A Self-Tuning Algorithm for Compositing Clear-Sky Reflectance Data at 250-Meter Spatial Resolution for the Fengyun-3D MERSI-II Land Bands Over China

Wenzhuo Li, Kaimin Sun, *Member, IEEE*, Hongjuan Zhang, Shunxia Miao, *Student Member, IEEE*, Fangyi Lv, Xiuqing Hu, and Hongya Zhang

***Abstract*—The MEdium Resolution Spectral Imager-II (MERSI-II) on board the Fengyun-3D satellite is an advanced imaging instrument with 250/1000 m spatial resolution. This instrument is a valuable resource for terrestrial remote sensing, such as clear-sky surface reflectance products. Clear sky composites are typically generated using fixed and empirical threshold values based on specific land surface type, climate, and weather conditions. Moreover, existing methods usually require cloud masks, which can be inaccurate and affect clear sky composites. We took inspiration from natural language processing and viewed the time-series reflectance data as per-pixel temporal sequences. A new self-tuning approach transforms the temporal reflectance data to contextualized embeddings in a high-dimensional feature space through pretraining, followed by a classification layer that is fine-tuned by a few labeled data to identify scene ID mask (cloud shadow, clear sky, thin cloud, and thick cloud). We then proposed a compositing strategy that takes into account the spatial connectivity and temporal consistency of clear-sky pixels. Comprehensive evaluations, including visual assessment, annual temporal profiles of NDVI/NDWI at a site scale, and NDVI spatial distribution at a national scale, of the proposed approach applied on 250 m bands were conducted by comparing it with four typical compositing methods (CCRS-BH, CCRS-FH, MaxNDVI, and MinRed), as well as the MODIS clear sky products. The proposed approach outperforms the four typical compositing methods and achieves a similar product level to MODIS composited products but with a higher spatial resolution.**

Manuscript received 3 March 2024; revised 16 April 2024; accepted 21 May 2024. Date of publication 4 June 2024; date of current version 18 June 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3902900 and in part by the National Natural Science Foundation of China under Grant 42301457 and Grant 42192583. *(Corresponding authors: Kaimin Sun; Hongjuan Zhang.)*

Wenzhuo Li is with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: alvinleelwz@gmail.com).

Kaimin Sun, Hongjuan Zhang, Shunxia Miao, and Fangyi Lv are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: sunkm@whu.edu.cn; hongjuanzhang@whu.edu.cn; shunxiamiao@whu.edu.cn; fangyilv@whu.edu.cn).

Xiuqing Hu is with the Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, China Meteorological Administration, National Satellite Meteorological Center, Beijing 100081, China (e-mail: huxq@cma.gov.cn).

Hongya Zhang is with the Changjiang River Scientific Research Institute of Changjiang Water Resources Commission, Wuhan 430010, China (e-mail: hongyazh@foxmail.com).

Digital Object Identifier 10.1109/JSTARS.2024.3408873

***Index Terms*—250-m spatial resolution, clear-sky, clear-sky composited, Fengyun-3D MEdium Resolution Spectral Imager-II (MERSI-II), surface reflectance products, temporal compositing.**

I. Introduction

LAUNCHED in November 2017, the MEdium Resolution Spectral Imager-II (MERSI-II) on board Fengyun-3D satellite has emerged as an advanced sensor that offers valuable capabilities for large-scale terrestrial applications [1]. The MERSI-II has global coverage every day in 25 spectral bands and can provide global true color images and thermal infrared data (twice a day) at 250-m spatial resolution [2]. Various MERSI-II land products can be produced at different levels of data processing from the MERSI-II image data [3]. The quality of images is of paramount importance and should be considered on par with the development of processing algorithms. Owing to the complex atmospheric conditions, a significant portion, approximately 70%, of the Earth's surface is frequently obscured by cloud cover, so most images cannot be directly used [4]. The objective of this article is to present a novel method for differentiating clear-sky pixels from those affected by clouds and cloud shadows, using MERSI-II data, to produce clear-sky composites that are well-suited for subsequent retrieval of diverse land parameters.

To produce clear-sky composites, the pixel level scene ID mask (clear versus cloudy) should be first derived to find clear-sky pixels from time-series data, which are composited afterward. The algorithms employed for scene identification can be broadly categorized into two main types: single-criteria methods and multicriteria methods [5], [6], [7]. The majority of existing single-criteria methods commonly utilize minimum or maximum rules in their approach. One notable example is the maximum-value composite (MVC) procedure, initially introduced by Holben [8] and applied to the multitemporal AVHRR data. However, the MVC technique exhibited significant instability when applied in semi-arid regions characterized by medium-to-low density vegetation cover [9], [10], [11], [12]. To mitigate this issue, several researchers have put forth the incorporation of minimum-value criteria to enhance the robustness of MVC for multitemporal compositing. These

approaches utilize minimum reflectance values in visible wavelengths, such as the blue or red bands, to effectively differentiate clear-sky pixels from cloud-contaminated pixels [12], [13]. But this criterion hardly distinguishes between cloud shadows and clouds. Some authors have recommended the inclusion of a third minimum value based on an assumption that cloud shadows hardly appeared more than twice in a given pixel [14]. While single-criteria methods are computationally straightforward, they have limitations in fully addressing the challenges posed by cloud contamination. These methods are only effective for specific surface cover conditions and lack explicit consideration of the spectral characteristics among different surface types.

Multicriteria methods typically exploit the known spectral variations of specific surface types to identify pixels by employing a sequence of threshold tests [6], [15]. Luo et al. [9] proposed a criteria matrix scheme, which is a decision-tree algorithm specifically designed for mid-latitude MODIS observations over land, using only visible and shortwave near-infrared bands. Building upon the similar bands available on the Thematic Mapper (TM) of Landsat-5, Enhanced TM Plus (ETM+) of Landsat-7, and additional bands of Landsat-8, Oreopoulos et al. improved the performance of the criteria matrix scheme. As a result, the scheme achieved a remarkable accuracy of approximately 93% in correctly distinguishing between clear and cloudy pixels [16], [17]. While multicriteria methods proved to be more effective compared to single-criterion methods, they still did not fully account for phenological variations in the land surface. To address this limitation, Li and Wang [18] specifically concentrated on cropland areas and developed a clear-sky composite approach for cropland using data from Landsat-7 and Landsat-8 with the help of the MODIS cloud mask products. Bian et al. [19] introduced a self-adaptive compositing approach aimed at generating global terrestrial clear-sky composites of VIIRS surface reflectance data. Frantz et al. [15] conducted a similar study wherein they dynamically adapted the selection process according to the specific land surface phenology of individual pixels. By adopting this approach, it becomes feasible to generate phenologically accurate composites for extensive regions at regular intervals, encompassing various phenological stages. However, the phenological variations of land surface covers are easily affected by climate and weather, and the sensors could also receive the signal of land surface through thin clouds [20]. Hence, fixed and empirical threshold values in the above works for determining surface cover conditions will lead to failure in some cases. In addition, multicriteria methods usually use cloud mask products [21], [22] as auxiliary data in the process. When compositing clear-sky imagery, a multicriteria approach was employed to select the most suitable pixel value from a pool of clear-sky candidates spanning the entire compositing period. However, this method may lead to the inclusion of clear-sky values captured under diverse geometric and weather conditions, consequently introducing spatial inconsistency in the final composite imagery.

In this article, we introduce an innovative approach that effectively tackles the aforementioned challenges associated with the generation of eight-day clear-sky composites at a spatial resolution of 250 m over China, utilizing MERSI-II data. Our primary contributions encompass the following key aspects.

1) A three-stage change pattern, corresponding to cloud shadow, clear sky, and cloud, was found in the temporal reflectance data sequence when organized in ascending order. This pattern inspired us to propose a self-tuning determination (ST-Det) approach for identifying scene ID masks (cloud shadow, clear sky, thin cloud, and thick cloud) for each pixel using time-series reflectance data from four land bands.
2) The proposed ST-Det approach employed a neural network model (NNM). To our best knowledge, ST-Det is the first NNM-based approach utilizing time-series analysis of reflectance data from green band (470 nm), blue band (550 nm), red band (650 nm), and NIR band (865 nm) to generate clear sky composites. First, abundant unlabeled temporal reflectance data in ascending order from visible bands and NIR bands were pretrained to achieve contextualized embeddings in a high-dimensional feature space that captured the temporal characteristics within and among the four bands. Then a classification layer was fine-tuned by a few labeled data to identify the scene ID mask automatically. The proposed approach got rid of the limitation of fixed and empirical threshold values in determining surface cover conditions, as well as the dependency on cloud mask products.
3) To alleviate substantial spatial inconsistency of compositing imagery, a compositing strategy is proposed considering both spatial connectivity and temporal consistency of clear-sky pixels.
4) The clear-sky composites produced by the proposed approach were comprehensively compared with four typical compositing methods, namely, CCRS-BH [9], CCRS-FH [9], MaxNDVI [8], MinRed [12], and with the MODIS MYD09A1 clear sky compositing products [23], in terms of visual assessment, annual temporal profiles of NDVI/NDWI at a site scale and NDVI spatial distribution at a national scale. Results proved that our approach outperformed the four typical compositing methods and achieved a similar product level to MODIS composited products but with a higher spatial resolution.

