sensors

MDPI

Article

An Improved Spatiotemporal Data Fusion Method for Snow-Covered Mountain Areas Using Snow Index and Elevation Information

Min Gao [1,2], Xingfa Gu [1,2,3], Yan Liu [1,*], Yulin Zhan [1], Xiangqin Wei [1], Haidong Yu [1,2], Man Liang [1,2], Chenyang Weng [1,2] and Yaozong Ding [1,2]

[1] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100049, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] School of Remote Sensing and Information Engineering, North China Institute of Aerospace Engineering, Langfang 065000, China
* Correspondence: liuyan@aircas.ac.cn

Abstract: Remote sensing images with high spatial and temporal resolution in snow-covered areas are important for forecasting avalanches and studying the local weather. However, it is difficult to obtain images with high spatial and temporal resolution by a single sensor due to the limitations of technology and atmospheric conditions. The enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) can fill in the time-series gap of remote sensing images, and it is widely used in spatiotemporal fusion. However, this method cannot accurately predict the change when there is a change in surface types. For example, a snow-covered surface will be revealed as the snow melts, or the surface will be covered with snow as snow falls. These sudden changes in surface type may not be predicted by this method. Thus, this study develops an improved spatiotemporal method ESTARFM (iESTARFM) for the snow-covered mountain areas in Nepal by introducing NDSI and DEM information to simulate the snow-covered change to improve the accuracy of selecting similar pixels. Firstly, the change in snow cover is simulated according to NDSI and DEM. Then, similar pixels are selected according to the change in snow cover. Finally, NDSI is added to calculate the weights to predict the pixels at the target time. Experimental results show that iESTARFM can reduce the bright abnormal patches in the land area compared to ESTARFM. For spectral accuracy, iESTARFM performs better than ESTARFM with the root mean square error (RMSE) being reduced by 0.017, the correlation coefficient (r) being increased by 0.013, and the Structural Similarity Index Measure (SSIM) being increased by 0.013. For spatial accuracy, iESTARFM can generate clearer textures, with Robert's edge (Edge) being reduced by 0.026. These results indicate that iESTARFM can obtain higher prediction results and maintain more spatial details, which can be used to generate dense time series images for snow-covered mountain areas.

Keywords: remote sensing; spatiotemporal data fusion; ESTARFM; NDSI; DEM

check for updates

Citation: Gao, M.; Gu, X.; Liu, Y.; Zhan, Y.; Wei, X.; Yu, H.; Liang, M.; Weng, C.; Ding, Y. An Improved Spatiotemporal Data Fusion Method for Snow-Covered Mountain Areas Using Snow Index and Elevation Information. *Sensors* **2022**, *22*, 8524. https://doi.org/10.3390/s22218524

Academic Editor: Alfred Stein

Received: 8 October 2022
Accepted: 2 November 2022
Published: 5 November 2022

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

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

1. Introduction

Snow cover is an important component of the cryosphere and has significant effects on regional water balance, local weather, atmospheric circulation, and surface hydrological processes [1–4]. Previous studies have shown that small changes in snow cover in mountain areas may have great thermal and dynamical influences on regional and even global circulation systems [5–8], which is an essential factor of energy balance. Remote sensing images with fine spatial and temporal resolutions contain much valuable information about the observed objects [9–12], which is a fundamental source for studying and monitoring the spatiotemporal distribution of snow cover [13–15]. However, due to technical and budget limitations, there is a trade-off between the swath width and the revisit cycle of satellites. So, it is difficult to obtain images with both high spatial and temporal resolution by a single

satellite [16], especially in the mountain region, where the spatial and temporal variability of the snow cover is particularly high [17]. Furthermore, optical remote sensing images are easily contaminated by clouds, cloud shadows, and atmospheric conditions, especially in cloudy and snowy areas such as tropical, subtropical, and high-altitude mountain areas [18]. As a result, the present remote sensing datasets cannot satisfy the need for high spatial time series observation.

The spatiotemporal data fusion method (STF) is a flexible, effective, and inexpensive solution to overcome these limitations. STF aims at fusing images with low spatial high temporal resolution and images with high spatial low temporal resolutions to generate images with high spatial and temporal resolution [19]. In the past decade, this method has been applied to a variety of research fields, such as crop monitoring, land cover classification, biomass estimation, and disturbance detection [20–24]. Meanwhile, a large number of STF methods have been developed in the remote sensing field. According to their principles, assumptions, and strategies, the STF methods can be divided into five categories [25–28]: weight function-based [29,30], unmixing-based [31,32], Bayesian-based [33], learning-based [34,35], and hybrid method [36,37]. Among these methods, weight-based methods are most widely applied in practical research because of their high prediction accuracy, good robustness, and flexibility [38–41]. The enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) was developed, aiming to improve applicability in heterogeneous regions [42]. This method is based on two pairs of fine and coarse images at base date and a coarse image at target time, which can obtain more spatial and temporal information. The ESTARFM can be reduced to three steps. Firstly, select similar pixels. The selection of similar pixels is based on the thresholds, which are estimated from the overall standard deviation of the whole image. Secondly, calculate weights and conversion coefficient. The weights are composed of spectral, temporal, and spatial distances. Finally, calculate the reflectance of the target pixel. Some experimental results show that ESTARFM can maintain fine spatial and spectral details as well as obtain high prediction accuracy in heterogeneous regions, such as agricultural areas and urban, coalfields [43–46]. However, ESTARFM still has several limitations and constraints in practical applications [36,40]. There are many studies that aim to solve the limitations in different fields. In order to overcome the uncertainty in selecting neighboring similar pixels, an improved ESTARFM method was developed based on the landcover endmember type [47]. To improve the prediction accuracy in non-shape-changing regions, Zhang et al. proposed an object-based method [48]. Aiming to improve the prediction accuracy of ESTARFM, a statistical method was proposed to adaptively determine the window size [42]. In order to improve the selection accuracy of similar pixels for the Paddy rice region, an enhanced vegetation index (EVI) was introduced for predicting reflectance [49]. To make it applicable to larger scales, an automatic filling cloud gaps framework was proposed to enhance the applicability in heterogeneous and cloud-prone landscapes [41]. Although these methods can obtain good prediction accuracy in the specific study areas, there is still a lack of specific methods in the snow-covered mountain regions where transient changes often occur due to snow and snowmelt [50].

There are some limitations of ESTARFM when applied to snow-covered mountain areas [51,52]. First, snow reflectance is higher than non-snow pixels in the visible band, which leads to a high threshold of selecting similar pixels, resulting in large errors. Similarly, for non-snow pixels, it could lead to a low threshold. Secondly, the type of pixels would change due to snow and snowmelt, while ESTARFM selects similar pixels based on the intersection of two fine images, which also leads to the error of selecting similar pixels. The above problems will lead to a decrease in prediction accuracy. Therefore, it is important to accurately capture snow-cover changes to improve prediction accuracy. The normalized snow index (NDSI) takes advantage of the characteristics of snow, which has a high reflection in the visible bands and strong absorption in the shortwave infrared, so a suitable threshold can distinguish snow and non-snow pixels [50,53]. For mountain areas, the variation in snow has a high correlation with elevation. So, the digital elevation model

(DEM) is essential for the analysis of the snow and glacier change in high mountain terrains [54]. Additionally, several studies have shown the applicability of morphometric parameters derived from DEM in the mapping of glaciers and snow cover [54–59]. Thus, NDSI and DEM are selected for estimating the change in snow.

This study proposes an improved ESTARFM (iESTARFM) for snow-covered mountain areas in the Nepal region by introducing NDSI and DEM information. Compared to the original ESTARFM, there are three improvements. Firstly, simulate snow-cover changes for the base time images using NDSI by taking advantage of the reflectance characteristics of snow. DEM data was introduced to simulate snow-cover for the target time images because of the high correlation between snowmelt and elevation. Secondly, select similar pixels. Similar pixels are selected according to the results of snow-covered changes. If the pixels at the target date are identified as snow, then similar pixels will be selected from the snow pixels in the window size. The same principle is applied to non-snow pixels. The thresholds are calculated separately according to the pixel type. Thirdly, calculate weights and target pixels. To reduce the error caused by the misclassification, NDSI is added to select similar pixels and calculate the weights; more similar pixels have more similar NDSI values and are assigned a larger weight value. Finally, the target pixels are calculated according to the similar pixels and their weights. The data used in this study are Landsat 8 surface reflectance images with high spatial resolution and MODIS surface reflectance products (MOD09A1) with a high temporal resolution, which are widely used in spatiotemporal data fusion methods [19,30,38,58]. The rest of the paper is organized as follows. Section 2 introduces the study area and datasets. Section 3 describes the details of the proposed iESTARFM method. Section 4 evaluates the performance of iESTARFM and compares it to the original ESTARFM. Sections 5 and 6 discuss and conclude the advantages and limitations of our method.

2. Materials

2.1. Study Area

Nepal experiences a wide range of climatic conditions that can be divided into two types: a dry winter period and a wet summer period. There are six bioclimatic zones that vary greatly in this climate, and they are tropical, subtropical, temperate, subalpine, alpine, and nival types [60]. The northern part of Nepal is a mountainous country that covers two-thirds of the Himalayan region with the eight highest mountains in the world [61]. Snow cover is one of the major types in Nepal [62], and there is seasonal snowfall in the region. The study area is located in the northwest part of Nepal in the mountain area, which is adjacent to China and the Himalayas, and the main land cover in this area is grass and snow. Snow cover changes rapidly in this study area, making the area suitable for testing the proposed method. Figure 1 shows the location of the study area and the Landsat image on 12 February 2020 using an RGB composite.

2.2. Satellite Data and Preprocessing

The experiment data include Landsat 8 surface reflectance (SR) products and MODIS surface reflectance products (MOD09A1). Because of the high spatial resolution of Landsat and the high temporal resolution of MODIS, they are widely used in spatiotemporal data fusion methods [19,30,38,63]. The datasets used in this paper were downloaded from Google Earth Engine (GEE). The Landsat 8 SR product contains five visible and near-infrared bands and two short-wave infrared (SWIR) bands. The dataset was processed to orthorectified surface reflectance with a spatial resolution of 30 m [64]. This product was generated using the Land Surface Reflectance Code (LaSRC). The MODIS surface reflectance product (MOD09A1) provides MODIS surface reflectance of seven bands at a resolution of 500 m, which was selected from the best L2G observations during an 8-day period [65]. The band information of Landsat and MODIS is shown in Table 1. The Advanced Spaceborne Thermal Emission and Reflection Radiometer global digital elevation model (ASTER GDEM) data were selected as digital elevation model (DEM) data.

ASTER GDEM provides a higher spatial resolution of 30 m and wider land coverage for estimating the extent of snow cover on the target dates [66].

