BI or IB: Which Better Generates High Spatiotemporal Resolution NDSI by Fusing Sentinel-2A/B and MODIS Data?

Linxin Dong, Haixi Zhou, Jiahui Xu, Yao Tang, Xiaolong Teng, Jingwen Ni, Bailang Yu, *Senior Member, IEEE*, Jianping Wu, and Yan Huang

***Abstract*—Snow cover is a sensitive indicator of climate change. Normalized difference snow index (NDSI) acquired from optical remote sensing data is usually used for monitoring snow cover, but the existing data are limited in spatiotemporal resolution. Here, we compared two blending strategies, blend-then-index (BI) and index-then-blend (IB), for generating high spatiotemporal resolution NDSI (daily, 20 m), and designed two groups of experiments (simulated and real) under three different snow cover periods over the Tibetan Plateau (TP). The flexible spatiotemporal data fusion (FSDAF) model was used as the fusion model. MODIS (daily, 500 m) and Sentinel-2A/B (2–5 days, 20 m) data were used as the inputs. The accuracy of the fused NDSI was evaluated from both spectral [root mean square error (RMSE), correlation coefficient (*R*), and average difference (AD)] and spatial (Robert's edge and local binary pattern) dimensions. Our results showed that the IB strategy produced more accurate NDSI results, with lower RMSE, higher *R*, and AD closer to zero compared to the BI strategy. In addition, there was no obvious difference in terms of texture between the two fusion strategies. Generally, the IB strategy is a better choice for generating a high spatiotemporal resolution NDSI through the FSDAF model under different snow cover periods on the TP. This study can provide effective guidelines for producing better high-resolution NDSI time series on the TP.**

***Index Terms*—Flexible spatiotemporal data fusion (FSDAF), normalized difference snow index (NDSI), spatiotemporal fusion, Tibetan Plateau (TP).**

I. Introduction

CLIMATE change is a serious challenge for human society in the context of globalization. The Tibetan Plateau (TP) is rich in snow resources and has a larger snow cover than any other mid-latitude region in the Northern Hemisphere [1], making it a hot spot and sensitive area to climate change [2]. The temperature on the TP increases by 0.3°C–0.4°C per decade on average [3], [4], which is over twice the global average temperature increase during the corresponding period, making it the region with the most significant uncertainty in environmental change [5], [6], [7]. Continued warming has decreased the extent of snow and increased snowmelt runoff in the TP [8]. This further inevitably altered the atmospheric circulation [9], [10], water balance [11], [12], and ecosystems stability [5], [13], [14] on the plateau [1]. As an important component of the cryosphere, snow cover plays an essential role in climate change in the TP and worldwide [15], [16].

Optical remote sensing data are widely used for monitoring snow cover. In particular, high demands for the spatial and temporal resolution of optical remote sensing data are placed on obtaining detailed and accurate information on snow distribution and its variability on the TP, which has a complex topography. However, there is no single satellite remote sensing data that can simultaneously satisfy the requirements of both high spatial and temporal resolution. For example, Sentinel-2A/B images have a spatial resolution of 20 m ("fine-resolution" images), which can provide more accurate and detailed spatial information on snow cover. However, its relatively long revisit period (2–5 days) may not accurately capture the dynamic changes in the snow cover on the TP. In contrast, MODIS images have a temporal resolution of 1 day, which can more accurately present dynamic changes of the snow cover on the TP. However, its spatial resolution is coarse ("coarse-resolution" images, 500 m) and cannot clearly present the details of snow information in a pixel.

Many studies (e.g., [17], [18], [19], and [20]) have used spatiotemporal fusion models to acquire remote sensing data with both high spatial and temporal resolution. Based on different assumptions and algorithmic principles, spatiotemporal fusion models can be categorized into three principal classes: weighted function-based models, unmixing-based models, and learning-based models. A weighted function-based model, the spatial and temporal adaptive reflectance fusion model (STARFM) was first proposed [17]. This model provides good results in capturing gradual change information but cannot predict transient perturbation events. To overcome this limitation, a spatial temporal adaptive algorithm for mapping reflectance change has been proposed [21], which can select an optimal reference date to improve accuracy. However, problems persisted in heterogeneous

Manuscript received 18 September 2023; revised 29 November 2023; accepted 20 December 2023. Date of publication 26 December 2023; date of current version 22 January 2024. This work was supported by the National Natural Science Foundation of China under Grant 42071306. *(Corresponding author: Yan Huang.)*

The authors are with the Key Laboratory of Geographic Information Science (Ministry of Education), the School of Geographic Sciences, East China Normal University, and the Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities (Ministry of Natural Resources), Shanghai 200241, China (e-mail: dong_linxin@163.com; holidayzhx@126.com; jiahuixu0826@163.com; tymax77777@gmail.com; teng_xiaolong@163.com; jingwen_ni@foxmail.com; blyu@geo.ecnu.edu.cn; jpwu@geo.ecnu.edu.cn; yhuang@geo.ecnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3347202

regions. Therefore, an enhanced STARFM (ESTARFM) [18] was proposed to compensate for this deficiency. This method effectively improves the simulation accuracy of the STARFM for regions with strong surface heterogeneity. However, in general, such methods cannot predict both transient perturbation events and the accurate boundaries of objects whose shapes change over time [18]. In addition, unmixing-based models have been proposed [22], [23], [24], in which the flexible spatiotemporal data fusion (FSDAF) model can address the above-mentioned problems [19]. The FSDAF model uses class-level temporal changes and thin-plate spline (TPS) interpolation for the temporal and spatial prediction of fine-resolution images at the prediction time. Through the allocation of residuals and usage of neighborhood information to refine the ultimate prediction, this method can more effectively simulate information in spatial heterogeneity regions. Subsequently, Liu et al. [20] proposed an improved FSDAF (IFSDAF) model based on the FSDAF model combined with the constrained least-squares method. However, the IFSDAF model requires the input of two fine-resolution images, making it more difficult to use in practice. Moreover, learning-based models have also advanced [25], [26]. Accompanied by the advancements in Artificial Intelligence and the proposal of new mathematical models, new spatiotemporal fusion models have been developed using various methods such as probabilistic statistics methods, convolutional neural networks, and generative adversarial network methods [27], [28], [29], [30]. For instance, the two-stream convolutional neural network [31] learning architecture not only predicts fine images from the structural similarity of coarse and fine image pairs, but also by adding temporal information in fine image sequences as strong priors. Meanwhile, the temporal constraints in the model also lead to better temporal consistency of the results. Chen et al. [32] used the cycle-generative adversarial networks to introduce an image fusion strategy with information gain for spatiotemporal image fusion by simulating and generating images for fusion. All of these methods can be effectively applied in their research directions and the spatiotemporal data fusion methods are constantly being enriched.

Among the various spatiotemporal fusion models described above, the FSDAF model stands out by virtue of its capability to identify surface changes automatically by analyzing the errors in the fusion process. It proposes a spatiotemporal data fusion framework that considers both gradual and abrupt variations in surface reflectance; thus, it performs well in regions with strong spatial heterogeneity [19]. In addition, the FSDAF model requires only one fine-resolution image and two coarse-resolution images to achieve highly accurate predictions with high efficiency. Influenced by the large topographic relief, the snow cover on the TP exhibits a spatially scattered distribution and fast-changing conditions, which is consistent with the application scope of the FSDAF model. Therefore, it is theoretically a more appropriate choice for the spatiotemporal fusion of the normalized difference snow index (NDSI) on the TP using the FSDAF model.

NDSI could highlight the snow characteristics and distinguish between snow and other features such as clouds [33]. Time-series NDSI data obtained from satellite imagery is essential for monitoring and analysis of snow cover distribution [34], [35]. At present, the NDSI snow cover data from MODIS Collection 6 (C6) are the most widely used NDSI product [36]. However, this product has data gaps caused by cloud cover. Jing et al. [37] utilized the spatio-temporal adaptive fusion method with error correction to fill the data gaps, and obtained a daily stretched seamless NDSI product for China with 500 m spatial resolution. Still, for the TP, which is characterized by complex topography and strong spatial heterogeneity snow cover, it is difficult to obtain detailed and accurate snow cover distribution or change information from such data, so the spatial and temporal resolution of optical remote sensing data is highly demanded. Therefore, the acquisition and production of higher spatiotemporal resolution NDSI is necessary. To obtain high spatial and temporal resolution snow cover data, the order of the NDSI calculations and data blending must be considered in addition to the spatiotemporal fusion model. Two blending strategies are available for consideration of the NDSI data [38], [39]. First, *blend-then-index (BI)*: The band reflectance data associated with the NDSI calculations are first blended and then the NDSI is calculated based on the fusion bands. Second, *index-then-blend (IB)*: The NDSI is first calculated using the original band reflectance data and then the obtained results are blended. Previously, the performance of the BI and IB strategies on several indices including NDVI and NDWI has been compared for vegetation and surface water monitoring [39], [40]. However, for NDSI, it remains uncertain which strategy produces superior outcomes and whether the optimal strategy differs during different snow cover periods (snow accumulation, snow stabilization, and snow ablation periods).

This study aimed to explore the optimal spatiotemporal fusion strategy for the NDSI on the TP. In the following sections, we first introduced our study area and data sources. Then, the fusion experiments were designed, and the specific steps were described. Subsequently, we discussed the uncertainties in the simulated NDSI under two different blending strategies (BI and IB), and their accuracy assessment under three snow cover periods. Finally, the results were presented. This study provides important guidelines for the spatiotemporal fusion of NDSI data on the TP.

II. Study Area and Data Sources

A. Study Area

Located in central Asia and southwest China, the TP is the highest area in the world, with an average elevation of approximately 4320 m. It is known as the "Roof of the World" and the "Third Pole" of the Earth [41]. It is located between 25°59′30″–41°1′0″ N and 67°40′37″–104°40′57″ E [42] (Fig. 1) with a total area of approximately 3.08×10^6 km^2. Throughout the year, the TP experiences low temperatures, strong solar radiation, and little rainfall [43]. The topography of the TP is complex and mountainous, with large differences in elevation between regions, showing characteristics of high spatial heterogeneity. It is rich in snow and has the largest snow cover area in the mid-latitudes of the Northern Hemisphere. Its snow cover is mostly patchy and shows the characteristics of

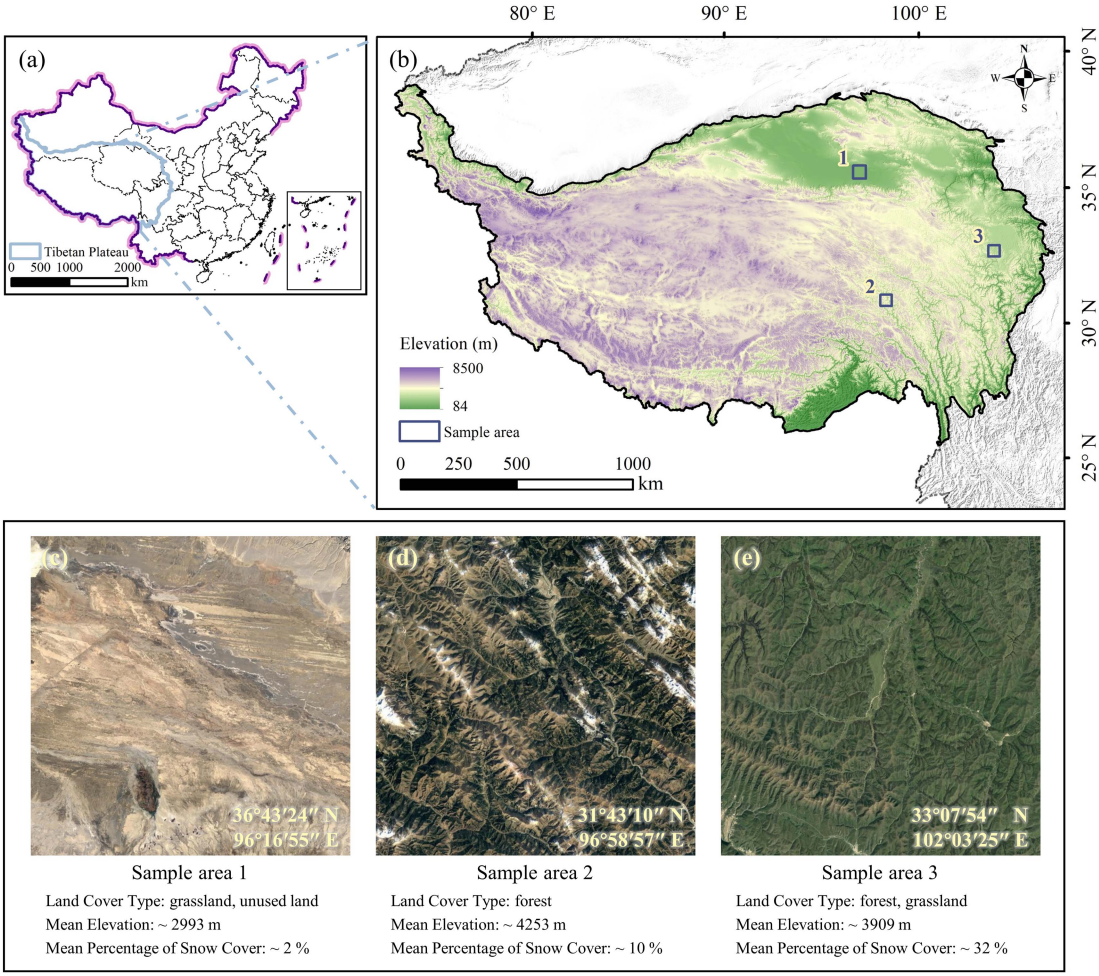

Fig. 1. Location and topography of the TP and sample areas. (a) Geographical location and (b) topography of the TP. (c)–(e) Basic information and location of the three sample areas.

scattered snow distribution. The TP has a unique and advanced perception of climate change, and its snow cover has an important impact on the hydrological cycle, regional climate change, and vegetation growth [44]. In this study, three sample areas on the TP were representatively selected as experimental areas for the spatiotemporal fusion of the NDSI (Fig. 1), considering the variability in land cover type, mean elevation, and mean percentage of snow cover.

