.; Wang, J.; Jiang, X.; Yang, Y. MODIS Ocean Color Product Downscaling via Spatio-Temporal Fusion and Regression: The Case of Chlorophyll-a in Coastal Waters. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 340–361. [[CrossRef](#)]

33. Wu, P.; Yin, Z.; Zeng, C.; Duan, S.B.; Shen, H. Spatially Continuous and High-Resolution Land Surface Temperature Product Generation: A Review of Reconstruction and Spatiotemporal Fusion Techniques. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 112–137. [[CrossRef](#)]
34. Peng, K.; Wang, Q.; Tang, Y.; Tong, X.; Atkinson, P.M. Geographically Weighted Spatial Unmixing for Spatiotemporal Fusion. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5404217. [[CrossRef](#)]
35. Ball, G.H.; Hall, J. *A Novel Method of Data Analysis and Pattern Classification*; Stanford Research Inst: Menlo Park, CA, USA, 1965.
36. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.* **1967**, *5.1*, 281–298.
37. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258. [[CrossRef](#)]
38. Abbas, A.W.; Minallh, N.; Ahmad, N.; Abid, S.A.R.; Khan, M.A.A. K-Means and ISODATA Clustering Algorithms for Landcover Classification Using Remote Sensing. *Sindh Univ. Res. J. SURJ (Sci. Ser.)* **2016**, *48*, 315–318.

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.