II. Methodology

A. Overall Concept

The literature overview highlights the need for considering spectral and phenological characteristics of land surface cover types when producing scene ID masks, for example, the reflectance of vegetation and water or snow/ice are quite different. To investigate the changing pattern of reflectance under different cloud conditions among different land surface cover types, we first analyze the temporal characteristics of reflectance over one year from the green band for different land surface cover type samples. After sorting the temporal reflectance in ascending order, we find a three-stage change pattern that can correspond to shadow, clear sky, and cloud, respectively, regardless of land surface cover types.

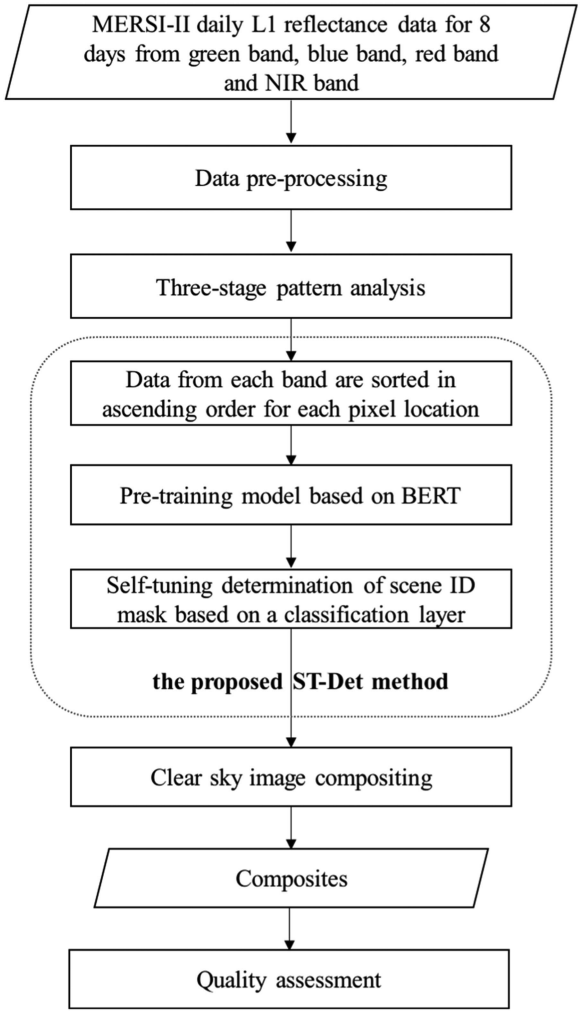

Fig. 1. Flowchart of our compositing strategy.

Based on the above-mentioned findings, we propose a ST-Det method of the scene ID mask, where the change patterns are pretrained using bidirectional encoder representations from transformer (BERT) model via abundant unlabeled data from the green band, blue band, red band, and NIR band. After pretraining, we achieve the contextualized embeddings in a high-dimensional feature space that captures the temporal characteristics within and among the four bands. A classification layer is used for the ST-Det of the scene ID mask for each pixel via labeled data. Finally, a compositing strategy is proposed considering both spatial connectivity and temporal consistency of clear-sky pixels to generate final composites. Note that land surface types are not needed in both the pretraining and ST-Det stages. The flowchart of our approach is shown in Fig. 1.

B. Data Preprocessing

In this work, a total volume of 2168 GB MERSI-II L1 data in daytime orbits of the year 2020 are processed. Obtained from the National Satellite Meteorological Center of China, the standard MERSI-II L1 product is provided in 5-min data blocks in HDF

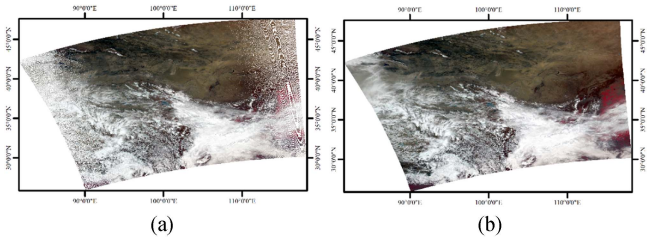

Fig. 2. NAN pixels fill before (left) and after (right) during reprojection.

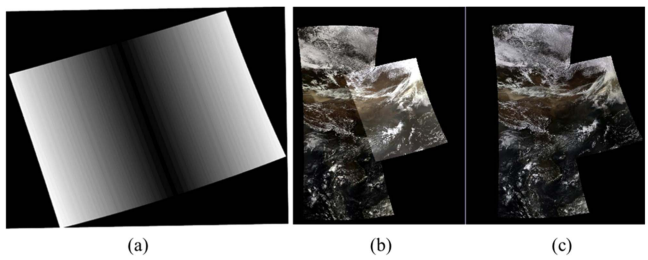

Fig. 3. (c) Mosaicking image based on the (a) satellite zenith angle mask in which the darker the color, the higher the zenith angle, using (b) raw images from different tracks.

format, including surface reflectance DN value, calibration parameters, and geolocation data. The ground surface reflectance is acquired using a simple atmospheric correction from the MODIS Corrected Reflectance Science Processing Algorithm [24], with appropriate adjustment for MERSI-II. We use the 1 km geodata with high geolocation accuracy [25] to reproject the reflectance data to the geographic latitude/longitude projection. To fill in the missing data of the reprojected image, the nearest-neighbor method is employed. In Fig. 2(a), we can see that the edges of the image have NAN pixels, which were filled during the reprojection process. The pixel size of the mosaicked image is 11804 rows by 20 549 columns for China-bound at a spatial resolution of 250 m, which is also the same for the final composites.

Since the swath of MERSI II is 2900 km, to cover China, typically we need two or three overlapping tracks as shown in Fig. 3(b), with each track containing about two to four 5-min data blocks. To keep spectral consistency in the overlapping areas when mosaicking the overlapping tracks into daily data over China as shown in Fig. 3(c), we select the pixel values with higher satellite zenith angles and closer to the nadir of the satellite based on a zenith angle mask, as shown in Fig. 3(a).

C. Developing Shadow, Clear-Sky, Thin Cloud, and Thick Cloud Mask

To composite clear-sky imagery, a per-pixel scene ID mask (clear versus cloudy) should be first derived to find clear-sky pixels from time-series data. In this work, we classify the masks into shadow, clear sky, thin cloud, and thick cloud.

1) Change Pattern Analysis of Reflectance for Different Land Surface Cover Types: To identify the changing pattern of reflectance of different land surface cover types, we choose five typical land surface cover types and take green band data as

TABLE I
SAMPLES FOR THE TYPICAL SURFACE COVER TYPES

Surface cover types	Sub classes	Description	Number of locations
water	PWT	permanent water without ice or snow	10
	WSI	water that may be covered by snow and ice during winter	10
barren land	BRL	bare soil, barren land	10
vegetation	DVT	deciduous vegetation in tropical areas	10
	DVN	deciduous vegetation in non-tropical areas that may be covered by snow during winter	10

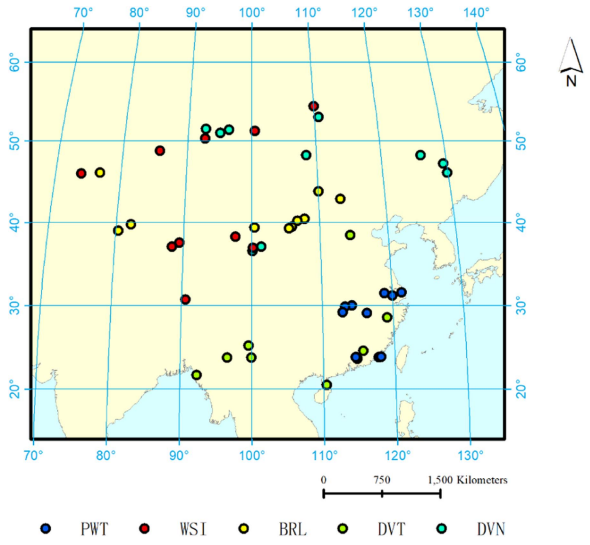

Fig. 4. Geographical distribution of the samples for different land surface cover types. See the subclasses in Table I for the surface type represented by each abbreviation.

an example. The five land surface cover types are described in Table I, including water, barren land, and vegetation. The green band data over one year from 10 locations are chosen for each land surface cover type. The distribution of the locations is illustrated in Fig. 4.

Daily reflectance data for all the samples of different land surface cover types from the green band of MERSI-II data in 2020 were analyzed in Fig. 5. In Fig. 5 left column, the grey lines are time series reflectance data over the whole year from 1 location for each surface cover type. We can see that the daily time series of reflectance shows chaotic fluctuation over time since the random clouds exhibit high reflectance values whereas the random cloud shadows represent low reflectance values. The chaotic fluctuation can be found for all selected samples and there are no distinct differences among the samples. In other words, it is hard to find clear-sky days and distinguish different land surface cover types from the daily time series reflectance data.