B. Data Sources

The FSDAF model required three input images: a pair of fine-resolution and coarse-resolution images at t_1, and a coarse-resolution image at t_2. The output image was a simulated fine-resolution image at t_2. The verification image we used was the fine-resolution image at t_2.

In this study, the fine-resolution images we used were the Sentinel-2A/B data, with a temporal resolution of 2–5 days and a spatial resolution of 10–60 m [45] for the 2020–2021 snow cover period. The coarse-resolution images we used were the MODIS Terra surface reflectance product (MOD09GA V061) [46] and snow cover product (MOD10A1 V061) [47], with a temporal resolution of 1 day and spatial resolution of 500 m, consistent with the overpass date of Sentinel-2A/B. The data were all obtained from the Google Earth Engine (GEE) cloud platform (https://earthengine.google.com/, accessed on 15 September 2022) and projected as WGS 84/UTM Zone 45N.

The green band (band 3, with the spatial resolution resampled to 20 m) and short-wave infrared band (SWIR1, band 11, with the spatial resolution maintained at 20 m) of Sentinel-2A/B were used for NDSI calculations (1) [48] and subsequent strategy arrangements. The green band (sur_refl_b04) and short-wave infrared band (sur_refl_b06) of MOD09GA were used as the basis for the NDSI calculations. We also used the NDSI band of MOD10A1. It reports snow cover, which is correlated with the presence or absence of snow within a pixel, and provides a precise description of snow monitoring [49].

$$NDSI = \frac{B_{Green} - B_{SWIR}}{B_{Green} + B_{SWIR}} \tag{1}$$

where B_{Green} and B_{SWIR} correspond to the green and short-wave infrared bands, respectively.

TABLE I
TIME OF INPUT DATA FOR THE STUDY SAMPLE AREAS

Period \ Study area	Sample area 1	Sample area 2	Sample area 3
Snow accumulation period	20201119 (S&M)	20201101 (S&M)	20201113 (S&M)
	20201124 (M)	20201111 (M)	20201120 (M)
Snow stabilization period	20210217 (S&M)	20210204 (S&M)	20210221 (S&M)
	20210222 (M)	20210224 (M)	20210226 (M)
Snow ablation period	20210304 (S&M)	20210316 (S&M)	20210313 (S&M)
	20210324 (M)	20210405 (M)	20210328 (M)

S indicates the Sentinel-2A/B image and M indicates the MODIS image.

TABLE II
DETAILED INFORMATION OF ACCURACY METRICS FOR ASSESSMENT OF THE SPATIOTEMPORAL FUSION MODEL

	Metric name	Equation	Range	Variable explanation
Spectral accuracy	RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(F_i - R_i)^2}{N}}$	[0,1]	N: the total number of pixels. F_i, R_i: the value of pixel i in the fused image and reference image. μ_F, μ_R: the mean pixel values of the fused image and reference image.
	R	$\frac{\sum_{i=1}^{N}(R_i - \mu_R)(F_i - \mu_F)}{\left(\sqrt{\sum_{i=1}^{N}(R_i - \mu_R)^2}\right)\left(\sqrt{\sum_{i=1}^{N}(F_i - \mu_F)^2}\right)}$	[0,1]	
	AD	$AD = \frac{\sum_{i=1}^{N}(F_i - R_i)}{N}$	[−1,1]	
Spatial accuracy	Robert's edge	$Edge = \lvert D_{i,j} - D_{i+1,j+1}\rvert + \lvert D_{i,j+1} - D_{i+1,j}\rvert$	[−1,1]	$D_{i,j}$: the value of pixels at i_{th} row and j_{th} column. Dc, Di: the value of the central pixel and the value of pixels surrounding the central pixel in a 3 × 3 moving window.
	LBP	$LBP = decimal(d1d2 \cdots d8)$ $di = \begin{cases} 1 & if \quad Di > Dc \\ 0 & otherwise \end{cases}$	[−1,1]	

Meanwhile, different snow cover periods are associated with different snow content. To analyze the differences in the optimal fusion strategy during different snow cover periods, we selected data from the snow accumulation (October to December of each year), snow stabilization (January to March of the following year), and snow ablation periods (March to April of the following year) for the three study sample areas. Input data were those with less than 20% cloud coverage under clear-sky conditions.

The inputs of our study for the FSDAF model were 18 pairs of Sentinel-2A/B and MODIS (S&M) data at t_1 and MODIS (M) data at t_2, where the selection of t_2 was mainly based on the availability of real Sentinel-2A/B images, which was convenient for verification (Table I).

III. METHODS

The technical workflow of this study is illustrated in Fig. 2. Based on the FSDAF model, we analyzed the most suitable fusion strategy (BI or IB) for a high spatiotemporal resolution NDSI for three different snow cover periods on the TP. This process involved data preparation, fusion strategies, and accuracy assessment.

A. Data Preparation

After temporal selection of the data based on the three snow cover periods, in the preprocessing step, cloud coverage screening, reprojection, and data downloading were first performed on GEE. Then, we georectified, cropped, and resampled the data locally. To maintain consistency with the Sentinel-2A/B images, the MODIS images were upsampled by the nearest neighbor method, ultimately achieving a spatial resolution of 20 m. The preprocessed MOD09GA and MOD10A1 data were used as input data for coarse-resolution images for the BI and IB strategies, respectively. Finally, we obtained data that met the input requirements of the FSDAF model.

First, we used real data to construct the experimental groups. In the real experimental group, the input fine-resolution images at t_1 were real Sentinel-2A/B images and the coarse-resolution

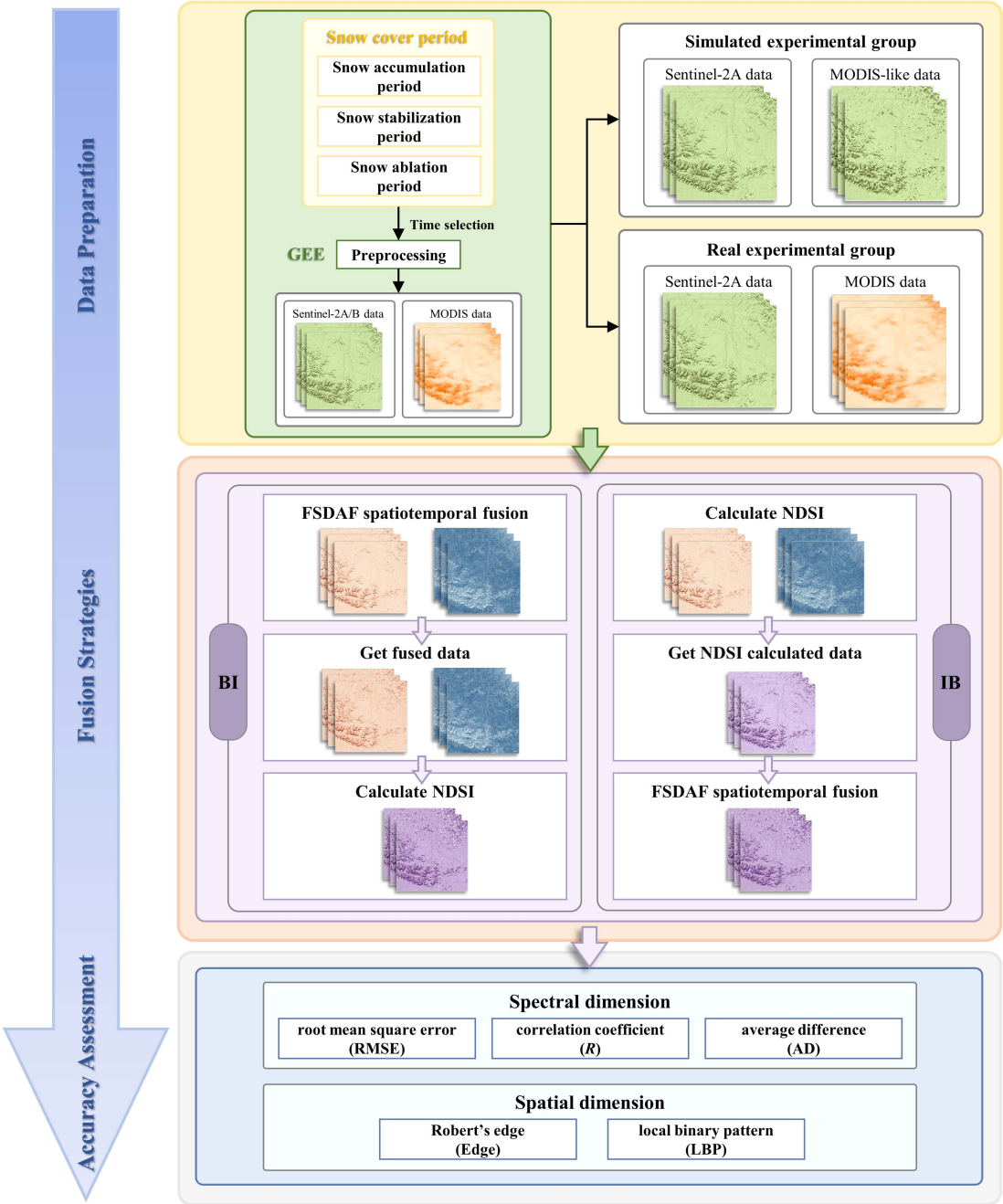

Fig. 2. Flowchart for analyzing the suitable spatiotemporal fusion strategy of NDSI under different snow cover periods on the TP.

images at t_1 and t_2 were real MODIS images. Then, to prevent radiation and geometric inconsistencies between Sentinel-2A/B and MODIS images from affecting the accuracy of the spatiotemporal fusion model, we also constructed a simulated experimental group. In the simulated experimental group, we selected MODIS-like images (resampling the spatial resolution of Sentinel-2A/B images to 500 m) as coarse-resolution input data for model testing [24]. In this group, the input fine-resolution images at t_1 were real Sentinel-2A/B images, and the coarse-resolution images at t_1 and t_2 were MODIS-like images. In this manner, we eliminated the inconsistencies between the Sentinel-2A/B and MODIS images, ensuring that the difference in accuracy was related only to the BI/IB strategies, thereby allowing a direct comparison of the performance of the two strategies.