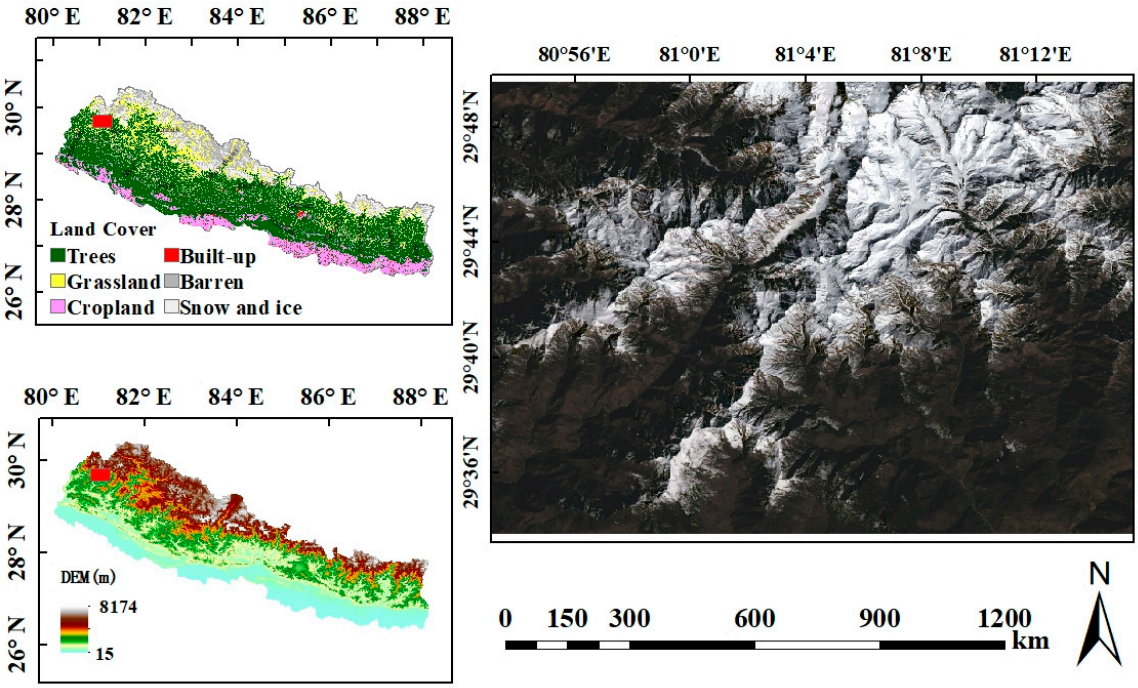

Figure 1. The location map of the study area. The top left image shows the land cover types of Nepal. The bottom left image shows the DEM of Nepal. The right image shows the RGB composites of the Landsat surface reflectance image on 12 February 2020.

Table 1. Information of corresponding bands of Landsat8 OLI, and MODIS.

Band	Landsat 8 OLI	Bandwidth (nm)	MODIS	Bandwidth (nm)
Blue	Band 2	450–510	Band 3	459–479
Green	Band 3	530–590	Band 4	545–565
Red	Band 4	630–690	Band 1	620–670
Near Infrared	Band 5	850–880	Band 2	841–876
Short-Wave Infrared 1 (SWIR1)	Band 6	1570–1650	Band 6	1628–1652
Short-Wave Infrared 2 (SWIR2)	Band 7	2110–2290	Band 7	2105–2155

The input data is the same as required by ESTARFM [29]. Our experiment requires two pairs of Landsat and MODIS SR images at the same date as well as a MODIS image at the target date as base images. A set of images with snow-cover changes from January to March were selected. Considering that the MODIS image is an eight-day composite product, the MODIS and Landsat images with the closest dates were selected. The images in January (t_1) and March (t_2) were used to predict the image in February (t_p). The data used in the experiment are listed in Table 2. The Landsat data in February were used as comparison data for accuracy verification. The cloud coverage of all the images is less than 5%, and the missing values were filled in with ENVI. The percentage of snow cover was calculated using the threshold method with NDSI > 0.4. Figure 2 shows the MODIS and Landsat surface reflectance images used in our experiment. It can be seen that the snow cover gradually decreases from January to March. MODIS images were resampled to the same spatial resolution of 30 m as Landsat. All the images were collected and cut to the same extent as the study area.

Table 2. Remote sensing data types and acquisition date.

Data Type	Spatial Resolution	Temporal Resolution	Acquisition Date	Expression	Use	Percentage of Snow	Percentage of Cloud
Landsat8 OLI	30 m	16 days	2020/01/11	F_{t_1}	Base Image	39.07	<5
	30 m		2020/02/12	F_{t_p}	Evaluation	25.44	<5
	30 m		2020/03/31	F_{t_2}	Base Image	22.71	<5
MODIS	500 m	daily	2020/01/09	C_{t_1}	Base Image	48.44	<5
	500 m		2020/02/10	C_{t_p}	Base Image	30.72	<5
	500 m		2020/03/29	C_{t_2}	Base Image	25.30	<5

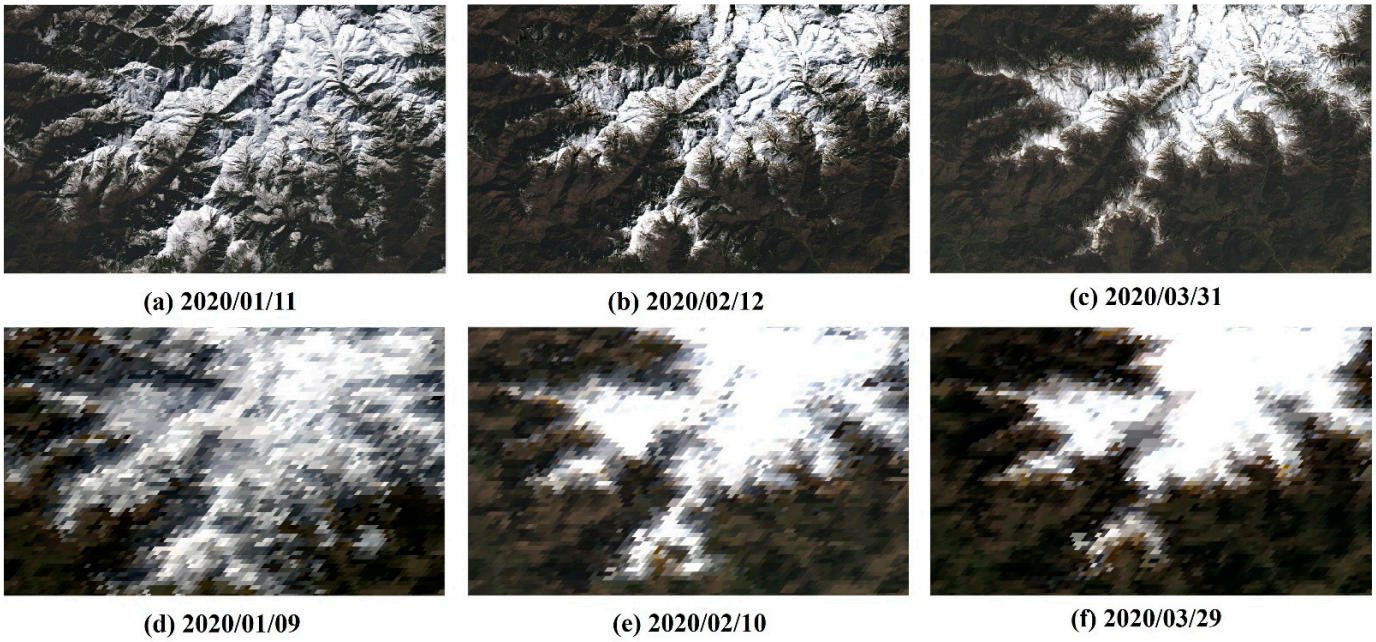

Figure 2. The RGB composites of MODIS and Landsat surface reflectance. (**a**–**c**) are Landsat surface reflectance images on 2020/01/11, 2020/02/12, and 2020/03/31, and the percentage of snow cover is 39.07%, 25.44%, and 22.71%. (**d**–**f**) are MODIS surface reflectance images on 2020/01/09, 2020/02/10, and 2020/03/29, and the percentage of snow cover is 48.44%, 30.72%, and 25.30%, respectively.

3. Methods

3.1. Description of the Improved ESTARFM

To describe the method more clearly, we explain some definitions in advance. Our experiment requires two pairs of Landsat and MODIS SR images at the same date as well as a MODIS image at the target date as base images. Landsat in January, February, and March are marked as F_{t_1}, F_{t_p}, and F_{t_2}. Meanwhile, MODIS in January, February, and March are marked as C_{t_1}, C_{t_p}, and C_{t_2}.

The flow chart of the improved ESTARFM (iESTARFM) is shown in Figure 3. Compared to the original ESTARFM, there are three improvements for snow-covered mountain regions in iESTARFM [42]. First, snow cover changes are simulated based on NDSI and DEM data by taking advantage of the spectral characteristics that snow has a high reflection in the visible bands and strong absorption in the shortwave infrared. A suitable threshold can distinguish between snow and non-snow pixels [50,53], and the details are shown in Section 3.2. DEM is important for estimating the volume change for inaccessible snow-covered mountain regions [9,13,14]. Considering that the variation in snow-covered mountain areas has a high correlation with elevation, it is feasible to simulate the snow cover based on a suitable DEM threshold, and the details are shown in Section 3.3. Second, similar pixels are selected according to the results of snow cover changes simulated by NDSI and DEM. Thirdly, the thresholds of selecting similar pixels are calculated separately

according to snow and non-snow pixels. To reduce the error caused by the misclassification, NDSI is added to select similar pixels and calculate the weights. The detailed descriptions of iESTARFM are given below. For more ESTARFM information please refer to [42].

(1) Simulate snow cover based on NDSI at the base date. The normalized difference snow index (NDSI) is widely used for snow identification by taking advantage of the spectral characteristics that snow has high reflectance in the green band and low reflectance in the short-wave infrared band. Based on this, the ratio of the two bands is calculated to highlight the characteristics of snow from others [53], and the calculation equation is as follows:

$$NDSI = \frac{\rho_{green} - \rho_{swir2}}{\rho_{green} + \rho_{swir2}} \quad (1)$$

where ρ_{green} is the reflectance of the green band, ρ_{swir2} is the reflectance of the SWIR2 band.

Firstly, by analyzing the NDSI distribution histogram and the true surface reflectance, the frequency histogram has two peaks, where the one with high values indicates the snow area and the one with low values indicates the non-snow area. Secondly, the NDSI threshold was determined by experimenting with different NDSI threshold values between the two peaks. Thirdly, the pixels with NDSI smaller than the threshold are considered as non-snow pixels and marked as 0, and those with NDSI larger than the threshold are considered as snow pixels and marked as 1. The experimental details are shown in Section 3.2, and the equation is as follows:

$$\begin{array}{ll} \text{If } |NDSI_F(x_i, y_i, t_k)| \leq \sigma_{snow} & Mask_{NDSI}(x_i,\ y_i, t_k) = 0 \\ \text{else} & Mask_{NDSI}(x_i, y_i, t_k) = 1 \end{array} \quad (2)$$

where σ_{snow} is the threshold to mask the snow-covered map, (x_i, y_i) is the coordinate of the ith pixel, t is the base date, t_k can be either t_1 or t_2, $Mask_{NDSI}(x_i,\ y_i, t_k)$ is the snow-covered mask, with 1 indicating snow pixels and 0 indicating non-snow pixels.