Therefore, we sort the daily time series reflectance in ascending order and segment the sorted line based on the scene ID mask, which is obtained by visual inspection (cloud shadow with blue line, clear sky with red line, and cloudy with green line) in Fig. 5 left column. After sorted, the data shows a three-stage change pattern that data first shows a short increase (the first stage, denoted as blue lines), then follows a relatively flat region (the second stage, denoted as red lines), and finally represents a steep increase (the third stage, denoted as green lines). To provide a more vivid illustration of this pattern, we calculate the averaged reflectance slope between adjacent data points for each stage

$$\text{mean reflectance slope} = \begin{cases} \frac{1}{n_1-1}\sum_{i=1}^{n_1-1}(x_{i+1}-x_i) \\ \frac{1}{n_2-1}\sum_{i=1}^{n_2-1}(x_{i+1}-x_i) \\ \frac{1}{n_3-1}\sum_{i=1}^{n_3-1}(x_{i+1}-x_i) \end{cases} \quad (1)$$

where n_1, n_2, and $n_3$are numbers of reflectance data during shadow, clear-sky, and cloud, respectively, and x represents the sorted reflectance data. The values of the mean reflectance slope are put above the blue, red, and green lines in Fig. 5 left column. Generally, the averaged slope of the second stage is less than 0.001, which is much smaller than the mean slopes of the first and third stages. Even for the land surface cover type that varies with time, such as the vegetated area losing the green leaves during the leaf-off season in DVT and DVN, a relatively flat clear-sky curve can be found at a one-year time scale.

This three-stage change pattern is closely related to the scene ID mask. Fig. 5 middle column and right column show the statistical distributions of reflectance data and reflectance slope for all the samples at 10 locations of different land surface cover types during the three stages. As expected in Fig. 5 left column, the reflectance is relatively low when there are cloud shadows, relatively high when there are clouds, and in the middle when the sky is clear, which can also be inferred from Fig. 5 middle column. In Fig. 5 right column, we can see that the mean and median reflectance slopes of clear-sky are much smaller than those of shadow and cloud. Moreover, the variance of the reflectance slop of clear-sky is also much smaller than those of shadow and cloud, which further proves that the clear-sky days should have middle reflectance values with a flat tendency.

Even though different land surface cover types show distinctive spectral characteristics, their sorted daily reflectance represents similar change patterns and statistical patterns under natural climate conditions with random cloudy or clear skies. The change pattern and statistical pattern provide good guide information for automatically choosing clear-sky days and identifying the scene ID mask simultaneously. In the following section, we input the sorted reflectance data to a neural network to learn the change pattern automatically, producing pixel-level shadow, clear-sky, thin cloud, and thick cloud mask.

2) ST-Det of Pixel-Level Scene ID Mask Based on BERT Model: The BERT model [26] was first introduced by Google in 2018, which was originally set up to deal with natural language processing (NLP) tasks. It is able to learn rich representations of words and sentences, capturing various syntactic and semantic properties of natural language. In BERT, words are defined by

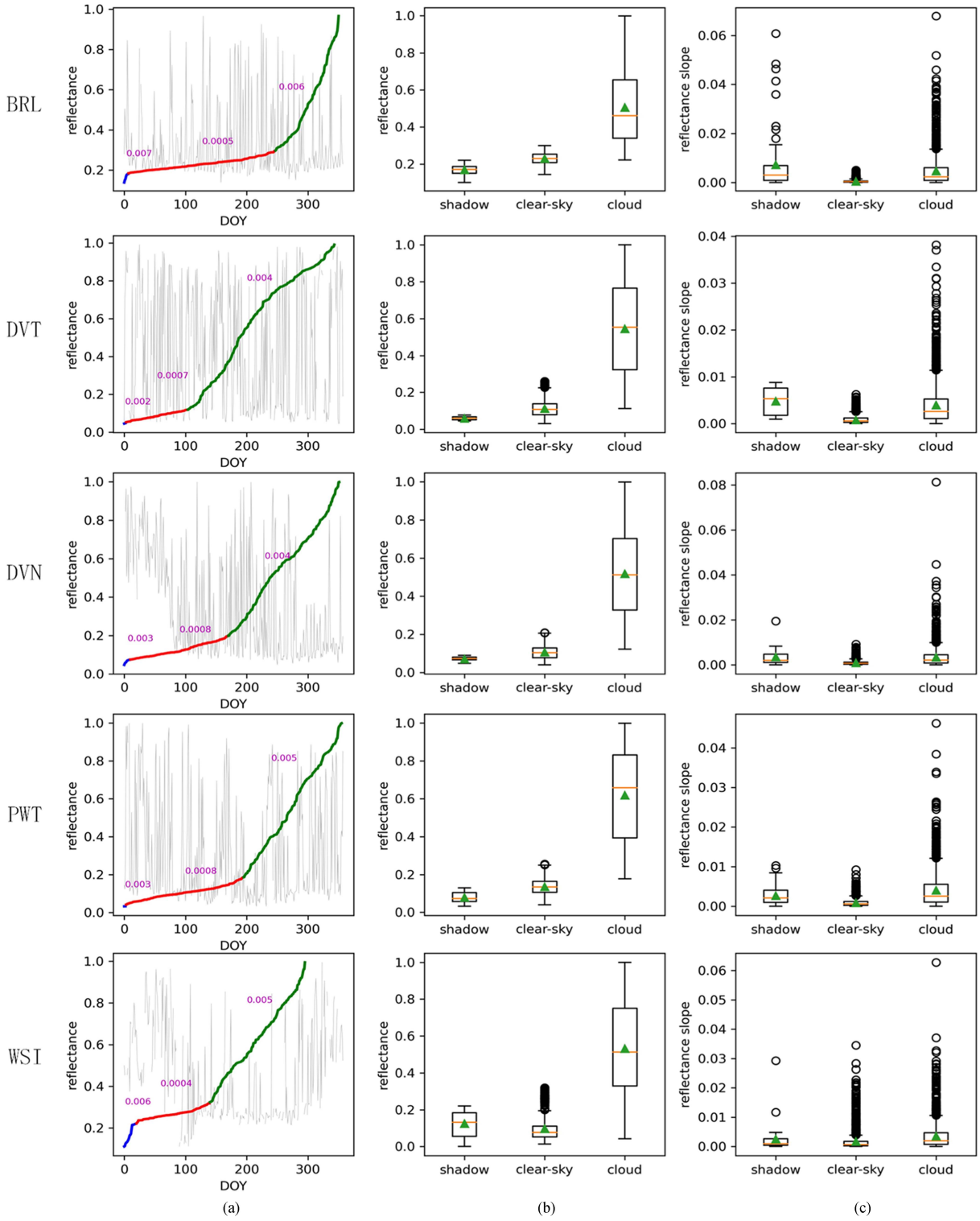

Fig. 5. Temporal reflectance from the green band for different land surface cover types over one year. (a) Left columns are reflectance data from one location, where grey lines are daily time series of reflectance, blue lines, red lines, and green lines represent reflectance sorted in ascending order, corresponding to temporal windows of shadow, clear sky, and cloud, respectively. The values on the lines of the three colors are the averaged slope between adjacent data points. (b) Middle columns and (c) right columns are boxplots of reluctance data from 10 locations, and boxplots of reflectance slope from 10 locations, respectively, where the orange line represents the median value and the green triangle represents the mean value. (a) 1 location. (b) 10 locations. (c) 10 locations.

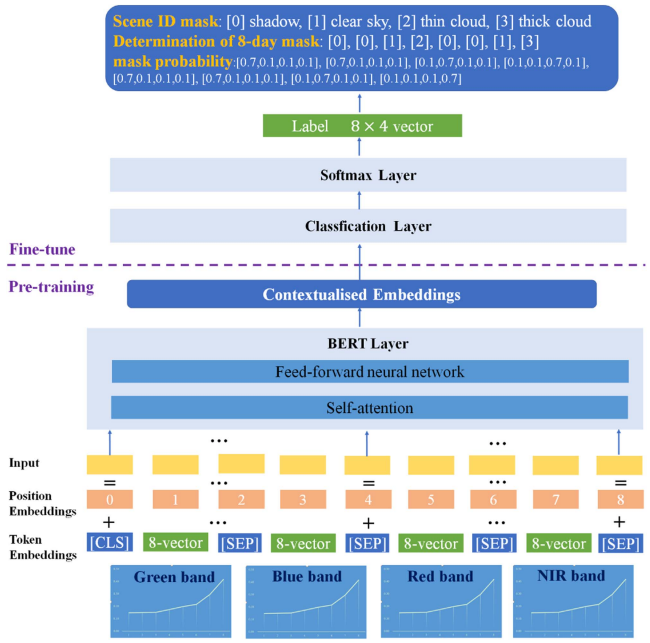

Fig. 6. Structure of the processing network.

their surroundings, not by a prefixed identity. Therefore, contrasted against the traditional methods of NLP, it was pretrained using only an unlabeled, plain text corpus.

There has been limited research on using this model for significant tasks in the field of satellite image processing, such as identifying clouds and creating clear-sky compositions. In Section II-C1, we find that although the specific patterns vary for different land surface cover types and are affected by cloud appearance rate, or pixel probability distribution in shadow, clear-sky, and cloudy conditions, their trend is consistent and statistical patterns are similar. Moreover, the BERT model is internally designed to encode these seemingly random sequential/temporal features and is highly suitable for downstream classification tasks. In NLP, the order of each word in a sentence may appear random. During pretraining with BERT, a masked language model is used where some words in the input text are randomly selected and replaced with a special [MASK] token. The model learns the representation of each word in the context, while also using the other words in the context to predict the masked words. Our time series reflectance data is similar to sentences of words. Therefore, the BERT model is suitable for our task, which can deal with the randomness of shadow, clear-sky, and cloudy conditions. The structure of the processing network is shown in Fig. 6.