B. FSDAF Model

The FSDAF model, originally proposed by Zhu et al. [19], can effectively solve the problem of data fusion in regions with gradual and abrupt surface reflectance changes and high spatial heterogeneity. There are six steps in generating a high spatiotemporal resolution NDSI using the FSDAF model. First, the fine-resolution images at t_1 are classified and then the ratio of

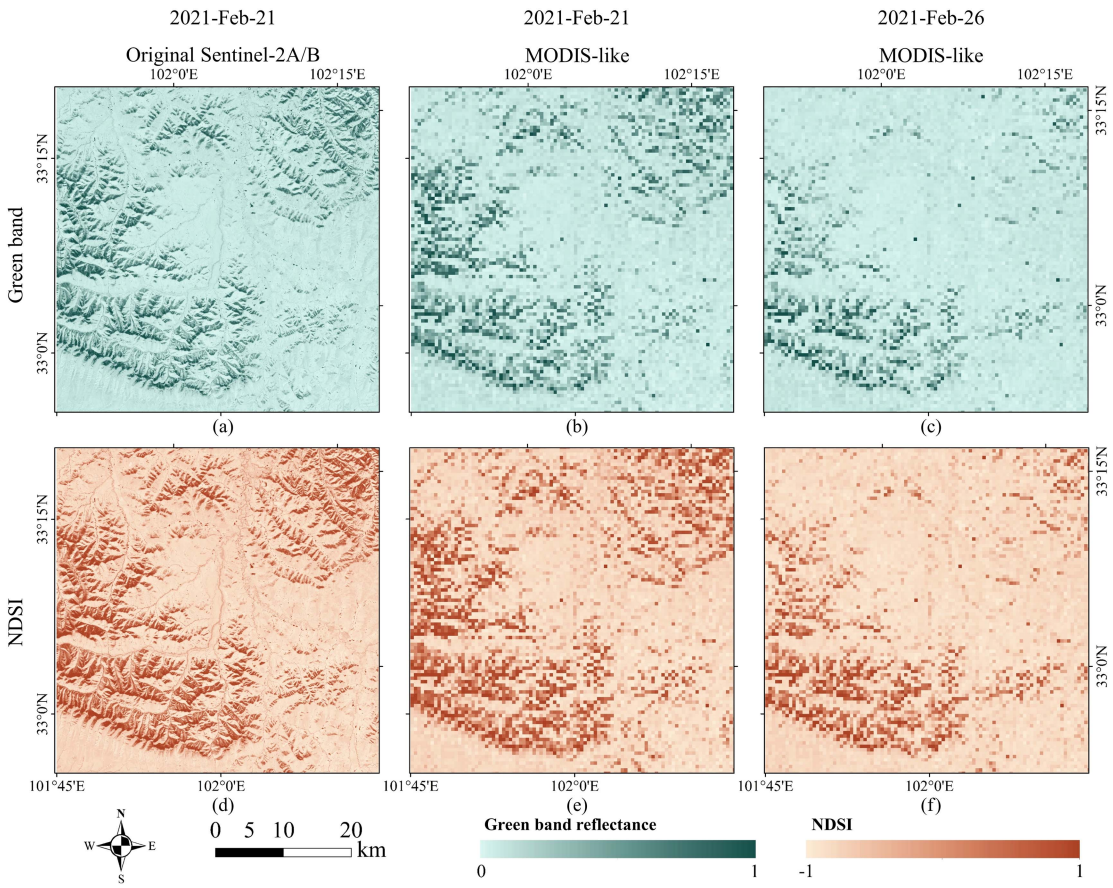

Fig. 3. Input data of the green band and NDSI in simulated experimental group under BI and IB strategies for sample area 3. (a) Green band data of Sentinel-2A/B image acquired on February 21, 2021. (b) and (c) Green band data of MODIS-like images aggregated from (a) and Sentinel-2A/B images acquired on February 26, 2021. (d) NDSI data of Sentinel-2A/B image acquired on February 21, 2021. (e) and (f) NDSI data of MODIS-like images aggregated from (d) and Sentinel-2A/B images acquired on February 26, 2021.

each class is calculated for each image pixel. Second, following the principles of linear decomposition of the mixed pixel theory, the temporal change from t_1 to t_2 for each class is calculated, where the temporal change of a pixel in the coarse-resolution image is equivalent to the weighted summation of the temporal changes across all classes within that pixel. Third, the fine-resolution image at t_2 is temporally predicted by the temporal change in each class, and the residual of each coarse-resolution image pixel is calculated as follows:

$$R\left(x_i, y_i, b\right) = \Delta C\left(x_i, y_i, b\right) - \frac{1}{m}\left[\sum_{j=1}^{m} F_2^{TP}\left(x_{ij}, y_{ij}, b\right) - \sum_{j=1}^{m} F_1\left(x_{ij}, y_{ij}, b\right)\right] \tag{2}$$

where for a given band b, i represents the index of a coarse pixel, j represents the index of a fine pixel within a coarse pixel, and m represents the total count of fine pixels encompassed within a single coarse pixel. $R(x_i, y_i, b)$ is the residual at (x_i, y_i) of the coarse pixel. $\Delta C(x_i, y_i, b)$ is the change in the (x_i, y_i) coarse pixel between t_1 and t_2, $F_1(x_{ij}, y_{ij}, b)$ is the value of the jth fine pixel within the coarse pixel at (x_i, y_i) observed at t_1, and $F_2^{TP}(x_{ij}, y_{ij}, b)$ is the temporal predicted value at fine pixels (x_i, y_i) at t_2.

Fourth, a spatial prediction of fine-resolution images at t_2 is performed based on coarse-resolution images at t_2 and the TPS interpolation method. Fifth, residuals to fine-resolution pixels are assigned. Combining homogenous and heterogeneous landscape factors, a homogeneity index (HI) is introduced to guide the distribution of residuals (3). The residuals distributed to the jth fine pixel can be obtained by combining the weights of the two cases using HI and normalizing them (4).

$$HI\left(x_{ij}, y_{ij}\right) = \left(\sum_{k=1}^{m} I_k\right) / m \tag{3}$$

where the kth fine pixel in the moving window has the same land cover type as that of the central fine pixel (x_{ij}, y_{ij}) with $I_k = 1$; otherwise, $I_k = 0$. The closer the HI is to 1, the more homogeneous is the landscape.

$$r\left(x_{ij}, y_{ij}, b\right) = m \times R\left(x_i, y_i, b\right) \times W\left(x_{ij}, y_{ij}, b\right) \tag{4}$$

where $W(x_{ij}, y_{ij}, b)$ is the normalized weight.

Through the aggregation of residual distributions and prediction of temporal changes, the value of the total change within a fine-resolution image pixel between t_1 and t_2 can be obtained

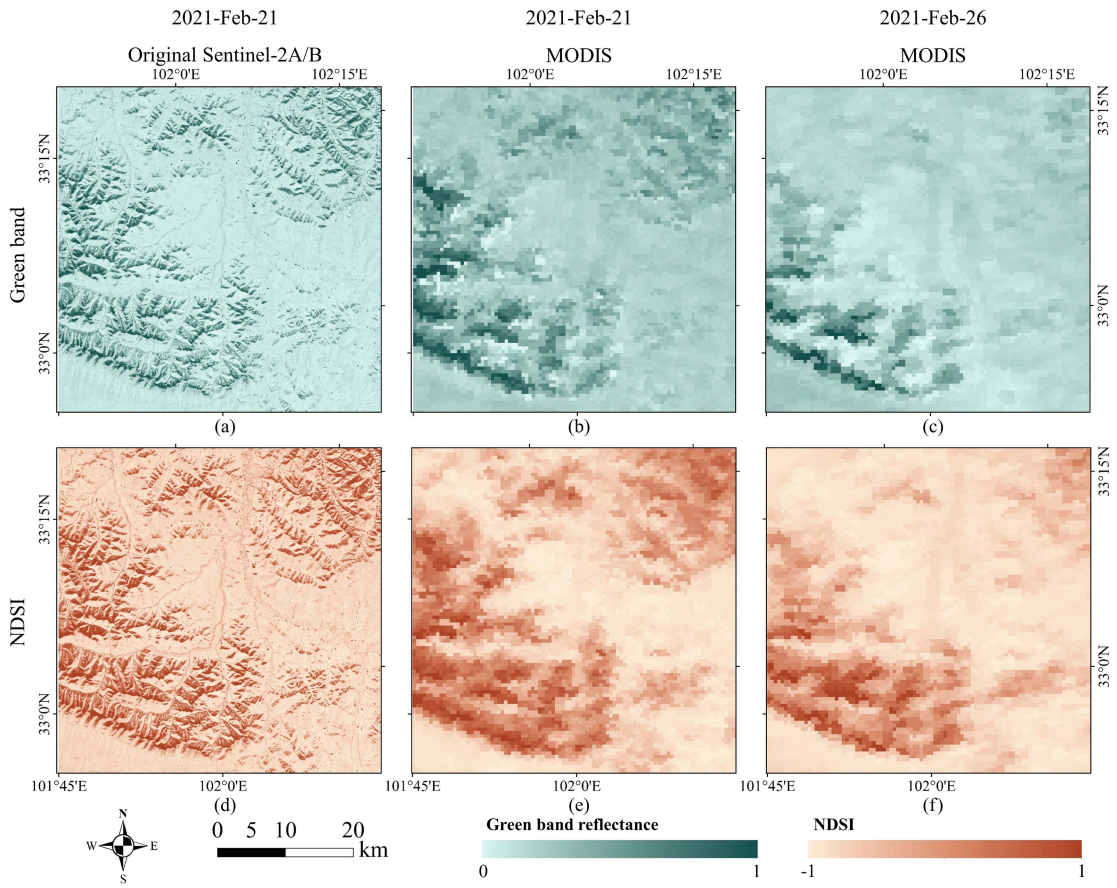

Fig. 4. Input data of the green band and NDSI in real experimental group under BI and IB strategies for sample area 3. (a) and (b) Green band data of Sentinel-2A/B and MOD09GA images acquired on February 21, 2021. (c) Green band data of MOD09GA image acquired on February 26, 2021. (d) and (e) NDSI data of Sentinel-2A/B and MOD10A1 images acquired on February 21, 2021. (f) NDSI data of MOD10A1 image acquired on February 26, 2021.

as follows:

$$\Delta F\left(x_{ij}, y_{ij}, b\right) = r\left(x_{ij}, y_{ij}, b\right) + \Delta F\left(c, b\right) \tag{5}$$

where $\Delta F(c, b)$ represents the change in the value of class c at a fine resolution between t_1 and t_2.

Finally, the neighborhood information is used to make the final predictions of the fine-resolution images. The change information of all similar pixels is weighted and summed to obtain the total change value of the target pixel. The final prediction results are as follows:

$$\hat{F}_2\left(x_{ij}, y_{ij}, b\right) = F_1\left(x_{ij}, y_{ij}, b\right) + \sum_{k-1}^{n} w_k \times \Delta F\left(x_k, y_k, b\right) \tag{6}$$

where w_k is the weight of the kth similar pixel.

C. Fusion Strategies

When using the FSDAF model for spatiotemporal fusion of the NDSI, an appropriate fusion strategy must be determined. The two strategies were applied in our study as follows [38]. First, *BI strategy*: Green (band 3 for Sentinel-2A/B and sur_refl_b04 for MOD09GA) and short-wave infrared (band 11 for Sentinel-2A/B and sur_refl_b06 for MOD09GA) bands about NDSI calculation were spatiotemporally fused using the FSDAF model. The two fused bands were then used to calculate the NDSI. Second, *IB strategy*: The NDSI was calculated using the reflectance of the green and short-wave infrared bands (bands 3 and 11 of Sentinel-2A/B and the NDSI band of MOD10A1). The calculated NDSI was then fused with the FSDAF model. Both strategies can achieve NDSI results with high spatiotemporal resolution, but the accuracy of the results is unknown.

For example, in the simulated experimental group, two sets of input data were used. One was under the BI strategy, using the green band (one of the calculated bands of NDSI) of Sentinel-2A/B data at t_1, the corresponding MODIS-like data at t_1 and t_2 of sample area 3 on February 21 and 26, 2021, during the snow stabilization period. The other was under the IB strategy, using the calculated NDSI by Sentinel-2A/B data at t_1 and the corresponding MODIS-like data at t_1 and t_2 under the same conditions (Fig. 3). In the real experimental group, we used the green band of the Sentinel-2A/B data at t_1 and the real MOD09GA data at t_1 and t_2 as inputs under the BI strategy. The NDSI calculated using the Sentinel-2A/B data at t_1 and the NDSI band of the MOD10A1 data at t_1 and t_2 were used as inputs under the IB strategy. We presented the data for the real experimental group under the same conditions as the simulated experimental group for a better comparison (Fig. 4).