(2) Simulate snow cover based on DEM at the target date. The elevation is an important factor affecting the spatiotemporal distribution of the snow cover in mountainous areas [54–59]. The temperature at high altitudes is low, which is suitable for snow accumulation; meanwhile, the relatively high temperatures at lower altitudes cause snow to melt faster. The distribution of snow is strongly correlated with the elevation. Therefore, the combination of NDSI products and DEM is an effective approach for studying the spatial and temporal distribution of snow in mountain areas. Firstly, the coarse snow-covered borders are extracted from MODIS NDSI by using the local binary pattern (LBP) operator [67]. Then, the DEM values located at the boundaries are counted, and the threshold value of DEM is obtained. Thirdly, the pixels with DEM smaller than the threshold are considered non-snow pixels and marked as 0, and those with DEM larger than the threshold are considered snow pixels and marked as 1. The details are shown in Section 3.3, and the equation is as follows:

$$\begin{array}{ll} \text{If } DEM(x_i, y_i) \leq \sigma_{DEM} & Mask_{DEM}(x_i,\ y_i, t_p) = 0 \\ \text{else} & Mask_{DEM}(x_i,\ y_i, t_p) = 1 \end{array} \quad (3)$$

where σ_{DEM} is the threshold of DEM to classify snow and non-snow pixels, $Mask_{DEM}(x_i,\ y_i, t_p)$ is the mask obtained by thresholding at the target date, with 1 indicating snow pixels and 0 indicating non-snow pixels.

(3) Select similar pixels. ESTARFM selects similar pixels based on spectral similarity [42]. The threshold is determined by the standard deviation of a population of pixels from the base image with a high spatial resolution. The improved method differs from ESTARFM in that it does not calculate thresholds based on the whole image but calculates the threshold for the snow and non-snow pixels separately to reduce errors in selecting similar pixels. Meanwhile, more similar pixels have more similar

NDSI values, so NDSI is used as an additional condition to improve the accuracy of selecting similar pixels. Finally, similar pixels are selected based on spectral and NDSI differences. The equation is as follows:

$$|F(x_i, y_i, t_k, B) - F(x_{w/2}, y_{w/2}, t_k, B)| \leq \sigma_{flag}(B) \times 2/m \quad (4)$$

$$|NDSI(x_i, y_i, t_k) - NDSI(x_{w/2}, y_{w/2}, t_k)| \leq \sigma_{NDSI} \quad (5)$$

where F is the spectral reflectance of fine images, w is the size of the searching window, $\sigma(B)$ is the standard deviation of reflectance for band B, m is the estimated number of classes. Flag means the pixels marked as snow or non-snow, with flag = 1 indicating snow pixels and flag = 0 indicating non-snow pixels. $\sigma_{flag}(B)$ is the threshold for non-snow or snow pixels, and σ_{NDSI} is the threshold based on NDSI standard deviation.

(4) Calculate weights and the conversion coefficient. The weight calculation in ESTARFM involves spectral, temporal, and spatial distances. Similarly, in iESTARFM, the spectral distance is calculated using the correlation coefficient between the fine image and the coarse image at t_k. The spatial distance is calculated using the geographic distance between the similar pixels and target pixels. The temporal distance is calculated as the spectral difference between two coarse images. The conversion coefficient is the ratio of the change in the pure pixels in the fine image to the change in pixels in the coarse image, and it is calculated by a linear regression model. iESTARFM adds NDSI to the weight calculation to reduce the error of snow and non-snow identification. The weights equation is defined as follows:

$$R_i = \frac{E[(F_i - E(F_i))(C_i - E(C_i))]}{\sqrt{D(F_i)}\sqrt{D(C_i)}} \quad (6)$$

$$d_i = 1 + \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_i)^2}/(w/2) \quad (7)$$

$$NDSI_i = |NDSI(x_i, y_i, t_k) - NDSI(x_{w/2}, y_{w/2}, t_k)| + 1 \quad (8)$$

$$D_i = (1 - R_i) \times d_i \times NDSI_i \quad (9)$$

$$W_i = (1/D_i) / \sum_{i=1}^{N} (1/D_i) \quad (10)$$

where R_i is the spectral correlation coefficient for the ith pixel, d_i is the spatial distance, E is the expected value, $NDSI_i$ is the spatial distance, D is the variance, d_i is the spatial distance, D_i is the normalized reciprocal of weight, and W_i is the normalized weight.

(5) Predict the target pixels. The prediction of the target pixels can be divided into two cases. Case 1: the pixels are marked with the same type at t_1, t_2, and t_p, indicating the type has not changed, so the target pixels can be calculated based on t_1 and t_2. Case 2: the pixels are marked with the same type only on one date as t_p, indicating that there is a change, so the target pixels are predicted based on the pixels marked with the same type. The equation is as follows:

$$\text{CASE1}: Mask_{NDSI}(x_i, y_i, t_1) = Mask_{NDSI}(x_i, y_i, t_2) = Mask_{DEM}(x_i, y_i, t_p) \quad (11)$$

$$F(x_{w/2}, y_{w/2}, t_p, B) = T_1 \times F_1(x_{w/2}, y_{w/2}, t_p, B) + T_2 \times F_2(x_{w/2}, y_{w/2}, t_p, B) \quad (12)$$

$$\text{CASE2}: Mask_{NDSI}(x_i, y_i, t_k) = Mask_{DEM}(x_i, y_i, t_p) \quad (13)$$

$$F(x_{w/2}, y_{w/2}, t_p, B) = F(x_{w/2}, y_{w/2}, t_k, B) + \sum_{i=1}^{N} W_i \times V_i \times (C(x_i, y_i, t_P, B) - C(x_i, y_i, t_K, B)) \quad (14)$$

where $Mask_{NDSI}(x_i, y_i, t_k)$ is the snow-covered mask obtained by calculating the NDSI threshold, and $Mask_{DEM}(x_i, y_i, t_p)$ is the snow-covered mask obtained by calculating the DEM threshold. $F(x_{w/2}, y_{w/2}, t_p, B)$ is the predicted surface reflectance

of the target pixels, F_1 is the prediction result calculated from the base image at time t_1, and F_2 is the prediction result calculated from the base image at time t_2.

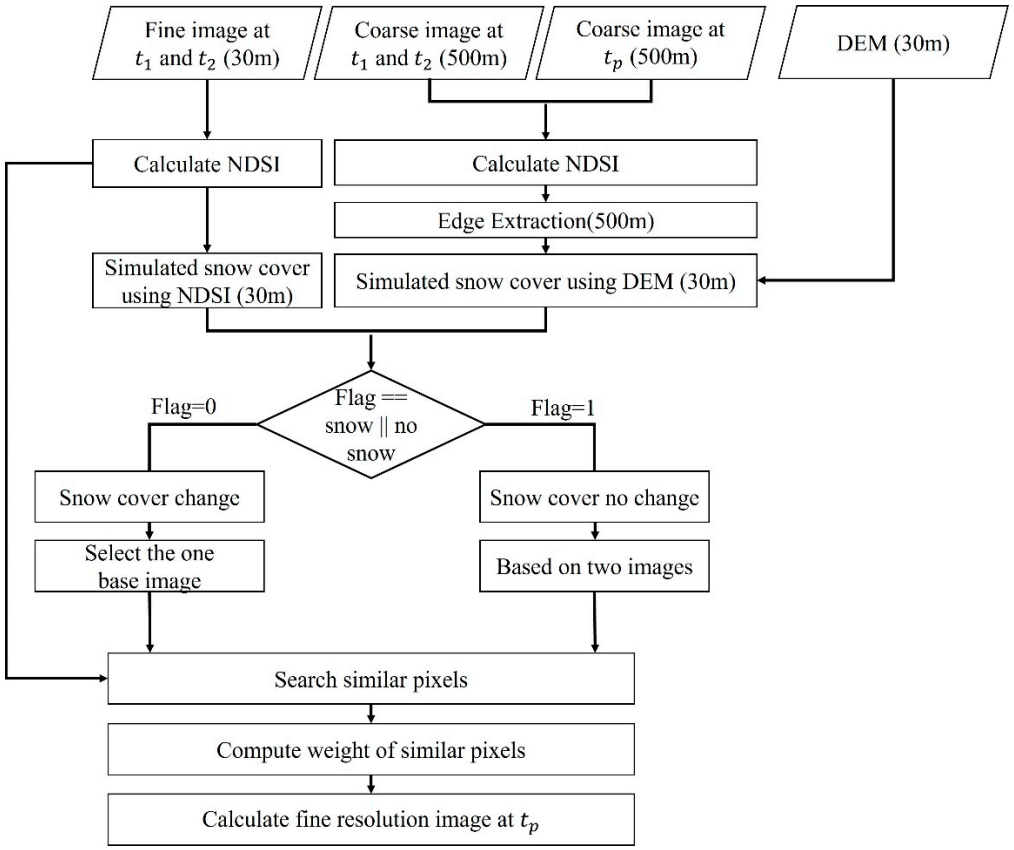

Figure 3. The flowchart of the iESTARFM algorithm.

3.2. Simulate Snow Cover Based on NDSI at the Base Date

The normalized snow index (NDSI) takes advantage of snow's high reflection in the visible bands and strong absorption in the shortwave infrared, so the selection of a suitable threshold can distinguish snow from other components [50,53]. When the NDSI of the pixels is greater than the threshold, the pixels are marked as snow; otherwise, they are marked as non-snow. In this study, five base images were selected for the experiment, including a pair of Landsat and MODIS images in January and March and a MODIS image in February, as shown in the first row of Figure 4a–e. The NDSI images corresponding to the five base images are shown in the second row of Figure 4f–j. The histogram of the distribution of NDSI values is shown in the third row of Figure 4k–o. By analyzing the histogram and surface reflectance images, it was found that the NDSI frequency distribution histogram has two peaks, where the peak of the low NDSI indicates the distribution of non-snow pixels and the peak of the high NDSI indicates the distribution of snow pixels. It can be found that the NDSI thresholds for snow-covered mapping are distributed between [−0.2, 0.5] for MODIS images and [0, 0.5] for Landsat images. Meanwhile, by analyzing the NDSI distribution histogram, the values that can distinguish between these two peaks are distributed in the range [0, 0.5]. Then, by experimenting with different NDSI threshold values, it was found that the NDSI threshold of 0.4 can distinguish snow and non-snow pixels well. Thus, the NDSI threshold of 0.4 was chosen to calculate the snow-covered mask, as shown in the fourth row of Figure 4p–t.