Taking advantage of the pretraining method of BERT, the sorted reflectance from eight days from the four bands at the pixel level is treated as unlabeled data. The pretraining of BERT is a self-supervised learning process, where a large amount of unlabeled data can be utilized by mapping temporal reflectance to predefined tokens. Each pixel position serves as a sample, allowing for the convenient acquisition of a substantial number of unlabeled samples for pretraining. In our work, the pretraining initially involves tokenization of reflectance values. The token construction involves multiplying the reflectance data by 1000, rounding to the nearest integer, and adding a shift for four bands with 0, 1000, 2000, and 3000, respectively, resulting in a vocabulary of 4000 words (tokens). This method helps distinguish the values for four spectral bands. For a single pixel sample, the temporal reflectance from four bands over eight days comprises 32 tokens. All unlabeled samples participating in pretraining are constructed following the same rules.

Then, the BERT model employs token embeddings and position embeddings as input. We put the unlabeled data as token embeddings by concatenating the four vectors of eight-day reflectance. The four vectors are separated by a SEP field. Position embeddings, on the other hand, provide information about the position of each token in the given data. They are used to capture the sequential structure of the input data by assigning each position in the data a fixed-size vector representation. This is achieved by adding a positional encoding vector to each token embedding. Together, the token embeddings and position embeddings allow the BERT model without explicitly recurrently encoding the sequence information.

After the pretraining of the BERT model, we achieve the contextualized embeddings in a high-dimensional feature space that captures the temporal characteristics within and among the four bands. In formulation, given a target pixel (the pixel for which we want to predict the eight-day scene ID mask), we want to map an input vector $X^{8\times4}$, which consists of eight-day reflectance from the four bands, to a feature vector $Y^{128\times1}$, where 128 is the dimension of a hidden layer of BERT. The network can be described as a mapping function, given by

$$Y^{128\times1} = f_{\text{FFN}}\left(f_{\text{MA}}\left(X^{8\times4};\theta\right);\vartheta\right) \tag{2}$$

where f_{MA} represents the multihead self-attention module with the corresponding parameter set θ and f_{FFN} refers to the feed-forward neural network with the corresponding parameter set ϑ.

Since the pretraining phase of BERT is an self-supervised method that only inputs the raw reflectance data without labels, it is effortless to obtain a large amount of unlabeled data. We use reflectance data from the year 2019 in different eight-day periods with almost one million pixels to complete the pretraining of BERT. Note that, the Mini-BERT model is used in this work to reduce the number of parameters. After pretraining, we use labeled data to fine-tune the classification head. The learned contextualized embeddings are then put into a fully connected layer (classification layer), followed by SoftMax for scene ID mask prediction. The cross-entropy loss is used for the model optimization in the fine-tuning stage

$$\text{Cross Entropy}_{\text{LOSS}} = -\sum_{j=1}^{4} p_j \log\left(\tilde{p}_j\right) \tag{3}$$

where j means the jth scene ID mask, which includes shadow, clear sky, thin cloud, and thick cloud, $\tilde{p}_j$ is the probability score inferred by the SoftMax function for class j, and p_j is the ground truth. We run the fine-tune of Mini-BERT for 50 epochs and save the model with minimal loss.

TABLE II
NUMBER OF LABELED DATA POINTS FOR FINE-TUNING AND VALIDATION COVERING FIVE LAND SURFACE COVER TYPES AND FOUR SCENE CLASSES

Land surface cover types	Number of labeled data points for fine-tuning				Number of labeled data points for validation			
	shadow	clear sky	thin cloud	thick cloud	shadow	clear sky	thin cloud	thick could
BRL	7872	28 864	12 370	37 486	1779	5858	2715	4224
DVT	9057	32 608	11 645	28 210				
DVN	7452	20 494	16 395	37 635				
PWT	132	14 238	6158	10 264				
WSI	251	15 219	7609	14 969				

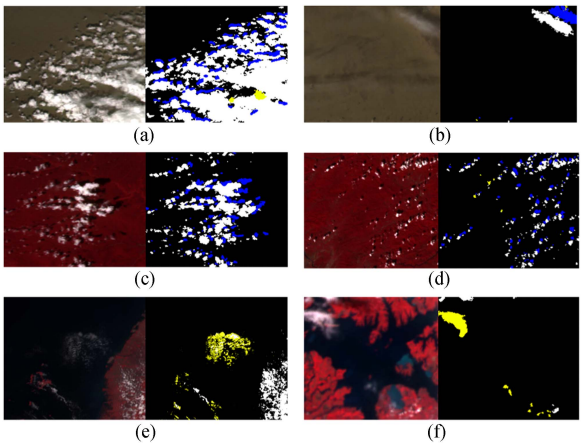

Fig. 7. Some samples of our labeled data that covers the five land surface cover types. The first and third columns are images, and the second and fourth columns are scene ID masks where blue, black, yellow, and white mean cloud shadow, clear sky, thin cloud, and thick cloud, respectively. (a) BRL. (b) BRL. (c) DVT. (d) DVN. (e) PWT. (f) WSI.

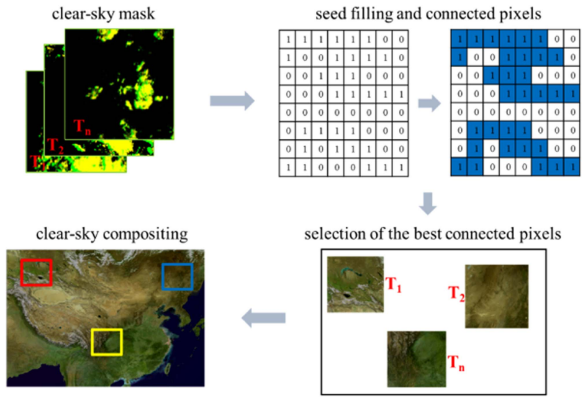

Fig. 8. Flowchart of clear-sky imagery compositing.

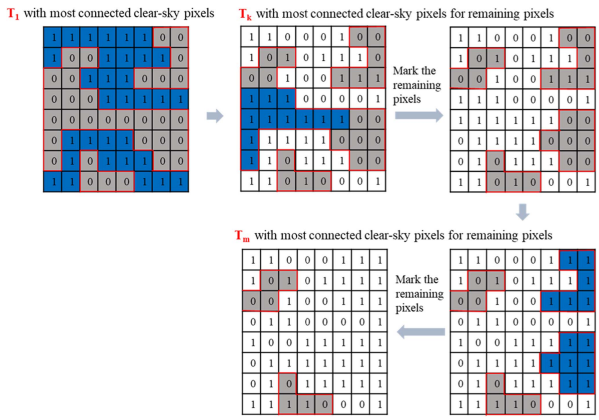

Fig. 9. Selection of clear-sky pixels for compositing.

In total, 39 866 labeled pixels covering the five land surface cover types are obtained by visual inspection from year 2019 in different eight-day period for fine-tuning. Table II shows the number of labeled data points including time-series data. To make our dataset more reasonable, we select enough data for the five land surface cover types and the four scene classes during fine-tuning. Fig. 7 shows some samples of our labeled data. In validation, we mainly focus on pixel-level classification results, therefore the dataset does not distinguish between land surface cover types. After the classification layer is fine-tuned, MERSI-II data over China for 2020 is input into our network to determine the daily scene ID mask for each pixel.

D. Clear-Sky Image Compositing

Traditional compositing methods apply several criteria for selecting the best pixel among various candidates at very diverse geometry, which introduces a substantial spatial inconsistency. To effectively minimize spatial inconsistency in the composites, we propose a novel compositing strategy (refer to Fig. 8) that incorporates the principles of spatial connectivity and temporal consistency of clear-sky pixels.

From Section II-C, we can get daily clear-sky and non-clear-sky masks for all pixels. Given T_n mask imageries, where $T_n = 8$ if generating eight-day composites, we first calculate spatially connected clear-sky pixels for each imagery by the seed filling algorithm, shown as in Fig. 8, where 1 means clear-sky mask and 0 is a cloudy mask. From T_n imageries, the imagery with the most connected clear-sky pixels is first selected, let us say T_1 in Fig. 9, and the clear-sky pixels with value 1 and blue background are chosen. The remaining nonclear-sky pixels are marked, denoted as grey areas with red boundaries in T_1. Then the imagery T_k is selected, which has the most connected clear-sky pixels in the marked pixels, and the connected clear-sky pixels with blue backgrounds are chosen. Again, we mark the remaining pixels, shown in the third subfigure in Fig. 9. And T_m is found with the most connected clear-sky pixels for the remaining pixels. Repeat the above-mentioned steps, until all pixels are filled by clear-sky pixels.