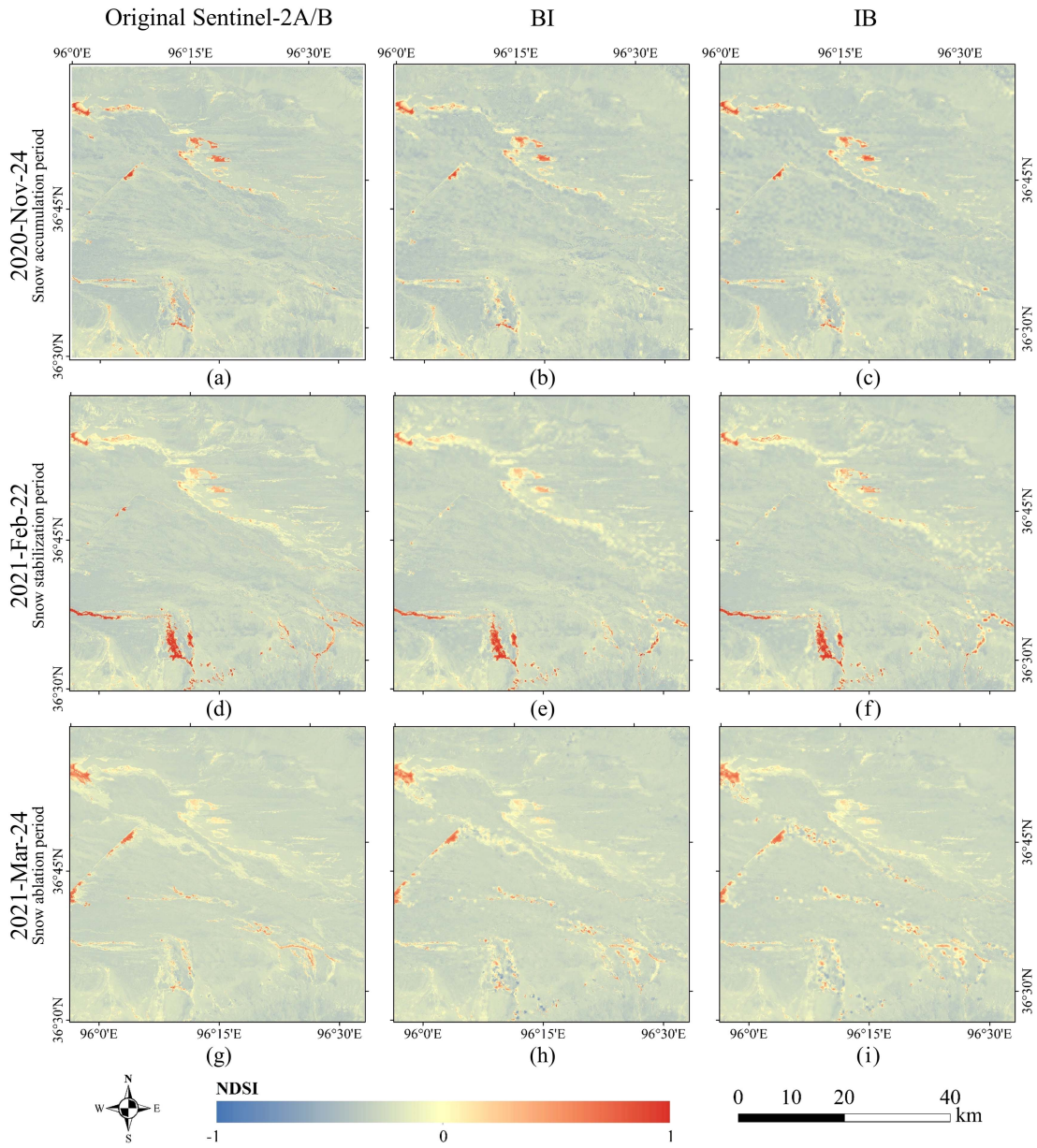

Fig. 5. FSDAF spatiotemporal fusion results of the simulated experimental group for sample area 1. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 24, 2020, February 22, 2021, and March 24, 2021, respectively.

D. Accuracy Assessment

In our study, the verification images were fine-resolution images at t_2 to quantitatively evaluate the accuracy of the fusion results. Zhu et al. [50] proposed a new framework that combines both spectral and spatial details for evaluating the performance of spatiotemporal fusion models. Their results suggested that the optimal combination of accuracy assessment metrics was the root mean square error (RMSE), average difference (AD), Robert's edge (Edge), and local binary pattern (LBP). This combination can not only provide a comprehensive and effective assessment of errors in both spectral and spatial dimensions but also significantly reduce the information redundancy associated with highly correlated metrics. Therefore, we used these four accuracy assessment metrics supplemented with the correlation coefficient (R) for quality assessment (Fig. 2, Table II).

IV. Results

A. Spatiotemporal Fusion Results

Figs. 5–7 show the results of the simulated experimental groups under the two strategies and three snow cover periods. We found that both strategies achieved good results for areas with gentle (sample area 1, Fig. 5) or large (sample areas 2 and 3, Figs. 6 and 7) topography compared with the original image. Visually, the texture details of the NDSI results fused by Sentinel-2A/B and MODIS-like data were close to the real

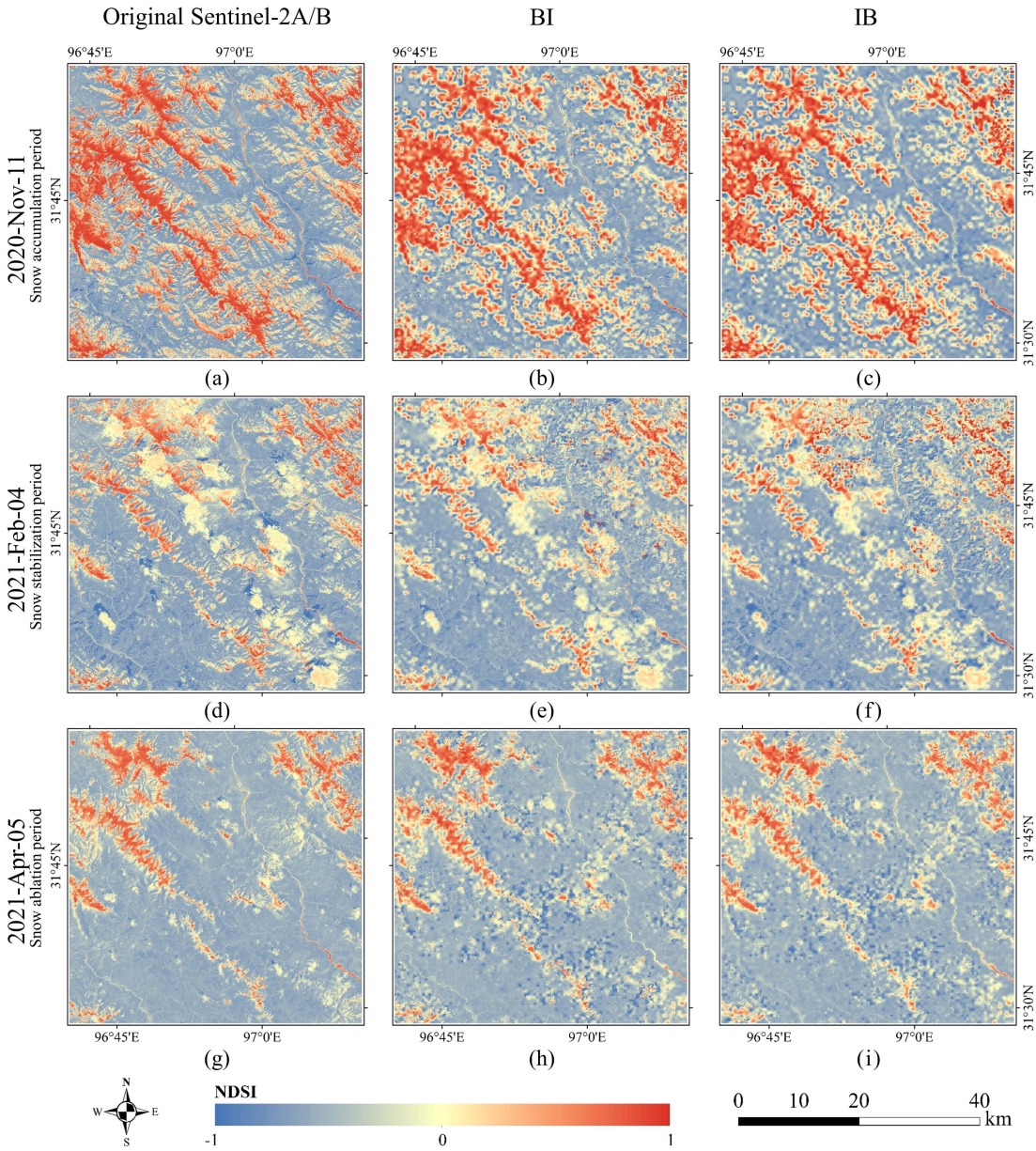

Fig. 6. FSDAF spatiotemporal fusion results of the simulated experimental group for sample area 2. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 11, 2020, February 4, 2021, and April 5, 2021, respectively.

images, which can reproduce the real NDSI well. Taking sample area 3 as an example, we found that the NDSI results were accurately predicted for areas with moderate topographic relief, such as the upper-right part of the image. However, this prediction tended to be smooth, resulting in blurred images. Comparatively, the NDSI results for the region with a large topographic relief in the lower-left part of the image were more obvious and better predicted. Compared to these strategies, the BI strategy predicted visually smoother results, resulting in the loss of more spatial detail, whereas the IB strategy had a better prediction of regional edge information.

Similarly, we obtained results for the real experimental group under the two strategies and three snow cover periods (Figs. 8 –10). The fusion results followed a pattern similar to that of the simulated experimental group for areas with gentle terrain (sample area 1, Fig. 8). Furthermore, the fusion results were more affected by clouds [sample area 2, Fig. 9(b), (c) and (e), (f)] and its shadow [Fig. 9(h) and (i)]. The NDSI results were largely underestimated with a large percentage of clouds. Similarly, for sample area 3 (Fig. 10), we found that the results of the IB strategy were generally better than those of the BI strategy, which presented more spatial details.

B. *Accuracy Assessment*

Tables III–V show the accuracy assessment results of the simulated experimental group under the two strategies for the three snow cover periods from spectral (RMSE, R, AD) and spatial (Edge and LBP) dimensions. The three study sample

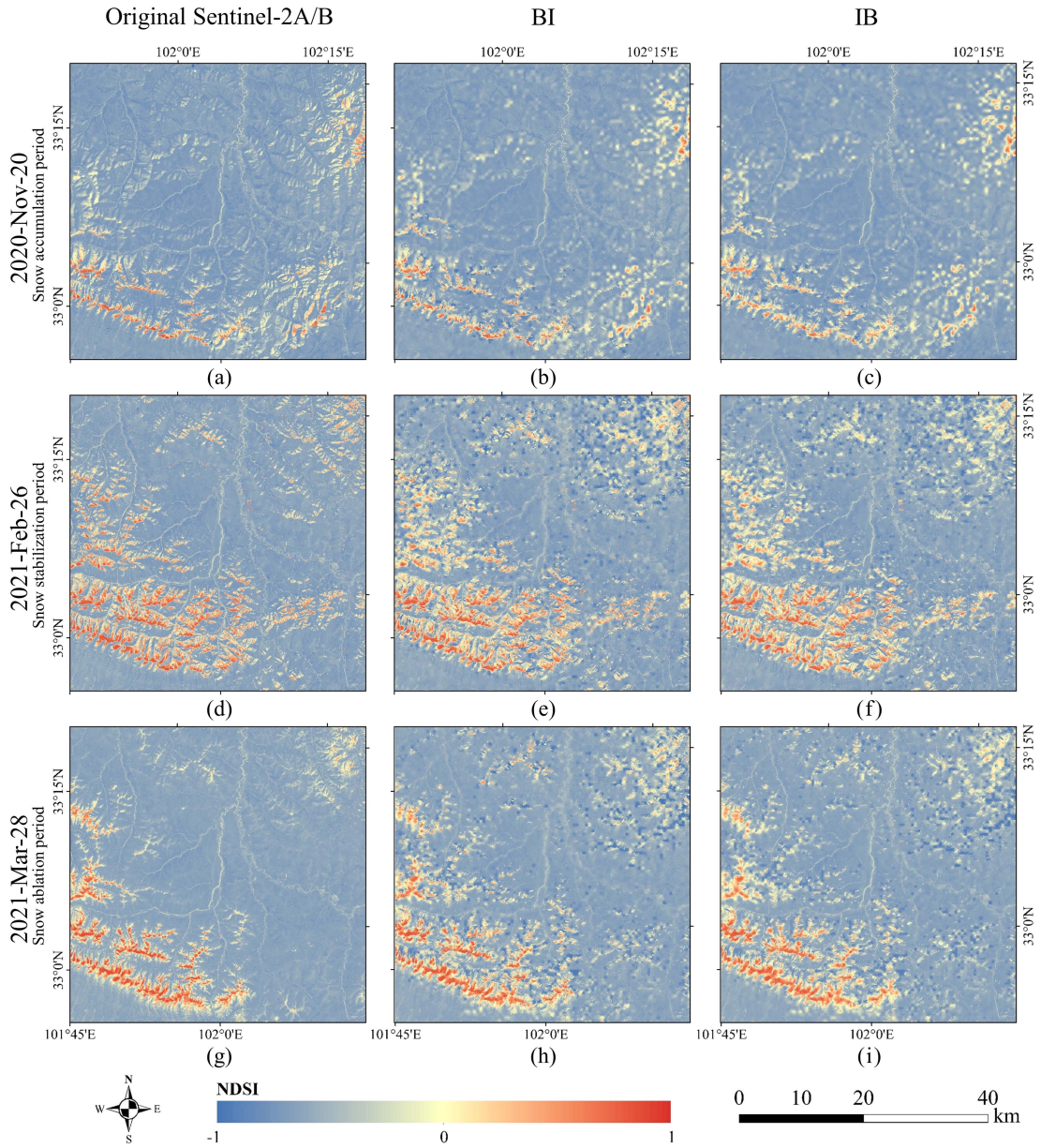

Fig. 7. FSDAF spatiotemporal fusion results of the simulated experimental group for sample area 3. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 20, 2020, February 26, 2021, and March 28, 2021, respectively.