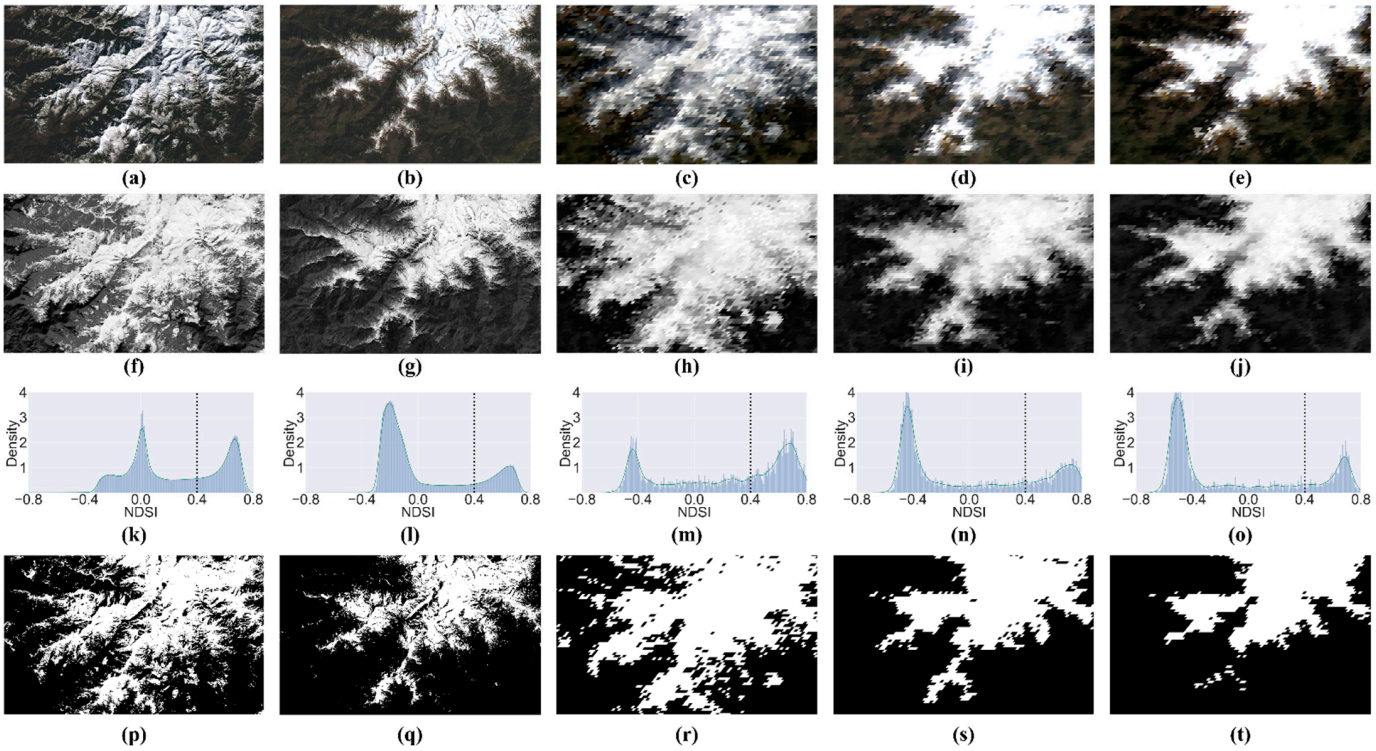

Figure 4. The process of generating the snow mask by using NDSI. Columns 1 and 2 show Landsat images at t_1 and t_2. Columns 3, 4, and 5 show MODIS images at t_1, t_p, and t_2. (**a**–**e**) are the surface reflectance images; (**f**–**j**) are the NDSI images; (**k**–**o**) are the NDSI frequency histogram; (**p**–**t**) are the snow mask based on a threshold of 0.4.

3.3. Simulate Snow Cover Based on DEM at the Target Date

To improve the prediction accuracy, it is necessary to obtain the type of pixels at the target time. However, the target data does not have a high-resolution image but only the MODIS low spatial resolution image. DEM is essential for analyzing the snow change in high mountain terrains [54]. As shown in Figure 5a–c are the true Landsat surface reflectance images on 2020/01/11, 2020/02/12, and 2020/03/31. There is an obvious change in snow melting from January to March. Figure 5d–f are the corresponding snow masks by setting different thresholds. It can be seen that there is a strong relationship between DEM and the snow cover boundary.

Considering that the variation in snow cover in mountainous areas has a high correlation with elevation, it is feasible to simulate the snow cover at the target time based on DEM data. Firstly, snow cover boundaries are extracted by using the coarse-resolution MODIS NDSI at the target time. The local binary pattern (LBP) operator is robust and computationally simple for texture analysis, and it applies the statistical and structural approach to texture analysis [68]. The LBP operator labels 3×3 neighborhood pixels for each central pixel of the image with a binary number by comparing gray values between the central pixel and each neighborhood pixel. In this way, the LBP operator can describe the structural information, thus providing excellent texture extraction for NDSI binary masks. The LBP operator is applied to the MODIS NDSI at the target time to extract the boundary of the snow-covered area, and the result is shown in Figure 6a. The equation is as follows:

$$LBP(x_c, y_c) = \sum_{P=0}^{P-1} 2^P s(i_p - i_c) \tag{15}$$

$$\text{where } s(x) = \begin{cases} 1 \; if \; x \geq 0 \\ 0 \; if \; x < 0 \end{cases} \tag{16}$$

where P is the pth pixel in the 3×3 window, i_p is the grayscale value of the pth pixel, i_c is the grayscale value of the central pixel.

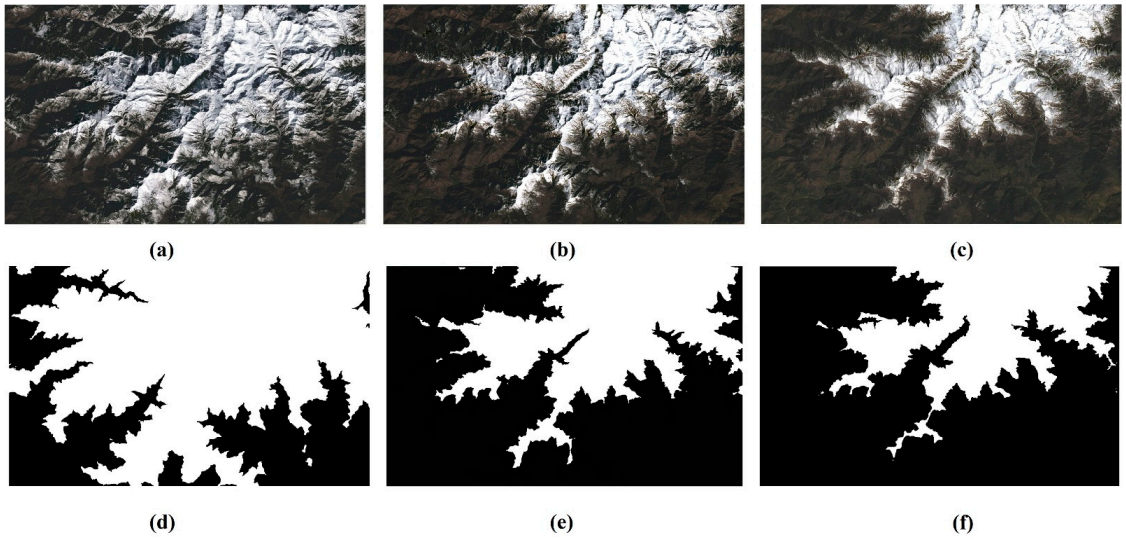

Figure 5. The Landsat surface reflectance image and snow mask using DEM. (**a**–**c**) are the true Landsat surface reflectance images on 2020/01/11, 2020/02/12, and 2020/03/31. (**d**–**f**) are the corresponding snow masks by setting different thresholds of DEM data. (**d**) A height of 3000 m is used as the threshold. (**e**) A height of 3600 m is used as the threshold. (**f**) A height of 3700 m is used as the threshold.

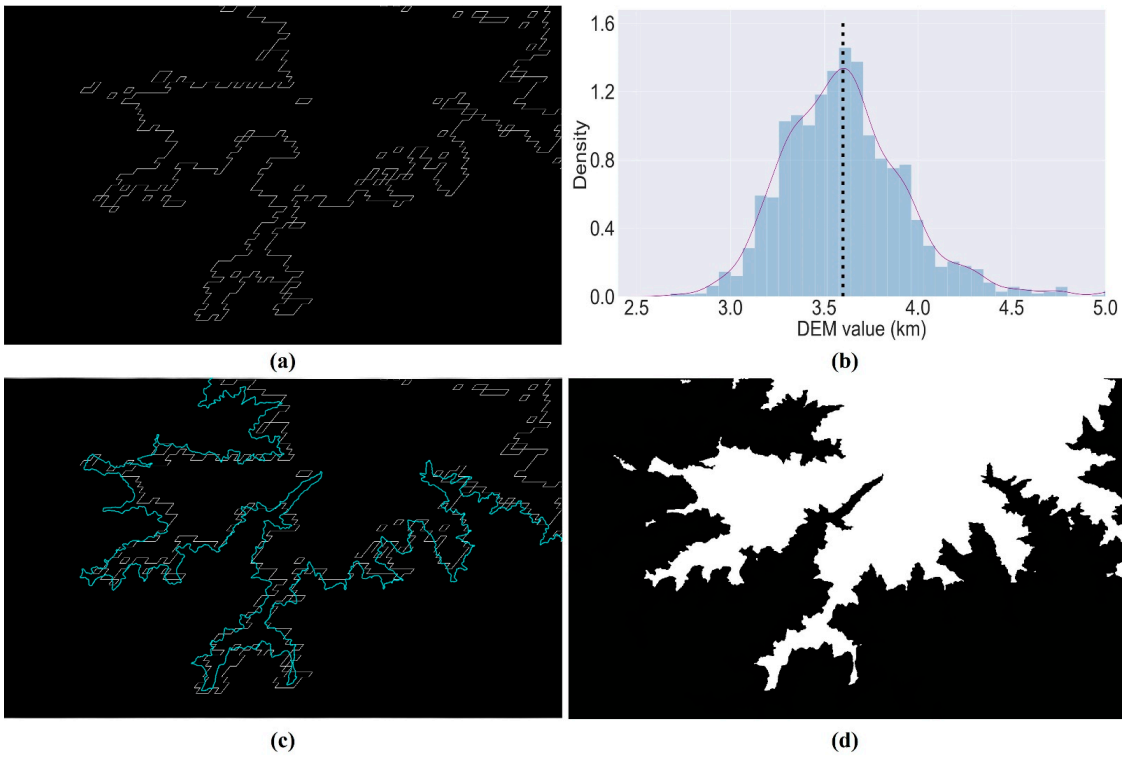

Figure 6. The process of making snow masks using DEM. (**a**) is the snow cover boundary extracted using MODIS NDSI. (**b**) is the frequency histogram of the DEM data located at the extracted snow cover boundary from MODIS NDSI. (**c**) The bright line is the extracted boundary overlaid on the DEM height of 3600 m. (**d**) is the snow mask extracted from DEM. The white part is snow, and the black part is non-snow.

Secondly, a statistical analysis is performed on the DEM data located at the boundary. The frequency histogram is shown in Figure 6b. It can be seen from the frequency histogram that the DEM height is mostly distributed in the range [3590, 3630]. So, the boundary extracted from MODIS NDSI is overlaid on the DEM height of 3600 m, as shown in Figure 6c. Finally, the DEM at a height of 3600 m is used as the threshold to extract the snow cover at the target time. To avoid misclassification of pixels at the boundary, this paper chooses a buffer of 500 m (one MODIS pixel). The snow-covered mask is shown in Figure 6d.

3.4. Data Quality Evaluation Metrics

Satellite images mainly contain spectral and spatial information. So, to quantitatively evaluate the accuracy of the proposed method, eight accuracy evaluation metrics including spectral and spatial evaluation metrics [27,69] were adopted to compare the predicted images with the true images [70]. The spectral accuracy metrics include mean square error (MAE), root mean square error (RMSE), the Pearson correlation coefficient (r), relative global dimensional synthesis error (ERGAS), Structural Similarity Index Measure (SSIM), Spectral Angle Mapper (SAM) and Peak Signal-to-Noise Ratio (PSNR). MAE and RMSE have similar meanings in the accuracy evaluation and are usually used to measure the difference between the predicted images and the true images. For MAE and RMSE, the value closer to 0 means that the predicted result is more similar to the real image, indicating a more accurate predicted result, while a larger value means that the predicted image deviates more from the real image. The Pearson correlation coefficient (r) indicates the linear relationship between the predicted image and the true image, and a value closer to 1 indicates a better correlation between the predicted and true values. ERGAS is used to evaluate the overall fusion result, and a value closer to zero indicates higher overall fidelity of the predicted image [71]. SSIM is an evaluation metric often used in computer vision to measure image similarity [72], which evaluates images in terms of luminance, contrast, and structure. In this study, this metric is used to evaluate the overall structural similarity of the image. A value closer to 1 indicates that the two images are more similar, and the larger the value, the better the image quality. SAM measures the spectral distortion of the fusion result. The smaller the value, the closer the image to the real image [73]. PSNR can evaluate the quality of the predicted result, and a higher value indicates that the predicted image has a better quality [74]. Spatial accuracy can be quantified by spatial characteristics, such as contrast and texture between the predicted and the true images. For spatial evaluation metrics, Robert's edge (Edge) was used to describe the spatial accuracy of the predicted images [75]. A value closer to 0 indicates a better image fusion result; a negative value indicates that the edge features are smoothed, and a positive value indicates that the edge features are sharpened. Table 3 shows the equations of accuracy metrics and the meaning of each variable.