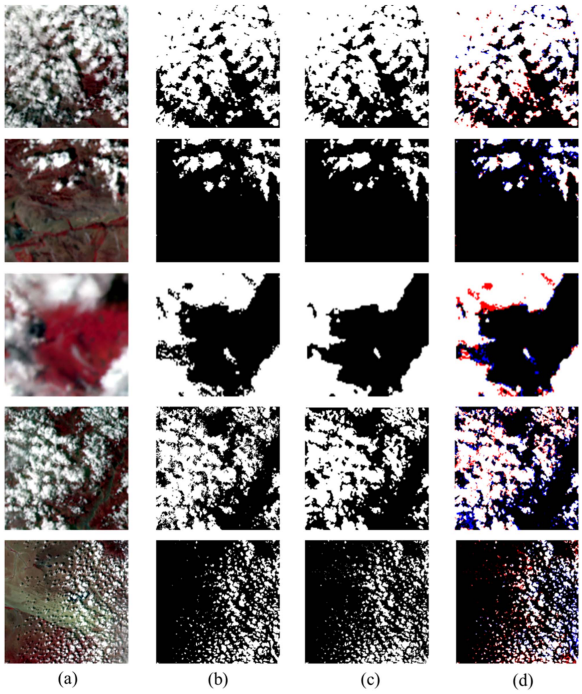

Fig. 10. Clear sky ID mask comparisons. Left to right. (a) Input image. (b) Ground truth where black means clear sky ID mask and white means nonclear sky ID mask. (c) BERT results where black means clear sky ID mask and white means nonclear sky ID mask. (d) Comparisons of (b) and (c), white, red, blue, and black mean the TP, FP, FN, and TN, respectively.

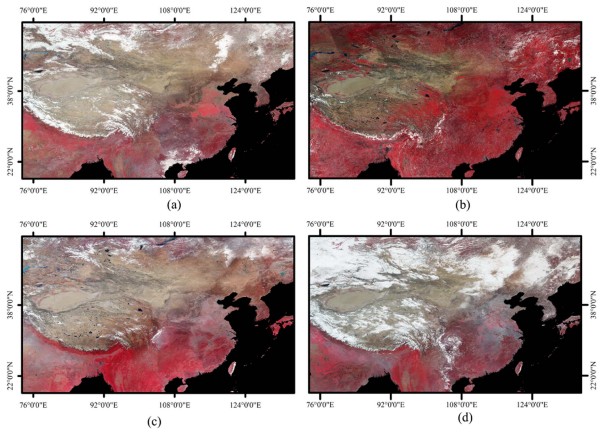

Fig. 11. Example eight-day MERSI-II composites over China using the proposed method for 2020. DOY: day of the year. (a) DOY 073-080. (b) DOY 241-248. (c) DOY 313-320. (d) 025-032.

E. Quality Assessment

Our eight-day composites were visually compared with the MODIS Collection 68-day composites with a spatial resolution of 500 m (MYD09A1). MYD09A1 was produced using the best observation method [23]. Even though MODIS has eight-day composites with a spatial resolution of 250 m (MYD09Q1), only NIR and red bands are involved. Therefore, we used the NDVI derived from MYD09Q1 for spectral indices comparison. In addition to visual assessment, we also examined the annual temporal profiles of selected samples to determine whether the desired observations were selected in each compositing period.

TABLE III
MISS RATE, FALSE RATE, AND PRECISION OF OUR MODEL FOR CLEAR SKY ID MASK CLASSIFICATION OVER THE OVERALL VALIDATION DATASET, THE DATASET ON FEB. 1, 2020, APR. 1, 2020, AND JUN. 1, 2020

	Miss rate (%) FN / (FN + TP)	False rate (%) FP / (FP + TN)	Precision (%) TP / (TP+FP)
Overall	**1.34**	**5.14**	**97.30**
20 200 201	3.40	6.21	98.06
20 200 401	0.04	7.54	96.71
20 200 601	0.20	3.44	98.49

20200201: Feb. 1st, 2020, 20200401: Apr. 1st, 2020, 20200601: Jun. 1st, 2020

III. RESULTS AND ANALYSIS

A. Clear Sky Detection Results Analysis of the ST-Det Method

As mentioned in Section II-C, even though our model is a four-classes classification model, we focus on the clear sky ID mask for clear-sky compositing. To access the clear-sky detection accuracy of the ST-Det method, we calculate the confusion matrix of the clear sky category including true positive (TP), false positive (FP), false negative (FN), and true negative (TN) based on the validation dataset. Table III shows the miss rate, false rate, and precision of the overall validation dataset, the dataset on Feb. 1, 2020, Apr. 1, 2020, and Jun. 1, 2020. Miss rate provides insight into the FN, thereby measuring how many of the actual clear sky data were missed. The false rate provides insight into the FP, thereby measuring how many of the non-clear sky data were incorrectly classified as clear sky. Precision indicates the proportion of classified truly clear sky pixels in real clear sky pixels. From the table, we see that the overall precision of clear sky identification of our model exceeds 97%, the miss rate is within 4%, and the false rate is no more than 8%. Fig. 10 shows the visual assessment of the clear sky results. It can be seen from the last column that incorrect classification usually occurs at cloud boundaries.

B. Overall Quality of Derived Composites Over China

After the clear sky ID mask is selected from the eight-day reflectance data for all pixels, a clear-sky composite is obtained using our compositing strategy. Fig. 11 shows the results of eight-day MERSI-II composites over China in four seasons, using the standard false-color scheme, where R is from the NIR band, G is from the Red band, and B is from the green band. Fig. 11(a) started from March 14th in spring, Fig. 11(b) started from August 28th in summer, Fig. 11(c) started from November 8th in autumn, and Fig. 11(d) started from January 25th in winter. Overall, the proposed method is effective, with only a small proportion of cloud-contaminated pixels. The geographical distribution and phenological characteristics of vegetation indicate that differences are related to the growing seasons of vegetation. The distribution of snow, mainly in high-latitude and high-altitude areas in spring and winter, is evident. The composites are generally good for bare land and sparse vegetation regions, such as the Turpan Basin in Xinjiang, the Mongolian Plateau, and the Qinghai-Tibet Plateau, where the aerosol optical thickness is typically low, and hence, the composites are clear. However,

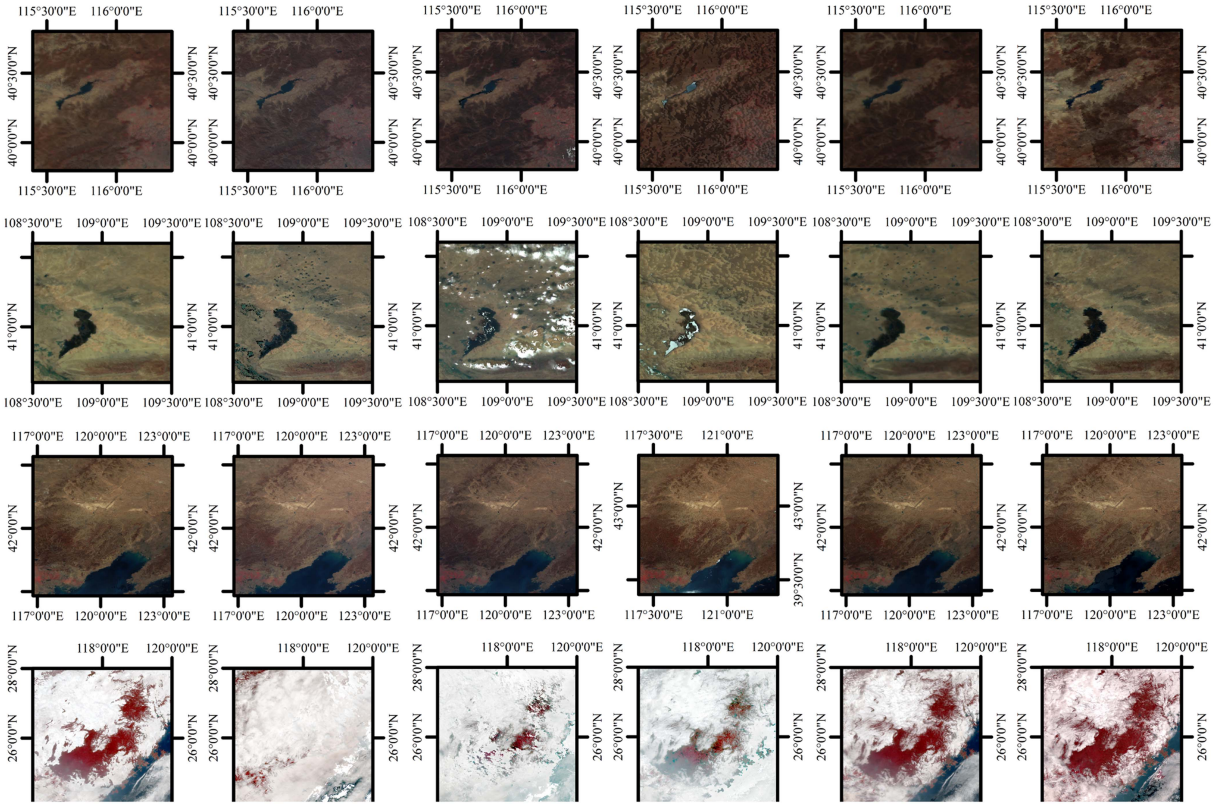

Fig. 12. MERSI-II eight-day composites comparison. (a) Proposed method. (b) CCRS-BH. (c) CCRS-FH. (d) MaxNDVI. (e) MinRed. (f) MODIS MYD09A1.