areas reached a similar conclusion that the IB strategy was better. For sample area 3 (Table V), we concluded that the IB strategy was better in the spectral dimension, whereas there was no significant pattern in the spatial dimension. Specifically, in the spectral dimension of our experiments, the RMSE values under the IB strategy for snow accumulation, snow stabilization, and snow ablation periods were 0.122, 0.162, and 0.153, respectively, which were lower than those of the BI strategy. The R values under the IB strategy for the three snow cover periods were higher than those of the BI strategy by approximately 0.012, 0.012, and 0.031, respectively. The AD values under the IB strategy were −0.001, −0.008, and −0.012, respectively, which were closer to 0 than the BI strategy. In the spatial dimension, there was no obvious rule for the values of Edge and LBP under different strategies in different snow cover periods, and the difference in the results was not significant. We also observed visually (Fig. 7) that the prediction results of the two strategies for the NDSI fused images in the spatial dimension were consistent with the accuracy assessment results.

Tables VI –VIII show the accuracy assessment results of the real experimental group under the two strategies for the three snow cover periods from spectral and spatial dimensions. We obtained the same results as in the simulated experimental group, with the IB strategy being better. Similarly, for sample area 3 (Table VIII), we concluded that the IB strategy was better in the spectral dimension, whereas the BI strategy tended to be better in the spatial dimension. In this group, the RMSE values of the IB strategy for the three snow cover periods were lower than

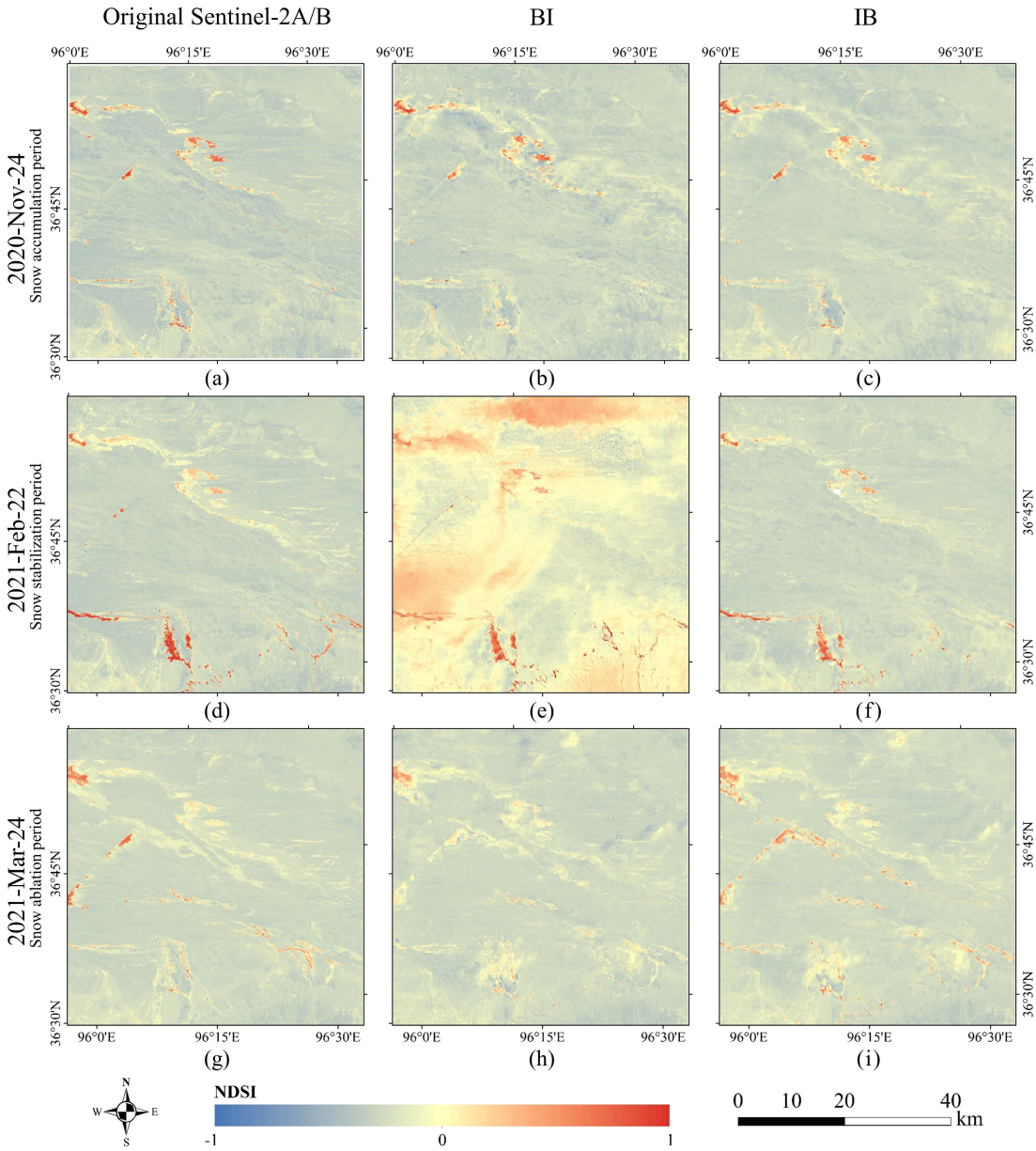

Fig. 8. FSDAF spatiotemporal fusion results of the real experimental group for sample area 1. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 24, 2020, February 22, 2021, and March 24, 2021, respectively.

those of the BI strategy by 0.010, 0.030, and 0.009, respectively. The R values of the IB strategy were 0.744, 0.833, and 0.753, respectively, which were higher than those of the BI strategy. The AD values of the IB strategy were 0.012, −0.010, and −0.034, respectively, with no significant tendency. In the spatial dimension, the Edge values of the BI strategy were higher than those of the IB strategy by 0.107, 0.036, and 0.023, respectively, which were closer to 0. The LBP values of the BI strategy were −0.025, −0.030, and 0.023, respectively, which were also closer to 0. Under these conditions, thus it was concluded that the BI strategy performed better in the spatial dimension. We also noticed (Fig. 10) that although the BI strategy was more accurate in assessment results, both strategies did well in terms of the details prediction. Overall, combining the fused images, accuracy assessment results, and the data conditions of the 3 sample areas (Figs. 8–10, Tables VI–VIII), it can be concluded that the IB strategy had stronger stability in the prediction of the spatial dimension, and its texture features and edge features can be well preserved.

Next, we performed an overall analysis of the snow cover periods to derive a final evaluation of the spatiotemporal fusion quality. For each of the three snow cover periods, we averaged the results of the accuracy assessment of the two experimental groups in the three sample areas under the two strategies, respectively (Table IX).

The IB strategy had better accuracy assessment results in the spectral dimension (Table IX). In the simulated experimental group, for the three snow cover periods, the mean RMSE values

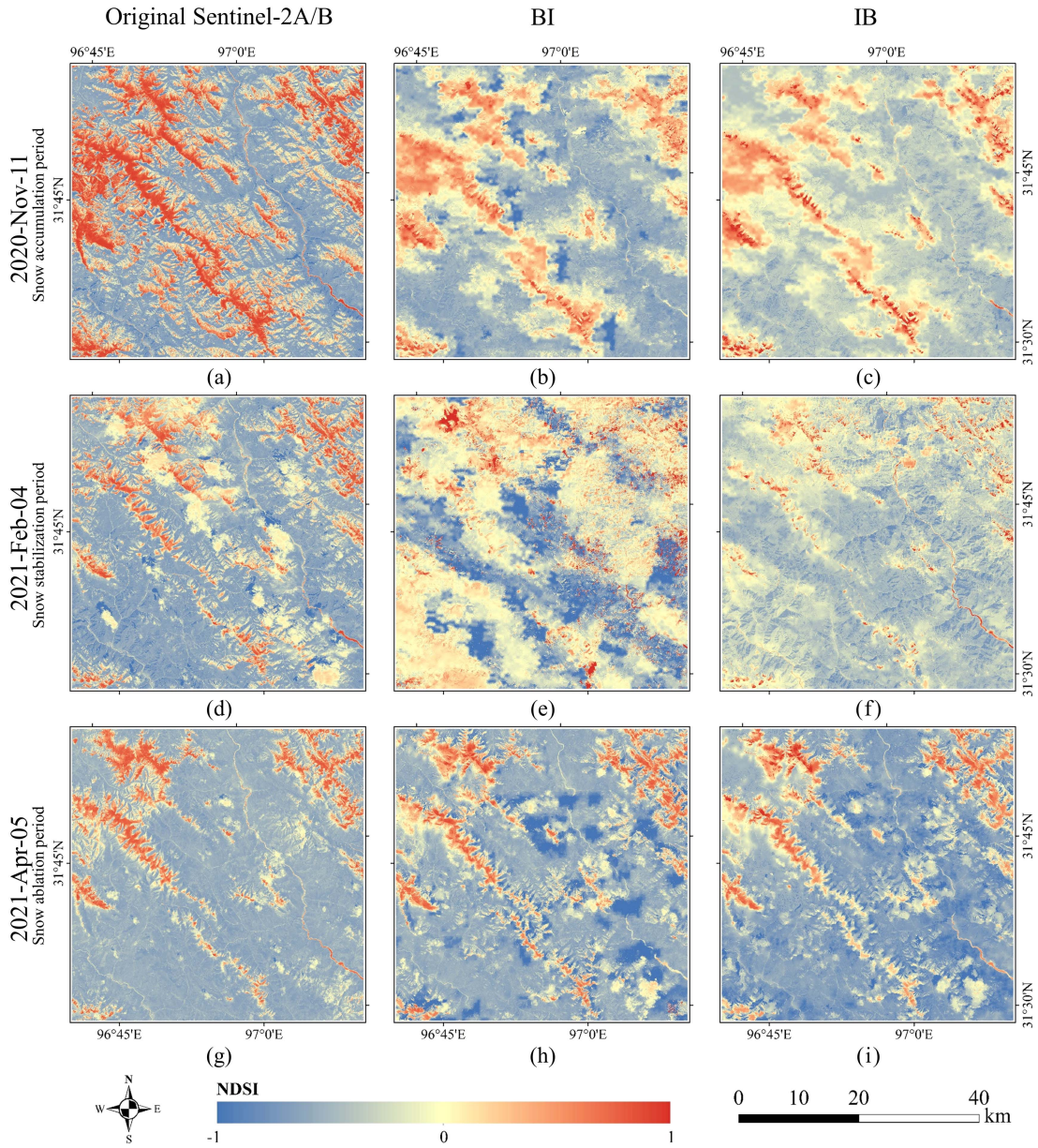

Fig. 9. FSDAF spatiotemporal fusion results of the real experimental group for sample area 2. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 11, 2020, February 4, 2021, and April 5, 2021, respectively.

for the IB strategy were lower than those for the BI strategy by 0.007, 0.026, and 0.014, respectively, with an overall average of 0.015. The mean R values of the IB strategy were higher than those of the BI strategy by 0.013, 0.054, and 0.022, respectively, with an overall average of 0.030. The mean AD values of the IB strategy were −0.002, −0.003, and −0.004, respectively, which were higher than those of the BI strategy with an overall average of 0.005. In the real experimental group, the mean RMSE, R, and AD values of the IB strategy were 0.072 lower, 0.080 higher, and 0.022 higher, respectively, than those of the BI strategy. In terms of spatial dimension, the accuracy assessment of the BI strategy was slightly higher than that of the IB strategy, but there were individual cases in which IB was higher and the texture and edge features were well preserved. The evaluated values of spatial dimension accuracy under the two strategies were closer, and the overall difference was not obvious. Therefore, based on a comprehensive consideration of the spectral and spatial dimensions, we concluded that it was better to fuse the NDSI using the IB strategy.