Table 3. Equations for accuracy metrics and the meaning of each variable.

Metric Name	Equation	Variable Explanation
MSE	$MSE = \frac{1}{N}\sum_{i=1}^{N}(F_i - R_i)^2$	R_i: the value of the ith pixel in the true image
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(F_i - R_i)^2}{N}}$	F_i: the value of the ith pixel in the predicted image
r	$r = \frac{\sum_{i=1}^{N}(R_i - \mu_R)(F_i - \mu_F)}{\left(\sqrt{\sum_{i=1}^{N}(R_i - \mu_R)^2}\right)\left(\sqrt{\sum_{i=1}^{N}(F_i - \mu_F)^2}\right)}$	μ_F the mean pixel values of the predicted image
ERGAS	$ERGAS = 100\sqrt{\frac{1}{B}\sum_{j=1}^{B}\left(\frac{RMSE(F_i)}{\mu_j}\right)^2}$	μ_R: the mean pixel values of the true image
SSIM	$SSIM = \frac{(2\mu_R\mu_F + c_1)(2\mu_{RF} + c_2)}{(\mu_R{}^2 + \mu_F{}^2 + c_1)(\sigma_R{}^2 + \sigma_F{}^2 + c_2)}$	σ_F: the variance of pixel values of the predicted image

Table 3. *Cont.*

Metric Name	Equation	Variable Explanation
SAM	$SAM = arccos\frac{\sum(F_i R_i)}{\sqrt{R_i^2 F_i^2}}$	σ_R: the variance of pixel values of the true image
PSNR	$PSNR = 10log_{10}\frac{255^2}{MSE}$	c_1, c_2: constants N: the total number of pixels
Edge	$Edge = \left\|D_{i,j} - D_{i+1,j+1}\right\| + \left\|D_{i,j+1} - D_{i+1,j}\right\|$ $Edge\ error = R_{Edge} - F_{Edge}$	D: the value of ith pixel in the moving window

4. Results

4.1. Qualitative Comparison

To compare the proposed method with the original ESTARFM method, the same input was used for both algorithms. Three pairs of Landsat8 OLI and MODIS images were acquired on January, February, and March. Figure 2 shows the images using RGB composites. The two pairs of Landsat8 OLI and MODIS images acquired in January and March were used to predict the image at the Landsat spatial resolution in February. Then, the predicted image was compared with an actual Landsat image acquired in February to evaluate the performance of the two methods. Figure 7 shows the images predicted by the ESTARFM and iESTARFM, respectively. Figure 7a shows the actual Landsat image, where the major land cover types are snow and land located in high-altitude mountain areas. At high latitudes, large areas are covered by snow with high reflectance in visible bands. As the altitude decreases, the amount of snow gradually decreases. Figure 7b,c present the predicted results of the ESTARFM and iESTARFM methods. A zoom-in area in the first and third rows is used to highlight the details between the predicted image and the actual image.

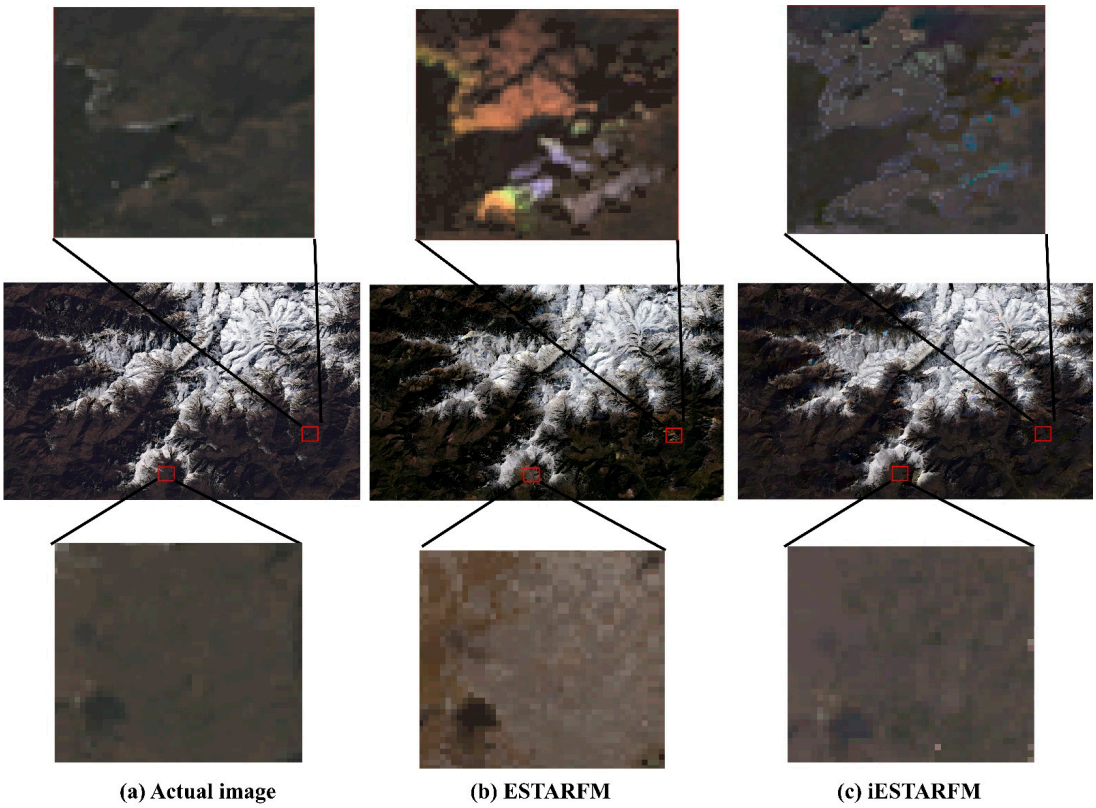

Figure 7. Comparison of the actual and predicted images. (**a**) is the actual image observed on 12 February 2020. (**b**) is the prediction image by ESTARFM. (**c**) is the prediction image by iESTARFM. The images of the first and third rows are the zoom-in areas of three images.

From the visual comparison of the overall image, the predicted snow cover boundaries of the two methods are similar to those of the actual Landsat image in Figure 7, indicating that the two methods can capture the major changes when the snow melts. However, both methods have common limitations, such as a blurring boundary between the snow-covered area and the land area and they also miss some tiny parts of the snow. Figure 7b presents the predicted image of ESTARFM, where bright abnormal patches are generated on the land area due to the wrong selection of similar pixels. The Figure 7c shows the predicted results of iESTARFM, where the over-bright patches and noise are reduced, and clearer texture structures are obtained. Furthermore, comparing the zoom-in area in the first row of the two methods in Figure 7b, it can be seen that bright noise is generated by the ESTARFM method, while iESTARFM can reduce such noise, and the predicted result is closer to the actual image. The third row of Figure 7 presents the areas with significant changes in snow melt. It can be seen that the ESTARFM cannot accurately predict the pixel values which the non-snow pixels are incorrectly predicted as snow. In contrast, iESTARFM is better at predicting the values. The visual comparison results indicate that the image predicted by iESTARFM is more similar to the actual image in terms of spatial details.

4.2. Quantitative Comparison

To further verify the effectiveness of the proposed method, eight accuracy metrics were used to evaluate the two methods, and the metrics are presented in Table 3. MSE, RMSE, r, ERGAS, SAM, PSNR, and SSIM were selected to evaluate the spectral accuracy, and they are sensitive to errors. EDGE was selected to evaluate spatial accuracy. The set of accuracy metrics was calculated for the two predicted results, where better prediction results are marked in bold, as shown in Table 4. The iESTARFM method performed better among the average of 6 bands of the evaluation metrics. To show the accuracy of evaluation results more clearly, a bar chart is used to show comparison results in Figure 8.

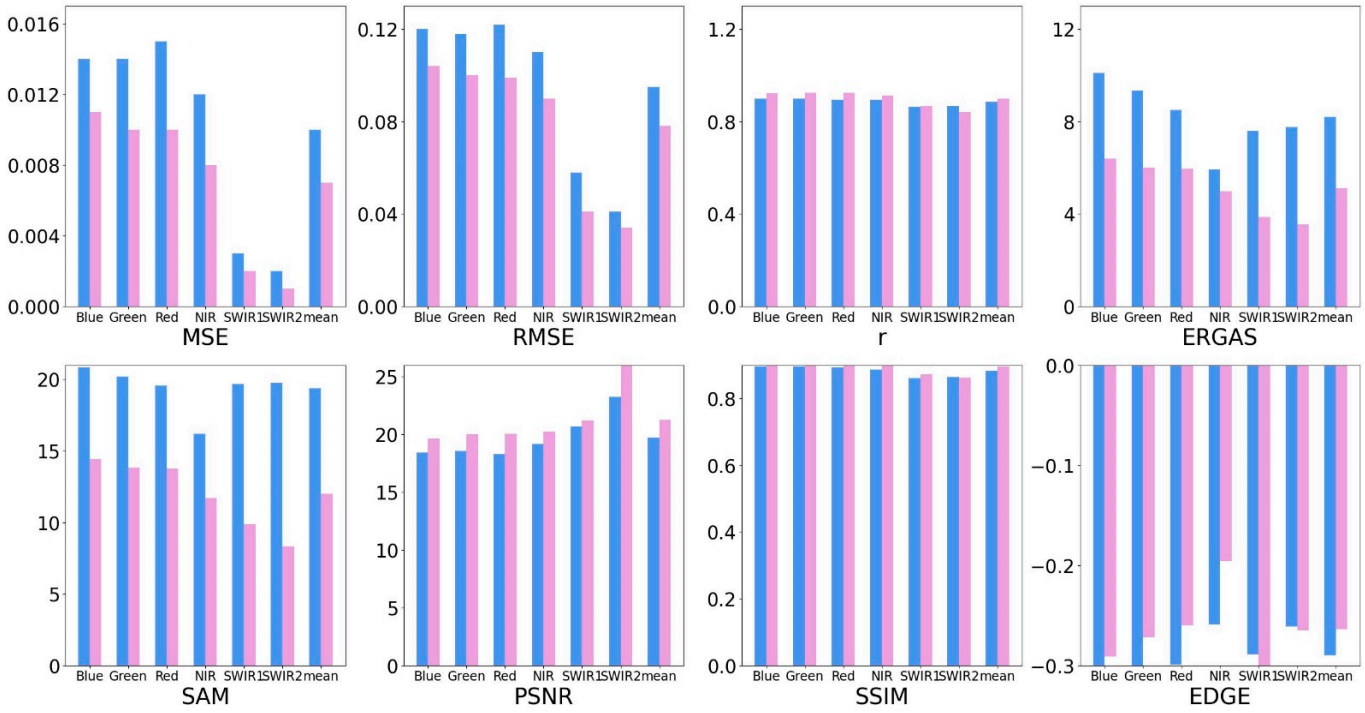

Figure 8. The accuracy evaluation results for ESTARFM and iESTARFM. The blue bars represent the result of the ESTARFM method, and the pink bars represent the result of the iESTARFM method.