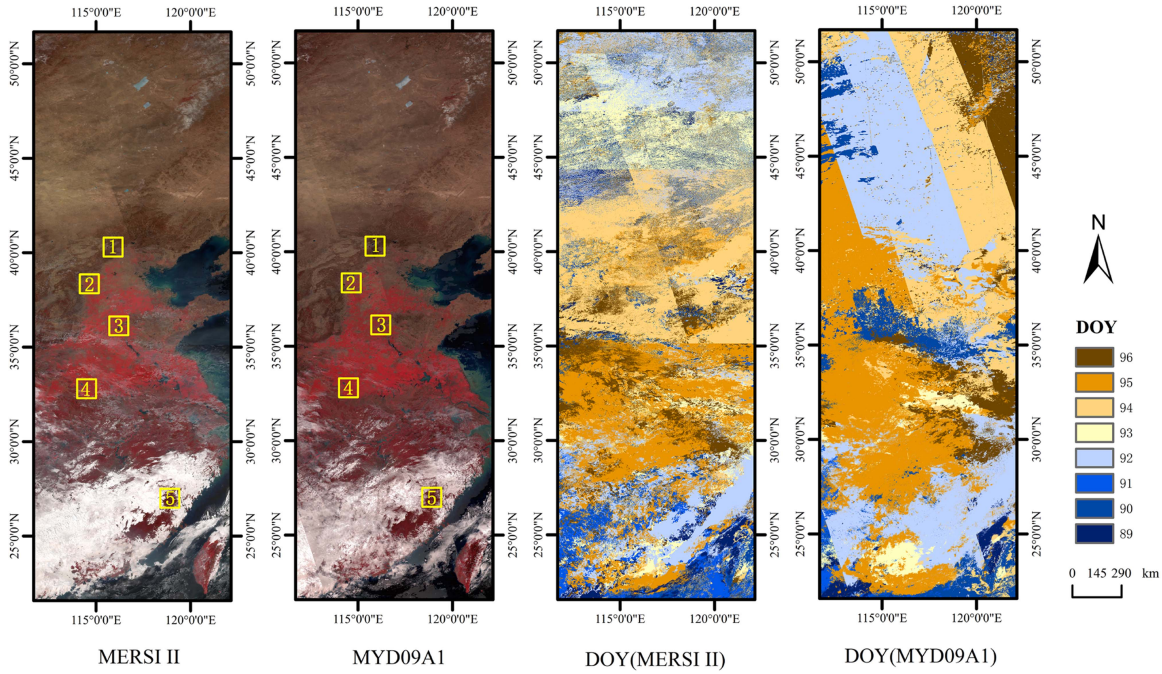

Fig. 13. Comparison between our MERSI II results and the MODIS MYD09A1. Left to right: Our proposed ST-Det method, MODIS MYD09A1 product, the DOY of DT-Det, and the DOY of MODIS MYD09A1.

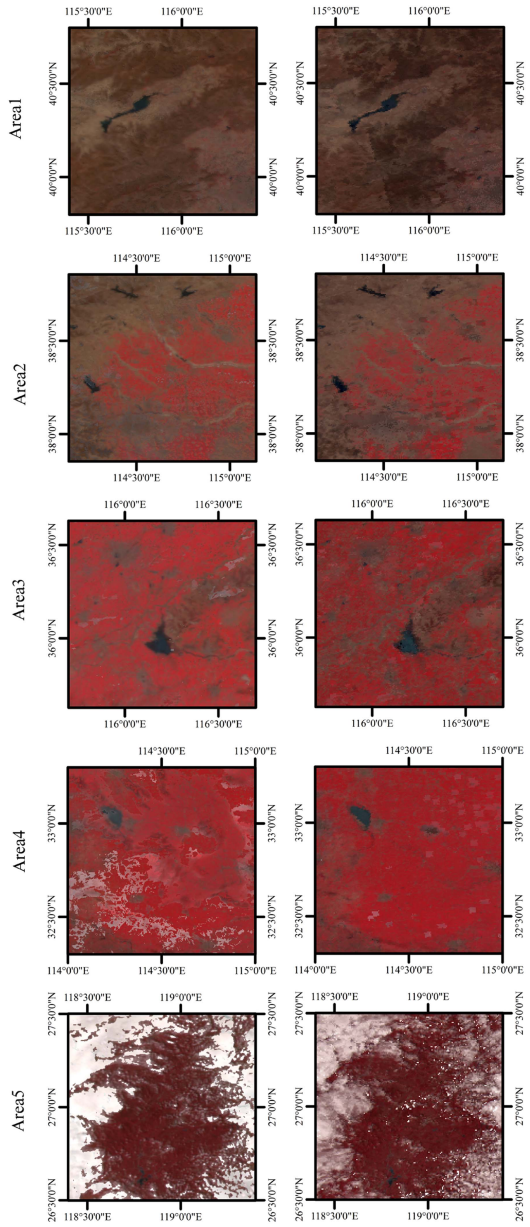

Fig. 14. Enlarged views of the results from our method (left column), and MODIS MYD09A1 (right column).

there is a small proportion of clouds in all four composites in the Sichuan Basin. In the spring composite in Fig. 11(a), a patch of cloud exists in the Beibuwan area in southern China, as well as in the northeast China of the summer composite in Fig. 11(b). We checked the original eight days' image at these sites and found that most areas of these sites have been cloud-covered for all eight days. In summary, the proposed method can effectively composite clear-sky images of various land types, which is crucial for classification and biophysical parameter retrieval.

C. Comparisons With Other Compositing Methods

Since both the MODIS Aqua satellite and the FY3D satellite are afternoon satellites, the MODIS Aqua eight-day clear-sky product MYD09A1 was selected for comparison with our FY3D MERSI-II eight-day composites, as well as other typical clear-sky compositing methods. Fig. 12 shows the enlarged details of six different methods, namely the proposed method, CCRS-BH, CCRS-FH, MaxNDVI, MinRed, and MODIS MYD09A1, including water bodies, bare land, and vegetation. These images are all displayed using the standard false-color scheme and the same stretch parameters.

It can be seen from the second row of Fig. 12 that the CCRS-FH performed the worst with a large number of residual clouds, whereas the CCRS-BH has a significant proportion of shadows. The MaxNDVI method can effectively remove clouds from vegetation land cover types, but cannot remove cloud contamination from water bodies and performs poorly on bare land. The MinRed method can effectively remove cloud contamination, but it cannot remove cloud shadows due to their low reflectance. The proposed method not only considers the spectral characteristics of different land surface cover types, clouds, and cloud shadows but also takes advantage of the temporal features in the eight-day time series within each band and among the four bands. The results show that our method performed much better than the MinRed method and the MaxNDVI method in removing clouds and cloud shadows for different land surface cover types, which means the proposed method significantly outperforms the existing single criterion methods in cloud/cloud shadow discrimination. Overall, our composites are similar to the MYD09A1 product, which also proves the effectiveness of our method.

D. Comparison With MODIS Composites

1) Visual Effects Comparison: A visual comparison was conducted over the eastern region of China from March 29 to April 6, 2020, as shown in Fig. 13. The same stretch parameters were used for both images. This region covers rich land cover types from south to north, with the southeastern part being heavily affected by clouds. In Fig. 13, we can see that the composite obtained by our proposed method was visually consistent with the MYD09A1 product. The subtropical range (∼35°N–60°N) was similar between the two, with few residual clouds and cloud shadows. Both products contain residual clouds in the latitude range of 20°N–35°N because it is challenging to distinguish clear sky pixels from the mostly cloud coverage within the eight-day composite period. From the right-most DOY (MYD09A1), compared to the DOY (MERSI II), there were more obvious swath boundaries. The composting results also present similar patterns. This difference was mostly due to the angle parameters and data coverage criteria used during compositing.