V. Discussion

A. Comparison with Other Spatiotemporal Fusion Models

This experiment focused on the differences in the fused NDSI results between the BI and IB spatiotemporal fusion strategies for different snow cover periods on the TP. To make the results more comparable, we chose the same FSDAF model for the strategy comparisons. However, there is doubt as to whether

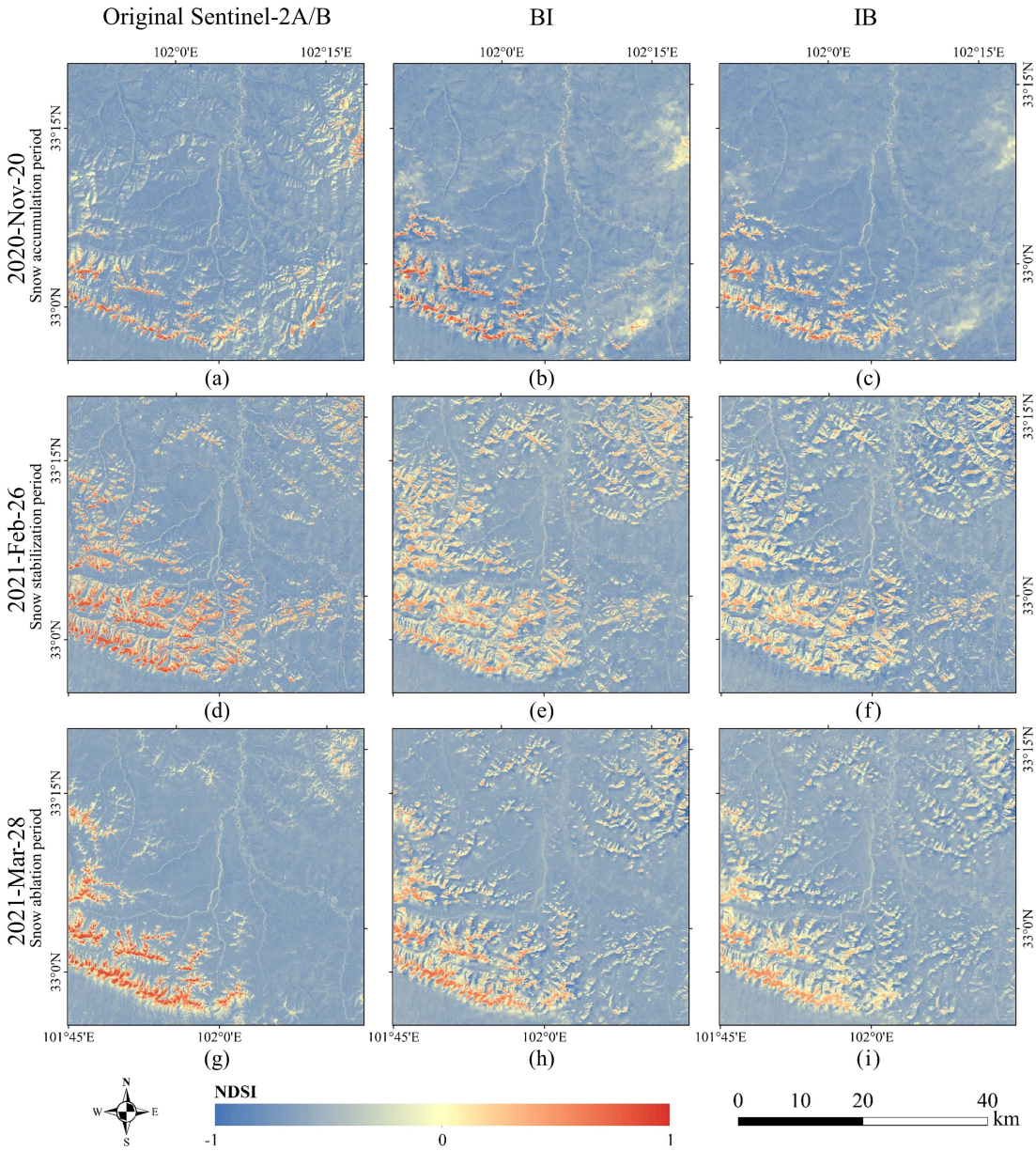

Fig. 10. FSDAF spatiotemporal fusion results of the real experimental group for sample area 3. (a), (d), and (g) Original Sentinel-2A/B images, and its predicted images by (b), (e), and (h) BI strategy and (c), (f), and (i) IB strategy of November 20, 2020, February 26, 2021, and March 28, 2021, respectively.

different spatiotemporal fusion models lead to different results. Therefore, we selected the other three models (STARFM, ESTARFM, and Fit-FC) to compare with the results of the FSDAF model. We used the simulated experimental group of sample area 3 on February 26, 2021, as an example (Fig. 11, Table X).

From Fig. 11, it can be observed that the results predicted by the STARFM [Fig. 11(a)–(d)] ignored some spatial details and exhibited poor stability. In contrast, the ESTARFM [Fig. 11(e)–(h)] and FSDAF [Fig. 7(e) and (f) and Fig. 10(e) and (f)] models predicted the results better, and it was evident that their fusion results retained more textural features and spatial details. However, the input data of the ESTARFM are two fine-resolution images and three coarse-resolution images, which are difficult to acquire and require more time than other models. Comparing the results of Fit-FC [Fig. 11(i)–(l)] and FSDAF models, it was found that the results of Fit-FC were better in the simulated experimental group, while the results of FSDAF were better in the real experimental group. This is primarily because Fit-FC must be implemented using radiative normalization, which is more sensitive to systematic radiometry [24]. Therefore, the Fit-FC model was a good choice for Sentinel-2A/B and its corresponding MODIS-like data in the simulated experimental group.

The results of all three models were in a reasonable range and had the same conclusions as in Section IV-B (Table X). Specifically, in the mean results of the simulated experimental group, the RMSE values (0.162 for BI and 0.151 for IB on average)

TABLE III
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE SIMULATED EXPERIMENTAL GROUP OF SAMPLE AREA 1

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201124	BI	0.046	0.891	−0.002	−0.263	−0.023
		IB	0.041	0.906	0.000	−0.355	−0.228
Snow stabilization period	20210222	BI	0.046	0.927	0.000	−0.212	−0.004
		IB	0.046	0.933	0.000	−0.203	−0.106
Snow ablation period	20210324	BI	0.058	0.811	−0.002	−0.172	0.033
		IB	0.055	0.825	−0.001	−0.212	−0.075

TABLE IV
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE SIMULATED EXPERIMENTAL GROUP OF SAMPLE AREA 2

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201111	BI	0.371	0.717	−0.021	−0.588	0.005
		IB	0.362	0.729	−0.005	−0.636	−0.042
Snow stabilization period	20210204	BI	0.319	0.593	−0.023	−0.445	0.001
		IB	0.261	0.737	−0.002	−0.426	−0.037
Snow ablation period	20210405	BI	0.216	0.813	0.008	−0.246	0.010
		IB	0.198	0.836	0.002	−0.339	−0.022

TABLE V
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE SIMULATED EXPERIMENTAL GROUP OF SAMPLE AREA 3

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201120	BI	0.130	0.776	−0.001	−0.277	−0.048
		IB	0.122	0.788	−0.001	−0.345	−0.140
Snow stabilization period	20210226	BI	0.181	0.829	−0.016	−0.231	0.057
		IB	0.162	0.841	−0.008	−0.330	0.004
Snow ablation period	20210328	BI	0.174	0.790	−0.018	−0.203	0.074
		IB	0.153	0.821	−0.012	−0.273	0.007

and *R* values (0.892 for BI and 0.900 for IB on average) of the spectral dimensions indicated that the IB strategy had a better effect, whereas AD values (0.007 for BI and 0.019 for IB on average) indicated that the BI strategy was slightly better. Overall, the IB performed better. In the spatial dimension, the accuracy of BI was slightly higher (−0.165 and −0.013 for BI, and −0.253 and −0.080 for IB, on average). In the results of the real experimental group, all five accuracy assessment metrics indicated the superiority of the IB strategy.

From the perspective of spatiotemporal fusion methods, ESTARFM slightly outperforms SRATFM, but not obviously. The results of both are influenced by the combined effects of temporal and spatial variances, which is consistent with previous

TABLE VI
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE REAL EXPERIMENTAL GROUP OF SAMPLE AREA 1

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201124	BI	0.133	0.818	−0.021	−0.245	0.014
		IB	0.048	0.853	−0.027	−0.348	−0.185
Snow stabilization period	20210222	BI	0.148	0.385	−0.244	−0.251	0.004
		IB	0.047	0.896	−0.013	−0.209	−0.054
Snow ablation period	20210324	BI	0.133	0.718	−0.005	−0.265	0.045
		IB	0.071	0.713	−0.016	−0.232	−0.036

TABLE VII
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE REAL EXPERIMENTAL GROUP OF SAMPLE AREA 2

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201111	BI	0.386	0.621	0.077	−0.746	−0.051
		IB	0.238	0.630	−0.015	−0.784	−0.116
Snow stabilization period	20210204	BI	0.408	0.467	−0.146	−0.483	−0.018
		IB	0.221	0.533	−0.065	−0.462	−0.048
Snow ablation period	20210405	BI	0.257	0.762	−0.035	−0.271	−0.007
		IB	0.236	0.812	−0.084	−0.347	−0.012

TABLE VIII
RESULTS OF THE ACCURACY ASSESSMENT UNDER THE REAL EXPERIMENTAL GROUP OF SAMPLE AREA 3

Period	Specific date	Strategy	Accuracy metrics				
			RMSE	*R*	AD	Edge	LBP
Snow accumulation period	20201120	BI	0.138	0.733	−0.007	−0.269	−0.025
		IB	0.128	0.744	0.012	−0.376	−0.107
Snow stabilization period	20210226	BI	0.194	0.813	−0.040	−0.290	−0.030
		IB	0.164	0.833	−0.010	−0.326	0.037
Snow ablation period	20210328	BI	0.167	0.732	−0.031	−0.252	0.023
		IB	0.158	0.753	−0.034	−0.275	0.038

studies [51]. For Fit-FC, it largely underestimates the spatial information of the real experimental group, but it is well suited for the studies of temporal dynamics [50]. And for the FSDAF model, from previous studies [19] and our results, its advantages are mainly reflected in the following points. First, it requires less input data. Second, it is highly accurate, spatially detailed, and suitable for areas of land cover type change. Third, it is suitable for regions with a high spatial heterogeneity. It then solves the problem of temporal changes in endmembers globally, uses the purest coarse pixels to form the system of linear equations, and refines them, providing stronger robustness than other methods based on unmixing. Finally, neighborhood information is introduced to obtain a more continuous spatial image [19]. As a significant complement to the collection of spatiotemporal

TABLE IX
MEAN VALUES OF ACCURACY EVALUATION OF FUSION RESULTS UNDER SIMULATED AND REAL EXPERIMENTAL GROUPS FOR THE THREE SNOW COVER PERIODS

Period	Strategy	Simulated experimental groups					Real experimental groups				
		RMSE	*R*	AD	Edge	LBP	RMSE	*R*	AD	Edge	LBP
Snow accumulation period	BI	0.182	0.795	−0.008	−0.376	−0.022	0.219	0.724	0.016	−0.420	−0.021
	IB	0.175	0.808	−0.002	−0.445	−0.137	0.138	0.742	−0.010	−0.503	−0.136
Snow stabilization period	BI	0.182	0.783	−0.013	−0.296	0.018	0.250	0.555	−0.143	−0.341	−0.015
	IB	0.156	0.837	−0.003	−0.320	−0.046	0.144	0.754	−0.029	−0.332	−0.022
Snow ablation period	BI	0.149	0.805	−0.004	−0.207	0.039	0.186	0.737	−0.024	−0.263	0.020
	IB	0.135	0.827	−0.004	−0.275	−0.030	0.155	0.759	−0.045	−0.285	−0.003

TABLE X
COMPARISON OF NDSI FUSION ACCURACY (RMSE, *R*, AD, EDGE, AND LBP VALUES) FOR STARFM, ESTARFM, AND FIT-FC MODELS UNDER BI AND IB STRATEGIES

Experimental group	Strategy	BI					IB				
	Method	RMSE	*R*	AD	Edge	LBP	RMSE	*R*	AD	Edge	LBP
Simulated	STARFM	0.205	0.855	–0.013	–0.104	–0.039	0.149	0.896	–0.003	–0.288	–0.172
	ESTARFM	0.149	0.911	0.033	–0.102	0.025	0.171	0.894	0.061	–0.138	0.007
	Fit-FC	0.133	0.911	0.001	–0.289	–0.026	0.133	0.910	–0.002	–0.333	–0.075
	Average	0.162	0.892	0.007	–0.165	–0.013	0.151	0.900	0.019	–0.253	–0.080
Real	STARFM	0.186	0.826	0.034	–0.385	–0.141	0.188	0.823	0.020	–0.281	–0.065
	ESTARFM	0.183	0.860	0.051	–0.149	0.031	0.171	0.885	0.060	–0.151	0.014
	Fit-FC	0.234	0.701	–0.022	–0.640	–0.217	0.235	0.759	–0.059	–0.666	–0.246
	Average	0.201	0.796	0.021	–0.391	–0.109	0.198	0.822	0.007	–0.366	–0.099

data fusion methods, it has become a widely used and stable spatiotemporal fusion model.