Table 4. Accuracy assessment of the ESTARFM and iESTARFM methods.

Metrics	Method	Band						
		Blue	**Green**	**Red**	**NIR**	**SWIR1**	**SWIR2**	**Average**
MSE	ESTARFM	0.014	0.014	0.015	0.012	0.003	0.002	0.010
	iESTARFM	**0.011**	**0.010**	**0.010**	**0.008**	**0.002**	**0.001**	**0.007**
RMSE	ESTARFM	0.120	0.118	0.122	0.110	0.058	0.041	0.095
	iESTARFM	**0.104**	**0.100**	**0.099**	**0.090**	**0.041**	**0.034**	**0.078**
r	ESTARFM	0.898	0.898	0.895	0.894	0.864	0.867	0.886
	iESTARFM	**0.923**	**0.924**	**0.925**	**0.913**	**0.866**	**0.841**	**0.899**
ERGAS	ESTARFM	10.113	9.348	8.505	5.935	7.605	7.764	8.212
	iESTARFM	**6.389**	**6.013**	**5.964**	**4.985**	**3.881**	**3.555**	**5.131**
SAM	ESTARFM	20.825	20.195	19.567	16.188	19.675	19.75	19.367
	iESTARFM	**14.428**	**13.827**	**13.763**	**11.717**	**9.885**	**8.351**	**11.995**
PSNR	ESTARFM	18.422	18.569	18.291	19.162	20.708	23.263	19.736
	iESTARFM	**19.634**	**19.997**	**20.045**	**20.256**	**21.226**	**26.494**	**21.275**
SSIM	ESTARFM	0.896	0.896	0.893	0.887	0.861	0.865	0.883
	iESTARFM	**0.911**	**0.914**	**0.915**	**0.902**	**0.872**	**0.863**	**0.896**
EDGE	ESTARFM	−0.328	−0.307	−0.299	−0.259	−0.289	−0.261	−0.29
	iESTARFM	**−0.291**	**−0.272**	**−0.260**	**−0.196**	**−0.3**	**−0.265**	**−0.264**

In the comparison of spectral accuracy, the iESTARFM provided the most accurate predictions with the smallest MSE, RMSE, and ERGAS, as well as the highest r, except for the SWIR2 band. The MSE and RMSE indicate the deviation between the true and predicted values. The prediction results by iESTARFM have smaller accuracy values in the six bands, where the mean value of MSE is 0.007 (reduced by 0.003 as compared to 0.010) and the mean value of RMSE is 0.078 (reduced by 0.017 as compared to 0.095), indicating that the spectral values of iESTARFM are closer to the true value. r reflects the linear correlation between the predicted and true values, and a value closer to 1 indicates a higher correlation. The accuracy of iESTARFM is greater than 0.9 in the visible and near-infrared bands, but less than 0.9 in the two short-wave infrared bands. The mean value of r is 0.899 in the six bands (increased by 0.013 as compared to 0.886), indicating that iESTARFM has a higher correlation in visible and near-infrared bands but a lower correlation in shortwave infrared bands. To describe the linear relationship more clearly, the scatter plots of the predicted and true values for each band of ESTARFM and iESTARFM are illustrated in Figures 9 and 10. The scatter plots of ESTARFM are more discrete, and those of iESTARFM are more aggregated, indicating that the predictions of iESTARFM are closer to the actual values than the ESTARFM method. A smaller ERGAS value indicates a higher fidelity of the prediction results. The ERGAS of iESTARFM is less than 6.4 in all six bands, and that of ESTARFM is greater than 7.0. The mean value of iESTARFM is 5.131 (reduced by 3.081 as compared to 8.212), which indicates that iESTARFM has a better texture performance. SAM was used to calculate the similarity between two images. A smaller SAM value indicates a higher similarity between two images. The iESTARFM has a smaller SAM value of less than 15.0 in the six bands with a mean value of 11.995, and the SAM value of ESTARFM is greater than 18.0, indicating that iESTARFM has better reconstruction results. A larger PSNR value indicates better image quality. The mean PSNR value of iESTARFM is 21.275 and that of ESTARFM is 19.736 among the six bands, indicating that iESTARFM can obtain better image quality. The SSIM value closer to 1 indicates that the two images are more similar. The SSIM value of iESTARFM is greater than 0.9 in the visible and near-infrared bands, with a mean value of 0.896. iESTARFM has better accuracy than ESTARFM except for the SWIR2 band, and the prediction result has a higher structural similarity to the true value. In the comparison of spatial accuracy, the EGDE value closer to 0 indicates that the predicted image texture features are more similar to the true value. The Edge values of ESTARFM and iESTARFM are less than 0 in the six bands, indicating that the predicted image of both methods shows a smooth effect compared to the true image. The Edge value of iESTARFM is less than −0.299 and that of ESTARFM is greater than −0.299 in the visible

and near-infrared bands, and for the two SWIR bands, ESTARFM has a smaller Edge value than iESTARFM. The mean Edge value of ESTARFM is −0.29, and that of iESTARFM is −0.264. Generally, iESTARFM can obtain more similar texture features.

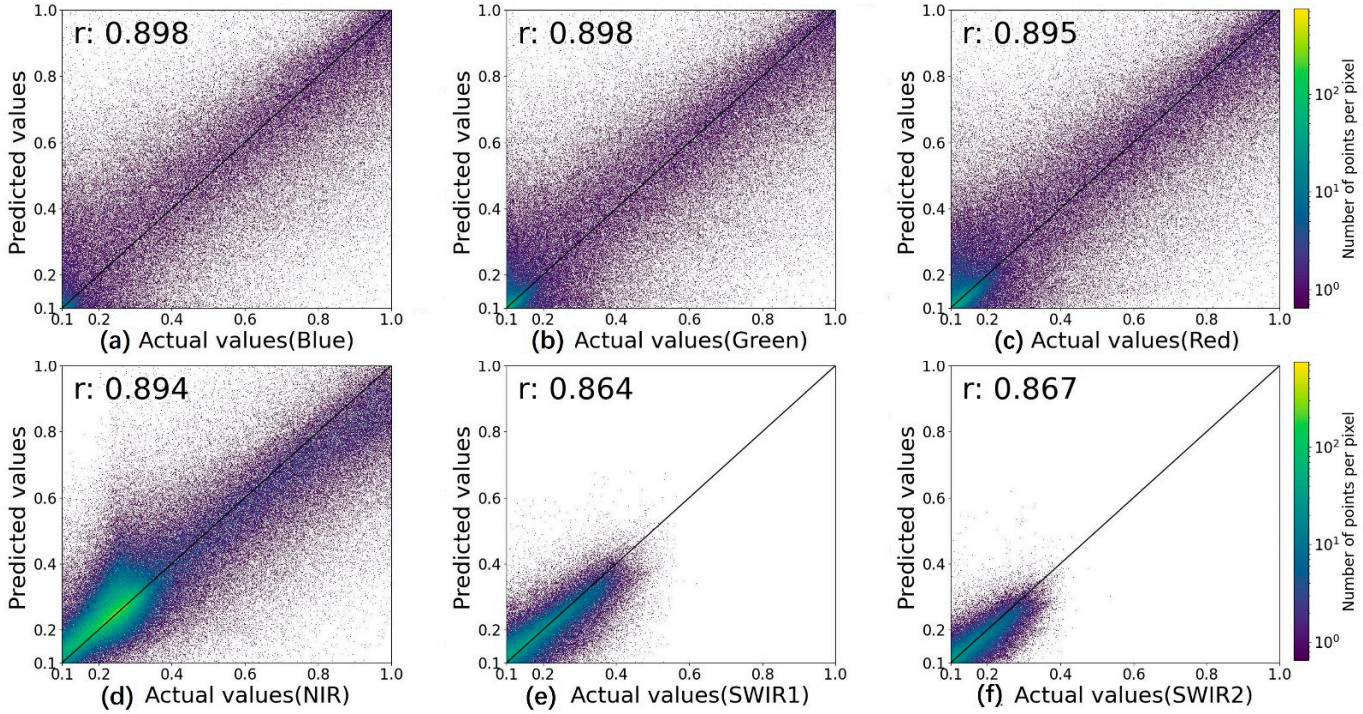

Figure 9. Scatter plots of the actual and predicted values for the six bands of ESTARFM. ((**a**–**f**) for different bands and the dark line is a 1:1 line).

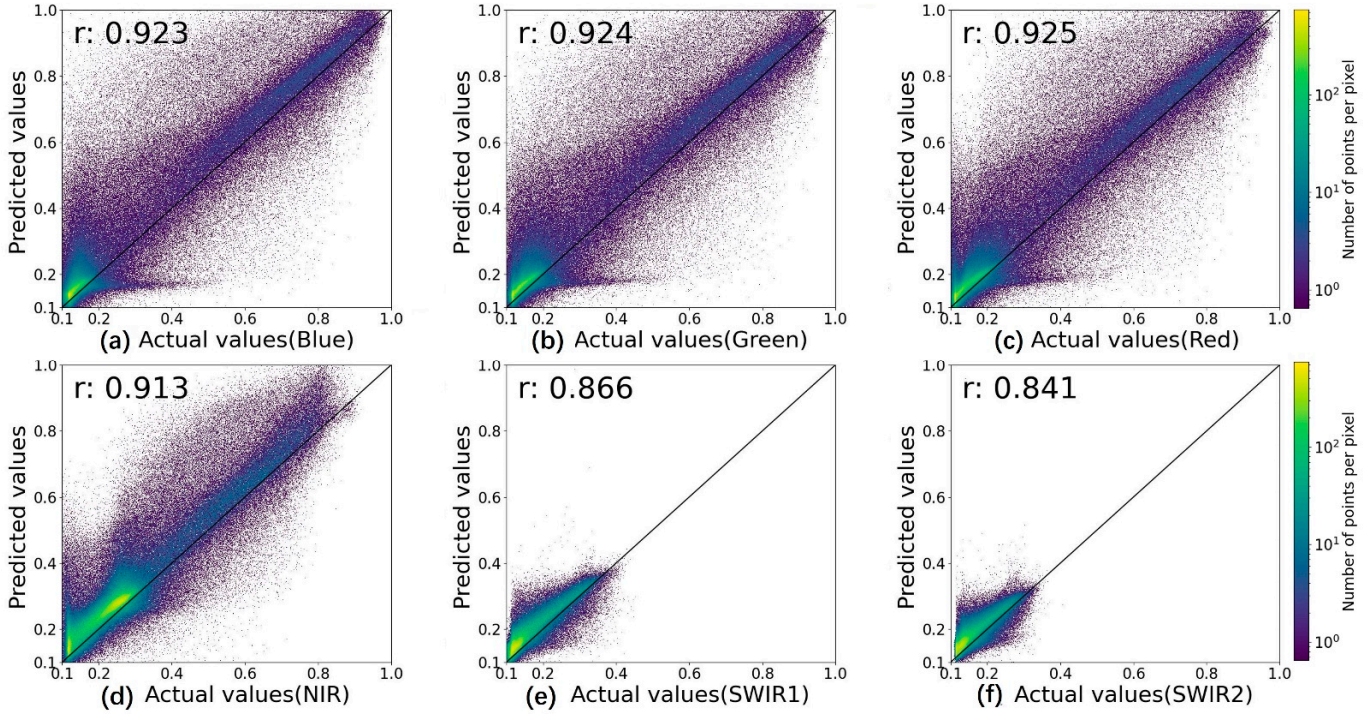

Figure 10. Scatter plots of the actual and predicted values for the six bands of iESTARFM. ((**a**–**f**) for different bands and the dark line is a 1:1 line).