Fig. 14 illustrates the five areas within the yellow rectangular box in Fig. 13. Our 250-m resolution product has clearer lake boundaries compared to the 500-m resolution MYD09A1 product in Area 1 to Area 3. The MYD09A1 projection is initially in SIN [9] and our MERSI-II products use latitude/longitude directly. Different map projections may also introduce different distortions of lake boundaries. Following a visual inspection, our method gained similar results with MYD09A1 for Area 1–Area 3, and outperformed the MYD09A1 in Area 5 with less cloud, whereas in Area 4, our method remains some residual

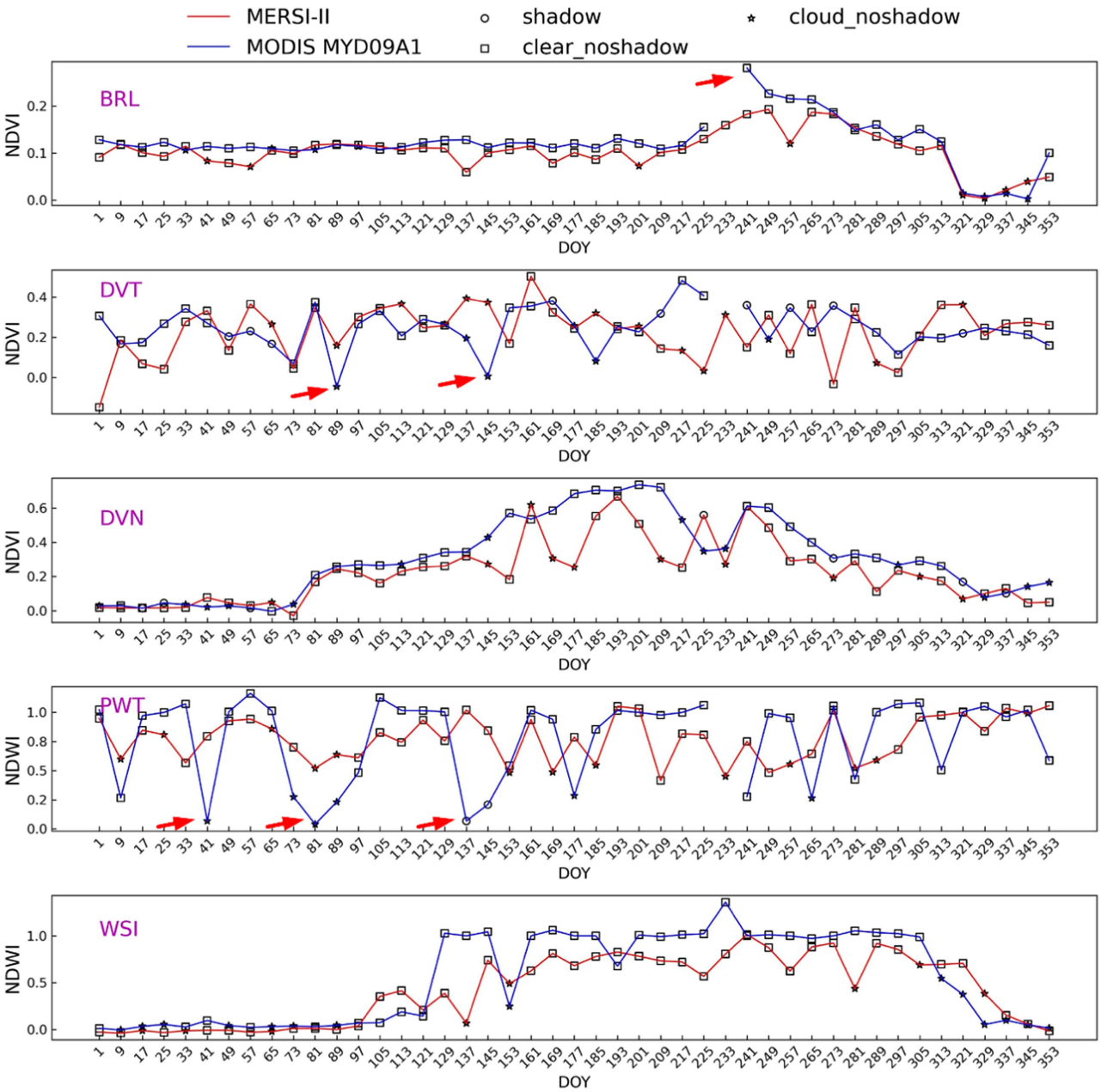

Fig. 15. Comparison of the temporal NDVI and NDWI for the five typical land cover types between our MERSI-II eight-day compositing product and MODIS MYD09A1 eight-day compositing product for the year 2020.

brighter aerosol/thin cloud pixels. In general, our time-series-based compositing method, which considers both temporal features and spatial connectivity, without the aid of auxiliary data, achieved a level of performance comparable to the MODIS eight-day clear-sky product, and in some cases, outperformed it.

In Region 5, which belongs to densely cloud-covered areas of southeast China, it can be seen that our method has fewer residual clouds and better spatial consistency. The MYD09A1 product has more isolated small cloud pixels compared to our method. This is because the MYD09A1 product uses the MODIS cloud mask product, which propagates cloud detection errors into the eight-day composite image. In contrast, our ST-Det method does not rely on auxiliary data such as cloud masks. Instead, We use time-based information to assess the state of the clouds, resulting in fewer residual clouds and better composite quality. On the other hand, our proposed composite criterion, utilizes the maximum connected area from the same day's imagery, thus taking into consideration the spatial distribution of the composite image.

2) Spectral Indices Comparison: In order to show the correctness of our compositing method in each eight-day period, we also extracted the spectral indices and compared the annual profiles of the five typical land surface cover types, namely,

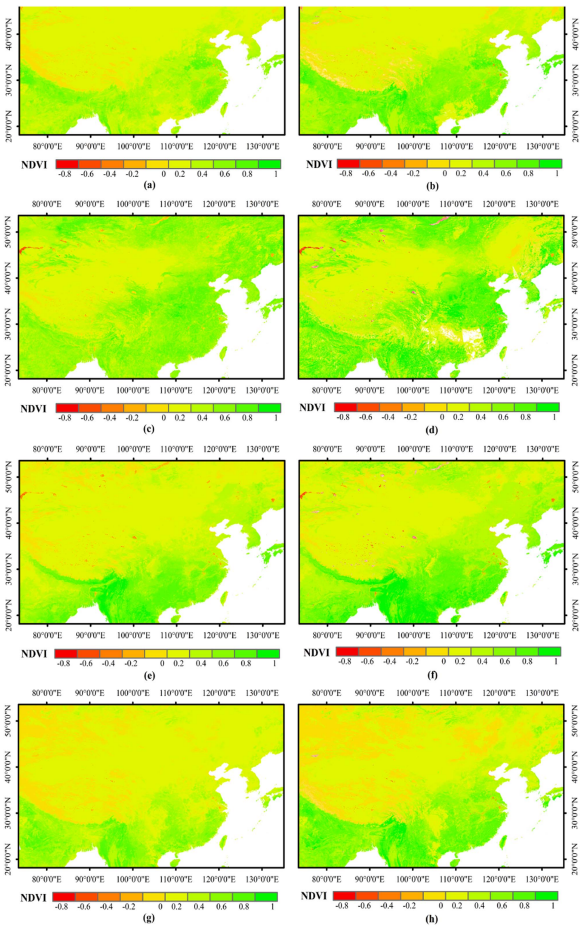

Fig. 16. Comparison of seasonal NDVI derived from our MERSI-II eight-day compositing product (left column) and MODIS MYD09Q1 product (right column) of the year 2020. Spring (DOY 73–80), summer (DOY 241–248), autumn (DOY 313–320), and winter (DOY 25–32) in row order.

BRL, DVT, DVN, PWT, and WSI. For BRL, DVT, and DVN, the NDVI is calculated from the equations provided in [27], and for PWT and WSI, the NDWI is calculated from [28]. The annual temporal NDVI and NDWI are shown in Fig. 15.

Overall, our composites and the MYD09A1 product show consistent spectral characteristics and temporal trends for different land cover types. Our method produces more stable and consistent time series curves, whereas MYD09A1 has more abrupt changes in its time series curves (red arrows in Fig. 15), indicating that those days were cloudy or in shadow. In addition, the MYD09A1 product has more missing data, while the MERSI II clear-sky compositing product has better temporal continuity.

For the BRL, both compositing products accurately capture the spectral characteristics and variations with small fluctuations in reflectance between 0 and 0.2. For the DVT sample, which is evergreen vegetation in the tropical area, the NDVI does not show seasonal change and it is larger than 0.2 on most days. However, for the DVN deciduous vegetation sample in the northern area, the NDVI values exhibit clear seasonal variations in both products, large in summer versus small in winter. Seasonal variations also can be found in the NDWI of WSI, which is ice in winter and water in summer. For the PWT sample, which is permanent water for the whole year in the tropical area, the NDWI should be close to 1. As the sample is located in low-latitude areas with high cloud cover probability, both compositing products exhibit a large fluctuation amplitude. The MYD09A1 product has sudden drops in its time series curves, denoted with red arrows, indicating incorrect pixels were selected. In contrast, our method successfully retains the water spectrum.

Fig. 16 shows seasonal NDVI derived from our MERSI-II eight-day composites (left column) versus NDVI derived from the MODIS MYD09Q1 product (right column) of the year 2020. The DOY of the first row is 73-80 starting from March 14th, the DOY of the second row is 241–248 starting from August 28th, the DOY of the third row is 313–320 starting from November 8th, and the DOY of the fourth row is 25–32 starting from January 25th. Overall, the distribution of the NDVI is consistent with each other in four seasonal eight-day products, with a slightly smaller value (less green) of our product. Our product exhibits better spatial continuity, such as the summer DOY in Fig. 16(c) and (d). In Fig. 16(d), the NDVI derived from MYD09Q1 has some missing data in southeast China. In northeast China, the majority part of the DNVI is close to zero, indicating that it is covered by cloud, as is confirmed by the inspection of the original MYD09Q1 reflectance product. There is no missing data in the southeast part of our product. And in north-east China, the NDVI has the right distribution for the summer vegetation. Thus further indicating that our product could ensure the maximum availability of clear-sky compositing data.

In the summer DOY composition product, the DNVI of MODIS in north-east China is cover by cloud whereas the NDVI value is moral in our product.