Although the FSDAF model has demonstrated strong performance in spatiotemporal fusion, it still has shortcomings. First, FSDAF selects coarse pixels with an interquartile range of 0.1−0.9 or narrower to participate in the calculation during the unmixing process in the spatiotemporal prediction step [52]. The selection of its change value is fixed and based only on experience, which is not strict. The likelihood of selecting coarse pixels with abundant boundary information also has a high probability of being selected. Once these pixels are selected, erroneous spectral information will be introduced into the unmixing calculation, which affects the accuracy of the results. Consequently, FSDAF reduces the contrast between different objects and predicts visually blurrier images. Second, in the spatial prediction step, FSDAF uses TPS interpolation to extract the change information at t_2 from the coarse-resolution image, but cannot accurately determine whether the land cover type of the fine-resolution image has changed [52]. As a result, the FSDAF model can neither determine a clear boundary for predicting land cover type changes nor estimate changed image pixel values. Third, Sentinel-2A/B data were used in our study to improve the spatial resolution, but this was still insufficient in study areas characterized by high cloud cover. The rationality of the FSDAF model in regions with high cloud cover needs to be explored further [53]. Finally, this study also has some limitations. Although we explored the optimal strategy, we just applied the conventional downsampling method to obtain MODIS-like data, which is also an important source of uncertainty in the results [54].

B. Selection of Fusion Strategy During Different Snow Cover Periods

Different snow cover periods imply different snow status, and their corresponding NDSI conditions are different, which may affect the accuracy of the prediction results. In our study, we investigated whether the optimal strategy for spatiotemporal fusion differed for the three snow cover periods. However, we found that the selection of snow cover periods did not have a significant effect on the determination of the optimal strategy. This conclusion is different from that of vegetation growth.

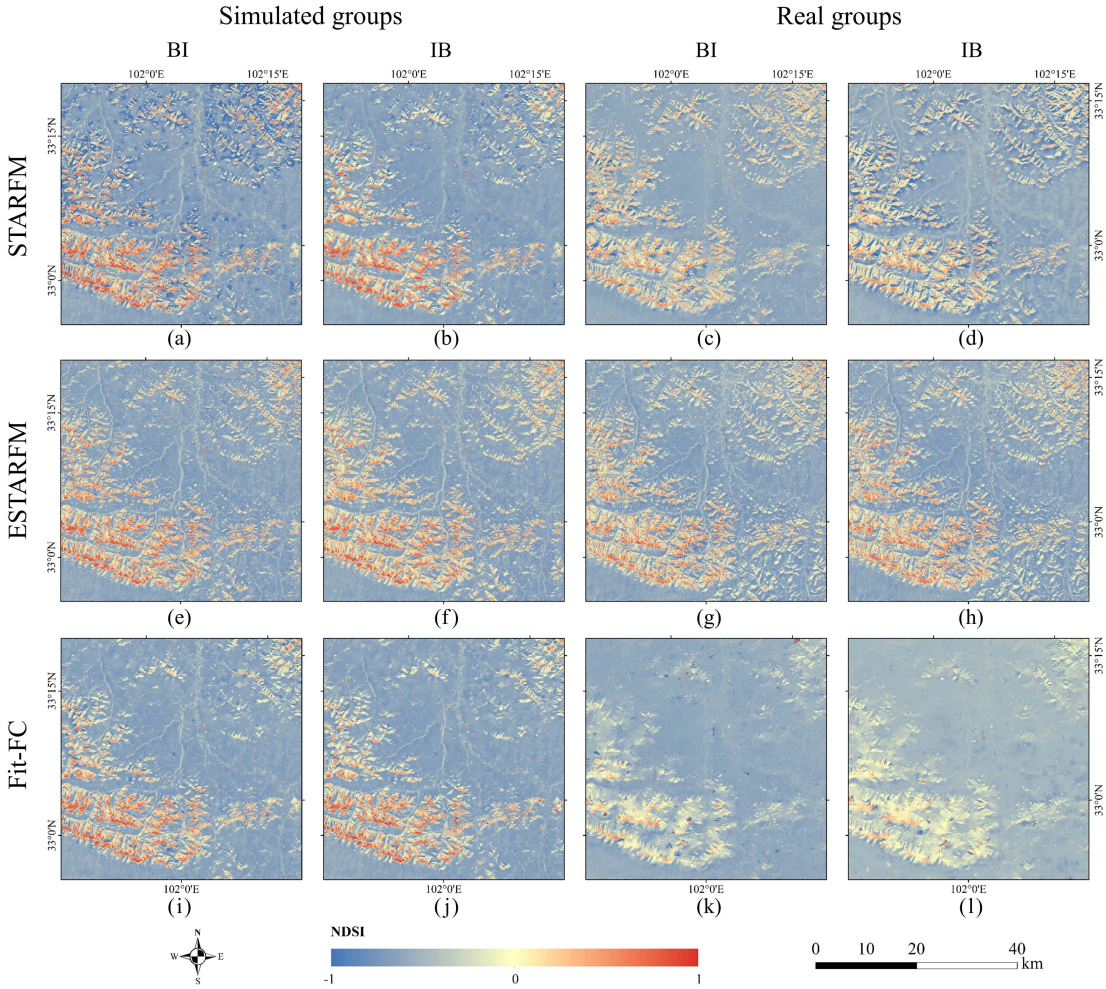

Fig. 11. NDSI fusion results for STARFM, ESTARFM, and Fit-FC under BI and IB strategies. Predicted images of simulated groups under BI and IB strategies by (a) and (b) STARFM, (e) and (f) ESTARFM, and (i) and (j) Fit-FC, respectively. Predicted images of real groups under BI and IB strategies by (c) and (d) STARFM, (g) and (h) ESTARFM, and (k) and (l) Fit-FC, respectively. All results were based on the input data from sample area 3 on February 26, 2021.

Related studies have shown that on the basis of distinguishing the vegetation growth stages (growth versus senescence), BI performed better when the predicted NDVI values were higher than the input NDVI values. Otherwise, IB performed better. When the images were homogeneous, both BI and IB can be used [38]. This may be related to the fact that the TP is mostly dominated by thin snow, and with multiple snow accumulation and ablation processes [55], [56]. Therefore, its tendency for strategy is not as obvious as that of vegetation. In addition, this is also influenced by cloud cover, so it needs to pay more attention to how to remove clouds at the same time with spatiotemporal fusion.

VI. Conclusion

In this study, we performed NDSI spatiotemporal fusion on the TP based on Sentinel-2A/B and MODIS data using the FSDAF model. We analyzed the effects of the BI and IB strategies on NDSI spatiotemporal fusion for the three snow cover periods. We also used simulated and real experimental groups to evaluate the accuracy of the NDSI fusion results in terms of five accuracy assessment metrics in spectral and spatial dimensions.

The results showed that the IB strategy was better in terms of spectral dimensions. In the simulated experimental group, the mean RMSE values of the fused NDSI under the IB strategy (0.175, 0.156, and 0.135) were lower than those under the BI strategy (0.182, 0.182, and 0.149), with an overall mean value 0.015 lower for the three snow cover periods. The overall mean values of R (0.824) and AD (-0.003) of the IB strategy were higher than those of the BI strategy ($R = 0.794$, AD $= -0.008$). In the real experimental group, the overall mean RMSE value under the IB strategy (0.146) was 0.072 lower than that of the BI strategy (0.218), and the overall mean values of R (0.752) and AD (-0.028) were higher than those of the BI strategy ($R = 0.672$, AD $= -0.050$) by approximately 0.080 and 0.022, respectively. From the spatial dimensions, there was no obvious pattern in the Edge and LBP values. In both experimental groups, the accuracy of the BI strategy was slightly higher than that of the IB strategy. However, there were cases where the accuracy of the IB strategy was higher in individual sample areas, whose texture and edge features could be well preserved in this context.

In addition, to validate the accuracy of the FSDAF model results, we conducted uncertainty analysis experiments on the results of three other spatiotemporal fusion models (STARFM, ESTARFM, and Fit-FC). The mean results of the simulated experimental group showed better results for the IB strategy (RMSE $= 0.151$, $R = 0.900$, AD $= 0.019$) in the spectral dimension and a slightly higher accuracy for BI in the spatial dimension (Edge $= -0.165$, LBP $= -0.013$). In the results of the real experimental group, all five accuracy assessment metrics indicated the superiority of the IB strategy. These results were in high agreement with those of the FSDAF model, indicating that the IB strategy was better.

Our study proved that the IB strategy is a more suitable fusion strategy for producing a high spatiotemporal resolution NDSI during different snow cover periods on the TP. This study provided an essential reference for the selection of an optimal strategy for the spatiotemporal fusion of NDSI on the TP and laid the foundation for producing better high-resolution snow cover products for long-time series on the TP.

References

[1] T. D. Yao et al., "Recent third pole's rapid warming accompanies cryospheric melt and water cycle intensification and interactions between monsoon and environment: Multidisciplinary approach with observations, modeling, and analysis," *Bull. Amer. Meteorological Soc.*, vol. 100, no. 3, pp. 423–444, Mar. 2019, doi: 10.1175/bams-d-17-0057.1.

[2] Q. L. You et al., "Review of snow cover variation over the Tibetan Plateau and its influence on the broad climate system," *Earth-Sci. Rev.*, vol. 201, Feb. 2020, Art. no. 103043, doi: 10.1016/j.earscirev.2019.103043.

[3] N. Pepin et al., "Elevation-dependent warming in mountain regions of the world," *Nature Climate Change*, vol. 5, no. 5, pp. 424–430, May 2015, doi: 10.1038/nclimate2563.

[4] M. Y. Dong, Y. Jiang, C. T. Zheng, and D. Y. Zhang, "Trends in the thermal growing season throughout the Tibetan Plateau during 1960–2009," *Agricultural Forest Meteorol.*, vol. 166, pp. 201–206, Dec. 2012, doi: 10.1016/j.agrformet.2012.07.013.

[5] D. Chen et al., "Assessment of past, present and future environmental changes on the Tibetan Plateau," *Chin. Sci. Bull.*, vol. 60, no. 32, pp. 3025–3035, Sep. 2015.

[6] H. Zou, J. H. Zhu, L. B. Zhou, P. Li, and S. P. Ma, "Validation and application of reanalysis temperature data over the Tibetan Plateau," *J. Meteorological Res.*, vol. 28, no. 1, pp. 139–149, Feb. 2014, doi: 10.1007/s13351-014-3027-5.

[7] Q. L. You et al., "Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences," *Earth-Sci. Rev.*, vol. 217, Jun. 2021, Art. no. 103625, doi: 10.1016/j.earscirev.2021.103625.

[8] A. F. Lutz, W. W. Immerzeel, A. B. Shrestha, and M. F. P. Bierkens, "Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation," *Nature Climate Change*, vol. 4, no. 7, pp. 587–592, Jul. 2014, doi: 10.1038/nclimate2237.

[9] S. C. Kang et al., "Linking atmospheric pollution to cryospheric change in the Third Pole region: Current progress and future prospects," *Nat. Sci. Rev.*, vol. 6, no. 4, pp. 796–809, Jul. 2019, doi: 10.1093/nsr/nwz031.

[10] T. Che, X. Hao, L. Dai, H. Li, X. Huang, and L. Xiao, "Snow cover variation and its impacts over the Qinghai-Tibet plateau," *Bull. Chin. Acad. Sci.*, vol. 34, no. 11, pp. 1247–1253, Oct. 2019.

[11] G. Q. Zhang et al., "Extensive and drastically different alpine lake changes on Asia's high plateaus during the past four decades," *Geophysical Res. Lett.*, vol. 44, no. 1, pp. 252–260, Jan. 2017, doi: 10.1002/2016gl072033.