5. Discussion

The change in snow cover significantly affects the exchange of energy between the atmosphere and land surface, which is important for theoretical studies and practical applications. However, due to technical and environmental limitations, it is hard to obtain images with both high spatial and temporal resolution just using a single satellite [16], especially in the mountain region, where the spatial and temporal variability of the snow-cover is particularly high [17]. The ESTARFM assumes no abrupt changes in the surface type, which limits its application in snow-covered mountain areas. In fact, the surface type may change abruptly due to snow and snow melting. This may cause problems with ESTARFM in the selection of similar pixels. The threshold of selecting similar pixels will be higher due to the presence of snow, which makes the wrong selection of pixels. Thus, the iESTARFM method has made some improvements based on the original ESTARFM method, considering the optical characteristics of snow and the high correlation between the change in snow cover and the elevation. The main idea of iESTARFM is to introduce NDSI and DEM to simulate the change in snow cover. By qualitative and quantitative comparison, the experiment found that iESTARFM has higher accuracy than ESTARFM

Although iESTARFM can predict good results, there are still several limitations and constraints. Firstly, we calculated the NDSI of the base images. The NDSI threshold is based on the histogram statistical method and surface reflectance. Although it can be used effectively and accurately to estimate snow cover information from satellite images, the disadvantage is that this approach is relatively subjective. The closer the snow-cover is to the real surface, the more accurate the selection of similar pixels. So, it is necessary to explore more accurate snow-cover mapping methods, such as the decision-tree-based classification model [76], the supervised fuzzy classification approach [77] and subpixel snow-cover mapping for automated mapping of snow cover. Secondly, we use DEM to simulate the snow cover on the target date. Although we can infer the change in snow cover from elevation, the simulation accuracy is still limited because there are many factors related to the snow-covered change, such as daytime air temperature, distance to significant open water bodies, topographic roughness and aspect, forest cover, and snow class. There is blurring at the boundaries of the snow-covered area because snow covers cannot be accurately identified based on NDSI and DEM. Finally, iESTARFM performs well in the visible and near-infrared bands, but in two shortwave infrared bands, the r, SSIM, and EDGE values are lower than those of ESTARFM. This may be because compared to non-snow pixels, the snow pixels have a particularly high reflectance in the visible and near-infrared bands, which can be easily distinguished. Furthermore, it is difficult to separate glacier from snow using optical remote sensing images.

In the future, there will still be much work to do to improve the accuracy of our method. Firstly, other useful information such as temperature data, slope, and aspect can be considered to obtain the snow change information. Secondly, the datasets with a higher resolution can be considered to capture more details of ground objects. Thirdly, it is necessary to find other study areas. Three conditions need to be met to select the study area. First, MODIS and Landsat images should be at the same time and place at high altitude. Second, the images should be largely free of clouds and shadows. Thirdly, there is a gradual trend of snow-covered change. Thus, the reliability of the method does need to be explored in more study areas. We will experiment with more similar study areas to explore the applicability of our approach and compare it with more spatiotemporal algorithms, as well as consider the advantages of other algorithms to improve the accuracy of data fusion not just in the high-altitude snow areas.

6. Conclusions

The spatial and temporal distribution of snow is important to the changes in climate. However, due to the limitations of technology and atmospheric conditions, it is hard to obtain images with both high spatial and temporal resolution just using a single satellite [16], especially in mountain regions. Thus, the iESTARFM is developed for snow-covered

mountain areas. The main idea of this method is to improve the accuracy of selecting similar pixels by introducing NDSI and DEM information that can simulate the change in snow cover. There are three main steps. Firstly, simulate snow-cover changes using NDSI and DEM information. Secondly, select similar pixels according to the results of snow-covered changes. Thirdly, the thresholds are calculated separately according to the pixels type. The prediction results of ESTARFM and iESTARFM methods are evaluated qualitatively and quantitatively. For the visual evaluation, both algorithms can simulate snow cover boundaries. However, the ESTARFM method could generate bright abnormal patches in the land area due to the wrong selection of similar pixels, while the iESTARFM made good predictions in the land area. For the quantitative analysis, eight evaluation metrics commonly used for the spatiotemporal fusion method were selected. The iESTARFM has better accuracy than ESTARFM in the visible, NIR, and SWIR bands. In addition to SWIR1 in r and EDGE evaluation metrics, SWIR2 in SSIM evaluation metric, ESTARFM has a better performance than iESTARFM. From the scatter plots iESTARFM is more focused and has a higher correlation coefficient. The correlation coefficients are greater than 0.9 in the visible and NIR bands, and less than 0.87 in short-wave infrared bands. Because snow has higher reflectance in the visible and near-infrared bands than short-wave infrared bands, the values have a larger distribution from 0 to 1 in the visible bands, and the values between 0 and 0.5 in the short-wave infrared bands. The evaluation metrics perform well in the visible and near-infrared bands, probably because the reflectance characteristics of snow are more distinguished compared to other objects in the visible and near-infrared bands. In the future, this method could be used to generate dense time series images for snow-covered mountain areas.

Author Contributions: Conceptualization, M.G., Y.L. and X.W.; methodology, M.G.; software, M.G.; validation, M.G., Y.L. and X.W. and Y.Z.; writing—review and editing, M.G., H.Y., M.L., C.W., X.G. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: The study is funded by the National Key R&D Program of China (2020YFE0200700 and 2019YFE0127300), the National Natural Science Foundation of China (41901367), and the Major Special Project the China High-Resolution Earth Observation System (30-Y30F06-9003-20/22).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank X. Zhu for making available the algorithm on the internet for our empirical analysis and comparison. The authors would like to thank the colleagues for their company and help. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

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

References

1. Pradhananga, D.; Pomeroy, J.W. Diagnosing changes in glacier hydrology from physical principles using a hydrological model with snow redistribution, sublimation, firnification and energy balance ablation algorithms. *J. Hydrol.* **2022**, *608*, 127545. [CrossRef]
2. Guo, S.; Du, P.; Xia, J.; Tang, P.; Wang, X.; Meng, Y.; Wang, H. Spatiotemporal changes of glacier and seasonal snow fluctuations over the Namcha Barwa–Gyala Peri massif using object-based classification from Landsat time series. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 21–37. [CrossRef]
3. Jin, H.; Chen, X.; Zhong, R.; Wu, P.; Ju, Q.; Zeng, J.; Yao, T. Extraction of snow melting duration and its spatiotemporal variations in the Tibetan Plateau based on MODIS product. *Adv. Space Res.* **2022**, *70*, 15–34. [CrossRef]
4. Ahluwalia, R.S.; Rai, S.P.; Meetei, P.N.; Kumar, S.; Sarangi, S.; Chauhan, P.; Karakoti, I. Spatial-diurnal variability of snow/glacier melt runoff in glacier regime river valley: Central Himalaya, India. *Quat. Int.* **2021**, *585*, 183–194. [CrossRef]
5. You, Q.; Cai, Z.; Pepin, N.; Chen, D.; Ahrens, B.; Jiang, Z.; Wu, F.; Kang, S.; Zhang, R.; Wu, T. Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences. *Earth-Sci. Rev.* **2021**, *217*, 103625. [CrossRef]
6. Guo, D.; Wang, H. The significant climate warming in the northern Tibetan Plateau and its possible causes. *Int. J. Climatol.* **2012**, *32*, 1775–1781. [CrossRef]

7. Wu, G.; Liu, Y.; Zhang, Q.; Duan, A.; Wang, T.; Wan, R.; Liu, X.; Li, W.; Wang, Z.; Liang, X. The influence of mechanical and thermal forcing by the Tibetan Plateau on Asian climate. *J. Hydrometeorol.* **2007**, *8*, 770–789. [CrossRef]
8. Zhang, H.; Immerzeel, W.W.; Zhang, F.; de Kok, R.J.; Chen, D.; Yan, W. Snow cover persistence reverses the altitudinal patterns of warming above and below 5000 m on the Tibetan Plateau. *Sci. Total Environ.* **2022**, *803*, 149889. [CrossRef]
9. Chen, Y.; Ge, Y.; Heuvelink, G.B.M.; An, R.; Chen, Y. Object-Based Superresolution Land-Cover Mapping From Remotely Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 328–340. [CrossRef]
10. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853. [CrossRef]
11. Gong, P.; Li, X.; Zhang, W. 40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. *Sci. Bull.* **2019**, *64*, 756–763. [CrossRef]
12. Chen, Y.; Shi, K.; Ge, Y.; Zhou, Y. Spatiotemporal Remote Sensing Image Fusion Using Multiscale Two-Stream Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
13. Kakareka, S.; Kukharchyk, T.; Kurman, P. Trace and major elements in surface snow and fresh water bodies of the Marguerite Bay Islands, Antarctic Peninsula. *Polar Sci.* **2022**, *32*, 100792. [CrossRef]
14. Yan, D.; Ma, N.; Zhang, Y. Development of a fine-resolution snow depth product based on the snow cover probability for the Tibetan Plateau: Validation and spatial–temporal analyses. *J. Hydrol.* **2022**, *604*, 127027. [CrossRef]
15. Wang, X.; Wu, C.; Peng, D.; Gonsamo, A.; Liu, Z. Snow cover phenology affects alpine vegetation growth dynamics on the Tibetan Plateau: Satellite observed evidence, impacts of different biomes, and climate drivers. *Agric. For. Meteorol.* **2018**, *256–257*, 61–74. [CrossRef]
16. Gao, F.; Masek, J.; Schwaller, M.; Hall, 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.
17. De Gregorio, L.; Callegari, M.; Marin, C.; Zebisch, M.; Bruzzone, L.; Demir, B.; Strasser, U.; Marke, T.; Gunther, D.; Nadalet, R.; et al. A Novel Data Fusion Technique for Snow Cover Retrieval. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2873–2888. [CrossRef]
18. Ju, J.; Roy, D.P. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* **2008**, *112*, 1196–1211. [CrossRef]
19. 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]
20. 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]
21. Zhang, B.; Zhang, L.; Xie, D.; Yin, X.; Liu, C.; Liu, G. Application of Synthetic NDVI Time Series Blended from Landsat and MODIS Data for Grassland Biomass Estimation. *Remote Sens.* **2016**, *8*, 10. [CrossRef]
22. Jia, K.; Liang, S.; Wei, X.; Yao, Y.; Su, Y.; Jiang, B.; Wang, X. Land Cover Classification of Landsat Data with Phenological Features Extracted from Time Series MODIS NDVI Data. *Remote Sens.* **2014**, *6*, 11518–11532. [CrossRef]
23. Chen, B.; Huang, B.; Xu, B. Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 27–39. [CrossRef]
24. Dhillon, M.S.; Dahms, T.; Kübert-Flock, C.; Steffan-Dewenter, I.; Zhang, J.; Ullmann, T. Ullmann, Spatiotemporal Fusion Modelling Using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria. *Remote Sens.* **2022**, *14*, 677. [CrossRef]
25. Zhou, J.; Chen, J.; Chen, X.; Zhu, X.; Qiu, Y.; Song, H.; Rao, Y.; Zhang, C.; Cao, X.; Cui, X. Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sens. Environ.* **2021**, *252*, 112130. [CrossRef]
26. Zhu, X.L.; Cai, F.Y.; Tian, J.Q.; Williams, T.K.A. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sens.* **2018**, *10*, 527. [CrossRef]
27. Zhu, X.; Zhan, W.; Zhou, J.; Chen, X.; Liang, Z.; Xu, S.; Chen, J. A novel framework to assess all-round performances of spatiotemporal fusion models. *Remote Sens. Environ.* **2022**, *274*, 113002. [CrossRef]
28. Chen, B.; Huang, B.; Xu, B. Comparison of Spatiotemporal Fusion Models: A Review. *Remote Sens.* **2015**, *7*, 1798–1835. [CrossRef]
29. Ping, B.; Meng, Y.; Su, F. An enhanced spatial and temporal adaptive reflectance fusion model based on optimal window. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017. [CrossRef]
30. Emelyanova, I.; Mcvicar, T.; Niel, T.V.; Li, L.; Dijk, A.V. *On Blending Landsat-MODIS Surface Reflectances in Two Landscapes with Contrasting Spectral, Spatial and Temporal Dynamics*; CSIRO: Canberra, Australia, 2012. [CrossRef]
31. Zhukov, B.; Oertel, D.; Lanzl, F.; Reinhackel, G. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [CrossRef]
32. 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. [CrossRef]
33. Xue, J.; Leung, Y.; Fung, T. A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. *Remote Sens.* **2017**, *9*, 1310. [CrossRef]