IV. Conclusion

In this article, with the help of our finding of the three-stage change pattern in the ascending sorted temporal reflectance data sequence, we proposed an approach to generate a 250-m clear-sky composite for MERSI-II land bands over China. The approach includes a ST-Det of pixel-level scene ID masks based on the BERT model and a compositing strategy that takes both spatial connectivity and temporal consistency of clear-sky pixels into consideration. Our pretrained BERT model could successfully handle different land cover types with a small amount of labels. Compositing result examples of the eight-day composites of all four seasons in 2020 were presented. The performances were evaluated and verified by comparing the composites obtained from our approach with those from four typical methods (CCRS-BH, CCRS-FH, MaxNDVI, and MinRed), as well as with the MODIS MYD09A1 products, in terms of visual assessments and annual temporal indices profiles of NDVI and NDWI, which showed consistent temporal trends for different land cover types between our method and MYD09A1. Evaluation results demonstrated that our approach performed better than the typical methods and could achieve a comparable product level to MODIS composited products. In certain cases, it even performed better than MYD09A1, with fewer residual clouds and cloud shadows. Moreover, our products were at 250-m

spatial resolution while the MYD09A1 were at 500-m spatial resolution, therefore, our clear-sky composites owned clearer ground boundaries.

During the data preprocessing stage, the raw MERSI-II data does not have the atmospheric aerosol removed. As a result, some areas with sparse aerosol coverage were treated as clear-sky pixels in the compositing results. When using the same stretching parameters as MODIS MYD09A1, these areas appear brighter. Our future research will focus on the aerosols/BRDF effects removal from raw land band data, with the help of more sophisticated atmospheric correction algorithms, to generate clear-sky composites, as well as improve the robustness of the proposed algorithm when applied to other regions of the world.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their many valuable comments and suggestions that helped improve the manuscript.

REFERENCES

[1] Y. Wang and G. Li, "Global land cover mapping using annual clear-sky composites from FY3D/MERSI-II," *Int. J. Remote Sens.*, vol. 43, no. 2, pp. 510–531, 2022.

[2] N. Xu et al., "Prelaunch calibration and radiometric performance of the advanced MERSI II on Fengyun-3D," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4866–4875, Aug. 2018.

[3] J. He and Z. Liu, "Water vapor retrieval from MERSI NIR channels of Fengyun-3B satellite using ground-based GPS data," *Remote Sens. Environ.*, vol. 258, 2021, Art. no. 112384.

[4] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, pp. 247–259, 2019.

[5] J. Bian, A. Li, Q. Wang, and C. Huang, "Development of dense time series 30-m image products from the Chinese HJ-1A/B constellation: A case study in Zoige plateau, China," *Remote Sens.*, vol. 7, no. 12, pp. 16647–16671, 2015.

[6] P. Griffiths, S. van der Linden, T. Kuemmerle, and P. Hostert, "A pixel-based Landsat compositing algorithm for large area land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2088–2101, Oct. 2013.

[7] D. P. Roy et al., "Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 35–49, 2010.

[8] B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens.*, vol. 7, no. 11, pp. 1417–1434, 1986.

[9] Y. Luo, A. P. Trishchenko, and K. V. Khlopenkov, "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sens. Environ.*, vol. 112, no. 12, pp. 4167–4185, 2008.

[10] E. Lindsay et al., "Multi-temporal satellite image composites in Google earth engine for improved landslide visibility: A case study of a glacial landscape," *Remote Sens.*, vol. 14, no. 10, 2022, Art. no. 2301.

[11] X. Fan, Y. Liu, G. Wu, and X. Zhao, "Compositing the minimum NDVI for daily water surface mapping," *Remote Sens.*, vol. 12, no. 4, 2020, Art. no. 700.

[12] E. Chuvieco, G. Ventura, M. P. Martín, and I. Gómez, "Assessment of multitemporal compositing techniques of MODIS and AVHRR images for burned land mapping," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 450–462, 2005.

[13] A. P. Trishchenko and Y. Luo, "Landfast ice mapping using MODIS clear-sky composites: Application for the Banks Island coastline in Beaufort Sea and comparison with Canadian ice service data," *Can. J. Remote Sens.*, vol. 47, no. 1, pp. 143–158, 2021.

[14] A. Cabral, M. J. De Vasconcelos, J. M. Pereira, É. Bartholomé, and P. Mayaux, "Multi-temporal compositing approaches for SPOT-4 VEGETATION," *Int. J. Remote Sens.*, vol. 24, no. 16, pp. 3343–3350, 2003.

[15] D. Frantz, A. Röder, M. Stellmes, and J. Hill, "Phenology-adaptive pixel-based compositing using optical earth observation imagery," *Remote Sens. Environ.*, vol. 190, pp. 331–347, 2017.

[16] M. J. Wilson and L. Oreopoulos, "Enhancing a simple MODIS cloud mask algorithm for the Landsat data continuity mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 723–731, Feb. 2013.

[17] L. Oreopoulos, M. J. Wilson, and T. Várnai, "Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 597–601, Jul. 2011.

[18] J. Li and S. Wang, "A new method for generating a clear-sky Landsat composite for cropland from cloud-contaminated Landsat-7 and Landsat-8 images," *Int. J. Digit. Earth*, vol. 11, no. 5, pp. 533–545, 2018.

[19] J. Bian, A. Li, C. Huang, R. Zhang, and X. Zhan, "A self-adaptive approach for producing clear-sky composites from VIIRS surface reflectance datasets," *ISPRS J. Photogrammetry Remote Sens.*, vol. 144, pp. 189–201, 2018.

[20] J. Li et al., "Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 373–389, 2020.

[21] X. Wang, H. Xie, and T. Liang, "Evaluation of MODIS snow cover and cloud mask and its application in Northern Xinjiang, China," *Remote Sens. Environ.*, vol. 112, no. 4, pp. 1497–1513, 2008.

[22] X. Wang, M. Min, F. Wang, J. Guo, B. Li, and S. Tang, "Intercomparisons of cloud mask products among Fengyun-4A, Himawari-8, and MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8827–8839, Nov. 2019.

[23] E. F. Vermote, J. C. Roger, and J. P. Ray, "MODIS Surface Reflectance User's Guide Collection 6," 2015.

[24] Goddard-Space-Flight-Center, *MODIS Level 2 Corrected Reflectance Science Processing Algorithm (CREFL_SPA) User's Guide; Version 1.7.1*, 2010. [Online]. Available: http://directreadout.sci.gsfc.nasa.gov/?id=software

[25] Z. Jing, S. Li, X. Hu, and F. Tang, "Sub-pixel accuracy evaluation of FY-3D MERSI-2 geolocation based on OLI reference imagery," *Int. J. Remote Sens.*, vol. 42, no. 19, pp. 7215–7238, 2021.

[26] J. Devlin, M. - W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training OF deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[27] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.

[28] S. K. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.

[29] F. Baret et al., "Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: Proposition of the CEOS-BELMANIP," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1794–1803, Jul. 2006.

Wenzhuo Li received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011, 2013, and 2017, respectively.

He is a Postdoctoral Researcher with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is currently a Researcher with the School of Information Science and Engineering, Wuhan University of Science and Technology. His current research interests include image processing, change detection, and ecological remote sensing monitoring.

Kaimin Sun (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1999, 2004, and 2008, respectively.

He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. His research interests include photogrammetry, applications of remote sensing in ecology and surface disaster, and time series analysis of remote sensing images.

Hongjuan Zhang received the B.S. and M.S. degrees in geodesy and survey engineering from Wuhan University, Wuhan, China, in 2011 and 2013, respectively, and the Ph.D. degree in science from RWTH Aachen University, Aachen, Germany, in 2018.

She is currently an Associate Researcher with State Key Laboratory of Information Engineering, Surveying, Mapping and Remote Sensing, Wuhan University. Her current research interests include remote sensing and multisource fusion navigation.

Shunxia Miao (Student Member, IEEE) received the B.S. degree in spatial information and digital technology from China University of Geosciences, Wuhan, China, in 2020, the M.S. degree in resource and environmental science from Wuhan University, Wuhan, in 2022, where she is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the School of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

Her research interests include the calibration of FengYun satellite instruments, applications of remote sensing in ecology and surface disaster, and image processing.

Fangyi Lv received the B.E. degree in surveying and mapping engineering in 2022, from the School of Geodesy and Geomatics, Wuhan University, Wuhan, China, where she is working toward the master's degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

Her major research interests include remote sensing applications and time-series analysis.

Xiuqing Hu received the B.Sc. degree in atmospheric science from Nanjing University, Nanjing, China, in 1996, the M.Sc. degree in cartography and geographical information systems from Beijing Normal University, Beijing, China, in 2004, and the Ph.D. degree in quantitative remote sensing science from the Institute of Remote Sensing Application, Chinese Academy of Sciences, Beijing, China, in 2012.

He is currently conducting research with the National Satellite Meteorological Center, China Meteorological Administration, Beijing. His research interests include calibration and validation for optical sensors, retrieval algorithms for aerosol/dust and water vapor, and climate data records from environment satellites.

Hongya Zhang received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011 and 2018, respectively.

She is currently a Senior Engineer with Changjiang River Scientific Research Institute of Changjiang Water Resources Commission. Her research interests include remote sensing applications, flood disaster prevention and control, and watershed planning.