[12] G. X. Wu, Y. M. Liu, B. He, Q. Bao, A. M. Duan, and F. F. Jin, "Thermal controls on the Asian summer monsoon," *Sci. Rep.*, vol. 2, May 2012, Art. no. 404, doi: 10.1038/srep00404.

[13] M. X. Yang, X. J. Wang, G. J. Pang, G. N. Wang, and Z. C. Liu, "The Tibetan Plateau cryosphere: Observations and model simulations for current status and recent changes," *Earth-Sci. Rev.*, vol. 190, pp. 353–369, Mar. 2019, doi: 10.1016/j.earscirev.2018.12.018.

[14] D. Qin, T. Yao, Y. Ding, and J. Ren, "The Cryospheric Science for sustainable development," *J. Geophysical Res.*, vol. 42, no. 1, pp. 1–10, Jun. 2020.

[15] W. F. Xu, L. J. Ma, M. N. Ma, H. C. Zhang, and W. P. Yuan, "Spatial-temporal variability of snow cover and depth in the Qinghai-Tibetan Plateau," *J. Climate*, vol. 30, no. 4, pp. 1521–1533, Feb. 2017, doi: 10.1175/jcli-d-15-0732.1.

[16] X. D. Huang, J. Deng, W. Wang, Q. S. Feng, and T. G. Liang, "Impact of climate and elevation on snow cover using integrated remote sensing snow products in Tibetan Plateau," *Remote Sens. Environ.*, vol. 190, pp. 274–288, Mar. 2017, doi: 10.1016/j.rse.2016.12.028.

[17] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006, doi: 10.1109/tgrs.2006.872081.

[18] X. L. Zhu, J. Chen, F. Gao, X. H. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, Nov. 2010, doi: 10.1016/j.rse.2010.05.032.

[19] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, Jan. 2016, doi: 10.1016/j.rse.2015.11.016.

[20] M. Liu et al., "An improved flexible spatiotemporal data fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sens. Environ.*, vol. 227, pp. 74–89, Jun. 2019, doi: 10.1016/j.rse.2019.03.012.

[21] T. Hilker et al., "A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, Aug. 2009, doi: 10.1016/j.rse.2009.03.007.

[22] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999, doi: 10.1109/36.763276.

[23] M. Q. Wu, J. Wang, Z. Niu, Y. Q. Zhao, and C. Y. Wang, "A model for spatial and temporal data fusion," *J. Infrared Millimeter Waves*, vol. 31, no. 1, pp. 80–84, Feb. 2012, doi: 10.3724/sp.J.1010.2012.00080.

[24] C. M. Gevaert and F. J. Garcia-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34–44, Jan. 2015, doi: 10.1016/j.rse.2014.09.012.

[25] B. Huang and H. H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012, doi: 10.1109/tgrs.2012.2186638.

[26] H. H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, Apr. 2013, doi: 10.1109/tgrs.2012.2213095.

[27] Z. X. Yin et al., "Spatiotemporal fusion of land surface temperature based on a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1808–1822, Feb. 2021, doi: 10.1109/tgrs.2020.2999943.

[28] Y. H. Zhang et al., "Spatial-temporal fraction map fusion with multi-scale remotely sensed images," *Remote Sens. Environ.*, vol. 213, pp. 162–181, Aug. 2018, doi: 10.1016/j.rse.2018.05.010.

[29] W. S. Li, C. Yang, Y. D. Peng, and X. Y. Zhang, "A multi-cooperative deep convolutional neural network for spatiotemporal satellite image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10174–10188, Sep. 2021, doi: 10.1109/jstars.2021.3113163.

[30] Y. H. Chen, K. X. Shi, Y. Ge, and Y. N. Zhou, "Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 4402112, doi: 10.1109/tgrs.2021.3069116.

[31] X. Liu, C. W. Deng, J. Chanussot, D. F. Hong, and B. J. Zhao, "StfNet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6552–6564, Sep. 2019, doi: 10.1109/tgrs.2019.2907310.

[32] J. Chen, L. Z. Wang, R. Y. Feng, P. Liu, W. Han, and X. D. Chen, "CycleGAN-STF: Spatiotemporal fusion via CycleGAN-based image generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5851–5865, Jul. 2021, doi: 10.1109/tgrs.2020.3023432.

[33] D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.*, vol. 54, no. 2, pp. 127–140, Nov. 1995, doi: 10.1016/0034-4257(95)00137-P.

[34] G. A. Riggs, D. K. Hall, and M. O. Roman, "Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records," *Earth Syst. Sci. Data*, vol. 9, no. 2, pp. 765–777, Oct. 2017, doi: 10.5194/essd-9-765-2017.
[35] O. Makarieva, A. Shikhov, N. Nesterova, and A. Ostashov, "Historical and recent aufeis in the Indigirka River basin (Russia)," *Earth Syst. Sci. Data*, vol. 11, no. 1, pp. 409–420, Mar. 2019, doi: 10.5194/essd-11-409-2019.
[36] G. A. Riggs, D. K. Hall, and M. O. Román, "MODIS snow products collection 6 user guide," Aug. 2016. [Online]. Available: https://modis-snow-ice.gsfc.nasa.gov/uploads/C6_MODIS_Snow_User_Guide.pdf
[37] Y. Jing, X. Li, and H. Shen, "STAR NDSI collection: A cloud-free MODIS NDSI dataset (2001–2020) for China," *Earth Syst. Sci. Data*, vol. 14, no. 7, pp. 3137–3156, Jul. 2022, doi: 10.5194/essd-14-3137-2022.
[38] X. Chen, M. Liu, X. Zhu, J. Chen, Y. Zhong, and X. Cao, "'Blend-then-index' or 'index-then-blend': A theoretical analysis for generating high-resolution NDVI time series by STARFM," *Photogrammetric Eng. Remote Sens.*, vol. 84, no. 2, pp. 66–74, Feb. 2018, doi: 10.14358/pers.84.2.65.
[39] A. A. Jarihani, T. R. McVicar, T. G. Van Niel, I. V. Emelyanova, J. N. Callow, and K. Johansen, "Blending Landsat and MODIS data to generate multispectral indices: A comparison of 'index-then-blend' and 'blend-then-index' approaches," *Remote Sens.*, vol. 6, no. 10, pp. 9213–9238, Oct. 2014, doi: 10.3390/rs6109213.
[40] C. Huang, Y. Chen, S. Q. Zhang, L. Y. Li, K. F. Shi, and R. Liu, "Surface water mapping from Suomi NPP-VIIRS imagery at 30 m resolution via blending with Landsat data," *Remote Sens.*, vol. 8, no. 8, Aug. 2016, Art. no. 631, doi: 10.3390/rs8080631.
[41] J. Qiu, "China: The Third Pole," *Nature*, vol. 454, no. 7203, pp. 393–396, Jul. 2008, doi: 10.1038/454393a.
[42] Y. Zhang, B. Li, L. Liu, and D. Zheng, "Redetermine the region and boundaries of Tibetan Plateau," *Geographical Res.*, vol. 40, no. 6, pp. 1543–1553, Jun. 2021.
[43] K. Y. Fang, P. Zhang, J. M. Chen, and D. L. Chen, "Co-varying temperatures at 200 hPa over the Earth's three poles," *Sci. China Earth Sci.*, vol. 64, no. 2, pp. 340–350, Feb. 2021, doi: 10.1007/s11430-020-9680-y.
[44] M. Xu, S. C. Kang, X. L. Chen, H. Wu, X. Y. Wang, and Z. B. Su, "Detection of hydrological variations and their impacts on vegetation from multiple satellite observations in the Three-River Source Region of the Tibetan Plateau," *Sci. Total Environ.*, vol. 639, pp. 1220–1232, Oct. 2018, doi: 10.1016/j.scitotenv.2018.05.226.
[45] SUHET, *Sentinel-2 User Handbook*. Jul. 2015. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook.pdf
[46] E. F. Vermote and R. E. Wolfe, "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061," NASA EOSDIS Land Processes Distributed Active Archive Center, 2021, doi: 10.5067/MODIS/MOD09GA.061.
[47] D. K. Hall and G. A. Riggs, "MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 61," NASA National Snow and Ice Data Center Distributed Active Archive Center, 2021, doi: 10.5067/MODIS/MOD10A1.061.
[48] V. V. Salomonson and I. Appel, "Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1747–1756, Jul. 2006, doi: 10.1109/tgrs.2006.876029.
[49] C. Liu, Z. Li, P. Zhang, J. Y. Zeng, S. Gao, and Z. J. Zheng, "An assessment and error analysis of MOD10A1 snow product using Landsat and ground observations over China during 2000–2016," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1467–1478, Apr. 2020, doi: 10.1109/jstars.2020.2983550.
[50] X. L. Zhu et al., "A novel framework to assess all-round performances of spatiotemporal fusion models," *Remote Sens. Environ.*, vol. 274, Jun. 2022, Art. no. 113002, doi: 10.1016/j.rse.2022.113002.
[51] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. van Dijk, "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.*, vol. 133, pp. 193–209, Jun. 2013, doi: 10.1016/j.rse.2013.02.007.
[52] D. Z. Guo, W. Z. Shi, M. Hao, and X. L. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 111973, doi: 10.1016/j.rse.2020.111973.
[53] Q. M. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, Jan. 2018, doi: 10.1016/j.rse.2017.10.046.
[54] J. Li, Y. F. Li, R. L. Cai, L. He, J. Chen, and A. Plaza, "Enhanced spatiotemporal fusion via MODIS-like images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5610517, doi: 10.1109/tgrs.2021.3106338.
[55] X. Huang, X. Li, C. Liu, M. Zhou, and J. Wang, "Remote sensing inversion of snow cover extent and snow depth/snow water equivalent on the Qinghai-Tibet Plateau: Advance and challenge," *J. Geophysical Res.*, vol. 41, no. 5, pp. 1138–1149, Oct. 2019.
[56] Y. L. Wang, X. D. Huang, J. S. Wang, M. Q. Zhou, and T. G. Liang, "AMSR2 snow depth downscaling algorithm based on a multifactor approach over the Tibetan Plateau, China," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111268, doi: 10.1016/j.rse.2019.111268.

Linxin Dong received the B.S. degree in remote sensing science and technology from Chengdu University of Technology, Chengdu, China, in 2022. She is currently working toward the M.S. degree in cartography and geography information systems with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China.

Her current research interests include spatiotemporal fusion strategies and cryosphere remote sensing.

Haixi Zhou received the B.S. degree in geographic information science in 2022 from East China Normal University, Shanghai, China, where he is currently working toward the M.S. degree in geomatics engineering with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences.

His current research interests include spatiotemporal fusion model and machine learning.

Jiahui Xu received the B.S. degree in geographic information science from Chongqing Normal University, Chongqing, China, in 2021. She is currently working toward the Ph.D. degree in cartography and geography information systems with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China.

Her current research interest is remote sensing of snow hydrology.

Yao Tang received the B.S. degree in geographic information science in 2021 from East China Normal University, Shanghai, China, where he is currently working toward the M.S. degree in geomatics engineering with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences.

His current research interest is cryosphere remote sensing.

Xiaolong Teng received the B.S. degree in human geography and urban-rural planning from Nanjing Agricultural University, Nanjing, China, in 2022. He is currently working toward the M.S. degree in geomatics engineering with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China.

His current research interests include polar remote sensing and cryosphere remote sensing.

Jingwen Ni received the B.S. degree in geographic information science in 2023 from East China Normal University, Shanghai, China, where she is currently working toward the M.S. degree in cartography and geography information systems with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences.

Her current research interests include vegetation remote sensing and cryosphere remote sensing.

Bailang Yu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in cartography and geographic information systems from East China Normal University, Shanghai, China, in 2002 and 2009, respectively.

He is currently a Professor with the Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, where he is also with the School of Geographic Sciences. His research interests include urban remote sensing, nighttime light remote sensing, light detection and ranging (LiDAR), and object-based methods.

Jianping Wu received the M.S. degree in cartography and remote sensing from Peking University, Beijing, China, in 1986, and the Ph.D. degree in regional geography from East China Normal University, Shanghai, China, in 1996.

He is currently a Professor with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University. His research interests include remote sensing and geographic information system.

Yan Huang received the B.S. degree in geographic information science from Chengdu University of Technology, Chengdu, China, in 2010, and the Ph.D. degree in cartography and geographic information systems from East China Normal University, Shanghai, China, in 2016.

She is currently an Associate Professor with the Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, where she is also with the School of Geographic Sciences. Her research interests include spatial and temporal modeling analysis of remote sensing images, cryosphere remote sensing, and global climate and environmental change.