34. Jia, D.; Song, C.; Cheng, C.; Shen, S.; Ning, L.; Hui, C. A novel deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions using a two-stream convolutional neural network. *Remote Sens.* **2020**, *12*, 698. [CrossRef]
35. Wang, X.; Wang, X. Spatiotemporal fusion of remote sensing image based on deep learning. *J. Sens.* **2020**, *2020*, 8873079. [CrossRef]
36. Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J.; Lefsky, M.A. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [CrossRef]
37. Jia, D.; Cheng, C.; Song, C.; Shen, S.; Ning, L.; Zhang, T. A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions. *Remote Sens.* **2021**, *13*, 645. [CrossRef]
38. Gevaert, C.M.; García-Haro, F.J. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [CrossRef]
39. Ma, J.; Zhang, W.; Marinoni, A.; Gao, L.; Zhang, B. Performance assessment of ESTARFM with different similar-pixel identification schemes. *J. Appl. Remote Sens.* **2018**, *12*, 025017. [CrossRef]
40. Liu, M.; Liu, X.; Dong, X.; Zhao, B.; Zou, X.; Wu, L.; Wei, H. An improved spatiotemporal data fusion method using surface heterogeneity information based on estarfm. *Remote Sens.* **2020**, *12*, 3673. [CrossRef]
41. Knauer, K.; Gessner, U.; Fensholt, R.; Kuenzer, C. An ESTARFM fusion framework for the generation of large-scale time series in cloud-prone and heterogeneous landscapes. *Remote Sens.* **2016**, *8*, 425. [CrossRef]
42. 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]
43. Ghosh, R.; Gupta, P.K.; Tolpekin, V.; Srivastav, S.K. An enhanced spatiotemporal fusion method—Implications for coal fire monitoring using satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102056. [CrossRef]
44. Bazrgar Bajestani, A.; Akhoondzadeh Hanzaei, M. ESTARFM Model for Fusion of LST Products of MODIS and ASTER Sensors to Retrieve the High Resolution Land Surface Temperature Map. *J. Geomat. Sci. Technol.* **2018**, *7*, 147–161.
45. Chen, M.; Li, C.; Guan, Y.; Zhou, J.; Wang, D.; Luo, Z. Generation and application of high temporal and spatial resolution images of regional farmland based on ESTARFM model. *Acta Agron. Sin.* **2019**, *45*, 1099–1110. [CrossRef]
46. Da Silva, B.B.; Mercante, E.; Kusminski, D.; Cattani, C.E.V.; Mendes, I.d.S.; Caon, I.L.; Ganascini, D.; Prior, M. Synthetic images to map daily evapotranspiration in field scale using SEBAL model and ESTARFM algorithm. *Aust. J. Crop Sci.* **2020**, *14*, 504–509. [CrossRef]
47. Liu, W.; Zeng, Y.; Li, S.; Pi, X.; Huang, W. An improved spatiotemporal fusion approach based on multiple endmember spectral mixture analysis. *Sensors* **2019**, *19*, 2443. [CrossRef]
48. Zhang, H.; Sun, Y.; Shi, W.; Guo, D.; Zheng, N. An object-based spatiotemporal fusion model for remote sensing images. *Eur. J. Remote Sens.* **2021**, *54*, 86–101. [CrossRef]
49. Liu, M.; Liu, X.; Wu, L.; Zou, X.; Jiang, T.; Zhao, B. A modified spatiotemporal fusion algorithm using phenological information for predicting reflectance of paddy rice in southern China. *Remote Sens.* **2018**, *10*, 772. [CrossRef]
50. Wang, X.; Xie, H. New methods for studying the spatiotemporal variation of snow cover based on combination products of MODIS Terra and Aqua. *J. Hydrol.* **2009**, *371*, 192–200. [CrossRef]
51. Tarigan, D.G.; Isa, S.M. A PSNR Review of ESTARFM Cloud Removal Method with Sentinel 2 and Landsat 8 Combination. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 189–198. [CrossRef]
52. Nietupski, T.C.; Kennedy, R.E.; Temesgen, H.; Kerns, B.K. Spatiotemporal image fusion in Google Earth Engine for annual estimates of land surface phenology in a heterogenous landscape. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *99*, 102323. [CrossRef]
53. Lin, J.; Feng, X.; Xiao, P.; Li, H.; Wang, J.; Li, Y. Comparison of snow indexes in estimating snow cover fraction in a mountainous area in northwestern China. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 725–729. [CrossRef]
54. Pandey, A.C.; Ghosh, S.; Nathawat, M.S.; Tiwari, R.K. Area Change and Thickness Variation over Pensilungpa Glacier (J&K) using Remote Sensing. *J. Indian Soc. Remote Sens.* **2012**, *40*, 245–255. [CrossRef]
55. Bhambri, R.; Bolch, T. Glacier mapping: A review with special reference to the Indian Himalayas. *Prog. Phys. Geogr. Earth Environ.* **2009**, *33*, 672–704. [CrossRef]
56. Rignot, E.; Rivera, A.; Casassa, G. Contribution of the Patagonia Icefields of South America to sea level rise. *Science* **2003**, *302*, 434–437. [CrossRef] [PubMed]
57. Surazakov, A.B.; Aizen, V.B. Estimating volume change of mountain glaciers using SRTM and map-based topographic data. *Ieee Trans. Geosci. Remote Sens.* **2006**, *44*, 2991–2995. [CrossRef]
58. Berthier, E.; Arnaud, Y.; Kumar, R.; Ahmad, S.; Wagnon, P.; Chevallier, P. Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India). *Remote Sens. Environ.* **2007**, *108*, 327–338. [CrossRef]
59. Bolch, T.; Kamp, U. Glacier Mapping in High Mountains Using DEMs, Landsat and ASTER Data. 2005. Available online: https://www.semanticscholar.org/paper/Glacier-mapping-in-high-mountains-using-DEMs%2C-and-Bolch-Kamp/123fcb9070bb27cce3b899bd53cbe787931bf25a (accessed on 10 May 2020).
60. Campbell, B. Biodiversity, livelihoods and struggles over sustainability in Nepal. *Landsc. Res.* **2018**, *43*, 1056–1067. [CrossRef]
61. Paudel, B.; Zhang, Y.-l.; Li, S.-c.; Liu, L.-s.; Wu, X.; Khanal, N.R. Review of studies on land use and land cover change in Nepal. *J. Mt. Sci.* **2016**, *13*, 643–660. [CrossRef]
62. Shrestha, A.B.; Joshi, S.P. Snow cover and glacier change study in Nepalese Himalaya using remote sensing and geographic information system. *J. Hydrol. Meteorol.* **2009**, *6*, 26–36. [CrossRef]

63. Emelyanova, I.V.; McVicar, T.R.; van Niel, T.G.; Li, L.T.; van Dijk, A.I.J.M. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sens. Environ.* **2013**, *133*, 193–209. [CrossRef]
64. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef] [PubMed]
65. Vermote, E. MYD09A1 MODIS/Aqua Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 [Data set]. 2015. Available online: https://lpdaac.usgs.gov/products/myd09a1v006/ (accessed on 10 May 2020).
66. Zhao, S.; Cheng, W.; Zhou, C.; Chen, X.; Zhang, S.; Zhou, Z.; Liu, H.; Chai, H. Accuracy assessment of the ASTER GDEM and SRTM3 DEM: An example in the Loess Plateau and North China Plain of China. *Int. J. Remote Sens.* **2011**, *32*, 8081–8093. [CrossRef]
67. Zhu, C.; Wang, R. Local multiple patterns based multiresolution gray-scale and rotation invariant texture classification. *Inf. Sci.* **2012**, *187*, 93–108. [CrossRef]
68. Zhou, Y.; Jiang, H.; Wang, Z.; Yang, X.; Geng, E. Assessment of Four Typical Topographic Corrections in Landsat Tm Data for Snow Cover Areas. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 157–162. [CrossRef]
69. Pan, Z.; Ma, W.; Guo, J.; Lei, B. Super-resolution of single remote sensing image based on residual dense backprojection networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7918–7933. [CrossRef]
70. Johnson, A.P.; Baker, C.L. First-and second-order information in natural images: A filter-based approach to image statistics. *JOSA A* **2004**, *21*, 913–925. [CrossRef]
71. Qian, D.; Gungor, O.; Jie, S. Performance evaluation for pan-sharpening techniques. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2005, IGARSS '05, Seoul, Korea, 29 July 2005. [CrossRef]
72. Li, X.; Foody, G.M.; Boyd, D.S.; Ge, Y.; Zhang, Y.; Du, Y.; Ling, F. SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. *Remote Sens. Environ.* **2020**, *237*, 111537. [CrossRef]
73. Yuhas, R.H.; Goetz, A.F.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In Proceedings of the JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop, Pasadena, CA, USA, 1–5 June 1992; Volume 1: AVIRIS Workshop.
74. Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Carli, M. Modified image visual quality metrics for contrast change and mean shift accounting. In Proceedings of the 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Polyana, Ukraine, 23–25 February 2011.
75. Li, W.; Cao, D.; Peng, Y.; Yang, C. MSNet: A multi-stream fusion network for remote sensing spatiotemporal fusion based on transformer and convolution. *Remote Sens.* **2021**, *13*, 3724. [CrossRef]
76. Zhu, L.; Xiao, P.; Feng, X.; Zhang, X.; Wang, Z.; Jiang, L. Support vector machine-based decision tree for snow cover extraction in mountain areas using high spatial resolution remote sensing image. *J. Appl. Remote Sens.* **2014**, *8*, 084698. [CrossRef]
77. Pepe, M.; Boschetti, L.; Brivio, P.A.; Rampini, A. Accuracy benefits of a fuzzy classifier in remote sensing data classification of snow. In Proceedings of the 2007 IEEE International Fuzzy Systems Conference, London, UK, 23–26 July 2007. [CrossRef]